Extremal words in the shift orbit closure of a morphic sequence

James Currie, Narad Rampersad, Kalle Saari

Department of Mathematics and Statistics
University of Winnipeg

Shift orbit closure

- a common object of study in combinatorics on words (or dynamical systems) is the shift orbit closure of an infinite word with some interesting combinatorial property
- the shift orbit closure of an infinite word $\mathbf{x}$ is the set of all infinite words $\mathbf{y}$ such that every factor of $\mathbf{y}$ is a factor of $\mathbf{x}$

Extremal words

- often the lexicographically least (or greatest) word in the orbit closure has some additional interesting properties
- given an infinite word, is there a simple way to describe the lexicographically least/greatest word in its orbit closure?

Sturmian words

- if $\mathbf{x}$ is a Sturmian word, the extremal words in the orbit closure of $\mathbf{x}$ are well-known:
- the lexicographically least word for the order $0 < 1$ is the word $0\mathbf{c}$, where $\mathbf{c}$ is the characteristic word with the same slope as $\mathbf{x}$
- the greatest word for the order $0 < 1$ is $10\mathbf{c}$
- for the order $1 < 0$, the extremal words are $1\mathbf{c}$ and $01\mathbf{c}$

The Thue–Morse word

The lexicographically least word $(0 < 1)$ in the orbit closure of the Thue–Morse word

$$0110100110010110\cdots$$

is the word

$$001011001101001\cdots$$

obtained by complementing the Thue–Morse word and deleting the initial 1.

The Rudin–Shapiro word

Currie showed that the least word in the orbit closure of the Rudin–Shapiro word

$$0001001000011101001\cdots$$

is obtained by simply adding a 0 to the beginning.

Automatic sequences

- Thue–Morse and Rudin–Shapiro are both automatic sequences
- Allouche, R., and Shallit showed that the lexicographically least word in the orbit closure of a k-automatic sequence is k-automatic

Generalizing this result

- the question we ask in this work is: Does the result of Allouche et al. generalize to morphic sequences?
- i.e., are the extremal words in the orbit closure of a morphic sequence morphic?
- we show that the answer is “yes” for a wide class of morphic sequences

Morphic sequences

- pure morphic word: an infinite word generated by iterating a prolongable morphism
- morphic word: an infinite word obtained by applying a coding to a pure morphic word

Morphic elements of the orbit closure

Theorem (Currie, R., Saari)

Let $\mathbf{x}$ be a pure morphic word generated by a morphism f. If a word $\mathbf{t}$ in the orbit closure of $\mathbf{x}$ is fixed by f (i.e., $f(\mathbf{t}) = \mathbf{t}$), then $\mathbf{t}$ is morphic.

Bounded letters vs. non-bounded letters

- the proof splits into two cases, depending on whether or not $\mathfrak{t}$ consists entirely of bounded letters
- a letter a is bounded under f if there is a constant k such that $|f^n(a)| < k$ for all n

Word $\mathbf{t}$ consists entirely of bounded letters

- Cassaigne and Nicolas (?) characterized the finite factors of a pure morphic word that consist of bounded letters
- from this characterization one can easily show that $\mathbf{t}$ is ultimately periodic and therefore morphic

Word $\mathfrak{t}$ contains some non-bounded letters

- in this case we use a characterization, due to Head and Lando, of the infinite fixed points of a morphism to conclude that $\mathfrak{t}$ is morphic

Extremal words

- for each letter a and each ordering σ of the alphabet, the lexicographically least word in the orbit closure of an infinite word $\mathbf{x}$ is denoted by $\mathbf{l}_{a,\sigma,\mathbf{x}}$
- any such word is called an extremal word of $\mathbf{x}$

Main Theorem

Theorem (Currie, R., Saari)

Let $\mathbf{x}$ be a pure morphic word generated by a morphism f belonging to a certain class $\mathcal{M}_{\mathbf{x}}$. Then the extremal words of $\mathbf{x}$ are morphic.

The class $\mathcal{M}_{\mathbf{x}}$

- What is the class $\mathcal{M}_{\mathbf{x}}$?
- let $\mathcal{S}_{\mathbf{x}}$ denote the orbit closure of $\mathbf{x}$
- f belongs to $\mathcal{M}_{\mathbf{x}}$ if for every letter b there is a word p_b such that:
  - if $\mathbf{y} \in \mathcal{S}_{\mathbf{x}}$ starts with b, then $f(\mathbf{y})$ starts with p_b; and
  - for distinct letters a, b, neither of p_a or p_b is a prefix of the other.
- e.g., any morphism whose images form a prefix code is of this type

A morphism in the class $\mathcal{M}_{\mathbf{x}}$

- let f map $0 \to 02$, $1 \to 02$, $2 \to 1$
- let $\mathbf{x} = 02102021021 \cdots$ be the fixed point
- if $0\mathbf{y} \in \mathcal{S}_{\mathbf{x}}$: $\mathbf{y}$ begins with 2 and $f(0\mathbf{y})$ begins with 021
- if $1\mathbf{y} \in \mathcal{S}_{\mathbf{x}}$: $\mathbf{y}$ begins with 0 and $f(1\mathbf{y})$ begins with 020
- finally, $f(2\mathbf{y})$ begins with 1
- we take $p_0 = 021$, $p_1 = 020$, $p_2 = 1$
- none of these are prefixes of another
- so $f \in \mathcal{M}_{\mathbf{x}}$

A morphism not in the class $\mathcal{M}_{\mathbf{x}}$

- let f map $0 \to 010$, $1 \to 21$, $2 \to 211$
- let $\mathbf{x} = 01021010211$ be the fixed point starting with 0
- then $f(10\cdots) = 210\cdots$ and $f(12\cdots) = 212\cdots$
- so p_1 must be a prefix of 21
- but 21 is a prefix of $f(2)$
- no matter the choice of p_2, one of p_1, p_2 is a prefix of the other
- so $f \notin \mathcal{M}_{\mathbf{x}}$

Binary morphisms

- the class $\mathcal{M}_{\mathbf{x}}$ may seem somewhat artificial, but it contains all non-trivial binary morphisms
- i.e., all binary morphisms with $f(01) \neq f(10)$ are in $\mathcal{M}_{\mathbf{x}}$

First step of the proof

First, we establish a "desubstitution" result for words in the orbit closure.

Lemma

Let $\mathbf{x}$ be an infinite word and f a morphism. If $\mathbf{y} \in \mathcal{S}_{f(\mathbf{x})}$, then there exist a letter a and an infinite word $\mathbf{z}$ such that $a\mathbf{z} \in \mathcal{S}_{\mathbf{x}}$ and $\mathbf{y} = uf(\mathbf{z})$, where u is a nonempty suffix of $f(a)$.

Second step of the proof

- the next step is to examine the properties of the preimage of an extremal word when applying this desubstitution
- idea: the preimage of a word, extremal with respect to a certain ordering of the alphabet, will again be extremal, but for a possibly different ordering of the alphabet
- here is where we need the condition $f \in \mathcal{M}_{\mathbf{x}}$
- recall: $\mathbf{l}_{a,\sigma,\mathbf{x}}$ is the least word starting with a in the orbit closure of $\mathbf{x}$ w.r.t. an order σ
- let $\mathbf{s}_{a,\sigma,\mathbf{x}}$ be the word obtained by erasing the initial a of $\mathbf{l}_{a,\sigma,\mathbf{x}}$

Preimages of extremal words

Lemma

Let f be a morphism and $\mathbf{x}$ a word such that $f \in \mathcal{M}_{\mathbf{x}}$. Let b be a letter and ρ an ordering of the alphabet of $f(\mathbf{x})$. Then

$$\mathbf{s}_{b,\rho,\mathbf{x}} = v f(\mathbf{s}_{a,\sigma,\mathbf{x}}),$$

where a is a letter, σ is an ordering of the alphabet and v is a proper suffix of $f(a)$.

Idea of the proof

- to show that $\mathbf{s}_{b,\rho,\mathbf{x}} = vf(\mathbf{s}_{a,\sigma,\mathbf{x}})$, we need to explain how to choose the ordering σ
- since $f \in \mathcal{M}_{\mathbf{x}}$, we have certain words p_z associated with every letter z of the alphabet
- no p_z is a prefix of another
- if $\mathbf{y} \in \mathcal{S}_{\mathbf{x}}$ begins with z then $f(\mathbf{y})$ begins with p_z
- so $z <_\sigma z'$ iff $p_z <_\rho p_{z'}$ is the desired ordering

Repeatedly taking preimages of extremal words

- now we iterate the previous lemma
- preimages of extremal words are extremal words (for a possibly different order)
- there are only finitely many ways to order the alphabet
- eventually this process repeats periodically
- we get

$$\mathbf{s}_{b,\rho,\mathbf{x}} = uf^k(\mathbf{s}_{a,\sigma,\mathbf{x}}) \qquad \text{and} \qquad \mathbf{s}_{a,\sigma,\mathbf{x}} = vf^m(\mathbf{s}_{a,\sigma,\mathbf{x}})$$

A characterization of the extremal words

- from this observation we get that

$$\mathbf{l}_{b,\rho,\mathbf{x}} = w\mathbf{t}$$

- where w is a finite word, $\mathbf{t} \in \mathcal{S}_{\mathbf{x}}$, and either

$$\mathbf{t} = f^m(\mathbf{t}) \qquad \text{or}$$

$$\mathbf{t} = x f^m(x) f^{2m}(x) \cdots , \qquad \text{for some finite } x.$$

The conclusion of the proof

- in the first case $\mathbf{t}$ is morphic by the theorem given earlier
- in the second case it is easy to see that $\mathbf{t}$ is morphic

Summary

- we have shown: If $\mathbf{x}$ is a pure morphic word generated by $f \in \mathcal{M}_{\mathbf{x}}$, then all extremal words in $\mathcal{S}_{\mathbf{x}}$ are morphic.
- in fact, all extremal words in any morphic image $g(\mathbf{x})$, where $g \in \mathcal{M}_{\mathbf{x}}$, are morphic
- the binary case: If $\mathbf{x}$ is a pure morphic binary word then all extremal words of $\mathbf{x}$ are morphic.

An example

- we applied this method to compute the extremal words for several classical sequences
- for example, let f map $0 \to 01$ and $1 \to 00$
- then $\mathbf{d} = f^{\omega}(0)$ is the period-doubling word
- let ρ be the ordering $0 < 1$ and $\overline{\rho}$ the opposite
- let $\mathbf{z}$ be the fixed point of the map $0 \to 0001$, $1 \to 0101$
- then

$$\mathbf{l}_{0,\rho,\mathbf{d}} = \mathbf{z} \qquad \mathbf{l}_{1,\rho,\mathbf{d}} = 1\mathbf{z}$$
$$\mathbf{l}_{1,\overline{\rho},\mathbf{d}} = 0^{-1} f(\mathbf{z}) \qquad \mathbf{l}_{0,\overline{\rho},\mathbf{d}} = f(\mathbf{z}).$$

Final remarks

- Luca Zamboni has discovered a clever argument to show that the extremal words of all primitive morphic sequences are primitive morphic.
- We conjecture that in general the extremal words of any morphic sequence are morphic.

The End