Distant words

Manuel Silva,
Universidade Nova de Lisboa

13th Combinatorics Days - Covilhã
July 2023.

Finite and infinite words

- Let A be a finite set (alphabet). Sequences $w = w_1 w_2 \cdots$, finite or infinite, with $w_i \in A$ are called words.
- Let A^* be the set of all finite words. A^* is the free monoid generated by A with the operation of concatenation of words. The empty word is the identity.
- The set of one-way infinite words over A is denoted by $A^{\mathbb{N}}$.
- We will say that u is a factor of the infinite word w if $u = w_i w_{i+1} \cdots w_j$, with $i \leq j$.

Thue-Morse infinite word

Let $h : A^* \to A^*$ satisfy $h(uv) = h(u)h(v)$ for all words u and v.
h is a morphism from the free monoid into itself.

Thue-Morse word:

Let $A = \{a, b\}$ and $\phi : A^* \to A^*$ be defined by
$\phi(a) = ab, \phi(b) = ba$.

Iterating this morphism, we obtain an infinite word
$\phi(a) = ab,\ \phi(ab) = abba,\ \phi(abba) = abbabaab,$

$$t = \lim_{i \to \infty} \phi^i(a) = abbabaabbaababba \cdots$$

A second example: Fibonacci word

Here is another example of an infinite word obtained by morphism iteration.

Fibonacci word:

Let $\psi : A^* \to A^*$ be defined by $\psi(a) = ab, \psi(b) = a$.

$$f = \lim_{i \to \infty} \phi^i(a) = abaababaabaab \cdots$$

Unavoidable regularities

An important topic in Combinatorics on Words is to decide which patterns are avoidable or not. Sometimes the answer depends on the size of the alphabet.

Definition

A power of order k is a pattern of the form u^k for some finite word u (e.g., a square if $k=2$, a cube if $k=3$).

Can we avoid squares, or cubes, inside some infinite word?

Avoiding Powers

The Thue-Morse word is an example of infinite binary word which avoids cubes: does not contain 3 equal consecutive blocks.

We can say that the pattern u^3 is avoidable in a two letter alphabet already.

The property must be preserved by the defining morphism ϕ.

A new definition

We introduce now a dual notion of repetition. Why not look for anti-repetitions instead?

Definition

An anti-power of order k is a pattern of the form $v_1 v_2 \cdots v_k$, with v_i all distinct and having the same length.

The finite word

$$u = 10110100$$

is an example of anti-power of order 4.

Powers or Anti-Powers

Theorem (Fici, Restivo, Zamboni, Silva)

Let w be an infinite word that avoids anti-powers of order k, for some $k > 1$. Then, there exists a finite word u such that u^n is a factor of w for each $n \geq 1$.

Corollary

Every infinite word w admits powers or anti-powers of arbitrarily large order.

Recurrence and uniform recurrence

Definition

An infinite word w is recurrent if all its factors show up twice.

Definition

An infinite word w is uniformly recurrent (almost periodic) if the distance between two consecutive occurrences of a given factor is bounded.

The importance of being uniformly recurrent

Theorem (Furstenberg - 1981)

Given an infinite word w, there exists an uniformly recurrent word x such that $F(x) \subset F(w)$.

- It is enough to understand the structure of the class of uniformly recurrent words.
- Periodic words, $w = u^{\infty}$, are (non-interesting) examples of uniformly recurrent words.

Aperiodic uniformly recurrent

Observe that an aperiodic uniformly recurrent word can never contain arbitrarily large powers of some factor u.

Theorem (Fici, Restivo, Zamboni, Silva)

Every aperiodic uniformly recurrent word w contains anti-powers of arbitrarily large order.

Factor complexity

- Given an infinite word w, let $F_w(n)$ be the set of factors of w with length n. $1 \leq |F_w(n)| \leq |A|^n, n \leq 1$.
- If w is periodic, then the function $F_w(n)$ is bounded .

Theorem (Hedlund, Morse - 1938)

Given an infinite aperiodic word w and $n \in \mathbb{N}$, we have

$$|F_w(n)| \geq n + 1.$$

The Thue-Morse word has linear factor complexity. A normal word has maximal complexity.

Abelian compelxity

- Given a finite word $u \in A^*$, the Parikh vector $\alpha(u) = (u_1, \cdots, u_k)$ counts the number of ocorrences of each letter $a_1, a_2, \cdots a_k$ from the finite alphabet A.
- We say that $u \sim v$ (abelian equivalent) if they share the same Parikh vector. v can be obtained from u by permutating the letters of u.
- We define the function $G_w(n)$ which counts the number of distinct Parikh vector for factors of length n.

Bounded abelian complexity

- u is a k-abelian power if it admits a decomposition $u = v_1 v_2 \cdots v_k$ with $|v_i| = |v_j|$ and $v_i \sim v_j, 1 \leq i < j \leq k$.
- A word w has bounded abelian complexity if any two factors of the same length u e v have at most L distinct letters.

Theorem (Richomme, Saari, Zamboni - 2011)

Let w be an infinite word with bounded abelian complexity. Then, w contains abelian powers of arbitrarily large order

Cyclic complexity

- Two words u and v are conjugated if $u = w_1 w_2$ and $v = w_2 w_1$. This is an equivalence relation.
- Given an infinite word w, the function $H_w(n)$ counts the number of conjugacy classes of factors of length n.

Theorem (Cassaigne, Fici, Sciortino, Zamboni - 2011)

Every aperiodic word w has unbounded cyclic complexity.

Complexity notions

Each one of the three notions of complexity is associated to actions of different subgroups of the symmetric group.

- Factor complexity is associated with the trivial group,
- Abelian complexity is associated with the symmetric group
- Cyclic complexity is associated with the cyclic subgroup generated by the permutation $(1, 2, 3, \cdots, n)$.

Far away and close by

Let w be an infinite word in a finite alphabet A and $a \in A$.

$S_1 = \{||u|_a - |v|_a| : |u| = |v|$ and uv is a factor of w$\}$.

$S_2 = \{||u|_a - |v|_a| : |u| = |v|$ and u and v are factors of w$\}$.

Proposition

S_1 is bounded if and only if S_2 is bounded.

This implies that an infinite word with unbounded abelian complexity will contain two consecutive blocks which are far apart.

Consecutive factors distant from each other

The distance (Hamming) between two equal-length words is the number of positions at which the corresponding symbols differ.

Proposition (Girão, Silva)

Given an infinite binary word w with an infinite number of zeros and ones and some $L > 0$, we can find two same length consecutive factors u and v such that the distance is at least L.

Theorem (Girão, Silva)

For every positive integers k,d, there exists $m(k,d)$ such that any word (on any alphabet) of size at least $m(k,d)$ contains either k-powers or k-anti-powers whose consecutive blocks are at distance at least d between each two.

The End

Thank you for your attention!

References

1. J. Cassaigne, G. Fici, M. Sciortino, L. Zamboni, *Cyclic complexity of words*, Journal Comb. Theory A (2017).
2. G. Fici, A. Restivo, M. Silva, L. Zamboni, *Anti-Powers in Infinite Words*, Journal Comb. Theory A, (2018).
3. G. Fici, M. Postic, M. Silva, *Abelian antipowers in infinite words*, Advances in Applied Mathematics (2019).
4. A. Girão, M. Silva, *Distant words*, in progress