

Avoiding Overlaps with the Greedy Algorithm

Mathieu Guay-Paquet ¹ Jeffrey Shallit ²

¹Department of Combinatorics & Optimization

²School of Computer Science

University of Waterloo

2nd Canadian Discrete and Algorithmic
Mathematics Conference

May 26, 2009

Outline

Background

Avoiding squares

Avoiding overlaps

Open questions

Outline

Background

Avoiding squares

Avoiding overlaps

Open questions

The objects

- ▶ Alphabet: a (usually finite) set of symbols.

$\{a, \dots, z\}, \{0, 1\}, \{0, 1, 2\}, \dots$

- ▶ Word: a finite or (right) infinite string of symbols.

word, ababca, 01101001, 010201202 \dots , \dots

- ▶ Pattern: a set of words.

$\{\text{English words}\}, \{xx \mid x \in \Sigma^*\}, \{axaxa \mid a \in \Sigma, x \in \Sigma^*\}, \dots$

The objects

- ▶ Alphabet: a (usually finite) set of symbols.

$\{a, \dots, z\}, \{0, 1\}, \{0, 1, 2\}, \dots$

- ▶ Word: a finite or (right) infinite string of symbols.

word, ababca, 01101001, 010201202 \dots , \dots

- ▶ Pattern: a set of words.

$\{\text{English words}\}, \{xx \mid x \in \Sigma^*\}, \{axaxa \mid a \in \Sigma, x \in \Sigma^*\}, \dots$

The objects

- ▶ Alphabet: a (usually finite) set of symbols.

$\{a, \dots, z\}, \{0, 1\}, \{0, 1, 2\}, \dots$

- ▶ Word: a finite or (right) infinite string of symbols.

word, ababca, 01101001, 010201202 \dots , \dots

- ▶ Pattern: a set of words.

$\{\text{English words}\}, \{xx \mid x \in \Sigma^*\}, \{axaxa \mid a \in \Sigma, x \in \Sigma^*\}, \dots$

The questions

- ▶ Given an alphabet and a pattern, are there long words avoiding it? Infinite words avoiding it?
- ▶ Given a pattern, how big of an alphabet do you need to avoid it?
- ▶ What is the lexicographically least infinite word avoiding a pattern?

The questions

- ▶ Given an alphabet and a pattern, are there long words avoiding it? Infinite words avoiding it?
- ▶ Given a pattern, how big of an alphabet do you need to avoid it?
- ▶ What is the lexicographically least infinite word avoiding a pattern?

The questions

- ▶ Given an alphabet and a pattern, are there long words avoiding it? Infinite words avoiding it?
- ▶ Given a pattern, how big of an alphabet do you need to avoid it?
- ▶ What is the lexicographically least infinite word avoiding a pattern?

Some answers

Thue (1906, 1912):

- ▶ Can avoid overlaps over a binary alphabet, but not a unary one.

0110100110010110...

- ▶ Can avoid squares over a ternary alphabet, but not a binary one.

0120210121020120...

Some answers

Thue (1906, 1912):

- ▶ Can avoid overlaps over a binary alphabet, but not a unary one.

0110100110010110...

- ▶ Can avoid squares over a ternary alphabet, but not a binary one.

0120210121020120...

Some tools

How do you describe an infinite word and prove things about it?

- ▶ Give a program to compute it. But it's hard to reason about properties of the word.
- ▶ Give it as a fixed point of a morphism. Then prove properties of the morphism.

0110100110010110...

0 \mapsto 01

1 \mapsto 10

0120210121020120...

0 \mapsto 012

1 \mapsto 02

2 \mapsto 1

Some tools

How do you describe an infinite word and prove things about it?

- ▶ Give a program to compute it. But it's hard to reason about properties of the word.
- ▶ Give it as a fixed point of a morphism. Then prove properties of the morphism.

0110100110010110...

0 \mapsto 01

1 \mapsto 10

0120210121020120...

0 \mapsto 012

1 \mapsto 02

2 \mapsto 1

Outline

Background

Avoiding squares

Avoiding overlaps

Open questions

Finite vs. infinite alphabets

The lexicographically least infinite squarefree word over $\{0, 1, 2\}$ is hard to compute.

0102012021012010...

The lexicographically least infinite squarefree word over \mathbb{N} is trivial to compute.

$w_2 = 0102010301020104 \dots$

Finite vs. infinite alphabets

The lexicographically least infinite squarefree word over $\{0, 1, 2\}$ is hard to compute.

0102012021012010...

The lexicographically least infinite squarefree word over \mathbb{N} is trivial to compute.

$w_2 = 0102010301020104\dots$

The ruler word

$$\mathbf{w}_2 = \gamma^\omega(0) = 01020103010201040102010301020105 \dots$$

where

$$\gamma(0) = 01$$

$$\gamma(1) = 02$$

$$\gamma(2) = 03$$

\vdots

$$\gamma(i) = 0 \cdot (i + 1)$$



Photo by Flickr user iliahi.

Irreducibility

The idea: we can generate w_2 by starting with the empty word and repeatedly appending the smallest letter which does not create a square.

A word is **irreducible** at a position (with respect to squares) if replacing the letter at that position by any smaller letter creates a squares ending at that position.

010201020304

Irreducibility

The idea: we can generate w_2 by starting with the empty word and repeatedly appending the smallest letter which does not create a square.

A word is **irreducible** at a position (with respect to squares) if replacing the letter at that position by any smaller letter creates a squares ending at that position.

010201020304

Squarefreeness, irreducibility, and the lex order

Key property: if v is irreducible (at every position) and w is squarefree, then $v \leq w$ in the lexicographic order (or w is a prefix of v).

So: the irreducible squarefree words are the prefixes of w_2 .

So: if γ is irreducible and squarefree, then $w_2 = \gamma^\omega(0)$.

Squarefreeness, irreducibility, and the lex order

Key property: if v is irreducible (at every position) and w is squarefree, then $v \leq w$ in the lexicographic order (or w is a prefix of v).

So: the irreducible squarefree words are the prefixes of \mathbf{w}_2 .

So: if γ is irreducible and squarefree, then $\mathbf{w}_2 = \gamma^\omega(0)$.

Squarefreeness, irreducibility, and the lex order

Key property: if v is irreducible (at every position) and w is squarefree, then $v \leq w$ in the lexicographic order (or w is a prefix of v).

So: the irreducible squarefree words are the prefixes of \mathbf{w}_2 .

So: if γ is irreducible and squarefree, then $\mathbf{w}_2 = \gamma^\omega(0)$.

Outline

Background

Avoiding squares

Avoiding overlaps

Open questions

Finite vs. infinite alphabets

Berstel (1994), then Allouche, Currie, Shallit (1998) describe the lexicographically least infinite overlap-free word over $\{0, 1\}$ with any prefix.

$$001001\bar{t} = 0010011001011001101001 \dots$$

M, Shallit (2008) describe the lexicographically least infinite overlap-free word over \mathbb{N} .

$$\mathbf{w}_{2+} = 00100110010020010011001002100 \dots$$

Finite vs. infinite alphabets

Berstel (1994), then Allouche, Currie, Shallit (1998) describe the lexicographically least infinite overlap-free word over $\{0, 1\}$ with any prefix.

$$001001\bar{t} = 0010011001011001101001 \dots$$

M, Shallit (2008) describe the lexicographically least infinite overlap-free word over \mathbb{N} .

$$\mathbf{w}_{2+} = 00100110010020010011001002100 \dots$$

The overlap-free word

The word $\mathbf{w}_{2+} = \varphi^\omega(0) = 00100110 \dots$ is generated by a morphism φ with

$$\varphi(0) = \psi(0, 1) = 001$$

$$\varphi(1) = \psi(1, 2) = 1001002$$

$$\varphi(2) = \psi(2, 3) = 200100110010020010011001003$$

\vdots

where $\psi(h, k)$ is the lexicographically least overlap-free word starting with h and ending with k for letters $h \leq k$. This morphism is overlap-free and irreducible.

The structure of $\psi(h, k)$

- ▶ It turns out that $\psi(h, k) = \varphi^{k-h}(h)$ for all $h \leq k$.
- ▶ We can construct $\psi(h, h+1)$ as

$$\begin{aligned}\psi(h, k) &= h \cdot \psi(0, h) \cdot \psi(0, h) \cdot h^{-1} \cdot (h+1) \\ &= S(\psi(0, h) \cdot \psi(0, h)) \cdot (h+1).\end{aligned}$$

- ▶ Solving a simple recurrence gives $|\psi(0, k)| = \lfloor 2^k k! \sqrt{e} \rfloor$.

The structure of $\psi(h, k)$

- ▶ It turns out that $\psi(h, k) = \varphi^{k-h}(h)$ for all $h \leq k$.
- ▶ We can construct $\psi(h, h+1)$ as

$$\begin{aligned}\psi(h, k) &= h \cdot \psi(0, h) \cdot \psi(0, h) \cdot h^{-1} \cdot (h+1) \\ &= S(\psi(0, h) \cdot \psi(0, h)) \cdot (h+1).\end{aligned}$$

- ▶ Solving a simple recurrence gives $|\psi(0, k)| = \lfloor 2^k k! \sqrt{e} \rfloor$.

The structure of $\psi(h, k)$

- ▶ It turns out that $\psi(h, k) = \varphi^{k-h}(h)$ for all $h \leq k$.
- ▶ We can construct $\psi(h, h+1)$ as

$$\begin{aligned}\psi(h, k) &= h \cdot \psi(0, h) \cdot \psi(0, h) \cdot h^{-1} \cdot (h+1) \\ &= S(\psi(0, h) \cdot \psi(0, h)) \cdot (h+1).\end{aligned}$$

- ▶ Solving a simple recurrence gives $|\psi(0, k)| = \lfloor 2^k k! \sqrt{e} \rfloor$.

The structure of $\psi(h, k)$

The word $\psi(0, k)$ is formed from many cyclically shifted copies of $\psi(0, k - 1)$.

$$\begin{aligned}\psi(0, 3) = & 0010011001002 \\ & 0010011001002 \\ & 1001002001001 \\ & 1001002001001 \\ & 2001001100100 \\ & 2001001100100 \quad 3\end{aligned}$$

In particular, it follows that $S(\psi(0, k)) = R(\psi(0, k))$.

The structure of $\psi(h, k)$

The word $\psi(0, k)$ is formed from many cyclically shifted copies of $\psi(0, k - 1)$.

$$\begin{aligned}\psi(0, 3) = & 0010011001002 \\ & 0010011001002 \\ & 1001002001001 \\ & 1001002001001 \\ & 2001001100100 \\ & 2001001100100 \quad 3\end{aligned}$$

In particular, it follows that $S(\psi(0, k)) = R(\psi(0, k))$.

Outline

Background

Avoiding squares

Avoiding overlaps

Open questions

Open questions

- ▶ For 1-letter prefixes, we have $\psi(h, \infty) = \varphi^\omega(h)$. What about longer prefixes?
- ▶ Similar constructions and arguments work for any integer or integer-plus power. What about avoiding rational powers over \mathbb{N} ? Do we always use every element of \mathbb{N} ?
- ▶ Are there other patterns P with irreducible P -free morphisms?

The end

Thank you!