TRANSCENDENCE OF NUMBERS WITH AN
EXPANSION IN A SUBCLASS OF COMPLEXITY $2N + 1$

TOMI KĂRKI

Abstract. We divide infinite sequences of subword complexity $2n + 1$
into four subclasses with respect to left and right special elements and
examine the structure of the subclasses with the help of Rauzy graphs.
Let $k \geq 2$ be an integer. If the expansion in base $k$ of a number is an
Arnoux-Rauzy word, then it belongs to Subclass I and the number is
known to be transcendental. We prove the transcendence of numbers
with expansions in the subclasses II and III.

Mathematics Subject Classification. 11J81, 68R15.

Introduction

For any natural number $k \geq 2$, the expansion in base $k$ of an irrational algebraic
number is conjectured to be normal in the sense that the expansion contains
every block of digits of length $n$ with a frequency asymptotic to $1/k^n$. Define the
complexity $p_v(n)$ to be the number of different blocks of digits of length $n$ in a fixed
expansion $v$. The conjecture implies that numbers with low complexity expansion
are either transcendental or rational. Hedlund and Morse [7] proved already in
the first half of the 20th century that, if there exists a natural number $n$ such
that $p_v(n) \leq n$, the expansion $v$ is ultimately periodic and the corresponding
number rational. In 1997 Ferenczi and Mauduit [6] proved that numbers with
an expansion $v$ in base $k$ of complexity $p_v(n) = n + l - 1$, where $2 \leq l \leq k$, are
transcendental. Such expansions are called Sturmian words. They also generalized
the transcendence result to so called Arnoux-Rauzy expansions, which form a
subclass of sequences with complexity $2n + 1$. The proof method was based on the
combinatorial translation of a number theoretical result of Ridout [8] concerning
rational approximations. The transcendence in the general case of complexity

Keywords and phrases. Transcendental numbers, subword complexity, Rauzy graph.

1 Department of Mathematics and Turku Centre for Computer Science, University of Turku,
20014 Turku, Finland; topeka@utu.fi

© EDP Sciences 2006
2n + 1 was left open. In this article we divide the sequences of this complexity into four subclasses as in [5]. Arnoux-Rauzy words belong to Subclass I. Our aim is to generalize the transcendence result to numbers with expansions in the subclasses II and III. We remark that these results follow from the more general result recently obtained by Adamczewski et al. [1]. By improving the combinatorial criterion for transcendence from [6], they proved that numbers with expansions of complexity $O(n)$ are either rational or transcendental. Our proof is based on the original combinatorial criterion of Ferenczi and Mauduit [6] and thus requires more detailed analysis of the combinatorial structure of the expansions in these specific subclasses. For other results related to this topic, see [2–4,9].

1. Basic definitions

An alphabet $\Sigma$ is a nonempty finite set of symbols and a word over $\Sigma$ is a (finite or infinite) sequence of symbols from $\Sigma$. A catenation of words is an operation defined as $a_1 \ldots a_n \cdot b_1 \ldots b_m = a_1 \ldots a_n b_1 \ldots b_m$ for $a_i, b_i \in \Sigma$. Denote by $\Sigma^*$, $\Sigma^+$ and $\Sigma^\omega$ the sets of all finite, finite nonempty and infinite words over $\Sigma$, respectively. A word $w$ is a factor of a word $u$ (resp. a left factor or a prefix, a right factor or a suffix), if there exist words $x$ and $y$ such that $u = xwy$ (resp. $u = wy, u = xw$). The length of $w$, denoted by $|w|$, is the total number of letters in $w$. The number of letters $a$ in $w$ is denoted by $|w|_a$. Let $L_n(w)$ be the set of all factors of $w$ of length $n$. The complexity function of $w$ is $p_w(n) = |L_n(w)|$. A sequence $v = (v_n) = v_1 v_2 \ldots v_n \ldots \in \Sigma^\omega$ is called recurrent when every factor of $v$ occurs infinitely many times in $v$. It is minimal when the factors occur also with bounded gaps, that is, the length of a word in $v$ between any two consecutive occurrences of factor $w$ cannot be arbitrarily large.

2. Combinatorial criterion for transcendence

We state here a combinatorial criterion for transcendence due to Ferenczi and Mauduit which is the basis of our transcendence proof. The proof of this theorem can be found, for example, in [6].

**Theorem 2.1** (combinatorial criterion for transcendence). If $\theta$ is an irrational number and, for every $n \in \mathbb{N}$, the expansion of $\theta$ in base $k$ begins by $0.U_nV_nV'_n$, where $U_n$ is a (possibly empty) word and $V_n$ is a nonempty word on an alphabet $\{0, \ldots, k-1\}$, $V'_n$ is a prefix of $V_n$ and $|V_n| \to \infty$, as $n \to \infty$, lim sup $|U_n|/|V_n| < \infty$ and $\lim \inf |V'_n|/|V_n| > 0$, then $\theta$ is a transcendental number.

In the following we use a modified form of this theorem. Suppose the expansion of $\theta$ is $0.v$, where $v \in \{0, \ldots, k-1\}^\omega$. The theorem implies that $\theta$ is transcendental if we find an infinite number of different word triplets $U, V$ and $V'$ satisfying the conditions

$$v = UV^2V' \ldots, \quad \frac{|U|}{|V|} \leq \frac{1}{h} \quad \text{and} \quad \frac{|V'|}{|V|} \geq \frac{1}{h}$$

(∗)
for some fixed positive integer $h$ depending on the sequence $v$. We say that we are in the situation of Theorem 2.1 for the words $V, V'$ and $U$, if conditions $(\ast)$ are satisfied.

3. Subclasses of complexity $2n + 1$

Let $v$ be an infinite recurrent word. A right special factor $X \in L_n(v)$ is a word with two or more extensions to the right. More precisely, there exist different letters $a_1$ and $a_2$ such that $Xa_1, Xa_2 \in L_{n+1}(v)$. Left special words are defined respectively. Denote by $\partial^+(X)$ (resp. $\partial^-(X)$) the number of different right (resp. left) extensions of $X$. Suppose now $p_v(n) = 2n + 1$ for every $n \in \mathbb{N}$. Then clearly

$$\sum_{X \in L_n(v)} (\partial^+(X) - 1) = \sum_{X \in L_n(v)} \partial^+(X) - \sum_{X \in L_n(v)} 1 = p_v(n + 1) - p_v(n) = 2.$$ 

This means that there are just two possibilities. There exist only one right special element $D \in L_n(v)$ with $\partial^+(D) = 3$ or two distinct elements $D_1$ and $D_2$ with $\partial^+(D_1) = \partial^+(D_2) = 2$. Since same conclusions can be made with $\partial^-$, we have four types of languages $L_n(v)$:

1. $\partial^+(D) = 3$ and $\partial^-(G) = 3$.
2. $\partial^+(D) = 3$ and $\partial^-(G_1) = \partial^-(G_2) = 2$, where $G_1 \neq G_2$.
3. $\partial^+(D_1) = \partial^+(D_2) = 2$, where $D_1 \neq D_2$, and $\partial^-(G) = 3$.
4. $\partial^+(D_1) = \partial^+(D_2) = 2$, where $D_1 \neq D_2$, and $\partial^-(G_1) = \partial^-(G_2) = 2$, where $G_1 \neq G_2$.

Note that it is possible that a right special element is also left special. Note also that ultimately $L_n(v)$ is of constant type. This can be easily verified. If there exists a word in $L_m(u)$ with three different extensions to the right, then deleting letters from the left we get elements of $L_n(u)$ with the same three extensions to the right for every $n < m$. Similar reasoning can be made for words with three extensions to the left. Thus, the possibilities are

1. for all $n \geq 0$, $L_n(u)$ is of type I;
2. there exists an integer $n_0$ such that $L_n(u)$ is of type I for $n < n_0$, and of type II for $n \geq n_0$;
3. there exists an integer $n_0$ such that $L_n(u)$ is of type I for $n < n_0$, and of type III for $n \geq n_0$;
4. there exist integers $n_0$ and $n_1$ such that $L_n(u)$ is of type I for $n < n_0$, of type II for $n_0 \leq n < n_1$, and of type IV for $n \geq n_1$;
5. there exist integers $n_0$ and $n_1$ such that $L_n(u)$ is of type I for $n < n_0$, of type III for $n_0 \leq n < n_1$, and of type IV for $n \geq n_1$.

Hence, the words $v$ of complexity $2n + 1$ can be divided into four subclasses with respect to the ultimate type of $L_n(v)$. Minimal words in Subclass I are the Arnoux-Rauzy words for which the transcendence result is known [6]. Our aim is to prove the transcendence of minimal words in Subclass II and in the symmetric case III.
4. Rauzy graphs

To examine the combinatorial structure of minimal words in subclass II we need the notion of Rauzy graphs. The following construction is presented for Arnoux-Rauzy words in [5] with slight modifications on notations. Let $v$ be an infinite recurrent word over $\Sigma$. The vertices of the Rauzy graph $\Gamma_n$ of $v$ are the factors of length $n$ in $v$ and there is an arrow going from vertex $E$ to vertex $F$ with label $a$ whenever $E = bhF = Ha$ with $a, b \in \Sigma$ and $bHa \in L_{n+1}(v)$. Now $\partial^+(X)$ (resp. $\partial^-(X)$) is the number of outgoing (resp. incoming) arrows of vertex $X$.

We call $n$-segment any finite sequence $(E_0, \ldots, E_k)$ of vertices of $\Gamma_n$ such that there exists an arrow from $E_i$ to $E_{i+1}$, $E_0$ and $E_k$ are right special and from each $E_{i}, 1 \leq i \leq k - 1$ leaves only one arrow. The name of the $n$-segment $(E_0, \ldots, E_k)$ is the catenation of the labels of the arrows $E_0 \rightarrow E_1 \rightarrow \ldots \rightarrow E_{k-1} \rightarrow E_k$. Because of the recurrence of $v$, the graph must be strongly connected, i.e. for every pair of vertices $X$ and $Y$, there exists a path from $X$ to $Y$ and from $Y$ to $X$. It follows that the graph $\Gamma_n$ of a minimal word in Subclass II has up to isomorphism one of the two forms of Figure 1.

The presentation of Rauzy graphs is simplified by representing only the right special element $D_n$ and left special elements $G_{1,n}$ and $G_{2,n}$. All the segments indicated in Figure 1 begin and end by $D_n$. We fix the following notation: $A_n$ is the name of the unique $n$-segment going through only one left special element $G_{1,n}$, $B_n$ and $C_n$ are the names of the $n$-segments going through both left special elements $G_{1,n}$ and $G_{2,n}$. Sometimes we simplify the notation and mark these $n$-segments without the subscript $n$.

In the first case, $G_{1,n} \neq D_n$. For any word $X \neq D_n$ in $L_n(v)$ there exists a unique word $X\alpha$ in $L_{n+1}(v)$ and $G_{1,n+1} = G_{1,n}\gamma_1$, $G_{2,n+1} = G_{2,n}\gamma_2$ and $D_{n+1} = \delta D_n$ are uniquely determined by the graph $\Gamma_n$. Thus, the graph $\Gamma_{n+1}$ is known entirely and we can easily see that $A_{n+1} = A_n$, $B_{n+1} = B_n$ and $C_{n+1} = C_n$. Note that the length of the path from $G_{1,n+1}$ to $D_{n+1}$ is one smaller than from $G_{1,n}$ to $D_n$. This means that finally $G_{1,n+l} = D_{n+l}$ for some $l \geq 1$ and we are in the second case.

The interesting case is the break down case, where $G_{1,n} = D_n$. Let the three $n$-segments be

\[
\begin{align*}
A_n & : (D_n, D'_n a_1, \ldots, b_1 D''_n, D_n), \\
B_n & : (D_n, D'_n a_2, \ldots, b_2 D''_n, D_n), \\
C_n & : (D_n, D'_n a_3, \ldots, b_2 D''_n, D_n).
\end{align*}
\]

We see that $G_{1,n+1} = G_{1,n}\gamma_1$, $G_{2,n+1} = G_{2,n}\gamma_2$ and $D_{n+1} = \delta D_n$, where $\gamma_1 \in \{a_1, a_2, a_3\}$, $\delta \in \{b_1, b_2\}$ and $\gamma_2$ is fixed by the graph $\Gamma_n$. Now $\Gamma_n$ does not determine $\Gamma_{n+1}$ entirely. Suppose, for example, that $\delta = b_1$. Since, by the recurrence of $v$, the graph is strongly connected, we must have $\gamma_1 = a_1$. This is illustrated by the dashed arrow in the left graph of Figure 2. From the figure we can also clearly see that $A_{n+1} = A_n$, $B_{n+1} = B_n A_n$ and $C_{n+1} = C_n A_n$. We call this break down type I.
Other possible break down types and the names of the new segments can be calculated similarly. They are presented in Table 1. Note that in the types 2 and 3 the roles of the left special elements change in such a way that \( G_{2,n+1} = G_{1,n+1} \gamma_1 \) and \( G_{1,n+1} = G_{2,n+1} \gamma_2 \) for some \( \gamma_1 \) and \( \gamma_2 \) in the alphabet \( \Sigma \).

For the \( n \)th break down graph of a word \( v \) we connect the break down type \( i_n \in \{1, 2, 3\} \) and the sequence \( (i_n) \) is called the **directive sequence of** \( v \). Thus, we have the following lemma.

**Lemma 4.1.** Let \( v \) be a recurrent word in Subclass II of complexity \( 2n + 1 \). Then there exist three words, \( A_1, B_1 \) and \( C_1 \) and a directive sequence of integers \( 1 \leq i_n \leq 3, n \geq 1 \), such that if the words \( A_n, B_n, C_n \), \( n \in \mathbb{N} \), are given by the recursion formulae of Table 1, then for any \( N \geq 1 \) and \( n \geq 1 \), the word \( v_0 v_1 \ldots v_{N-1} \) is of the form \( X_0 X_1 \ldots X_m \), where \( X_1, X_2, \ldots, X_m \) are equal to

<table>
<thead>
<tr>
<th>( i_n )</th>
<th>( A_{n+1} )</th>
<th>( B_{n+1} )</th>
<th>( C_{n+1} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( A_n )</td>
<td>( B_n A_n )</td>
<td>( C_n A_n )</td>
</tr>
<tr>
<td>2</td>
<td>( C_n )</td>
<td>( A_n B_n )</td>
<td>( B_n )</td>
</tr>
<tr>
<td>3</td>
<td>( B_n )</td>
<td>( A_n C_n )</td>
<td>( C_n )</td>
</tr>
</tbody>
</table>
5. TRANSCENDENCE OF NUMBERS WITH THE EXPANSIONS IN SUBCLASSES II AND III

Let \( \theta \) be a number with expansion \( 0.v \), where the sequence \( v \) is a minimal word belonging to Subclass II of complexity \( 2n + 1 \). Our aim is now to show that for such a sequence \( v \) we are in the situation of Theorem 2.1 for an infinite number of different word triplets \( V, V', U \) satisfying conditions (\( * \)) with some fixed positive integer \( h \). We examine the directive sequence \( (i_n) \) given in Lemma 4.1.

We have three different cases:

- **Case I:** The sequence \( (i_n) \) contains an infinite number of occurrences of the factor 11.
- **Case II:** Ultimately, the sequence \( (i_n) \) does not contain any occurrences of 11, but it contains infinitely many occurrences of the factor 1.
- **Case III:** Ultimately, the sequence \( (i_n) \) consists only of integers 2 and 3.

First, we take an example. Suppose that \( l \geq 3, i_{m+1} = \ldots i_{m+l} = 1 \) and \( i_{m+l+1} = 2 \), then

\[
A_{m+l+2} = B_mC_m^l, \quad B_{m+l+2} = C_mA_mB_mC_m^l, \quad C_{m+l+2} = A_mB_mC_m^l.
\]

Clearly, every segment contains now \( C_mC_mC_m^l \) and we may choose \( V = V' = C_m \). Consider then all possible words \( U \) such that \( UC_m^l \) is a prefix of \( v \) but \( C_m^l \) is not an internal factor of \( UC_m^l \). We use the notation of Lemma 4.1. If the cube can be found in \( X_0 \), then \( U \) is a suffix of \( B_mC_mA_mB_m \) or \( A_mB_m \). Otherwise, \( X_0 \) is a strict suffix of \( C_m^l \) and \( U \) is a strict suffix of \( C_m^l B_m, C_m \) or \( C_m^l A_mB_m \).

Hence, if there exists a fixed integer \( h' \) depending on the sequence \( v \) such that \( h'|C_m^l \geq |A_mB_m| \) and \( h'|C_m^l \geq |B_mC_m| \) for infinitely many \( m \) and situations similar to our example, then the conditions (\( * \)) are satisfied infinitely often with \( h = 3 + 2h' \). In the following we are going to prove the existence of such an integer \( h' \).

In order to find repetitions and fixed positive integers \( h' \) and \( h \) in every above mentioned case our strategy is the following. We find prefixes \( UVV' \) of \( v \) with the help of useful factors \( u \) of \( (i_n) \) as in the previous example. For the length conditions we divide factors of the sequence \( (i_n) \) into three blocks. We introduce suitable words \( s \), which together with finite prefixes of \( (i_n) \) act as beginning blocks and useful words act as end blocks. Each middle block \( t \) consists of at most one inequality preserving and one inequality changing word in this order. This construction is illustrated in Figure 3. More precise descriptions of these concepts are given later.

We say that we **execute** a word \( w = i_{j_1} \ldots i_{j_r} \), when we apply recursion rules \( i_{j_1}, \ldots, i_{j_r} \) in this order to the names of the segments \( A_j, B_j \) and \( C_j \). By a suitable
word we mean a sequence $13^{2l+1}1, 23^{2l+1}2, or 23^{2l}2$, where $l \geq 0$. By executing suitable words we have the following lemma.

**Lemma 5.1.** Let $i_{j_0} \ldots i_{j_1-1}$ be a suitable word for $0 \leq j_0 < j_1$. Then we have the inequality

$$2|AC_j| - |B_j| \geq |C_j| \geq |A_j|.$$  

**Proof.** Denote the names of the segments $A_{j_0}, B_{j_0}, C_{j_0}$ by small letters $a, b, c$ and $A_{j_1}, B_{j_1}, C_{j_1}$ by capital letters $A, B$ and $C$. The names of the segments and some length calculations are presented in Table 2. Note that, by the recursion formulae of Table 1, the inequality $|b| \geq |c|$ is always valid. Thus, by the calculations, after the execution of suitable words we have $2|AC| - |B| \geq 0$ for every $l \geq 0$. Examining the names of the segments in Table 2, we can also see that $|C| \geq |A|$. Hence, the inequality $2|AC| \geq |B| \geq |C| \geq |A|$ holds. □

The suitable words $s$ are beginning blocks in the factor $stu$ of $(i_n)$. Now we look into the structure of the middle block $t$. Suppose that $s$ is the nearest suitable word before the end block $u$. We assume that this suitable word does not overlap with the end block. Since the word $s$ ends with either 1 or 2, the sequence $t$ must begin by $3^{2l}$ or $3^{2l+1}$ or $t = 3^l$ for some $l \geq 0$. Otherwise we would have a suitable word closer to $u$ than our $s$, which is impossible by definition. Since the

| $i_{j_0} \ldots i_{j_1-1}$ | Names of the segments | $2|AC| - |B|$ |
|------------------------|----------------------|-----------------|
| $13^{2l+1}1$           | $A = ba(ca)^l$       | $(2(2l+3) - (2l+3))|a| + (4-1)|b|$ |
|                        | $B = aca(ca)^lba(ca)^l$ | $+(2(2l+1) - (2l+1))|c|$ |
|                        | $C = cab(ca)^l$       | $=(2l+3)|a| + 3|b| + (2l+1)|c|$ |
| $23^{2l+1}2$           | $A = abbl$           | $(4-1)|a| + (2(2l+3) - (2l+2))|b|-|c|$ |
|                        | $B = cblabb$         | $=3|a| + (2l+4)|b|-|c|$ |
|                        | $C = babb$           | $=(2l+2) - (2l+2)|a| + (2-1)|b|$ |
| $13^{2l}2$             | $A = ca$             | $(2(l+2) - (2l+2))|a| + (2-1)|b|$ |
|                        | $B = aca^lb(a)^l$    | $+(2(l+1) - (2l))|c|$ |
|                        | $C = ba(ca)^l$       | $=2|a| + |b| + 2|c|$ |
| $23^{2l}2$             | $A = b$              | $(2-1)|a| + (2(l+2) - (2l+1))|b|-|c|$ |
|                        | $B = cb^labb$        | $=|a| + 3|b|-|c|$ |
|                        | $C = abbl$           | $=|a| + 3|b|-|c|$ |

**Figure 3.** Illustration of blocks in the sequence $(i_n)$.

**Table 2.** Length calculations after executing suitable words.
Table 3. Length calculations after executing inequality preserving words.

| $i_{j_1}, \ldots, i_{j_2-1}$ | Names of the segments | $r|AC| - |B|$ |
|---|---|---|
| $3^{2l}$ | $A = ac^l$ | $(2r-1)|a| - |b| + (r(2l+1) - 2l)|c| - (2r-1)|a| - |b| + ((r-1)(2l+r))|c|$ |
| | $B = bc^l ac^l$ | |
| | $C = cac^l$ | |
| $3^{2l+1}$ | $A = c$ | $(r-1)|a| - |b| + (r(l+2) - (2l+1))|c|$ |
| | $B = bc^l ace^l$ | |
| | $C = ace^l$ | |

first two cases end with 1 or 2, we can apply the previous reasoning to the end of the middle block. This shows that the word $t$ between the suitable word $s$ and the useful word $u$ belongs to $\{(33)^*1 \cup (33)^*3\}^*$. We call a word in the beginning part $\{(33)^*1 \cup (33)^*3\}^*$ inequality preserving. The following lemma justifies this term. The end part of $t$ belonging to $3^*$ is called inequality changing.

Lemma 5.2. Let $i_{j_1} \ldots i_{j_2-1}$ be an inequality preserving word for $0 \leq j_1 < j_2$. Suppose $r|A_{j_1}C_{j_1}| \geq |B_{j_1}| \geq |C_{j_1}| \geq |A_{j_1}|$ for a positive integer $r \geq 2$. Then we have

$$r|A_{j_2}C_{j_2}| \geq |B_{j_2}| \geq |C_{j_2}| \geq |A_{j_2}|.$$  \hspace{1cm} (2)

Proof. Denote the names of the segments $A_{j_1}, B_{j_1}, C_{j_1}$ by small letters $a, b, c$ and $A_{j_2}, B_{j_2}, C_{j_2}$ by capital letters $A, B$ and $C$. Remember that always $|B| \geq |C|$. Regardless of the lengths of $a, b$ and $c$, Table 3 shows that after executing a word $3^{2l}$ or $3^{2l+1}$ for any $l \geq 0$, we have $|C| \geq |A|$. Since $2r-1 \geq r+1$, when $r \geq 2$, we can use the assumption $r|ac| \geq |b|$ to conclude $r|AC| - |B| \geq 0$ in the case $3^{2l}$. For the case $3^{2l+1}$ we also need to note that $(2r-1)|c| + (r-1)|a| \geq r|ac|$, because $|c| \geq |a|$. Thus, after any number of executions of words $3^{2l}$ and $3^{2l+1}$ our inequality (2) holds.

If the suitable words do not occur infinitely many times as factors of $i_{n}$, then the whole sequence is a catenation of inequality preserving words, at least after a finite (possibly empty) prefix. This prefix combined with a word of the form $3^{2l}$ or $3^{2l+1}$ is considered as the beginning block in this case. Whatever the situation is, Lemma 5.3 shows that a beginning block followed by a middle block is convenient for our purposes.

Lemma 5.3. Let $0 \leq j_0 < j_1 \leq j_2 \leq j_3$. Suppose $i_{j_0} \ldots i_{j_1-1}$ is a beginning block and middle block $t$ consists of inequality preserving word $i_{j_1} \ldots i_{j_2-1}$ and inequality changing word $i_{j_2} \ldots i_{j_3-1}$. (Word $i_{j_2}i_{j_2-1}$ means the empty word.) Then after executing the beginning block and the middle block we are either in the situation of Theorem 2.1, or $h'|B_{j_3}| \geq h'|C_{j_3}| \geq |X_{j_3}|$, for every $X \in \{A, B, C\}$ and for a fixed integer $h' \geq 2$. 

Proof. After executing any beginning block we have

\[ r|A_j, C_j,| \geq |B_j,| \geq |C_j,| \geq |A_j,|, \]

for some fixed integer \( r \geq 2 \). Namely, if the beginning block is a suitable word, then by Lemma 5.1 we may choose \( r = 2 \). Otherwise, the beginning block is the non-inequality-preserving finite prefix of \((i_n)\) combined with a word of the form \( 3^2l \) or \( 3^{2l+1}2 \). After executing \( 3^2l \) or \( 3^{2l+1}2 \) we have \( |C_j,| \geq |A_j,| \), as seen in Lemma 5.2, and by the finiteness of the beginning part, the desired number \( r \) surely exists.

As before, \( |B_j,| \geq |C_j,| \). Therefore (3) holds. After the execution of an inequality preserving word, inequality (2) is valid by Lemma 5.2 or trivially, if \( j_1 = j_2 \). The question is, what happens in the inequality changing part \( i_{j_2} \ldots i_{j_3-1} = 3^l \in 3^* \).

We may suppose that \( j_3 > j_2 \). In order to simplify notations denote the names of the segments \( A_j, B_j, C_j \) by small letters \( a, b, c \) and \( A_{j_3}, B_{j_3}, C_{j_3} \) by capital letters \( A, B, C \). Then we have \( A = ac^l, B = bc^l \) and \( C = c, \) if \( l = 2l' \), or \( A = bc^l, B = ac^l \) and \( C = c, \) if \( l = 2l' + 1 \). If \( l' \geq 3 \), we may choose \( V = V' = C = c \). Now segments \( A \) and \( B \) contain the word \( V^2V' = c^3 \) and using the notations of Lemma 4.1, one or the other must appear after \( X_0, X_0C \) or \( X_0CC \ldots \). We may suppose that \( X_0 \) is a strict suffix of \( CCC \).

Otherwise, \( |U| \leq \max\{|b|, |ac|\} \). Using equation (2) we also conclude that \( |c| \geq |a| \) and \( 2r|c| \geq |b| \). Hence, \( |U| \leq |X_0CC| + \max\{|b|, |ac|\} \leq (5 + 2r)|c| = (5 + 2r)|V| \)

and after the execution of the middle block we are in the situation of Theorem 2.1 with \( h = 5 + 2r \). Thus, we may now suppose that \( l \leq 5 \). Then, after executing the inequality changing word, \((2r + 2)|C| \geq |A| \) and \((2r + 2)|C| \geq |B| \). We choose \( l' = 2r + 2 \) and the lemma follows, since \( |B| \geq |C| \) as noted before.

Now the only thing to do is to introduce the useful words for the three cases mentioned in the beginning of this section, find the words \( U, V, V' \) and prove that the conditions (*) of Theorem 2.1 are satisfied by applying Lemma 5.3. From now on useful word \( u = i_{j_3} \ldots i_{j_4-1} \) and small letters \( a, b, c \) and capital letters \( A, B, C \) denote the names of the segments before and after the execution of \( u \), respectively.

Lemma 5.4. If there exist infinitely many occurrences of the factor 11 in \((i_n)\), we are in the situation of Theorem 2.1 for infinitely many different word triplets \( U, V, V' \).

Proof. First we note that there must be infinitely many \( i_n \in \{2, 3\} \). Otherwise, for some \( m \geq 1 \), we have \( A_{m+i} = A_m, B_{m+i} = B_mA_m^l \) and \( C_{m+i} = C_mA_m^l \) for every \( l \geq 0 \). This implies an ultimately periodic sequence, which is not possible by our assumptions on the complexity of \( v \). Thus, we have infinitely many occurrences of \( 21^l, 21^l3, 31^l2 \) or \( 31^l3 \), where \( l \geq 2 \). These are the useful words for Case I. The execution of \( 21^l2 \) and \( 21^l3 \) are represented in Table 4. By the recursion formulae of Table 1, the other cases are obtained from these by replacing \( b \) by \( c \) and vice versa.

Suppose first that \( l \geq 3 \). The case \( 21^l2 \) is analysed in the example in the beginning of this section with \( j_3 = m \) and \( j_4 - 1 = m + l + 1 \). The analysis for the
Table 4. Names of the segments after executing $21^22$ or $21^33$.

<table>
<thead>
<tr>
<th>$i_{j_2} \ldots i_{j_{k-1}}$</th>
<th>Names of the segments</th>
</tr>
</thead>
<tbody>
<tr>
<td>$21^22$</td>
<td>$A = bc^j$</td>
</tr>
<tr>
<td></td>
<td>$B = cabc^j$</td>
</tr>
<tr>
<td></td>
<td>$C = abc^j$</td>
</tr>
<tr>
<td>$21^33$</td>
<td>$A = abc^j$</td>
</tr>
<tr>
<td></td>
<td>$B = cbc^j$</td>
</tr>
<tr>
<td></td>
<td>$C = bc^j$</td>
</tr>
</tbody>
</table>

case $21^33$ is similar. By Lemma 5.3 we are either in the situation of Theorem 2.1 after the execution of the middle block or

$$h' |b| \geq h' |c| \geq |x|,$$

for every $x \in \{a, b, c\}$ and a fixed integer $h'$. Thus, by our example, we find a cube $V^3 = C^3$ and the conditions ($\ast$) are satisfied with $h = 3 + 2h'$.

Secondly, suppose $l = 2$ and inequality (4) holds. Now none of the segments seems to contain a cube, but we take advantage of the fact that all the segments share a common suffix $bc^2$. If this is followed by segment $B$, we have $V^2V' = c^3$ in $bc^2B$. Also squares over the alphabet $\{A, C\}$ in $X_2X_3X_4 \ldots$ allow us to have $bc^2(ybc^2)^2 = (bc^2y)^2bc^2 = V^2V'$ as a factor of the sequence $v$ for some word $y \in \{a, b, c\}^*$. Because we cannot avoid squares in binary alphabet $\{A, C\}$ and $bc^2B$ contains $c^3$, we cannot avoid repetitions $V^2V'$ and the length of $U$ must be bounded. This is illustrated in Figure 4. Since in any case $V$ and $V'$ contain $c$ and inequality (4) holds, we can approximate $|V| \leq (|V|_a + |V|_b + |V|_c) \cdot h' |V'|$, where $|V|_x$ is the number of occurrences of letter $x$ in the representation of $V$. Hence, $|V| \leq 7h' |V'|$ as, for example, in the case $V^2V' = bc^2CACA = bc^2(abc^2bc^2)^2 = (bc^2abc^2)^2bc^2$. The length of $U$ can also be approximated, because $|X_i| \leq |cabcc| \leq 5h' |V|$ and $V^2V'$ must occur at least before $X_3$. Thus $|U| \leq 5 \cdot 5h' |V|$. Hence, the conditions ($\ast$) are obtained with $h = 25h'$. The analysis for the other cases are similar, since the approximations can be done equivalently also for $V$ and $V'$ containing $b$. Thus, the existence of infinite number of occurrences of the factor 11 in the directive sequence, allows us to be infinitely often in the situation of Theorem 2.1.

Lemma 5.5. If the sequence $(i_n)$ does not ultimately contain any occurrences of 11, but does contain infinitely many occurrences of the factor 1, we are in the situation of Theorem 2.1 for infinitely many different word triplets $U, V$ and $V'$.

Proof. By the assumption, we have infinitely many words $2121, 2122, 2123, 2131, 2132, 2133, 3121, 3122, 3123, 3131, 3132$ or $3133$. These are the useful words for Case II. We analyse those beginning with integer 2. The other cases are similar; the segment $b$ must be replace by $c$ and vice versa. From Table 5 we conclude that...
in every case all the segments end with $c$ and one of the segments, say $Y = xc$, is a suffix of the others. More precisely, the other segments end with $cxc$. It means that $Y$ takes the role of segment $B$ in the previous lemma. For example, in the case $2122$, $Y = abc = A$ and there is $V^2V' = (cab)^2c$ in $v$ if we find a sequence $X_iA(AA)(i \geq 1)$, $BA$ or $CA$. Also any segment concatenated with a square over the alphabet $\{A, B, C\}\setminus Y$ contains word $c(ye)^2c$ for some word $y$ in $\{a, b, c\}^*$. Using Lemma 5.3 we conclude as in Lemma 5.4 that either we are in the situation of Theorem 2.1 after the middle block or inequality (4) holds. In the latter case we use the repetitions mentioned above. Since $V$ and $V'$ contain $c$, we calculate as in Lemma 5.4 that $|V| < 11h' |V'|$ for the fixed $h' \geq 2$. Also $|X_i| < 6h' |V|$ and $|U| < 6 \cdot 6h' |V'|$, because $V^2V'$ must now occur before $X_0$. Since there are infinitely many occurrences of the factor 1, we are infinitely often in the situation of Theorem 2.1 with $h = 36h'$. □
Table 6. Names of the segments after executing 3222, 3223, 2332, 2333, 3232 or 2222.

<table>
<thead>
<tr>
<th>Case</th>
<th>$i_j \ldots i_{j+1}$</th>
<th>Names of the segments</th>
<th>$i_j \ldots i_{j+1}$</th>
<th>Names of the segments</th>
</tr>
</thead>
<tbody>
<tr>
<td>1°</td>
<td>3222</td>
<td>$A = bac$</td>
<td>3223</td>
<td>$A = cbac$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$B = acbac$</td>
<td></td>
<td>$B = acbac$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$C = cbac$</td>
<td></td>
<td>$C = bac$</td>
</tr>
<tr>
<td>2°</td>
<td>2332</td>
<td>$A = b$</td>
<td>2333</td>
<td>$A = abb$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$B = cbabb$</td>
<td></td>
<td>$B = cbb$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$C = abb$</td>
<td></td>
<td>$C = b$</td>
</tr>
<tr>
<td>3°</td>
<td>3232</td>
<td>$A = ac$</td>
<td>3232</td>
<td>$A = cab$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$B = baccac$</td>
<td></td>
<td>$B = abceab$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$C = cac$</td>
<td></td>
<td>$C = bcab$</td>
</tr>
</tbody>
</table>

Lemma 5.6. If the sequence $(i_n)$ consists ultimately only of integers 2 and 3, we are in the situation of Theorem 2.1 for infinitely many different word triplets $U, V$ and $V'$.

Proof. We divide this examination into subcases:

1°: The sequence $(i_n)$ contains infinitely many occurrences of factors 22 and 3.

2°: Ultimately, the sequence $(i_n)$ does not contain any occurrences of 22, but the sequence contains infinitely many occurrences of factors 33 and 2.

3°: Ultimately, the sequences $(i_n)$ is $(23)\omega$, $(32)\omega$, 2\omega or 3\omega.

First we note that the sequence 3\omega is impossible. Otherwise, for every $l \geq 1$, after executing $3^{2l}$ we have $A = ac^l$, $B = bc^l$ and $C = c$. This implies periodicity, which contradicts with the complexity of $v$. Table 6 introduces the useful sequences and corresponding names of the segments. One of the useful words must occur infinitely often. As in Lemma 5.5, in every case the segments $A, B$ and $C$ end with the same letter which is either $b$ or $c$. One segment $Y = xc$ or $Y = xb$ is a suffix of the other segments, which end with $cxc$ or $bxb$, respectively. Now Lemma 5.3 is valid and we are either in the situation of Theorem 2.1 before the end block or the repetition $V^2V'$ is found analysing the segments in $v = X_0X_1X_2\ldots$ same way as in Lemma 5.5. In the latter case $|V| < 10h'|V'|$, $|U| < 6 \cdot 6h'|V|$ and we choose $h = 36h'$.

Finally, we state our result.

Theorem 5.7. Let $\theta$ be a number with expansion $0.v$, where the sequence $v$ is a minimal word belonging to Subclass II of complexity $2n + 1$. Then $\theta$ is transcendental.

Proof. This is a straightforward consequence of Lemmata 5.4–5.6 and Theorem 2.1, the combinatorial criterion for transcendence.
Corollary 5.8. Let $\theta$ be a number with expansion $0.v$, where the sequence $v$ is a minimal word belonging to Subclass III of complexity $2n + 1$. Then $\theta$ is transcendental.

Proof. The Rauzy graphs of Subclass III can be obtained from the graphs in Figure 1 by converting all the arrows and replacing $G$ by $D$ and vice versa. Using similar considerations as in Section 4, we easily find out that the recursion formulae for the names of the segments are inverses of those of Table 1. All the lemmata are valid, since length calculations are exactly the same and repetitions $V^2V'$ can be found similarly. Note that now the segments have a common prefix instead of a common suffix in Lemma 5.5-5.6. □

6. Future work

Our aim is to search concrete examples of transcendental numbers with expansions in these subclasses. We are also going to examine, how the methods introduced here are suitable for the words in Subclass IV. However, this case seems to be quite complicated and, naturally, this result follows from the work of Adamczewski et al. [1].

Acknowledgements. The problem of this paper was introduced to me by J. Karhumäki. I would also wish to thank J. Cassaigne, V. Halava, T. Harju and A. Renvall for the useful discussions during the course of this work. The reference for [1] was brought to my attention by the referee of the conference version of this paper. Finally, I thank the referees for helpful comments and suggestions, especially for mentioning the references [2–4, 9].

REFERENCES