Extremal Square-Free Words and Variations

Lucas Mol

THE UNIVERSITY OF WINNIPEG

Including joint work with Narad Rampersad and Jeffrey Shallit

One World Combinatorics on Words Seminar
June 29, 2020

SQUARES AND OVERLAPS

- A *square* is a word of the form *xx*, where *x* is a nonempty word.
  - Examples: `murmur`, `hotshots`
- An *overlap* is a word of the form *axaxa*, where *a* is a letter and *x* is a (possibly empty) word.
  - Example: `alfalfa`
- A word is square-free (overlap-free) if it contains no squares (overlaps) as factors.
  - The word `apple` is not square-free, but it is overlap-free.
  - The word `banana` is not overlap-free.
  - The word `clementine` is square-free.

The origin of combinatorics on words

Axel Thue
(1863-1922)

- There are arbitrarily long overlap-free binary words.
- There are arbitrarily long square-free ternary words.

EXTENSIONS OF A WORD

- Let w be a word over a fixed alphabet Σ.
- An *extension* of w is a word of the form $w_1 a w_2$, where $a \in \Sigma$, and $w = w_1 w_2$.
  - Note that w_1 and w_2 could be empty!
- Some extensions of the English word `pans` are:

`spans`   `plans`   `pawns`   `pants`   `pansy`

left extension   internal extensions   right extension

- Question: Given a square-free word w over a fixed alphabet, can we extend w to a longer square-free word? Are there square-free words that can't be extended?
  - One could replace "square-free" with any property of interest.

EXTREMAL SQUARE-FREE WORDS

- A square-free word *w* is called *extremal* if every extension of *w* contains a square.
- Introduced by Grytczuk, Kordulewski, and Niewiadomski (2020).
- Over the binary alphabet $\{\mathtt{a}, \mathtt{b}\}$, only the words `aba` and `bab` are extremal square-free.
- Over the ternary alphabet $\{\mathtt{a}, \mathtt{b}, \mathtt{c}\}$, the following is a shortest extremal square-free word:

$$\mathtt{abcabacbcabcbabcabacbcabc}$$

- Again, one can replace "square-free" with any property of interest.

EXTREMAL SQUARE-FREE WORDS

- A square-free word w is called *extremal* if every extension of w contains a square.
- Introduced by Grytczuk, Kordulewski, and Niewiadomski (2020).
- Over the binary alphabet $\{\mathtt{a}, \mathtt{b}\}$, only the words aba and bab are extremal square-free.
- Over the ternary alphabet $\{\mathtt{a}, \mathtt{b}, \mathtt{c}\}$, the following is a shortest extremal square-free word:

  aabcabacbcabcbabcabacbcabc

- Again, one can replace "square-free" with any property of interest.

EXTREMAL SQUARE-FREE WORDS

- A square-free word w is called *extremal* if every extension of w contains a square.
- Introduced by Grytczuk, Kordulewski, and Niewiadomski (2020).
- Over the binary alphabet $\{\mathtt{a}, \mathtt{b}\}$, only the words `aba` and `bab` are extremal square-free.
- Over the ternary alphabet $\{\mathtt{a}, \mathtt{b}, \mathtt{c}\}$, the following is a shortest extremal square-free word:

$$\color{red}{\mathtt{b}}\mathtt{abcabacbcabcbabcabacbcabc}$$

- Again, one can replace "square-free" with any property of interest.

EXTREMAL SQUARE-FREE WORDS

- A square-free word *w* is called *extremal* if every extension of *w* contains a square.
- Introduced by Grytczuk, Kordulewski, and Niewiadomski (2020).
- Over the binary alphabet $\{\mathtt{a}, \mathtt{b}\}$, only the words $\mathtt{aba}$ and $\mathtt{bab}$ are extremal square-free.
- Over the ternary alphabet $\{\mathtt{a}, \mathtt{b}, \mathtt{c}\}$, the following is a shortest extremal square-free word:

$$\underbrace{\mathtt{babcabacbcabc}}\underbrace{\mathtt{babcabacbcabc}}$$

- Again, one can replace “square-free” with any property of interest.

Extremal square-free words

- A square-free word w is called *extremal* if every extension of w contains a square.
- Introduced by Grytczuk, Kordulewski, and Niewiadomski (2020).
- Over the binary alphabet $\{\mathtt{a}, \mathtt{b}\}$, only the words `aba` and `bab` are extremal square-free.
- Over the ternary alphabet $\{\mathtt{a}, \mathtt{b}, \mathtt{c}\}$, the following is a shortest extremal square-free word:

$$\textcolor{red}{\mathtt{c}}\mathtt{abcabacbcabcbabcabacbcabc}$$

- Again, one can replace "square-free" with any property of interest.

EXTREMAL SQUARE-FREE WORDS

- A square-free word *w* is called *extremal* if every extension of *w* contains a square.
- Introduced by Grytczuk, Kordulewski, and Niewiadomski (2020).
- Over the binary alphabet $\{\mathtt{a}, \mathtt{b}\}$, only the words `aba` and `bab` are extremal square-free.
- Over the ternary alphabet $\{\mathtt{a}, \mathtt{b}, \mathtt{c}\}$, the following is a shortest extremal square-free word:

$$\underbrace{\mathtt{cab}}\underbrace{\mathtt{cab}}\mathtt{acbcabcbabcabacbcabc}$$

- Again, one can replace “square-free” with any property of interest.

EXTREMAL SQUARE-FREE WORDS

- A square-free word *w* is called *extremal* if every extension of *w* contains a square.
- Introduced by Grytczuk, Kordulewski, and Niewiadomski (2020).
- Over the binary alphabet $\{\mathtt{a}, \mathtt{b}\}$, only the words `aba` and `bab` are extremal square-free.
- Over the ternary alphabet $\{\mathtt{a}, \mathtt{b}, \mathtt{c}\}$, the following is a shortest extremal square-free word:

`aabcabacbcabcbabcabacbcabc`

- Again, one can replace "square-free" with any property of interest.

EXTREMAL SQUARE-FREE WORDS

- A square-free word w is called *extremal* if every extension of w contains a square.
- Introduced by Grytczuk, Kordulewski, and Niewiadomski (2020).
- Over the binary alphabet $\{\mathtt{a}, \mathtt{b}\}$, only the words `aba` and `bab` are extremal square-free.
- Over the ternary alphabet $\{\mathtt{a}, \mathtt{b}, \mathtt{c}\}$, the following is a shortest extremal square-free word:

$$\mathtt{abbcabacbcabcbabcabacbcabc}$$

- Again, one can replace "square-free" with any property of interest.

Extremal square-free words

- A square-free word w is called *extremal* if every extension of w contains a square.
- Introduced by Grytczuk, Kordulewski, and Niewiadomski (2020).
- Over the binary alphabet $\{\mathtt{a}, \mathtt{b}\}$, only the words `aba` and `bab` are extremal square-free.
- Over the ternary alphabet $\{\mathtt{a}, \mathtt{b}, \mathtt{c}\}$, the following is a shortest extremal square-free word:

  `acbcabacbcabcbabcabacbcabc`

- Again, one can replace "square-free" with any property of interest.

EXTREMAL SQUARE-FREE WORDS

- A square-free word *w* is called *extremal* if every extension of *w* contains a square.
- Introduced by Grytczuk, Kordulewski, and Niewiadomski (2020).
- Over the binary alphabet $\{\mathtt{a}, \mathtt{b}\}$, only the words `aba` and `bab` are extremal square-free.
- Over the ternary alphabet $\{\mathtt{a}, \mathtt{b}, \mathtt{c}\}$, the following is a shortest extremal square-free word:

$$\underbrace{\mathtt{acbcab}}\,\underbrace{\mathtt{acbcab}}\,\mathtt{cbabcabacbcabc}$$

- Again, one can replace "square-free" with any property of interest.

EXTREMAL SQUARE-FREE WORDS

- A square-free word w is called *extremal* if every extension of w contains a square.
- Introduced by Grytczuk, Kordulewski, and Niewiadomski (2020).
- Over the binary alphabet $\{\mathtt{a}, \mathtt{b}\}$, only the words `aba` and `bab` are extremal square-free.
- Over the ternary alphabet $\{\mathtt{a}, \mathtt{b}, \mathtt{c}\}$, the following is a shortest extremal square-free word:

$$\mathtt{abacabacbcabcbabcabacbcabc}$$

- Again, one can replace "square-free" with any property of interest.

EXTREMAL SQUARE-FREE WORDS

- A square-free word *w* is called *extremal* if every extension of *w* contains a square.
- Introduced by Grytczuk, Kordulewski, and Niewiadomski (2020).
- Over the binary alphabet $\{\mathtt{a}, \mathtt{b}\}$, only the words $\mathtt{aba}$ and $\mathtt{bab}$ are extremal square-free.
- Over the ternary alphabet $\{\mathtt{a}, \mathtt{b}, \mathtt{c}\}$, the following is a shortest extremal square-free word:

$$\underbrace{\mathtt{abac}}\,\underbrace{\mathtt{abac}}\,\mathtt{bcabcbabcabacbcabc}$$

- Again, one can replace “square-free” with any property of interest.

Extremal square-free words

- A square-free word w is called *extremal* if every extension of w contains a square.
- Introduced by Grytczuk, Kordulewski, and Niewiadomski (2020).
- Over the binary alphabet $\{\mathtt{a}, \mathtt{b}\}$, only the words `aba` and `bab` are extremal square-free.
- Over the ternary alphabet $\{\mathtt{a}, \mathtt{b}, \mathtt{c}\}$, the following is a shortest extremal square-free word:

  `abbcabacbcabcbabcabacbcabc`

- Again, one can replace “square-free” with any property of interest.

Extremal square-free words

- A square-free word w is called *extremal* if every extension of w contains a square.
- Introduced by Grytczuk, Kordulewski, and Niewiadomski (2020).
- Over the binary alphabet $\{\mathtt{a}, \mathtt{b}\}$, only the words `aba` and `bab` are extremal square-free.
- Over the ternary alphabet $\{\mathtt{a}, \mathtt{b}, \mathtt{c}\}$, the following is a shortest extremal square-free word:

$$\mathtt{abccabacbcabcbabcabacbcabc}$$

- Again, one can replace “square-free” with any property of interest.

EXTREMAL SQUARE-FREE WORDS

- A square-free word *w* is called *extremal* if every extension of *w* contains a square.
- Introduced by Grytczuk, Kordulewski, and Niewiadomski (2020).
- Over the binary alphabet $\{\mathtt{a}, \mathtt{b}\}$, only the words `aba` and `bab` are extremal square-free.
- Over the ternary alphabet $\{\mathtt{a}, \mathtt{b}, \mathtt{c}\}$, the following is a shortest extremal square-free word:

$$\mathtt{abcabacbcabcbabcabacbcabc}$$

- Again, one can replace "square-free" with any property of interest.

Plan

Extremal Square-Free Words

Extremal Overlap-Free Words

Extremal β-free words

EXTREMAL SQUARE-FREE WORDS

1 None
2 None
3 None
4 None
5 None
6 None
7 None
8 None
9 None
10 None
11 None
12 None
13 None
14 None
15 None
16 None
17 None
18 None
19 None
20 None
21 None
22 None
23 None
24 None
25 0120102120121012010212012
26 None
27 None
28 None
29 None
30 None
31 None
32 None
33 None
34 None
35 None
36 None
37 None
38 None
39 None
40 None
41 01021012021020121021201021012021020121021
42 None
43 None
44 None
45 None
46 None
47 None
48 010212012102010212012101202120121020102120121020
49 None
50 01021201021012021020121012021201021012021020121020
51 None
52 None
53 None
54 None
55 None
56 None
57 None
58 None
59 None
60 None

Theorem (GKN 2020): There are arbitrarily long extremal square-free ternary words.

Theorem (GKN 2020): There are arbitrarily long extremal square-free ternary words.

Square-Free Lemma (GKN 2020): Let $f : \Sigma^* \to \Delta^*$ be a morphism, and let $u \in \Sigma^*$ be a square-free word. Then the word $f(u)$ is square-free if all of the following are satisfied:

1. For every factor v of u of length at most 3, the word $f(v)$ is square-free.
2. For every $a, b, c \in \Sigma$:
   - If $f(a)$ is a factor of $f(b)$, then $a = b$.
   - If $f(a)f(b) = pf(c)s$ for some words $p, s \in \Delta^*$, then $p = \varepsilon$ or $s = \varepsilon$.
   - If $f(a) = A'A''$, $f(b) = B'B''$, and $f(c) = A'B''$, then $c = a$ or $c = b$.

Building blocks

- A square-free word is *nearly extremal* if it has at most two square-free extensions; one left extension, and one right extension.

Lemma: If u and v are nearly extremal square-free words over $\{\mathrm{a}, \mathrm{b}, \mathrm{c}\}$, and uv is square-free, then uv is nearly extremal square-free.
Proof: It suffices to show that adding a letter between u and v produces a square.

- Write $u = u'x$ and $v = yv'$, where $x, y \in \{\mathrm{a}, \mathrm{b}, \mathrm{c}\}$.

<table>
<tr><td colspan="2" align="center">u</td><td colspan="2" align="center">v</td></tr>
<tr><td>u'</td><td>x</td><td>y</td><td>v'</td></tr>
</table>

- Since uv is square-free, we must have $x \neq y$, and uy must be the unique square-free right extension of u.
- Let $z \neq x, y$ be the other letter – then uz contains a square.

BUILDING BLOCKS

- Fact: The word

$$N = \mathtt{abacbabcabacbcacbabcabacabcbabcabacbcabcb}$$

  is *nearly extremal square-free*.
- For every permutation π of the set $\{\mathtt{a}, \mathtt{b}, \mathtt{c}\}$, let N_π denote the word obtained from N by applying π.
- Let $N_{\tilde{\pi}}$ denote the reversal of the word N_π.
- In this way, the word N gives rise to twelve distinct nearly extremal square-free words:

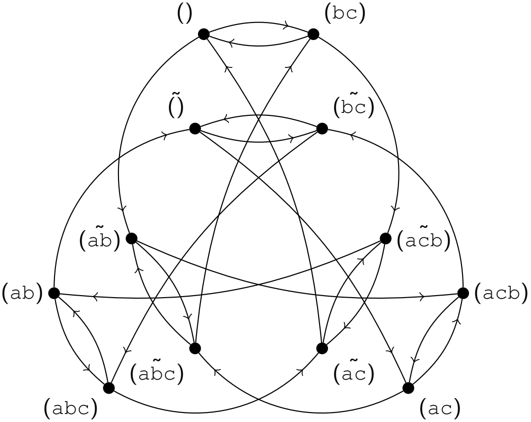
$$N_{()}, N_{(\mathtt{ab})}, N_{(\mathtt{ac})}, N_{(\mathtt{bc})}, N_{(\mathtt{abc})}, N_{(\mathtt{abc})}$$
$$N_{\widetilde{()}}, N_{\widetilde{(\mathtt{ab})}}, N_{\widetilde{(\mathtt{ac})}}, N_{\widetilde{(\mathtt{bc})}}, N_{\widetilde{(\mathtt{abc})}}, N_{\widetilde{(\mathtt{abc})}}$$

A USEFUL DIGRAPH

- Fact: The words $N_{()}N_{(\mathtt{bc})}$ and $N_{()}N_{(\mathtt{a}\tilde{\mathtt{b}})}$ are square-free, and hence nearly extremal square-free.

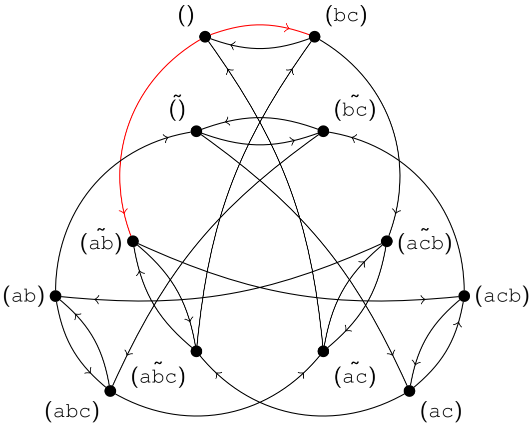
A USEFUL DIGRAPH

- Fact: The words $N_{()}N_{(\mathtt{bc})}$ and $N_{()}N_{(\tilde{\mathtt{ab}})}$ are square-free, and hence nearly extremal square-free.
- We are led to the following digraph D, that has an arc from π to ρ whenever $N_\pi N_\rho$ is square-free.

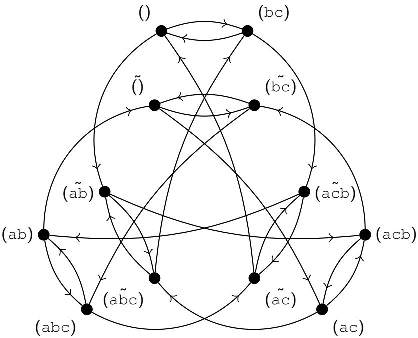


A USEFUL DIGRAPH

- **Fact:** The words $N_{()}N_{(\mathtt{bc})}$ and $N_{()}N_{(\tilde{\mathtt{ab}})}$ are square-free, and hence nearly extremal square-free.
- We are led to the following digraph D, that has an arc from π to ρ whenever $N_\pi N_\rho$ is square-free.

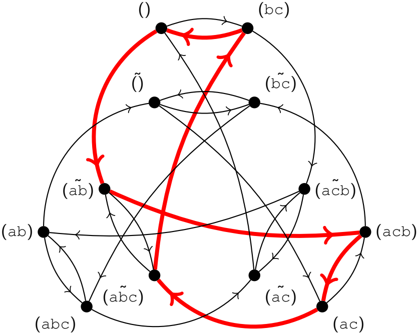


LONG SQUARE-FREE WALKS



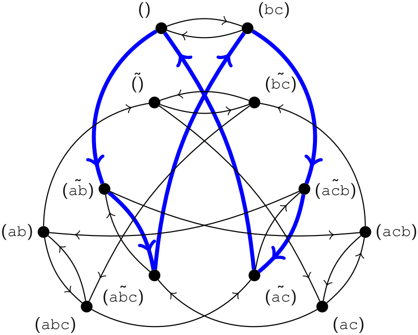
- Lemma: There are arbitrarily long square-free walks in D.

Long square-free walks



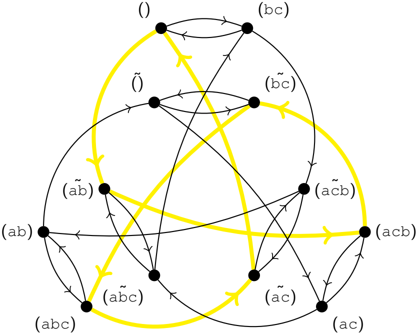
- **Lemma:** There are arbitrarily long square-free walks in D.

LONG SQUARE-FREE WALKS



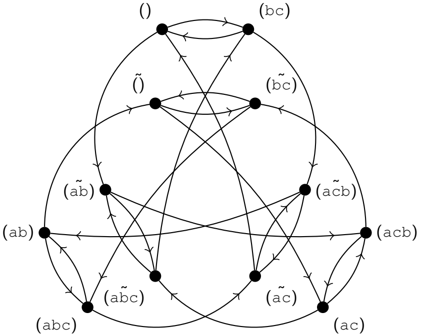
- Lemma: There are arbitrarily long square-free walks in D.

LONG SQUARE-FREE WALKS



- Lemma: There are arbitrarily long square-free walks in D.

LONG NEARLY EXTREMAL SQUARE-FREE WORDS



- Let $f : V(D)^* \to \{\mathrm{a}, \mathrm{b}, \mathrm{c}\}^*$ be the morphism defined by $f(\pi) = N_\pi$ for all $\pi \in V(D)$.
- Lemma: If w is a square-free walk in D, then $f(w)$ is square-free.

Long extremal square-free words

- Finally, there are two short words P and S that will serve as “bookends”.

$$P = \texttt{cbacbcabacbabcabacbcabcbacbc}$$
$$S = \texttt{acabcacbabcabacbcabcbacabacbcabcb}$$

- Specifically, the following hold:
  - The word $PN_{()}$ has exactly one square-free extension, and it is on the right.
  - The word $N_{()}S$ has exactly one square-free extension, and it is on the left.
- Lemma: If w is a square-free walk in D starting and ending at $()$, then $Pf(w)S$ is square-free.
  - From the properties of P, N, and S, it follows that $Pf(w)S$ is extremal square-free.
- So there are arbitrarily long extremal square-free ternary words!

WE WANT MORE!

Question: Are there extremal square-free ternary words of every sufficiently large length?

Theorem (MR 2020+): There is an extremal square-free ternary word of length n if and only if n is in the set

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

IS ANYONE ELSE CRAVING MCNUGGETS?

- Chicken McNuggets come in packs of 6, 9, and 20.
- Oh, you want exactly 16? Too bad.
- But it is not hard to prove that the largest number that cannot be purchased is 43.
- This is a particular instance of the *Frobenius Problem*.

All we really need is the following lemma:

- McNugget Lemma: 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

$$n = ap + bq.$$

TWEAKING THE CONSTRUCTION

- We replace the word N with a set of two nearly extremal square-free words Q and R:

  $Q =$ abacbabcacbacabacbcacbacabcbabcabacbcabcb

  $R =$ abacabcacbacabcbabcacbacabacbcacbacabcbabcabacbcabcb

- We extend the Square-Free Lemma to substitutions.
- Since $|Q| = 41$ and $|R| = 52$, the McNugget Lemma guarantees that there is a nearly extremal square-free word of every length $n \geq 40 \cdot 51 = 2040$.
- Finally, we find bookends P and S, both of length 49:

  $P =$ abacbcabcbacabacbcabcbabcacbcabcbacabacbcabcbacbc

  $S =$ acabacbabcacbacabcbacbcabacbabcacbacabcbabcacbaca

- So there is an extremal square-free word of every length $n \geq 2040 + 2 \cdot 49 = 2138$.

Summary: Extremal square-free words

- There are extremal square-free ternary words of every length $n \geq 87$.
- One can also show that the number of extremal square-free ternary words of length n grows exponentially in n.

SUMMARY: EXTREMAL SQUARE-FREE WORDS

- There are extremal square-free ternary words of every length $n \geq 87$.
- One can also show that the number of extremal square-free ternary words of length n grows exponentially in n.

- Conjecture (GKN 2020): Let Σ be a fixed alphabet of size at least 4. Then there are no extremal square-free words over Σ.
  - i.e., over every alphabet of size at least 4, *every* square-free word can be extended to a longer square-free word!

Plan

Extremal Square-Free Words

Extremal Overlap-Free Words

Extremal β-free words

EXTREMAL OVERLAP-FREE WORDS

- An overlap-free word w is called *extremal* if every extension of w contains an overlap.
- Are there arbitrarily long extremal overlap-free binary words?
- If so, are there extremal overlap-free binary words of every sufficiently large length?

Extremal overlap-free words

- An overlap-free word w is called *extremal* if every extension of w contains an overlap.
- Are there arbitrarily long extremal overlap-free binary words?
- If so, are there extremal overlap-free binary words of every sufficiently large length?
- Theorem (MRS 2020+): There is an extremal overlap-free binary word of length n if and only if n is in the set

$$\mathcal{N} := \{10, 12\} \cup \{2k : k \geq 10\} \cup \left\{2^k + 1 : k \geq 5\right\} \cup \left\{3 \cdot 2^k + 1 : k \geq 3\right\}.$$

Preliminaries

- Let $\mu : \Sigma_2^* \to \Sigma_2^*$ denote the Thue-Morse morphism, defined by

$$\mu(\texttt{0}) = \texttt{01}$$
$$\mu(\texttt{1}) = \texttt{10}$$

- We call $\mu^\omega(\texttt{0})$ the Thue-Morse word.
- Lemma: Let $w \in \Sigma_2^*$ be an overlap-free word of length at least 10, and write $w = w_1 w_2$ with $|w_1|, |w_2| \geq 5$. Then for every letter $a \in \Sigma_2$, the extension $w_1 a w_2$ contains an overlap.
- Proof: It suffices to check the lemma statement for every overlap-free word of length exactly 10.

EVEN LENGTHS

Theorem (MRS 2020+): There is an extremal overlap-free binary word of length n if and only if n is in the set

$$\mathcal{N} := \{10, 12\} \cup \{2k : \ k \geq 10\} \cup \left\{2^k + 1 : \ k \geq 5\right\} \cup \left\{3 \cdot 2^k + 1 : \ k \geq 3\right\}.$$

Earmarked words

- Definition: A word $w \in \Sigma_2^*$ is called *earmarked* if all of the following conditions are satisfied:
  1. w is overlap-free;
  2. the length 4 prefix of w is in $\{\texttt{0010}, \texttt{1101}\}$; and
  3. the length 4 suffix of w is $\texttt{0100}$.

Earmarked words

- Definition: A word $w \in \Sigma_2^*$ is called *earmarked* if all of the following conditions are satisfied:
    1. w is overlap-free;
    2. the length 4 prefix of w is in $\{\texttt{0010}, \texttt{1101}\}$; and
    3. the length 4 suffix of w is $\texttt{0100}$.
- Consider the following earmarked word:

$$u = \texttt{0010}u'\texttt{0100}$$

Earmarked words

- Definition: A word $w \in \Sigma_2^*$ is called *earmarked* if all of the following conditions are satisfied:
  1. w is overlap-free;
  2. the length 4 prefix of w is in $\{\texttt{0010}, \texttt{1101}\}$; and
  3. the length 4 suffix of w is $\texttt{0100}$.
- Consider the following earmarked word:

$$u = \texttt{0010}u'\texttt{0100}$$

- Let $v = \mu(u)$, and let w be the word obtained from v by complementing the first and last letters.

$$v = \mu(u) = \texttt{01011001}\mu(u')\texttt{01100101}$$
$$w = \texttt{11011001}\mu(u')\texttt{01100100}$$

Earmarked words

- Definition: A word $w \in \Sigma_2^*$ is called *earmarked* if all of the following conditions are satisfied:
  1. w is overlap-free;
  2. the length 4 prefix of w is in $\{\texttt{0010}, \texttt{1101}\}$; and
  3. the length 4 suffix of w is $\texttt{0100}$.
- Consider the following earmarked word:

$$u = \texttt{0010}u'\texttt{0100}$$

- Let $v = \mu(u)$, and let w be the word obtained from v by complementing the first and last letters.

$$\begin{aligned} v = \mu(u) &= \texttt{01011001}\mu(u')\texttt{01100101} \\ w &= \texttt{11011001}\mu(u')\texttt{01100100} \end{aligned}$$

- Claim: The word w is earmarked and extremal overlap-free.

EARMARKED WORDS

- Definition: A word $w \in \Sigma_2^*$ is called *earmarked* if all of the following conditions are satisfied:
  1. w is overlap-free;
  2. the length 4 prefix of w is in $\{\texttt{0010}, \texttt{1101}\}$; and
  3. the length 4 suffix of w is $\texttt{0100}$.
- Consider the following earmarked word:

$$u = \texttt{0010}u'\texttt{0100}$$

- Let $v = \mu(u)$, and let w be the word obtained from v by complementing the first and last letters.

$$\begin{aligned} v = \mu(u) &= \texttt{01011001}\mu(u')\texttt{01100101} \\ w &= \texttt{11011001}\mu(u')\texttt{01100100} \end{aligned}$$

- Claim: The word w is earmarked and extremal overlap-free.

EARMARKED WORDS

- Definition: A word $w \in \Sigma_2^*$ is called *earmarked* if all of the following conditions are satisfied:
  1. w is overlap-free;
  2. the length 4 prefix of w is in $\{\texttt{0010}, \texttt{1101}\}$; and
  3. the length 4 suffix of w is $\texttt{0100}$.
- Consider the following earmarked word:

$$u = \texttt{0010}u'\texttt{0100}$$

- Let $v = \mu(u)$, and let w be the word obtained from v by complementing the first and last letters.

$$\begin{aligned} v = \mu(u) &= \texttt{01011001}\mu(u')\texttt{01100101} \\ w &= \texttt{11011001}\mu(u')\texttt{01100100} \end{aligned}$$

- Claim: The word w is earmarked and extremal overlap-free.

EARMARKED WORDS

- Definition: A word $w \in \Sigma_2^*$ is called *earmarked* if all of the following conditions are satisfied:
  1. w is overlap-free;
  2. the length 4 prefix of w is in $\{\texttt{0010}, \texttt{1101}\}$; and
  3. the length 4 suffix of w is $\texttt{0100}$.
- Consider the following earmarked word:

$$u = \texttt{0010}u'\texttt{0100}$$

- Let $v = \mu(u)$, and let w be the word obtained from v by complementing the first and last letters.

$$\begin{aligned} v = \mu(u) &= \texttt{01011001}\mu(u')\texttt{01100101} \\ w &= \texttt{11011001}\mu(u')\texttt{01100100} \end{aligned}$$

- Claim: The word w is earmarked and extremal overlap-free.

Earmarked words

- Definition: A word $w \in \Sigma_2^*$ is called *earmarked* if all of the following conditions are satisfied:
  1. w is overlap-free;
  2. the length 4 prefix of w is in $\{\texttt{0010}, \texttt{1101}\}$; and
  3. the length 4 suffix of w is $\texttt{0100}$.
- Consider the following earmarked word:

$$u = \texttt{0010}u'\texttt{0100}$$

- Let $v = \mu(u)$, and let w be the word obtained from v by complementing the first and last letters.

$$\begin{aligned} v = \mu(u) &= \texttt{01011001}\mu(u')\texttt{01100101} \\ w &= \texttt{11011001}\mu(u')\texttt{01100100} \end{aligned}$$

- Claim: The word w is earmarked and extremal overlap-free.

EARMARKED WORDS

- So from every earmarked word of length n, we can find an earmarked word of length $2n$ that is extremal overlap-free.
- This suggests an inductive approach, but we still need to find earmarked words of odd length to get things started.
- For this, we use `Walnut`.
- `Walnut` is a free software program (written by Hamoon Mousavi) that can prove or disprove certain statements, written in first-order logic, about automatic sequences.
- `Walnut` is available at:

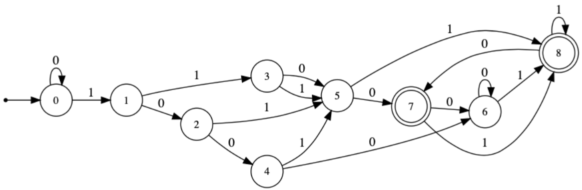
  `https://github.com/hamousavi/Walnut`

EARMARKED WORDS

Lemma: Let $n \geq 10$ be an integer satisfying $n \not\equiv 0 \pmod 4$. Then there is an earmarked word of length n.

Proof: We ask `Walnut`:

- For which n does the Thue-Morse word contain a factor u of length $n - 4$ such that the word $u\texttt{0100}$ is earmarked?



Corollary: If n is even, then there is an extremal overlap-free binary word of length n if and only if $n \in \{10, 12\}$ or $n \geq 20$.

ODD LENGTHS

Theorem (MRS 2020+): There is an extremal overlap-free binary word of length n if and only if n is in the set

$$\begin{aligned}\mathcal{N} :=& \{10, 12\} \cup \{2k\colon\ k \geq 10\} \\ & \cup \left\{2^k + 1\colon\ k \geq 5\right\} \cup \left\{3 \cdot 2^k + 1\colon\ k \geq 3\right\}.\end{aligned}$$

OVERLAP-FREE BINARY SQUARES

- Define $A = \{\mathtt{00}, \mathtt{11}, \mathtt{010010}, \mathtt{101101}\}$ and $\mathcal{A} = \bigcup_{k \geq 0} \mu^k(A)$.
- Theorem (Shelton and Soni, 1985): The overlap-free binary squares are the conjugates of the words in $\mathcal{A}$.
- Hence, every overlap-free binary square has length 2^k or $3 \cdot 2^k$ for some k.
- Claim: Let u be an extremal overlap-free binary word of odd length. Then $u = vva$ or $u = avv$ for a word $v \in \Sigma_2^*$ and a letter $a \in \Sigma_2$.

Summary: Extremal overlap-free words

Theorem (MRS 2020+): There is an extremal overlap-free binary word of length n if and only if n is in the set

$$\begin{aligned}\mathcal{N} :=& \{10, 12\} \cup \{2k : \ k \geq 10\} \\ & \cup \left\{2^k + 1 : \ k \geq 5\right\} \cup \left\{3 \cdot 2^k + 1 : \ k \geq 3\right\}.\end{aligned}$$

SUMMARY: EXTREMAL OVERLAP-FREE WORDS

Theorem (MRS 2020+): There is an extremal overlap-free binary word of length n if and only if n is in the set

$$\mathcal{N} := \{10, 12\} \cup \{2k : \; k \geq 10\} \cup \left\{2^k + 1 : \; k \geq 5\right\} \cup \left\{3 \cdot 2^k + 1 : \; k \geq 3\right\}.$$

Some Questions:

- Can the extremal overlap-free binary words themselves be characterized?
- Are there extremal overlap-free words over any larger alphabets?
- Are there extremal cube-free words over the binary alphabet? What about fractional powers between two and three?

Plan

Extremal Square-Free Words

Extremal Overlap-Free Words

Extremal β-free words

FRACTIONAL POWERS

- Let $w = w_1 w_2 \cdots w_n$ be a word.
- We say that w has *period* $p > 0$ if $w_{i+p} = w_i$ for all $1 \leq i \leq n - p$.
  - E.g., the word `alfalfa` has periods 3, 6, and 7.
- The *exponent* of w is its length divided by its minimal period.
  - E.g., the word `alfalfa` has exponent $7/3$.

β-FREE WORDS

- Let $r > 1$ be a real number.
- We say that w is r-free if it contains no factor of exponent greater than or equal to r.
  - Note: Square-free words are 2-free words.
- We say that w is r^+-free it it contains no factor of exponent strictly greater than r.
  - Note: Overlap-free words are 2^+-free words.
- Henceforth, we let β be a member of the "extended real numbers", denoted $\mathbb{R}_{\text{ext}}$, consisting of all real numbers together with all real numbers with a plus.

Extremal β-free words

- A β-free word *w* is called *extremal* if every extension of *w* contains a factor of exponent at least β.
- Question: Over a fixed alphabet, for which β are there arbitrarily long extremal β-free words?
- Let's start with the binary alphabet.
- We have already seen that there are arbitrarily long extremal 2^{+}-free binary words.

(α, β)-EXTREMAL WORDS

- We noticed that in some of the extremal overlap-free words that we constructed, every extension actually contained not just an overlap, but a very short overlap.
- In fact, we found an infinite family of overlap-free words for which every extension contains a factor of exponent at least $7/3$.

Definition: Let *w* be a word over a fixed alphabet Σ, and let $\alpha, \beta \in \mathbb{R}_{\text{ext}}$ satisfy $1 < \alpha \leq \beta$. We say that *w* is (α, β)*-extremal* if *w* is α-free, and every extension of *w* contains a factor of exponent at least β.

Claim: Suppose that *w* is (α, β)-extremal, and that $\alpha \leq \gamma \leq \beta$. Then *w* is extremal γ-free.

Theorem (MRS 2020+): Let $\beta \in \mathbb{R}_{\text{ext}}$ satisfy $2 < \beta \leq 8/3$. Then there are arbitrarily long extremal β-free binary words.

Proof: We show that all of the following hold.

1. There are arbitrarily long $(2^+, 7/3)$-extremal binary words.
2. There are arbitrarily long $(7/3^+, 17/7)$-extremal binary words.
3. There are arbitrarily long $(17/7^+, 5/2)$-extremal binary words.
4. There are arbitrarily long $(5/2^+, 18/7)$-extremal binary words.
5. There are arbitrarily long $(18/7^+, 8/3)$-extremal binary words.

- For parts 2-5, we go back to the "building blocks" and "bookends" idea.

- For example, we find some short binary words that are $7/3^+$-free for which every internal extension contains a factor of exponent at least $17/7$.
- Define $f : \Sigma_3^* \to \Sigma_2^*$ by

$$
\begin{aligned}
f(\mathtt{0}) &= \mathtt{001011001101100100110100110110010011} \\
f(\mathtt{1}) &= \mathtt{001011001101100100110110010110010011} \\
f(\mathtt{2}) &= \mathtt{001011001101100101100100110110010011}.
\end{aligned}
$$

- Define “bookends” $r = \mathtt{1100110010011}$ and $s = \mathtt{001001}$.
- Claim: Let $u \in \Sigma_3^*$ be a square-free word of length at least 3, and write $u = avb$, where $a, b \in \Sigma_3$. Then

$$rf(v)s$$

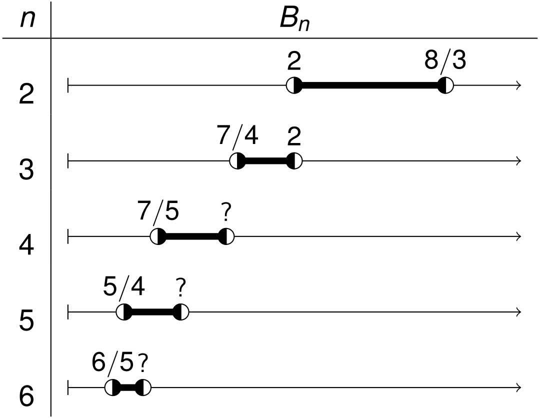
is $(7/3^+, 17/7)$-extremal.

A VAGUE GUESS

- What if $\beta \geq 8/3^+$?
- What about larger alphabets?

A VAGUE GUESS

- What if $\beta \geq 8/3^+$?
- What about larger alphabets?
- Let $B_n = \{\beta \in \mathbb{R}_{\text{ext}} :$ there are arbitrarily long extremal β-free words over $\Sigma_n\}$.



CONCLUSION: OUR MOST BURNING QUESTIONS

- Are there any extremal square-free words over a four letter alphabet?
- Are there any extremal overlap-free words over the ternary alphabet?
- Are there any extremal $8/3^+$-free words over the binary alphabet?
- Are there arbitrarily long extremal $\mathrm{RT}(n)^+$-free words over n letters?

Thank you!