

Infinite Smooth Lyndon Words

Geneviève Paquin

Laboratoire de Combinatoire et d'Informatique Mathématique,
Université du Québec à Montréal,
Montréal (QC) CANADA H3C 3P8,
genevieve.paquin@gmail.com

Abstract. In a recent paper, Brlek et al. showed that some extremal infinite smooth words are also infinite Lyndon words. This result raises a natural question: what are the infinite smooth words that are also infinite Lyndon words? In this paper, we give the answer: the only infinite smooth Lyndon words are $m_{\{a < b\}}$, with a, b even, and $m_{\{1 < b\}}$, with b odd, where $m_{\mathcal{A}}$ is the minimal infinite smooth word with respect to lexicographic order over the numerical alphabet \mathcal{A} .

Keywords: Lyndon words, smooth words, Kolakoski sequence

1 Introduction

Lyndon words were introduced by Lyndon in [9] for constructing bases of the lower central series for free groups. The authors proved that any finite word can be expressed as a unique non-increasing product of Lyndon words. Later, Lyndon words were studied by Duval [11,12]. He gave an algorithm that generates Lyndon words of bounded length for a finite alphabet and another one that computes the Lyndon factorization in linear time. Siromoney et al [26] defined infinite Lyndon words in order to introduce Lyndon factorization of infinite words. Lyndon words also appeared in [18,20,22]. This factorization gives nice properties about the structure of words. Since a few years, a wide literature is devoted to Lyndon words: [2,13,23,24,25]. For instance, Melançon [19] studied Lyndon factorization of Sturmian infinite words.

Smooth infinite words over $\mathcal{A} = \{1, 2\}$ form an infinite class \mathcal{K} of infinite words containing the well-known Kolakoski word K [17] defined as one of the two fixed points of the run-length encoding function Δ , that is

$$\Delta(K) = K = 2211212212211211221211212211211212212211212212 \dots$$

They are characterized by the property that the orbit obtained by iterating Δ is contained in $\{1, 2\}^*$. In the early work of Dekking [10], there are some challenging conjectures on the structure of K that still remain unsolved despite the efforts devoted to the study of patterns in K . For instance, we know from Carpi [8] that K and more generally, any word in the infinite class \mathcal{K} of smooth words over $\mathcal{A} = \{1, 2\}$, contain only a finite number of squares, implying by direct inspection that K and any $w \in \mathcal{K}$ are cube-free. Weakley [17] showed that the number of factors of length n of \mathcal{K} is polynomially bounded. In [6], a connection was established between the palindromic complexity and the recurrence of K . Then, Berthé et al. [3] studied smooth words over arbitrary alphabets and obtained a new characterization of the infinite Fibonacci word F . Relevant work may also be found in [1] and in [3,16], where generalized Kolakoski words are studied for arbitrary alphabets. The authors investigated in [7] the extremal infinite smooth words, that is the minimal and the

maximal ones w.r.t. the lexicographic order, over $\{1, 2\}$ and $\{1, 3\}$: a surprising link is established between F and the minimal infinite smooth word over $\{1, 3\}$.

More recently, Brlek et al. [5] studied the extremal smooth words for any 2-letter alphabet and they showed the existence of infinite smooth words that are also Lyndon words: the minimal smooth word over an even alphabet and the one over the alphabet $\{1, b\}$, with b odd, are Lyndon words. Then a natural question arises: are there other infinite smooth words that are infinite Lyndon words?

In this paper, we show that the minimal smooth words that are also Lyndon words given in [5] are the only smooth Lyndon words. In order to prove it, we study the words over a 2-letter alphabet depending on the parity of the letters. The paper is organized as follow. In Section 2, we recall the basic definitions in combinatorics on words, we state the notation we will use next and we recall useful known results. Section 3 is devoted to the characterization of infinite smooth Lyndon words. It is divided in 4 subsections. In Section 3.1, we study the case of an alphabet $\mathcal{A} = \{a < b\}$, with a even and b odd. We show that there is no infinite Lyndon words that is also smooth. In Section 3.2, we are interested in even alphabets. We show that only the minimal smooth word is a Lyndon word. Section 3.3 is devoted to odd alphabet. We prove that only $m_{\{1,b\}}$ is a Lyndon word. Finally, Section 3.4 studies the words over an alphabet $\{a < b\}$ with a odd and b even. In this last case, we show that there is no infinite Lyndon words that are also smooth.

Notice that some proofs are omitted for lack of space and will appear in a full paper.

2 Preliminaries

Throughout this paper, \mathcal{A} is a finite *alphabet* of *letters* equipped with a total order $<$. A *finite word* w is a finite sequence of letters $w = w[0]w[1] \cdots w[n-1]$, where $w[i] \in \mathcal{A}$ denotes its $(i+1)$ -th letter. Its length is n and we write $|w| = n$. The set of n -length words over \mathcal{A} is denoted by \mathcal{A}^n . By convention the *empty* word is denoted by ε and its length is 0. The free monoid generated by \mathcal{A} is defined by $\mathcal{A}^* = \bigcup_{n \geq 0} \mathcal{A}^n$ and $\mathcal{A}^* \setminus \varepsilon$ is denoted \mathcal{A}^+ . The set of right infinite words, also called infinite words for short, is denoted by \mathcal{A}^ω and $\mathcal{A}^\infty = \mathcal{A}^* \cup \mathcal{A}^\omega$. Adopting a consistent notation for finite words over the infinite alphabet \mathbb{N} , $\mathbb{N}^* = \bigcup_{n \geq 0} \mathbb{N}^n$ is the set of finite sequences and \mathbb{N}^ω is that of infinite ones. Given a word $w \in \mathcal{A}^*$, a *factor* f of w is a word $f \in \mathcal{A}^*$ satisfying

$$\exists x, y \in \mathcal{A}^*, w = xfy.$$

If $x = \varepsilon$ (resp. $y = \varepsilon$) then f is called a *prefix* (resp. *suffix*). Note that by convention, the empty word is suffix and prefix of any word. A *block* of length k is a maximal factor of the particular form $f = \alpha^k$, with $\alpha \in \mathcal{A}$. The set of all factors of w , also called the *language* of w , is denoted by $F(w)$, and those of length n is $F_n(w) = F(w) \cap \mathcal{A}^n$. We denote by $\text{Pref}(w)$ (resp. $\text{Suff}(w)$) the set of all prefixes (resp. suffixes) of w .

Over an arbitrary 2-letter alphabet $\mathcal{A} = \{a, b\}$, there is a usual length preserving morphism, the *complementation*, defined by $\bar{a} = b$, $\bar{b} = a$, which extends to words as follows. The complement of $u = u[0]u[1] \cdots u[n-1] \in \mathcal{A}^n$ is the word $\bar{u} = \bar{u}[0] \bar{u}[1] \cdots \bar{u}[n-1]$. The *reversal* of u is the word $\tilde{u} = u[n-1] \cdots u[1]u[0]$.

For $u, v \in \mathcal{A}^*$, we write $u < v$ if and only if u is a proper prefix of v or if there exists an integer k such that $u[i] = v[i]$ for $0 \leq i \leq k-1$ and $u[k] < v[k]$. The relation \leq defined by $u \leq v$ if and only if $u = v$ or $u < v$, is called the *lexicographic order*.

That definition holds for \mathcal{A}^∞ . Note that in general the complementation does not preserve the lexicographic order. Indeed, when u is not a proper prefix of v then

$$u > v \iff \bar{u} < \bar{v}. \quad (1)$$

A word $u \in \mathcal{A}^*$ is a *Lyndon word* if $u < v$ for all proper non-empty suffixes v of u . For instance, the word 11212 is a Lyndon word while 12112 is not since $112 < 12112$. A word of length 1 is clearly a Lyndon word. The set of Lyndon words is denoted by \mathcal{L} .

From Lothaire [18], we have the following theorem.

Theorem 1. [9] *Any non empty finite word w is uniquely expressed as a non increasing product of Lyndon words*

$$w = \ell_0 \ell_1 \cdots \ell_n = \bigodot_{i=0}^n \ell_i, \text{ with } \ell_i \in \mathcal{L}, \text{ and } \ell_0 \geq \ell_1 \geq \cdots \geq \ell_n. \quad (2)$$

Siromoney et al. [26] extended Theorem 1 to infinite words. The set \mathcal{L}_∞ of *infinite Lyndon words* consists of infinite words smaller than any of their suffixes.

Theorem 2. [26] *Any infinite word w is uniquely expressed as a non increasing product of Lyndon words, finite or infinite, in one of the two following forms:*

- i) *either there exists an infinite sequence $(\ell_k)_{k \geq 0}$ of elements in \mathcal{L} such that*

$$w = \ell_0 \ell_1 \ell_2 \cdots \text{ and for all } k, \ell_k \geq \ell_{k+1}.$$
- ii) *there exist a finite sequence ℓ_0, \dots, ℓ_m ($m \geq 0$) of elements in \mathcal{L} and $\ell_{m+1} \in \mathcal{L}_\infty$ such that*

$$w = \ell_0 \ell_1 \cdots \ell_m \ell_{m+1} \text{ and } \ell_0 \geq \cdots \geq \ell_m > \ell_{m+1}.$$

Let us recall from ([18] Chapter 5.1) a useful property concerning Lyndon words.

Lemma 3. *Let $u, v \in \mathcal{L}$. We have $uv \in \mathcal{L}$ if and only if $u < v$.*

A direct corollary of this lemma is:

Corollary 4. *Let $u, v \in \mathcal{L}$, with $u < v$. Then $w^n, u^n v \in \mathcal{L}$, for all $n \geq 0$.*

The widely known *run-length encoding* is used in many applications as a method for compressing data. For instance, the first step in the algorithm used for compressing the data transmitted by Fax machines consists of a run-length encoding of each line of pixels. Let $\mathcal{A} = \{a < b\}$ be an ordered alphabet. Then every word $w \in \mathcal{A}^*$ can be uniquely written as a product of factors as follows:

$$w = a^{i_0} b^{i_1} a^{i_2} \cdots \quad \text{or} \quad w = b^{i_0} a^{i_1} b^{i_2} \cdots$$

with $i_k \geq 1$ for $k \geq 0$. The operator giving the size of the blocks appearing in the coding is a function $\Delta : \mathcal{A}^* \longrightarrow \mathbb{N}^*$, defined by $\Delta(w) = i_0, i_1, i_2, \dots$ which is easily extended to infinite words as $\Delta : \mathcal{A}^\omega \longrightarrow \mathbb{N}^\omega$.

For instance, let $\mathcal{A} = \{1, 3\}$ and $w = 13333133111$. Then

$$w = 1^1 3^4 1^1 3^2 1^3 \quad \text{and} \quad \Delta(w) = [1, 4, 1, 2, 3].$$

When $\Delta(w) \subseteq \{1, 2, \dots, 9\}^*$, the punctuation and the parentheses are often omitted in order to manipulate the more compact notation $\Delta(w) = 14123$. This example is a

special case where the coding integers do not coincide with the alphabet on which is encoded w , so that Δ can be viewed as a partial function $\Delta : \{1, 3\}^* \longrightarrow \{1, 2, 3, 4\}^*$.

From now on, we only consider 2-letter alphabets $\mathcal{A} = \{a < b\}$, with $a, b \in \mathbb{N} \setminus \{0\}$.

Recall from [6] that Δ is not bijective since $\Delta(w) = \Delta(\bar{w})$, but commutes with the reversal ($\widetilde{}$), is stable under complementation ($\bar{}$) and preserves palindromicity. Since Δ is not bijective, pseudo-inverse functions

$$\Delta_a^{-1}, \Delta_b^{-1} : \mathcal{A}^* \longrightarrow \mathcal{A}^*$$

are defined for 2-letter alphabets by

$$\Delta_\alpha^{-1}(u) = \alpha^{u[1]}\bar{\alpha}^{u[2]}\alpha^{u[3]}\bar{\alpha}^{u[4]}\dots, \quad \text{for } \alpha \in \{a, b\}.$$

Note that the pseudo-inverse function Δ^{-1} also commutes with the mirror image, that is,

$$\widetilde{\Delta_\alpha^{-1}(w)} = \Delta_\beta^{-1}(\widetilde{w}), \quad (3)$$

where $\beta = \alpha$ if $|w|$ odd and $\beta = \bar{\alpha}$ if $|w|$ is even.

The operator Δ may be iterated, provided the process is stopped when the coding alphabet changes or when the resulting word has length 1.

Example 5. Let $w = 1333111333133311133313133311133313331113331$. The successive application of Δ gives :

$$\begin{aligned} \Delta^0(w) &= 1333111333133311133313133311133313331113331; \\ \Delta^1(w) &= 1333133311133313331; \\ \Delta^2(w) &= 131333131; \\ \Delta^3(w) &= 1113111; \\ \Delta^4(w) &= \mathbf{313}; \\ \Delta^5(w) &= \mathbf{111}; \\ \Delta^6(w) &= \mathbf{3}. \end{aligned}$$

The set of *finite smooth words* over the alphabet \mathcal{A} is defined by

$$\Delta_{\mathcal{A}}^+ = \{w \in \mathcal{A}^* \mid \exists n \in \mathbb{N}, \Delta^n(w) = \alpha^i, \alpha \in \mathcal{A}, i \leq \beta \text{ and } \forall k \leq n, \Delta^k(w) \in \mathcal{A}^*\},$$

with β the greatest letter of the alphabet.

The operator Δ extends to infinite words (see [6]). Define the set of *infinite smooth words* over $\mathcal{A} = \{a, b\}$ by

$$\mathcal{K}_{\mathcal{A}} = \{w \in \mathcal{A}^\omega \mid \forall k \in \mathbb{N}, \Delta^k(w) \in \mathcal{A}^\omega\}.$$

The well-known [17] *Kolakoski word* denoted K is defined as the fix-point starting with the letter 2 of the operator Δ over the alphabet $\{1, 2\}$:

$$K = 22112122122112112212112211211212212211 \dots$$

More generally, the operator Δ has two fix-points in $\mathcal{K}_{\mathcal{A}}$, namely

$$\Delta(K_{(a,b)}) = K_{(a,b)} \quad \text{and} \quad \Delta(K_{(b,a)}) = K_{(b,a)},$$

where $K_{(a,b)}$ is the generalized Kolakoski word [16] over the alphabet $\{a, b\}$ starting with the letter a .

Example 6. The Kolakoski word over $\mathcal{A} = \{1, 2\}$ starting with the letter 2 is $K = K_{(2,1)}$. We also have $K_{(2,3)} = 22332223332233223332 \dots$ and $K_{(3,1)} = 33311133313133311133 \dots$.

A bijection $\Phi : \mathcal{K}_{\mathcal{A}} \longrightarrow \mathcal{A}^{\omega}$ is defined by

$$\Phi(w) = \Delta^0(w)[0]\Delta^1(w)[0]\Delta^2(w)[0] \dots = \prod_{i \geq 0} \Delta^i(w)[0]$$

and its inverse is defined as follows. Let $u \in \mathcal{A}^k$, then $\Phi^{-1}(u) = w_k$, where

$$w_n = \begin{cases} u[k-1], & \text{if } n = 1; \\ \Delta_{u[k-n]}^{-1}(w_{n-1}), & \text{if } 2 \leq n \leq k. \end{cases}$$

Then for $k = \infty$, $\Phi^{-1}(u) = \lim_{k \rightarrow \infty} w_k = \lim_{k \rightarrow \infty} \Phi^{-1}(u[0..k-1])$.

Remark 7. With respect to the usual topology defined by

$$d((u_n)_{n \geq 0}, (v_n)_{n \geq 0}) := 2^{-\min\{j \in \mathbb{N}, u_j \neq v_j\}},$$

the limit exists because each iteration is a prefix of the next one.

Example 8. For the word $w = 1333111333133311133313133311133313331113331$ of Example 5, $\Phi(w) = \mathbf{1111313}$.

Note that since Φ is a bijection, the set of infinite smooth words is infinite. Moreover, given a prefix of $\Phi(w)$, for w a smooth word, we can construct a prefix of w as in the following example.

Example 9. Let $p = 1221$ be a prefix of $\Phi(w)$, with $w \in \{1, 2\}^{\omega}$ an infinite smooth word. Then we compute from bottom to top, using the operator Δ^{-1} :

$$\Delta^0(w) = 11221221 \dots$$

$$\Delta^1(w) = 2212 \dots$$

$$\Delta^2(w) = 21 \dots$$

$$\Delta^3(w) = 1 \dots$$

Note that in $\Delta^2(w)$, the letter 1 is obtained by deduction, since $\Delta^3(w)$ indicates that the first block of letters of $\Delta^2(w)$ has length 1. The last written letter of every line is deduced by the same argument.

We recall from [7] the useful *right derivative* $D_r : \mathcal{A}^* \rightarrow \mathbb{N}^*$ defined by

$$D_r(w) = \begin{cases} \varepsilon & \text{if } \Delta(w) = \alpha, \alpha < b \text{ or } w = \varepsilon, \\ \Delta(w) & \text{if } \Delta(w) = xb, \\ x & \text{if } \Delta(w) = x\alpha, \alpha < b, \end{cases}$$

where $\alpha \in \mathbb{N}$ and $x \in \mathcal{A}^*$. A word w is *r-smooth* (also said *smooth prefix*) if $\forall k \geq 0$, $D_r^k(w) \in \mathcal{A}^*$. In other words, if a word w is *r-smooth*, then it is a prefix of at least one infinite smooth word (see [4] for more details).

Example 10. Let $w = 112112212$. Then $\Delta(w) = 212211$, $\Delta^2(w) = 1122$, $\Delta^3(w) = 22$ and $D_r(w) = 21221$, $D_r^2(w) = 112$, $D_r^3(w) = 2$.

Similarly, the operator D is defined over the alphabet $\{a < b\}$ by

$$D(w) = \begin{cases} \varepsilon & \text{if } \Delta(w) < b \text{ or } w = \varepsilon, \\ \Delta(w) & \text{if } \Delta(w) = bxb \text{ or } \Delta(w) = b, \\ bx & \text{if } \Delta(w) = bxu, \\ xb & \text{if } \Delta(w) = uxb, \\ x & \text{if } \Delta(w) = uxv, \end{cases}$$

where u and v are blocks of length $< b$. A finite word is called a *smooth factor* (also called a C^∞ -word in [8,14,15,27]) if there exists $k \in \mathbb{N}$ such that $D^k(w) = \varepsilon$ and $\forall j < k, D^j(w) \in \mathcal{A}^*$.

The *minimal* (resp. the *maximal*) *infinite smooth word* over the alphabet \mathcal{A} is the smallest (resp. biggest) infinite smooth word, with respect to the lexicographic order. It is denoted by $m_{\mathcal{A}}$ (resp. $M_{\mathcal{A}}$).

An alphabet $\mathcal{A} = \{a < b\}$ is called an *odd alphabet* (resp. *even alphabet*) if both a and b are odd (resp. even). The extremal smooth words satisfy the following properties established in a previous paper.

Proposition 11. [5] *Let $\mathcal{A} = \{a, b\}$ be a 2-letter alphabet with $a < b$. Then the following properties hold:*

- i) *If a and b are both even we have :*
 $\Phi(M_{\{a,b\}}) = b^\omega; \quad \Phi(m_{\{a,b\}}) = ab^\omega; \quad \text{and} \quad m_{\{a,b\}} \in \mathcal{L}_\infty.$
- ii) *If a and b are both odd we have :*
 $\Phi(M_{\{a,b\}}) = (ba)^\omega; \quad \Phi(m_{\{a,b\}}) = (ab)^\omega; \quad \text{and} \quad m_{\{a,b\}} \in \mathcal{L}_\infty \iff a = 1.$

Let us recall some known results about smooth words.

Lemma 12. [3] *Let u, v be finite smooth words. If there exists an index m such that, for all $i, 0 \leq i \leq m$, the last letter of $\Delta^i(u)$ differs from the first letter of $\Delta^i(v)$, and $\Delta^i(u) \neq 1, \Delta^i(v) \neq 1$, then*

- i) $\Phi(uv) = \Phi(u)[0..m] \cdot \Phi \circ \Delta^{m+1}(uv);$
- ii) $\Delta^i(uv) = \Delta^i(u)\Delta^i(v).$

The following properties follow immediately from the definitions. For more details, the reader is referred to [21].

Recall from [4] that in the case of the alphabet $\mathcal{A} = \{1, 2\}$, every finite word $w \in \Delta_{\mathcal{A}}^+$ can be easily extended to the right in a smooth word by means of the function Φ as follows:

$$\forall u \in \mathcal{A}^\infty, w \in \text{Pref}(\Phi^{-1}(\Phi(w) \cdot u)).$$

Its generalization to arbitrary alphabets is immediate (see [21]).

Proposition 13. *Let $\mathcal{A} = \{a < b\}$. Then the following properties hold.*

- i) *Any smooth prefix can be arbitrarily right extended to an infinite smooth word.*
- ii) *Let $u = \Phi(w)$, with $w \in \mathcal{A}^\omega$ an infinite smooth word. If $u = u'u''$, then $\Phi^{-1}(u')$ is prefix of w .*

3 Characterization of infinite smooth Lyndon words

In this section, we prove our main result: the only infinite smooth words that are also infinite Lyndon words are $m_{\{2a < 2b\}}$ and $m_{\{1 < 2b+1\}}$, for $a, b \in \mathbb{N}$. In order to prove it, we study the four possible combinations of the parity of the letters.

Let us consider all the possible words p of a fixed length $\leq n$ such that $\Phi^{-1}(p)$ is prefix of an infinite smooth word w . We suppose that w is also an infinite Lyndon word. In the following, for each word p , either we show that $\Phi^{-1}(p)$ can not be a prefix of a Lyndon word by showing the existence of a smaller suffix, or we describe an infinite smooth Lyndon word having $\Phi^{-1}(p)$ as prefix.

Lemma 14. $p[0] = a$.

Proof. Follows from the equality $p[0] = w[0]$ and since a Lyndon word w must start by the smallest letter. \square

Lemma 14 will be used in this section to exclude the cases numbered (0) in the proofs.

3.1 Over \mathcal{A} with a even and b odd

In this section, we prove the following result.

Theorem 15. *Over the alphabet $\{a < b\}$, with a even and b odd, there is no infinite smooth word that is also a Lyndon word.*

Proof. Figure 1 illustrates the 5 possible cases to consider, using a tree. The leaves correspond to the first letter of $\Phi(w)$ that leads to a contradiction: the prefix $\Phi^{-1}(p)$ obtained can not be the prefix of an infinite Lyndon word. We will prove it by showing that there exists a factor f of w not prefix of $\Phi^{-1}(p)$ such that $f < w$. For clarity issues, the first letter of f is underlined.

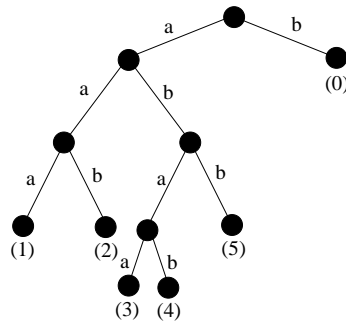


Figure 1. Possible cases for an even-odd alphabet

Case (1) If $p = aaa$, then
 $\Delta^0(w) = (a^a b^a)^{\frac{a}{2}} (\underline{a}^b b^b)^{\frac{a}{2}} \dots$
 $\Delta^1(w) = a^a b^a \dots$
 $\Delta^2(w) = aa \dots$

Since w has the prefix $a^a b^a$ and the factor $f = a^b$, it can not be a Lyndon word.

Remark 16. Since the smallest letter a of the alphabet is even, $p[3] \geq a \geq 2$. That allows us to assume that $\Delta^2(w)$ starts with a block of length at least 2. This argument holds for $\Delta^i(w)$, $i \geq 0$, and will be used for almost all cases considered in this paper.

Remark 17. In the previous case, we construct $\Delta^0(w)$ from $\Delta^2(w)$, applying Δ^{-1} twice. We will always proceed this way.

Case (2) If $p = aab$, then

$$\Delta^0(w) = (a^a b^a)^{\frac{b-1}{2}} a^a (b^b \underline{a}^b)^{\frac{b-1}{2}} b^b \dots$$

$$\Delta^1(w) = a^b b^b \dots$$

$$\Delta^2(w) = bb \dots$$

w has the factor $f = a^b$ smaller than its prefix $a^a b^a$.

Case (3) If $p = abaa$, then

$$\Delta^0(w) = ((a^b b^b)^{\frac{a}{2}} (a^a b^a)^{\frac{a}{2}})^{\frac{a}{2}} ((\underline{a}^b b^b)^{\frac{b-1}{2}} a^b (b^a a^a)^{\frac{b-1}{2}} b^a)^{\frac{a}{2}} \dots$$

$$\Delta^1(w) = (b^a a^a)^{\frac{a}{2}} (b^b a^b)^{\frac{a}{2}} \dots$$

$$\Delta^2(w) = a^a b^a \dots$$

$$\Delta^3(w) = aa \dots$$

w has the factor $f = (a^b b^b)^{\frac{b-1}{2}} a^b$.

Case (4) If $p = abab$, then

$$\Delta^0(w) = ((a^b b^b)^{\frac{a}{2}} (a^a b^a)^{\frac{a}{2}})^{\frac{b-1}{2}} (a^b b^b)^{\frac{a}{2}} ((a^a b^a)^{\frac{b-1}{2}} a^a (b^b \underline{a}^b)^{\frac{b-1}{2}} b^b)^{\frac{b-1}{2}} (a^a b^a)^{\frac{b-1}{2}} a^a \dots$$

$$\Delta^1(w) = (b^a a^a)^{\frac{b-1}{2}} b^a (a^b b^b)^{\frac{b-1}{2}} a^b \dots$$

$$\Delta^2(w) = a^b b^b \dots$$

$$\Delta^3(w) = bb \dots$$

w has the prefix $(a^b b^b)^{\frac{a}{2}} a^a b^a$ and the smaller factor $f = (a^b b^b)^{\frac{b-1}{2}}$ contained in $(b^b a^b)^{\frac{b-1}{2}} b^b$.

Remark 18. Since $b > a$ and a is even, $b > a \geq 2$. Thus, $b \geq 3$ and $\frac{b-1}{2} \geq 1$. This insures that the factor $(b^b a^b)^{\frac{b-1}{2}} b^b$ occurs at least once.

Case (5) If $p = abb$, then

$$\Delta^0(w) = (a^b b^b)^{\frac{b-1}{2}} \underline{a}^b (b^a a^a)^{\frac{b-1}{2}} b^a \dots$$

$$\Delta^1(w) = b^b a^b \dots$$

$$\Delta^2(w) = bb \dots$$

w has the factor $f = a^b b^a a^a$.

Using Proposition 13, we conclude. □

3.2 Over an even alphabet

Let us now consider the case of an alphabet \mathcal{A} with even letters.

Theorem 19. *Over the alphabet $\{a < b\}$, with a and b even, the only smooth word that is also an infinite Lyndon word is $m_{\{a < b\}}$.*

Proof. We proceed similarly as in the previous section. The 4 possibilities are illustrated in Figure 2.

Case (1) If $p = aax$, with $x \in \mathcal{A}$, then

$$\Delta^0(w) = (a^a b^a)^{\frac{x}{2}} (\underline{a}^b b^b)^{\frac{x}{2}} \dots$$

$$\Delta^1(w) = a^x b^x \dots$$

$$\Delta^2(w) = xx \dots$$

w has the factor $f = a^b$.

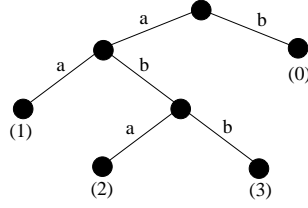


Figure 2. Possible cases for an even alphabet

Case (2) If $p = abax$, with $x \in \mathcal{A}$, then

$$\Delta^0(w) = ((a^b b^b)^{\frac{a}{2}} (a^a b^a)^{\frac{a}{2}})^{\frac{x}{2}} ((a^b b^b)^{\frac{b}{2}} (a^a b^a)^{\frac{b}{2}})^{\frac{x}{2}} \dots$$

$$\Delta^1(w) = (b^a a^a)^{\frac{x}{2}} (b^b a^b)^{\frac{x}{2}} \dots$$

$$\Delta^2(w) = a^x b^x \dots$$

$$\Delta^3(w) = xx \dots$$

w has the factor $f = (a^b b^b)^{\frac{b}{2}}$.

Case (3) Recall that the minimal smooth word $\Phi^{-1}(ab^\omega)$ is a Lyndon word. Let us show that this is the only smooth word that is also a Lyndon word. In order to prove it, let us suppose that we can write $p = ab^k ay$, with $k \geq 2$ maximal (since Case (2) has already excluded the possibility $k = 1$) and $y \in \mathcal{A}^*$. Let us compute $u = \Phi^{-1}(bbax)$, with $x \in \mathcal{A}$. We get

$$\Delta^0(u) = ((b^b a^b)^{\frac{a}{2}} (b^a a^a)^{\frac{a}{2}})^{\frac{x}{2}} ((b^b a^b)^{\frac{b}{2}} (b^a a^a)^{\frac{b}{2}})^{\frac{x}{2}}$$

$$\Delta^1(u) = (b^a a^a)^{\frac{x}{2}} (b^b a^b)^{\frac{x}{2}}$$

$$\Delta^2(u) = a^x b^x$$

$$\Delta^3(u) = xx$$

Since a and b are even and using Lemma 12, $\Phi^{-1}(b^k ay)$ can be written as

$$(w_1^{\frac{a}{2}} w_2^{\frac{a}{2}})^{\frac{y}{2}} (w_1^{\frac{b}{2}} w_2^{\frac{b}{2}})^{\frac{y}{2}} s,$$

with $w_1 = \Delta_b^{-(k-2)}(b^b a^b)$, $w_2 = \Delta_b^{-(k-2)}(b^a a^a)$ and $s \in \mathcal{A}^*$.

Moreover, since $\Phi^{-1}(b^\omega)$ is the maximal smooth word, $\Phi^{-1}(b^k)$ (resp. $\Phi^{-1}(b^{k-1}a)$) is prefix of w_1 (resp. w_2), we have that $w_1 > w_2$ and w_1 is not prefix of w_2 . Furthermore for $k \geq 2$, using Equation (1) we get

$$\Delta_b^{-1}(w_1) > \Delta_b^{-1}(w_2) \Leftrightarrow \Delta_a^{-1}(w_1) < \Delta_a^{-1}(w_2),$$

that implies

$$\Delta_a^{-1}(w_1^{\frac{b}{2}}) < \Delta_a^{-1}(w_1^{\frac{a}{2}} w_2^{\frac{a}{2}}).$$

Thus $\Phi^{-1}(ab^k ay)$ is not a Lyndon word.

The only Lyndon smooth word over an even 2-letter alphabet is the minimal smooth word $m_{\mathcal{A}}$ with $\Phi(m_{\mathcal{A}}) = ab^\omega$. \square

3.3 Over an odd alphabet

In this section, we prove the following result.

Theorem 20. *Over the alphabet $\{a < b\}$, with a and b odd, there exists an infinite smooth Lyndon word if and only if $a = 1$. More precisely, the smooth Lyndon word is the minimal smooth word $m_{\{1 < b\}}$, with $b \in 2\mathbb{N} + 1$.*

Before proving Theorem 20, some results are required.

Lemma 21. *Let $\mathcal{A} = \{a < b\}$ be an odd alphabet. Let w, w' be two factors of a smooth word such that $w < w'$ and $w = xay$, $w' = xby'$, with $x, y, y' \in \mathcal{A}^*$. Then, if $|x|$ is even,*

$$\Delta_{\alpha}^{-1}(w) < \Delta_{\alpha}^{-1}(w') \iff \bar{\alpha} < \alpha,$$

with $\alpha \in \mathcal{A}$ and $\bar{\alpha}$ its complement. If $|x|$ is odd, then

$$\Delta_{\alpha}^{-1}(w) < \Delta_{\alpha}^{-1}(w') \iff \alpha < \bar{\alpha}.$$

Proof. Assume $|x|$ even. By direct computation, we have the following equations:

$$\Delta_{\alpha}^{-1}(w) = \Delta_{\alpha}^{-1}(xay) = \Delta_{\alpha}^{-1}(x)\Delta_{\alpha}^{-1}(a)\Delta_{\bar{\alpha}}^{-1}(y) = \Delta_{\alpha}^{-1}(x)\alpha^a\Delta_{\bar{\alpha}}^{-1}(y)$$

and

$$\Delta_{\alpha}^{-1}(w') = \Delta_{\alpha}^{-1}(xby') = \Delta_{\alpha}^{-1}(x)\Delta_{\alpha}^{-1}(b)\Delta_{\bar{\alpha}}^{-1}(y') = \Delta_{\alpha}^{-1}(x)\alpha^b\Delta_{\bar{\alpha}}^{-1}(y').$$

Then $\Delta_{\alpha}^{-1}(w) < \Delta_{\alpha}^{-1}(w')$ if and only if $\Delta_{\bar{\alpha}}^{-1}(y)[0] < \alpha$. We conclude using $\Delta_{\bar{\alpha}}^{-1}(y)[0] = \bar{\alpha}$. A similar argument holds for $|x|$ odd. \square

Let us now prove 2 sub-cases of Theorem 20: $a \neq 1$ (Theorem 22) and $a = 1$ (Theorem 25).

Theorem 22. *Over the alphabet $\{a < b\}$, with a, b odd and $a \neq 1$, there is no infinite smooth word that is also a Lyndon word.*

Proof. As in Sections 3.1 and 3.2, we proceed by inspection of the different possible prefixes of $\Phi(w)$ (see Figure 3) for an infinite smooth word w .

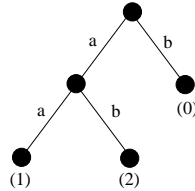


Figure 3. Possible cases for an odd alphabet, with $a \neq 1$

Case (1) If $p = aax$, then

$$\Delta^0(w) = (a^a b^a)^{\frac{x-1}{2}} a^a (b^b a^a)^{\frac{x-1}{2}} b^b (a^a b^a)^{\frac{x-1}{2}} a^a \dots$$

$$\Delta^1(w) = a^x b^x a^x \dots$$

$$\Delta^2(w) = xxx \dots$$

Since $x \geq a > 1$ and x is odd, $\frac{x-1}{2} \geq 1$. Thus, w has the factor $f = a^b$.

Remark 23. In the same way as in Remark 16, we can suppose that $\Delta^i(w)$ starts by a block of length at least 3.

Case (2) If $p = abx$, then

$$\Delta^0(w) = (a^b b^b)^{\frac{x-1}{2}} a^b (b^a a^a)^{\frac{x-1}{2}} b^a (a^b b^b)^{\frac{x-1}{2}} a^b \dots$$

$$\Delta^1(w) = b^x a^x b^x \dots$$

$$\Delta^2(w) = xxx \dots$$

w has the factor $f = a^b b^a a^a$.

Since in the 3 cases, it is possible to find a factor smaller than the prefix, we conclude that there is no smooth Lyndon word over an odd alphabet \mathcal{A} , with $a \neq 1$. \square

Proposition 24. *Let $w \in \{1 < b\}^\omega$ be an infinite smooth word that is also a Lyndon word. Then $1b \in \text{Pref}(\Phi(w))$.*

We know from [5] that the minimal smooth word over an odd 2-letter alphabet $\{1 < b\}$ is a Lyndon word. The next theorem shows that this is the only infinite Lyndon word over the alphabet $\{1 < b\}$.

Theorem 25. *Over the alphabet $\{1 < b\}$, with b odd, the only infinite smooth word that is also a Lyndon word is $m_{\{1 < b\}}$.*

Proof. Recall that $\Phi(m_{\{1, b\}}) = (1b)^\omega$. Let us show that for an infinite smooth word w , any prefix p of $\Phi(w)$ that is not prefix of $(1b)^\omega$ can not be such that $\Phi^{-1}(p)$ is prefix of an infinite Lyndon word. By Proposition 24, p starts by $1b$. We proceed by inspection of the different possibilities (see Figure 4).

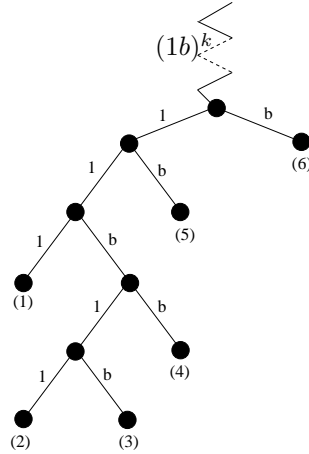


Figure 4. Possible prefixes of p starting by $(1b)^k$, $k \geq 1$

Case (1): $p = (1b)^k 111$. Let us consider the prefix $\Phi^{-1}(1b111)$ of u :

$$\Delta^0(u) = 1^b((b1)^{\frac{b-1}{2}}b(1^bb)^{\frac{b-1}{2}}\underline{1}^b)^{\frac{b-1}{2}}(b1)^{\frac{b-1}{2}}b1^bb1 \dots$$

$$\Delta^1(u) = b(1^bb)^{\frac{b-1}{2}}1^bb1 \dots$$

$$\Delta^2(u) = 1b^b1 \dots$$

$$\Delta^3(u) = 1b \dots$$

$$\Delta^4(u) = 1 \dots$$

u has the factor $f = 1^b(b1)^{\frac{b-1}{2}}b1^bb1$. Using Lemma 21 $2(k-1)$ times, we conclude that the prefix p does not describe a smooth Lyndon word. In Cases (2), (3), (4) and (6), we get the same conclusion with a similar argument. Thus the only prefix leading to a smooth Lyndon word is the one considered in Case (5): $(1b)^\omega$. \square

Proof. (of Theorem 20) Follows from Theorems 22 and 25. \square

3.4 Over \mathcal{A} with a odd and b even

In this section, we consider infinite smooth words over an alphabet $\{a < b\}$, with a odd and b even. We prove that over this alphabet, there is no infinite smooth word that is also a Lyndon word. In order to prove it, we consider 2 cases, $a \neq 1$ and $a = 1$, that have to be analysed separately.

Theorem 26. *Over the alphabet $\{a < b\}$, with $a \neq 1$ odd and b even, there is no smooth infinite word that is a Lyndon word.*

Proof. There are 5 possibilities to consider, illustrated in Figure 5 (a).

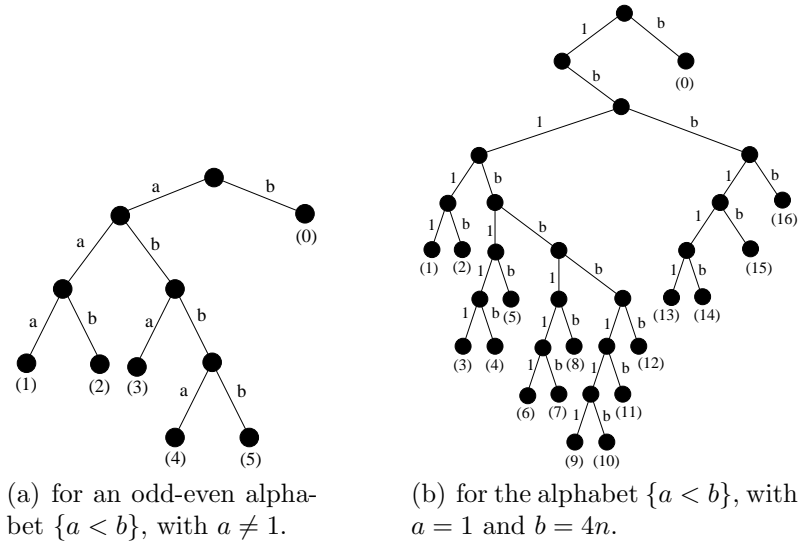


Figure 5. Possible cases...

In each case, it is possible to find a factor f smaller than the smooth word. Thus, there is no smooth Lyndon word. \square

Theorem 27. *Over the alphabet $\{a < b\}$, with $a = 1$ and $b = 4n$, there is no infinite smooth word that is a Lyndon word.*

Proof. Figure 5 (b) shows the different cases to consider. For each of the 16 cases, it is again possible to find a factor of $\Phi^{-1}(p)$ in order to prove that it is not a prefix of an infinite Lyndon word. \square

Theorem 28. *Over the alphabet $\{a < b\}$, with $a = 1$ and $b = 2(2n + 1)$, there is no infinite smooth word that is a Lyndon word.*

Proposition 24 can be generalized to an alphabet $\{1, b\}$, with b even. This result will be used in the following proof.

Proof. Figure 6 shows the different cases to consider. Cases numbered less or equal to 16 are the same as in Theorem 27. For the other cases, it is possible to find a factor in $\Phi^{-1}(p)$ smaller than its prefix, following that the word is not a Lyndon word. \square

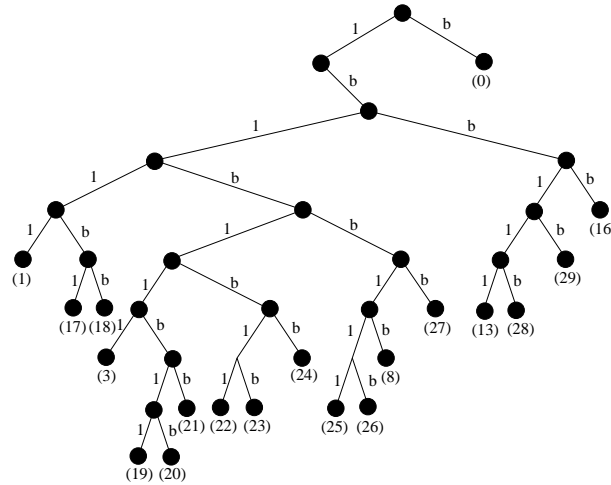


Figure 6. Different cases for the alphabet $\{a < b\}$, with $a = 1$ and $b = 2(2n + 1)$

4 Summary and concluding remarks

The next theorem summarizes the results of Section 3.

Theorem 29. *Over any 2-letter alphabet, the only infinite smooth words that are also infinite Lyndon words are $m_{\{2a < 2b\}}$ and $m_{\{1 < 2b+1\}}$, for $a, b \in \mathbb{N} \setminus \{0\}$.*

Recall that for the alphabet $\{1, 2\}$, it is conjectured [4] that in any infinite smooth word, any smooth factor appears. From this conjecture follows that no infinite smooth word is a Lyndon word. This is exactly what we have proved for the alphabet $\{1, 2\}$. Moreover, the existence of infinite smooth Lyndon words over the alphabets $\{2a < 2b\}$ and $\{1 < 2b + 1\}$ leads to the following corollary.

Corollary 30. *Let \mathcal{A} be a 2-letter alphabet such that $\mathcal{A} = \{2a < 2b\}$ or $\mathcal{A} = \{1 < 2b + 1\}$. Then, any infinite smooth words $w \in \mathcal{A}^\omega$ does not contain every smooth factors.*

Otherwise, no infinite Lyndon word would exist: a factor smaller than the prefix necessarily occurs. It is also interesting to notice that our main result completely characterized the trivial finite Lyndon factorization of infinite smooth words: the only infinite smooth words that have a finite Lyndon factorization composed of only one factor are $m_{\{2a < 2b\}}$ and $m_{\{1 < 2b+1\}}$. It is still an open problem to characterized infinite smooth words that have a non trivial finite Lyndon factorization. Giving an explicit computation of the Lyndon factorization, finite or infinite, of any infinite smooth words, as Melançon did for standard Sturmian words [19] is still a challenging problem.

Acknowledgements. The author would like to thank Pierre Lalonde for his interesting question during the LaCIM seminar that leads to this paper and Srečko Brlek for his comments.

References

1. A. BERGERON-BRLEK, S. BRLEK, A. LACASSE, AND X. PROVENÇAL: *Patterns in smooth tilings*, in Proceedings of WORDS'03, vol. 27 of TUCS Gen. Publ., Turku Cent. Comput. Sci., Turku, 2003, pp. 370–381.

2. J. BERSTEL AND A. DE LUCA: *Sturmian words, Lyndon words and trees*. Theoret. Comput. Sci., 178(1-2) 1997, pp. 171–203.
3. V. BERTHÉ, S. BRLEK, AND P. CHOQUETTE: *Smooth words over arbitrary alphabets*. Theoret. Comput. Sci., 341 2005, pp. 293–310.
4. S. BRLEK, S. DULUCQ, A. LADOUCEUR, AND L. VUILLON: *Combinatorial properties of smooth infinite words*. Theoret. Comput. Sci., 352 2006, pp. 306–317.
5. S. BRLEK, D. JAMET, AND G. PAQUIN: *Smooth words on 2-letter alphabets having same parity*. Theoret. Comput. Sci., 393 2008, pp. 166–181.
6. S. BRLEK AND A. LADOUCEUR: *A note on differentiable palindromes*. Theoret. Comput. Sci., 302 2003, pp. 167–178.
7. S. BRLEK, G. MELANÇON, AND G. PAQUIN: *Properties of the extremal infinite smooth words*. Discrete Math. Theoret. Comput. Sci., 9(2) 2007, pp. 33–49 (electronic).
8. A. CARPI: *On repeated factors in C^∞ -words*. Information Processing Letters, 52 1994, pp. 289–294.
9. K. T. CHEN, R. H. FOX, AND R. C. LYNDON: *Free differential calculus, IV. The quotient groups of the lower central series*. Ann. of Math., 68 1958, pp. 81–95.
10. F. M. DEKKING: *On the structure of self generating sequences*. Séminaire de théorie des nombres de Bordeaux, exposé 31, 1980–1981.
11. J.-P. DUVAL: *Factorizing words over an ordered alphabet*. Journal of Algorithms, 4 1983, pp. 363–381.
12. J.-P. DUVAL: *Génération d’une section des classes de conjugaison et arbre de mots de Lyndon de longueur bornée*. Theoret. Comput. Sci., 60 1988, pp. 255–283.
13. C. HOHLWEG AND C. REUTENAUER: *Lyndon words, permutations and trees*. Theoret. Comput. Sci., 307 2003, pp. 173–178.
14. Y. B. HUANG: *On the condition of powers of C^∞ -word to be a C^∞ -word*. Applied Mathematics - A Journal of Chinese Universities (Ser. A), 2 1997, pp. 243–246.
15. Y. B. HUANG: *On the number of C^∞ -words of form $\tilde{w}xw$* . Theoret. Comput. Sci., 393 2008, pp. 280–286.
16. D. JAMET AND G. PAQUIN: *Discrete surfaces and infinite smooth words*, in FPSAC’05 - 17th annual International conference on Formal Power Series and Algebraic Combinatorics (Taormina, Italie) June 20–25, 2005.
17. W. KOLAKOSKI: *Self generating runs*, Problem 5304. Amer. Math. Monthly, 72 1965, p. 674.
18. M. LOTHAIRE: *Combinatorics on words*, Addison Wesley, Reading MA, 1983.
19. G. MELANÇON: *Lyndon factorization of Sturmian words*. Discrete Appl. Math., 210 2000, pp. 137–149.
20. G. MELANÇON AND C. REUTENAUER: *Lyndon words, free algebras and shuffles*. Canad. J. Math., 4 1989, pp. 577–591.
21. G. PAQUIN: *Mots équilibrés et mots lisses*, PhD thesis, Université du Québec à Montréal, 2008.
22. C. REUTENAUER: *Free Lie algebras*, vol. 7, London Math. Soc. Monographs New Ser., 1993.
23. C. REUTENAUER: *Mots de Lyndon généralisés*. Sémin. Lothar. Combin., 54 2005/07, pp. Art. B54h, 16 pp. (electronic).
24. G. RICHOMME: *Conjugacy of morphisms and Lyndon decomposition of standard Sturmian words*. Theoret. Comput. Sci., 380(3) 2007, pp. 393–400.
25. P. SÉÉBOLD: *Lyndon factorization of the Prouhet words*. Theoret. Comput. Sci., 307 2003, pp. 179–197.
26. R. SIROMONEY, L. MATTHEW, V. R. DARE, AND K. G. SUBRAMANIAN: *Infinite lyndon words*. Information Processing Letters, 50 1994, pp. 101–104.
27. W. D. WEAKLEY: *On the number of C^∞ -words of each length*. J. Combin. Theory Ser. A, 51(1) 1989, pp. 55–62.