Least Periods of $k$-Automatic Sequences

Daniel Goč and Jeffrey Shallit
School of Computer Science
University of Waterloo
Waterloo, ON N2L 3G1
Canada
dgoc@cs.uwaterloo.ca, shallit@cs.uwaterloo.ca

May 2, 2014

Abstract

Currie and Saari initiated the study of least periods of infinite words, and they showed that every integer $n \geq 1$ is a least period of the Thue-Morse sequence. We generalize this result to show that the characteristic sequence of least periods of a $k$-automatic sequence is (effectively) $k$-automatic. Through an implementation of our construction, we confirm the result of Currie and Saari, and we obtain similar results for the period-doubling sequence, the Rudin-Shapiro sequence, and the paperfolding sequence.

1 Introduction

In a recent paper, Currie and Saari [5] initiated the study of the least periods of infinite words. If $x = a_1 \cdots a_n$ is a finite word, then we say $x$ has period $p \geq 1$ if $a_i = a_{i+p}$ for $1 \leq i \leq n - p$. For example, alfalfa has period 3 and entanglement has period 9. If $x = a_0a_1a_2\cdots$ is an infinite word, then a factor of $x$ is a contiguous subword of the form $a_ia_{i+1}\cdots a_j$ for some $i, j$ with $0 \leq i \leq j + 1$; we write it as $x[i..j]$. (If $i = j + 1$ then the factor is $\epsilon$, the empty string.)

Currie and Saari were interested in the set of all positive integers that can be the least period of some finite nonempty factor of $x$. They explicitly computed the set of least periods for some famous infinite words, such as the Thue-Morse sequence.

The Thue-Morse sequence $t = 0110100110010110\cdots$ is defined by letting $t[n]$ be the sum of the bits in the binary expansion of $n$, taken modulo 2. They proved that every positive integer can be the least period of the Thue-Morse sequence.

The Thue-Morse sequence is one of a much larger class of infinite words called “automatic”. Roughly speaking, an infinite word $x$ is $k$-automatic if there is a deterministic finite
automaton taking the base-$k$ representation of $n$ as input, with transitions leading to a state with output $x[n]$. For more details, see [4, 2].

In this note, we prove that if $x$ is $k$-automatic, then so is the characteristic sequence of the least periods of $x$. Our method gives an explicit way to construct the automaton accepting the base-$k$ representation of the least periods of $x$. Using an implementation developed by the first author, we then reprove the Currie-Saari result for Thue-Morse using a short finite computation, and we find similar results for three other classic sequences.

2 The main result

**Theorem 1.** If $x$ is a $k$-automatic sequence, then the characteristic sequence of least periods of $x$ is (effectively) $k$-automatic.

**Proof.** Using the method developed in [1, 3], it suffices to construct a predicate $L(n)$ that is true if $n$ is a least period and false otherwise, using a logical language restricted to addition, subtraction, indexing into $x$, comparisons, logical operations, and the existential and universal quantifiers.

It is easy to express the predicate $P$ that $n$ is a period of the factor $x[i..j]$, as follows:

$$P(n, i, j) = x[i..j - n] = x[i + n..j] = \forall t \text{ with } i \leq t \leq j - n \text{ we have } x[t] = x[t + n].$$

Using this, we can express the predicate $LP$ that $n$ is the least period of $x[i..j]$:

$$LP(n, i, j) = P(n, i, j) \text{ and } \forall n' < n \neg P(n', i, j).$$

Finally, we can express the predicate that $n$ is a least period as follows

$$L(n) = \exists i, j \geq 0 \text{ with } 0 \leq i + n \leq j - 1 \ LP(n, i, j).$$

The construction is effective, and there is an algorithm that, given the automaton generating $x$, will produce an automaton generating the characteristic sequence of least periods of $x$.  

3 Computations

Currie and Saari [5, Thm. 2] proved

**Theorem 2.** For each integer $n \geq 1$, the Thue-Morse word has a factor of period $n$.

We implemented the algorithm in [6] to convert the automaton generating a $k$-automatic sequence $x$ to the automaton accepting the characteristic sequence of least periods of $x$. Using this, we were able to verify the result above using a short computation. (In contrast, Currie and Saari used four pages of rather intricate case reasoning.)

We also carried out the same computation for three other famous infinite words:
• The period-doubling sequence \( d = d_0d_1 \cdots \) defined by \( d_n = 1 \) if \( t[n] \neq t[n + 1] \), and 0 otherwise;

• The paperfolding sequence \( p \), defined as the limit of the finite words \( p_0 = 0 \) and \( p_{i+1} = p_i 0 \overline{p_i}R \), where \( \overline{0} = 1 \), \( \overline{1} = 0 \), and \( wR \) denotes the mirror image or reversal of the word \( w \);

• The Rudin-Shapiro sequence \( r = r_0r_1 \cdots \) defined by counting the number of (possibly overlapping) occurrences of 11 in the binary representation of \( n \), taken modulo 2.

Our results can be summarized as follows:

**Theorem 3.** For each integer \( n \geq 1 \), the period-doubling sequence and the Rudin-Shapiro sequence have a factor of least period \( n \).

For the paperfolding sequence, the least periods are given by the integers whose base-2 representations are accepted by the automaton below. The least omitted least period is 18, and there are infinitely many. In the limit, exactly \( 57/64 \) of all integers are least periods of the paperfolding sequence.

![Figure 1: A finite automaton accepting least periods of the paperfolding sequence](image)

**Proof.** The first results were obtained through our algorithm. A summary of our computations appears below:
For the result about the paperfolding sequence, we take the automaton computed by the algorithm (displayed in Figure 1) and compute the transition matrices \( M_a, a \in \{0, 1\} \), containing a 1 in row \( i \) and column \( j \) if there is a transition on \( a \) from state \( i \) to state \( j \). Then \((M^n)_{i,j}\), where \( M := M_0 + M_1 \), gives the number of words taking the automaton from state \( i \) to state \( j \). A short computation gives that each row of \(\lim_{n \to \infty} 2^{-n}M^n\) equals

\[
\frac{1}{64}[0, 16, 8, 4, 2, 10, 5, 4, 2, 6, 3].
\]

All states except 7 and 11 are accepting, so the density of least periods is given by \((64 - 4 - 3)/64 = 57/64\), as claimed.

\[\square\]

References

[1] J.-P. Allouche, N. Rampersad, and J. Shallit. Periodicity, repetitions, and orbits of an automatic sequence. *Theor. Comput. Sci.* **410** (2009), 2795–2803.

[2] J.-P. Allouche and J. Shallit. *Automatic Sequences: Theory, Applications, Generalizations*, Cambridge, 2003.

[3] E. Charlier, N. Rampersad, and J. Shallit. Enumeration and decidable properties of automatic sequences. In G. Mauri and A. Leporati, eds., *Developments in Language Theory — 15th International Conference, DLT 2011*, Vol. 6795, Springer, 2011, pp. 165–179.

[4] A. Cobham. Uniform tag sequences. *Math. Systems Theory* **6** (1972), 164–192.

[5] J. D. Currie and K. Saari. Least periods of factors of infinite words. *RAIRO – Theor. Inf. Appl.* **43** (2009), 165–178.

[6] D. Goč, D. Henshall, and J. Shallit. Automatic theorem-proving in combinatorics on words. In N. Moreira and R. Reis, eds., *CIAA 2012*, Lect. Notes in Comput. Sci., Vol. 7381, Springer, 2012, pp. 180–191.