Decidable properties of automatic sequences

Narad Rampersad

University of Liège

The Thue–Morse sequence

- the prototypical 2-automatic sequence:

$$0110100110010110\cdots$$

- generated by iterating the map

$$0 \to 01, \quad 1 \to 10$$

Properties of the Thue–Morse sequence

- aperiodic
- uniformly recurrent
- contains no block of the form xxx
- contains at most $4n$ blocks of length $n+1$ for $n \geq 1$
- etc.

Decidable properties

- We present algorithms to decide if an automatic sequence
  - is aperiodic
  - is recurrent
  - avoids repetitions
  - etc.
- We also describe algorithms to calculate its
  - complexity function
  - recurrence function
  - critical exponent
  - etc.

Automatic sequences

- A sequence is k-automatic if it is generated by first iterating a k-uniform morphism and then renaming some of the symbols.

The characteristic sequence of the powers of 2

- Iterate the 2-uniform morphism

$$a \to ab, b \to bc, c \to cc$$

to get the infinite sequence

$$abbcbcccbcccccccbcccccccccccccccbcc \cdots .$$

- Now recode by $a, c \to 0$; $b \to 1$:

$$01101000100000001000000000000000100 \cdots .$$

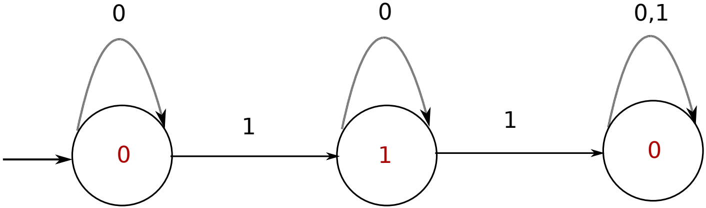
Determining periodicity

- Given a k-automatic sequence, can we tell if it is ultimately periodic?
- Honkala (1986) gave an algorithm.
- This result was often reproved: Muchnik (1991), Fagnot (1997), Allouche, R., and Shallit (2009).
- Leroux (2005) gave a polynomial time algorithm.

An automaton-based characterization

- The proof of Allouche et al. is perhaps the simplest.
- It is based on another characterization of automatic sequences:
- A sequence $\mathbf{a}$ is k-automatic if there exists a finite automaton with output that, when given the base-k representation of n as input, outputs the $(n+1)$-th term of $\mathbf{a}$.
- This is the original definition of an automatic sequence; the equivalence with the morphism-based definition is due to Cobham.

An automaton for the powers of 2



A logic-based characterization

- Another important characterization (Büchi–Bruyère):
- Let $V_k(x)$ denote the largest power of k that divides x.
- A sequence $\mathbf{a}$ is k-automatic if it is definable in the logical structure $\langle \mathbb{N}, +, V_k \rangle$.
- I.e., for each alphabet symbol b, there exists a first-order formula φ_b of $\langle \mathbb{N}, +, V_k \rangle$ such that

$$\mathbf{a}^{-1}(b) = \{n \in \mathbb{N} : \langle \mathbb{N}, +, V_k \rangle \models \varphi_b(n)\}.$$

Defining the powers of 2 using logic

- The characteristic sequence $\mathbf{a}$ of the powers of 2 has a simple definition in this formulation:

$$
\begin{aligned}
\mathbf{a}^{-1}(1) &= \{n \in \mathbb{N} : \langle \mathbb{N}, +, V_k \rangle \models (V_2(n) = n)\} \\
\mathbf{a}^{-1}(0) &= \{n \in \mathbb{N} : \langle \mathbb{N}, +, V_k \rangle \models \neg(V_2(n) = n)\}
\end{aligned}
$$

Decidability

Theorem (Bruyère 1985)

The first order theory of $\langle \mathbb{N}, +, V_k \rangle$ is decidable.

Putting all these ideas together

Theorem (Charlier, R., Shallit 2011)

If we can express a property of a k-automatic sequence $\mathbf{x}$ using quantifiers, logical operations, integer variables, the operations of addition, subtraction, indexing into $\mathbf{x}$, and comparison of integers or elements of $\mathbf{x}$, then this property is decidable.

Applying these ideas

- We can now apply these ideas to obtain algorithms to determine periodicity, recurrence, etc.
- A sequence $\mathbf{a}$ is ultimately periodic if and only if there exist integers $p \geq 1$ and $n \geq 0$ such that $\mathbf{a}(i) = \mathbf{a}(i + p)$ for all $i \geq n$.
- Hence there exists a decision procedure for determining the periodicity of k-automatic sequences.

Recurrence

- An infinite word is recurrent if every factor that occurs at least once in it occurs infinitely often.
- Equivalently, for each occurrence of a factor there exists a later occurrence of that factor.
- Equivalently, for every $n \geq 0$, $r \geq 1$, there exists $m > n$ such that $\mathbf{a}(n+j) = \mathbf{a}(m+j)$ for $0 \leq j < r$.

Uniform recurrence

- An infinite word is uniformly recurrent if every factor that occurs at least once occurs infinitely often with bounded gaps between consecutive occurrences.
- Equivalently, for every $r \geq 1$ there exists $t > 0$ such that for every $n \geq 0$ there exists $m \geq 0$ with $n < m < n + t$ such that $\mathbf{a}(n+i) = \mathbf{a}(m+i)$ for $0 \leq i < r$.

Deciding recurrence

- We obtain another proof of the following result:

Theorem (Nicolas and Pritykin 2009)

There is an algorithm to decide if a k-automatic sequence is recurrent or uniformly recurrent.

The k-kernel

- We now look at enumeration results.
- Recall that we have three equivalent characterizations of k-automatic sequences: uniform morphisms, automata, and logic.
- The k-kernel of a sequence $(a(n))_{n \geq 0}$ is the set

$$\{(a(k^e n + c))_{n \geq 0} \; : \; e \geq 0, \; 0 \leq c < k^e\}.$$

- A sequence is k-automatic if and only if its k-kernel is finite (Eilenberg).

k-regular sequences

- With this definition we can generalize the notion of k-automatic to the class of sequences over infinite alphabets.
- A sequence $(a(n))_{n\geq 0}$ is k-regular if the module generated by the set

$$\{(a(k^e n + c))_{n\geq 0} \; : \; e \geq 0, \; 0 \leq c < k^e\}$$

is finitely generated.

Factor complexity

- The following result generalizes slightly a result of Mossé (1996).
- Carpi and D'Alonzo (2010) proved a slightly more general result.

Theorem (Charlier, R., Shallit 2011)

Let $\mathbf{x}$ be a k-automatic sequence. Let $b(n)$ be the number of distinct factors of length n in $\mathbf{x}$. Then $(b(n))_{n\geq 0}$ is a k-regular sequence.

Palindrome complexity

- The following result generalizes a result of Allouche, Baake, Cassaigne and Damanik (2003).
- Carpi and D'Alonzo (2010) proved a slightly more general result.

Theorem (Charlier, R., Shallit 2011)

Let $\mathbf{x}$ be a k-automatic sequence. Let $c(n)$ be the number of distinct palindromes of length n in $\mathbf{x}$. Then $(c(n))_{n\geq 0}$ is a k-regular sequence.

Other numeration systems

- The previous results hold in a slightly more general setting.
- The automaton-based formulation of k-automatic sequences used numeration in base k.
- We can also consider other non-standard numeration systems.

Positional numeration systems

- A positional numeration system is an increasing sequence of integers $U = (U_n)_{n\geq 0}$ such that
  - $U_0 = 1$ and
  - $C_U := \sup_{n\geq 0} \lceil U_{n+1}/U_n \rceil < \infty$.
- It is linear if it satisfies a linear recurrence over $\mathbb{Z}$.

Greedy representations

- A greedy representation of a non-negative integer n is a word $w = w_{\ell-1} \cdots w_0$ over $\{0, 1, \dots, C_U - 1\}$ such that

$$\sum_{i=0}^{\ell-1} w_i U_i = n,$$

and for all j

$$\sum_{i=0}^{j-1} w_i U_i < U_j.$$

- $(n)_U$ denotes the greedy representation of n with $w_{\ell-1} \neq 0$.

U-automatic sequences

- An infinite sequence **x** is U-automatic if it is computable by a finite automaton taking as input the U-representation $(n)_U$ of n, and having $\mathbf{x}(n)$ as the output associated with the last state encountered.

The Fibonacci word

- Let $U = (1, 2, 3, 5, 8, 13, \ldots)$ be the sequence of Fibonacci numbers.
- Greedy U-representations do not contain 11.
- The well-known Fibonacci word

$$0100101001001010010100100101001\cdots$$

generated by the morphism $0 \to 01$, $1 \to 0$ is U-automatic.
- The $(n+1)$-th term is 1 exactly when the U-representation of n ends with a 1.

Pisot systems

- A Pisot number is a real algebraic integer greater than one such that all of its algebraic conjugates have absolute value less than one.
- A Pisot system is a linear numeration system whose characteristic polynomial is the minimal polynomial of a Pisot number.

Recognizability of addition

Theorem (Frougny and Solomyak 1996)

Addition is recognizable in all Pisot systems U, i.e., it can be performed by a finite letter-to-letter transducer reading U-representations with least significant digit first.

An equivalent logical formulation

Theorem (Bruyère and Hansel 1997)

Let U be a Pisot system. A sequence is U-automatic if and only if it is U-definable, i.e., it is expressible as a predicate of $\langle \mathbb{N}, +, V_U \rangle$, where $V_U(n)$ is the smallest U_i occurring in $(n)_U$ with a nonzero coefficient.

Passing to this more general setting

- By virtue of these results, all of our previous reasoning applies to U-automatic sequences when U is a Pisot system.
- Hence, there exist algorithms to decide periodicity, recurrence, etc. for sequences defined in such systems.
- Next we return again to the base-k setting.

Linear recurrence

- An infinite word $\mathbf{w}$ is linearly recurrent if it is recurrent and there exists a constant R such that for each factor u of $\mathbf{w}$, the distance between consecutive occurrences of u in $\mathbf{w}$ is at most $R|u|$.
- Given an automatic sequence, can we decide if it is linearly recurrent?
- Can we compute the constant R?

Representing pairs of integers

- the binary representation of 12 is 1100
- the binary representation of 37 is 100101
- we represent the pair $(12, 37)$ by

$$[0, 1], [0, 0], [1, 0], [1, 1], [0, 0], [0, 1]$$

- the sequence of first components gives 001100
- the sequence of second components gives 100101
- we denote the representation of (x, y) in base k by $(x, y)_k$

A technical result

Theorem (Shallit 2011)

Let $X \subseteq \mathbb{N}^2$ and let $k \geq 2$. Suppose that

$$\{(x, y)_k : (x, y) \in X\}$$

is accepted by a finite automaton. The quantity

$$\sup\{x/y : (x, y) \in X\}$$

is either rational or infinite and can be effectively computed.

Linear recurrence

- For a k-automatic sequence $\mathbf{a}$, one can construct a finite automaton to accept the set X of all pairs $(n, l)_k$ such that:
  - there exists $i \geq 0$ such that for all j, $0 \leq j < l$ we have $\mathbf{a}(i + j) = \mathbf{a}(i + n + j)$, and
  - there exists no t, $0 < t < n$ such that for all j, $0 \leq j < l$ we have $\mathbf{a}(i + j) = \mathbf{a}(i + t + j)$.
- The constant of linear recurrence is

$$\sup\{x/y : (x, y) \in X\}.$$

Decidability of linear recurrence

Theorem (Shallit 2011)

Given a k-automatic sequence, there is an algorithm to decide if it is linearly recurrent, and if so, to compute its recurrence constant.

Critical exponent

- A word w with period p has an exponent $|w|/p$.
- The exponent of w is its largest exponent.
- The critical exponent of an infinite word is the supremum of the exponents of its finite factors.
- The Thue–Morse word has critical exponent 2.
- The Fibonacci word has critical exponent $2 + \varphi$.

An expression for the critical exponent

- Krieger showed that the critical exponent of the fixed point of a uniform morphism is either rational or infinite.
- For a sequence $\mathbf{a}$, let X be the set of all pairs (q, p) such that there exists a factor of $\mathbf{a}$ of length q with period p.
- If $\mathbf{a}$ is k-automatic, we can construct a finite automaton to accept $\{(q, p)_k : (q, p) \in X\}$.
- The critical exponent is $\sup\{q/p : (q, p) \in X\}$.

Calculating the critical exponent

Theorem (Shallit 2011)

Given a k-automatic sequence, its critical exponent is either rational or infinite and can be effectively computed.

What remains to be done

- Recall: automatic sequences are generated by uniform morphisms (with some possible recoding of the alphabet)
- The general case consists of morphic sequences: those generated by possibly non-uniform morphisms (again with a final recoding of the alphabet).
- Our techniques do not seem to apply in this setting.
- Some partial results are known (typically for purely morphic sequences).
- Finding decision procedures for periodicity, etc. in the general setting remains an open problem.

The End