Automata and nested recurrences

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

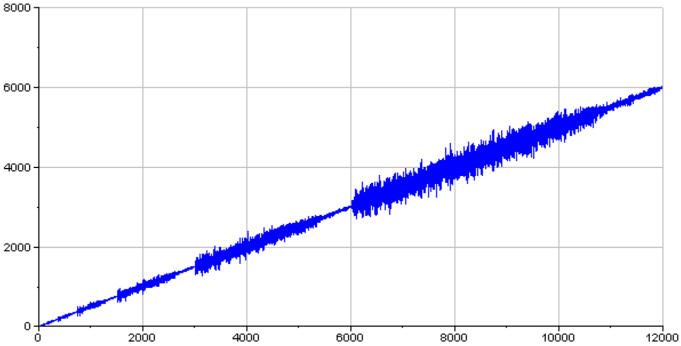
Joint work with Jean-Paul Allouche.

Hofstadter's recurrence

Douglas Hofstadter, *Gödel, Escher, Bach*, 1979:

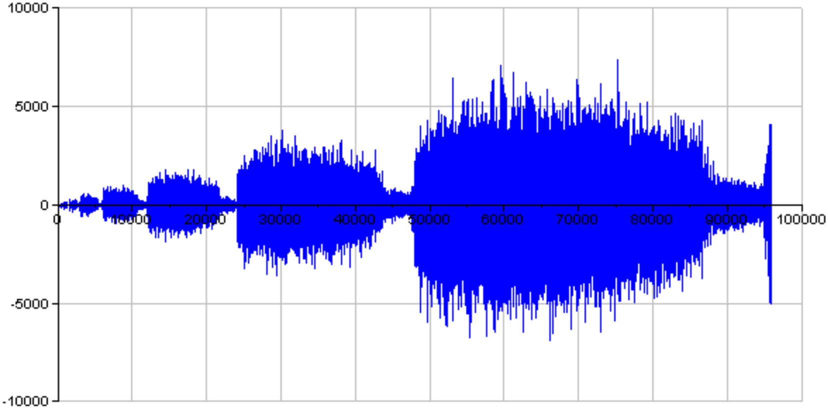
$$Q(n) = Q(n - Q(n-1)) + Q(n - Q(n-2))$$

for $n \geq 2$ and $Q(1) = Q(2) = 1$.



1st differences of Hofstadter's Q-sequence

Larger and larger hedgehogs



Hofstadter and Huber (1999):

$$Q_{r,s}(n) = Q_{r,s}(n - Q_{r,s}(n-r)) + Q_{r,s}(n - Q_{r,s}(n-s))$$

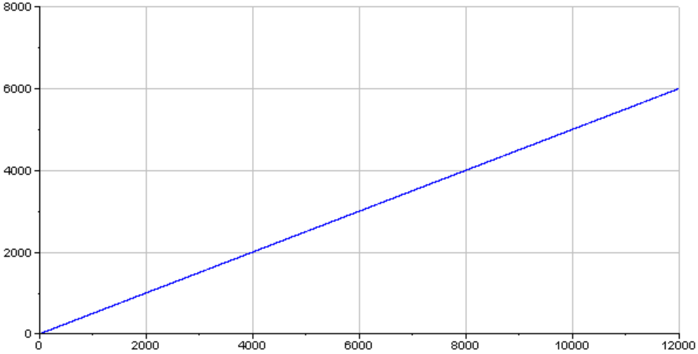
for $n > s > r$.
Balamohan, Kuznetsov and Tanny (2007): a nearly complete analysis of the sequence $Q_{1,4}$ (called V in their paper). It is defined by

$$V(1) = V(2) = V(3) = V(4) = 1, \text{ and}$$

$$\forall n > 4, \; V(n) := V(n - V(n-1)) + V(n - V(n-4)).$$

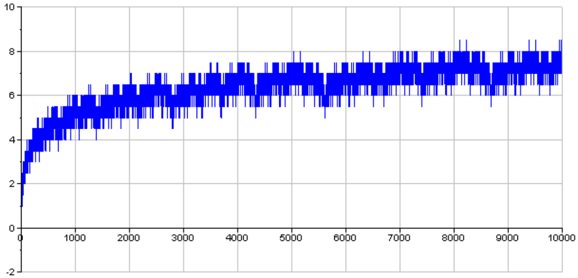
The V-sequence

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14
$V(n)$	1	1	1	1	2	3	4	5	5	6	6	7	8	8



The V-sequence

- is monotone increasing
- successive terms differ by 0 or 1
- no number appears more than four times (and only 1 appears that many)
- grows approximately like $n/2$



$V(n) - n/2$

Frequency sequence

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14
$V(n)$	1	1	1	1	2	3	4	5	5	6	6	7	8	8

Balamohan, Kuznetsov, and Tanny (2007): a precise description of the "frequency" sequence $F(n)$ defined by

$$F(a) := \#\{n : V(n) = a\}.$$

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14
$F(n)$	4	1	1	1	2	2	1	2	2	1	3	2	1	2

Explicit description

Theorem (Balamohan, Kuznetsov, Tanny)

There exist two (explicit) maps g, h, *with* $g, h : \{1, 2, 3\}^4 \to \{1, 2, 3\}$, *such that, for all* $a > 3$

$$\begin{aligned} F(2a) &= g(F(a-2), F(a-1), F(a), F(a+1)) \\ F(2a+1) &= h(F(a-2), F(a-1), F(a), F(a+1)). \end{aligned}$$

Corollary of our main result

Theorem. (Allouche & JOS, 2012) The sequence $(F(n))_{n\geq 1}$ is 2-automatic.

- This means that $F(n)$ can be computed "in a simple way" from the base-2 representation of n.
- In particular, it can be computed in $O(\log n)$ time.
- Furthermore, we can compute the automaton explicitly.

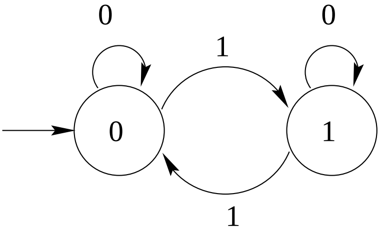
What is an automatic sequence?

- 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
- Statistics like the limiting frequency of occurrence of any letter are easy to compute

The main result

Theorem.
If a sequence F satisfies a recurrence "like"

$$\begin{aligned} F(2a) &= g(F(a-2), F(a-1), F(a), F(a+1)) \\ F(2a+1) &= h(F(a-2), F(a-1), F(a), F(a+1)). \end{aligned}$$

then F is 2-automatic.

Proof idea. Find a finite class of sequences of which F is a member, and which is closed under the transformations

$$n \to 2n; \qquad n \to 2n+1.$$

The proof is explicit, but the automaton it constructs in our case is too large to deal with.

Constructing the automaton

We constructed the automaton by "guessing" it and checking it.
The states are named by binary strings.
We guess that x and y correspond to the same state if $F([xz])$ and $F([yz])$ are the same for all the z we can reasonably test (say $|z| \leq 20$).
Actually, this is not quite good enough; we actually deal with the 4-tuple

$$(F([x]-2), F([x]-1), F([x]), F([x]+1)).$$

This gives us a 33-state automaton that we can now verify by a very tedious induction on $|x|$, using the recurrence we know.
Now, throwing away all but the 3rd of 4 outputs for each state, we get a smaller automaton by minimization.

A conjecture

Conjecture. V', the first difference sequence of $V(n)$, is also 2-automatic.

For example, up to the limit we have checked (tens of thousands of terms)

$$\begin{aligned}
V'([10110x]) &= V'([10001x]) \\
V'([11000x]) &= V'([10011x]) \\
V'([100011x]) &= V'([11110x]) \\
V'([101010x]) &= V'([100000x]) \\
V'([101011x]) &= V'([100001x]) \\
V'([101110x]) &= V'([100100x]) \\
V'([101111x]) &= V'([100101x])
\end{aligned}$$

However, if it is 2-automatic, then the smallest automaton (msd first) has at least 900 states.

Morphic sequences: a generalization of automatic sequences

A *morphism* is a map h from a finite alphabet $\Sigma^* \to \Delta^*$ satisfying

$$h(xy) = h(x)h(y)$$

for all strings x, y.

If $\Sigma = \Delta$ we can iterate h.

If $h(a) = ax$ for a letter a then

$$h^\omega(a) = a\,x\,h(x)\,h^2(x)\,\ldots .$$

Note that $h^\omega(a)$ is a fixed point of h.

If a sequence arises by applying a coding (a renaming of the letters) to $h^\omega(a)$ then it is called *morphic*.

Slow sequences

Recently Allouche has proposed studying “slow” sequences, that is, sequences of natural numbers where the first differences are either 0 or 1.

Theorem. Let $\mathbf{x} = (x_n)_{n \geq 1}$ be a slow sequence. Then the corresponding frequency sequence $\mathbf{F}(\mathbf{x})$ is bounded and morphic iff the the first difference sequence $\Delta\mathbf{x}$ is morphic and 1’s appear there with bounded gaps.

Corollary The first difference sequence of V is morphic.

Proof of the theorem

Proof. Apply the morphism that sends i to $1\overbrace{0\,0\,\cdots\,0}^{i-1}$ to the frequency sequence.

Example: original sequence $\mathbf{x}$ is

$$1\,1\,1\,1\,2\,3\,4\,5\,5\,6\,6\,7\,8\,8\,9\,9\,10\,11\,11\,11\cdots$$

Frequency sequence is

$$41112212213\cdots$$

Apply morphism to that to get

$$10001111010110101100\cdots,$$

which, except for the first term, is $\Delta\mathbf{x}$.

Proof of the other direction

For the other direction, apply a finite-state transducer that measures the distance between consecutive 1's:

Example: if the first difference sequence is

$$(1)000111101011010110 01\cdots,$$

then the transducer outputs

$$41112212213\cdots.$$

However, these implications need not hold for k-automatic sequences...

Example 1

Suppose the frequency sequence $F(\mathbf{x}) = (f(n))_{n\geq 1}$ is given by

$$f(n) = \begin{cases} 2, & \text{if } n \text{ is one less than a power of 2;} \\ 1, & \text{otherwise.} \end{cases}$$

So $(f(n))_{n\geq 1}$ is

$$212112111111211111111111111121\cdots$$

Then $\mathbf{x}$ is

$$1\,1\,2\,3\,3\,4\,5\,6\,7\,7\,8\,9\,10\,11\,12\,13\,14\,15\,15\,16\,17\cdots$$

and the first difference of this is

$$011011110111111110111111111111\cdots$$

with 0's in positions

$$1\,4\,9\,18\,35\,68\cdots$$

and generally in positions $2^n + n - 2$. This is not 2-automatic.

Example 2

On the other hand if we define the first difference sequence by

$$g(n) = \begin{cases} 0, & \text{if } n \text{ is a power of 2;} \\ 1, & \text{otherwise.} \end{cases}$$

Then $(g(n))_{n \geq 1}$ is automatic and equals

$$0010111011111110 1111 \cdots$$

so the original sequence is

$$1\,1\,1\,2\,2\,3\,4\,5\,5\,6\,7\,8\,9\,10\,11\,12\,12\,13\,14\,15\,16\,17\,18\,19 \cdots$$

and the corresponding frequency sequence is

$$321121111112111111111111112 \cdots$$

with 2's in positions

$$2\,5\,12\,27\,58 \cdots$$

and in general $2^n - n$, so this is not automatic either.

For further reading

1. J.-P. Allouche and J. Shallit, A variant of Hofstadter's sequence and finite automata, *J. Aust. Math. Soc.* **93** (2012), 1–8.
2. B. Balamohan, A. Kuznetsov, and S. Tanny, On the behavior of a variant of Hofstadter's Q-sequence, *J. Integer Seq.* **10** (2007) Article 07.7.1.
3. K. Pinn, Order and chaos in Hofstadter's $Q(n)$ sequence, *Complexity* **4** (1999), 41–46.