TUCS

Tomi Kärki

Transcendence of numbers with an expansion in a subclass of complexity $2n+1$

Turku Centre for Computer Science

TUCS Technical Report
No 654, December 2004

Transcendence of numbers with an expansion in a subclass of complexity $2n+1$

Tomi Kärki
Department of Mathematics
University of Turku
FIN-20014 Turku, Finland
topeka@utu.fi

TUCS Technical Report
No 654, December 2004

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 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.

Keywords: transcendental numbers, subword complexity, Rauzy graph

TUCS Laboratory
Discrete Mathematics for Information Technology

1 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 [4] 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 [3] 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 [5] concerning rational approximations. The transcendence in the general case of complexity $2n + 1$ was left open. In this article we divide the sequences of this complexity into four subclasses as in [2]. 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 they proved that numbers with expansions of complexity $\mathcal{O}(n)$ are transcendental. Our proof is based on the original combinatorial criterion of Ferenczi and Mauduit [3] and thus requires more detailed analysis of the combinatorial structure of the expansions in these specific subclasses.

2 Basic definitions

Alphabet Σ is a nonempty finite set of symbols and a *word* over Σ is a (finite or infinite) sequence of symbols from Σ. *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 Σ^*, Σ^+ and Σ^ω the sets of all finite, finite nonempty and infinite words over Σ, respectively. Word w is a *factor* of 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 $p_w(n) = \#L_n(w)$. 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.

3 Combinatorial criterion for transcendence

We state here a combinatorial criterion for transcendence which is the basis of our transcendence proof. The proof of this theorem can be found, for example, in [3].

Theorem 1. (Combinatorial criterion for transcendence) *If θ is an irrational number and, for every $n \in \mathbb{N}$, the expansion of θ in base k begins by $0.U_nV_nV_nV'_n$, where U_n is a possibly empty 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 $n \to \infty$, $\limsup |U_n|/|V_n| < \infty$ and $\liminf |V'_n|/|V_n| > 0$, then θ is a transcendental number.*

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

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

for some fixed positive integer h depending on the sequence v. We say that we are *in the situation of Theorem* 1 for the words V, V' and U, if conditions ($*$) are satisfied.

4 Subclasses of complexity $2n + 1$

Let v be an infinite recurrent word. *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 ∂^-, we have four types of languages $L_n(v)$:

I: $\partial^+(D) = 3$ and $\partial^-(G) = 3$.
II: $\partial^+(D) = 3$ and $\partial^-(G_1) = \partial^-(G_2) = 2$, where $G_1 \neq G_2$.
III: $\partial^+(D_1) = \partial^+(D_2) = 2$, where $D_1 \neq D_2$, and $\partial^-(G) = 3$.
IV: $\partial^+(D_1) = \partial^+(D_2) = 2$, 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 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 [3]. Our aim is to prove the transcendence of minimal words in Subclass II and in the symmetric case III.

5 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 [2] with slight modifications on notations. Let v be an infinite recurrent word over Σ. The vertices of the Rauzy graph Γ_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 $E = bH, F = Ha$ 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 Γ_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 $E_0 \to E_1 \to \ldots \to E_{k-1} \to E_k$. 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 Γ_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 $\mathcal{D}_n$ and left special elements $\mathcal{G}_{1,n}$ and $\mathcal{G}_{2,n}$. All the segments indicated in Figure 1 begin and end by

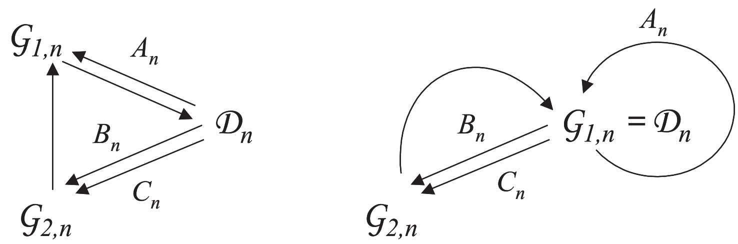

Figure 1: Graph types of Subclass II

$\mathcal{D}_n$. We fix the following notation: A_n is the name of the unique n-segment going through only one left special element $\mathcal{G}_{1,n}$. B_n and C_n are the names of the n-segments going through both left special elements $\mathcal{G}_{1,n}$ and $\mathcal{G}_{2,n}$. Sometimes we simplify the notation and mark these n-segments without the subscript n.

In the first case, $\mathcal{G}_{1,n} \neq \mathcal{D}_n$. For any word $X \neq \mathcal{D}_n$ in $L_n(v)$ there exists a unique word Xa in $L_{n+1}(v)$ and $\mathcal{G}_{1,n+1} = \mathcal{G}_{1,n}\gamma_1$, $\mathcal{G}_{2,n+1} = \mathcal{G}_{2,n}\gamma_2$ and $\mathcal{D}_{n+1} = \delta\mathcal{D}_n$ are uniquely determined by the graph Γ_n. Thus, the graph Γ_{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 $\mathcal{G}_{1,n+1}$ to D_{n+1} is one smaller than from $\mathcal{G}_{1,n}$ to D_n. This means that finally $\mathcal{G}_{1,n+l} = \mathcal{D}_{n+l}$ for some $l \geq 1$ and we are in the second case.

The interesting case is the *break down case,* where $\mathcal{G}_{1,n} = \mathcal{D}_n$. Let the three n-segments be

$$\begin{aligned} A_n &: (\mathcal{D}_n, \mathcal{D}'_n a_1, \ldots, b_1\mathcal{D}''_n, \mathcal{D}_n), \\ B_n &: (\mathcal{D}_n, \mathcal{D}'_n a_2, \ldots, b_2\mathcal{D}''_n, \mathcal{D}_n), \\ C_n &: (\mathcal{D}_n, \mathcal{D}'_n a_3, \ldots, b_2\mathcal{D}''_n, \mathcal{D}_n). \end{aligned}$$

We see that $\mathcal{G}_{1,n+1} = \mathcal{G}_{1,n}\gamma_1$, $\mathcal{G}_{2,n+1} = \mathcal{G}_{2,n}\gamma_2$ and $\mathcal{D}_{n+1} = \delta\mathcal{D}_n$, where $\gamma_1 \in \{a_1, a_2, a_3\}$, $\delta \in \{b_1, b_2\}$ and γ_2 is fixed by the graph Γ_n. Now Γ_n does not determine Γ_{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 1* .

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 $\mathcal{G}_{2,n+1} = \mathcal{G}_{1,n}\gamma_1$ and $\mathcal{G}_{1,n+1} = \mathcal{G}_{2,n}\gamma_2$ for some γ_1 and γ_2 in the alphabet Σ.

For the nth 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.

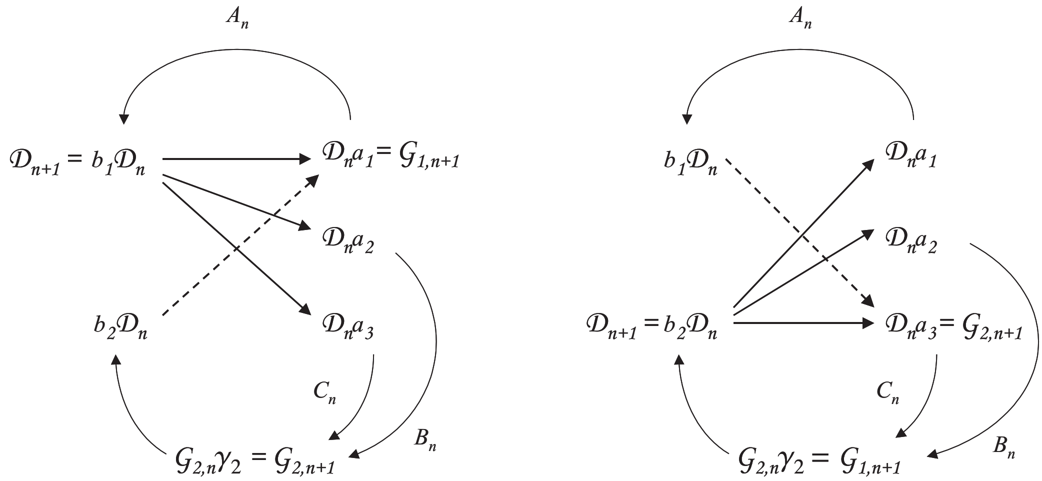

Figure 2: Rauzy graph Γ_{n+1} in break down types 1 and 3

i_n	A_{n+1}	B_{n+1}	C_{n+1}
1	A_n	B_nA_n	C_nA_n
2	C_n	A_nB_n	B_n
3	B_n	A_nC_n	C_n

Table 1: Recursion formulae of the names of the n-segments

Lemma 1. *Let v be a recurrent word in Subclass II of complexity $2n+1$. Then there exists 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 $v_0v_1 \dots v_{N-1}$ is of the form $X_0X_1 \dots X_m$, where $X_1, X_2, \dots X_{m-1}$ are equal to A_n, B_n or C_n; X_0 is a (possibly empty) suffix of A_n, B_n or C_n ; and X_m is a (possibly empty) prefix of A_n, B_n or C_n.*

6 Transcendence of numbers with the expansions in subclasses II and III

Let θ 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 1 for an infinite number of different word triplets V, V' and U satisfying conditions $(*)$ with some fixed positive integer h. We examine the directive sequence (i_n) given in Lemma 1. We have tree 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 $l \geq 3$, $i_m = 2$, $i_{m+1} = \ldots i_{m+l} = 1$ and $i_{m+l+1} = 2$, then

$$\begin{aligned} A_{m+l+2} &= B_m C_m^l, \\ B_{m+l+2} &= C_m A_m B_m C_m^l, \\ C_{m+l+2} &= A_m B_m C_m^l. \end{aligned}$$

Clearly, every segment contains now $C_m C_m C_m$ and we may choose $V = V' = C_m$. Consider then all possible prefixes U of v not containing the cube C_m^3. We use the notation of Lemma 1. If the cube can be found in X_0, then U is a suffix of B_m, $C_m A_m B_m$ or $A_m B_m$. Otherwise, X_0 is a strict suffix of C_m^3 and U is a strict suffix of $C_m^3 B_m$, C_m or $C_m^3 A_m B_m$. Hence, if there exists a fixed integer h' depending on the sequence v such that $h'|C_m| \geq |A_m|$ and $h'|C_m| \geq |B_m|$ 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 UV^2V' 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 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.

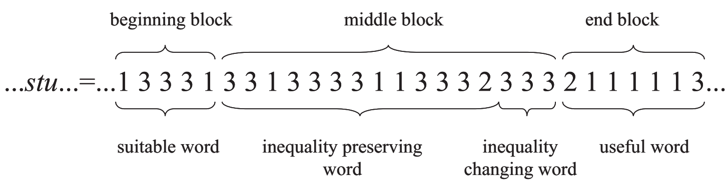

Figure 3: Illustration of blocks in the sequence (i_n).

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

Lemma 2. *Let $i_{j_0} \dots i_{j_1 - 1}$ be a suitable word for $0 \le j_0 < j_1$. Then we have the inequality*

$$2|A_{j_1} C_{j_1}| \ge |B_{j_1}| \ge |C_{j_1}| \ge |A_{j_1}|. \tag{1}$$

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 $|b| \ge |c|$ is always valid. Thus, by the calculations, after the execution of suitable words we have $2|AC| - |B| \ge 0$ for every $l \ge 0$. Examining the names of the segments in Table 2, we can also see that $|C| \ge |A|$. Hence, the inequality $2|AC| \ge |B| \ge |C| \ge |A|$ holds. □

$i_{j_0} \dots i_{j_1 - 1}$	Names of the segments	$2\lvert AC\rvert - \lvert B\rvert$
$13^{2l+1}1$	$A = ba(ca)^l$ $B = aca(ca)^l ba(ca)^l$ $C = caba(ca)^l$	$(2(2l+3) - (2l+3))\lvert a\rvert + (4-1)\lvert b\rvert$ $+(2(2l+1) - (2l+1))\lvert c\rvert$ $= (2l+3)\lvert a\rvert + 3\lvert b\rvert + (2l+1)\lvert c\rvert$
$23^{2l+1}1$	$A = abb^l$ $B = cbb^l abb^l$ $C = babb^l$	$(4-1)\lvert a\rvert + (2(2l+3) - (2l+2))\lvert b\rvert - \lvert c\rvert$ $= 3\lvert a\rvert + (2l+4)\lvert b\rvert - \lvert c\rvert$
$13^{2l}2$	$A = ca$ $B = a(ca)^l ba(ca)^l$ $C = ba(ca)^l$	$(2(l+2) - (2l+2))\lvert a\rvert + (2-1)\lvert b\rvert$ $+(2(l+1) - (2l))\lvert c\rvert$ $= 2\lvert a\rvert + \lvert b\rvert + 2\lvert c\rvert$
$23^{2l}2$	$A = b$ $B = cb^l abb^l$ $C = abb^l$	$(2-1)\lvert a\rvert + (2(l+2) - (2l+1))\lvert b\rvert - \lvert c\rvert$ $= \lvert a\rvert + 3\lvert b\rvert - \lvert c\rvert$

Table 2: Length calculations after executing suitable words

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 doesn't overlap with the end block. Since the word s ends with either 1 or 2, the sequence t must begin by $3^{2l}1$ or $3^{2l+1}2$ or $t = 3^l$ for some $l \ge 0$. Otherwise we would have a suitable word closer to u than our s, which is impossible by definition. Since the 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 $\{(33)^*1 \cup (33)^*32\}^*3^*$. We call a word in the beginning part $\{(33)^*1 \cup (33)^*32\}^*$ *inequality preserving.* The following lemma justifies this term. The end part of t belonging to 3^* is *called inequality changing.*

Lemma 3. *Let $i_{j_1} \dots i_{j_2 - 1}$ be an inequality preserving word for $0 \le j_1 < j_2$. Suppose $r|A_{j_1} C_{j_1}| \ge |B_{j_1}| \ge |C_{j_1}| \ge |A_{j_1}|$ for a positive integer $r \ge 2$. Then we have*

$$r|A_{j_2} C_{j_2}| \ge |B_{j_2}| \ge |C_{j_2}| \ge |A_{j_2}|. \tag{2}$$

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 $3^{2l}1$ or $3^{2l+1}2$ 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 $3^{2l}1$. For the case $3^{2l+1}2$ 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 $3^{2l}1$ and $3^{2l+1}2$ our inequality (2) holds. $\square$

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

Table 3: Length calculations after executing inequality preserving words

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 $3^{2l}1$ or $3^{2l+1}2$ is considered as the beginning block in this case. Whatever the situation is, Lemma 4 shows that a beginning block followed by a middle block is convenient for our purposes.

Lemma 4. *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 $i_j i_{j-1}$ means the empty word.) Then after executing the beginning block and the middle block we are either in the situation of Theorem 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

$$r|A_{j_1} C_{j_1}| \geq |B_{j_1}| \geq |C_{j_1}| \geq |A_{j_1}|, \tag{3}$$

for some fixed integer $r \geq 2$. Namely, if the beginning block is a suitable word, then by Lemma 2 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 $3^{2l}1$ or $3^{2l+1}2$. After executing $3^{2l}1$ or $3^{2l+1}2$ we have $|C_{j_1}| \geq |A_{j_1}|$, as seen in Lemma 3, and by the finiteness of the beginning part, the desired number r surely exists. As before, $|B_{j_1}| \geq |C_{j_1}|$. Therefore (3) holds. After the execution of an inequality preserving word, inequality (2) is valid by Lemma 3 or trivially, if $j_1 = j_2$. The question is, what happens in the inequality changing part $i_{j_2} \dots i_{j_3-1} = 3^l \in 3^*$. We may suppose that $j_3 > j_2$. In order to simplify notations denote the names of the segments $A_{j_2}, B_{j_2}, C_{j_2}$ by small letters a, b, c and $A_{j_3}, B_{j_3}, C_{j_3}$ by capital letters A, B and C. Then we have $A = ac^{l'}, B = bc^{l'}$ and $C = c$, if $l = 2l'$, or $A = bc^{l'}, B = acc^{l'}$ 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 1, one or the other must appear after X_0, X_0C or X_0CC or we have $v = X_0CCC\dots$. 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 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 $h' = 2r + 2$ and the lemma follows, since $|B| \geq |C|$ as noted before. $\square$

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 1 are satisfied by applying Lemma 4. From now on useful word $u = i_{j_3} \dots i_{j_4-1}$ 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. *If there exist infinitely many occurrences of the factor* 11 *in* (i_n), *we are in the situation of Theorem* 1 *for infinitely many different word triplets* U, V *and* V'.

Proof. First we note that there must be infinitely many $i_n \in \{2, 3\}$. Otherwise, for some $m \geq 1$, we have $A_{m+l} = A_m, B_{m+l} = B_m A_m^l$ and $C_{m+l} = C_m A_m^l$ 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^l2, 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.

$i_{j_3} \dots i_{j_4-1}$	Names of the segments
21^l2	$A = bc^l$ $B = cabc^l$ $C = abc^l$
21^l3	$A = abc^l$ $B = cbc^l$ $C = bc^l$

Table 4: Names of the segments after executing 21^l2 or 21^l3.

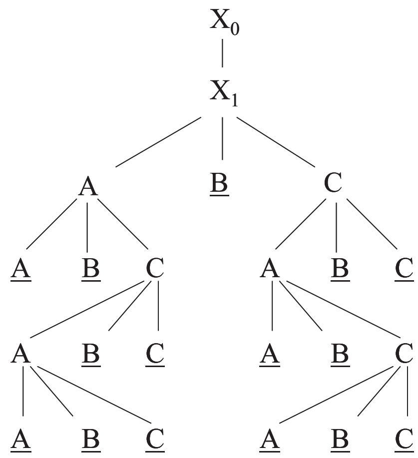

Figure 4: Tree representation of prefixes of the sequence v: Segment B and end symbols of squares are underlined.

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 case 21^l3 is similar. By Lemma 4 we are either in the situation of Theorem 1 after the execution of the middle block or

$$h'|b| \geq h'|c| \geq |x|, \tag{4}$$

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 ($*$) 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 can not 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_5. Thus $|U| \leq 5 \cdot 5h'|V|$. Hence, the conditions ($*$) 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 1. $\square$

Lemma 6. *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* 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_iAA ($i \geq 1$), BA or CA. Also any segment catenated with a square over the alphabet $\{A, B, C\}\backslash Y$ contains word $c(yc)^2 = (cy)^2c$ for some word y in $\{a, b, c\}^*$. Using Lemma 4 we conclude as in Lemma 5 that either we are in the situation of Theorem 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 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_6. Since there are infinitely many occurrences of the factor 1, we are infinitely often in the situation of Theorem 1 with $h = 36h'$. ☐

$i_{j_3} \ldots i_{j_4 - 1}$	Names of the segments	$i_{j_3} \ldots i_{j_4 - 1}$	Names of the segments
2121	$A = bc$ $B = cabcbc$ $C = abcbc$	2131	$A = abc$ $B = cbcabc$ $C = bcabc$
2122	$A = abc$ $B = bccabc$ $C = cabc$	2132	$A = bc$ $B = abccbc$ $C = cbc$
2123	$A = cabc$ $B = bcabc$ $C = abc$	2133	$A = cbc$ $B = abcbc$ $C = bc$

Table 5: Names of the segments after executing $2121, 2122, 2123, 2131, 2132$ or 2133.

Lemma 7. *If the sequence* (i_n) *consists ultimately only of integers* 2 *and* 3*, we are in the situation of Theorem* 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^ω.

First we note that the sequence 3^ω 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 in each subcase.

Case	$i_{j_3} \ldots i_{j_4-1}$	Names of the segments	$i_{j_3} \ldots i_{j_4-1}$	Names of the segments
1°	3222	$A = bac$ $B = accbac$ $C = cbac$	3223	$A = cbac$ $B = acbac$ $C = bac$
2°	2332	$A = b$ $B = cbabb$ $C = abb$	2333	$A = abb$ $B = cbb$ $C = b$
3°	3232	$A = ac$ $B = baccac$ $C = cac$	2222	$A = cab$ $B = abbcab$ $C = bcab$

Table 6: Names of the segments after executing 3222, 3223, 2332, 2333, 3232 or 2222.

As in Lemma 6, 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 4 is valid and we are either in the situation of Theorem 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 6. In the latter case $|V| < 10h'|V'|$, $|U| < 6 \cdot 6h'|V|$ and we choose $h = 36h'$. □

Finally, we state our result.

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

Proof. This is a straightforward consequence of Lemmata 5-7 and Theorem 1, the combinatorial criterion for transcendence. □

Corollary 1. *Let θ be a number with expansion $0.v$, where the sequence v is a minimal word belonging to Subclass III of complexity $2n+1$. Then θ 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 $\mathcal{G}$ by $\mathcal{D}$ and vice versa. Using similar considerations as in Section 5, 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 Lemmata 6-7. $\square$

7 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, the transcendence result follows from the work of Adamczewski *et al.* [1].

Acknowledgement: 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.

References

[1] B. Adamczewski, Y. Bugeaud, and F. Luca, Sur la complexité des nombres algébriques. C. R. Acad. Sci. Paris, ser. I **339** (2004) 11–14.

[2] P. Arnoux and G. Rauzy, Représentation géométrique de suites de complexité $2n+1$. Bull. Soc. Math. France **119** (1991) 199–215.

[3] S. Ferenczi and C. Mauduit, Transcendence of numbers with a low complexity expansion. J. Number Theory **67** (1997) 146–161.

[4] G.A. Hedlund and M. Morse, Symbolic dynamics II: Sturmian trajectories. Amer. J. Math. **62** (1940) 1–42.

[5] D. Ridout, Rational approximations to algebraic numbers. Mathematika **4** (1957) 125–131.

TURKU

CENTRE *for*

COMPUTER

SCIENCE

Lemminkäisenkatu 14 A, 20520 Turku, Finland | www.tucs.fi

University of Turku

- Department of Information Technology
- Department of Mathematics

Åbo Akademi University

- Department of Computer Science
- Institute for Advanced Management Systems Research

Turku School of Economics and Business Administration

- Institute of Information Systems Sciences

ISBN 952-12-1488-0
ISSN 1239-1891