Infinite words containing squares at every position

James Currie and Narad Rampersad

Department of Mathematics and Statistics
University of Winnipeg

27 August 2008

The Main Problem

We consider the following question of Richomme:

> *What is the infimum of the real numbers $\alpha > 2$ for which there exists an infinite word that contains no α-power as a subword, yet contains arbitrarily large squares beginning at every position?*

As we shall see, over the binary alphabet, the answer to Richomme's question is $\alpha = 7/3$.

Definitions

First we recall some basic definitions.

square:	A word of the form xx (like tintin).
cube:	A word of the form xxx.
α-power:	For a real number α, the shortest prefix of length $\geq \alpha\|x\|$ of some infinite word $xxx\cdots$.
α^+-power:	For a real number α, the shortest prefix of length $> \alpha\|x\|$ of some infinite word $xxx\cdots$.
overlap:	A 2^+-power.
α-power-free word:	A word containing no α-power as a subword.
α^+-power-free word:	A word containing no α^+-power as a subword.

Aperiodic Words with Arbitrarily Large Squares

- There exist aperiodic infinite binary words that contain arbitrarily large squares starting at every position.
- For instance, all Sturmian words have this property (see, for example, Allouche, Davison, Queffélec, and Zamboni 2001).
- Certain Sturmian words additionally avoid α-powers for some real number α.
- The Fibonacci word

$$\mathbf{f} = 0100101001001010010100\cdots$$

contains no $(2+\varphi)$-powers, where $\varphi = (1+\sqrt{5})/2$ (Mignosi and Pirillo 1992; see also Krieger 2007).

Aperiodic Words with Arbitrarily Large Squares

- However, $\mathbf{f}$ contains arbitrarily large squares at each position.
- For example, $\mathbf{f}$ begins with the squares
  - 010010,
  - 0100101001,
  - 0100101001001010, etc.
- By contrast, the Thue–Morse word

$$\mathbf{t} = 0110100110010110100101 10 \cdots$$

is overlap-free (Thue 1912) but does not contain squares beginning at every position.

The Squares in the Thue–Morse Word

- Let μ denote the Thue–Morse morphism, which maps

$$0 \to 01, \quad 1 \to 10.$$

- Let $A = \{00, 11, 010010, 101101\}$ and

$$\begin{aligned} \mathscr{A} &= \bigcup_{k \geq 0} \mu^k(A) \\ &= \{00, 11, 0101, 1010, 010010, 101101, 01100110, \ldots\}. \end{aligned}$$

- Then $\mathscr{A}$ is the set of squares appearing in the Thue–Morse word (Pansiot 1981; Brlek 1989).
- The overlap-free binary squares in general are the conjugates of the words in $\mathscr{A}$ (Thue 1912; Shelton and Soni 1985).
- (A conjugate of x is a word y such that $x = uv$ and $y = vu$ for some u, v.)

The Squares in Overlap-free Words

Using this characterization of the overlap-free binary squares, we generalize the previous observation that the Thue–Morse word does not contain squares starting at every position.

Theorem

If $\mathbf{w}$ is an infinite overlap-free binary word, then there is a position i such that $\mathbf{w}$ does not contain a square beginning at position i.

The Squares in Overlap-free Words

- Check by computer that any overlap-free word of length greater than 36 contains 010011.
- E.g., in the Thue–Morse word:

$$011\textcolor{red}{010011}0010110\cdots.$$

- Let i denote any position at which 010011 occurs in $\mathbf{w}$.
- We claim that no square begins at position i.
- Suppose to the contrary that xx is such a square.
- Except for $00, 11, 010010, 101101$ and their conjugates, every overlap-free square xx has $|x|$ even.
- It follows that $xx = 0y10y1$, where $y \in \{01, 10\}^\ell$ for some ℓ.
- This forces xx to be followed by 0 in $\mathbf{w}$, so that the overlap $xx0$ occurs in $\mathbf{w}$, a contradiction.

A Word with Large Squares at Every Position

Theorem

There exists an infinite $(7/3)^+$-power-free binary word that contains arbitrarily large squares beginning at every position.

- We illustrate the idea of the proof with the following example, due to Richomme.
- Let f be the map $0 \to 001$, $1 \to 011$.
- f maps cubefree words to cubefree words: iterating f gives the infinite cubefree word

$$\mathbf{x} = f^{\omega}(0) = 001001011001001011001011011 \cdots .$$

The Construction Continued

- $\mathbf{x}$ starts with 0010010.
- $\mathbf{x}$ therefore starts with $f^n(001)f^n(001)f^n(0)$, for $n \geq 0$.
- $f^n(0)$ is a prefix of $f^n(001)$.
- Every subword of length $|f^n(001001)|$ starting at positions 1 to $|f^n(0)|$ is thus a square.
- For instance, when $n = 1$:

$$001001011001001011001011011 \cdots .$$

- Thus $\mathbf{x}$ is cubefree, but contains arbitrarily large squares starting at every position.

The Construction Continued

- $\mathbf{x}$ starts with 0010010.
- $\mathbf{x}$ therefore starts with $f^n(001)f^n(001)f^n(0)$, for $n \geq 0$.
- $f^n(0)$ is a prefix of $f^n(001)$.
- Every subword of length $|f^n(001001)|$ starting at positions 1 to $|f^n(0)|$ is thus a square.
- For instance, when $n = 1$:

$$\color{red}{001001011\,001001011}\,001011011\cdots.$$

- Thus $\mathbf{x}$ is cubefree, but contains arbitrarily large squares starting at every position.

The Construction Continued

- $\mathbf{x}$ starts with 0010010.
- $\mathbf{x}$ therefore starts with $f^n(001)f^n(001)f^n(0)$, for $n \geq 0$.
- $f^n(0)$ is a prefix of $f^n(001)$.
- Every subword of length $|f^n(001001)|$ starting at positions 1 to $|f^n(0)|$ is thus a square.
- For instance, when $n = 1$:

$$0\,010010110\,010010110\,01011011 \cdots .$$

- Thus $\mathbf{x}$ is cubefree, but contains arbitrarily large squares starting at every position.

The Construction Continued

- $\mathbf{x}$ starts with 0010010.
- $\mathbf{x}$ therefore starts with $f^n(001)f^n(001)f^n(0)$, for $n \geq 0$.
- $f^n(0)$ is a prefix of $f^n(001)$.
- Every subword of length $|f^n(001001)|$ starting at positions 1 to $|f^n(0)|$ is thus a square.
- For instance, when $n = 1$:

$$00\,100101100\,100101100\,1011011 \cdots .$$

- Thus $\mathbf{x}$ is cubefree, but contains arbitrarily large squares starting at every position.

The Construction Continued

- $\mathbf{x}$ starts with 0010010.
- $\mathbf{x}$ therefore starts with $f^n(001)f^n(001)f^n(0)$, for $n \geq 0$.
- $f^n(0)$ is a prefix of $f^n(001)$.
- Every subword of length $|f^n(001001)|$ starting at positions 1 to $|f^n(0)|$ is thus a square.
- For instance, when $n = 1$:

$$001\,\color{red}{001011001}\,\color{red}{001011001}\,011011\cdots.$$

- Thus $\mathbf{x}$ is cubefree, but contains arbitrarily large squares starting at every position.

The Desired Result

- To improve from cubefree to $(7/3)^+$-power-free, we instead iterate g, which maps

$$\begin{aligned} 0 &\rightarrow 011010011011001101001 \\ 1 &\rightarrow 100101100110110010110. \end{aligned}$$

- Then g generates an infinite $(7/3)^+$-power-free word (Kolpakov, Kucherov, and Tarannikov 1999) , but contains 0110110.
- As previously, iterates of g on the overlap 0110110 create arbitrarily large squares at every position.
- The details are a little more involved, but conceptually similar to the previous argument.

Optimality of the Previous Result

Theorem

If $\mathbf{w}$ is an infinite $7/3$-power-free binary word, then there is a position i such that $\mathbf{w}$ does not contain arbitrarily large squares beginning at position i.

- The idea is to find an occurrence of 010011 at position i in $\mathbf{w}$ and argue that there cannot exist arbitrarily large squares at position i.
- The proof requires a characterization of the $7/3$-power-free binary squares.
- As we shall show later, the $7/3$-power-free binary squares are exactly the overlap-free binary squares.

Words with Squares Starting at Every Position

- What if we remove the requirement that our words contain *arbitrarily large* squares at every position?
- What if we simply want a square at every position?

Theorem

There exists an infinite $7/3$-power-free binary word that contains squares beginning at every position.

- Define the sequence: $A_0 = 00$ and $A_{n+1} = 0\mu^2(A_n)$, $n \geq 0$.
- The first few terms are

$$\begin{aligned} A_0 &= 00 \\ A_1 &= 001100110 \\ A_2 &= 0011001101001100101100110100110010110 \\ &\vdots \end{aligned}$$

Words with Squares Starting at Every Position

- As $n \to \infty$, A_n tends to an infinite limit word

$$\mathbf{w} = 00110011010011001011001101001100101 10 \cdots .$$

- Note: $\mathbf{w}$ contains the squares
  - $0\,0$ at position 1,
  - $0110\,0110$ at position 2,
  - $110011010011001 0\,1100110100110010$ at position 3, etc.
- In general, $\mathbf{w}$ contains a square starting at every position.
- A similar, but slightly more complicated, construction yields the stronger result:

Theorem

For every real number $\alpha > 2$, there exists an infinite α-power-free binary word that contains squares beginning at every position.

Characterizing the $7/3$-power-free Squares

- One of our main results generalizes Shelton and Soni's characterization of the overlap-free binary squares.
- Recall the sets $A = \{00, 11, 010010, 101101\}$ and

$$\begin{aligned} \mathscr{A} &= \bigcup_{k \geq 0} \mu^k(A) \\ &= \{00, 11, 0101, 1010, 010010, 101101, 01100110, \ldots\}. \end{aligned}$$

Theorem

The $7/3$-power-free binary squares are the conjugates of the words in $\mathscr{A}$.

- In other words, the $7/3$-power-free binary squares are exactly the overlap-free binary squares.

The Factorization Theorem

- Recall the morphism μ, which maps $0 \to 01$ and $1 \to 10$.
- We need the factorization theorem:

Theorem (Karhumäki and Shallit 2004)

Let $x \in \{0,1\}^*$ be α-power-free, $2 < \alpha \leq 7/3$. Then there exist $u, v \in \{\epsilon, 0, 1, 00, 11\}$ and an α-power-free $y \in \{0,1\}^*$ such that $x = u\mu(y)v$.

- For example, if $x = 0010011001$, then $x = 00\mu(1010)$.
- Here, $u = 00$, $y = 1010$, and $v = \epsilon$.

Some Lemmas

Lemma

Let $xx \in \{0,1\}^*$ *be* $7/3$*-power-free. If* $xx = \mu(y)$*, then* $|y|$ *is even. Consequently,* y *is a square.*

- We also need a special version of the factorization theorem just for squares:

Lemma

Let $xx \in \{0,1\}^*$ *be* $7/3$*-power-free. If* $|xx| > 8$*, then either*

(a) $xx = \mu(y)$*, where* $y \in \{0,1\}^*$*; or*

(b) $xx = \overline{a}\mu(y)a$*, where* $a \in \{0,1\}$ *and* $y \in \{0,1\}^*$*.*

- For example, if $xx = 010110010110$, then $xx = \mu(001001)$.
- And if, $xx = 11001100$, then $xx = 1\mu(101)0$.

A Sketch of the Proof

- Let xx be a minimal $7/3$-power-free square that is not a conjugate of a word in $\mathscr{A}$.
- A computer check verifies that $|xx| > 8$.
- Case 1: $xx = \mu(y)$. Then y is a square.
- Furthermore, y is not a conjugate of a word in $\mathscr{A}$, contradicting the minimality of xx.
- Case 2: $xx = \overline{a}\mu(y)a$. Then $a\overline{a}\mu(y) = \mu(ay)$ is also a square zz.
- We proceed by showing that zz is $7/3$-power-free, and consequently, that ay is a $7/3$-power-free square, contradicting the minimality of xx.
- The details of showing that zz is $7/3$-power-free are somewhat technical, and we omit them.

Summary of Results

- An infinite overlap-free binary word cannot contain squares beginning at every position.
- There exists an infinite $(7/3)^+$-power-free binary word that contains arbitrarily large squares beginning at every position.
- An infinite $7/3$-power-free binary word cannot contain arbitrarily large squares beginning at every position.
- For every real number $\alpha > 2$, there exists an infinite α-power-free binary word that contains squares beginning at every position.
- The $7/3$-power-free binary squares are exactly the overlap-free binary squares.

Open Problems

- We have only considered words over a binary alphabet. Do similar results hold over a larger alphabet?
- For instance, does there exist an infinite overlap-free ternary word that contains squares beginning at every position?
- Richomme observed that over any alphabet there cannot exist an infinite overlap-free word containing infinitely many squares at every position.
- Apply the following result of Ilie (2007): In any word, if vv and uu are two squares at position i and ww is a square at position $i+1$, then either $|w| = |u|$ or $|w| = |v|$ or $|w| \geq 2|v|$.
- Consequently, in any infinite word, if infinitely many distinct squares begin at position i and ww is a square beginning at position $i+1$, then $|w| = |u|$ for some square uu occurring at position i, so there is an overlap at position i.

Thank you!