Circularly squarefree words and unbordered conjugates: a new approach

Trevor Clokie, Daniel Gabric, and Jeffrey Shallit

School of Computer Science
University of Waterloo
Waterloo, Ontario N2L 3G1
Canada
{trevor.clokie,dgabric,shallit}@uwaterloo.ca

Abstract. Using a new approach based on automatic sequences, logic, and a decision procedure, we reprove some old theorems about circularly squarefree words and unbordered conjugates in a new and simpler way. Furthermore, we prove two new results about unbordered conjugates: we complete the classification, due to Harju and Nowotka, of binary words with the maximum number of unbordered conjugates, and we prove that for every possible number, up to the maximum, there exists a word having that number of unbordered conjugates.

1 Introduction

Throughout this paper, Σ_k denotes the alphabet $\{0, 1, \ldots, k-1\}$.

Two finite words are said to be *conjugate* if one is a cyclic shift of the other, as in the English words `enlist` and `listen`.

A finite word w has a *border* x if $x \notin \{\epsilon, w\}$ and x is both a prefix and suffix of w; the two occurrences of x are allowed to overlap each other. For example, `alfa` is a border of `alfalfa`. A finite word w is said to be *bordered* if it has a border, and otherwise, it is *unbordered*. A finite word w if bordered iff it has a border of length $\leq |w|/2$, for if a word has a longer border y, then the nonempty overlap of the two occurrences of y — one as prefix and one as suffix — provides a shorter border. For example, `alfalfa` is also bordered by `a`.

A finite word w is said to be a *square* if $w = xx$ for some nonempty word x. An example in French is the word `couscous`. A word (finite or infinite) is *squarefree* if no nonempty factor is a square. Let μ be the *Thue-Morse morphism*, defined by $\mu(0) = 01$ and $\mu(1) = 10$. The Thue-Morse word $\mathbf{t} = \texttt{01101001}\cdots$ is the infinite fixed point, starting with 0, of μ. Thue [12,13,4] proved that there exist infinite squarefree words over a three-letter alphabet; also see [2]. A famous example of such a word can be obtained from the Thue-Morse word as follows: count the number of 1's between two consecutive 0's in $\mathbf{t}$. This gives the so-called *ternary Thue-Morse word*

$$\mathbf{c} = 210201 \cdots ,$$

and is squarefree. An alternative description of $\mathbf{c}$ is as follows [3]: it is the image, under τ of the fixed point of the morphism φ defined below:

$$\begin{aligned} \varphi(0) &= 01 & \tau(0) &= 2 \\ \varphi(1) &= 20 & \tau(1) &= 1 \\ \varphi(2) &= 23 & \tau(2) &= 0 \\ \varphi(3) &= 02 & \tau(3) &= 1 \end{aligned}$$

A word w is *circularly squarefree* if every one of its conjugates is squarefree. For example, `outshout` is squarefree, but not circularly squarefree. Clearly we have

Proposition 1. *A word is circularly squarefree iff all its conjugates are unbordered.*

We now turn to a description of what we do in this paper. Using a complicated case-based argument, Currie [6] proved that there exist circularly squarefree ternary words of every length n, except for $\{5, 7, 9, 10, 14, 17\}$. The first of our main results is a new proof of Currie's theorem, based on the following result:

Theorem 1. *For all natural numbers $n > 3$, except $5, 7, 9, 10, 14, 17, 21$, and 28, there exists a factor $x = x(n)$ of the ternary Thue-Morse word $\mathbf{c}$ that is either*

(a) of length $n - 3$, and $x021$ is circularly squarefree;
(b) of length $n - 4$, and $x2120$ is circularly squarefree.

The virtues of our proof are (a) it requires very little work—just setting up the appropriate logical predicates—and (b) it gives specific examples of the desired words that are very easy to describe and compute. For a completely different approach, which has the virtue of allowing one to give a good estimate for the number of circularly squarefree words of length n, see Shur [11].

We now turn to unbordered conjugates. In two fundamental papers, Harju and Nowotka [7,8] studied the unbordered conjugates of a word. In particular, letting $\mathrm{nuc}(w)$ denote the number of unbordered conjugates of w, and $\mathrm{mnuc}_k(n)$ denote the maximum number of unbordered conjugates of a length-n word over a k-letter alphabet, they proved that

(a) for binary words w of length $n \geq 4$ we have $\mathrm{nuc}(w) \leq n/2$;
(b) for $n > 2$ even, there exists a binary word of length n having $n/2$ unbordered conjugates iff $n = 2^k$ or $n = 3 \cdot 2^k$ for some $k \geq 1$.

In other words, they explicitly computed $\mathrm{mnuc}_2(n)$ for all even n and bounded it above for odd n. We complete the understanding of $\mathrm{mnuc}_2(n)$ by proving that $\mathrm{mnuc}_2(n) = \lfloor n/2 \rfloor$ for all odd $n > 3$. Our strategy is to show that the maximum of $\mathrm{nuc}(w)$, over all words of length n, is actually achieved by a factor of the Thue-Morse word.

More precisely, we prove

Theorem 2. *For all $n \geq 1$, there exists a length-n factor w of the Thue-Morse word $\mathbf{t}$ with $\mathrm{nuc}(w) = \mathrm{mnuc}_2(n)$. Furthermore, such a factor is guaranteed to occur starting at a position $\leq n$ in $\mathbf{t}$.*

2 Circularly squarefree ternary words via `Walnut`

Since the ternary Thue-Morse word **c** is squarefree, it is reasonable to hope its factors might be a good source of circularly squarefree words. Unfortunately, **c** contains circularly squarefree words of length n for only about 1/8 of all natural numbers n, as the following result shows.

Theorem 3. *There is a length-n factor of* **c** *that is circularly squarefree iff* $(n)_2$ *is accepted by the automaton in Figure 1.*

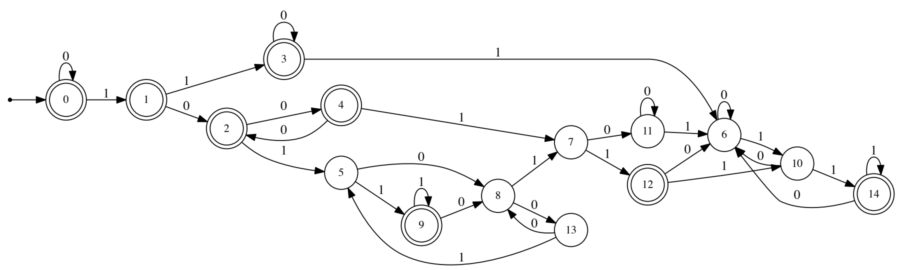

Fig. 1. Automaton accepting lengths $(n)_2$ of circularly squarefree words occurring in **c**

To prove this result, we make use of the fact that many first-order statements concerning claims about k-automatic sequences are decidable [5]. Furthermore, there is free software called `Walnut` available to decide these claims [9].

Let $(n)_k$ denote the canonical base-k representation of n, starting with the most significant digit, having no leading zeroes. A sequence $(a_n)_{n\geq 0}$ is *k-automatic* if there is a deterministic finite automaton with output (DFAO) taking $(n)_k$ as input, and reaching a state with a_n as output. For example, Figure 2 illustrates the DFAO generating the sequence **c**. The notation q/a in a state means the name of the state is q and the output is a. For more about automatic sequences, see [1].

Proof. We can use the ideas in [10], adapted for our case. We create first-order logical predicates crep, facge2, and circsf as follows:

- crep(i, m, p, n, s) evaluates to `true` iff in the length-n word (considered circularly) starting at position s of the word **c**, there is a factor w of length m and (not necessarily least) period $p \geq 1$ starting at position i;
- facge2(n, s) evaluates to `true` iff in the length-n word (considered circularly) starting at position s of the word **c** there is a square or higher power;

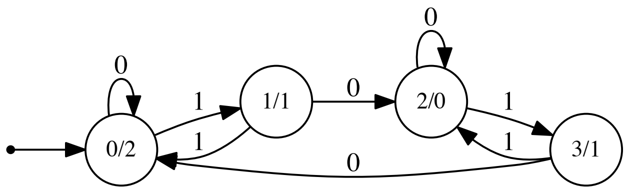

Fig. 2. DFAO computing the sequence **c**

– circsf(n) evaluates to `true` iff some length-n factor (considered circularly) of the word **c** has no squares.

$$\begin{aligned}
&\mathrm{crep}(i,m,n,p,s) := \exists j\, ((j \geq i) \wedge (j+p < s+n) \wedge (j+p < i+m)) \implies \mathbf{c}[j] = \mathbf{c}[j+p]) \wedge \\
&(\forall j\, ((j \geq i) \wedge (j < s+n) \wedge (j+p \geq s+n) \wedge (j+p < i+m)) \implies \mathbf{c}[j] = \mathbf{c}[j+p-n]) \wedge \\
&(\forall j\, ((j \geq i) \wedge (j \geq s+n) \wedge (j+p < i+m)) \implies \mathbf{c}[j-n] = \mathbf{c}[j+p-n])
\end{aligned}$$

$$\mathrm{facge2}(n,s) := \exists i,m,p\ (p \geq 1) \wedge (m \leq n) \wedge (i \geq s) \wedge (i < s+n) \wedge (m \geq 2p) \wedge \mathrm{crep}(i,m,n,p,s)$$

$$\mathrm{circsf}(n) := \exists s\, \neg\, \mathrm{facge2}(n,s)$$

When we evaluate these predicates in `Walnut`, we get the automaton depicted in Figure 1. It accepts those $(n)_2$ for which `circsf` evaluates to `true`.

Remark 1. All the `Walnut` code for the theorems in this paper is available at

https://cs.uwaterloo.ca/~shallit/papers.html .

The reader can therefore verify our results.

Corollary 1. *The number of lengths ℓ, with $2^n \leq \ell < 2^{n+1}$ and $n \geq 4$, such that **c** contains a circularly squarefree factor of length ℓ, is $2^{n-3} - F_{n-3} + 2$, where F_n is the n'th Fibonacci number.*

Proof. By standard techniques, by determining the roots of the characteristic polynomial of the 15×15 matrix encoding transitions of the automaton in Fig. 1.

So while the factors of the ternary Thue-Morse word alone do not suffice for our purpose, it turns out that a small modification of them do. We now give the proof of our first main result, Theorem 1.

Proof. (of Theorem 1) Let $n \geq 4$ and $w \in \{x021, y2120\}$, where x, y are factors of the ternary Thue-Morse word **c** of lengths $n-3$ and $n-4$, respectively.

First, we create a predicate sq021(i,n,p,s) which evaluates to `true` if $w' := x021x02$ contains a square of order p with $p \geq 1$ and $2p \leq n$ beginning at index

$i - s$, where $x = \mathbf{c}[s..s + n - 4]$. We do this by defining $w[j]$ for all j such that $i \le j < i + p$ as follows:

$$w[j] = \begin{cases} \mathbf{c}[j], & \text{if } j < s + n - 3; \\ 0, & \text{if } j \in \{s + n - 3,\ s + 2n - 3\}; \\ 2, & \text{if } j \in \{s + n - 2,\ s + 2n - 2\}; \\ 1, & \text{if } j = s + n - 1; \\ \mathbf{c}[j - n], & \text{if } s + n \le j < s + 2n - 3. \end{cases}$$

The goal is that sq021 should represent the implication

$$\forall j\, ((i \le j) \wedge (j < i + p)) \implies w[j] = w[j + p].$$

It is formed by constructing the conjunction of the predicates

$$\forall j\, ((i \le j) \wedge (j < i + p) \wedge (w[j] = \alpha) \wedge (w[j + p] = \beta)) \implies \alpha = \beta$$

for each possible combination j and $j + p$, and simplifying.

Next, we create a second predicate $\mathrm{sqfree021}(n, s)$, which evaluates to `true` if there exists x where $w = x021$ is circularly squarefree, for the given values of n and s:

$$\begin{aligned} \mathrm{sqfree021}(i, n, p, s) := (n > 3) \wedge (\forall i, p\, ((1 \le p) \wedge (2p \le n) \wedge (s \le i) \wedge (i < s + n)) \\ \implies \neg(\mathrm{sq021}(i, n, p, s))). \end{aligned}$$

Similarly, we create the analogous predicates $\mathrm{sq2120}(i, n, p, s)$ and $\mathrm{sqfree2120}(n, s)$ for the word $w' := y2120y212$.

Finally, the predicates

$$\begin{aligned} \mathrm{test021}(n) &:= \exists s\ \mathrm{sqfree021}(n, s) \\ \mathrm{test2120}(n) &:= \exists s\ \mathrm{sqfree2120}(n, s) \end{aligned}$$

return `true` if there exists a length-n squarefree word formed by concatenating some factor of $\mathbf{c}$ with 021 (respectively, 2120). The automaton for $\mathrm{test021}(n)$ is depicted in Figure 3; the automaton for $\mathrm{test2120}(n)$ is omitted for space considerations.

When we now evaluate the predicate

$$\mathrm{currie}(n) := \mathrm{test021}(n) \vee \mathrm{test2120}(n)$$

with `Walnut`, we get the automaton depicted in Figure 4.

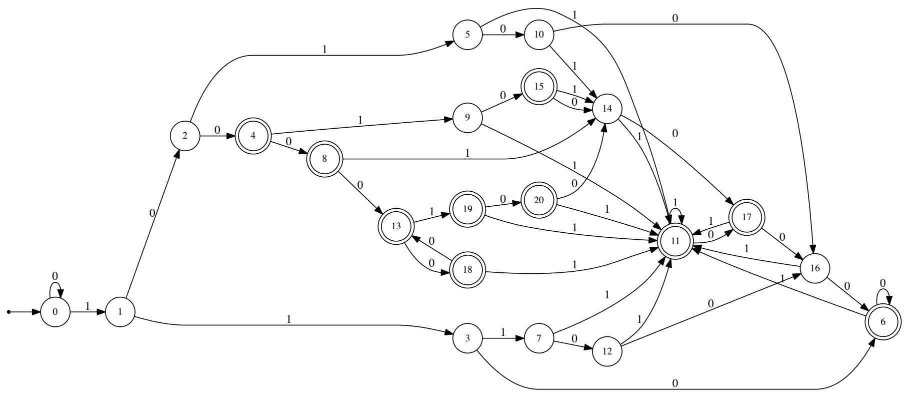

Fig. 3. DFA computing $\exists s$ sqfree021(n, s)

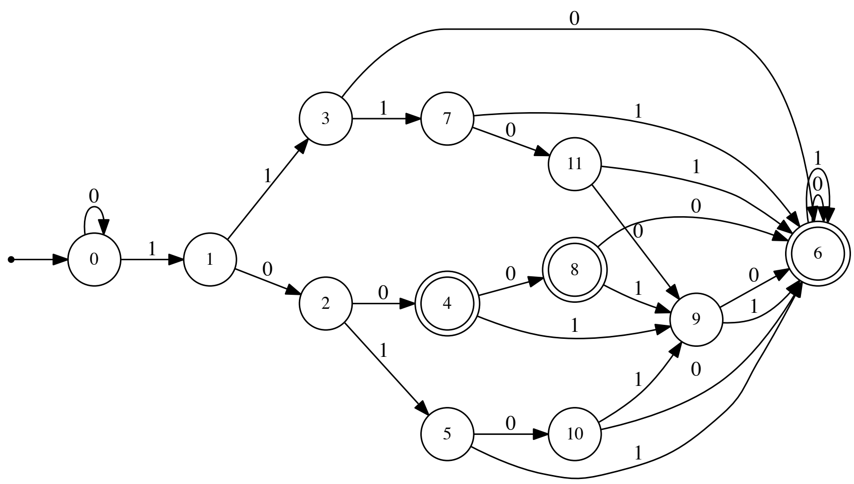

Fig. 4. DFA computing acceptable n

By inspection we easily see that the automaton in Figure 4 accepts the base-2 representation of all n except $0, 1, 2, 3, 5, 7, 9, 10, 14, 17, 21, 28$.

As a consequence we now get Currie's theorem:

Corollary 2. *There exist circularly squarefree ternary words of every length n, except for $n \in \{5, 7, 9, 10, 14, 17\}$.*

Proof. Theorem 1 gives the result for all but finitely many n. It is easy to verify by a short computation that there are cyclically squarefree words of lengths $0, 1, 2, 3, 21, 28$, and none for lengths $5, 7, 9, 10, 14, 17$.

Remark 2. These calculations were done in `Walnut` on a Linux machine (2 CPU — Intel E5-2697 v3 Xeon, 256 GB of RAM). Computing the automaton for sq021 took 115.505 seconds, and the automaton for sq2120 took 124.908 seconds.

3 Unbordered conjugates

Let $\sigma : \Sigma_k^* \to \Sigma_k^*$ denote the *cyclic shift function*, where $\sigma(\epsilon) = \epsilon$, $\sigma(cw) = wc$ for $w \in \Sigma_k^*$ and $c \in \Sigma_k$. Let $\sigma^0(w) = w$ and $\sigma^i(w) = \sigma^{i-1}(\sigma(w))$ for $i \geq 1$.

Suppose w is a binary word of length n. Let $\beta : \Sigma_k^* \to \Sigma_k^*$ be the *border correlation function* of a word (introduced by Harju and Nowotka [7]), and defined as follows: $\beta(w) = a_0 a_1 \cdots a_{n-1}$, where

$$a_i = \begin{cases} u, & \text{if } \sigma^i(w) \text{ is unbordered;} \\ b, & \text{if } \sigma^i(w) \text{ is bordered.} \end{cases}$$

For example, $\beta(0001) = ubbu$ since 0001 is unbordered, while 0010, and 0100 are both bordered, and 1000 is unbordered. Let $u, v \in \Sigma_k^*$. We say u is the *i'th cyclic shift* of v if $\sigma^i(v) = u$.

A result from Harju and Nowotka [7] shows that a binary word has no two consecutive cyclic shifts that are unbordered. This result immediately tells us that a binary word of length n can have at most $\lfloor n/2 \rfloor$ unbordered conjugates. For a binary word w of even length to achieve this bound, every other cyclic shift must be unbordered, or, in other words either $\beta(w) = (ub)^{|w|/2}$ or $\beta(w) = (bu)^{|w|/2}$. Harju and Nowotka [7] showed that the only words of even length that achieve this bound are the circularly overlap-free words, which are of length $3 \cdot 2^i$ and 2^i for $i \geq 1$.

Let w be a binary word. Suppose w is of even length and is not circularly overlap-free. Clearly w cannot have $|w|/2$ unbordered conjugates, but it could potentially have $|w|/2-1$ unbordered conjugates. Then $\beta(w) = (ub)^i b (ub)^{|w|/2-i-1} b$ for some $i \geq 0$, up to conjugation. Now suppose w is of odd length. No circularly overlap-free words exist of odd length, so it makes sense to think that w could contain a maximum of $\lfloor |w|/2 \rfloor$ unbordered conjugates. Then $\beta(w) = (ub)^{\lfloor |w|/2 \rfloor} b$, up to conjugation.

Let w be a bordered binary word. Then $w = uvu$ for some words u and v. By the *left border* of w we mean the occurrence of u that begins at position 1 of w, and by the *right border* we mean the occurrence of u that begins at position $|w| - |u| + 1$ of w.

Now we prove Theorem 2.

Proof. When $n = 1, 2, 3$ the maximum number of unbordered conjugates $\mathrm{mnuc}_2(n)$ is achieved by the words 0, 01, and 011 respectively. Specifically we have that $\mathrm{mnuc}_2(1) = 1$, $\mathrm{mnuc}_2(2) = 2$, and $\mathrm{mnuc}_2(3) = 2$. It is readily verified that each of these words occur as a factor of the Thue-Morse word at position $\leq n$.

Let w be a length-n word at position m of the Thue-Morse word. The first step is to create a first-order predicate $\mathrm{isBorder}(l, m, n)$ that asserts that a cyclic shift of w has a border of a certain length. More specifically, we want to know whether the l'th cyclic shift of w has a border of length k. There are three cases to consider.

1. When a prefix of the right border is a suffix of w and a suffix of the right border is a prefix of w. In other words, $w = yuvx$ for words u, v, x, y where $xy = u$, $|y| = l$, and $|u| = k$. This predicate is denoted by $\mathrm{isBorderC1}(k, l, m, n)$.
2. When both borders are completely contained inside of w. In other words, $w = yuux$ for words y, u, x where $|yu| = l$, and $|u| = k$. This predicate is denoted by $\mathrm{isBorderC2}(k, l, m, n)$.
3. When a prefix of the left border is a suffix of w and a suffix of the left border is a prefix of w. In other words, $w = yvux$ for words u, v, x, y where $xy = u$, $|yvu| = l$, and $|u| = k$. This predicate is denoted by $\mathrm{isBorderC3}(k, l, m, n)$.

$$\begin{aligned}\mathrm{isBorderC1}(k, l, m, n) := {} & ((k + l > n) \Rightarrow ((\forall i(i < n - l) \Rightarrow T[m + l + i] = T[m + l - k + i]) \\ & \wedge (\forall i(i < k + l - n) \Rightarrow T[m + i] = T[m + n - k + i])))\end{aligned}$$

$$\begin{aligned}\mathrm{isBorderC2}(k, l, m, n) := {} & (((k + l \leq n) \wedge (l \geq k)) \Rightarrow (\forall i\, (i < k) \Rightarrow \\ & T[m + l + i] = T[m + l - k + i]))\end{aligned}$$

$$\begin{aligned}\mathrm{isBorderC3}(k, l, m, n) := {} & (((k + l \leq n) \wedge (l < k)) \Rightarrow ((\forall i\, (i < k - l) \Rightarrow T[m + n - k + l + i] \\ & = T[m + l + i]) \wedge (\forall i\, (i < l) \Rightarrow T[m + i] = T[m + k + i])))\end{aligned}$$

$$\mathrm{isBorder}(k, l, m, n) := \mathrm{isBorderC1}(k, l, m, n) \wedge \mathrm{isBorderC2}(k, l, m, n) \wedge \mathrm{isBorderC3}(k, l, m, n).$$

We define the predicate $\mathrm{isBordered}(l, m, n)$ that asserts that the l'th cyclic shift of a length-n word at position m in the Thue-Morse word is bordered. We can create this predicate by checking whether this word has a border of size $\leq n/2$.

$$\mathrm{isBordered}(l, m, n) := \exists i(2i \leq n \wedge i \geq 1 \wedge \mathrm{isBorder}(i, l, m, n)).$$

Recall that when $|w|$ is odd and w has a maximum number of unbordered conjugates, we have that $\beta(w) = (ub)^{\lfloor |w|/2 \rfloor} b$, up to conjugation. So we have exactly one pair of adjacent bordered cyclic shifts, and the rest of the cyclic shifts of w alternate between bordered and unbordered. The predicate $\mathrm{isAlternating0}(l, m, n)$ asserts that all of the cyclic shifts of a length-n word at position m in the Thue-Morse word alternate between unbordered and bordered, except for the l'th and $l + 1$'th cyclic shifts, which are both bordered.

$$\begin{aligned}& \mathrm{isAlternating0}(l, m, n) := \\ & \forall i(((i \neq l \wedge i < n - 1) \Rightarrow (\mathrm{isBordered}(i, m, n) = \neg\, \mathrm{isBordered}(i + 1, m, n))))\wedge \\ & (((i \neq l) \wedge (i = n - 1)) \Rightarrow (\mathrm{isBordered}(n - 1, m, n) = \neg\, \mathrm{isBordered}(0, m, n))).\end{aligned}$$

Now we create a predicate hasMNUCO(m, n) that asserts that a length-n word at position m in the Thue-Morse word achieves the maximum number of unbordered conjugates.

$$\text{hasMNUCO}(m,n) := \exists i(((i < n-1 \wedge \text{isBordered}(i,m,n) \wedge \text{isBordered}(i+1,m,n)) \vee (i = n-1 \wedge \text{isBordered}(n-1,m,n) \wedge \text{isBordered}(0,m,n))) \wedge \text{isAlternating0}(i,m,n)).$$

Similarly, recall that when $|w|$ is even and w has a maximum number of unbordered conjugates, we have that $\beta(w) = (ub)^i b(ub)^{|w|/2-i-1}b$ for some $i \geq 0$ or $\beta(w) = (ub)^{|w|/2}$, up to conjugation. So we have that either all of the cyclic shifts of w alternate between bordered and unbordered, or there are exactly two pairs of adjacent bordered cyclic shifts, and the rest of the cyclic shifts of w alternate between bordered and unbordered. The predicate

$$\text{isAlternatingE}(e, l, m, n)$$

asserts that all of the cyclic shifts of a length-n word at position m in the Thue-Morse word alternate between unbordered and bordered, except for the l'th, $l+1$'th, e'th, and $e+1$'th cyclic shifts, which are all bordered. Note that isAlternatingE(n, n, m, n) asserts that all of the cyclic shifts of a length n word at position m in the Thue-Morse word alternate between unbordered and bordered.

$$\begin{aligned}\text{isAlternatingE}(e,l,m,n) := (\forall i\,(((i \neq l \wedge i \neq e \wedge i < n-1) \Rightarrow (\text{isBordered}(i,m,n) \Leftrightarrow\\ \neg\,\text{isBordered}(i+1,m,n)))) \wedge (((i \neq l) \wedge (i \neq e) \wedge (i = n-1)) \Rightarrow\\ (\text{isBordered}(n-1,m,n) \Leftrightarrow \neg\,\text{isBordered}(0,m,n))))\end{aligned}$$

Now we create a predicate hasMNUCE(m, n) that asserts that a length-n word at position m in the Thue-Morse word achieves the maximum number of unbordered conjugates.

$$\begin{aligned}\text{hasMNUCE}(m,n) := (\exists i,j\,((i<j) \wedge (i < n-1 \wedge \text{isBordered}(i,m,n) \wedge \text{isBordered}(i+1,m,n)) \wedge\\ ((j = n-1 \wedge \text{isBordered}(n-1,m,n) \wedge \text{isBordered}(0,m,n)) \vee ((j < n-1) \wedge\\ \text{isBordered}(j,m,n) \wedge \text{isBordered}(j+1,m,n))) \wedge \text{isAlternatingE}(i,j,m,n))) \vee\\ \text{isAlternatingE}(n,n,m,n).\end{aligned}$$

With these predicates we can write a predicate asserting that the Thue-Morse word contains factors of every length $n > 3$ that are maximally unbordered and occur at position $\leq n$. We split the computation into cases, one for even length words, and one for odd:

$$\begin{aligned}&\forall n\,((n \geq 2) \implies (\exists i \text{ hasMNUCE}(i, 2n)) \wedge i \leq 2n)\\ &\forall n\,((n \geq 2) \implies (\exists i\, \text{hasMNUCO}(i, 2n+1)) \wedge i \leq 2n+1),\end{aligned}$$

and `Walnut` evaluates these predicates to be true.

Thus we have that

$$\text{mnuc}_2(n) = \begin{cases} 1, & \text{if } n = 1;\\ 2, & \text{if } n = 2 \text{ or } n = 3;\\ n/2, & \text{if } n \in \{2^{i+1}, 3 \cdot 2^i : i \geq 1\};\\ n/2 - 1, & \text{if } n > 3 \text{ even and } n \notin \{2^i, 3 \cdot 2^i : i \geq 1\};\\ \lfloor n/2 \rfloor, & \text{if } n > 3 \text{ odd.}\end{cases}$$

As a corollary, we easily get the following.

Corollary 3. *Let* $f(n) = \text{mnuc}_2(n) - \lfloor n/2 \rfloor$. *Then* f *is a* 2*-automatic sequence.*

4 More about unbordered conjugates

In this section we show that there exist binary words of length n that have exactly i unbordered conjugates where $1 < i \leq \text{mnuc}_2(n)$.

The general idea behind the proof is to pick some $i > 1$ and then pick a word w of odd length such that $\text{nuc}(w) = i$ and $\text{mnuc}_2(|w|) = i$. Furthermore we only consider such words w such that one of w's conjugates contain 000 as a factor. Then we keep adding 0's to w precisely where 000 first occurs. This keeps the number of unbordered conjugates the same. Then we can keep increasing the size of w in this way until we hit the length we want.

Lemma 1. *For* $n > 4$ *odd, there exists a word* $w \in \Sigma_2^n$ *such that* $\text{nuc}(w) = \lfloor n/2 \rfloor$ *and* 000 *is a factor of some conjugate of* w.

Proof. By Theorem 2, such a word w exists as a factor of the Thue-Morse word. It is well known that the Thue-Morse word is overlap-free. So 000 cannot be a factor of such a word w. But it is possible that $w = 0u00$, or $w = 00u0$ for some word u. We can check whether this is the case for all odd $n > 4$ by modifying our predicate from the proof of Theorem 2:

$$\forall n\,((n \geq 2) \implies (\exists i \text{ hasMNUCO}(i, 2n{+}1)) \wedge ((T[i] = 0 \wedge T[i{+}1] = 0 \wedge T[2n{+}i] = 0) \\ \vee (T[i] = 0 \wedge T[2n-1+i] = 0 \wedge T[2n+i] = 0))),$$

which evaluates to true.

Lemma 2. *Let* $n > 4$ *be odd and* w *be a binary word of length* n *such that a conjugate of* w *has* 000 *as a factor and* $\text{nuc}(w) = \lfloor n/2 \rfloor$. *Then every conjugate of* w *contains at most one distinct occurrence of* 000 *as a factor.*

Proof. Suppose, contrary to what we want to prove that a conjugate of w contains at least two distinct occurrences of 000 as a factor. Call this conjugate w'.

If the two occurrences of 000 overlap, then we can write $w' = s0000t$ for some words s, t. Then the cyclic shifts $0ts000$, $00ts00$, and $0ts000$ are bordered. This means that only $\lfloor |ts|/2 \rfloor + 1$ of the remaining cyclic shifts of w can be unbordered since any unbordered cyclic shift must be followed by a bordered one. But $\lfloor |ts|/2 \rfloor + 1 = \lfloor (n-4)/2 \rfloor + 1 < \lfloor n/2 \rfloor$, so the two occurrences of 000 cannot overlap.

If the two occurrences of 000 do not overlap, then we can write $w' = s000t000$ for some words s, t where s, and t are non-empty. Then the conjugates $00t000s0$, $0t000s00$, $00s000t0$, and $0s000t00$ are bordered. By the same argument as above, of the remaining cyclic shifts, a maximum of $\lfloor |st|/2 \rfloor + 2$ of them can be unbordered. But $\lfloor |st|/2 \rfloor + 2 = \lfloor (n-6)/2) \rfloor + 2 < \lfloor n/2 \rfloor$, a contradiction.

Lemma 3. *Let* $n > 4$ *be odd and* w *be a binary word of length* n *such that a conjugate* w' *of* w *has* 000 *as a prefix and* $\text{nuc}(w) = \lfloor n/2 \rfloor$. *Then* $\text{nuc}(w) = \text{nuc}(w') = \text{nuc}(0^i w')$ *for all* $i \geq 0$.

Proof. Let $i \geq 0$ be an integer. We can write $w' = 000u$ for some word u. It is clear that $0^j u 0^{i+3-j}$ is bordered for all $1 \leq j \leq i+2$. Therefore, it suffices to prove that $s000t$ is bordered if and only if $s0^{i+3}t$ is bordered where $u = ts$.

First we prove the forward direction. Suppose $s000t$ is bordered. By Lemma 2 we have that $s000t$ contains only one occurrence of 000 as a factor. So 000 is neither a prefix of $s00$ nor a suffix of $00t$. Thus, any border of $s000t$ must of length $\leq \min\{|s|, |t|\} + 2$. But such a border would also be a border of $s0^{i+3}t$.

A similar argument works for the reverse direction. Therefore $\mathrm{nuc}(w) = \mathrm{nuc}(w') = \mathrm{nuc}(0^i w')$ for all $i \geq 0$.

Theorem 4. *For all* $1 < i \leq \mathrm{mnuc}_k(n)$ *there exists* $w \in \Sigma_k^n$ *such that* $\mathrm{nuc}(w) = i$.

Proof. Let $C = \{5, 7, 9, 10, 14, 17\}$. For $k \geq 4$, Harju and Nowotka [8] showed that for all integers i with $1 < i \leq n$ there exists a word $w \in \Sigma_k^n$ such that $\mathrm{nuc}(w) = i$. For $k = 3$, Harju and Nowotka [8] showed that if $n \notin C$ then for all integers i with $1 < i \leq n$ there exists a word $w \in \Sigma_k^n$ such that $\mathrm{nuc}(w) = i$, and if $n \in C$ then for all integers i with $1 < i < n$ there exists a word $w \in \Sigma_k^n$ such that $\mathrm{nuc}(w) = i$.

To the best of the authors' knowledge, there is no known proof of the existence of such words for $k = 2$. Suppose $k = 2$. By Theorem 2 there exists a $w \in \Sigma_2^n$ such that w is a factor of the Thue-Morse word and $\mathrm{mnuc}_2(n) = \mathrm{nuc}(w)$. So assume $i < \mathrm{mnuc}_2(n)$. By Lemma 1 there exists a binary word u of odd length m such that $\mathrm{nuc}(u) = i = \lfloor m/2 \rfloor$ and 000 is a factor of some conjugate of u. Let u' be the conjugate of u such that 000 is a prefix of u'. Lemma 3 tells us $\mathrm{nuc}(u) = \mathrm{nuc}(u') = \mathrm{nuc}(0^{n-m}u')$. Since $\mathrm{nuc}(0^{n-m}u') = i$ and $|0^{n-m}u'| = n$, we have that for all $1 < i \leq \mathrm{mnuc}_2(n)$, there exists a $w \in \Sigma_2^n$ such that $\mathrm{nuc}(w) = i$.

5 Conclusions

We want to emphasize that our experience shows that rephrasing problems in combinatorics on words using the first-order logical theory of automatic sequences can be a useful tool in solving these problems. We encourage others to adopt this approach.

Acknowledgments

We thank Dirk Nowotka for helpful discussions. We are very grateful to the referees for helpful comments, and for reminding us about the paper [11].

References

1. Allouche, J.P., Shallit, J.: Automatic Sequences: Theory, Applications, Generalizations. Cambridge University Press (2003)

2. Allouche, J.P., Shallit, J.O.: The ubiquitous Prouhet-Thue-Morse sequence. In: Ding, C., Helleseth, T., Niederreiter, H. (eds.) Sequences and Their Applications, Proceedings of SETA '98, pp. 1–16. Springer-Verlag (1999)
3. Berstel, J.: Sur la construction de mots sans carré. Séminaire de Théorie des Nombres pp. 18.01–18.15 (1978–1979)
4. Berstel, J.: Axel Thue's Papers on Repetitions in Words: a Translation. No. 20 in Publications du Laboratoire de Combinatoire et d'Informatique Mathématique, Université du Québec à Montréal (February 1995)
5. Bruyère, V., Hansel, G., Michaux, C., Villemaire, R.: Logic and p-recognizable sets of integers. Bull. Belgian Math. Soc. **1**, 191–238 (1994), corrigendum, Bull. Belg. Math. Soc. **1**, 577 (1994)
6. Currie, J.: There are ternary circular square-free words of length n for $n \geq 18$. Electronic J. Combinatorics **9**, #N10 (2002), available at https://www.combinatorics.org/v9i1n10
7. Harju, T., Nowotka, D.: Border correlation of binary words. J. Combin. Theory Ser. A **108**, 331–341 (2004)
8. Harju, T., Nowotka, D.: Bordered conjugates of words over large alphabets. Electronic J. Combinatorics **15**, #N41 (2008), available at https://www.combinatorics.org/ojs/index.php/eljc/article/view/v15i1n41
9. Mousavi, H.: Automatic theorem proving in `Walnut` (2016), arxiv preprint, available at https://arxiv.org/abs/1603.06017. Software available at https://github.com/hamousavi/Walnut
10. Shallit, J., Zarifi, R.: Circular critical exponents for Thue-Morse factors. RAIRO Inform. Théor. App. **53**, 37–49 (2019)
11. Shur, A.M.: On ternary square-free circular words. Electronic J. Combinatorics **17**, #R140 (2010), available at https://www.combinatorics.org/ojs/index.php/eljc/article/view/v17i1r140
12. Thue, A.: Über unendliche Zeichenreihen. Norske vid. Selsk. Skr. Mat. Nat. Kl. **7**, 1–22 (1906), reprinted in *Selected Mathematical Papers of Axel Thue*, T. Nagell, editor, Universitetsforlaget, Oslo, 1977, pp. 139–158
13. Thue, A.: Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen. Norske vid. Selsk. Skr. Mat. Nat. Kl. **1**, 1–67 (1912), reprinted in *Selected Mathematical Papers of Axel Thue*, T. Nagell, editor, Universitetsforlaget, Oslo, 1977, pp. 413–478