Avoiding Overlaps with the Greedy Algorithm

Mathieu Guay-Paquet [1]   Jeffrey Shallit [2]

[1]Department of Combinatorics & Optimization

[2]School 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

The objects

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

$$\{\text{a}, \ldots, \text{z}\}, \{0, 1\}, \{0, 1, 2\}, \ldots$$

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

$$\text{word}, \text{ababca}, 01101001, 010201202 \cdots, \ldots$$

- Pattern: a set of words.

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

The objects

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

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

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

$$\text{word}, \text{ababca}, 01101001, 010201202\cdots, \ldots$$

- Pattern: a set of words.

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

The objects

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

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

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

$$\text{word}, \text{ababca}, 01101001, 010201202\cdots, \ldots$$

- Pattern: a set of words.

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

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.

$$0110100110010110\cdots$$

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

$$0120210121020120\cdots$$

Some answers

Thue (1906, 1912):

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

$$0110100110010110\cdots$$

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

$$0120210121020120\cdots$$

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.

$$0110100110010110\cdots \qquad 0120210121020120\cdots$$

$$0 \mapsto 01 \qquad 0 \mapsto 012$$

$$1 \mapsto 10 \qquad 1 \mapsto 02$$

$$2 \mapsto 1$$

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.

$$0110100110010110\cdots$$

$$0 \mapsto 01$$
$$1 \mapsto 10$$

$$0120210121020120\cdots$$

$$0 \mapsto 012$$
$$1 \mapsto 02$$
$$2 \mapsto 1$$

Outline

Finite vs. infinite alphabets

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

$$0102012021012010\cdots$$

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

$$\mathbf{w}_2 = 010201030102010 4\cdots$$

Finite vs. infinite alphabets

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

$$0102012021012010\cdots$$

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

$$\mathbf{w}_2 = 0102010301020104\cdots$$

The ruler word

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

where

$$\gamma(0) = 01$$
$$\gamma(1) = 02$$
$$\gamma(2) = 03$$
$$\vdots$$
$$\gamma(i) = 0 \cdot (i+1)$$

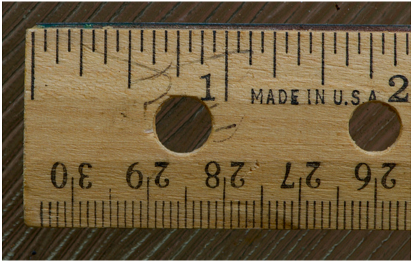

Photo by Flickr user iliahi.

Irreducibility

The idea: we can generate $\mathbf{w}_2$ 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 $\mathbf{w}_2$ 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

Squarefreedom, 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 $\mathbf{w}_2 = \gamma^\omega(0)$.

Squarefreedom, 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 $\mathbf{w}_2 = \gamma^\omega(0)$.

Squarefreedom, 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

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.

$$001001\bar{t} = 001001100101100110100 1\cdots$$

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

$$\mathbf{w}_{2+} = 0010011001002001001100100 2100\cdots$$

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.

$$001001\bar{t} = 001001100101100110100 1\cdots$$

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

$$\mathbf{w}_{2+} = 001001100100200100110010 02100\cdots$$

The overlap-free word

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

$$\begin{aligned}
\varphi(0) &= \psi(0,1) = 001 \\
\varphi(1) &= \psi(1,2) = 1001002 \\
\varphi(2) &= \psi(2,3) = 2001001100100200100110010 03 \\
&\vdots
\end{aligned}$$

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 $\psi(0, k-1)$.

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

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 $\psi(0, k-1)$.

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

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

Outline

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!