Vertical Representation of $C^\infty$-words

Jean-Marc Fédou\textsuperscript{a}, Gabriele Fici\textsuperscript{a,∗}

\textsuperscript{a}Laboratoire d’Informatique, Signaux et Systèmes de Sophia-Antipolis
CNRS \& Université Nice Sophia Antipolis
2000, Route des Lucioles - 06903 Sophia Antipolis cedex, France

Abstract

We present a new framework for studying $C^\infty$-words, that are arbitrarily many times differentiable words over the alphabet $\Sigma = \{1, 2\}$. After introducing an equivalence relation on $C^\infty$-words, whose classes are called minimal classes and represent all the $C^\infty$-words, we define a vertical coding of these words based on a three-letter alphabet, and a set of functions operating over this representation. We show that the minimal classes of $C^\infty$-words can be represented on an infinite directed acyclic graph, which, as well as all the functions introduced for the vertical coding, can be defined recursively with no explicit reference to $C^\infty$-words. This new representation adequately expresses the combinatorial structure of $C^\infty$-words, and brings new perspectives in the study of the Kolakoski word and its factors.

Keywords: Kolakoski word, $C^\infty$-words, directed acyclic graph, recursive function.

1. Introduction

The Kolakoski word \cite{Kim} is the unique right infinite word $K$ over the alphabet $\Sigma = \{1, 2\}$ starting with 2 and coinciding with its run-length encoding\textsuperscript{5}:

$$K = 2 2 1 1 2 2 1 2 2 1 1 2 1 2 2 1 1 2 1 2 2 1 2 2 1 1 2 1 2 2 1 1 2 1 ...$$

This mysterious self-generating word is far to be well understood, and several longstanding conjectures on its structure remain unproved. Kimberling \cite{Kim} asked whether the Kolakoski word is recurrent and whether the set of its factors is closed under complement (swapping of 1’s and 2’s). Dekking \cite{Dek} observed that the latter condition implies the former, and introduced an operator on finite words, called the derivative, that consists in discarding the first and/or the last run if these have length 1 and then applying the run-length encoding. For example, the derivative of 2122 is 12. The set of words that are derivable arbitrarily many times over $\Sigma$, denoted by $C^\infty$, is then closed under complement and reversal, and contains the set of factors of the Kolakoski word. Therefore, one of the most important open problems about the Kolakoski word is to decide whether all the words in $C^\infty$ occur as factors in the Kolakoski word:

Conjecture 1.1. $\text{Fact}(K) = C^\infty$.

Actually, the set $C^\infty$ contains the set of factors of any right-infinite word over $\Sigma$ having the property that an arbitrary number of applications of the run-length encoding still produces a word over $\Sigma$. Such

\textsuperscript{5}Some of the results in this paper were presented at the Conference for the 20th Anniversary of LaCIM, 29-31 August 2010, Montreal, Canada \textsuperscript{5}.

\textsuperscript{∗}Corresponding author.

\textit{Email addresses: fedou@i3s.unice.fr} (Jean-Marc Fédou), \textit{fici@i3s.unice.fr} (Gabriele Fici)

\textsuperscript{1}The run-length encoding is the operator that counts the lengths of the maximal blocks of consecutive identical symbols in a string.
words are called smooth words [3, 1]. Nevertheless, the existence of a smooth word such that the set of its factors is equal to the whole set $C^\infty$ is still an open question.

Another renowned open problem is to decide whether the Kolakoski word (or any other smooth word) is recurrent (every factor appears infinitely often) or even uniformly recurrent (consecutive occurrences of the same factor appear with bounded gap). Should Conjecture 1.1 be true, the Kolakoski word would be recurrent [6].

In addition to the aforementioned problems, there is a conjecture of Keane [9] stating that the frequencies of 1’s and 2’s in the Kolakoski sequence exist and are equal to 1/2. Chvátal [5] showed that if these limits exist, they are very close to 1/2. Indeed, Carpi [4] proved that for every $C\infty$ of 1’s and 2’s in the Kolakoski sequence exist and are equal to 1/2. (see also [12] and [2]).

Up to now, only few combinatorial properties of $C^\infty$-words have been established. Weakley [13] started a classification of $C^\infty$-words and obtained significant results on their complexity function. Carpi [2] proved that the set $C^\infty$ contains only a finite number of squares, and does not contain cubes (see also [12] and [2]). This result generalizes to repetitions with gap, i.e., to the words in $C^\infty$ of the form $uzu$, for a non-empty $z$. Indeed, Carpi [4] proved that for every $k > 0$, only finitely many $C^\infty$-words of the form $uzu$ exist with $z$ not longer than $k$. In a recent paper [2], we proved that for any $u \in C^\infty$, there exists a $z$ such that $uzu \in C^\infty$, and $|uzu| \leq C|u|^{72}$, for a suitable constant $C$. In the same paper, we proposed the following conjecture:

**Universal Glueing Conjecture.** For any $u, v \in C^\infty$, there exists $z$ such that $uzu \in C^\infty$.

Despite the Universal Glueing Conjecture being a weaker assumption than Conjecture 1.1, it remains an open question. Its validity would imply, among other things, that for any integer $n > 0$, there exists a $C^\infty$-word containing as factors all the $C^\infty$-words of length $n$.

Let $w$ be a $C^\infty$-word. Recall that any word $v$ such that $w$ is the derivative of $v$ is called a primitive of $w$. For example, the primitives of the word 21 are 2212, 1121, 12212 and 21121. The $C^\infty$-words having the property that each derivative is obtained from the shortest primitive are called minimal words. Analogously, $C^\infty$-words having the property that each derivative is obtained from the longest primitive are called maximal words. Moreover, a $C^\infty$-word is single-rooted if its last non-empty derivative is a word of length one (that is, 1 or 2), or double-rooted if its last non-empty derivative is a word of length two (that is, 12 or 21).

In this paper, we mostly focus on single-rooted minimal words. Indeed, we show in Theorem 2.3 that the first single-rooted minimal factor of height $k$ appearing in a $C^\infty$-word $w$ is called the minimal part of $w$. Thus, every $C^\infty$-word is an extension of its minimal part preserving the height. This allows us to consider classes of $C^\infty$-words having the same minimal part (that we call minimal classes).

We recall a framework for dealing with $C^\infty$-words that we introduced in a recent paper [7], based on a three-letter alphabet. We define, for every $C^\infty$-word $w$, two words $U_0$ and $V_0$ over the alphabet $\Sigma_0 = \{0, 1, 2\}$, respectively the left frontier and the right frontier of $w$, which code respectively the sequences of the first and the last symbols of the derivatives of $w$. The pair $(U_0, V_0)$ uniquely determines $w$, and is called the vertical representation of $w$ (Theorem 3.1).

Minimal words do not have 0’s in their vertical representation, that is, are coded by left and right frontiers over the alphabet $\Sigma = \{1, 2\}$. As a consequence, single-rooted minimal words (and hence minimal classes) are in bijection with $\Sigma^\ast$.

We then define the functions $\Gamma_u$ and $\Gamma_d$, which map the left frontier of a $C^\infty$-word into its right frontier. More precisely, for any $U \in \Sigma^\ast$, $\Gamma_u(U)$ is the word $V \in \Sigma^\ast$ such that $(U, V)$ is the vertical representation of the single-rooted minimal word having $U$ as left frontier, whereas $\Gamma_d(U)$ is the word $V' \in \Sigma^\ast$ such that $(U, V')$ is the vertical representation of the double-rooted minimal word having $U$ as left frontier. The functions $\Gamma_u$ and $\Gamma_d$ are idempotent and therefore establish bijections of $\Sigma^\ast$. We also define the compositions $\Theta = \Gamma_u \circ \Gamma_d$ and $\Pi = \Gamma_d \circ \Gamma_u$, for which we are able to find a very compact recursive form (Corollary 3.3), that can be defined independently from the context of $C^\infty$-words (Theorem 4.4). The functions $\Gamma_u$, $\Gamma_d$, and therefore $\Theta$ and $\Pi$, can be naturally extended to words with 0’s, that is, words coding the frontiers of any $C^\infty$-word.

We then show that any single-rooted minimal word $u$ having left frontier $U \in \Sigma^\ast$ and height $k > 0$ has three extensions to the right of minimal length having height $k + 1$: the two single-rooted minimal words

2
Remark 1. Since \( \omega \) runs identical symbols (called encoding over \( \Sigma \) is denoted by \( \Sigma^* \)) of a word \( w \) is a prefix of a suffix of \( w \). Let \( w \) be a word over \( \Sigma \). Then \( w \in \Sigma^* \). The set of all finite words over \( \Sigma \) having length \( n \) is denoted by \( \Sigma^n \). The \( i+1 \)-th symbol of a word \( w \) is denoted by \( w[i] \). So, we write a word \( w \) of length \( n > 0 \) as \( w = w[0]w[1] \cdots w[n-1] \).

Let \( w \in \Sigma^* \). If \( w = uv \) for some \( u, v \in \Sigma^* \), we say that \( u \) is a prefix of \( w \) and \( v \) is a suffix of \( w \). A factor of \( w \) is a prefix of a suffix of \( w \) (or, equivalently, a suffix of a prefix). The reversal of \( w \) is the word \( \overline{w} \) obtained by writing the letters of \( w \) in the reverse order. For example, the reversal of \( w = 11212121 \) is \( \overline{w} = 21211212 \). The complement of \( w \) is the word \( \overline{\overline{w}} \) obtained by swapping the letters of \( w \), i.e., by changing the 1’s in 2’s and the 2’s in 1’s. For example, the complement of \( w = 11212121 \) is \( \overline{\overline{w}} = 22122122 \).

A right-infinite word over \( \Sigma \) is a non-ending sequence of letters from \( \Sigma \). The set of all right-infinite words over \( \Sigma \) is denoted by \( \Sigma^\omega \). Let \( w \) be a word over \( \Sigma \). Then \( w \) can be uniquely written as a concatenation of maximal blocks of identical symbols (called runs), i.e., \( w = x_1^{i_1}x_2^{i_2} \cdots x_n^{i_n} \), with \( x_j \in \Sigma, x_j \neq x_{j+1}, \text{ and } i_j > 0 \). The run-length encoding of \( w \), noted \( \Delta(w) \), is the sequence of exponents \( i_j \), i.e., one has \( \Delta(w) = i_1i_2 \cdots i_n \). The run-length encoding extends naturally to right-infinite words.

Definition 1. \( \square \) A right infinite word \( w \in \Sigma^\omega \) is called a smooth word over \( \Sigma \) if for every integer \( k > 0 \) one has that \( \Delta^k(w) \) is still a word over \( \Sigma \).

The operator \( \Delta \) on right infinite words over \( \Sigma \) has two fixed points, namely the Kolakoski word

\[
K = 2211212211221211121221211121212211212112212211212112121121212121 \ldots
\]

and the word \( 1K \).

In this paper, we focus on the set \( \Sigma^\omega \) of factors of smooth words. We start by recalling some definitions.

Definition 2. \( \square \) A word \( w \in \Sigma^* \) is differentiable if \( \Delta(w) \) is still a word over \( \Sigma \).

Remark 1. Since \( \Sigma = \{1,2\} \) we have that \( w \) is differentiable if and only if neither 111 nor 222 appear in \( w \).
Definition 3. [6] The derivative is the function \( D : \Sigma^* \to \Sigma^* \) defined on the differentiable words by:

\[
D(w) = \begin{cases} 
\varepsilon & \text{if } \Delta(w) = 1 \text{ or } w = \varepsilon, \\
\Delta(w) & \text{if } \Delta(w) = 2x2 \text{ or } \Delta(w) = 2, \\
x2 & \text{if } \Delta(w) = 1x2, \\
2x & \text{if } \Delta(w) = 2x1, \\
x & \text{if } \Delta(w) = 1x1.
\end{cases}
\]

In other words, the derivative \( D(w) \) is obtained from \( \Delta(w) \) by erasing the first and/or the last symbol if they are equal to 1.

Let \( k \geq 0 \). A word \( w \) is \( k \)-differentiable on \( \Sigma \) if \( D^k(w) \) is defined. By Remark 1, a word \( w \) is \( k \)-differentiable if and only if for every \( 0 \leq j < k \) the word \( D^j(w) \) does not contain 111 nor 222 as factor. We use the convention that \( D^0(w) = w \). Clearly, if a word is \( k \)-differentiable, then it is also \( j \)-differentiable for every \( 0 \leq j \leq k \).

We denote by \( C^k \) the set of \( k \)-differentiable words, and by \( C^\infty \) the set of words which are differentiable arbitrarily many times. Therefore, \( C^\infty = \bigcap_{k \geq 0} C^k \). A word in \( C^\infty \) is also called a \( C^\infty \)-word [6].

As a direct consequence of the definition, we have that the set \( C^\infty \) and, for any \( k \geq 0 \), the set \( C^k \), are closed under reversal and complement.

Definition 4. [6] The height of a \( C^\infty \)-word is the least integer \( k \) such that \( D^k(w) = \varepsilon \).

Definition 5. [7] Let \( w \in C^\infty \) of height \( k > 0 \). The root of \( w \) is \( D^{k-1}(w) \). Therefore, the root of \( w \) belongs to \( \{1, 2, 12, 21\} \). Consequently, \( w \) is single-rooted if its root has length one or double-rooted if its root has length two.

Definition 6. [6] A primitive of a word \( w \) is any word \( w' \) such that \( D(w') = w \).

It is easy to see that any \( C^\infty \)-word has two, four or eight distinct primitives (actually, it has two primitives if it starts and ends with 1, eight primitives if it starts and ends with 2, and four primitive else). For example, the word \( w = 22 \) has eight primitives (1122, 21122, 11221, 211221, 2211, 12211, 22112, 122112), whereas the word \( w = 121 \) has only two primitives (121121, 212212).

However, every \( C^\infty \)-words admits exactly two primitives of minimal (maximal) length, one being the complement of the other.

Definition 7. [7] Let \( w \) be a \( C^\infty \)-word of height \( k > 1 \). We say that \( w \) is minimal (resp. maximal) if for every \( 0 \leq j < k-2 \), \( D^j(w) \) is a primitive of \( D^{j+1}(w) \) of minimal (resp. maximal) length. The words of height \( k = 1 \) are assumed to be at the same time minimal and maximal.

We can define minimal and maximal words on one side only, in the following way.

Definition 8. [7] A word \( w \in C^\infty \) is left minimal (resp. left maximal) if it is a prefix of a minimal (resp. maximal) word. Analogously, \( w \) is right minimal (resp. right maximal) if it is a suffix of a minimal (resp. maximal) word.

Clearly, a word is minimal (resp. maximal) if and only if it is both left minimal and right minimal (resp. left maximal and right maximal).

Example 1. The word 2211 is minimal, since 2211 is a primitive of 22 of minimal length and 22 is a primitive of 2 of minimal length; the word 22121211 is maximal, since 22121211 is a primitive of 2 of maximal length and 2211 is a primitive of 2 of maximal length; the word 2122112 is left maximal but not right maximal. Note that 2211 is a proper factor of 21221121 and that the two words have the same height and the same root.

| \( w \)     | 2211  | 21221121 | 21221121 |
|------------|-------|----------|----------|
| \( D(w) \) | 2     | 122      | 1221     |
| \( D^2(w) \)| 2     | 2        | 2        |
Weakley started a classifications of $C^\infty$-words based on the extendability. Indeed, any $C^\infty$-word has arbitrary long left and right extensions in $C^\infty$. Indeed, for any $C^\infty$-word $w$ at least one between $1w$ and $2w$ is a $C^\infty$-word. Analogously, at least one between $w2$ and $2w$ is a $C^\infty$-word.

**Definition 9.** [13] A word $w \in C^\infty$ is left doubly extendable (resp. right doubly extendable) if $1w$ and $2w$ (resp. $w1$ and $w2$) are both in $C^\infty$. Otherwise, $w$ is left simply extendable (resp. right simply extendable).

A word $w \in C^\infty$ is fully extendable if $1w1$, $1w2$, $2w1$ and $2w2$ are all in $C^\infty$.

It is worth noticing that a word can be at the same time left doubly extendable and right doubly extendable but not fully extendable. This is the case, for example, for the word $w = 1$.

Note that, by definition, any $C^\infty$-word $w$ can be extended to the left and to the right with simple extensions in a unique way, up to a left-and-right doubly extendable (but not necessarily fully extendable) word $w'$.

Based on the previous definitions and on a result of Weakley ([13], Proposition 3) one can establish the following structural result.

**Theorem 2.1.** [7] Let $w \in C^\infty$. The following conditions are equivalent:

1. $w$ is fully extendable (resp. $w$ is left doubly extendable, resp. $w$ is right doubly extendable).
2. $w$ is double-rooted maximal (resp. $w$ is left maximal, resp. $w$ is right maximal).
3. $w$ and all its derivatives (resp. $w$ and all its derivatives longer than one) begin and end (resp. begin, resp. end) with two distinct symbols.

**Example 2.** Consider the word $w = 121$. By Theorem [2,7] $w$ is left doubly extendable and right doubly extendable. Nevertheless, $w$ is not fully extendable, since it is single-rooted. Indeed, the word $2w2$ is not in $C^\infty$.

We report here a lemma given in [7] with a slightly different statement. Basically, it states that extending to the right by one letter a right maximal word (resp. extending to the left by one letter a left minimal word) results in a right minimal word (resp. a left minimal word). Recall that, by Theorem 2.1 a right (resp. left) maximal word is right (resp. left) doubly extendable.

**Lemma 2.2.** [7] Let $w \in C^\infty$ be a right maximal word (resp. a left maximal word). Then the words $w1$ and $w2$ are right minimal words (resp. $1w$ and $2w$ are left minimal words).

Conversely, if $wx$, $x \in \Sigma$, is a right minimal word (resp. $xw$ is a left minimal word), then $w$ is a right maximal word (resp. $w$ is a left maximal word).

Let us summarize the previous results. Let $v$ be a $C^\infty$-word of height $k > 0$. Then we know that $v$ can be extended, to the left and to the right, with simple extensions (therefore, in a unique way) up to reaching a word $\hat{v}$, which is left doubly extendable and right doubly extendable (so, $\hat{v}$ is a maximal word). Moreover, $\hat{v}$ has the same root and the same height as $v$. Now, two substantially different cases arise: if $v$ is double-rooted, then so is $\hat{v}$, and therefore $\hat{v}$ is fully extendable, that is, the four words $1\hat{v}1$, $1\hat{v}2$, $2\hat{v}1$ and $2\hat{v}2$ are all $C^\infty$-words. In particular, there is a choice of $x, y \in \Sigma$ such that the words $x\hat{v}y$ and $\hat{x}\hat{v}\hat{y}$ are minimal words of height $k + 2$ and root $2$, and the words $x\hat{v}\hat{y}$ and $\hat{x}\hat{v}y$ are double-rooted minimal words of height $k + 1$. If instead $v$ is single-rooted, then so is $\hat{v}$, and therefore $\hat{v}$ is not fully extendable. In fact, only three among the four words $1\hat{v}1$, $1\hat{v}2$, $2\hat{v}1$ and $2\hat{v}2$ are $C^\infty$-words. In particular, there is a choice of $x, y \in \Sigma$ such that the word $x\hat{v}y$ is a minimal word of height $k + 1$ and root $1$. The words $x\hat{v}\hat{y}$ and $\hat{x}\hat{v}y$ are instead words of height $k + 1$ and root $2$, but they are minimal on one side only, while the word $\hat{x}\hat{v}\hat{y}$ is not a $C^\infty$-word.

**Example 3.** The word $v = 112212$ is a single-rooted minimal word of height 3. It extends to the single-rooted maximal word (of height 3) $\hat{v} = 12112212$. Table 1 shows the extensions of $\hat{v}$.

**Example 4.** The word $v = 1122122$ is a double-rooted minimal word of height 3. It extends to the double-rooted maximal word (of height 3) $\hat{v} = 1211221221$. Table 2 shows the extensions of $\hat{v}$.

---

2 It is in fact a minimal forbidden word for $C^\infty$. 

### Table 1: Left and right extensions of the single-rooted word 1122.

|   | $\hat{v}$ | $x\hat{v}$ | $\hat{v}y$ | $\hat{v}\hat{y}$ |
|---|---------|------------|----------|--------------|
| $D_0^1(\hat{v})$ | 1122 | 112112212 | 212112212 | 121122122 |
| $D_1^1(\hat{v})$ | 22 | 1212 | 2121 | 1221 |
| $D_2^1(\hat{v})$ | 2 | 12 | 22 | 22 |
| $D_3^1(\hat{v})$ | 2 | 12 | 22 | 22 |

|   | $\hat{v}$ | $x\hat{v}$ | $\hat{v}y$ | $\hat{v}\hat{y}$ |
|---|---------|------------|----------|--------------|
| $D_0^2(\hat{v})$ | 1121122122 | 2121122122 | 1121122122 | 2121122122 |
| $D_1^2(\hat{v})$ | 2212 | 212212 | 212212 | 112212 |
| $D_2^2(\hat{v})$ | 121 | 221 | 222 | 222 |
| $D_3^2(\hat{v})$ | 1 | 2 | 2 | 2 |

### Table 2: Left and right extensions of the double-rooted word 112212.

|   | $\hat{v}$ | $x\hat{v}$ | $\hat{v}y$ | $\hat{v}\hat{y}$ |
|---|---------|------------|----------|--------------|
| $D_0^3(\hat{v})$ | 112212 | 121212212 | 1211221221 | 1211221221 |
| $D_1^3(\hat{v})$ | 2212 | 221212 | 221212 | 2212 |
| $D_2^3(\hat{v})$ | 121 | 221 | 221 | 221 |
| $D_3^3(\hat{v})$ | 1 | 2 | 2 | 2 |

|   | $\hat{v}$ | $x\hat{v}$ | $\hat{v}y$ | $\hat{v}\hat{y}$ |
|---|---------|------------|----------|--------------|
| $D_0^4(\hat{v})$ | 112112212211 | 212112212211 | 112112212211 | 212112212211 |
| $D_1^4(\hat{v})$ | 221212 | 221212 | 221212 | 221212 |
| $D_2^4(\hat{v})$ | 121 | 221 | 221 | 221 |
| $D_3^4(\hat{v})$ | 1 | 2 | 2 | 2 |

Table 2: Left and right extensions of the double-rooted word 112212.
One of the aims of this paper is to show that single-rooted minimal words are all one needs for dealing with $C^\infty$-words.

**Theorem 2.3.** Let $w \in C^\infty$ of height $k > 0$. If $w$ is single-rooted, then $w$ contains exactly one single-rooted minimal factor $u$ of height $k$. If $w$ is double rooted, then $w$ contains exactly two single-rooted minimal factors $u$ and $u'$ of height $k$.

Theorem 2.3 allows us to state the following definition, which is fundamental for the rest of the paper.

**Definition 10.** Let $w \in C^\infty$ of height $k > 0$. The first single-rooted minimal factor of height $k$ appearing in $w$ is called the *minimal part* of $w$.

**Example 5.** Let $w = 21221211221221121$, which is a double-rooted word of height 4. Then, $w$ contains two single-rooted minimal factors of height 4: $u = 212122$ and $u' = 112212211$. The word $u$ is the minimal part of $w$.

**Corollary 2.4.** Any $C^\infty$-word of height $k > 0$ can be obtained from its minimal part by extensions, to the left and to the right, preserving the height $k$.

The definition of minimal part induces an equivalence relation on the set of $C^\infty$-words (defined by the property of having the same minimal part). We call a class of $C^\infty$-words w.r.t. this equivalence a *minimal class*. As a direct consequence of Theorem 2.3, there is a one-to-one correspondence between minimal classes and single-rooted minimal words. We will show that there is also a one-to-one correspondence between the set of single-rooted minimal words and the set $\Sigma^*$. Consequently, there are exactly $2^k$ single-rooted minimal words of height $k$ (and so $2^k$ minimal classes of height $k$).

Weakley [13] conjectured that for every $k \geq 0$, any double-rooted maximal word (that is, by Theorem 2.1, any fully extendable word) of height $k$ is shorter than any double-rooted maximal word of height $k + 1$. Weakley’s conjecture can also be rephrased in terms of minimal words, i.e. it is equivalent to the following

**Conjecture 2.5.** Any minimal word of root 2 and height $k$ is shorter than any minimal word of root 2 and height $k + 1$.

Indeed, any minimal word of root 2 and height $k > 2$ is of the form $xwy$, with $x, y \in \Sigma$ and $w$ is a double-rooted maximal word of height $k - 2$.

One can wonder whether Weakley’s conjecture holds true when double-rooted minimal words are replaced with single-rooted maximal words, i.e., whether any single-rooted maximal word of height $k$ is shorter than any single-rooted maximal word of height $k + 1$. Note that any minimal word of root 1 and height $k > 1$ is of the form $xwy$, with $x, y \in \Sigma$ and $w$ is a single-rooted maximal word of height $k - 1$. Therefore, one can equivalently wonder whether any minimal word of root 1 and height $k$ is shorter than any minimal word of root 1 and height $k + 1$. However, the answer to the above question is negative. For example, there exist minimal words $w, w'$ of root 1 and height, respectively, 19 and 20, such that $|w| = 3858$ and $|w'| = 3851$.

It is worth noticing that Weakley’s conjecture is also equivalent to the following.
Conjecture 2.6. Any double-rooted minimal word of height \( k \) is shorter than any double-rooted minimal word of height \( k + 1 \).

Indeed, as a consequence of Lemma 2.2 \( xw, x \in \Sigma \), is a minimal word of root 2 and height \( k + 1 \) if and only if \( Tw \) is a double-rooted minimal word of height \( k \).

3. Vertical representation of \( C^\infty \)-words

We recall here the definition of vertical representation of a \( C^\infty \)-word \( [7] \).

We define a function \( \Psi \) for representing a \( C^\infty \)-word on a three-letter alphabet \( \Sigma_0 = \{0, 1, 2\} \). This function is a generalization of the function \( \Phi \) considered in \([3]\), that associates to any \( C^\infty \)-word \( w = w[0]w[1]...w[n-1] \) the sequence of the first symbols of the derivatives of \( w \), that is, the function defined by \( \Phi(w)[i] = D^i(w)[0] \) for \( 0 \leq i < k \), where \( k \) is the height of \( w \).

If one takes the first and the last symbol of each derivatives of a \( C^\infty \)-word \( w \), that is, the pair \( \Phi(w), \Phi(\tilde{w}) \), one gets a representation of \( C^\infty \)-words that is not injective. For example, take the two \( C^\infty \)-words \( w = 2211 \) and \( w' = 21121221 \). Then one has \( \Phi(w) = \Phi(w') = 222 \) and \( \Phi(\tilde{w}) = \Phi(\tilde{w'}) = 122 \). In order to obtain an injective representation, we need an extra symbol. We thus introduce the following definition.

We set \( \Sigma \) = \( \{0, 1, 2\} \). We also set

\[ \Sigma_0^+ = \{ U \in \Sigma_0^* : U = \varepsilon \text{ or the first symbol of } w \text{ is different from } 0 \} \]

Clearly, \( \Sigma^* \subset \Sigma_0^+ \).

Definition 11. Let \( w = w[0]w[1]...w[n-1] \) be a \( C^\infty \)-word of height \( k > 0 \). The left frontier of \( w \) is the word \( \Psi(w) \in \Sigma_0^+ \) of length \( k \) defined by: \( \Psi(w)[0] = w[0] \) and for \( 0 < i < k \)

\[ \Psi(w)[i] = \begin{cases} 0 & \text{if } D^i(w)[0] = 2 \text{ and } D^{i-1}(w)[0] \neq D^{i-1}(w)[1], \\ D^i(w)[0] & \text{otherwise.} \end{cases} \]

For the empty word, we set \( \Psi(\varepsilon) = \varepsilon \).

The right frontier of \( w \) is defined as \( \Psi(\tilde{w}) \). If \( U \) and \( V \) are respectively the left and right frontier of \( w \), we call \( U|V \) the vertical representation of \( w \).

In other words, to obtain the left (resp. the right) frontier of \( w \), one has to take the first (resp. the last) symbol of each derivative of \( w \) and replace a 2 by a 0 whenever the primitive above is not left minimal (resp. is not right minimal).

Example 6. Let \( w = 21221211221 \). We have:

\[
\begin{array}{c|c}
D^0(w) & 21221211221 \\
D^1(w) & 1221122 \\
D^2(w) & 122 \\
D^3(w) & 2 \\
\end{array}
\]

The word \( D^2(w) = 122 \) is not a left minimal primitive of the word \( D^3(w) = 2 \), and therefore \( \Psi(w)[3] \), the fourth symbol of the left frontier of \( w \), is a 0; analogously, the word \( w = 21221211221 \) is not a right minimal primitive of \( D(w) = 121122 \), and therefore \( \Psi(\tilde{w})[1] \), the second symbol of the right frontier of \( w \), is a 0. Hence, the vertical representation of \( w \) is \( \Psi(w)|\Psi(\tilde{w}) = 21101022 \).

The following theorem is a direct consequence of the definition of vertical representation.
Theorem 3.1. Any $C^\infty$-word is uniquely determined by its vertical representation; that is, the map defined on $C^\infty$ by

$$w \mapsto (\Psi(w), \Psi(\overline{w}))$$

is injective.

Remark 2. In what follows, uppercase letters ($U, V, W, \ldots$) will denote vertical words, that are words over $\Sigma_0^+$ coding the (left or right) frontier of a $C^\infty$-word; lowercase letters ($u, v, w, \ldots$) will still denote $C^\infty$-words.

Remark 3. If $w$ is a minimal word, then, by definition, the left and the right frontier of $w$ are vertical words belonging to $\Sigma^*$. That is, they do not contain 0. In particular, if $w$ is a single-rooted (or a double-rooted) minimal word, then the left frontier univocally determines the right frontier and vice versa.

For minimal words, the knowledge of one frontier is sufficient to reconstruct the word. For example, let $U = U[0]U[1]\ldots U[k-1] \in \Sigma^k$. Then, the unique single-rooted $C^\infty$-word having $U$ as left frontier is the word $w$ such that $D^i(w)$ is the shortest primitive of $D^{i+1}(w)$ which begins with the symbol $U[i]$, for every $0 \leq i < k$. Analogously, one can construct the unique double-rooted minimal word having $U$ as left frontier.

Hence, we can state the following

Proposition 3.2. There is a one-to-one correspondence between the set of single-rooted minimal words and $\Sigma^*$. This correspondence is given by the map $\Psi$.

Therefore, there are exactly $2^k$ single-rooted minimal words of height $k$.

Analogously, the map $\Psi$ gives also a one-to-one correspondence between the set of double-rooted minimal words and $\Sigma^*$, and therefore there are exactly $2^k$ double-rooted minimal words of height $k$.

Clearly, if a smooth word contains all the single-rooted minimal words of height $k+1$ as factors, then it will also contain as factors all the $C^\infty$-words of height $k$, by Theorem 2.3. So, in order to prove that a smooth word $W$ over $\Sigma$ contains as factors all the single-rooted minimal words,

we end this section with a conjecture.

Conjecture 3.3. There exists a linear function $f(k)$ such that any $C^\infty$-word of height $f(k)$ contains as factors all the single-rooted minimal words of height $k$.

Should Conjecture 3.3 be true, any smooth word over $\Sigma$ would contain all the $C^\infty$-words as factors and would be uniformly recurrent. In particular, this would be true for the Kolakoski word.

4. Functions on the frontiers

Let $U \in \Sigma^*$. Then, by Proposition 3.2, $U$ is the left frontier of a unique single-rooted minimal word $w$. We denote by $\Gamma_s(U)$ the right frontier of $w$. Analogously, $U$ is the left frontier of a unique double-rooted minimal word $w'$. We denote by $\Gamma_d(U)$ the right frontier of $w'$.

Example 7. Let $U = 2122$. Then $\Gamma_s(U) = 2222$ and $\Gamma_d(U) = 1221$. The situation is depicted in Fig. 2.

Clearly, $\Gamma_s^2(U) = \Gamma_2^2(U) = U$ for any $U \in \Sigma^*$.

Lemma 4.1. For any $U \in \Sigma^*$ one has $\overline{\Gamma_d(U)} = \Gamma_d(\overline{U})$.

We also define the compositions $\Theta = \Gamma_s \circ \Gamma_d$, and

$\Pi = \Gamma_d \circ \Gamma_s$.

Therefore, $\Theta(\Pi(U)) = \Pi(\Theta(U)) = U$ for any $U \in \Sigma^*$. 9
The functions $\Pi$ and $\Theta$ allow us to describe, from the point of view of the vertical representation, how single-rooted minimal words can be extended, respectively to the left and to the right, into double-rooted minimal words of the same height. Let $w \in C^\infty$ be a double-rooted minimal word and let $U \in \Sigma^*$ be its left frontier. Then $w$ can be viewed as the overlap between two single rooted minimal words $u$ and $u'$. The left frontiers of $u$ and $u'$ are then $U$ and $\Theta(U)$ respectively. Clearly, $\Pi$ is the inverse of $\Theta$.

**Example 8.** Let $U = 2122$. Then $\Theta(U) = 1221$ and $\Pi(U) = 1121$. The situation is depicted in Fig. 3.

**Remark 4.** Let $k \geq 0$. Then $\Gamma_s$, $\Gamma_d$, $\Theta$ and $\Pi$ are bijections of $\Sigma^k$. Moreover, $\Gamma_s$ and $\Gamma_d$ are involutory\(^3\).

We now show that the functions $\Gamma_s$, $\Gamma_d$, and therefore $\Pi$ and $\Theta$, can be defined recursively and independently from the context of $C^\infty$-words.

We clearly have

$$\Gamma_s(\varepsilon) = \Gamma_d(\varepsilon) = \Pi(\varepsilon) = \Theta(\varepsilon) = \varepsilon.$$  

**Lemma 4.2.** Let $U \in \Sigma^*$. Then:

1. $\Gamma_s(U) = \Gamma_d \circ \Gamma_s \circ \Gamma_d(U)$
2. $\Gamma_s(U) = \Gamma_d(U)$
3. $\Gamma_d(U) = \Gamma_d \circ \Gamma_s \circ \Gamma_d \circ \Gamma_s \circ \Gamma_d(U)$
4. $\Gamma_d(U) = \Gamma_d(U)$

The lemma can be proved by simple observations on the compositions of the functions $\Gamma_s$ and $\Gamma_d$ (see Fig. 4 for an example).

From the recursive formulae for $\Gamma_s$ and $\Gamma_d$ we can derive recursive formulae for $\Pi$ and $\Theta$.

---

\(^3\)A map is involutory if the composition with itself is the identity map.
Corollary 4.3. Let $U \in \Sigma^*$. Then:

1. $\Pi(U1) = \Pi(U)2$
2. $\Pi(U2) = \Pi^2(U)1$
3. $\Theta(U1) = \Theta^2(U)2$
4. $\Theta(U2) = \Theta(\overline{U})1$

Note that the function $\Pi$ (or, analogously, the function $\Theta$) can be defined recursively by the formulae given in Corollary 4.3 with no explicit reference to $C^\infty$-words. Once the function $\Pi$ has been so defined, one can derive a recursive definition of $\Gamma_s$ and $\Gamma_d$. This allows us to give a very compact description of single-rooted minimal words, presented in the following theorem.

Theorem 4.4. Let $U \in \Sigma^*$. Then:

1. $\Pi(U1) = \Pi(U)2$
2. $\Pi(U2) = \Pi^2(U)1$
3. $\Gamma_s(UX) = \Pi(\Gamma_s(U)X)$, for any $X \in \Sigma$
4. $\Gamma_d(U) = \Pi(\Gamma_s(U))$

The theorem is a consequence of the recursive definition of $\Pi$ given in Corollary 4.3 and of simple arguments on the compositions of the functions over the frontiers (see Fig. 5).

Theorem 4.4 together with Theorem 2.3 gives an essential representation of $C^\infty$-words. Most of the algorithmic operations on $C^\infty$-words can be implemented by means of the formulae given in Theorem 4.4. This reduces significantly the space needed for storing a $C^\infty$-word, since the ratio between the height of a $C^\infty$-word and its length is logarithmic.

The functions $\Gamma_s$, $\Gamma_d$, $\Theta$, and $\Pi$ can be easily extended to $\Sigma_{++}^+$, as described below.

Let $U_0 \in \Sigma_0^+$ be a non-empty word, and let $u$ be the shortest single-rooted $C^\infty$-word having $U_0$ as left frontier (hence $u$ is a right minimal word). We define the function $\Gamma_s : \Sigma_0^+ \rightarrow \Sigma^*$ by

$$\Gamma_s(U_0) = \Psi(\overline{u}),$$

that is, $\Gamma_s(U_0)$ is the right frontier of $u$.

Let $v$ be the shortest double-rooted $C^\infty$-word having $U_0$ as left frontier (hence $v$ is a right minimal word). We define the function $\Gamma_d : \Sigma_0^+ \rightarrow \Sigma^*$ by

$$\Gamma_d(U_0) = \Psi(\overline{v}),$$

that is, $\Gamma_d(U_0)$ is the right frontier of $v$.

Now we can define $\Theta : \Sigma_0^+ \rightarrow \Sigma^*$ and $\Pi : \Sigma_0^+ \rightarrow \Sigma^*$ by setting

$$\Theta = \Gamma_s \circ \Gamma_d,$$

and

$$\Pi = \Gamma_d \circ \Gamma_s.$$
For any word $U_0 \in \Sigma_0^+$, one has $(\Pi \circ \Theta)(U_0) = (\Theta \circ \Pi)(U_0) = \Gamma_d^2(U_0) = \Gamma_s^2(U_0) = U \in \Sigma^*$. So $\Pi$ and $\Theta$ are not bijections on $\Sigma_0^{++}$ (neither are $\Gamma_d^2$ and $\Gamma_s^2$). They are instead projections of $\Sigma_0^{++}$ on $\Sigma^*$.

Recursive formulae for $\Gamma_s$, $\Gamma_d$, $\Pi$ and $\Theta$ as functions on words over $\Sigma_0^{++}$ are derived below.

We clearly have

$$\Gamma_s(\varepsilon) = \Gamma_d(\varepsilon) = \Pi(\varepsilon) = \Theta(\varepsilon) = \varepsilon.$$

**Lemma 4.5.** Let $U_0 \in \Sigma_0^+$. Then:

1. $\Gamma_s(U_01) = \Gamma_d \circ \Gamma_s \circ \Gamma_d(U_0)1$
2. $\Gamma_s(U_02) = \Gamma_d(U_0)2$
3. $\Gamma_d(U_00) = \Gamma_d \circ \Gamma_s \circ \Gamma_d(U_0)2$
4. $\Gamma_d(U_01) = \Gamma_d \circ \Gamma_s \circ \Gamma_d \circ \Gamma_s \circ \Gamma_d(U_0)2$
5. $\Gamma_d(U_02) = \Gamma_d \circ \Gamma_s \circ \Gamma_d \circ \Gamma_s \circ \Gamma_d(U_0)1$
6. $\Gamma_d(U_00) = (\Gamma_d \circ \Gamma_s \circ \Gamma_d \circ \Gamma_s \circ \Gamma_d(U_0))1$
7. $\Pi(U_01) = \Pi(U_0)2$
8. $\Pi(U_02) = \Pi(U_0)2$
9. $\Pi(U_00) = \Pi(U_0)1$
10. $\Theta(U_01) = \Theta(U_0)2$
11. $\Theta(U_02) = \Theta(U_02)1 = \Gamma_d(\Gamma_s(U_0))1$
12. $\Theta(U_00) = \Theta(U_00)1$

**5. Right extensions**

In this section, we describe how a $C^\infty$-word of height $k$ can be extended to the right into a $C^\infty$-word of height $k+1$. We do this from the point of view of the vertical representation. Indeed, an extension to the right of a $C^\infty$-words eventually results into an extension of one letter of its left frontier when the height of the word becomes $k+1$. We restrict our attention to single-rooted minimal words. An analogous approach can be applied to the extensions on the left and the right frontier.

Let $u$ be a single-rooted minimal word with left frontier $U \in \Sigma^*$, and let $k > 0$ be the height of $u$, that is the length of $U$.

What are the shortest $C^\infty$-words of height $k+1$ having $u$ as a prefix? These are exactly the shortest $C^\infty$-words of height $k+1$ whose left frontiers have $U$ as a prefix. That is, they are the shortest $C^\infty$-words having left frontier, respectively, $U1$, $U2$ and $U0$. We call these three words, respectively, the right $1$-extension, right $2$-extension and the right $0$-extension of the word $u$. 

---

Figure 5: The recursive definition of $\Gamma_s$ by means of $\Pi$ in Theorem 4.4: $\Gamma_s(U) = \Pi(\Gamma_s(U)\overline{X})$, for any $X \in \Sigma$. Recall that $\Pi = \Gamma_d \circ \Gamma_s$.
Note that the right 1-extension and the right 2-extension are single-rooted minimal words, whereas the right 0-extension is not minimal, since its left frontier does not belong to $\Sigma^*$. So, what is the single-rooted minimal word corresponding to the right 0-extension of $u$? That is, what is the minimal part of the right 0-extension of $u$? The following theorem provides the answer to this question.

**Theorem 5.1.** Let $u$ be a single-rooted minimal word of height $k > 0$, and $U \in \Sigma^k$ its left frontier. The word $u$ can be extended to the right into two single-rooted minimal words of height $k+1$: the right 1-extension of $u$, which is the single-rooted minimal word having left frontier $U1$, and the right 2-extension of $u$, which is the single-rooted minimal word having left frontier $U2$; and into a single-rooted (not minimal) word having left frontier $U0$, called the right 0-extension of $u$.

Moreover, the minimal part of the right 0-extension of $u$ is the single-rooted minimal word having left frontier $\Theta(U)2$.

**Remark 6.** The right 0-extension and the right 1-extension of a word $u$ are words that differ only on the last letter.

**Example 9.** Let $u = 2121122$, having left frontier $U = 2122$. The right 2-extension of $u$ is the word 2121221212212 having left frontier $U2$; the right 1-extension of $u$ is the word 21211221212 having left frontier $U1$; the right 0-extension of $u$ is the word 21211221211 having left frontier $U0$, whose minimal part is the word 112212211211 having left frontier $\Theta(U)2$. The situation is depicted in Fig. 6.

The left 1-extension, the left 2-extension and the left 0-extension are defined analogously.

We can now define an infinite directed acyclic graph for representing the minimal classes of $C^\infty$-words accordingly with the three right extensions (see Fig. 7).

The graph of $C^\infty$-words is the graph $G$ is defined by
where \( \mathcal{V} = \Sigma^* \) and the set \( \mathcal{E} \) of labeled edges is partitioned into three subsets:

- \( \mathcal{E}_1 = \{(U,1,U1), \text{ solid edges}\} \)
- \( \mathcal{E}_2 = \{(U,2,U2), \text{ solid edges}\} \)
- \( \mathcal{E}_0 = \{((U),0,\Theta(U)2), \text{ dashed edges}\} \)

Thus, \( G \) is obtained by adding to an infinite complete binary tree (with edge labels in \( \Sigma \) and node labels in \( \Sigma^* \)) one additional edge outgoing from each node (labeled by 0).

A partial diagram of the graph \( G \) is depicted in Fig. 9. The edges in \( \mathcal{E}_0 \) are dashed.

Figure 8: The minimal classes of the right extensions with \( \Pi \).

Clearly, one can define the graph \( G \) by means of \( \Pi \) instead of \( \Theta \) (see Fig. 8), by defining

\[
G' = (\mathcal{V}, \mathcal{E}'),
\]

where \( \mathcal{V} = \Sigma^* \) and the set \( \mathcal{E}' \) of labeled edges is partitioned into three subsets:

- \( \mathcal{E}_1' = \{(U,1,U1), \text{ solid edges}\} \)
- \( \mathcal{E}_2' = \{(U,2,U2), \text{ solid edges}\} \)
- \( \mathcal{E}_0' = \{((\Pi(U)),0,U2), \text{ dashed edges}\} \)

Since \( \Theta = \Pi^{-1} \) over \( \Sigma^* \), one has \( G = G' \).

Remark 7. As a consequence of Theorem 4.4, the graph \( G \) can be defined recursively and with no explicit reference to \( C^\infty \)-words.

Let \( w \) be a \( C^\infty \)-word and \((U_0,V_0)\) its vertical representation, \( U_0, V_0 \in \Sigma_0^\infty \). Let \( u \) be the minimal part of \( w \) and \((U,\Gamma_s(U))\) the vertical representation of \( u, U \in \Sigma^* \). If \( w \) is double-rooted, then \( w \) contains a double-rooted minimal factor \( u' \) having vertical representation \((U,\Gamma_d(U))\), see Fig. 4.

By Theorem 2.3 the word \( w \) is a simple extension on the left of \( u \), and it is a simple extension on the right of \( u \) (if \( w \) is single-rooted) or of \( u' \) (if \( w \) is double-rooted).

The left frontier \( U_0 \) of \( w \) is the label of a path in \( G \) starting at the origin and ending in \( U \). Analogously, the right frontier \( V_0 \) of \( w \) is the label of a path in \( G \) starting at the origin and ending in \( \Gamma_s(U) \) (if \( w \) is single-rooted) or in \( \Gamma_d(U) \) (if \( w \) is double-rooted).
Figure 9: The graph $G$ cut at height 4.
Conversely, any word $U_0 \in \Sigma_0^+$ labeling a path in $G$ starting at the origin and ending in $U \in \Sigma^*$ is the left frontier of a simple extension on the left of any $C^\infty$-word having $U$ as left frontier, and, symmetrically it is also the right frontier of a simple extension on the right of any $C^\infty$-word having $U$ as right frontier. An example is given in Fig. 10.

Thus, we have the following characterization of the minimal classes of $C^\infty$ words by means of the graph $G$.

**Theorem 5.2.** Let $U \in \Sigma^*$ be the left frontier of a single-rooted minimal word $u$. The $C^\infty$-words having $u$ as minimal part are exactly the words $w$ having vertical representation $(U_0, V_0)$ such that $U_0$ labels a path in $G$ starting at the origin and ending in $U$, and $V_0$ labels a path in $G$ starting at the origin and ending in $\Gamma_s(U)$ (if $w$ is single-rooted) or $\Gamma_d(U)$ (if $w$ is double-rooted).

In particular, the number of distinct paths in $G$ starting at the origin and ending in $U \in \Sigma^*$ is equal to the number of simple extensions on the left (resp. on the right) of a single-rooted minimal word having $U$ as left (resp. right) frontier.

This allows us to show that there exists a close relation between the paths in $G$ and the length of the single-rooted minimal words.

Let $|U|$ denote the length of the single-rooted minimal word $u$ having left frontier $U$ and $\|U\|$ the number of distinct paths in $G$ starting at the origin and ending in $U$.

From the definition of the graph $G$ (see Fig. 5), one has $||U_1|| = |U|\|$ and $||U_2|| = |U| + |\Pi(U)|$.

Since $||U||$ is the number of single-extensions to the left of $u$, we have that the length of the left 2-extension of $u$ (that is, the single-rooted minimal word having left frontier $\Pi(U)2$) is equal to $|U| + ||U||$ (see Fig. 6). Symmetrically, the length of the right 2-extension of $u$ (which is the single-rooted minimal word having left frontier $U2$) is equal to $|\Gamma_s(U)| + ||\Gamma_s(U)||$. Analogous considerations hold for the (left and right) 1-extensions. Note also that one has $|U| = |\Gamma_s(U)|$, $|\Theta(U)| = |\Gamma_d(U)|$, $|U_2| = |U_2|$ and $|U_1| = |U_0|$.

We therefore deduce the following recursive formulae:

1. $|U_2| = |U| + |\Gamma_s(U)|| = |\Theta(U)| + |\Theta(U)||$
2. $|U_1| = |\Theta(U)| + |\Theta(U)| = |\Gamma_d(U)|$

We think that these formulae show the interest of dealing with the graph $G$ when considering problems on the $C^\infty$-words. For example, analogous formulae can be applied to compute the number of 1’s and 2’s...
in a $C^\infty$-word, thus giving a possible direction to investigate Keane’s conjecture on the frequency of letters in the Kolakoski word.

6. Conclusion and open problems

The vertical representation is a compact representation of $C^\infty$-words that allows one to represent any $C^\infty$-word of length $n$ by means of two words whose length is logarithmic in $n$ (the frontiers). In this paper, we showed that this representation can be defined in terms of two simple recursive functions, that are naturally represented by a directed acyclic infinite graph $G$. The recursive definitions of the functions on the frontiers leads to a recursive definition of the graph $G$, therefore independent from the context of $C^\infty$-words.

Besides being more compact, we believe that the new representation presented here will allow the use of results from graph theory or poset theory in the study of the Kolakoski word. As an illustration, we formulate below two new conjectures on the graph $G$ which, if proven, would imply the validity of important conjectures on $C^\infty$-words.

- We conjecture that for any $U_1, U_2 \in \Sigma^*$ it is possible to find two words $V_1, V_2 \in \Sigma^*_0$ such that $U_1V_1$ and $U_2V_2$ label two paths in $G$ starting at the origin and ending in the same node (see Fig. 11). This would imply that for any $u, v \in C^\infty$ there exists $z$ such that $uzv \in C^\infty$ (see Fig. 12). This latter conjecture was referred to as the Universal Glueing Conjecture [7].
We also conjecture that there exists a linear function \( f(k) \) such that for every word \( Z \in \Sigma^{f(k)} \) and every word \( U \in \Sigma^k \), there exists a word \( V \in \Sigma_0^+ \) such that \( UV \) labels a path in \( G \) starting at the origin and ending in \( Z \). This would imply that any smooth word is uniformly recurrent. In particular, this would be true for the Kolakoski word.

The key step underlying the construction presented here is the recognition of the fundamental role of minimal words (and associated classes) in the structure of \( C^\infty \)-words. The simplicity of the representation of \( C^\infty \)-words in terms of their vertical representation strongly suggests that these are at the heart of the structure of the entire set. This is supported by the fact that, quite surprisingly, also the study of the densities of 1’s and 2’s in a \( C^\infty \)-word can be carried out using combinatorial properties of the graph \( G \).

7. Acknowledgments

The authors acknowledge a financial support from the French National Research Agency project EMC (ANR-09-BLAN-0164).

References

[1] V. Berthé, S. Brlek, and P. Choquette. Smooth words over arbitrary alphabets. Theoret. Comput. Sci., 341(1):293–310, 2005.
[2] S. Brlek, S. Dulucq, A. Ladouceur, and L. Vuillon. Combinatorial properties of smooth infinite words. Theoret. Comput. Sci., 352(1):306–317, 2006.
[3] S. Brlek and A. Ladouceur. A note on differentiable palindromes. Theoret. Comput. Sci., 302(1-3):167–178, 2003.
[4] A. Carpi. On repeated factors in \( C^\infty \)-words. Inform. Process. Lett., 52:289–294, 1994.
[5] V. Chvátal. Notes on the Kolakoski sequence. Technical Report 93-84, DIMACS, March 1993. Revised.
[6] F. M. Dekking. On the structure of self-generating sequences. Séminaires de théorie des nombres de Bordeaux, Exposé 31, 1980-81.
[7] J. M. Fédou and G. Fici. Automata and differentiable words. Theoretical Computer Science, to appear. Preprint available at http://arxiv.org/abs/1102.0913.
[8] Jean-Marc Fédou and Gabriele Fici. Vertical representation of smooth factors. In Conference for the 20th anniversary of LaCIM. Montreal, Canada, August 29–31, 2010. Local proceedings, 10 pages, 2010.
[9] M. S. Keane. Ergodic theory and subshifts of finite type. In T. Bedford, M. Keane, and C. Series, editors, Ergodic Theory, Symbolic Dynamics, and Hyperbolic Spaces, pages 35–70. Oxford University Press, 1991.
[10] C. Kimberling. Advanced problem 6281. Amer. Math. Monthly, 86:793, 1979.
[11] W. Kolakoski. Elementary problem 5304. Amer. Math. Monthly, 72:674, 1965. Solution, 73 (1966), 681–682.
[12] A. Lepistö. Repetitions in Kolakoski sequence. In G. Rozenberg and A. Salomaa, editors, Developments in Language Theory, pages 130–143. World Scientific, 1994.
[13] W. D. Weakley. On the number of \( C^\infty \)-words of each length. J. Combin. Theory, Ser. A, 51:55–62, 1989.