The lexicographically least word in the orbit closure of the Rudin-Shapiro word

James Currie*
Department of Mathematics and Statistics
University of Winnipeg
515 Portage Avenue
Winnipeg, Manitoba R3B 2E9 (Canada)
j.currie@uwinnipeg.ca

June 1, 2009

Abstract

We show that the lexicographically least word in the orbit closure of
the Rudin-Shapiro word $w$ is $0w$.

Keywords: Combinatorics on words, Rudin-Shapiro word, morphic
words, automatic words.

Let $f : \{a, b, c, d\}^* \rightarrow \{0,1\}^*$ and $g : \{a, b, c, d\}^* \rightarrow \{a, b, c, d\}^*$ be given
respectively by

\begin{align*}
f(a) &= 0 \\
f(b) &= 0 \\
f(c) &= 1 \\
f(d) &= 1
\end{align*}

*The author is supported by an NSERC Discovery Grant.
and
\[
\begin{align*}
g(a) &= ab \\
g(b) &= ac \\
g(c) &= db \\
g(d) &= dc.
\end{align*}
\]

Let \( u = g^\omega(a) \). The Rudin-Shapiro word \( w \) is given by \( w = f(u) \). Thus
\[
w = 00010010000111 \cdots
\]

The Rudin-Shapiro word has been the subject of much study in combinatorics on words. A standard reference is [1]. An alternative characterization of the Rudin-Shapiro word is as follows: For each non-negative integer \( n \), let \( P(n) \) denote the parity of the number of times 11 appears in the binary representation of \( n \). For example, 59 has binary representation 111011, which contains 3 occurrences of 11, so that \( P(59) = 1 \equiv 3 \pmod 2 \). The Rudin-Shapiro word is the infinite binary word whose \( i \)th bit (starting at \( i = 0 \) on the left) is \( P(i) \).

**Remark 1.** From this second characterization, it follows that if \( p \) is any finite prefix of \( w \), then \( 0p \) is a factor of \( w \); indeed, choose odd \( s > |p| \). Then the binary representation of \( 2^s - 1 \) is a string of 1’s of length \( s \), whence \( P(2^s - 1) = 0 \). On the other hand, \( P(i) = P(2^s + i) \) for \( 0 \leq i \leq |p| - 1 \), so that \( 0p \) appears in \( w \), starting at bit \( 2^s - 1 \).

The orbit closure of a right infinite word \( v \) is the set of those right infinite words whose every finite prefix is a factor of \( v \). Our remarks of the previous paragraph show that \( 0w \) is in the orbit closure of \( w \). Recently it was conjectured [2] that

**Conjecture 1.** Word \( 0w \) is the lexicographically least word in the orbit closure of \( w \).

The purpose of our note is to prove this conjecture.

**Remark 2.** Morphism \( g \) is order-preserving; i.e., \( g(x) \leq g(y) \) if and only if \( x \leq y \). One also notices that \( f \circ g \) is order preserving.
Remark 3. We say that word $x$ appears with index $i$ in word $y$ if we can write $y = pxq$ for some words $p$ and $q$ where $|p| = i$. Note that letter a only ever appears in $u$ with an even index. In particular, $aa$ is not a factor of $u$. At this point we also note that if $f(g(x)) = 00$, then $x = a$.

Lemma 2. Suppose $p0000$ is a prefix of $w$. Then $|p|$ is odd. Thus $0000$ only appears in $w$ with odd index.

Proof: Otherwise, write $p0000 = f(g(q))$ where $q$ is a prefix of $u$. But this forces $q$ to have $aa$ as a suffix, which is impossible. □

Denote by $\pi_n$ the length $n$ prefix of $w$.

Lemma 3. For $i \leq 12$, $0\pi_i$ is the lexicographically least factor of $w$ of length $i + 1$.

Remark 4. One can effectively list all factors of $w$ of length 13 or less, establishing the truth of the lemma. However, a briefer proof follows:

Proof: Certainly each $0\pi_i$ is a factor of $w$ by Remark 1. Clearly our result holds for $i \leq 3$. By Lemma 2, word 00000 cannot be a factor of $w$ so that the result holds for $i \leq 6$.

Suppose that $0\pi_60$ is a factor of $w$. Then one of $0\pi_600$ and $0\pi_601$ is a factor of $w$. Since these words have 0000 as a prefix, by Lemma 2 they could only appear in $w$ with odd index, forcing one of $\pi_600$ and $\pi_601$ to have a length 4 suffix of the form $f(g(v))$, some factor $v$ of $u$. This hypothetical $v$ would then have the form $abaa$ (in the case where $f(g(v)) = \pi_600$), or $abab$ (in the case where $f(g(v)) = \pi_601$). In either case, $f(v) = 0000$, forcing $v$ to have odd index in $u$. However, then the initial $a$ of $v$ has odd index in $u$, which is impossible.

It follows that $0\pi_61$ is the lexicographically least length 8 factor of $w$, and the result holds for $i \leq 11$. Since 00000 is not a factor of $w$, our lemma holds for $i = 12$ also. □

Proof of Conjecture 1: Suppose that $z$ is lexicographically least in the orbit closure of $w$, but $z \neq 0w$. It follows that a prefix $0q$ of $z$ is lexicographically less than $0\pi_{|q|}$. Since $z$ is in the orbit closure of $w$, $0q$ is a factor of $w$. From the fact that $0q$ is lexicographically less that $0\pi_{|q|}$, it follows from Lemma 3 that $\pi_{12}$ is a proper prefix of $q$. Using the fact that 0000 is a factor with odd index in $\pi_{12}$, we deduce that $q$ has even index in $w$. Let $q$ be the shortest factor of $w$ having the property that
1. $q$ is lexicographically less than $\pi_{|q|}$

2. $q$ appears in $w$ with even index.

Let $q'$ be a shortest factor of $w$ with prefix $q$ and even length. (Thus $|q'| \leq |q| + 1$.) Write $q' = f(g(v))$, some factor $v$ of $u$. Let $v'$ be a shortest factor of $w$ with prefix $v$ and even length. (Thus $|v'| \leq |v| + 1$.) Let $u'$ be the prefix of $u$ of length $v'$. Since $q$ is a prefix of $f(g(v'))$, and $q$ is lexicographically less than the prefix $\pi_{|q|}$ of $f(g(u'))$, we see that $f(g(v'))$ is lexicographically less than $f(g(u'))$. Since $f \circ g$ is order-preserving, $v'$ is lexicographically less than $u'$. Write $u' = g(U)$, where $U$ is a prefix of $u$.

Suppose that an occurrence of $v$ has even index in $u$. Then we can write $v' = g(V)$, some factor $V$ of $u$ with $|V| = |v'|/2$. Since $g$ is order-preserving, $V$ is lexicographically less than $U$. It follows that $f(v') = f(g(V))$ is lexicographically less than $f(g(U))$, which is a prefix of $w$. However the index of $f(v')$ in $w$ is just the index of $v'$ in $u$, which is even. Thus

1. $f(v')$ is lexicographically less than $\pi_{|f(v')|}$

2. $f(v')$ appears in $w$ with even index.

However, $|f(v')| \leq |v'| + 1 = |q'|/2 + 1 \leq (|q| + 1)/2 + 1 = |q|/2 + 3/2 < |q|$, since $|q| > |\pi_{12}| = 12$. This contradicts the minimality of $|q|$. We conclude that $v$ cannot have even index in $u$.

We now use the fact that $\pi_{12}$ is a proper prefix of $q' = f(g(v))$. From $\pi_{12} = 000100100001$, we deduce that $g(v) = abacabdbabac$. This implies that $v = abacab$. Since $v$ begins with an $a$, it must have even index in $u$. This is a contradiction. $\square$

References

[1] Jean-Paul Allouche and Jeffrey Shallit, *Automatic Sequences: Theory, Applications, Generalizations*. Cambridge University Press, 2003.

[2] Jean-Paul Allouche, Narad Rampersad & Jeffrey Shallit, Periodicity, repetitions, and orbits of an automatic sequence, *Theoret. Comput. Sci.* To appear.