Some further results on squarefree arithmetic progressions in infinite words

James Currie, Tero Harju, Pascal Ochem, Narad Rampersad

Some further results on squarefree arithmetic progressions in infinite words

James Currie and Narad Rampersad*

Department of Mathematics and Statistics

University of Winnipeg

`{j.currie,n.rampersad}@uwinnipeg.ca`

Tero Harju

Department of Mathematics and Statistics

University of Turku, Finland

`harju@utu.fi`

Pascal Ochem†

LIRMM - CNRS

University of Montpellier, France

`ochem@lirmm.fr`

January 21, 2019

Abstract

In a recent paper, one of us posed three open problems concerning squarefree arithmetic progressions in infinite words. In this note we solve these problems and prove some additional results.

1 Introduction

The study of infinite words avoiding squares is a classical problem in combinatorics on words. A *square* is a word of the form xx, such as `tartar`. One of the most well-studied squarefree words [16, 9] is the word $\mathtt{vtm} = 012021012102012\cdots$ obtained by iterating the map $0 \mapsto 012$, $1 \mapsto 02$, $2 \mapsto 1$.

A number of authors, such as Carpi [5], Currie and Simpson [8], and Kao et al. [12], have studied squarefree arithmetic progressions in infinite words. Let p be a positive integer. If

---

1

$w = w_0 w_1 w_2 \cdots$ is an infinite word where each w_i is a single letter, then $[w]_p$ denotes the infinite word $w_0 w_p w_{2p} \cdots$. Harju [10] studied the following question and showed that it has a positive solution for all $p \geq 3$:

> Given p, does there exist an infinite squarefree word w over a ternary alphabet such that $[w]_p$ is squarefree?

At the end of his paper, Harju posed three open problems:

Problem 1. Does there exist a squarefree word w over a ternary alphabet such that for every $p \geq 3$, the subsequence $[w]_p$ contains a square?

Problem 2. Do there exist pairs (p, q) of relatively prime integers such that there exists a squarefree word w over a ternary alphabet for which both $[w]_p$ and $[w]_q$ are squarefree?

Problem 3. It is true that for all squarefree words v over a ternary alphabet, there exists a squarefree word w and an integer $p \geq 3$ such that $[w]_p = v$?

In this paper we answer each of these questions. Regarding Problem 3, we resolve this problem by instead studying the following much stronger version of this problem.

Problem 4. Let $p \geq 2$ be an integer and let v be *any* infinite ternary word. Does there exist an infinite ternary squarefree word w such that $[w]_p = v$?

For $6 \leq p \leq 29$ and $v = 00000 \cdots$, we give a positive answer to Problem 4. For $p \geq 30$, we give a positive answer for all infinite words v. In fact, we prove an even stronger result by showing that, given any sequence $(p_i)_{i \geq 0}$ of positions such that $p_{i+1} - p_i \geq 30$, it is possible to construct w such that v appears as the subsequence of w indexed by $(p_i)_{i \geq 0}$.

2 Preliminaries

Let A be a finite alphabet of letters. For a word w over A (i.e., $w \in A^*$), let $|w|$ denote its length, i.e., the number of occurrences of letters in w. A word u is a *factor* of w, if $w = xuy$ where x and/or y may be empty. If x (y, resp.) is empty then u is a *prefix* (a *suffix*, resp.) of w.

A finite or infinite word w over A is *squarefree* if it does not have any factors of the form $u^2 = uu$ for nonempty words $u \in A^*$. A morphism $h \colon A^* \to A^*$ is said to be *squarefree*, if it preserves squarefreeness of words, i.e., if $h(w)$ is squarefree for all squarefree words w. A morphism $h \colon A^* \to A^*$ is *uniform* if the images $h(a)$ have the same length: $|h(a)| = n$ for all $a \in A$ and for some positive n called the *length* of h.

An infinite word w is a *fixed point* of a morphism h if $h(w) = w$. This happens if w begins with the letter a, and w is obtained by iterating h on the first letter a of w: $h(a) = au$ and $w = auh(u)h^2(u) \cdots$. In this case we denote the fixed point w by $h^\omega(a)$.

For the next algorithmically helpful result we refer to Crochemore [6].

Theorem 1. *Let $h \colon A^* \to A^*$ be a morphism.*

1. *If $|A| = 3$ then h is squarefree if and only if it preserves squarefreeness of all squarefree words of length five.*

2. *If h is uniform then h is squarefree if and only if it preserves squarefreeness of all squarefree words of length three.*

In the rest of this paper we work over the ternary alphabet $T = \{0, 1, 2\}$. Let $\tau \colon T^* \to T^*$ be the morphism defined by

$$\tau(0) = 012, \ \tau(1) = 02 \text{ and } \tau(2) = 1 \,. \tag{1}$$

The morphism τ is not squarefree since it does not preserve squarefreeness of the words 010 and 02120. The images of these are 01(2020)12 and 01(210210)12, respectively, with the squares indicated. Nevertheless, the word obtained by iterating the morphism on 0 gives an infinite squarefree word as a limit, called the *Thue word*:

$$\texttt{vtm} = 012021012102012 \cdots \ (= \tau^{\omega}(0)).$$

(Here we follow [2] in using $\texttt{vtm}$, for *variant of the Thue–Morse word*, to denote this word.) For the next basic result, see [4, 9, 11, 14, 16] and [13]:

Lemma 2. *The Thue word $\texttt{vtm}$ is squarefree and it does not contain 010 or 212 as factors.*

Blanchet-Sadri et al. [2, Theorem 3] proved the following:

Lemma 3. *For each odd k and each factor w of $\texttt{vtm}$, the set of positions at which w occurs in $\texttt{vtm}$ contains all congruence classes modulo k.*

3 Problem 1

In this section we show that the word $\texttt{vtm}$ gives a positive solution to Problem 1.

Theorem 4. *For each $k \geq 2$ the word $[\texttt{vtm}]_k$ contains either the square 00 or the square 22.*

Proof. The first part of the proof relies on the fact that $\texttt{vtm}$ is a 2-automatic sequence. Berstel [1] studied several different ways to generate the sequence $\texttt{vtm}$; in particular, he showed that $\texttt{vtm}$ is generated by the 2-DFAO (*deterministic finite automaton with output*) in Figure 1. The automaton takes the binary representation of n as input (reading from most significant digit to least significant digit), and if the computation ends in a state labeled a, the automaton outputs a, indicating that $v_n = a$.

Since $\texttt{vtm}$ is an automatic sequence, we can use Walnut [15] to verify that it has certain combinatorial properties. We verify with Walnut that for every $k \geq 2$, the sequence $\texttt{vtm}$ contains a length $k + 1$ factor of the form $0u0$ or $2u2$. The Walnut command to do this is:

```
eval same_first_last ''Ei (VTM[i]=@0 & VTM[i+k]=@0)|(VTM[i]=@2 &
                        VTM[i+k]=@2)'';
```
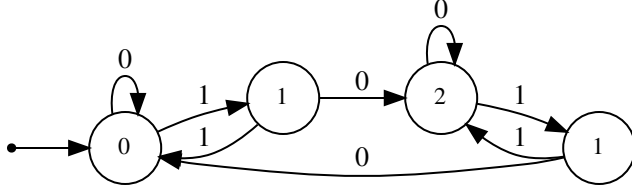
Figure 1: 2-DFAO for `vtm`
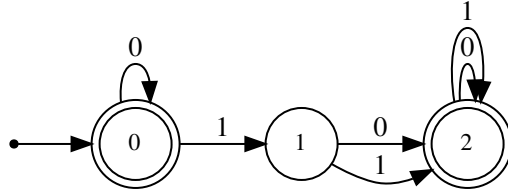
(k): Ei (VTM[i] = @0 & VTM[i+k] = @0) | (VTM[i] = @2 & VTM[i+k] = @2)

Figure 2: Walnut output automaton

The Walnut output for this command is the automaton in Figure 2, which shows that the given predicate holds for all $k \geq 2$.

Suppose k is odd. To complete the proof, it suffices to show that `vtm` contains an occurrence of the length $k+1$ factor $0u0$ or $2u2$ at a position congruent to 0 modulo k. This follows immediately from Lemma 3 and so we have established the claim for all odd $k \geq 2$.

If k is even, write $k = 2^a k'$, where k' is odd. Suppose that $k' \geq 3$. We have already seen that `vtm` contains an occurrence of a length $k' + 1$ factor $0u0$ or $2u2$ at a position $i \equiv 0$ (mod k'). From the automaton generating `vtm`, we see that if $v_i = 0$ (resp. $v_i = 2$), then $v_{2^a i} = 0$ (resp. $v_{2^a i} = 2$), which establishes the claim for $k' \geq 3$.

Finally, suppose that k is a power of 2. Then the binary representations of k and $2k$ have the form 10^ℓ and $10^{\ell+1}$ respectively, for some $\ell \geq 1$. From the automaton generating `vtm`, we see that $v_k = v_{2k} = 2$, as required. This completes the proof. $\square$

4   Problem 2

In this section we give positive solutions to Problem 2 for two different pairs (p, q); i.e. $(p, q) = (3, 11)$ and $(p, q) = (5, 6)$. We first introduce some notation. Given a morphism $h : A^* \to A^*$ and a positive integer p such that p divides $|h(a)|$ for all $a \in A$, let h_p be the

morphism defined by $h_p(a) = [h(a)]_p$.

Theorem 5. *For $(p, q) = (3, 11)$ there exists an infinite ternary squarefree word w such that $[w]_p$ and $[w]_q$ are both squarefree.*

We first define the following morphism h on T^*:

$$
\begin{aligned}
0 \mapsto\ & 01210120210201021201210120102120210120102012021012102012021012010210 \\
& 12102120102012021201021012102012021201210201021012102012021012010210 \\
& 12021201210201021201210120102120210201021201210120210201021012010212 \\
1 \mapsto\ & 01210120210201021201210120102120210120102012021012102012021012010210 \\
& 12021201210201021201210120102120210201021201210120210201021012010212 \\
2 \mapsto\ & 01021012102120102012102120210120102120210201210120210201021012010212
\end{aligned}
$$

The next result gives a condition for a morphism g to map $\mathtt{vtm}$ to a squarefree word. We will use it to show that $h(\mathtt{vtm})$, $h_3(\mathtt{vtm})$, and $h_{11}(\mathtt{vtm})$ are all squarefree, which is sufficient to establish Theorem 5.

Theorem 6. *Suppose that $g : T^* \to A^*$ is such that*

1. *$g(u)$ is square-free for every factor u of $\mathtt{vtm}$ of length 5.*

2. *The only solution of*

$$
\begin{aligned}
g(a) &= uv \\
g(b) &= zv \\
g(c) &= zw \\
a, b, c &\in T,
\end{aligned}
$$

   such that $v \neq w$ and $u \neq z$ is $abc = 010$.

3. *For $a \in T$ we can write $g(a) = uv_a$ such that whenever $w \in T^*$ and $g(w) = xv_a z$, then $z \in g(T^*)$.*

Then $g(\mathtt{vtm})$ is square-free.

Remark 7. Let h, h_3, h_{11} be given as above. Let $\hat{h}_{11}$ be defined on letters by $\hat{h}_{11}(a) = 0^{-1}h_{11}(a)0$. The theorem's conditions hold for $g = h, h_3, \hat{h}_{11}$.

- For $g = h$, condition 3 is witnessed by $v_a = 02102010210120212$ for $a = 0, 1, 2$.

- For $g = h_3$, condition 3 is witnessed by $v_a = 210212$ for $a = 0, 1, 2$.

- For $g = \hat{h}_{11}$, condition 3 is witnessed by $v_0 = v_1 = 1210$, $v_2 = 20210$.

Thus, in order to establish Theorem 5, it remains to give the proof of Theorem 6.

Proof of Theorem 6. Let $a_0 a_1 \cdots a_m$ be a shortest factor of $\mathtt{vtm}$ such that $g(a_0 a_1 \cdots a_m)$ contains a square. Some non-empty xx is a factor of $g(a_0 a_1 \cdots a_m)$. Since $m \geq 4$ by condition 1, and m is as small as possible, it follows that $g(a_i)$ is a factor of x for some i.

Using condition 3 repeatedly, write

$$A_0'' A_1 \cdots A_{j-1} A_j' = A_j'' A_{j+1} \cdots A_{2j-1} A_{2j}'$$

where

$$
\begin{aligned}
A_i &= g(a_i), 0 \leq i \leq 2j, \\
A_i &= A_i' A_i'', i = 0, j, 2j \\
A_0'' &= A_j'' \\
A_j' &= A_{2j}'
\end{aligned}
$$

and $m = 2j$.

By condition 1, g is 1-1, so that $a_i = a_{j+i}$, $1 \leq i \leq j-1$. Since $\mathtt{vtm}$ is square-free, $A_0' \neq A_j'$ and $A_j'' \neq A_{2j}''$. By condition 2, this implies $a_0 = a_{2j} = 0$, $a_j = 1$. Again, since $\mathtt{vtm}$ is square-free, $a_0 \neq a_1$, $a_{j-1} \neq a_j \neq a_{j+1}$, and $a_{2j-1} \neq a_{2j}$. Now $0 = a_0 \neq a_1 = a_{j+1} \neq a_j = 1$; thus $a_{j+1} = 2$. Similarly, $0 = a_{2j} \neq a_{2j-1} = a_{j-1} \neq a_j = 1$; thus $a_{j-1} = 2$. Then $a_{j-1} a_j a_{j+1} = 212$ is a factor of $\mathtt{vtm}$. This is a contradiction. $\square$

Remark 8. We also make the following observation concerning the morphism h_{11}, which is defined by

$$
\begin{aligned}
0 &\mapsto 0201021 \quad 012021 \quad 20121 \\
1 &\mapsto 0201021 \quad 20121 \\
2 &\mapsto 012021.
\end{aligned}
$$

By splitting the images in this way, we see that $h_{11} = x_{11} \circ \tau$, and hence that $h_{11}(\mathtt{vtm}) = x_{11}(\mathtt{vtm})$, where x_{11} is the morphism defined by

$$0 \mapsto 0201021, \quad 1 \mapsto 012021, \quad 2 \mapsto 20121.$$

One now verifies (using Theorem 1) that x_{11} is a squarefree morphism, which implies that $h_{11}(\mathtt{vtm})$ is squarefree.

Theorem 9. *For $(p, q) = (5, 6)$ there exists an infinite ternary squarefree word w such that $[w]_p$ and $[w]_q$ are both squarefree.*

Proof. To prove the result it suffices to find a morphism h such that h, h_5, and h_6 are all squarefree (which we can verify using Theorem 1). Such a morphism is given below.

$0 \mapsto$ 012102120210201021012021201210120210201021012102120121012021201020
120210201021012010201210120212012102120210201202120102012101201021
012102120121012021201020120210201021012010201210120212010201210212
021012102001021012010201202101210212021020102120121012010201202102 0
102120210120102012021201210212021012102001202120121012010201202102 01
021202101201020120210121021202102001021201210120102120210201021012 1
020120212012101201021202102001021201210120102012021020102120210121 0
212012101201021202102001021012102001202120121021020102012021020102 1201
020120210201021202101210201202120121021201021012102010201201210120 21
201020120210201021012010201210120212

$1 \mapsto$ 012102120210201021012021201210120210201021012102120121012021201020
120210201021012010201210120212012102120210201202120102012101201021
012102120121012021201020120210201021012010201210120212010201210212
021012102001021012010201202101210212021020102120121012010201202102 01
021012102012021201210201021012102120210120102012021201210212021012
102012021201210120102101210201021202101201020120210121021202102010
212012101201021202102001021012102010202120121021020102120210201021 2012
101201020120210201021202101210201202120121021021020120102102120121 2012
101201021202102001021012102001202120121021020102120210201202101020 120
210201021012102120102012101201021012102120210201210120212

$2 \mapsto$ 012102120210201021012021201210120210201021012102120121012021201020
120210201021012010201210120212012102120210201202120102012101201021
012102120121012021201020120210201021012010201210120212010201210212
021012102001021012010201202101210212021020102120121012010201202102 01
021012102012021201210201021012102120210120102012021201210212021012
102012021201210120102101210201021202101201020120210121021202102010 120
212012101201021202102001021012102010202120121021020102120210201021 2012
101201020120210201021202102012102120121021020102102012021012102012 02120121
012010210121020102120121021202102001202102010210120102012101201021 0120212
012102120210201202120102012101201021012102120210201210120212

Remark 10. For the pair $(p, q) = (5, 8)$ we searched via a backtracking algorithm for words of the desired form. The backtrack appears to get "stuck" around length 1200, suggesting that there may not be an infinite word for the pair $(5, 8)$. This would be somewhat surprising, given the existence of an infinite word for the pair $(5, 6)$.

5   Problems 3 and 4

Problem 3 is a special case of Problem 4, so we focus on Problem 4. We shall show that the least positive answer for Problem 4 in the ternary case is $p = 4$. We also state some positive and negative examples for small integers p for Problem 4.

5.1   Problem 4 for $2 \leq p \leq 5$ and constant v

Lemma 11. *Problem 4 has no solution for the moduli $p = 2$, $p = 3$ and $p = 5$.*

Proof. Indeed, suppose first that $w \in T^\omega$ is an infinite squarefree word such that $[w]_p = 00 \cdots$ is unary for $p \in \{2, 3\}$. Then necessarily $w \in \{01, 02\}^\omega$ or $w \in \{012, 021\}^\omega$, respectively. However, there are no infinite squarefree words in the binary case, and hence the moduli 2 and 3 have no solutions.

We then show that Problem 4 has no solution for the modulus $p = 5$. In this case if $[w]_5$ is constant then
$$w \in \{01021, 01201, 02102, 02012\}^\omega.$$
The longest words having an arithmetic progression of zeros modulo $p = 5$ are the following three words of length 40:

$$0102101201020120210201021012010201202101,$$
$$0102101201020120210201021012010201202120,$$
$$0120102012021020102101201020120210201210,$$

where there are 'malapropos' factors 02120 and 01210 at the end of the two last words.   □

Remark 12. If we allow four letters in the alphabet, then Problem 4 becomes trivial for constant $v = 0000 \cdots$. Indeed, let $p \geq 2$, and let $w \in \{1, 2, 3\}^\omega$ be any squarefree ternary word. The quaternary word obtained by adding 0 at the beginning and after every $p - 1$ letters of w has an arithmetic progression of 0's modulo p. For arbitrary v, determining precisely the values of p for which Problem 4 has a solution over a four-letter alphabet remains an open problem.

For modulus $p = 4$ we have a solution for constant v.

Theorem 13. *The Thue word* vtm *has an infinite arithmetic progression of 1's modulo 4.*

Proof. We rely on the 3rd power of the morphism τ defined in Eq. (1):
$$\tau^3(0) = 012021012102$$
$$\tau^3(1) = 01202102$$
$$\tau^3(2) = 0121.$$

Since $\tau^3(\mathtt{vtm}) = \mathtt{vtm}$, it is immediate that the letters of $\mathtt{vtm}$ at positions $n \equiv 1 \pmod 4$ are all equal to 1. $\qquad\square$

5.2   Problem 4 for $6 \le p \le 29$ and constant v

Theorem 14. *For all $p \ge 6$ divisible by an integer $m \le 29$, there exists an infinite squarefree word $w \in T^\omega$ such that $[w]_p = 000\cdots$.*

Before going to the constructions we state some aspects related to the problem. For constant v, Theorem 14 will settle the cases for $6 \le p \le 29$.

Remark 15. Currie [7] showed that there exist uniform squarefree morphisms $T^* \to T^*$ for all lengths $n \ge 23$. However, the constructed morphisms h in [7] are *cyclic shift morphisms* for which $h(1) = \pi(h(0))$ and $h(2) = \pi(h(1))$ where π is the permutation $(0\,1\,2)$ of T. In particular the images of the letters start with different letters. Because of this they do not provide solutions to our main problem; see Theorem 16 below.

The following result narrows the candidates of infinite squarefree words with arithmetic progressions obtained by morphisms. It also gives infinitely many examples of integer sequences $a_1 < a_2 < \cdots$ avoiding infinite arithmetic progressions with the property that the differences $a_{i+1} - a_i$ form a bounded sequence. Indeed, the bound is always 4. (Here a_i will mark the place of the ith letter 0 in the sequence $h^\omega(0)$.)

Theorem 16. *Let $h\colon T^* \to T^*$ be a uniform squarefree morphism of prime number length q such that*

$$0 \text{ is a prefix of } h(0) \text{ but not of } h(1) \text{ and } h(2). \tag{2}$$

Then there does not exist any infinite arithmetic progressions of 0's in $w = h^\omega(0)$.

Proof. Assume that there exists an integer p such that w has an arithmetic progression of 0's modulo p. Let p be chosen to be the least of these integers. Then the common length q of the images $h(a)$, $a \in T$, divides p. Indeed, otherwise $\gcd(p, q) = 1$, as q is a prime number, and therefore the multiples of p form a complete residue system modulo q. In this case w is the all zero word; a contradiction.

Let $w = W_1 W_2 \cdots$ where $|W_i| = p$. In particular, each W_i begins with the letter 0, and, by the above, for each i there exists a unique word $U_i \in T^*$ of length p/q such that $h(U_i) = W_i$. By (2), each W_i begins with 0. Now, since w is a fixed point of h, it is also the case that $w = U_1 U_2 \cdots$. However, $|U_i| = p/q < p$ contradicts the minimality assumption on p. $\qquad\square$

Example 17. The least length for a uniform squarefree morphism is 11; see Brandenburg [3]. The following uniform morphism of length 11 is squarefree and satisfies the conditions of Theorem 16:

$$
\begin{aligned}
h(0) &= 02120102012\,, \\
h(1) &= 10201021012\,, \\
h(2) &= 10212021012\,.
\end{aligned}
$$

Proof of Theorem 14. We proceed by constructing a uniform squarefree morphism h to witness the claim for each p with $6 \le p \le 29$. For this we need only to give the morphisms for values of p that are not divisible by 4 or by any q with $6 \le q < p$. The morphism h is then applied to any infinite squarefree word (such as the Thue word `vtm`). The squarefreeness of the morphisms listed below can be checked by Theorem 1 aided by a computer.

Case $p = 6$. Image length 36:
010212012102010212021012021201021012
010212012102012021012102010212021012
010212012102012021020121021201021012

Case $p = 7$. Image length 28:
0121012021012102010210120212
0121012021012102120210120212
0121012021020102120210120102

Case $p = 9$. Image length 18:
010210121020120212
010210121020121012
010210121021202102

Case $p = 10$. Image length 20:
01021012010201202102
01021012010201210212
01021012010212012102

Case $p = 11$. Image length 11:
01021012102
01021202102
01210120212

Case $p = 13$. Image length 26:
01201020120210120102120121
01201020120210121021201021
01201020120210201210212021

Case $p = 15$. Image length 30:
010201202101201021012102010212
010201202101201021201210120212
010201202101201021202101210212

Case $p = 17$. Image length 34:
0102012021020102101201021201210212
0102012101202120102101202102010212
0102012101202120102101202120121012

Case $p = 19$. Image length 19:
0102012021020121012
0102012021201021012
0102012102120121012

Case $p = 23$. Image length 23:
01020120210121020120212
01020120210201210120212
01020121012010210120212

Case $p = 25$. Image length 25:
0102012021012010201210212
0102012021012010210120212
0102012021012102120121012

Case $p = 29$. Image length 29:
01020120210120102012021201021
01210201202120102012102010212
01210201202120102120210120212

$\square$

We note that there are only cyclic shift morphisms for lengths 13 and 17, and there are no uniform morphisms of length 15. We also note that the morphisms given above all have the property that the images all have a common non-empty prefix. This raises an interesting question concerning the existence of n-uniform squarefree morphisms with this property and the maximum length of the common prefix. Given a morphism $h : T^* \to T^*$, let $\mathrm{lcp}(h)$ denote the length of the longest common prefix of the images $\{h(a) : a \in T\}$. Given a positive integer n, let $\mathrm{lcp}(n)$ denote the maximum value of $\mathrm{lcp}(h)$ over all n-uniform squarefree morphisms $h : T^* \to T^*$. We have computed the value of $\mathrm{lcp}(n)$ for $18 \le n \le 70$. The results of this computation suggest the following conjecture:

Conjecture 18. For every $n \ge 30$, there exists an n-uniform squarefree morphism $h : T^* \to T^*$ such that $\mathrm{lcp}(h) = n - 6$.

Example 19. For $n = 70$ the morphism h with images

$$h(0) = 0120210201021012010201202101201021012021020102120210120102012021201021$$
$$h(1) = 0120210201021012010201202101201021012021020102120210120102012021020121$$
$$h(2) = 0120210201021012010201202101201021012021020102120210120102012021012102$$

is squarefree and the images have a common prefix of length 64.

We also note that the $n - 6$ in Conjecture 18 would be optimal. Using a computer we verified that for every triple of distinct squarefree words (u, v, w) with $|u| = |v| = |w| = 5$, there do not exist words p and s with $|p| = |s| = 8$ such that pus, pvs, and pws are all squarefree. It follows that the $n - 6$ in the conjecture cannot be replaced with $n - 5$.

5.3 Problem 4 for $p \geq 30$ and arbitrary v

The next result implies a positive solution to Problem 4 for $p \geq 30$ and arbitrary v, and also implies a positive solution to Problem 3.

Theorem 20. *Let $(p_i)_{i \geq 0}$ be a sequence of positions such that $p_{i+1} - p_i \geq 30$ and let v be any infinite ternary word. There exists an infinite squarefree ternary word w such that v appears as the subsequence of w indexed by $(p_i)_{i \geq 0}$.*

Proof. We use the squarefree multi-valued morphism h defined by

$$h(0) = \begin{cases} 01210212021012021201210 \\ 01210212021020102021201210 \\ 012102120210201202021201210 \\ 01210212021020121021201210, \end{cases}$$

$h(1) = \pi(h(0))$, and $h(2) = \pi(h(1))$, where π is the permutation $(0\,1\,2)$ of T. To show that h is squarefree, we apply Theorem 1(a). It is easy to check that the proof of Theorem 1 from [6] works for multi-valued morphisms as well; however, when applying the theorem, for each squarefree word x of length 5 we have to check that all 4^5 words in the image $h(x)$ are squarefree. Note that the images of each letter have length 23, 24, 25, or 26. Also, they share a common prefix of length 12 and a common suffix of length 9, so that they only differ by their middle factors (in bold).

We start with any infinite ternary squarefree word, say the Thue word `vtm`. We apply h to `vtm` with images of length 26 only, which gives an infinite ternary squarefree word $t = t_0 t_1 t_2 \cdots$.

Suppose that $t_{p_i} = v_i$; we will show how to modify t so that $t_{p_{i+1}} = v_{i+1}$. If $t_{p_{i+1}} \neq v_{i+1}$, then t contains the forced letter v_{i+1} at position $p_{i+1} + 1$, $p_{i+1} + 2$, or $p_{i+1} + 3$. Notice that every factor of t of length $26 + 5 - 1 = 30$ contains a full occurrence of a middle factor of length 5. Since $p_{i+1} - p_i \geq 30$, there exist a full occurrence of a middle factor of length 5 between positions $p_i + 1$ and p_{i+1}. If the match to v_{i+1} is at position $p_{i+1} + 1$ (resp. $p_{i+1} + 2$, $p_{i+1} + 3$), then we can replace this middle factor by the middle factor of

11

length 4 (resp. 3, 2), corresponding to the alternative h-image of length 25 (resp. 24, 23). This ensures that in the resulting word, the letter at position p_{i+1} now matches v_{i+1}. This procedure can be repeated for all i and the resulting word is the desired word w. $\qquad\square$

6 Future work

One obvious open problem is to characterize the pairs (p, q) for which Problem 2 has a positive solution. Another open problem is to prove Conjecture 18. Regarding Problem 4, we have only shown the existence of a positive solution for $p \geq 30$. It remains to determine what happens for $p \leq 29$. For example, for $p = 4$, one can verify by computer that the infinite word $v = 010101 \cdots$ cannot be obtained as a mod 4 subsequence of any squarefree ternary word. Are there similar counterexamples for other small values of p, such as $p = 6$?

References

[1] J. Berstel, "Sur la construction de mots sans carré", Séminaire de Théorie des Nombres 1978âĂŞ1979, Exp. No. 18, 15 pp., CNRS, Talence, 1979.

[2] F. Blanchet-Sadri, J. Currie, N. Fox, N. Rampersad, "Abelian complexity of fixed point of morphism $0 \mapsto 012$, $1 \mapsto 02$, $2 \mapsto 1$", INTEGERS 14 (2014), A11.

[3] F.-J. Brandenburg, "Uniformly growing k-th power-free homomorphisms", Theoret. Comput. Sci. 23 (1983), 69–82.

[4] C. Braunholtz, "An infinite sequence of three symbols with no adjacent repeats", Amer. Math. Monthly 70 (1963), 675–676.

[5] A. Carpi, "Multidimensional unrepetitive configurations", Theoret. Comput. Sci. 56 (1988), 233–241.

[6] M. Crochemore, "Sharp characterizations of squarefree morphisms", Theoret. Comput. Sci. 18 (1982), 221–226.

[7] J. Currie, "Infinite ternary square-free words concatenated from permutations of a single word", Theoret. Comput. Sci. 482 (2013), 1–8.

[8] J. Currie, J. Simpson, "Non-repetitive tilings", Electron. J. Combinatorics 9 (2002) #R28.

[9] M. Hall, Jr., "Generators and relations in groups–The Burnside problem". In *Lectures on Modern Mathematics, Vol. 2*, Wiley, New York, pp. 42–92.

[10] T. Harju, "On square-free arithmetic progressions in infinite words". To appear in Theoret. Comput. Sci. Currently online at `https://doi.org/10.1016/j.tcs.2018.09.032`.

[11] S. Istrail, "On irreducible languages and nonrational numbers", Bull Math. Soc. Sci. Math. R. S. Roumanie (N.S.) 21(69)(3-4) (1977), 301–308.

[12] J.-Y. Kao, N. Rampersad, J. Shallit, M. Silva, "Words avoiding repetitions in arithmetic progressions", Theoret. Comput. Sci. 391 (2008), 126–137.

[13] M. Lothaire, *Combinatorics on Words*, vol. 17 of Encyclopedia of Mathematics, Addison-Wesley, Reading, MA, 1983.

[14] M. Morse and G. Hedlund, "Unending chess, symbolic dynamics and a problem in semigroups", Duke Math. J. 11 (1944), 1–7.

[15] H. Mousavi, Automatic theorem proving in Walnut. Preprint: `https://arxiv.org/abs/1603.06017` (2016). Walnut Prover itself is available at `https://github.com/hamousavi/Walnut`.

[16] A. Thue, "Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen", Norske Vid. Skrifter I. Mat.-Nat. Kl., Christiana 1 (1912), 1–67.