Contributions to Discrete Mathematics

Volume 16, Number 1, Pages 8–19
ISSN 1715-0868

LENGTHS OF EXTREMAL SQUARE-FREE TERNARY WORDS

LUCAS MOL AND NARAD RAMPERSAD

Abstract. A square-free word w over a fixed alphabet Σ is *extremal* if every word obtained from w by inserting a single letter from Σ (at any position) contains a square. Grytczuk et al. recently introduced the concept of extremal square-free word and demonstrated that there are arbitrarily long extremal square-free ternary words. We find all lengths which admit an extremal square-free ternary word. In particular, we show that there is an extremal square-free ternary word of every sufficiently large length. We also solve the analogous problem for circular words.

1. Introduction

Throughout, we use standard definitions and notations from combinatorics on words (see [10]). The word u is a *factor* of the word w if we can write $w = xuy$ for some (possibly empty) words x, y. A word is *square-free* if it contains no factor of the form xx, where x is a nonempty word. Early in the twentieth century, Norwegian mathematician Axel Thue demonstrated that one can construct arbitrarily long square-free words over a ternary alphabet (see [2]). Thue's work is recognized as the beginning of the field of combinatorics on words [3].

Let w be a word over a fixed alphabet Σ. A *left (right) extension* of w is a word of the form aw (wa, respectively), where $a \in \Sigma$. We say that a square-free word w is *maximal* if both every left extension of w contains a square, and every right extension of w contains a square. Bean, Ehrenfeucht, and McNulty [1] demonstrated that every square-free word over a fixed alphabet Σ is a factor of a maximal square-free word over Σ. (In fact, Bean, Ehrenfeucht, and McNulty established this result not only for square-free words, but for kth-power free words for every integer $k \geq 2$.)

Received by the editors January 31, 2020, and in revised form May 4, 2020.

2010 *Mathematics Subject Classification.* 68R15.

Key words and phrases. square-free word; extremal square-free word.

The work of Narad Rampersad is supported by the Natural Sciences and Engineering Research Council of Canada (NSERC), [funding reference number 2019-04111].

A corollary is that there are arbitrarily long maximal square-free words over any alphabet of size at least 3.

Grytczuk et al. [7] recently introduced a variant of maximal square-free words, in which extensions not just at the beginning and the end, but at any point in the interior of the word, are considered. Let w be a word over a fixed alphabet Σ. An *extension* of w is a word of the form $w'aw''$, where $a \in \Sigma$ and $w'w'' = w$. We say that w is *extremal square-free* if w is square-free, and there is no square-free extension of w.

Grytczuk et al. [7] demonstrated that there are arbitrarily long extremal square-free ternary words. In this article, we describe exactly those integers n for which an extremal square-free ternary word of length n exists. In particular, we find that there is an extremal square-free ternary word of every sufficiently large length. This confirms a conjecture of Jeffrey Shallit [13].

Theorem 1.1. *Let n be a nonnegative integer. Then there is an extremal square-free word of length n over the alphabet $\Gamma = \{\mathtt{a}, \mathtt{b}, \mathtt{c}\}$ if and only if n is in the set*

$$\mathcal{A} = \{25, 41, 48, 50, 63, 71, 72, 77, 79, 81, 83, 84, 85\} \cup \{m\colon\ m \geq 87\}.$$

We also consider the analogous problem for *circular words*. The words u and v are *conjugates* if there exist words x and y such that $u = xy$ and $v = yx$, i.e., if u and v are cyclic shifts of one another. Let $w \in \Sigma^*$. The *circular word* formed from w, denoted $\langle w \rangle$, is the set of all conjugates of w. For a set of words L, the word u is a *factor* of L if u is a factor of some word in L, and the set L is *square-free* if every word in L is square-free. In particular, the word u is a factor of the circular word $\langle w \rangle$ if and only if u is a factor of some conjugate of w, and the circular word $\langle w \rangle$ is square-free if and only if every conjugate of w is square-free. The following theorem was first proven by Currie [5], and has since been reproven by several different methods [6, 14].

Theorem 1.2 (Currie [5]). *For every integer $n \geq 18$, there is a square-free circular word of length n over the alphabet $\{\mathtt{0}, \mathtt{1}, \mathtt{2}\}$.*

By Theorem 1.2 and a finite search, the only lengths which do not admit square-free ternary circular words are 5, 7, 9, 10, 14, and 17.

Let w be a word over a fixed alphabet Σ. An *extension* of the circular word $\langle w \rangle$ is a circular word of the form $\langle w'aw'' \rangle$, where $a \in \Sigma$ is a letter and $w = w'w''$. The circular word $\langle w \rangle$ is *extremal square-free* if $\langle w \rangle$ is square-free, and every extension of $\langle w \rangle$ contains a square. We prove the following theorem concerning the attainable lengths of extremal square-free ternary circular words. This can be regarded as a strengthening of Theorem 1.2.

Theorem 1.3. *Let n be a nonnegative integer. Then there is an extremal square-free circular word of length n over the alphabet $\Gamma = \{\mathtt{a}, \mathtt{b}, \mathtt{c}\}$ if and* *only if n is in the set*

$$\mathcal{B} = \{4, 6, 8, 13, 15, 16, 18, 20, 21, 22, 23, 24, \\ 28, 30, 32, 33, 34, 35, 36\} \cup \{m\colon \ m \geq 38\}.$$

The layout of the remainder of the article is as follows. In Section 2, we present some preliminaries which are used to prove both of our main results. In Section 3, we prove Theorem 1.1. In Section 4, we prove Theorem 1.3. We conclude with a discussion of some related problems.

2. Preliminaries

We will need the following well-known lemma, attributed to Sylvester. See [11, Section 2.1] for several different proofs.

Lemma 2.1. *Let p and q be relatively prime positive integers. For every integer $n \geq (p-1)(q-1)$, there exist nonnegative integers a and b such that $n = ap + bq$.*

We will also need the following corollary of Lemma 2.1.

Corollary 2.2. *Let p and q be relatively prime positive integers, exactly one of which is even. For every integer $n \geq pq + (p-1)(q-1)$, there exist nonnegative integers a and b such that $n = ap + bq$, and the sum $a + b$ is even.*

Proof. Suppose without loss of generality that p is even and q is odd. Let $n \geq pq+(p-1)(q-1)$. Then we have $n-pq \geq (p-1)(q-1)$. By Lemma 2.1, there are nonnegative integers α and β such that $n - pq = \alpha p + \beta q$. If $\alpha + \beta$ is odd, then we can write $n = (\alpha + q)p + \beta q$, and $\alpha + q + \beta$ is even. If $\alpha + \beta$ is even, then we can write $n = \alpha p + (\beta + p)q$, and $\alpha + \beta + p$ is even. $\square$

Next, we prove a theorem which essentially extends a result of Grytczuk et al. [7, Theorem 2] from a morphism to a multi-valued substitution. We note that many results similar to [7, Theorem 2] have appeared before in the literature (see [12, Section 4.2.5] for a summary). However, most of these results give conditions on a morphism $f : \Sigma^* \to \Delta^*$ which guarantee that $f(w)$ is square-free for *every* square-free word $w \in \Sigma^*$. By contrast, the result of Grytczuk et al. gives conditions on a morphism $f : \Sigma^* \to \Delta^*$ and a square-free word $w \in \Sigma^*$ which guarantee that the word $f(w)$ is square-free, i.e., the conditions depend explicitly on the word w. We note that arguments similar to those used in the proof of the following theorem have appeared before (see the proof of [9, Lemma 8], for example), but the entire proof is included for completeness.

Theorem 2.3. *Let $f : \Sigma^* \to 2^{\Delta^*}$ be a substitution, and let $u \in \Sigma^*$ be a square-free word. Then the set $f(u)$ is square-free if all of the following conditions are satisfied:*

(I) *For every factor v of u of length at most 3, the set $f(v)$ is square-free.*

(II) *For every $a, b, c \in \Sigma$, and every $A \in f(a)$, $B \in f(b)$, and $C \in f(c)$:*
   (i) *If A is a factor of B, then $a = b$ and $A = B$.*
   (ii) *If $AB = pCs$ for some words $p, s \in \Delta^*$, then $p = \varepsilon$ or $s = \varepsilon$.*
   (iii) *If $A = A'A''$, $B = B'B''$, and $C = A'B''$, then $c = a$ or $c = b$.*

Proof. Suppose towards a contradiction that conditions (I) and (II) are satisfied, but that some word in $f(u)$ contains a square. Let $w = a_1 a_2 \cdots a_n$ be a minimal factor of u such that some word $W = A_1 A_2 \cdots A_n$ contains a square, where $A_i \in f(a_i)$ for all $i \in \{1, \ldots, n\}$. Write $W = XYYZ$. By the minimality of w, the word X must be a proper prefix of A_1, and the word Z must be a proper suffix of A_n. By condition (I), we have $n \geq 4$.

Suppose that $n = 4$. Then we can write

$$W = A_1' A_1'' A_2' A_2'' A_3 A_4' A_4'',$$

where $A_i = A_i' A_i''$ for all $i \in \{1, 2, 4\}$, and $Y = A_1'' A_2' = A_2'' A_3 A_4'$, or we can write

$$W = A_1' A_1'' A_2 A_3' A_3'' A_4' A_4'',$$

where $A_i = A_i' A_i''$ for all $i \in \{1, 3, 4\}$, and $Y = A_1'' A_2 A_3' = A_3'' A_4'$. Assume the former. (The latter is handled similarly.) Then the word A_3 is a factor of $A_1'' A_2'$, and hence of $A_1 A_2$. By condition (II)(ii), we must have either $A_1' = A_2'' = \varepsilon$, or $A_4' = \varepsilon$. Since $A_4' = \varepsilon$ is impossible by the minimality of w, we must have $A_1' = A_2'' = \varepsilon$. Then $A_1 A_2 = A_3 A_4'$, and hence either A_1 is a factor of A_3, or vice versa. By condition (II)(i), we must have $a_1 = a_3$ and $A_1 = A_3$. It follows that A_2 is a factor of A_4', and hence $a_2 = a_4$. But then w contains the square $(a_1 a_2)^2$, an impossibility.

So we may assume that $n \geq 5$. For some $j \in \{2, \ldots, n-1\}$, we can write

$$W = A_1 A_2 \cdots A_n = A_1' A_1'' A_2 \cdots A_{j-1} A_j' A_j'' A_{j+1} \cdots A_{n-1} A_n' A_n'', \tag{2.1}$$

where $A_i = A_i' A_i''$ for all $i \in \{1, j, n\}$, $X = A_1'$, $Z = A_n''$, and

$$Y = A_1'' A_2 \cdots A_{j-1} A_j' = A_j'' A_{j+1} \cdots A_{n-1} A_n'.$$

By the minimality of w, we must have $|A_1''|, |A_n'| > 0$, and we may assume without loss of generality that $|A_j''| > 0$.

Suppose that $|A_1''| > |A_j''|$. If $j = 2$, then $A_1'' A_2' = A_2'' A_3 \cdots A_{n-1} A_n'$. But since $n \geq 5$, we see that either A_3 must be a proper factor of A_1'', or A_4 must be a proper factor of A_2'. By condition (II)(i), this is impossible. So we may assume that $j > 2$. By condition (II)(i), we have that A_2 is not a factor of A_{j+1}, so A_{j+1} must be a factor of $A_1'' A_2$. In particular, this implies that $j + 1 < n$. Write $A_{j+2} = A_{j+2}' A_{j+2}''$ so that $A_1' A_2 = A_j'' A_{j+1} A_{j+2}'$. By condition (II)(ii), we must have either $|A_j''| = 0$ or $|A_{j+2}'| = 0$. Since $|A_j''| > 0$ by assumption, we must have $|A_{j+2}'| = 0$. By condition (II)(i), we must have $A_2 = A_{j+1}$, and hence $A_1'' = A_j''$. This contradicts the assumption that $|A_1''| > |A_j''|$.

Now suppose that $|A_j''| > |A_1''|$. If $j = n - 1$, then $A_1'' A_2 \cdots A_{n-2} A_{n-1}' = A_{n-1}'' A_n'$. But since $n \geq 5$, we see that either A_2 must be a proper factor of A''_{n-1}, or A_3 must be a proper factor of A'_n. By condition (II)(i), this is impossible. So we may assume that $j < n-1$. By condition (II)(i), we have that A_{j+1} is not a factor of A_2, so A_2 must be a factor of $A''_j A_{j+1}$. Write $A_3 = A'_3 A''_3$, where $A''_1 A_2 A'_3 = A''_j A_{j+1}$. By condition (II)(ii), we must have $|A''_1| = 0$ or $|A'_3| = 0$. Since $|A''_1| > 0$ by assumption, we must have $|A'_3| = 0$. By condition (II)(i), we must have $A_2 = A_{j+1}$, and hence $A''_1 = A''_j$. This contradicts the assumption that $|A''_j| > |A''_1|$.

So we may assume that $|A''_1| = |A''_j|$, and hence $A''_1 = A''_j$. Then either A_2 is a factor of A_{j+1}, or vice versa. By condition (II)(i), we conclude that $a_2 = a_{j+1}$ and $A_2 = A_{j+1}$. Applying this argument repeatedly, we find $a_2 a_3 \cdots a_{j-1} = a_{j+1} a_{j+2} \cdots a_{n-1}$, and $A_2 A_3 \cdots A_{j-1} = A_{j+1} A_{j+2} \cdots A_{n-1}$. Finally, we see that $A'_j = A'_n$. But then $A_j = A'_j A''_j = A'_n A''_1$. By condition (II)(iii), we conclude that $a_j = a_1$ or $a_j = a_n$. But then the word u contains either the square $(a_1 a_2 \cdots a_{j-1})^2$ or the square $(a_2 a_3 \cdots a_j)^2$, respectively. $\square$

3. Extremal square-free words

In this section, we prove Theorem 1.1. We first summarize the method of Grytczuk et al. [7] used to construct arbitrarily long extremal square-free ternary words. The proof of Theorem 1.1 is obtained by a similar method. We first introduce some notation and terminology.

Let w be a word over a fixed alphabet Σ. We say that w is *nearly extremal square-free* if w is square-free, and there are only two square-free extensions of w; one left extension, and one right extension. We say that w is *left (right) extremal square-free* if every square-free extension of w is a right (left, respectively) extension.

Let $\mathbb{S}_\Gamma$ denote the symmetric group on the alphabet $\Gamma = \{\mathtt{a}, \mathtt{b}, \mathtt{c}\}$. We represent the permutations of $\mathbb{S}_\Gamma$ using cycle notation, but for ease of notation, we omit the commas. We denote the identity permutation by the empty cycle $()$. We treat every member of $\mathbb{S}_\Gamma$ as both a permutation, and as a letter, depending on context. For every permutation $\pi \in \mathbb{S}_\Gamma$, we let $\tilde{\pi}$ be another letter, which we refer to as the *mirror image* of π. Let $\tilde{\mathbb{S}}_\Gamma = \{\tilde{\pi} \colon\ \pi \in \mathbb{S}_\Gamma\}$. Since every permutation of Γ can be written as either a single nontrivial cycle or the empty cycle, the parentheses serve as delimiters for letters in words over the alphabet $\mathbb{S}_\Gamma \cup \tilde{\mathbb{S}}_\Gamma$.

Let D be the digraph with vertex set $V(D) = \mathbb{S}_\Gamma \cup \tilde{\mathbb{S}}_\Gamma$ that is shown in Figure 1. Let

$$N = \mathtt{abacbabcabacbcacbabcabacabcbabcabacbcabcb}.$$

It is easily checked by computer that the word N is nearly extremal square-free. Now for every permutation $\pi \in \mathbb{S}_\Gamma$, let N_π denote the word obtained by permuting the letters of N by π, and let $N_{\tilde{\pi}}$ denote the reversal of N_π. Define the morphism $f : V(D)^* \to \Gamma^*$ by $f(x) = N_x$ for all $x \in V(D)$.

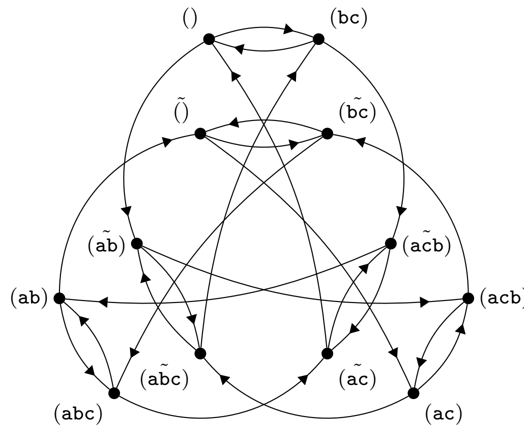

FIGURE 1. The Digraph D.

First of all, Grytczuk et al. show that if w is a square-free walk in D (where walks are treated as words over the vertex set), then the word $f(w)$ is square-free. Next, they show that there are arbitrarily long square-free walks in D that begin and end at the vertex (). It follows that there are arbitrarily long nearly extremal square-free words over Γ that have N as both a prefix and a suffix. Finally, Grytczuk et al. provide two short words that can be added to the beginning and the end of any such word to form an extremal square-free word.

In order to prove Theorem 1.1, we replace the morphism f with a multi-valued substitution δ. Using Lemma 2.1, we can then construct nearly extremal square-free words of every sufficiently large length. Finally, we find words of a single fixed length that can be added to the beginning and end of every such nearly extremal square-free word to form an extremal square-free word. This guarantees the existence of extremal square-free words of every sufficiently large length, and the smaller lengths are handled computationally.

Let

$$P = \mathtt{abacbcabcbacabacbcabcbabcacbcabcbacabacbcabcbacbc},$$

$$Q = \mathtt{abacbabcacbacabacbcacbacabcbabcabacbcabcb},$$

$$R = \mathtt{abacabcacbacabcbabcacbacabacbcacbacabcbabcabacbcabcb}, \text{ and}$$

$$S = \mathtt{acabacbabcacbacabcbacbcabacbabcacbacabcbabcacbaca}.$$

We note that the words P and S have length 49, the word Q has length 41, and the word R has length 52. The word R can be obtained from Q by inserting the word abcacbacabc after the 4th letter. By computer check, the words Q and R are both nearly extremal square-free, the words PQ and PR are left extremal square-free, and the words QS and RS are right extremal square-free. Although the word Q has the same length as the word

N used by Grytczuk et al., we note that Q cannot be obtained from N by permutation of the alphabet and/or reversal, i.e., for every $x \in V(D)$, we have $N_x \neq Q$.

For every letter $x \in V(D)$, let p_x and s_x be two new letters. Let $\hat{D}$ be the graph obtained from D by adding, for every $x \in V(D)$, the vertices p_x and s_x, as well as arcs from p_x to x and from x to s_x. For every $x \in V(D)$, define the words P_x, Q_x, R_x, and S_x analogously to N_x. Now define the substitution $\delta : V(\hat{D})^* \to 2^{\Gamma^*}$ by

- $\delta(x) = \{Q_x, R_x\}$ for all $x \in \mathbb{S}_\Gamma \cup \tilde{\mathbb{S}}_\Gamma$;
- $\delta(p_x) = \{P_x\}$ and $\delta(s_x) = \{S_x\}$ for all $x \in \mathbb{S}_\Gamma$; and
- $\delta(p_x) = \{S_x\}$ and $\delta(s_x) = \{P_x\}$ for all $x \in \tilde{\mathbb{S}}_\Gamma$.

Since Q and R are nearly extremal square-free, it follows immediately that Q_x and R_x are nearly extremal square-free for every $x \in V(D)$. We also have the following fact.

Lemma 3.1. *For every $x \in V(D)$, every word in the set $\delta(p_x x)$ is left extremal square-free, and every word in the set $\delta(x s_x)$ is right extremal square-free.*

Proof. Let $x \in V(D)$. We show that every word in the set $\delta(p_x x)$ is left extremal square-free. The proof that every word in the set $\delta(x s_x)$ is right extremal square-free is similar. If $x \in \mathbb{S}_\Gamma$, then $\delta(p_x x) = \{P_x Q_x, P_x R_x\}$. Since the words PQ and PR are left extremal square-free, so are $P_x Q_x$ and $P_x R_x$. On the other hand, if $x \in \tilde{\mathbb{S}}_\Gamma$, then write $x = \tilde{\pi}$. Then $\delta(p_x x) = \{S_{\tilde{\pi}} Q_{\tilde{\pi}}, S_{\tilde{\pi}} R_{\tilde{\pi}}\}$. Note that $S_{\tilde{\pi}} Q_{\tilde{\pi}}$ is the reversal of the word $Q_\pi S_\pi$. Since QS is right extremal square-free, so is the word $Q_\pi S_\pi$. It follows that the word $S_{\tilde{\pi}} Q_{\tilde{\pi}}$ is left extremal square-free. The proof that $S_{\tilde{\pi}} R_{\tilde{\pi}}$ is left extremal square-free is analogous. □

Using Theorem 2.3, the next lemma can be verified by a computer check. (We check condition (I) for all square-free walks of length 3 in $\hat{D}$.)

Lemma 3.2. *If w is a square-free walk in the digraph $\hat{D}$, then the set $\delta(w)$ is square-free.*

Together, Lemma 3.1 and Lemma 3.2 yield the following corollary.

Corollary 3.3. *Let w be a square-free walk in D of length at least 2, and write $w = xw'y$, where $x, y \in V(D)$. Then every word in the set $\delta(p_x w s_y)$ is extremal square-free.*

Proof. Since w is a square-free walk in D (and not $\hat{D}$), it contains neither p_x nor s_y, and hence the walk $p_x w s_y$ is square-free. By Lemma 3.2, the set $\delta(p_x w s_y)$ is square-free. Note that for every $z \in V(D)$, both words in the set $\delta(z)$, namely Q_z and R_z, are nearly extremal square-free. It follows easily that every word in the set $\delta(w)$ is nearly extremal square-free. By Lemma 3.1, every word in $\delta(p_x x)$ is left extremal square-free, and every word in $\delta(ys_y)$ is right extremal square-free. It follows that every word in the set $\delta(p_x w s_y)$ is extremal square-free. $\square$

We are now ready to prove Theorem 1.1.

Proof of Theorem 1.1. If $n \notin \mathcal{A}$, then we used a standard backtracking algorithm to show that there is no extremal square-free ternary word of length n.

Suppose otherwise that $n \in \mathcal{A}$. Suppose first that $n \geq 2138$. Then $n-98 \geq 2040$, and by Lemma 2.1, we can write $n-98 = 41a+52b$ for some nonnegative integers a and b. Let w be a square-free walk in D of length $a+b$, and write $w = xw'y$ for some $x, y \in V(D)$. Since $\delta(z)$ contains a word of length 41 and a word of length 52 for every $z \in V(D)$, it is evident that there is a word W of length $n-98$ in $\delta(w)$. Now the unique word W' in $\delta(p_x)W\delta(s_y)$ has length n, and by Corollary 3.3, the word W' is extremal square-free.

So we may assume that $n < 2138$. In this case, we found an extremal square-free circular word of length n by computer search. $\square$

4. Extremal square-free circular words

In this section, we prove Theorem 1.3. A *circumnavigation* of a circular word $\langle w\rangle$ is a linear word of the form ava, where a is a letter, and av is a conjugate of w. We begin with an elementary lemma about the circumnavigations of square-free circular words.

Lemma 4.1. *Let $w \in \Sigma^*$ be a word of length at least 2. If $\langle w\rangle$ is square-free, then every circumnavigation of $\langle w\rangle$ is square-free.*

Proof. Let u be a circumnavigation of $\langle w\rangle$. Since $|w| \geq 2$, we have $|u| \geq 3$. Suppose towards a contradiction that u contains a square. Write $u = ava$, where a is a letter, and av is a conjugate of w. Then va is also a conjugate of w. Since $\langle w\rangle$ is square-free, neither conjugate av nor va contains a square. Hence, it must be the case that $u = xx$ for some word x. Evidently, the word x begins and ends in a, and has length at least 2, since $|u| \geq 3$. Write $x = ax'a$ for some word x'. Then $u = ax'aax'a$, and we conclude that $\langle w\rangle$ contains the square aa, a contradiction. $\square$

Define the substitution $h\colon \{\mathtt{0},\mathtt{1},\mathtt{2}\}^* \to 2^{V(D)^*}$, where D is the digraph shown in Figure 1, by

$$\mathtt{0} \to \Big\{()(\tilde{\mathtt{ab}})(\mathtt{acb})(\mathtt{ac})(\tilde{\mathtt{abc}})(\mathtt{bc})\Big\}$$

$$\mathtt{1} \to \Big\{()(\tilde{\mathtt{ab}})(\tilde{\mathtt{abc}})(\mathtt{bc})(\tilde{\mathtt{acb}})(\tilde{\mathtt{ac}})\Big\}$$

$$\mathtt{2} \to \Big\{()(\tilde{\mathtt{ab}})(\mathtt{acb})(\tilde{\mathtt{bc}})(\mathtt{abc})(\tilde{\mathtt{ac}}), ()(\tilde{\mathtt{ab}})(\mathtt{acb})(\tilde{\mathtt{bc}})(\mathtt{abc})(\mathtt{ab})(\mathtt{abc})(\tilde{\mathtt{ac}})\Big\}.$$

Let $w = w_0 w_1 \cdots w_{n-1}$ be a word over the alphabet $V(D)$, where the w_i's are letters. We say that the circular word $\langle w\rangle$ is *walkable* in D if every conjugate of w is a valid walk in D. Equivalently, the circular word $\langle w\rangle$ is walkable in D if there is an arc from w_i to w_{i+1} for every $i \in \{0,1,\dots,n-1\}$, with indices taken modulo n.

Lemma 4.2. *Let $\langle w\rangle$ be a square-free circular word of length at least 2 over the alphabet $\{\mathtt{0},\mathtt{1},\mathtt{2}\}$, and let $W \in h(w)$. Then the circular word $\langle W\rangle$ is square-free and walkable in the digraph D.*

Proof. We first show that $\langle W\rangle$ is square-free. Let U be a conjugate of W. Then U is a factor of the set $h(u)$ for some circumnavigation u of w. By Lemma 4.1, the circumnavigation u is square-free. Using Theorem 2.3, we verify by computer that $h(u)$ is square-free. (We check condition (I) for every square-free word $v \in \{\mathtt{0},\mathtt{1},\mathtt{2}\}^*$ of length 3.) We conclude that the word U is square-free. Since U was an arbitrary conjugate of W, we conclude that the circular word $\langle W\rangle$ is square-free.

It remains to show that $\langle W\rangle$ is walkable in D. Note that every word $A \in h(\{0,1,2\})$ begins in the identity permutation (). So it suffices to check that for all $A \in h(\{0,1,2\})$, the word $A()$ is walkable in D, and this is easily done by inspection. □

Lemma 4.3. *For every even positive integer n, there is a square-free circular word of length n that is walkable in the digraph D.*

Proof. Let n be an even positive integer. First suppose that $n \geq 6\cdot 18 = 108$. Then we may write $n = 6m + r$ for some $m \geq 18$ and $r \in \{0,2,4\}$. By Theorem 1.2, there is a circular square-free word $\langle u\rangle$ of length m over the alphabet $\{\mathtt{0},\mathtt{1},\mathtt{2}\}$. Since every square-free word on $\{\mathtt{0},\mathtt{1}\}^*$ has length at most 3, we must have $|u|_{\mathtt{2}} \geq 2$. Note that for every $a \in \{\mathtt{0},\mathtt{1},\mathtt{2}\}$, there is a word in $h(a)$ of length 6. Further, the set $h(\mathtt{2})$ contains a word of length 8. Thus, there is a word $U \in h(u)$ of length $n = 6m + r$, obtained by using the word of length 8 in $h(\mathtt{2})$ exactly 0, 1, or 2 times (for r equal to 0, 2, or 4, respectively). By Lemma 4.2, the circular word $\langle U\rangle$ is square-free, and is walkable in the digraph D.

Now we may assume that $n < 108$, and in this case we verify the statement by means of a computer search. □

Let

$$Q' = \mathtt{abacabcacbacabcbabcacbacabacbcacbacabcbacbcabacbabcabacbcabcb}, \text{ and}$$
$$R' = \mathtt{abacabcacbacabcbabcacbacabacbcacbacabcbabcacbcabacbabcabacbcabcb}.$$

Note that Q' has length 61, and R' has length 64. The word Q' can be obtained from the word R by adding the factor $\mathtt{cbcabacba}$ after the 40th letter, and the word R' can be obtained from the word Q' by adding the factor $\mathtt{bca}$ after the 40th letter. For every $x \in \mathbb{S}_\Gamma \cup \tilde{\mathbb{S}}_\Gamma$, define Q'_x and R'_x analogously to N_x. Define the substitution $\delta' : V(D)^* \to 2^{\Gamma^*}$ by $\delta'(x) = \{Q_x, R_x, Q'_x, R'_x\}$ for all $x \in \mathbb{S}_\Gamma \cup \tilde{\mathbb{S}}_\Gamma$. So for every $x \in V(D)$, the set $\delta'(x)$ contains a word of length m for all $m \in \{41, 52, 61, 64\}$. The proof of the following lemma is analogous to the first paragraph of the proof of Lemma 4.2.

Lemma 4.4. *Let $\langle w \rangle$ be a square-free circular word that is walkable in the digraph D, and let $W \in \delta'(w)$. Then the circular word $\langle W \rangle$ is square-free.*

We are now ready to prove Theorem 1.3.

Proof of Theorem 1.3. If $n \notin \mathcal{B}$, then we used a standard backtracking algorithm to show that there is no extremal square-free ternary circular word of length n.

Suppose otherwise that $n \in \mathcal{B}$. First suppose that $n \geq 470$. Then there are nonnegative integers a, b, c, and d such that $41a + 52b + 61c + 64d = n$, and the sum $a + b + c + d$ is even. (For $n \geq 4172$, this claim follows from Lemma 2.2, and we verified the remaining cases by computer.) By Lemma 4.3, there is a square-free circular word $\langle w \rangle$ of length $a + b + c + d$ that is walkable in the digraph D. Evidently, there is a word W in $\delta'(w)$ of length n. By Lemma 4.4, the circular word $\langle W \rangle$ is square-free. Since all words in the set $\delta'(x)$ are nearly extremal square-free for every $x \in V(D)$, it follows that the circular word $\langle W \rangle$ is extremal square-free.

So we may assume that $n < 470$. In this case, we found an extremal square-free circular word of length n by computer search. □

Note that the proof of Theorem 1.3 presented in this section can be adapted to provide yet another alternate proof of Theorem 1.2. (Note that while Theorem 1.2 was used in the proof of Lemma 4.3, an inductive argument could be used instead.)

5. Conclusion

We have completely described the attainable lengths of extremal square-free ternary words and extremal square-free ternary circular words. It is well-known that the number of square-free ternary words of length n grows exponentially in n [4]; currently, the best-known bounds on the growth rate are due to Shur [15]. It is also known that the number of square-free ternary circular words of length n grows exponentially in n [14]. Using these results together with the results of this paper, one can show that both the number of extremal square-free ternary words of length n, and the number of extremal square-free ternary circular words of length n, grow exponentially in n.

Surprisingly, over larger alphabets, it appears that there are no extremal square-free words. The following is a minor variant of a conjecture of Grytczuk et al. [7, Conjecture 2].

Conjecture. *Let Σ be a fixed alphabet of size at least 4. Then there are no extremal square-free words over Σ.*

In other words, Conjecture 1 says that every square-free word over an alphabet Σ of size greater than 3 has at least one square-free extension over Σ. While it would be most interesting to establish Conjecture 1 in the case $|\Sigma| = 4$, establishing the conjecture for larger alphabets would also be interesting.

Finally, we note that Harju [8] recently introduced the related notion of *irreducibly square-free words*; these are square-free words in which the removal of any interior letter produces a square. In particular, Harju demonstrated that there are irreducibly square-free ternary words of every sufficiently large length, just as we have shown for extremal square-free ternary words. However, the situation appears to be quite different over larger alphabets. Let $n \geq 4$, and let $\Sigma_n = \{\mathtt{1}, \mathtt{2}, \ldots, \mathtt{n}\}$. For every $3 \leq k \leq n$, define $u_k = \mathtt{k2k}$. Then define

$$w_n = \mathtt{121}u_3\mathtt{121}u_4 \cdots \mathtt{121}u_n\mathtt{121} = \mathtt{121323121424} \cdots \mathtt{121n2n121}.$$

It is straightforward to verify that w_n is irreducibly square-free. So there are irreducibly square-free words over any fixed alphabet, while Conjecture 1 suggests that this is not the case for extremal square-free words.

References

1. D. R. Bean, A. Ehrenfeucht, and G. F. McNulty, *Avoidable patterns in strings of symbols*, Pacific J. Math. **85** (1979), 261–294.
2. J. Berstel, *Axel Thue's papers on repetitions in words: A translation*, Publications du LaCIM (Université du Québec à Montréal), vol. 20, 1995.
3. J. Berstel and D. Perrin, *The origins of combinatorics on words*, European J. Combin. **28** (2007), 996–1022.
4. F. J. Brandenburg, *Uniformly growing k-th power-free homomorphisms*, Theoret. Comput. Sci. **23** (1983), 69–82.
5. J. D. Currie, *There are ternary circular square-free words of length n for $n \geq 18$*, Electron. J. Combin. **9** (2002), #N10.
6. T. Clokie, D. Gabric, and J. Shallit, *Circularly squarefree words and unbordered conjugates: A new approach*, in Proc. WORDS 2019, Lect. Notes in Comput. Sci., Vol. 11682, Springer, 2019, pp. 133–144.
7. J. Grytczuk, H. Kordulewski, and A. Niewiadomski, *Extremal square-free words*, Electron. J. Combin. **27**, 2020, #P1.48.
8. T. Harju, *Disposability in square-free words*, Theor. Comput. Sci. (2020). Available at `https://doi.org/10.1016/j.tcs.2020.07.030`.
9. J. Karhumäki and J. Shallit, *Polynomial versus exponential growth in repetition-free binary words*, J. Combin. Theory Ser. A **105** (2004), 335–347.
10. M. Lothaire, *Algebraic Combinatorics on Words*, Encyclopedia of Mathematics and Its Applications, vol. 90, Cambridge University Press, 2002.
11. J. L. Ramírez Alfonsín, *The Diophantine Frobenius Problem*, Oxford Lecture Series in Mathematics and Its Applications, vol. 30, Oxford University Press, 2005.
12. N. Rampersad and J. Shallit, *Repetitions in words*, in V. Berthé and M. Rigo, eds., *Combinatorics, Words and Symbolic Dynamics*, Encyclopedia of Mathematics and Its Applications, vol. 159, Cambridge University Press, 2016, pp. 101–150.
13. J. Shallit, personal communication, October 2019.
14. A. M. Shur, *On ternary square-free circular words*, Electron. J. Combin. **17** (2010), #R140.
15. ———, *Growth properties of power-free languages*, Comput. Sci. Rev. **6** (2012), 187–208.

Department of Mathematics and Statistics, The University of Winnipeg

515 Portage Ave. Winnipeg, MB, Canada R3B 2E9
E-mail address: `l.mol@uwinnipeg.ca`

Department of Mathematics and Statistics, The University of Winnipeg
515 Portage Ave. Winnipeg, MB, Canada R3B 2E9
E-mail address: `n.rampersad@uwinnipeg.ca`