Kolakoski Sequence and Parity of Subwords

Alessandro Della Corte*

March 16, 2020

Abstract

The Kolakoski sequence $S$ is the unique element of $\{1, 2\}^\omega$ starting with 1 and coinciding with its own run length encoding. We use the parity of the lengths of particular subclasses of initial words of $S$ as a unifying tool to address the main open questions. By means of this we prove some results and give sufficient conditions which would imply that the density of 1s is $\frac{1}{2}$.

KEYWORDS: Kolakoski sequence; Combinatorics on Words; Symbolic Dynamics.

MSC2010: 68R15, 05A99, 37B10

1 Introduction

In 1939 Oldenburger considered, within the context of symbolic dynamics, a sequence having the property of coinciding with its own run length encoding \[11\]. If we choose the alphabet $\{1, 2\}$ there are two such sequences, the second of which being simply the first one without the initial element. The two sequences start as

$$1221121221 \ldots$$

and

$$2211212212 \ldots$$

In 1965 Kolakoski rediscovered the sequence \[10\], and it was easily established that it is not eventually periodic. Besides this, very little is known about the sequence. In particular, it is still not known whether it is recurrent, mirror invariant or reversal invariant and whether the asymptotic density of 1s exists and equals $\frac{1}{2}$, a conjecture formulated by Keane \[9\]. A sharp bound ($0.5 \pm 0.00084$) for the density of 1s (assuming its existence) has been provided \[3\]. Concerning other properties of the sequence, it has been proven that it is cube-free, which is a particular case of a more general result on repetitions \[2\]. Moreover, a measure conjectured to completely describe the densities of all subwords of the sequence has been introduced, and the conjecture has been proved under fairly natural additional hypotheses \[6\]. Recursive formulas for the $n$-th element of the sequence are also known \[15\] \[8\].

*Mathematics Division, School of Science and Technology, University of Camerino (Italy). Email:alessandro.dellacorte@unicam.it
The sequence is relevant for applications concerning optical properties of aperiodic structures [16, 7, 13], but probably its most interesting features are linked to the unique combination of the simplicity of its definition and the difficulty of the problems it raises.

The sequence is nowadays indexed as A000002 in Sloane’s online Encyclopedia of Integer Sequences.

In this paper we study the open problems trying to identify a unifying concept, i.e. the parity of the integrals (the converse transform of run length encoding) of subwords of the sequence. After introducing some notation and terminology in Section 2, the main open problems are reformulated in terms of parity of integrals of prefixes in Section 3 while in Section 4 we prove some results and provide sufficient conditions implying that the asymptotic frequency of 1s is $\frac{1}{2}$. In Section 5 we describe a constructive procedure showing the existence of arbitrarily long recurrent subwords and identify places where they must occur in the structure of $S$. Finally, in Section 6 we formulate some conjectures arising from the proposed approach.

We tried to make the article completely self-contained. For this reason we used at times slightly different definitions of concepts already used in the literature.

2 Notation and preliminary definitions

Let $A^*$ be the set of the finite words on the alphabet

$$A = \{1, 2\}$$

and $A^\infty$ the set $A^* \cup A^\omega$ of all finite or infinite words on $A$. We let $\epsilon$ denote the empty word and we set $A^+ := A^* \setminus \{\epsilon\}$.

The concatenation of the finite word $w = a_1a_2\cdots a_n$ and the (possibly infinite) word $v = b_1b_2\cdots$, i.e. the word $a_1a_2\cdots a_nb_1b_2\cdots$, will be written as $wv$.

The sets $A^*$ and $A^+$ have respectively the structure of a free monoid and a free semigroup with the internal operation defined as the concatenation of words.

For every $w \in A^+$, by $\overline{w}$ we mean the mirror word of $w$, i.e. the transform of $w$ under the substitutions $1 \rightarrow 2$ and $2 \rightarrow 1$.

We also set $\overline{\epsilon} := \epsilon$. If $w = a_1\cdots a_n$ is a word in $A^+$, we set $\Sigma w := \sum_{i=1}^{n}a_i$ and $\overline{w} := a_{n}a_{n-1}\cdots a_1$, calling the former the sum of $w$ and the latter the inverse word of $w$ (we set $\overline{\epsilon} := \epsilon$). We indicate by $|w|$ the length of $w$, i.e. the positive integer $n$ (we set $|\epsilon| := 0$).

We say that $v \in A^+$ is a subword of $w = a_1a_2\cdots \in A^\infty$ if there exist a positive integer $k$ and a non-negative integer $h$ such that $v = a_ka_{k+1}\cdots a_{k+h}$. In the following it will be handy to have a term concisely referring to a particular occurrence of a subword. Therefore, we will call the pair $(v, k)$ a subrow of $w$ and will say that $(v, k)$ is an occurrence of $v$ in $w$, and that two subrows $(v_1, k)$ and $(v_2, h)$ coincide as subwords if $v_1 = v_2$. When we want to emphasize the initial and final elements of the subrow $a_ka_{k+1}\cdots a_{k+h}$, we will write it as $w_{k, k+h}$. To lighten the notation, if there is no possibility of confusion we may use the same
symbol for the subrow \((v, k)\) and the subword \(v\). We let \(\mathcal{SR}(S)\) denote the set of all the finite subrows of \(S\).

We say that \(v\) is a prefix of a (possibly infinite) word \(w = a_1 a_2 \cdots\) if there is a positive integer \(k\) such that \(v = a_1 \cdots a_k\). We say that \(v\) is a suffix of a finite word \(w = a_1 \cdots a_n\) if there is a positive integer \(k < n\) such that \(v = a_{n-k} \cdots a_n\).

Let us define a map from \(\mathcal{A}^\infty\) to itself by means of alternating substitution rules. Specifically, for every nonempty \(w \in \mathcal{A}^\infty\) we define the following substitution rules:

\[
\begin{align*}
1 & \rightarrow 1 & 2 & \rightarrow 11 & \quad \text{for the elements having odd index in } w \\
1 & \rightarrow 2 & 2 & \rightarrow 22 & \quad \text{for the elements having even index in } w
\end{align*}
\]

We let \(w^{-1}\) denote the transform of \(w\) under the substitutions. We also set \(\epsilon^{-1} := \epsilon\). For every \(w \in \mathcal{A}^\infty\), we define inductively:

\[
\begin{align*}
\quad w^0 & := w \\
\quad w^{-k} & := (w^{-(k-1)})^{-1}
\end{align*}
\]

We will refer to the \((-)^{-1}\) map as the integration map.

Alternating substitution rules are quite well investigated and have also been generalized [5]. Relative results have been used specifically to study the Kolakoski sequence. It was indeed proved that, even if (1) are very close to the simplest possible case of alternating substitution rules, its fixed point (i.e., Kolakoski sequence) cannot be obtained by iteration of a simple substitution [4].

The existence and uniqueness of the Kolakoski sequence are established by means of the following Lemma.

**Lemma 1.** There exists a unique element \(S\) of \(\{1, 2\}^\omega\) such that \(S = S^{-1}\). Moreover, indicating the \(n\)-th element of \(S\) by \(s_n\), for every positive integer \(n\) there exists a positive integer \(h\) such that \(s_n\) is the \(n\)-th element of \((12)^{-k}\) for every integer \(k\) such that \(k \geq h\).

**Proof.** Take a finite word \(u\) such that

\[u^{-1} = uv_1\]

with \(v_1 \neq \epsilon\). Integrating both sides of the previous equality one gets

\[u^{-2} = (uv_1)^{-1} = u^{-1}v_2 = uv_1v_2\]

where \(v_2\) equals \(v_1^{-1}\) or \(\overline{v_1^{-1}}\) according to \(|u|\) being respectively even or odd. Iterating the argument it follows that, for every non-negative integer \(k\),

\[u^{-k} = uv_1 \cdots v_k\]

where, for \(2 \leq h \leq k\), \(v_h = v_{h-1}^{-1}\) or \(v_h = \overline{v_{h-1}^{-1}}\) according to \(|v_1 \cdots v_{h-2}|\) being respectively even or odd. Since the words \(v_i\) are nonempty, an arbitrarily long

\[\text{The reason of this choice is that the converse transformation, which we will introduce later, is usually denoted derivative-like, with positive integers as exponents. Moreover, it will come handy when we will have to write relatively long concatenations of words which are iterations of the integration map.}\]
prefix of $u^{-k}$ remains unaltered by further integrations. More precisely, if we write $u^{-k}$ as

$$u^{-k} = a_1^k a_2^k \cdots a_{|u^{-k}|}^k$$

(where $a_i^h \in \{1, 2\}$), for every positive integer $n$ there is $h$ such that $a_j^h = a_j^{2h}$ for every $j_1, j_2 \geq h$. Hence we can define the limit sequence $S$ of the right hand side of (3) for $k \to \infty$, which will clearly verify $S = S^{-1}$. Taking $u = 12$ one gets the existence of $S$. As for the uniqueness, it follows immediately observing that the only word of length $2$ which is a prefix of its integral is $12$.

**Definition 1.** The sequence $S$ is called the Kolakoski sequence.

By the arbitrariness of the prefix $u$ in the previous proof, we easily get by induction the following

**Lemma 2.** If $p$ is a prefix of $S$, such is $p^{-k}$ for every non-negative integer $k$.

We now want to adapt the previous definition of integral so that it applies nicely to subrows of $S$, meaning that we will be able to identify which subrow of $S$ can be naturally seen as the integral of a given subrow. We thus want a version of the integration map which maps $SR(S)$ to itself (while $(\cdot)^{-1}$ maps $A^\infty$ to itself). Therefore we introduce the following

**Definition 2.** Let $w$ be a subrow of $S$ and $u$ the prefix such that $S = uw\ldots$. We define the $S$-integral of the subrow $w$ as the subrow $w_S^{-1} := S_{h,k}$ where

$$h = |u^{-1}| + 1 \quad \text{and} \quad k = |u^{-1}| + |w^{-1}|$$

We also define inductively $w_S^{-n} := (w_S^{-n+1})^{-1}$.

**Remark 1.** Since $(\cdot)^{-1}$ is a non-morphic map, the $S$-integral of a subrow $w$ does not coincide always with its integral as a subword, defined by means of the substitution rules (1). Indeed, considering $w_S^{-1}$ as a subword, we have

$$w_S^{-1} = w^{-1}$$

if $|u|$ is even and

$$w_S^{-1} = \tilde{w}^{-1}$$

if $|u|$ is odd. Notice also that in general, for a subrow $w$ which is not a prefix and for $k$ large, $w^{-k}$ and $w_S^{-k}$ will be different words which are not linked in any trivial way.

Next we want to define the property of a subrow of having $S$-integrals of even length up to a certain order, starting from the order 0 (that is, from the length of the subrow itself).

More precisely, we introduce the following

**Definition 3.** We say that the subrow $w$ is $k$-regular if $|w_S^{-h}|$ is even for $0 \leq h \leq k$. We will say that a subrow is $k$-normal if it is $k$-regular but not $(k+1)$-regular. We will say that a subrow is $\infty$-regular if it is $k$-regular for every non-negative integer $k$.
Notice that in case \( w \) is a prefix, \( k \)-regularity reduces to requiring that \( |w^{-h}| \) is even for \( 0 \leq h \leq k \).

We indicate by \( k-R \) the subset of \( SR(S) \) consisting of all the \( k \)-regular subrows of \( S \), by \( k-N \) the subset of \( SR(S) \) consisting of all the \( k \)-normal subrows of \( S \) and by \( \infty-R \) the subset of \( SR(S) \) consisting of all the \( \infty \)-regular subrows of \( S \). It is easily proved the following

**Lemma 3.**

1. For every non-negative integer \( k \), \( k-N \subset k-R \).
2. \( k \neq h \implies k-N \cap h-N = \emptyset \).
3. For every non-negative integer \( k \), \( w \in k-N \implies w \notin \infty-R \).
4. For positive integers \( a < b < c \), \( S_{a,b}, S_{(b+1),c} \in k-R \implies S_{a,c} \in k-R \).
5. For positive integers \( a < b < c \), \( S_{a,b}, S_{(b+1),c} \in k-N \implies S_{a,c} \in (k+1)-R \).

Next we want to introduce the converse operation of integration, i.e. the so-called \textit{derivative} for words in \( \mathcal{A}^* \), which, roughly speaking, coincides with a run-length counting operation. However, we should take care to avoid the ambiguity arising when a subword starts or ends with a single digit not belonging to a pair of equal elements of the alphabet, as in that case we cannot know the length of its run without looking outside the subword. For this reason it is usual [6] to cut off those single digits, whether they are present.

More precisely, for every \( w = a_1 \cdots a_n \in \mathcal{A}^* \) we define \( w' \) as the unique finite word such that \( (w')^{-1} \) equals

\[
\begin{align*}
a_1 \cdots a_n & \quad \text{if } a_1 = a_2 \text{ and } a_{n-1} = a_n \\
a_2 \cdots a_n & \quad \text{if } a_1 \neq a_2 \text{ and } a_{n-1} = a_n \\
a_1 \cdots a_{n-1} & \quad \text{if } a_1 = a_2 \text{ and } a_{n-1} \neq a_n \\
a_2 \cdots a_{n-1} & \quad \text{if } a_1 = a_2 \text{ and } a_{n-1} = a_n
\end{align*}
\]

We also set \( 1' = 2' = \epsilon' := \epsilon \). Notice that this implies \((12)' = (21)' = \epsilon \). We define the derivative of an infinite word \( v = a_1 a_2 \cdots \in \mathcal{A}^* \) as the limit sequence \((a_1 \cdots a_n)'\) when \( n \to \infty \). Finally, we define inductively \( w^{(n)} := (w^{(n-1)})' \).

Adopting the usual convention, we indicate by \( C^k \) the set of words which belong to \( \mathcal{A}^\infty \) together with their first \( k \) derivatives, and by \( C^\infty \) the set \( \bigcap_{k \in \mathbb{N}} C^k \).

Similarly to what done before for integrals, we want now to adapt the definition of derivative so that it works for subrows. Therefore we introduce the following

**Definition 4.** Let \( w \) be a subrow of \( S \). If there exists a subrow \( v \) such that \( v^{-1}_S = w \), we set \( w'_S := v \). We call \( v \) the \( S \)-derivative of \( w \).

We also define inductively \( w^{(n)}_S := (w^{(n-1)}_S)'_S \), of course if \( w^{(k)}_S \) admits an \( S \)-derivative for every \( k \leq n \).

**Remark 2.** Notice that not every subrow has an \( S \)-derivative. For instance there is no subrow \( u \) such that \( s_3 s_4 = u^{(1)}_S \).

The following Lemma characterizes the subrows admitting an \( S \)-derivative.
Lemma 4. Let $n$ and $m$ be two positive integers such that $n < m$. If $w = s_n \ldots s_m$ is a subrow of $S$, it admits an $S$-derivative if and only if $s_{n-1} \neq s_n$ and $s_m \neq s_{m+1}$. Moreover, if $w'_S = s_h \ldots s_{h+k}$ then $h \leq n$ and $h+k \leq m$.

Proof. Suppose that $w$ admits an $S$-derivative. Then $w$ is the transform under substitutions (1) (or the mirror of the transform under (1)) of another subrow $v = s_h \ldots s_j$. This means in particular that $(s_j)^{-1}_S = s_m$ or $(s_j)^{-1}_S = s_{m-1}s_m$, so that $(s_{j+1})^{-1} = s_{m+1}$ or $(s_{j+1})^{-1} = s_{m+1}s_{m+2}$. Since $j$ and $j+1$ cannot be both even or both odd, it follows that $s_m \neq s_{m+1}$. The proof proceeds analogously for $s_n$ if $s_{n-1}$ exists, otherwise (i.e. if $w$ is a prefix), the thesis is vacuously true.

Conversely, suppose that $s_{n-1} \neq s_n$ and $s_m \neq s_{m+1}$. Then, by definition of $S$, $w$ is the transform under (1) (or the mirror of the transform under (1)) of some subrow $v$.

Finally, if $w'_S = s_h \ldots s_k$, the inequalities $h \leq n$ and $h+k \leq m$ easily follow from the fact that, for every word $w$, $|w^{-1}| \geq |w|$.

Since $w'$ is always a subword of $w'_S$, the following Lemma follows from the previous one:

Lemma 5. If $w = s_ns_{n+1} \ldots s_{n+m}$ is a subword of $S$ with nonempty derivative, there exists $h$ and $k$ such that $w' = s_h s_{h+1} \ldots s_{h+k}$.

The previous Lemma immediately implies that

Lemma 6. If $w$ is a subword of $S$, then $w \in C^\infty$.

Finally we introduce some definitions concerning asymptotic frequencies of subwords. We indicate by $|w|_v$ the number of occurrences of the subword $w$ in $v$, and by $f_v(w)$ the frequency of $w$ in $v$, i.e. the number $|w|_v / |v|$. We set

$$f_\infty(w) := \lim_{n \to \infty} \frac{|w|_{S_1,n}}{n}$$

whether the limit exists. The most famous conjecture concerning Kolakoski sequence is Keane’s conjecture [9]:

$$f_\infty(1) \text{ exists and equals } \frac{1}{2}$$

3 Reformulation of the problems

In this section we will reformulate some open questions concerning $S$ in terms of regularity of subrows. Let us recall that, according to (2), elements with even (odd) index in $S$ are mapped by the integration in 2 or 22 (1 or 11). We will use systematically this fact (usually without mentioning it explicitly) throughout.

In the following we will need a (rough) estimate of the relative length of $w$, $w'$ and $w^{-1}$ when $w$ is a subword of $S$, ensuring in particular that, for every finite word $w$ with more than one element,

$$|w^{(k)}| \to 0 \quad \text{if } k \text{ diverges}$$

and

$$|w^{-k}| = |w_S^{-k}| \to \infty \quad \text{if } k \text{ diverges}$$
and that for every positive integer \( k \),

\[ |w^{(k)}| \to \infty \quad \text{if} \quad |w| \ \text{diverges} \quad (6) \]

This is obtained observing that \( |w^{-1}| = \sum w \), and that the maximum and minimum local density of 2s in a \( C^\infty \) word are achieved respectively by 11211 and 22122. From this (recalling that the derivative cuts off single digits at both ends) the following Lemma is easily proved.

**Lemma 7.** If \( w \) is a subrow of \( S \) and \( |w| \geq 3 \), then

\[
\frac{6}{5}|w| \leq |w^{-1}| = |w_S^{-1}| \leq \frac{9}{5}|w|,
\]

\[
\frac{5}{9}|w| \leq |w'| \leq \frac{5}{6}|w| \quad \text{(if \ w \ admits an S-derivative)}
\]

and

\[
\frac{1}{4}|w| \leq |w'| \leq \frac{3}{4}|w|
\]

The asymptotic behaviors (4), (5) and (6) immediately follow from Lemma 7 (notice that, if \( |w| = 2 \), \( |(ws)^{-2}| \geq 3 \)).

We add some definitions: a sequence \( w \in \mathcal{A}_\omega \) is called **recurrent** if every finite subword of it is repeated (and therefore every finite subword is repeated infinitely many times). It is called **uniformly recurrent** if it is recurrent and the gaps between disjoint consecutive occurrences of every given finite subword are bounded. Moreover, \( w \) is called **mirror invariant** (reversal invariant) if the set of its finite subwords is closed under the mirror operation: \( v \to \tilde{v} \) (inverse operation: \( v \to ^{-1}v \)).

It is a well known result that, for \( S \), mirror invariance implies recurrence. We will provide in Theorem 2. For this, though, we need some preliminary result.

The links between recurrence, mirror invariance and regularity/normality of subwords are established in the following Lemmas.

**Lemma 8.** \( S \) is recurrent if and only if for every positive integer \( k \) there is a \( k \)-regular prefix of \( S \).

**Proof.** Suppose that \( w \) is a \( k \)-regular prefix of \( S \).

From Lemma 2 recalling the substitution rules (1) and the \( k \)-regularity of \( w \), it follows that \( w^{-1}1 \) and \( w^{-2}12 \) are both prefixes for \( S \). Then integrating further (and again recalling that \( w \in k-R \)) we find a prefix of the form

\[
(w^{-2}12)^{-k+2} = w^{-k}(12)^{-k+2} \quad (7)
\]

By Lemma 3 the last factor in the right hand side of (7) coincides with a prefix of \( S \). Recalling (5), this prefix is arbitrarily long if \( k \) is large enough, which is sufficient to conclude that \( S \) is recurrent.

Conversely, suppose that \( S \) is recurrent. This implies, in particular, that arbitrarily long prefixes of \( S \) are repeated infinitely many times, thus for every positive integer \( N \) there is a prefix \( w \) and a prefix of the form \( www \) such that
$|w| > N$ and $|v| > N$. We can assume that the last element of $w$ is not equal to the first element of $v$. Moreover, since $w$ starts with 12211, by Lemma 8 the last element of $v$ has to be different from the first element of $w$, as otherwise $vw \notin C^\infty$. Therefore, by Lemma 4 there exists a nonempty word $u_1$ such that, setting $p_1 := (wS)'$, the word $p_1u_1p_1$ is also a prefix for $S$.

We then define recursively

$$p_{i+1} := (S_{k, |p_i|})'_S$$

and

$$u_{i+1} := (u_i)'_S$$
in case $s_{|p_i|} \neq s_{|p_i|+1}$. If instead $s_{|p_i|} = s_{|p_i|+1}$, we replace, in the definition of $p_{i+1}$ and $u_{i+1}$, $p_i$ with the largest of its prefixes admitting an $S$-derivative and $u_i$ with the smallest subrows having $u_i$ as a suffix and admitting an $S$-derivative.

Since $|v|$ can be arbitrarily large and recalling Lemma 9, we have that

$$p_ku_kp_k$$
is a prefix of $S$ (with $u_k \neq \epsilon$) for every $k$ such that $|p_k| \geq 2$. Therefore $p_k$ starts with 1 for every $k$ such that $|p_k| \geq 2$. Since $p_k$ also follows the prefix $p_ku_k$ in $S$, this means, recalling the substitution rules (1), that the first $k - 1$ integrals of the prefix $p_ku_k$ have even length. By Lemma 7 it follows that $k \to \infty$ when $|w| \to \infty$.

Lemma 9. $S$ is mirror invariant if and only if for every positive integer $n$ there is a $k$-normal prefix of $S$ with $k > n$.

Proof. Suppose that $w$ is a $k$-normal prefix of $S$. As seen in the proof of Lemma 8 $w^{-k}(12)^{-k+2}$ is also a prefix for $S$ for every $h \leq k + 1$. Since $|w^{-k-1}|$ is odd, integrating further and recalling Lemma 8 one gets the prefix $w^{-k-2}v$, where $v = (12)^{-k}$. By Lemma 1 $v$ is also a prefix of $S$, and $|v|$ is arbitrarily large if $k$ is large enough. Since concatenation commutes with the mirror operation, this is sufficient to conclude that $S$ is mirror invariant.

Conversely, suppose that $S$ is mirror invariant and let $w$ be a prefix of $S$ such that its last element is not equal to the following element of $S$. By mirror invariance there will be a prefix of the form $wvw$. Let us define the subwords $p_n$ and $u_n$ as done in the previous proof, and let $\bar{n}$ be the largest integer for which $|p_{\bar{n}}| \geq 2$. Since $S$-derivatives of mirror words coincide as subwords, there will be prefixes of the form

$$p_nu_np_n$$

with $u_n$ nonempty for every positive integer $n \leq \bar{n}$ (notice that this means that $S$ is recurrent). As $p_n$ is a prefix of $w$ and so starts with 1 for every $n \leq \bar{n}$, Lemma 8 ensures that the prefix $p_nu_n$ is $k$-regular for arbitrarily large positive integers $k$ if $|w|$ and $|v|$ are chosen large enough. Moreover, since $\bar{w}$ starts with 2 and $(p_1u_1p_1)^{-1} = w\bar{w}$ by hypothesis, $|p_1u_1|$ has to be odd, and therefore the prefix $p_nu_n$ is $(\bar{n} - 2)$-normal, where $\bar{n} - 2$ can be arbitrarily large if $|w|$ is chosen large enough.

It is well known that mirror invariance is equivalent to: every $C^\infty$ finite word occurs in $S$. One implication is obvious, while the other easily follows from the fact that mirror invariance implies that, if $w$ does not occur in $S$, neither does $w'$. This result and Lemma 9 mean that
Lemma 10. Every $C^\infty$ word is a subword of $S$ if and only if for every positive integer $n$ there is a $k$-normal prefix of $S$ with $k > n$.

Concerning uniform recurrence, we have the following

Lemma 11. $S$ is uniformly recurrent if and only if, for every positive integer $n$, $S$ can be written as an infinite concatenation of $k$-regular subwords of bounded length with $k > n$.

Proof. Suppose that, for every positive integer $N$, there exists a sequence of subwords $w_i (i \in \mathbb{N})$ and a positive integer $M$ such that $|w_i| < M$ for every $i$ and

$$S = w_1 w_2 \cdots$$  \hspace{1cm} (8)

with every $w_i \in k$-R and $k > N$. Then integrating (8) $k$ times yields

$$S = w_1^{-k} w_2^{-k} \cdots$$  \hspace{1cm} (9)

Since $|w_i^{-k}|$ is even for every $h \leq k$, the word $(12)^{-k+2}$ is a prefix of $w_i^{-k}$ for every $i \geq 2$, and by Lemma 1 it is also a prefix of $S$. Since, by Lemma 7 $|w_i^{-k}| < M (\frac{5}{6})^k$, it follows that $S$ is uniformly recurrent.

Conversely, suppose that $S$ is uniformly recurrent. Then, for every prefix $w$, $S$ can be written as

$$S = w_1 w_2 w \cdots$$  \hspace{1cm} (10)

with $2 < |u_i| < M$ for every $i$ for some $M > 0$. We can assume that $w$ ends with 11 or 22, so that its last element is not equal to the first element of $u_i$ for every $i$. Since $w$ starts with 12211, every $u_i$ has to end with 2, otherwise a subword which is not $C^\infty$ would occur in $S$, which is not possible by Lemma 6.

By Lemma 4 we then have

$$S = p_k w_1 w_2 \cdots$$  \hspace{1cm} (11)

If $w$ is long enough, defining the subwords $p_i$ as in the proof of Lemmas 8 and 9 and the subrows $u_j$ accordingly, we can iterate $k$ times the argument, so as to obtain

$$S = p_k u_1^k p_k u_2^k p_k \cdots$$  \hspace{1cm} (12)

As $w$ is a prefix of $S$, $p_k$ begins with 1 for every positive integer $k$. Since $M$ can be chosen arbitrarily large (by simply neglecting a suitable number of occurrences of $w$ in $S$ if needed), $|u_k|$ can be made nonempty for arbitrarily large $k$ because of (13). Therefore, each prefix of the the form

$$p_k u_1^k \cdots p_k u_n^k$$

is $k$-regular with arbitrarily large $k$, and thus so is every subrow $S_{a,a+b}$ where $a = |p_k u_1^k| + 1$ and $b = |p_k u_b^k|$. By Lemma 7 every such subrow has length smaller than $\left(\frac{5}{6}\right)^k M|w|$, which concludes the proof.

Remark 3. Lemmas 8, 9, 10 and 11 can be straightforwardly adapted to generalized Kolakoski sequences defined over binary alphabets $\{m, n\}$ other than $A$ if we define analogously the concepts of regularity and normality of subrows.

On generalized Kolakoski words we mention the works by Sing [13, 14] and, in an interesting but slightly different direction compared to typical Kolakoski literature, by Shen [12].
4 Main results

Let us start by observing that the existence of an $\infty$-regular prefix of $S$ would have strong consequences on its structure and properties, as $S$ would then be recurrent and would have a rigidly fractal structure.

More precisely, we establish the following

**Theorem 1.** Suppose that $S$ has an $\infty$-regular prefix $w$ and let $k$ be a positive integer large enough so that $|w| > |(12)^{-k-2}|$. Then, for every positive integer $n$, $S$ has a prefix with the following structure:

$w^{-nk}w^{-(n-1)k} \ldots w^{-k}w \quad (13)$

In particular, $S$ is recurrent.

**Proof.** Since the prefix $w$ is $\infty$-regular, there exists a positive integer $\bar{h}$ such that

$w^{-h}(12)^{-h}w$

is also a prefix for every $h \geq \bar{h}$. Therefore arbitrarily long prefixes of $S$ are repeated, which is enough to have recurrence. In particular, if $k$ is such that $|w| > |(12)^{-k-2}|$, then, by Lemma 1, $w^{-k}w$ is a prefix. Integrating further for $k$ times, and recalling that $w \in \infty\text{-R}$, we get the prefix

$w^{-2k}w^{-k}w$

and continuing the integrations for further $(n-2)k$ times we get (13). Notice that, since $w^{-nk}$ is a prefix for every $n$, it has to begin with $w^{-hk}$ for every $h < k$.

**Remark 4.** Theorem 1 can be applied to generalized Kolakoski words. Its application to Kolakoski words over binary alphabets $\{m, n\}$ in which $m$ and $n$ are both even or both odd immediately implies that those sequences are recurrent, which is a well-known result [14].

As already said, it is a known result that if $S$ is mirror invariant, then it is recurrent [6] (notice that Lemmas 8 and 9 imply that). The converse implication is obtained with an additional hypothesis in the following

**Theorem 2.** $\infty\text{-R} = \emptyset \Rightarrow (S$ is recurrent $\iff S$ is mirror invariant$).$

**Proof.** We only have to prove that, under the hypothesis, recurrence implies mirror invariance.

Suppose that $\infty\text{-R} = \emptyset$ and that $S$ is recurrent. Then by Lemma 8 there will be a strictly increasing sequence of positive integers $k_n$ such that $S$ has a $k_n$-regular prefix $w_n$ for every $n$. Since $w_n \not\in \infty\text{-R}$, there is a positive integer $k$ which is the least integer such that $|w_n^{-k}|$ is odd. As seen in the proof of Lemma 8 $w_n^{-k}(12)^{-k+2}$ is also a prefix of $S$, and integrating once more (recalling Lemma 2) we get

$S = w_n^{-k-1}(12)^{-k+1} \ldots$

Recalling that $(12)^{-k+1}$ is also a prefix of $S$ by Lemma 1, and that $k$ can be taken arbitrarily large by suitably choosing $w_n$, we can conclude. □
Theorem 3. $S$ is recurrent if and only if $S$ is reversal invariant.

Proof. The implication from reversal invariance to recurrence is a known result holding for all elements of $A^\omega$ (for the application to differentiable sequences see [1], where recurrence of $S$ is shown to be implied by the existence of arbitrarily long palindromes).

Conversely, suppose that $S$ is recurrent. Then, by Lemma 8 for every integer $k$ the sequence $S$ has a $k$-regular prefix $w_k$. We have

$$S = w_k^{-2}12\ldots$$

where it is easily seen that $w_k^{-2}$ must end with 2. Defining $v$ by $v2 = w_k^{-2}$ we can write

$$S = v212\ldots$$

Integrating (15), and recalling that $v2 \in (k-2)$-R, we have

$$S = (v2)^{-1}(12)^{-1}\ldots$$

and as $|v2|$ and $|v212|$ are both even, the 2 appearing as the last element of $v2$ is transformed by the rules (1) in the same way as the 2 appearing as the last element of $v212$. Therefore prefix $(v2)^{-1}$ can be rewritten as

$$v^{-1}(12)^{-1}$$

Proceeding by induction, suppose that the prefix $(v2)^{-h}$ can be rewritten as

$$v^{-h}(12)^{-h}$$

Integrating $h$ times (15), and recalling that $v2$ is $(k-2)$-regular, we get the prefix

$$(v2)^{-h}(12)^{-h}$$

Let be $n := |(v2)^{-h}|$. Comparing (18) and (19), it follows that, for every integer $j$ such that $0 \leq j \leq |(12)^{-h}|$, the $(n-j)$th element of $(v2)^{-h}$ is equal to and has always the same parity of the $(j+2)$th element of $(12)^{-h}$, and therefore it transformed by the rules (1) in the same way. Therefore, since integrating (19) we get $(v2)^{-h-1}(12)^{-h-1}$, integrating (18) we get the prefix

$$v^{-h-1}(12)^{-h-1}$$

Since $(12)^{-k}$ is a prefix of $S$ for every $k$ by Lemma 11 the arbitrariness of $k$, and thus of $h$, allows us to conclude that $S$ is reversal invariant.

A natural question is whether the property of a finite word of being $k$-regular for arbitrarily large $k$ is compatible with the requirement of being $C^\infty$. In fact it is possible to prove more, i.e., that $S$ can be eventually written as a concatenation of arbitrarily regular words. More precisely, we have the following

Theorem 4. For every non-negative integer $n$, there exist a finite word $u_n$ and finite words $w_i$ ($i = 1, 2\ldots$) such that

$$S = u_n w_1 w_2 \ldots$$

where $w_i \in k$-R for every $i$ and $k \geq n$. 

11
Proof. We proceed by induction. Let us suppose that $S$ has the form \[(21)\] and that $w_i \in k$-R $\forall i$. Let $M$ be the set of positive integers $i_m$ such that $w_{i_m} \notin (k+1)$-R. Clearly the subrows $w_{i_m}$ are $k$-normal. If $M$ is finite, we define $p := \max \{j \in \mathbb{N}^+ : j \in M\}$ and $u_{n+1} := u_n w_1 \ldots w_p$ so that

$$S = u_{n+1}w_{p+1}w_{p+2} \ldots \quad (p = 1, 2 \ldots)$$

which is the desired result.

If $M$ is infinite, $i_m$ is a subsequence of $i$, so for every positive integer $m$ we can define the words $v_m := w_{i_m}w_{(i_m + 1)} \ldots w_{(i_m + 1)}$. Every $v_m$ is a concatenation of the $k$-normal word $w_{i_m}$, the (possibly empty) word formed by the $i_{(m + 1)} - i_m - 1$ words $w_{i_m + h} \ (h = i_m + 1 \ldots i_{m+1} - 1)$, which are $(k+1)$-regular, and the $k$-normal word $w_{(i_{m+1})}$. Therefore, by Lemma \[8\] $v_m \in (k+1)$-R for every $m$, and therefore defining $p := \min M$ and the word $u_{n+1} := u_n w_1 \ldots w_{p-1}$, we have

$$S = u_{n+1}v_1v_2 \ldots v_{2m+1} \ldots \quad (m = 1, 2 \ldots) \quad (22)$$

which is the desired result.

Finally, $S$ is obviously written as a concatenation of 0-regular words, so the proof is concluded. \[\square\]

Remark 5. In the previous Lemma, if we start the inductive construction of the words $w_i$ from $u_0 := \epsilon$ and $w_1 := s_{2i-1}s_{2i}$ ($i = 1, 2 \ldots$), we can have two different possibilities:

1. $u_n = \epsilon$ for every non-negative integer $n$. In this case $S$ is written as a concatenation of $k$-regular words for every $k$, and therefore it is recurrent and reversal invariant by Lemmas \[8\] and \[4\].

2. At some step $\bar{n}$ of the inductive procedure we have $\epsilon \neq u_{\bar{n}} \in (\bar{n} - 1)$-N. By Lemma \[8\] it follows that in this case, continuing the inductive procedure, we will have $u_n \in (\bar{n} - 1)$-N for every $n > \bar{n}$.

Remark 6. We can apply the iterative procedure of Theorem \[4\] starting from an arbitrary element of $S$, as no special properties of the beginning of $S$ were used. This means that, for every positive integer $k$, the same conclusion of the Lemma applies to the sequence $s_k s_{k+1} s_{k+2} \ldots$.

To proceed further we need one more definition, as we want to assign a special name to the subrows $v_i$ in \[3\]. We recall that, for every prefix $w$ of $S$ we have, by Lemma \[2\] that $w^{-k}$ is also a prefix.

Definition 5. For every prefix $w$ such that $|w| > 1$, and for every positive integer $k$, we define the $k$-th block generated by the prefix $w$ as the unique subrow $b_k$ such that $w^{-k+1}b_k = w^{-k}$.

We have clearly

$$S = w b_1 b_2 b_3 \ldots \quad (23)$$

and

$$(w b_1 \ldots b_{k-1})^{-1} = w^{-k} = w b_1 \ldots b_k \quad (24)$$

Notice that, for every $k$, $b_{k+1} = (b_k)^{-1}$ so that, for every $k$:

$$b_k = (b_{k-1})^{-1} \quad \text{if } |w b_1 \ldots b_{k-2}| \text{ is even}$$
\[ b_k = \left( \frac{b_{k-1}}{2} \right)^{-1} \text{ if } |w| \text{ is odd} \]

Since \(|w| = |w^*|\) for every finite word \(w\), we have by Lemma 7 that
\[
\frac{6}{5} |b_{k-1}| \leq |b_k| \leq \frac{9}{5} |b_{k-1}| 
\]
and therefore
\[
\left( \frac{6}{5} \right)^{k-1} |b_1| \leq |b_k| \leq \left( \frac{9}{5} \right)^{k-1} |b_1| 
\]
A block has the property of being always “not small” with respect to whatever comes before it in the sequence, and therefore the asymptotic frequency of the element 1 and 2, if they exist, have to be reached uniformly on the blocks. More precisely, we have the following

**Lemma 12.** Let \(w\) be a prefix \(|w| \geq 2\) of \(S\) and \(b_k (k = 1, 2, \ldots)\) the blocks generated by \(w\). Suppose \(f_\infty(1)\) exists. Then for every \(\epsilon > 0\) there is an integer \(n\) such that \(|f_{b_k}(1) - f_\infty(1)| < \epsilon\) for every \(k \geq n\).

**Proof.** Let us define the prefix \(u_k := wb_1 \ldots b_k\). Using (25), it is easily shown that there exist two real numbers \(c_1\) and \(c_2\), with \(0 < c_1 < c_2 < 1\) such that, for every \(k\) large enough,
\[
|b_k| \geq c_1 |u_k| \quad \text{and} \quad |b_k| \leq c_2 |u_k| 
\]
We have
\[
f_{u_k}(1) = f_{u_{k-1}}(1) \frac{|u_{k-1}|}{|u_k|} + f_{b_k}(1) \frac{|b_k|}{|u_k|} \geq f_{u_{k-1}}(1)(1 - c_2) + f_{b_k}(1) c_1 
\]
Since \(u_k\) is a prefix for every \(k\) and by (25) \(|u_k| \rightarrow \infty\) when \(k \rightarrow \infty\), if there exists \(f_\infty(1)\) then for every \(\epsilon > 0\) there is \(h\) so large that, for every \(k > h\),
\[
f_{u_{k-1}}(1) = f_\infty(1) + \epsilon_1 
\]
and
\[
f_{u_k}(1) = f_\infty(1) + \epsilon_2 
\]
with \(\max \{\epsilon_1, \epsilon_2\} < \epsilon\). Therefore, from (28) we have
\[
f_{b_k}(1)c_1 \leq f_{u_k}(1) - f_{u_{k-1}}(1)(1 - c_2) = \epsilon_2 - \epsilon_1 + f_\infty(1)c_2 + \epsilon_1 c_2 
\]
whence
\[
f_{b_k}(1)c_1 - f_\infty(1)c_2 \leq \epsilon_2 + \epsilon_1(1 - c_2) 
\]
so that
\[
(f_{b_k}(1) - f_\infty(1))c_1 \leq \epsilon_2 + \epsilon_1(1 - c_2) + f_\infty(1)(c_2 - c_1) 
\]
If there exists \(f_\infty(1)\) the difference \(c_2 - c_1\) becomes arbitrarily small when \(k\) diverges. Indeed:
\[
\frac{|b_k|}{|u_k| - |u_{k-1}|} = f_{u_{k-1}}(1) + 2 \left( 1 - f_{u_{k-1}}(1) \right) 
\]
The right hand side of (32) tends to \(2 - f_\infty(1) = 1 + f_\infty(2)\) when \(k \rightarrow \infty\). Therefore also \(\frac{|b_k|}{|u_k| - |u_{k-1}|}\) converges to a limit when \(k\) diverges, as it is obviously \(|u_k| = |u_{k-1}| + |b_k|\). Therefore, we can take \(c_1\) and \(c_2\) such that \(c_2 - c_1\) is arbitrarily small if \(k\) is large enough, and thus (31) implies that \(|f_{b_k}(1) - f_\infty(1)|\) also becomes arbitrarily small when \(k\) diverges.
Remark 7. Let us take another prefix \( v \) with \( |v| > |w| \) and let \( d_k \) denote the blocks generated by \( v \). Then instead of (32) we have

\[
\frac{|d_k|}{|u_k|} = f_{p_{k-1}}(1) + 2 \left( 1 - f_{p_{k-1}}(1) \right)
\]

(33)

where \( p_k := vd_1 \ldots d_k \). Clearly the right hand side of (33) converges to \( 1 + f_\infty(2) \) faster than the right hand side of (32), as for every \( k \) we have \( |p_k| > |u_k| \). From this it easily follows that for every \( \epsilon > 0 \), if \( n \) satisfies Lemma 13 for a given prefix \( w \), then it satisfies Lemma 17 for \( v \), i.e. \( |f_{b_k}(1) - f_\infty(1)| < \epsilon \) for every \( k \geq n \) implies \( |f_{d_k}(1) - f_\infty(1)| < \epsilon \) for every \( k \geq n \).

Definition 6. We say that a prefix is \( k \)-minimal if it is the shortest \( k \)-normal prefix of \( S \).

Clearly for every positive integer \( k \) there is at most one \( k \)-minimal prefix.

We want now to provide sufficient conditions (as weak as possible) implying that Keane’s conjecture is true. Specifically, we will require the existence of arbitrarily normal prefixes as well as a uniformity property, i.e. that a sufficiently large portion of every sufficiently advanced block is representative of the frequency of 1s on that block. This portion, however, is allowed to become arbitrarily large for blocks which are advanced enough in \( S \).

More precisely, we have the following

Theorem 5. Suppose that

1. There exists \( f_\infty(1) \);
2. There is a strictly increasing sequence of positive integers

\[
K := k_1, k_2, k_3, \ldots
\]

such that there exists a \( k_n \)-normal prefix of \( S \) for every \( n \);
3. for every \( \epsilon > 0 \), \( |f_{b_{k_n}}(1) - f_{c_{k_n}}(1)| < \epsilon \) for every \( n \) large enough, where

- \( b_{k_n} \) are the blocks generated by the \( k_n \)-minimal prefix \( p \),
- \( c_{k_n} \) is a prefix of the block \( b_{k_n} \) such that \( \frac{|b_{k_n}|}{|c_{k_n}|} < |p| \left( \frac{1}{2} \right)^{k_n+2} \)

Then \( f_\infty(1) = \frac{1}{2} \).

Remark 8. Notice that \( |c_{k_n}| \) is allowed to become arbitrarily large with \( k_n \).

Proof. Take \( \epsilon > 0 \) and consider a prefix \( w \) so large that

\[
|f_v(1) - f_\infty(1)| < \epsilon
\]

(34)

for every prefix \( v \) such that \( |v| \geq |w| \). We can find a \( k_n \)-minimal prefix \( p \) with \( k_n \) being the smallest element of the sequence \( K \) such that \( w \) is a prefix of \( (12)^{-k_n+2} \), which implies that \( p^{-k_n} w \) is a prefix of \( S \). By Lemma 12 there will be a positive integer \( m \) such that the \( m \)-th element \( k_m \) of the sequence \( K \) has the property that \( |f_{b_j}(1) - f_\infty(1)| < \epsilon \) for every \( j \geq k_m \), where \( b_j \) are the blocks generated by \( p \). There are two possibilities:
1. \( k_m > k_n \).

Then by hypothesis 2, we can find a \( k_s \)-minimal prefix \( q \) with \( k_s \) being the smallest element of the sequence \( K \) such that \( k_s \geq k_m \) and \( |q| > |p| \) (of course the latter is verified for some \( k_s \) because there are only finitely many prefixes which are shorter than \( p \)). Clearly \( q^{-k_s}w \) is also a prefix. Moreover, recalling Remark 7 and denoting by \( d_j \) the blocks generated by \( q \), we have

\[
|f_{d_j}(1) - f_\infty(1)| < \epsilon \quad (35)
\]

for every \( j \geq k_m \) and thus in particular for \( j \geq k_s \). We can assume that \( k_s \geq 2 \), so that \( |q| \geq 16 \), \( |d_1| \geq 8 \) and thus

\[
q^{-k_s}d_{k_s+1} = q^{-k_s}wu \quad (36)
\]

with \( u \) nonempty.

Since \( |q^{-k_s-1}| \) is odd, integrating two more times we get the prefix

\[
q^{-k_s-2}d_{k_s+3} = q^{-k_s-2}w^{-2}u_S^{-2}
\]

where \( u_S^{-2} \) is nonempty, so that \( w^{-2} \) is a prefix of \( d_{k_s+3} \) and does not coincide with it. If \( \delta \) is the least integer such that \( w^{-2} \) is a prefix of \( (12)^{-k_s-2} \), setting \( k_s - k = \delta \), we have that \( w^{-2} \) is a prefix of \( d_{k_s+3} \) (notice that it can be \( \delta = 0 \)). Since \( w^{-2} \) is a prefix of \( S \) longer than \( w \), by assumption both the following inequalities are verified:

\[
|f_{d_{k_s+3}}(1) - f_\infty(1)| < \epsilon \quad (37)
\]

\[
|f_{w^{-2}}(1) - f_\infty(1)| < \epsilon \quad (38)
\]

We can bound \( |d_{k_s+3}| \) from above by means of (26) and Lemma 7 whence, recalling that \( |q^{-1}| = |q| + |d_1| \), we obtain

\[
|d_{k_s+3}| \leq \left( \frac{9}{5} \right)^{k_s+2} |d_1| \leq \left( \frac{9}{5} \right)^{k_s+2} \frac{4}{5} |q| \quad (39)
\]

Moreover, we can bound from below \( |w^{-2}| \). Indeed we have \( |w| > |(12)^{-k_s-1}| \), thus

\[
|w^{-2}| > |(12)^{-k_s-1}| \geq 2 \left( \frac{6}{5} \right)^{-k_s-1}
\]

where the last inequality holds because of Lemma 4.

Therefore the ratio \( \frac{|d_{k_s+3}|}{|w^{-2}|} \) is bounded from above by \( \left( \frac{9}{5} \right)^{k_s+2} |q| \), and therefore, by hypothesis 3., for every \( \theta > 0 \),

\[
|f_{w^{-2}}(1) - f_{d_{k_s+3}}(1)| < \theta
\]

if \( n \) is large enough. Combining the last inequality with (37) and (38), we get that the difference \( |f_{w^{-2}}(1) - f_{w^{-2}}(2)| \) is arbitrarily small if \( n \) is large enough, and therefore so is, by definition of mirror word, the difference \( |f_{w^{-2}}(1) - f_{w^{-2}}(2)| \), from which we can conclude.
2. \( k_n \geq k_m \).

Then we proceed as above with the prefix \( p \) instead of \( q \) and \( k_n \) instead of \( k_m \).

\[ \square \]

5 An iterative procedure providing arbitrarily long recurrent subwords

In this section we want to use the iterative construction shown in the proof of Theorem 4 to establish constructively the existence of recurrent subwords of arbitrary length and identify places where they must appear in the structure of \( S \). Before this, let us explicitly remark that:

- the existence of at least one recurrent subword of any given finite length is shown by a trivial cardinality argument;
- the existence of two recurrent subwords of any given length is easily proved using the fact that \( S \) cannot be eventually periodic;
- the existence of more than two recurrent subwords of any given length is also easy. Indeed, assuming that only two recurrent subwords \( w_1 \) and \( w_2 \) of length \( L \) exist, one can deduce the existence of other recurrent subwords which are subwords of \( w_1 w_2 \) or of \( w_2 w_1 \), thus obtaining a contradiction.

However, all these arguments are completely non-constructive.

We start by associating to every subrow of \( S \) a sequence over \( \{0, 1\}^\omega \) describing the parity of all its \( S \)-integrals. More precisely, we introduce the following definition:

**Definition 7.** For every subrow \( w \) we define \( P_0(w) = 0 \) if \( |w| \) is even, \( P_0(w) = 1 \) otherwise. We define inductively \( P_n(w) = 0 \) if \( |w_{-n}^{n+1}| \) is even, \( P_n(w) = 1 \) otherwise. We will call the sequence \( P_n(w) \) the history of parity of the integrals of \( w \).

In the particular case in which \( w \) is a prefix, \( P_n(w) \) is simply the sequence describing the parities of \( |w^{n+1}| \).

**Lemma 13.** Suppose that \( u_1 \) and \( u_2 \) are distinct prefixes of \( S \) such that \( S = u_1 w_1 \ldots \) and \( S = u_2 w_2 \ldots \) with the subrows \( w_1 \) and \( w_2 \) coinciding as subwords. If there exists \( \bar{n} \) such that \( P_n(u_1) \neq P_n(u_2) \), then \( (w_1)^{-k} \neq (w_2)^{-k} \) for every \( k \geq \bar{n} \).

**Proof.** Suppose that \( \bar{n} \) is the least integer for which \( P_n(u_1) \neq P_n(u_2) \). Then \( (w_1)^{\bar{n}+1} = (w_2)^{-\bar{n}+1} \), while \( (w_1)^{\bar{n}} = (w_2)^{\bar{n}} \neq (w_2)^{-\bar{n}} \). To conclude it is enough to observe that if \( w \) and \( v \) are subwords, \( w \neq v \) implies that the four words \( w^{-1}, \bar{w}^{-1}, v^{-1} \) and \( \bar{v}^{-1} \) are all distinct. \[ \square \]

Let us now start the inductive procedure described in the proof of Theorem 4 with the empty prefix \( u_0 = \epsilon \) and \( w_i := s_{2i-1} s_{2i} \) (\( i = 1, 2 \ldots \)). Suppose that \( \bar{n} \) is the least integer for which \( u_{\bar{n}} \neq \epsilon \) (we recall that, by Lemma 8, if \( u_{\bar{n}} = \epsilon \) for
every positive integer $n$, then every subword of $S$ is recurrent). It follows from the construction of Theorem 4 that it has to be $u_{n+k} ∈ \bar{h}-N$ for every positive integer $k$. By direct inspection it can be seen that $\bar{n}$ is not smaller than 2, as the prefix $s_1 s_2 \ldots s_{16} = 1221121221221121$ is 2-regular. Therefore, according to Theorem 4 we have that, for every positive integer $k$,

$$S = u_k w_1 w_2 \ldots$$

(40)

with $u_k ∈ \bar{h}-N (h ≥ 2)$ and $w_i ∈ k-R$ for every $i$. Since $u_k$ is 1-regular, we have that, writing $S$ as

$$S = u_k^{-2}(w_1)^{-2}(w_2)^{-2} \ldots$$

(41)

the subrows $(w_i)^{-2}$ all begin with 12. Therefore integrating $k - 2$ times (41) and suitably defining the subrows $\bar{v}_i$, we obtain for $S$ the structure

$$S = \bar{v}_0(z_1)^{-k+2} \bar{v}_1(z_2)^{-k+2} \ldots$$

(42)

were $z_i = 12$ for every integer $i$, and the subrows $(z_i)^{-k+2}$ are all coinciding as subwords since, for every $j ≤ k$, the parity of $|u_k w_1 w_2 \ldots w_n|^{-j}$ is the same for every $n > 0$. Finally, since $k$ is arbitrarily large by Theorem 4, (5) implies that the recurrent subrows $(z_i)^{-k+2}$ are arbitrarily long.

We can also use the same iterative procedure to identify other arbitrarily long recurrent subrows which, in general, will be not coinciding with the previous ones. Indeed, recalling Remark 6 we can also apply the iterative construction starting right after any given prefix of $S$. Taking, for instance, the 1-normal prefix $p := 1221$, for every positive integer $k$ we can write

$$S = p\tilde{u}_k \tilde{w}_1 \tilde{w}_2 \ldots$$

(43)

where, noticing that $s_5 \ldots s_8 ∈ 2-R$, we have $\tilde{u}_k ∈ \bar{h}-N (\bar{h} ≥ 2)$ and $\tilde{w}_i ∈ k-R$ for every $i$. Since $p\tilde{u}_k$ is 1-regular by Lemma 3 we have that, writing $S$ as

$$S = (p\tilde{u}_k)^{-2}(\tilde{w}_1)^{-2}(\tilde{w}_2)^{-2} \ldots$$

(44)

the subrows $(\tilde{w}_i)^{-2}$ all begin with 12. Therefore, integrating $k - 2$ times and suitably defining the subrows $\bar{v}_i$, we obtain for $S$ the structure

$$S = \bar{v}_0(z_1)^{-k+2} \bar{v}_1(z_2)^{-k+2} \ldots$$

(45)

were $\bar{z}_i = 12$ for every integer $i$, and the subrows $(\bar{z}_i)^{-k+2}$ are all coinciding as subrows since, for every $j ≤ k$, the parity of $|(p\tilde{u}_k \tilde{w}_1 \tilde{w}_2 \ldots \tilde{w}_n)^{-j}|$ is the same for every $n > 0$. Since by construction $P_n(p\tilde{u}_k) ≠ P_n(p\tilde{u}_k)$, Lemma 13 ensures that $(\bar{z}_i)^{-k+2} ≠ (\bar{z}_i)^{-k+2}$, while again since $k$ can be arbitrarily large (from Theorem 4), the recurrent subrows $(\bar{z}_i)^{-k+2}$ have arbitrarily large length by 5.

6 More open questions

How do $k$-regular prefixes (or, in general, subrows) look? This is a difficult question. Let us take a look at the first cases. A prefix $w = s_1 \ldots s_n$ is

- 0-regular if $n$ is even;
• 1-regular if it is 0-regular and $\Sigma w$ is even;
• 2-regular if it is 1-regular and $\Sigma_{i=0}^{2i-1} s_{2i+1}$ is even;
• 3-regular if it is 2-regular and
\[
\left| \{s_j : s_j = 2 \text{ and } j \text{ is odd} \} \right| + \left| \{s_j : s_j = 1, j \text{ is odd and } \Sigma_{i=1}^{j-1} s_i \text{ is even} \} \right|
\]
is even.

With some effort one can go a bit further, but it is not easy to see where the thing is going. Since the number of conditions that a finite word has to satisfy to be $k$-regular seems to increase with $k$, the following conjecture arises naturally.

**Conjecture 1.** There are no $\infty$–regular subrows in $SR(S)$.

A consequence of this conjecture is seen in Theorem 2. It is also natural, in our view, to formulate a stronger conjecture, namely that two subrows whose integrals have exactly the same history of parity, must coincide. More precisely, we state the following

**Conjecture 2.** If $u$ and $w$ are two subrows and $P_n(u) = P_n(w)$ for every $n \geq 0$, then $u = w$.

If this is true, then Conjecture 1 follows, as if $w$ is an $\infty$-regular subrow, we can split it in two subrows with the same history of parity, which contradicts Conjecture 2.

### 7 Acknowledgments

I am deeply grateful to Lucio Russo (who introduced me to the problem), Stefano Isola and Riccardo Piergallini for many fruitful discussions.

### References

[1] Srecko Brlek and A Ladouceur. A note on differentiable palindromes. *Theoretical Computer Science*, 302(1-3):167–178, 2003.

[2] Arturo Carpi. On repeated factors in C-infinity words. *Information Processing Letters*, 52(6):289–294, 1994.

[3] Vašek Chvátal. Notes on the Kolakoski sequence. *Rapport, DIMACS Techn. Rep.*, 1994.

[4] K Culik, J Karhumäki, and A Lepistö. Alternating iteration of morphisms and the kolakovski sequence. In *Lindenmayer Systems*, pages 93–106. Springer, 1992.

[5] Karel Culik and Juhani Karhumäki. Iterative devices generating infinite words. In *Annual Symposium on Theoretical Aspects of Computer Science*, pages 529–543. Springer, 1992.
[6] Frederik Michel Dekking. What is the long range order in the Kolakoski sequence? NATO ASI Series C Mathematical and Physical Sciences-Advanced Study Institute, 489:115–126, 1997.

[7] Volodymyr I Fesenko, Vladimir R Tuz, and Igor A Sukhoivanov. Terahertz aperiodic multilayered structure arranged according to the Kolakoski sequence. In Terahertz and Mid Infrared Radiation: Detection of Explosives and CBRN (Using Terahertz), pages 25–32. Springer, 2014.

[8] Abdallah Hammam. Some new formulas for the Kolakoski Sequence A000002. Turkish Journal of Analysis and Number Theory, 4(3):54–59, 2016.

[9] Michael S Keane et al. Ergodic theory and subshifts of finite type. Ergodic theory, symbolic dynamics and hyperbolic spaces, pages 57–66, 1991.

[10] William Kolakoski. Problem 5304: self generating runs. Amer. Math. Monthly, 72(6):674, 1965.

[11] Rufus Oldenburger. Exponent trajectories in symbolic dynamics. Transactions of the American Mathematical Society, 46(3):453–466, 1939.

[12] Bobby Shen. The Kolakoski sequence and related conjectures about orbits. Experimental Mathematics, 2018.

[13] Bernd Sing. Kolakoski sequences–an example of aperiodic order. Journal of non-crystalline solids, 334:100–104, 2004.

[14] Bernd Sing. More Kolakoski sequences. arXiv preprint arXiv:1009.4061, 2010.

[15] Bertran Steinsky. A recursive formula for the Kolakoski sequence A000002. J. Integer Sequences, 9(3):06–3, 2006.

[16] V Tuz, V Fesenko, and I Sukhoivanov. Optical characterization of the aperiodic multilayered anisotropic structure based on Kolakoski sequence. Integrated optics: physics and simulations, 8781:87811C, 2013.