Critical exponents of k-automatic words and generalizations

Jeffrey Shallit
School of Computer Science
University of Waterloo
Waterloo, Ontario N2L 3G1
Canada
shallit@cs.uwaterloo.ca
http://www.cs.uwaterloo.ca/~shallit

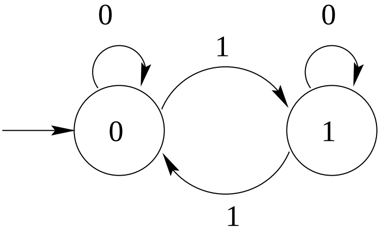
What is an automatic sequence?

- An infinite sequence

$$\mathbf{a} = a_0 a_1 a_2 \cdots$$

  over a finite alphabet of letters, generated by a finite-state machine (automaton)
- The automaton, given n as input, computes a_n as follows:
  - n is represented in some fixed integer base $k \geq 2$
  - The automaton moves from state to state according to this input
  - Each state has an output letter associated with it
  - The output on input n is the output associated with the last state reached

The canonical example: the Thue-Morse automaton



This automaton generates the *Thue-Morse sequence*

$$\mathbf{t} = (t_n)_{n \geq 0} = 0110100110010110 \cdots .$$

Why automatic sequences?

- A nontrivial class of self-similar sequences
- Many "naturally-occurring" sequences are automatic
- Halfway between periodic and chaotic
- Provide canonical examples for various kinds of avoidance problems

Repetitions in words

- A *square* is a word made up of two consecutive identical blocks, like the French word `chercher`.
- An *overlap* is a word made up of two consecutive identical blocks, followed by the first letter of the first block, like the French word `entente`.
  - Also called a 2^+-power.
- Thue proved (1912) that the Thue-Morse word is overlap-free.
- More generally, we can talk about n'th powers in words for integer $n \geq 2$.
- And fractional (p/q)-powers: a word of length p of period q.
  - e.g., `entente` is a $\frac{7}{3}$-power; the exponent of `entente` is $\frac{7}{3}$.

Critical exponents

- The *critical exponent* of a word **w** is the supremum, over all factors x of **w**, of the exponent of x.
- The critical exponent of **t** is 2.

Decidability

- Many questions about automatic sequences **a** are decidable
- For example, does **a** contain an overlap?
- Decidability sketch:
  - Phrase question as predicate $P(n)$ in $\mathrm{Th}(\mathbb{N}, +, V_k)$ where $V_k(n) = \sup\{k^e \;:\; k^e \mid n\}$
  - Use ideas of Büchi, Bruyère, Villemaire, Hodgson, ...
  - Implement predicate as automaton accepting $\{n \;:\; P(n) \text{ is true }\}$

Automatic predicate

Theorem

There is an algorithm that, given a predicate phrased using only the universal and existential quantifiers, indexing into a given automatic sequence $\mathbf{a}$*, addition, subtraction, logical operations, and comparisons, will decide the truth of that proposition.*

We call such a predicate an *automatic predicate*.

Repetitions

- Using our technique, we can express the property of having an overlap *axaxa* beginning at position N with $|ax| = p$, as follows: $\mathbf{a}[N..N+p] = \mathbf{a}[N+p..N+2p]$.
- So the corresponding automatic predicate for $\mathbf{t}$ is

$$\exists p \geq 1,\ N \geq 0\ \ \mathbf{t}[N..N+p] = \mathbf{t}[N+p..N+2p],$$

or, in other words,

$$\exists p \geq 1,\ N \geq 0\ \ \forall i,\ 0 \leq i \leq p\ \ \mathbf{t}[N+i] = \mathbf{t}[N+p+i].$$

Critical exponents

- We can check if a k-automatic sequence has α-powers, using the following predicate:

$$\exists N \geq 0,\ p, q \geq 1\ \ \mathbf{a}[N..N+p-q-1] = \mathbf{a}[N+q..N+p-1] \text{ and } p = \alpha q.$$

- However, this observation alone does not suffice to compute the critical exponent of **a**
- It turns out that the critical exponent is also computable for automatic sequences....

Representing rational numbers

- Represent rational number $\alpha = p/q$ by pair of integers (p, q), represented in base k; pad shorter with leading zeroes
- So representations of rationals are over the alphabet $\Sigma_k \times \Sigma_k$
- For example, if $w = [3, 0][5, 0][2, 4][6, 1]$ then $[w]_{10} = (3526, 41)$.
- Define $\text{quo}_k(x) = [\pi_1(x)]_k / [\pi_2(x)]_k$, where π_i is the projection onto the i'th coordinate
- So $\text{quo}_{10}(w) = 3526/41 = 86$.
- Canonical representations lack leading $[0, 0]$'s
- Every rational has infinitely many canonical representations, e.g., as $(1, 2), (2, 4), (3, 6), \ldots$, etc.

Automatic sets over $\mathbb{Q}^{\geq 0}$

- $\text{quo}_k(L) = \bigcup_{x \in L} \{\text{quo}_k(x)\}$

- $A \subseteq \mathbb{Q}^{\geq 0}$ is a k-automatic set of rationals if $A = \text{quo}_k(L)$ for some regular language $L \subseteq (\Sigma_k \times \Sigma_k)^*$.
- *not* the same notion as the automatic reals of Boigelot, Brusten, and Bruyère

Examples

Example 1. Let $k = 2$, $B = \{[0,0],[0,1],[1,0],[1,1]\}$, and consider

$$L_1 := B^*\{[0,1],[1,1]\}B^*.$$

Then L_1 consists of all pairs of integers where the second component has at least one nonzero digit — the point being to avoid division by 0. Then $\mathrm{quo}_k(L) = \mathbb{Q}^{\geq 0}$, the set of all non-negative rational numbers.

Example 2. Consider

$$L_2 = \{w \in (\Sigma_k^2)^* \ : \ \pi_1(w) \in 0^* C_k \text{ and } \pi_2(w) \in 0^* 1\}.$$

Then $\mathrm{quo}_k(L_2) = \mathbb{N}$.

Examples

Example 3. Let $k = 3$, and consider the language

$$L_3 := [0,1]\{[0,0],[2,0]\}^*.$$

Then $\mathrm{quo}_k(L_3)$ is the 3-*adic Cantor set*, the set of all rational numbers in the "middle-thirds" Cantor set with denominators a power of 3.

Example 4. Let $k = 2$, and consider

$$L_4 := [0,1]\{[0,0],[0,1]\}^*\{[1,0],[1,1]\}.$$

Then the numerator encodes the integer 1, while the denominator encodes all positive integers that start with 1. Hence

$$\mathrm{quo}_k(L_4) = \{\frac{1}{n} : n \geq 1\}.$$

Examples

Example 5. Let $k = 4$, and consider

$$S := \{0, 1, 3, 4, 5, 11, 12, 13, \dots\}$$

of all non-negative integers that can be represented using only the digits $0, 1, -1$ in base 4. Consider the language

$$L_5 = \{(p, q)_4 \; : \; p, q \in S\}.$$

It is not hard to see that L_5 is $(\mathbb{Q}, 4)$-automatic.
The main result in Loxton & van der Poorten [1987] can be rephrased as follows: $\mathrm{quo}_4(L_5)$ contains every odd integer.
In fact, an integer t is in $\mathrm{quo}_4(L_5)$ if and only if the exponent of the largest power of 2 dividing t is even.

Example 6. Consider

$$L_6 = \{w \in (\Sigma_k^2)^* \ : \ \pi_2(w) \in 0^*1^+0^*\}.$$

An easy exercise using the Fermat-Euler theorem shows that that $\text{quo}_k(L_6) = \mathbb{Q}^{\geq 0}$.

Example 7. For a word x and letter a let $|x|_a$ denote the number of occurrences of a in x. Consider the regular language

$$L_7 = \{w \in (\Sigma_2^2) \ : \ |\pi_1(w)|_1 \text{ is even and } |\pi_2(w)|_1 \text{ is odd}\}.$$

Then it follows from a result of Schmid [1984] that

$$\text{quo}_2(L_7) = \mathbb{Q}^{\geq 0} - \{2^n \ : \ n \in \mathbb{Z}\}.$$

Basic decidability properties

Given a DFA M accepting a language L representing a set of rationals S, can decide

- if $S = \emptyset$
- given $\alpha \in \mathbb{Q}^{\geq 0}$, whether there exists $x \in S$ with $x = \alpha$ (resp., $x < \alpha$, $x \leq \alpha$, $x > \alpha$, $x \geq \alpha$, $x \neq \alpha$, etc.)
- if $|S| = \infty$
- given a finite set $F \subseteq \mathbb{Q}^{\geq 0}$, if $F \subseteq S$ or if $S \subseteq F$
- given $\alpha \in \mathbb{Q}^{\geq 0}$, if α is an accumulation point of S

sup A is rational or infinite

Given a DFA M accepting $L \subseteq (\Sigma_k \times \Sigma_k)^*$ representing a set of rationals $A \subseteq \mathbb{Q}^{\geq 0}$, what can we say about $\sup A$?

Theorem. $\sup A$ is rational or infinite, and is computable.

Proof ideas: $\text{quo}_k(uv^i w)$ forms a monotonic sequence. Defining

$$\gamma(u, v) := \frac{[\pi_1(uv)]_k - [\pi_1(u)]_k}{[\pi_2(uv)]_k - [\pi_2(u)]_k}$$

one of the following three cases must hold:

(i) $\text{quo}_k(uw) < \text{quo}_k(uvw) < \text{quo}_k(uv^2 w) < \cdots < U$;
(ii) $\text{quo}_k(uw) = \text{quo}_k(uvw) = \text{quo}_k(uv^2 w) = \cdots = U$;
(iii) $\text{quo}_k(uw) > \text{quo}_k(uvw) > \text{quo}_k(uv^2 w) > \cdots > U$.

Furthermore, $\lim_{i\to\infty} \text{quo}_k(uv^i w) = U$.

$\sup A$ is rational or infinite

It follows that if $\sup A$ is finite, and the DFA M has n states, then $\sup A = \max T$, where

$$T = T_1 \ \cup \ T_2$$

and

$$\begin{aligned} T_1 &= \{\mathrm{quo}_k(x) \ : \ |x| < n \text{ and } x \in L\}; \\ T_2 &= \{\gamma(u,v) \ : \ |uv| \le n, \ |v| \ge 1, \ \delta(q_0, u) = \delta(q_0, uv), \\ &\qquad \text{and there exists } w \text{ such that } uvw \in L\}. \end{aligned}$$

$\sup A$ is computable

We know that $\sup A$ lies in the finite computable set T.

For each of $t \in T$, we can check to see if $t \geq \sup A$ by checking if $A \cap (t, \infty)$ is empty.

Then $\sup A$ is the least such t.

Computing the critical exponent

- Previously known to be computable for fixed points of uniform morphisms (Krieger)

Theorem. If $\mathbf{w}$ is a k-automatic sequence, then its critical exponent is rational or infinite. Furthermore, it is computable from the DFAO M generating w.

Proof sketch. Given M, we can transform it into another automaton M' accepting

$$\{(m,n) \; : \; \text{there exists } i \geq 0 \text{ such that } \mathbf{w}[i..i+m-1] \text{ has period } n\}.$$

We then apply our algorithm for computing $\sup(\mathrm{quo}_k(L))$ to $L(M')$.

The Leech word

Leech [1957] showed that the fixed point $\mathbf{l}$ of the morphism

$$\begin{aligned} 0 &\rightarrow 0121021201210 \\ 1 &\rightarrow 1202102012021 \\ 2 &\rightarrow 2010210120102 \end{aligned}$$

is squarefree.

We used our method to compute the critical exponent of this word. It is $15/8$.

Furthermore, if x is a $15/8$-power occurring in $\mathbf{l}$, then $|x| = 15 \cdot 13^i$ for some $i \geq 0$.

Generalization: fixed points of morphisms over $\mathbb{Z}$

- Example: morphism $h : i \to (0, i + 1)$
- Then $h^\omega(0) = 0102010301020104 \cdots$, the so-called "ruler function"
  - Naturally arises as the lexicographically least squarefree word over $\mathbb{N}$
- The critical exponent of $h^\omega(0)$ is 2

Nørgård's "infinity series"

- The sequence $(0, 1, -1, 2, 1, 0, -2, 3, \ldots)$
- Fixed point of $i \to (-i, i+1)$
- Used by Danish composer Per Nørgård in many of his musical compositions
- Critical exponent is $\frac{4}{3}$

Open Questions

- Extend these ideas to morphic sequences (fixed points of possibly non-uniform morphisms, followed by a coding)
  - Some ideas are extendable to, e.g., the Fibonacci word
  - Carton & Thomas proved that $(\mathbb{N}, <, \text{morphic word})$ is decidable
- Which predicates for automatic sequences (like squarefreeness) are decidable in polynomial time? Leroux has proved it for ultimate periodicity.

More open problems

- Extend these ideas to "infinite state" automata (i.e., fixed points of morphism like $n \to (an + b, cn + d)$) or prove undecidability
- Is $\sup\{x/y : (x, y)_k \in L\}$ computable for context-free languages L?
- Given a regular language $L \subseteq (\Sigma_k \times \Sigma_k)^*$ representing a a set $S \subseteq \mathbb{N} \times \mathbb{N}$ of pairs of natural numbers, is it decidable if S contains a pair (p, q) with $p \mid q$?
  - This is a question of $\exists^1(\mathbb{N}, +, V_k, |)$; of course $\mathrm{Th}(\mathbb{N}, +, |)$ is undecidable and $\exists^1(\mathbb{N}, +, |)$ is decidable (Lipshitz)

More Open Questions

- Prove or disprove: if L is a regular language with $\text{quo}_k(L) = \mathbb{Q}^{\geq 0}$, then L contains infinitely many distinct representations for infinitely many distinct rational numbers.

Which of the following questions is decidable? Given L representing a set of rationals S,

- Is there some rational $p/q \in S$ having infinitely many distinct representations in L?
- Are there infinitely many distinct rationals $p/q \in S$ having infinitely many distinct representations in L?