Two results on automatic sequences

Narad Rampersad

(from discussions with J.-P. Allouche, J. Currie, and J. Shallit)

Department of Mathematics and Statistics

University of Winnipeg

- Recently, Jeffrey Shallit asked the two following questions by email to myself, James Currie, and Jean-Paul Allouche.
- Are there any k-automatic infinite overlap-free binary sequences for k not a power of 2?
- Can a k-automatic sequence have arbitrarily large factors in common with a Sturmian sequence?

- The answer to the first question is "no". Here is the proof.
- Suppose there exists some 3-automatic (say) binary overlap-free word $\mathbf{x}$.
- Then the set of lengths m such that $\mathbf{x}$ contains a square of length m is a 3-automatic set.
- However, the set of lengths of squares occurring in $\mathbf{x}$ is $\{2^n : n \geq 1\} \cup \{3 \cdot 2^n : n \geq 1\}$.
- This is not a 3-automatic set. Contradiction.

- OK, let's give some more details.
- A square is a repetition of the form xx, such as **tartar**.
- An overlap is a repetition of the form $axaxa$, such as **entente**.
- A word is overlap-free (resp. squarefree) if no factor is an overlap (resp. square).

- Overlap-free words over the binary alphabet are highly structured.
- The prototype is the Thue–Morse word

$$0110100110010110\cdots$$

- It is 2-automatic: it is the fixed point of the 2-uniform morphism

$$\mu : 0 \to 01, 1 \to 10.$$

- We want to show no infinite overlap-free binary word can be 3-automatic.
- In the proof sketch we derived this from the fact that the lengths of squares in an infinite overlap-free binary word are powers of 2 or 3 times a power of 2.
- Why is this?

Theorem (Restivo–Salemi)

Let x be an overlap-free word over $\{0, 1\}$. Then $x = u\mu(y)v$, where y is an overlap-free word and $u, v \in \{\epsilon, 0, 1, 00, 11\}$.

e.g., $1011010011001011 0 = 1\mu(01101001)$

- If $\mathbf{x}$ is infinite, then we have $\mathbf{x} = u\mu(\mathbf{y})$.
- We can iterate:

$$\mathbf{x} = u_1\mu(u_2)\mu^2(u_3)\mu^3(u_4)\cdots\mu^k(\mathbf{y}').$$

- "Locally" an infinite overlap-free binary word looks like the Thue–Morse word.

Recall the exercise: The squares in the Thue–Morse word are of the form $\mu^k(xx)$, where $x \in \{0, 1, 010, 101\}$.

Theorem

The overlap-free binary squares are the conjugates (cyclic shifts) of the squares in the Thue–Morse word.

- The proof is similar to the exercise, but uses the Restivo–Salemi Theorem.
- So the set of lengths of squares in an overlap-free binary word is $\{2^n : n \geq 1\} \cup \{3 \cdot 2^n : n \geq 1\}$.

A set A of non-negative integers is k-automatic if $\{\langle n \rangle_k : n \in A\}$ is a regular language (i.e., accepted by a deterministic finite automaton).

Theorem

Let $\mathbf{x}$ be a k-automatic sequence. Then

$$\{n : \mathbf{x} \text{ contains a square of length } n\}$$

is a k-automatic set.

The proof is the software Walnut. How does it work?

Theorem (Büchi–Bruyère)

A set is k-automatic if and only if it is k-definable.

- k-definable means definable in the first-order theory of $\langle \mathbb{N}, +, V_k \rangle$.
- V_k is the largest power of k that divides n.

- Given a k-automatic sequence $\mathbf{x}$, consider the set of all integers n that satisfy

$$(\exists i)(\forall j),(j < n) \Rightarrow (\mathbf{x}[i+j] = \mathbf{x}[i+n+j])$$

- Then the set of such n is the set of n for which "there is a square of length n starting at position i in $\mathbf{x}$".
- The previous theorem implies that there is an automaton that accepts the base-k representations of this set.

- Walnut implements an algorithm that builds precisely the automaton accepting this set.
- Using this one can prove that an automatic sequence is squarefree, overlap-free, etc.

- This also implies that if $\mathbf{x}$ is 3-automatic and overlap-free, then

$$\{n : \mathbf{x} \text{ contains a square of length } n\}$$

is a 3-automatic set.
- So $\{2^n : n \geq 1\} \cup \{3 \cdot 2^n : n \geq 1\}$ would be a 3-automatic set.
- Now this contradicts a classical theorem on automatic sets.

Theorem (Cobham)

Let k, ℓ be multiplicatively independent. Then a set A is k-automatic and ℓ-automatic if and only if A is a finite union of arithmetic progressions.

- Clearly $\{2^n : n \geq 1\} \cup \{3 \cdot 2^n : n \geq 1\}$ is 2-automatic.
- If it were 3-automatic, it would be a finite union of arithmetic progressions, which clearly it is not.
- The contradiction proves the result: If $\mathbf{x}$ is a k-automatic overlap-free binary word, then k is a power of 2.

Theorem

Let $\mathbf{x}$ be a k-automatic sequence and let $\mathbf{a}$ be a Sturmian sequence. There exists a constant C (depending on $\mathbf{x}$ and $\mathbf{a}$) such that if $\mathbf{x}$ and $\mathbf{a}$ have a factor in common of length n then $n \leq C$.

Here is a sketch of the proof:

- Since $\mathbf{x}$ and $\mathbf{a}$ agree on a long segment, they agree on a long arithmetic progression.
- Use the finiteness of the k-kernel of $\mathbf{x}$ to show that $\mathbf{a}$ contains repeated occurrences of a long factor whose starting positions are "close together" in this arithmetic progression.
- Show that this contradicts Kronecker's Theorem.

Let $\mathbf{x} = x_0 x_1 \cdots$ and $\mathbf{a} = a_0 a_1 \cdots$. Without loss of generality, suppose that $\mathbf{a}$ is a characteristic word. Then there exists an irrational α such that $\mathbf{a}$ is defined by the following rule:

$$a_n = \begin{cases} 0 & \text{if } \{(n+1)\alpha\} < \alpha \\ 1 & \text{otherwise.} \end{cases}$$

Here $\{\cdot\}$ denotes the fractional part of a real number.

Suppose that for some L, the words $\mathbf{x}$ and $\mathbf{a}$ have a factor of length L in common: i.e., for some $i \leq j$

$$x_i \cdots x_{i+L-1} = a_j \cdots a_{j+L-1}.$$

Suppose that the k-kernel

$$\{(x_{nk^r+s})_{n\geq 0} : r \geq 0 \text{ and } 0 \leq s < k^r\}$$

has Q distinct elements. Let r satisfy $k^r > Q$. Then there exist $s_1 < s_2$ such that

$$(x_{nk^r+s_1})_{n\geq 0} = (x_{nk^r+s_2})_{n\geq 0}.$$

Write

$$d_1 = s_1 + j - i + 1$$
$$d_2 = s_2 + j - i + 1.$$

For all n satisfying $i \leq nk^r + s_1$ and $i + L - 1 \leq nk^r + s_2$ we have

$$x_{nk^r+s_1} = a_{nk^r+d_1-1} \qquad \text{and} \qquad x_{nk^r+s_2} = a_{nk^r+d_2-1}.$$

Since $x_{nk^r+s_1} = x_{nk^r+s_2}$, we have

$$a_{nk^r+d_1-1} = a_{nk^r+d_2-1}.$$

This means that either both

$$\{(nk^r + d_1)\alpha\} < \alpha \text{ and } \{(nk^r + d_2)\alpha\} < \alpha \tag{1}$$

or both

$$\{(nk^r + d_1)\alpha\} \geq \alpha \text{ and } \{(nk^r + d_2)\alpha\} \geq \alpha. \tag{2}$$

If L is arbitrarily large, then there exist arbitrarily large sets I of consecutive positive integers such that every $n \in I$ satisfies either (1) or (2).

- Without loss of generality, suppose that $\{d_2\alpha\} > \{d_1\alpha\}$.
- Choose $\epsilon > 0$ such that $\epsilon < \{d_2\alpha\} - \{d_1\alpha\}$.
- Note that $d_2 - d_1 = s_2 - s_1$, so ϵ does not depend on L (or I).
- Since $k^r\alpha$ is irrational, if I is sufficiently large then by Kronecker's Theorem there exists $N \in I$ such that

$$\{Nk^r\alpha + d_2\alpha\} \in [\alpha, \alpha + \epsilon].$$

By the choice of ϵ, this implies that

$$\{Nk^r\alpha + d_2\alpha\} \geq \alpha \text{ and } \{Nk^r\alpha + d_1\alpha\} < \alpha,$$

contradicting the assumption that N satisfies one of (1) or (2). The contradiction means that L must be bounded by some constant C, which proves the theorem.

The End