# Automata and Rational Numbers

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

# Representations of integers

- $\Sigma_k = \{0, 1, \ldots, k-1\}$
- Numbers are represented in base $k$ using digits in $\Sigma_k$
- So numbers are represented by words in $\Sigma_k^*$
- Canonical representation of $n$ denoted $(n)_k$, without leading zeroes
- Set of all canonical representations of integers is

$$C_k = \Sigma_k^* \setminus 0\Sigma_k^*$$

- If $w \in \Sigma_k^*$ then $[w]_k$ is the integer represented by $w$
- Sometimes useful to use least-significant-digit-first representation; sometimes most-significant-digit-first.

# Automatic sets over $\mathbb{N}$

- A $k$-automatic set of (non-negative) integers $A$ corresponds to a regular (rational) subset of $\Sigma_k^* \setminus 0\Sigma_k^*$
- Example:

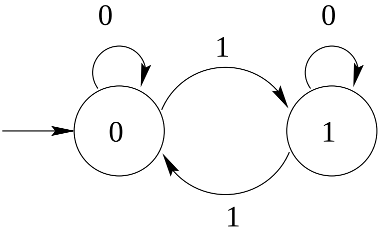
$$\mathbf{t} = 0110100110010110\cdots$$

- Let $T$ be the positions of the 1's in $\mathbf{t}$:

$$T = \{1, 2, 4, 7, 8, 11, 13, 14, \ldots\},$$

  i.e., those integers with an odd number of 1's in their base-2 representation
- $T$ is accepted by a DFA reading numbers expressed in base 2

# The Thue-Morse automaton



# Closure properties

The class of $k$-automatic sets is

- Closed under complement
  - interchange "finality" of states
- Closed under union, intersection:
  - use "direct product" construction
- Closed under the sum operation $A + B = \{a + b \,:\, a \in A,\ b \in B\}$.
  - On input $n$, "guess" $a$ and $b$, add and check if equal to $n$
- Closed under multiplication by constants

# Decidability

Given a $k$-automatic set $A \subseteq \mathbb{N}$, can we decide properties of $A$?

When we say "given a $k$-automatic set $A$", we really mean given a DFA $M$ accepting $A$.

- Given $A$ and $n$, we can decide if $n \in A$
- We can easily decide if $A = \emptyset$ (use DFS)
- We can easily decide if $|A| < \infty$ (look for useful cycles)

# 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 $\mathrm{quo}_k(x) = [\pi_1(x)]_k / [\pi_2(x)]_k$, where $\pi_i$ is the projection onto the $i$'th coordinate
- So $\mathrm{quo}_k(w) = 3526/41 = 86$.
- 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 $\mathbb{Q}^{\geq 0}$

- $\mathrm{quo}_k(L) = \bigcup_{x \in L} \{\mathrm{quo}_k(x)\}$
- $A \subseteq \mathbb{Q}^{\geq 0}$ is a *$k$-automatic set of rationals* if $A = \mathrm{quo}_k(L)$ for some regular language $L \subseteq (\Sigma_k \times \Sigma_k)^*$.

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

**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

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

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

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

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 $\mathrm{quo}_k(L_6) = \mathbb{Q}^{\geq 0}$.

# Examples

**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

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

# Intersection

Note that $\text{quo}_k(L_1 \cup L_2) = \text{quo}_k(L_1) \cup \text{quo}_k(L_2)$ but the analogous identity involving intersection need not hold.

**Example 8.** Consider $L_1 = \{[2,1]\}$ and $L_2 = \{[4,2]\}$. Then $\text{quo}_{10}(L_1 \cap L_2) = \emptyset \neq \{2\} = \text{quo}_{10}(L_1) \cap \text{quo}_{10}(L_2)$.

# Not the same as previous models of automatic rationals

– e.g., Boigelot-Brusten-Bruyère or Adamczewski-Bell

**Example 9.** Define

$$S = \{(k^m - 1)/(k^n - 1) \,:\, 1 \leq m < n\};$$

this is easily seen to be a $k$-automatic set of rationals.

However, the set of its base-$k$ expansions is of the form

$$\bigcup_{0<m<n<\infty} 0.(0^{n-m}\,(k-1)^m)^\omega,$$

where by $x^\omega$ we mean the infinite word $xxx\cdots$.

A simple argument using the pumping lemma shows that no Büchi automaton can accept this language.

# Don't demand lowest terms

Why not consider only representations $p/q$ with $\gcd(p, q) = 1$?

Two problems:

- the language of all such representations

$$\{w \in (\Sigma_k^2)^* \ : \ \gcd(\pi_1(w), \pi_2(w)) = 1\}$$

is not even context-free

- given a DFA accepting $(S)_k$, where $S \subseteq \mathbb{N} \times \mathbb{N}$, can we decide if $\gcd(p, q) = 1$ for all $(p, q) \in S$? Not known to be decidable!

# Don't demand unique or finite number of representations

No known way to represent $\mathbb{Q}^{\geq 0}$ as a regular language with every rational represented only a finite number of times.

# Closure properties

Automatic sets of rationals are closed under

- union;
- $S \to S + \alpha$ for $\alpha \in \mathbb{Q}^{\geq 0}$
- $S \to S \dot{-} \alpha$ for $\alpha \in \mathbb{Q}^{\geq 0}$;
- $S \to \alpha \dot{-} S$ for $\alpha \in \mathbb{Q}^{\geq 0}$;
- $S \to \alpha S$ for $\alpha \in \mathbb{Q}^{\geq 0}$;
- $S \to \{\frac{1}{x} : x \in S \setminus \{0\}\}$.

# 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 $\alpha$ is an accumulation point of $S$

# Deciding if $S \subseteq \mathbb{N}$

Important concept: let $S \subseteq \mathbb{N} \setminus \{0\}$. We say $S$ is **$k$-finite** if there exist

- an integer $n \geq 0$,
- $n$ positive integers $g_1, g_2, \ldots, g_n$ and
- $n$ ultimately periodic sets $W_1, W_2, \ldots, W_n \subseteq \mathbb{N}$ such that

$$S = \bigcup_{1 \leq i \leq n} \{g_i k^j \ : \ j \in W_i\}.$$

**Theorem.** Suppose there is a finite set of prime numbers $D$ such that each element of $S \subseteq \mathbb{N}$ is factorable into a product of powers of elements of $D$.

Then $S$ is $k$-automatic if and only if $S$ is $k$-finite.

# Deciding if $S \subseteq \mathbb{N}$

Furthermore, there is an algorithm that, given the DFA $M$ accepting $(S)_k$, will determine if $S$ is $k$-finite and if so, will produce the decomposition

$$S = \bigcup_{1 \le i \le n} \{g_i k^j \ : \ j \in W_i\}.$$

(Use reversed representations; follow path from start state on 0's; from each such state there can only be finitely many paths to an accepting state.)

# Deciding if $S \subseteq \mathbb{N}$

Suppose $M$ is a DFA with $n$ states accepting $L$ representing a set of rationals $S$.

Case 1: If $\alpha \in S$ and $\alpha$ is "small", say $\alpha \leq k^n$, then we can intersect $S$ with $[0, k^n]$, remove all representations of integers $0, 1, \ldots, k^n$, and see if any word is left. If so, then $S$ is not a subset of $\mathbb{N}$.

# Deciding if $S \subseteq \mathbb{N}$

Case 2: If $\alpha \in S$ and $\alpha$ is "big", say $\alpha > k^n$, then the numerator can be pumped, but the denominator stays the same. So the denominator must divide $[uvw]_k - [uw]_k$. But this is $k^{|w|}([uv]_k - [u]_k)$, and $|uv| \leq n$, so each denominator must divide a bounded number, times a power of $k$. So the set of all prime factors of all denominators is finite.

So the projection onto the set of denominators is $k$-finite.

We can easily remove powers of $k$. For the remaining finite set of denominators we can check if each denominator divides the numerator.

# 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.

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

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

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

# $\sup A$ 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 $t$.

# Applications 1: Critical exponent

We say a word $w$ is a $p/q$ power if we can write

$$w = \overbrace{xx \cdots x}^{n} x'$$

where

- $n = \lfloor p/q \rfloor$;
- $x'$ is prefix of $x$; and
- $p/q = n + |x'|/|x|$.

For example, the French word `entente` is a $\frac{7}{3}$-power.

The *exponent* of a finite word $w$ is defined to be the largest rational number $\alpha$ such that $w$ is an $\alpha$ power.

Given an infinite word $\mathbf{w}$, its *critical exponent* is defined to be the supremum, over all finite factors $x$ of $\mathbf{w}$, of $\exp(x)$.

# Critical exponents

Examples of critical exponent:

- the Thue-Morse word

$$\mathbf{t} = 0110100110010110\cdots$$

has critical exponent 2 (Thue)

- the Fibonacci word

$$\mathbf{f} = 01001010\cdots$$

has critical exponent $(5+\sqrt{5})/2$ (Mignosi & Pirillo)

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

**Theorem.** If **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 $i \geq 0$.

# Applications 2: Diophantine exponent

The *Diophantine exponent* of an infinite word $w$ is defined to be the supremum of the real numbers $\beta$ for which there exist arbitrarily long prefixes of **w** that can be expressed as $uv^e$ for finite words $u, v$ and rationals $e$ such that $|uv^e|/|uv| \geq \beta$.

(concept due to Adamczewski, Bugeaud)

**Theorem.** The Diophantine exponent of a $k$-automatic sequence is either rational or infinite. Furthermore, it is computable.

# Applications 3: Linear recurrence

A sequence $\mathbf{a}$ is said to be *recurrent* if every factor that occurs in $\mathbf{a}$ occurs infinitely often.

It is linearly recurrent if consecutive occurrences of factors of length $\ell$ appear at distance bounded by $C\ell$, for some constant $C$ independent of $\ell$.

For example, the Thue-Morse word

$$\mathbf{t} = 0110100110010110 \cdots$$

is linearly recurrent, but the Barbier infinite word

$$\mathbf{b} = 110111001011101111000 \cdots$$

is recurrent but not linearly recurrent.

**Theorem.** It is decidable if a given $k$-automatic sequence is linearly recurrent. If so, the optimal constant of linear recurrence is computable.

# Recurrence constant for the Rudin-Shapiro sequence

The Rudin-Shapiro sequence

$$0001001000011101\cdots$$

counts the parity of the number of 11's occurring in the binary representation of $n$.

We used our method to compute the recurrence constant for this sequence; it is 41.

# How about lim sup?

I don't know how to compute $\limsup(\mathrm{quo}_k(L))$ in general, where $L \subseteq (\Sigma_k \times \Sigma_k)^*$ is a regular language.

However, the largest *special point* can be computed.

A real number $\beta$ is a special point if there exists an infinite sequence $(x_j)_{j\geq 1}$ of distinct words of $L$ such that $\lim_{j\to\infty} \mathrm{quo}_k(x_j) = \beta$.

Thus, a special point is either an accumulation point of $\mathrm{quo}_k(L)$, or a rational number with infinitely many distinct representations.

Luckily, special points suffice for most applications involving lim sup.

# Application 4: Goldstein quotients

Let $\mathbf{x}$ be an infinite word and $\rho_{\mathbf{x}}(n)$ its *subword complexity*, the number of distinct factors of length $n$.

I. Goldstein [2011] showed that in some cases the quantities

$$\limsup_{n\geq 1} \frac{\rho_{\mathbf{x}}(n)}{n} \quad \text{and} \quad \liminf_{n\geq 1} \frac{\rho_{\mathbf{x}}(n)}{n}$$

are computable.

We can show that these are computable when $\mathbf{x}$ is a $k$-automatic sequence.

We can construct a DFA accepting

$$L := \{(\rho_{\mathbf{x}}(n), n)_k \ : \ n \geq 1\}$$

and then find the largest (resp., smallest) special point. This corresponds to the lim sup (resp., lim inf).

# Open Problems 1

Are the following problems decidable?

Given a DFA $M$ accepting a language $L$ representing a set of pairs of integers $S \subseteq \mathbb{N} \times \mathbb{N}$:

- does there exist a pair $(p, q) \in S$ with $p \mid q$?
- does there exist infinitely many pairs $(p, q) \in S$ with $p \mid q$?

# Open Problems 2

Is there a regular language $L \subseteq (\Sigma_k \times \Sigma_k)^*$ representing $\mathbb{Q}^{\geq 0}$ such that each rational has only finitely many representations?

# Open Problem 3: Cobham's theorem

Prove or disprove: let $S \subseteq \mathbb{Q}^{\geq 0}$ be a set of non-negative rational numbers. Then $S$ is simultaneously $k$-automatic and $\ell$-automatic, for multiplicatively independent integers $k, \ell \geq 2$, if and only iff there exists a semilinear set $A \subseteq \mathbb{N}^2$ such that $S = \{p/q \, : \, [p, q] \in A\}$.

# For Further Reading


1. Luke Schaeffer and Jeffrey Shallit, The critical exponent is computable for automatic sequences, *Int. J. Found. Comput. Sci.* **23** (2012), 1611–1626.
2. Eric Rowland and Jeffrey Shallit, $k$-automatic sets of rational numbers, *Proc. LATA 2012*, Lect. Notes in Computer Science, Vol. 7183, pp. 490–501.

# Thanks, JPA, for 28 years of collaboration!