Automatic Sequences

and why I am interested

Wieb Bosma
Radboud Universiteit Nijmegen

Streams Seminar
25 January 2011
Most material taken from book *Automatic Sequences* by Jean-Paul Allouche and Jeffrey Shallit (Cambridge University Press, 2003), and paper *Number theory and formal languages* by J.O. Shallit, (Springer 1999).

*Automatic sequences form a class of sequences somewhere between simple order and chaotic disorder.*
PART I

automata and sequences
A **deterministic finite automaton** starts in a distinguished initial state, reads letters from a finite input word, which determine transitions to finite number of states, to end in one of the final states, which is either accepting or rejecting for the input word.

$$A = (Q, \Sigma, \delta, q_0, F):$$

- $Q$ finite set of states,
- $\Sigma$ finite alphabet,
- $\delta : Q \times \Sigma \rightarrow Q$ transitions,
- $q_0 \in Q$ initial state,
- $F \subset Q$ accepting states.
4.1 Finite Automata

Figure 4.1: DFA Accepting Strings with No Two Consecutive 1's.

Figure 4.5: Minimal DFA for $L = \{0, 1\}^* \{0, 1\}$. 
Theorem 1  A language is accepted by a deterministic finite state automaton if and only if it can be specified by a regular expression.

Theorem 2  There is a unique deterministic finite state automaton with minimal number of states accepting a given regular language.
A non-deterministic finite state automaton allows any finite number (including zero) of transitions from a state for a given input letter. A word is accepted if and only if there exists a choice of transitions leading to an accepting state.

A deterministic finite state automaton with output produces an output symbol at final states (instead of a Boolean), and hence defines a finite state function from $\Sigma^*$ to (a possibly different output alphabet) $\Delta$.

A transducer produces a word over the output alphabet $\Delta$ for every transition $(q, a) \in Q \times \Sigma$. It is uniform if all output words have the same length.
examples

Figure 5.6: Building the Regular Paperfolding Sequence: Twelve Folds.

Figure 5.7: 2-DFAO for the Paperfolding Sequence.
examples

Figure 4.6: DFAO Computing the mod-2 Sum of its Input Bits.

Figure 4.7: A Finite-State Transducer.
A sequence \((a_n)_{n \geq 0}\) is a **k-automatic sequence** if there exists a deterministic finite state automaton which produces for all \(n \geq 0\) on input the base-\(k\) representation \([n]_k\) of \(n\), the element \(a_n\) as output.
Define the Thue-Morse sequence \((t_n)_{n \geq 0}\) by \(t_n = \sum_{i=0}^{\infty} d_i \mod 2\) if \(n = \sum_{i=0}^{\infty} d_i 2^i\) is the binary expansion of \(n\). It starts \(0, 1, 1, 0, 1, 0, 0, 1, 1, 0, \ldots\).

**Theorem 3** The Thue-Morse sequence is 2-automatic.

\[\text{Figure 5.1: Automaton Generating the Thue–Morse Sequence.}\]
Define the **Rudin-Shapiro** sequence \((r_n)_{n \geq 0}\) by \(r_n = 1\) or \(-1\) according to whether the number of pairs of consecutive 1’s in the binary expansion of \(n\) is even or odd. It starts 1, 1, 1, −1, 1, 1, −1, 1, 1, 1, −1, ….

**Theorem 4** *The Rudin-Shapiro sequence is 2-automatic.*

![DFAO Generating the Rudin–Shapiro Sequence.](image-url)
Define the Baum-Sweet sequence \((b_n)_{n \geq 0}\) by \(b_n = 1\) or \(0\) according to whether the number of blocks of odd length of consecutive 0’s in the binary expansion of \(n\) is zero or positive. It starts \(1, 1, 0, 1, 1, 0, 0, 1, 0, 1, 0, 0, 1, 0, 0, 1, \ldots\).

**Theorem 5**  *The Baum-Sweet sequence is 2-automatic.*
For positive integer $k$ and output letter $d$, define the fiber $I_k(a,d)$ of the sequence $a$ as the set of input words $[n]_k$ such that $a_n = d$.

**Theorem 6** $a$ is $k$-automatic if and only if every $I_k(a,d)$ is a regular language.

Define the $k$-kernel of $a$ to be the set of subsequences

$$\{(a_{kn+i}.n+j)_{n \geq 0} : i \geq 0, 0 \leq j < k^i\}.$$  

**Theorem 7** $a$ is $k$-automatic if and only if the $k$-kernel is finite.
Theorem 8  If \((a_n)_{n \geq 0}\) is ultimately periodic, then it is \(k\)-automatic for all \(k \geq 2\).

Theorem 9  \((a_n)_{n \geq 0}\) is ultimately periodic if and only if it is \(1\)-automatic.

Theorem 10 (Cobham)  \((a_n)_{n \geq 0}\) is ultimately periodic if it is \(k\)- and \(l\)-automatic, for multiplicatively independent \(k, l \geq 2\).
A \emph{k-uniform morphism} is a map $\phi : \Sigma^* \to \Sigma^*$ such that $\phi(xy) = \phi(x)\phi(y)$ for all words $x, y$ and such that the length of $\phi(a)$ is $k$ for all letters $a \in \Sigma$.

If $\phi(a) = ax$ for some letter $a$ then

$$\phi^\omega(a) = ax\phi(x)\phi^2(x) \cdots$$

is the unique fixed point of $\phi$ starting with $a$.

**Theorem 11 (Cobham)** $a$ is a $k$-automatic sequence if and only if it is fixed point of a $k$-uniform morphism (followed by a coding).
The Thue-Morse sequence is the fixed point of $\phi$ defined by $\phi(0) = 01$ and $\phi(1) = 10$. Indeed,

$$01\phi(1)\phi^2(1)\phi^3(1) \cdots = 011010011001011010 \cdots .$$

The Baum-Sweet sequence is obtained from the fixed point of

$$\phi(a) = ab, \phi(b) = cb, \phi(c) = bd, \phi(d) = dd,$$

followed by replacing $a, b$ by 1 and $c, d$ by 0. Indeed, we obtain

$$abcbbdcbcbddbdcb \cdots ,$$

which becomes

$$110110010100100101 \cdots .$$
Theorem 12  The image of an automatic sequence under a uniform transducer is again automatic.

A morphic sequence is the fixed point of a morphism (after a coding).

The infinite Fibonacci word is the morphic sequence that is the fixed point of $\phi(0) = 01$ and $\phi(1) = 0$. We get

$$f = 010010100100101001010\ldots.$$  

Theorem 13  The image of a morphic sequence under a transducer is finite or again a morphic sequence.
PART II

numbers and functions
Conway's prime-producing machine

Figure 7. The Seven States and Fourteen Transitions.
**Regular continued fractions** for real numbers are of the form

\[
a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cdots}}},
\]

with positive integers \(a_i\). Denoted \([a_0, a_1, a_2, a_3, \ldots]\).

(In)finite continued fractions represent (ir)rational numbers.

For irrational \(x\) the infinite sequence of \(a_i\) is obtained by putting \(x_0 = x\) and repeating

\[
a_i = \lfloor x_i \rfloor, \quad x_{i+1} = \frac{1}{x_i - a_i}.
\]

Alternative (semi-regular) continued fraction expansions are obtained by rounding differently.
Important because this furnishes infinite sequence of (best) rational approximations by the *convergents*

\[
\frac{p_n}{q_n} = [a_0, a_1, a_2, \ldots, a_n],
\]

each satisfying

\[
\left| x - \frac{p_n}{q_n} \right| < \frac{1}{q_n^2}.
\]

**Theorem 14** *The regular continued fraction expansion of* \(x\) *is ultimately periodic if and only if* \(x\) *is a quadratic irrational number.*

For example,

\[
\sqrt{67} = [8, 5, 2, 1, 1, 7, 1, 1, 2, 5, 16].
\]
transcendental

This automaton generates the continued fraction expansion of

\[ f(B) = \sum_{k=0}^{\infty} \frac{1}{B^{2^k}}, \]

a transcendental number, for any integer \( B \geq 3 \).
The Thue-Morse number

\[ \sum_{i=0}^{\infty} t_i 2^{-i} \]

is also transcendental.
Thus we find two classes and lots of examples of real numbers having bounded partial quotients: the rational numbers (finite) and the quadratic irrationalities (ultimately periodic) and some transcendental numbers.

**Conjecture 1** If a real number $x$ has bounded partial quotients then it is either rational, quadratic irrational, or transcendental.

Note: this is no longer true for complex continued fractions!

There are also interesting questions related to sums of real numbers with bounded partial quotients: M. Hall proved that every real number in the unit interval is the sum of two reals with partial quotients at most 4.
Unfortunately, continued fraction representations are hopeless for doing arithmetic. The main result is by Raney.

**Theorem 15**  For integers $a, b, c, d$ with $ad - bc \neq 0$, there exits a finite transducer that, on input the regular continued fraction representation of $x$, produces the continued fraction representation of

$$y = \frac{ax + b}{cx + d}.$$
example
Raney uses the \( LR \)-representation, which is closely related to Farey-fractions and the Stern-Brocot representation for reals.

The numerators \( p_n \) and denominators \( q_n \) for the convergents of \( x = [a_0, a_1, a_2, \ldots] \) satisfy the same recursion

\[
p_n = a_n p_{n-1} + p_{n-2}, \quad q_n = a_n q_{n-1} + q_{n-2},
\]

if we put

\[
p_{-1} = 0, p_{-2} = 1, \quad q_{-1} = 1, q_{-2} = 0.
\]
It turns out that matrices like

\[
\begin{pmatrix}
p_{n+1} & p_n \\
q_{n+1} & q_n
\end{pmatrix}
\]

can be written as product

\[L^{a_0} R^{a_1} L^{a_2} \ldots L^{a_n}\]

(or a slight variant, depending on the parity of \(n\)) where

\[
L = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}, \quad R = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix},
\]

so

\[
L^{a_i} = \begin{pmatrix} 1 & 0 \\ a_i & 1 \end{pmatrix}, \quad R^{a_j} = \begin{pmatrix} 1 & a_j \\ 0 & 1 \end{pmatrix}.
\]
\[ A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} 4 & 1 \\ 1 & 2 \end{pmatrix}, \quad C = \begin{pmatrix} 3 & 1 \\ 2 & 3 \end{pmatrix}, \quad D = \begin{pmatrix} 4 & 3 \\ 3 & 4 \end{pmatrix}, \quad C' = \begin{pmatrix} 3 & 2 \\ 1 & 3 \end{pmatrix}, \quad B' = \begin{pmatrix} 2 & 1 \\ 1 & 4 \end{pmatrix}, \quad A' = \begin{pmatrix} 1 & 0 \\ 0 & 7 \end{pmatrix} \]

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>C'</th>
<th>B'</th>
<th>A'</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>L^7:R</td>
<td>LR:R^3</td>
<td>L^2R:R^2</td>
<td>L^3R:R</td>
<td>L^4R:RL</td>
<td>L^5R:RL^2</td>
</tr>
<tr>
<td>D</td>
<td>L:LR^2</td>
<td>R:RL^3</td>
<td>L:R</td>
<td>R:R</td>
<td>LR:RL</td>
<td>L^2R:R</td>
</tr>
<tr>
<td>A'</td>
<td>R^6L:LR^6</td>
<td>R^3L:LR^2</td>
<td>R^4L:LR</td>
<td>R^3L:R</td>
<td>R^2L:L^2</td>
<td>R^1L:R</td>
</tr>
</tbody>
</table>

Transition table for \( \mathcal{F}_{7,1} \)

State-graph for \( \mathcal{F}_{7,1} \)
Let \((a_n)_{n \geq 0}\) be a sequence over the finite alphabet \(\Delta\), let \(p\) be a prime number, and \(\mathbb{F}_{p^m}\) the finite field of \(p^m\) elements.

**Theorem 16 (Christol)** \((a_n)_{n \geq 0}\) is \(p\)-automatic if and only if there exist a positive integer \(m\) and an injection \(\iota\) of \(\Delta\) in \(\mathbb{F}_{p^m}\) such that

\[
\sum_{k=0}^{\infty} \iota(a_k)X^k
\]

is algebraic over \(\mathbb{F}_{p^m}(X)\).
Let \((t_n)_{n \geq 0}\) be the Thue-Morse sequence again, and

\[
T(X) = \sum_{k=0}^{\infty} t_k X^k \in \mathbb{F}_2[[X]].
\]

Since \(t_{2n+1} = t_n + 1\) we find

\[
T(X) = \sum_{k=0}^{\infty} t_{2k} X^{2k} + X \sum_{k=0}^{\infty} (t_k + 1) X^{2k}
\]

and hence

\[
T(X) = T(X)^2 + XT(X)^2 + X \frac{1}{1 - X^2},
\]

so \(T\) satisfies

\[
(1 + X)^3 T^2 + (1 + X)^2 T + X = 0.
\]


