Smooth Words on a 2-letter alphabets having same parity

Srecko Brlek, Damien Jamet, Geneviève Paquin

Smooth Words on 2-letter alphabets having same parity ⋆

S. Brlek [a], D. Jamet [b], G. Paquin [*,a]

[a] *LaCIM, Université du Québec à Montréal, C. P. 8888 Succursale "Centre-Ville", Montréal (QC), CANADA H3C 3P8*

[b] *LORIA - Campus Scientifique - BP 239 54506 Vandoeuvre-lès-Nancy Cedex, FRANCE*

Abstract

In this paper, we consider smooth words over 2-letter alphabets $\{a, b\}$, where a, b are integers having same parity, with $0 < a < b$. We show that all are recurrent and that the closure of the set of factors under reversal holds for odd alphabets only. We provide a linear time algorithm computing the extremal words, w.r.t. lexicographic order. The minimal word is an infinite Lyndon word if and only if either $a = 1$ and b odd, or a, b are even. A connection is established between generalized Kolakoski words and maximal infinite smooth words over even 2-letter alphabets revealing new properties for some of the generalized Kolakoski words. Finally, the frequency of letters in extremal words is $1/2$ for even alphabets, and for $a = 1$ with b odd, the frequency of b's is $1/(\sqrt{2b-1}+1)$.

Key words: Smooth words, Kolakoski word, Lyndon factorization, letter frequency

1 Introduction

Smooth infinite words over $\Sigma = \{1, 2\}$ form an infinite class $\mathcal{K}$ of infinite words containing the well known Kolakoski word K [10] defined as one of the two fixed points of the run-length encoding function Δ, that is

$$\Delta(K) = K = 2211212212211211221211212211211212212211212212\cdots .$$

⋆ with the support of NSERC (Canada)

* Corresponding author.

Email addresses: brlek.srecko@uqam.ca (S. Brlek), jamet@loria.fr (D. Jamet), genevievepaquin@gmail.com (G. Paquin).

They are characterized by the property that the orbit obtained by iterating Δ is contained in $\{1, 2\}^*$. As a discrete dynamical system, $(\mathcal{K}, \Delta)$ is topologically conjugate of the full shift (Σ^*, σ) where σ is the shift operator. In the early work of Dekking [8] 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 [6] that K does contain only a finite number of squares, implying by direct inspection that K is cube-free. This result was extended in [5] to the infinite class $\mathcal{K}$ of smooth words over $\Sigma = \{1, 2\}$. Weakley [16] showed that the complexity function (number of factors of length n) of $\mathcal{K}$ is polynomially bounded. In [4], a connection was established between the palindromic complexity and the recurrence of K. More recently, Berthé et al. [2] studied smooth words over arbitrary alphabets and obtained a new characterization of the infinite Fibonacci word. Relevant work may also be found in [1] and in [2,9], where generalized Kolakoski words are studied for arbitrary alphabets. Finally, in [13], the authors studied the extremal infinite smooth words, that is the minimal and the maximal ones w.r.t. the lexicographic order, over the alphabets $\{1, 2\}$ and $\{1, 3\}$: a surprising link was established between the minimal infinite smooth word over $\{1, 3\}$ and the Fibonacci word.

Here, we deal with smooth words over 2-letter alphabets $\{a, b\}$ where $a < b$ are positive integers having same parity. The paper is organized as follows. In Section 2, we borrow from Lothaire [11] all the basic notions on combinatorics on words, while in Section 3, we briefly sketch the computation of extremal infinite smooth words and recall the main results of Paquin et al. [13]. Section 4 deals with the extremal smooth words over odd alphabets. We generalize a result of [13] about the extremal words over $\{1, 3\}$: we show that $\Phi(m_{\{a,b\}}) = (ab)^\omega$ where $m_{\{a,b\}}$ is the minimal smooth word over the alphabet $\{a, b\}$ and Φ is a natural bijection (Theorem 12), giving linear time algorithm for computing the extremal words (Corollary 14). A recurrent definition of extremal smooth words over the alphabet $\{1, b\}$ is given and it provides the letter frequencies (Theorem 20). Next, we prove that the set $F(w)$ of factors of an infinite smooth word w is closed under reversal, and consequently, that w is recurrent (Proposition 15). Finally, we show that the minimal infinite smooth word is an infinite Lyndon word if and only if $a = 1$ and then, that the Lyndon factorization of $\Delta(m_{\{a,b\}})$ is an infinite sequence of finite Lyndon words (Theorem 19). Section 5 is devoted to even alphabets, in which case $\Phi(m_{\{a,b\}}) = ab^\omega$ (Theorem 22 and Corollary 23), yielding in turn a linear time algorithm to generate the extremal words. From the algorithm, we deduce that the frequency of the letters a and b is $\frac{1}{2}$. Moreover, smooth words over even alphabets are recurrent (Proposition 25) despite the fact that the set of factors is not closed under reversal (Proposition 26). Minimal smooth words are infinite Lyndon words (Theorem 29), and a connection is established between generalized Kolakoski words and maximal infinite smooth words. It provides new properties for some generalized Kolakoski words which are still open problems for the alphabet $\{1, 2\}$.

2 Preliminaries

Throughout this paper Σ is a finite *alphabet* of *letters* equipped with an order $<$. A finite *word* is a finite sequence of letters

$$w : [1..n] \longrightarrow \Sigma, n \in \mathbb{N}$$

of length n, and $w[i]$ denotes its i-th letter. The set of n-length words over Σ is denoted by Σ^n. By convention the *empty* word is denoted by ε and its length is 0. The free monoid generated by Σ is defined by $\Sigma^* = \bigcup_{n\geq 0} \Sigma^n$ and $\Sigma^* \setminus \varepsilon$ is denoted Σ^+. The set of right infinite words, also called infinite words for short, is denoted by Σ^ω and $\Sigma^\infty = \Sigma^* \cup \Sigma^\omega$. 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 $\mathbb{N}^\omega$ is that of infinite ones. Given a word $w \in \Sigma^*$, a *factor* f of w is a word $f \in \Sigma^*$ satisfying

$$\exists x, y \in \Sigma^*, w = xfy.$$

If $x = \varepsilon$ (resp. $y = \varepsilon$) then f is called a *prefix* (resp. *suffix*). A *block* of length k is a factor of the particular form $f = \alpha^k$, with $\alpha \in \Sigma$. The set of all factors of w, also called the language of w, is denoted by $F(w)$, and those of length n is $F_n(w) = F(w) \cap \Sigma^n$, while $\mathrm{Pref}(w)$ (resp. $\mathrm{Suff}(w)$) denotes the set of all prefixes (resp. suffixes) of w. The length of a word w is $|w|$, and the number of occurrences of a factor $f \in \Sigma^*$ is $|w|_f$. For a finite word w, the frequency of the letter a is defined by $d_a(w) = |w|_a/|w|$. For an infinite word w, we follow [14] and define the frequency of a letter a in w by

$$d_a(w) = \lim_{n\to\infty} \frac{1}{n} |w[1..n]|_a$$

whenever this limit exists. An infinite word w is said *recurrent* if $|w|_f$ is infinite for every factor $f \in F(w)$.

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

For $u, v \in \Sigma^*$, we write $u \prec v$ 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 $1 \leq i \leq k - 1$ and $u[k] < v[k]$. The relation $\preceq$ defined by $u \preceq v$ if and only if $u = v$ or $u \prec v$, is called the *lexicographic order*. That definition holds for Σ^∞. Note that in general, the complementation does not preserve the lexicographic order. Indeed, when u is not a proper prefix of v then

$$u \succ v \iff \overline{u} \prec \overline{v}. \tag{1}$$

A word $u \in \Sigma^*$ is a *Lyndon word* if $u \prec v$ for all proper non-empty suffixes v of u. For instance, the word 11212 is a Lyndon word while 12112 is not since $112 \prec 12112$. A word of length 1 is clearly a Lyndon word. The set of Lyndon words is denoted by $\mathcal{L}$. From Lothaire [11], we take the following theorem.

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

$$w = \ell_1\ell_2\cdots\ell_n = \bigodot_{i=1}^{n} \ell_i, \ \textit{where} \quad \ell_i \in \mathcal{L}, \ \textit{and} \quad \ell_1 \succeq \ell_2 \succeq \cdots \succeq \ell_n. \tag{2}$$

Siromoney et al. [15] extended Theorem 1 to infinite words. The set $\mathcal{L}_\infty$ of infinite Lyndon words consists of words smaller than any of their suffixes.

Theorem 2 [15] *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 $(\ell_k)_{k\geq 1}$ of elements in $\mathcal{L}$ such that*
$$w = \ell_1\ell_2\ell_3\cdots \ \textit{and for all } k, \ \ell_k \succeq \ell_{k+1}.$$
(ii) *there exist a finite sequence $\ell_1, \ldots, \ell_m \,(m \geq 0)$ of elements in $\mathcal{L}$ and $\ell_{m+1} \in \mathcal{L}_\infty$ such that*
$$w = \ell_1\ell_2\cdots\ell_m\ell_{m+1} \ \textit{and} \ \ell_1 \succeq \cdots \succeq \ell_m \succ \ell_{m+1}.$$

Let recall from ([11] 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 $u \prec v$.*

A direct corollary of this lemma is:

Corollary 4 *Let $u, v \in \mathcal{L}$, with $u \prec v$. Then uv^n, $u^nv \in \mathcal{L}$, 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. It also was used for the enumeration of factors in the Thue-Morse sequence [3]. Let $\Sigma = \{a, b\}$ be an ordered alphabet. Then every word $w \in \Sigma^*$ can be uniquely written as a product of factors as follows:

$$w = a^{i_1}b^{i_2}a^{i_3}\cdots$$

with $i_1 \geq 0$ and $i_k \geq 1$ for $k \geq 2$. The operator giving the size of the blocks appearing in the coding is a function $\Delta : \Sigma^* \longrightarrow \mathbb{N}^*$, defined by $\Delta(w) = i_1, i_2, i_3, \cdots$ which is easily extended to infinite words as $\Delta : \Sigma^\omega \longrightarrow \mathbb{N}^\omega$.

For instance, let $\Sigma = \{1, 3\}$ 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 $\Delta(w) \subseteq \{1, 2, \cdots, 9\}^*$, 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 $\Delta : \{1, 3\}^* \longrightarrow \{1, 2, 3, 4\}^*$.

Remark 5 *From now on, we only consider 2-letter alphabets $\Sigma = \{a, b\}$, with $a < b$.*

Recall from [4] that Δ is not bijective since $\Delta(w) = \Delta(\overline{w})$, but commutes with the reversal $(\widetilde{\ })$, is stable under complementation $(\overline{\ })$ and preserves palindromicity. Since Δ is not bijective, pseudo-inverse functions

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

are defined for 2-letter alphabets by

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

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

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

where $\beta = \alpha$ if $|w|$ odd, and $\beta = \overline{\alpha}$ 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. Let $w = 1333111333133311133313133311133313331113331$. The successive application of Δ gives :

$\Delta^0(w) = \mathbf{1}333111333133311133313133311133313331113331$;
$\Delta^1(w) = \mathbf{1}333133311133313331$;
$\Delta^2(w) = \mathbf{1}31333131$;
$\Delta^3(w) = \mathbf{1}113111$;
$\Delta^4(w) = \mathbf{3}13$;
$\Delta^5(w) = \mathbf{1}11$;
$\Delta^6(w) = \mathbf{3}$.

The operator Δ extends to infinite words (see [4]). Define the set of infinite *smooth* words over $\Sigma = \{a, b\}$ by

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

In $\mathcal{K}_\Sigma$ the operator Δ has two fixpoints, namely

$$\Delta(K_{(a,b)}) = K_{(a,b)}, \qquad \Delta(K_{(b,a)}) = K_{(b,a)},$$

where $K_{(a,b)}$ is the generalized Kolakoski word [9] over the alphabet $\{a, b\}$ starting with the letter a.

Example. The Kolakoski word [10] over $\Sigma = \{1, 2\}$ and starting with the letter 2 is $K = K_{(2,1)}$. We also have $K_{(2,3)} = 2233222333223322333222\cdots$, and $K_{(3,1)} = 3331113331313331113331\cdots$.

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

$$\Phi(w)[j + 1] = \Delta^j(w)[1], \text{for } j \geq 0,$$

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

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

Then, for $k = \infty$, $\Phi^{-1}(u) = \lim_{k\to\infty} w_k = \lim_{k\to\infty} \Phi^{-1}(u[1..k])$. Such a bijection also exists for k-letters alphabet, but an additional parameter is required for recording the letter written, in order to avoid writing 0-blocks.

Remark 6 *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. For the word $w = 133311133313331113331313331113331333111333 1$ of Example 2, $\Phi(w) = \mathbf{1111313}$.

Note that since Φ is a bijection, the set of infinite smooth words is infinite, and conjugate of the full shift Σ^ω (in the terminology of symbolic dynamics). For later use we borrow from [2] the following powerful lemma:

Lemma 7 [Glueing Lemma] *Let u, $v \in \Delta^*(\Sigma)$. 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) $\Delta^i(uv) = \Delta^i(u)\Delta^i(v)$.

We recall from [13] the useful *right derivative* $D_r : \Sigma^* \rightarrow \mathbb{N}^*$ such that:

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

where $\alpha \in \mathbb{N}$. A word w is *r-smooth* (also said *smooth prefix*) if $\forall k \geq 0,\ D_r^k(w) \in \Sigma^*$. In other words, if a word w is *r-smooth*, then it is a prefix of at least one infinite smooth word (see [5] for more details).

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

3 Computation of extremal smooth words

Let $m_{\{a,b\}}$ (resp. $M_{\{a,b\}}$) be the minimal (resp. maximal) infinite smooth word over the alphabet $\Sigma = \{a, b\}$ w.r.t the lexicographic order. From (1), it easily follows that $M_{\{a,b\}} = \overline{m}_{\{a,b\}}$, so that the computation of $m_{\{a,b\}}$ also yields $M_{\{a,b\}}$, by simply exchanging the order on the alphabet. The naive algorithm for computing the minimal infinite smooth word over an alphabet Σ consists in computing the minimal smooth prefixes of increasing length. At each step, the minimal letter of the alphabet Σ which makes the word a smooth prefix is added. The smoothness condition is checked with the right derivative operator D_r, and ensures that the prefix computed is the prefix of at least one infinite smooth word. If we assume $a < b$, the corresponding algorithm is:

Algorithm 1
input : $\Sigma = \{a, b\}$, *MaxLength ;*

```
m_{a,b} := a;
loop
    if isSmooth (m_{a,b} · a) then m_{a,b} := m_{a,b} · a;
    else m_{a,b} := m_{a,b} · b;
    end if;
    exit when length(m_{a,b})=MaxLength;
end loop
```

Observe that Algorithm 1 does not depend on letter parities. For different alphabets and for $MaxLength = 47$, we obtain the following words:

$$
\begin{aligned}
m_{\{1,2\}}[1..47] &= 11211221211212211211212212112212211211212211211, \\
M_{\{1,2\}}[1..47] &= 21211221211212211211212212112212211211212211211, \\
m_{\{1,3\}}[1..47] &= 11131113131113111313111313111311131311131113131, \\
M_{\{1,3\}}[1..47] &= 33313331313331333131333131333133313133313331313, \\
m_{\{2,4\}}[1..47] &= 22224444222244442244224422224444222244442244224, \\
m_{\{3,5\}}[1..47] &= 33333555553333355533355533333555553333355533355, \\
m_{\{2,3\}}[1..47] &= 22233222332233222333222332233322233223322233333, \\
m_{\{3,4\}}[1..47] &= 33334444333344433344433333444433333444333444433334.
\end{aligned}
$$

With the naive algorithm, the computation of a n-length prefix of $m_{\{a,b\}}$ takes $\mathcal{O}(n^2 \log(n))$ steps: indeed, for every newly added letter to the current prefix of $m_{\{a,b\}}$, we have to check smoothness by applying the D_r operator. To improve the amount of D_r operations, it is convenient to add more than one letter at each step. That was already done for $m_{\{1,2\}}$ in [13] by using the De Bruijn graphs. The same idea can be applied to extremal smooth words for other alphabets, but we shall prove in the next sections that more efficient algorithms exist for computing them.

3.1 Extremal smooth words over $\{1,2\}$ *and* $\{1,3\}$

We recall some results established in a previous paper [13]. First, extensive computations yield

$$
\begin{aligned}
\Phi(m_{\{1,2\}}) &= 1212212112221121112112221111221211112222\cdots \\
\Phi(M_{\{1,2\}}) &= 2212212112221121112112221111221211112222\cdots
\end{aligned}
$$

No characaterization is known, so that we do not know whether $\Phi(m_{\{1,2\}})$ and $\Phi(M_{\{1,2\}})$ are periodic or not. Nevertheless, the minimal smooth word $m_{\{1,2\}} \notin \mathcal{L}_\infty$ [13].

In [2], Berthé et al. showed that the infinite Fibonacci word F, defined as

$$F = \lim_{n\to\infty} F_n \quad \text{where} \quad F_0 = 2, \quad F_1 = 1, \quad \text{and} \quad \forall n \geq 2,\ F_n = F_{n-1}F_{n-2},$$

is not smooth over the alphabet $\Sigma = \{1, 2\}$, but smooth over the alphabet $\Sigma = \{1, 2, 3\}$. More precisely, they proved that $\Phi(F) = 112(13)^\omega$, the periodicity meaning that $\Delta^k(F) = \Delta^{k+2}(F)$ for all $k \geq 3$. In [13], the link between the Fibonacci word and the minimal infinite smooth word over $\Sigma = \{1, 3\}$ is established:

Theorem 8 [[13] Theorem 6] $m_{\{1,3\}} = \Delta^3(F)$.

Since F and $m_{\{1,3\}}$ are in the same orbit of the Δ operator, Corollary 9 follows immediately from properties established for the Fibonacci orbit in [2].

Corollary 9 [[13] Cor. 8] *The extremal infinite smooth words over* $\Sigma = \{1, 3\}$ *satisfy the conditions:*

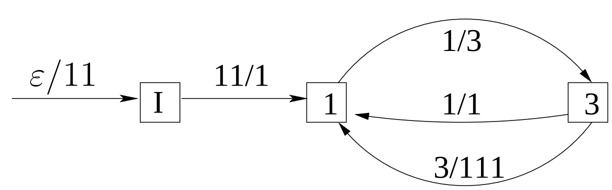
(i) $\Delta^k(m_{\{1,3\}}) = \Delta^{k+2}(m_{\{1,3\}})$, *for all* $k \geq 0$;
(ii) $\Phi(m_{\{1,3\}}) = (13)^\omega$ *and* $\Phi(M_{\{1,3\}}) = 3(31)^\omega$;
(iii) 33 *and* $31313 \notin F(m_{\{1,3\}})$; 11 *and* $13131 \notin F(M_{\{1,3\}})$;
(iv) *Let* $m_{\{1,3\}} = 11u$, *then* $\Delta(m_{\{1,3\}}) = 3u$.

The close relation between the Fibonacci word and the minimal infinite smooth word also provides a recursive definition for $m_{\{1,3\}}$:

Proposition 10 [[13] Prop. 9] *Let* $m_{\{1,3\}} = 11u$. *Then* u *is defined as*

$$u = \lim_{n\to\infty} u_n \quad \text{where} \quad u_0 = 11, \quad u_1 = 13, \quad \text{and} \quad \forall n \geq 2,\ u_n = u_{n-1}u_{n-2}.$$

Finally, from property (iv) of Corollary 9, the following transducer computing the minimal infinite smooth word $m_{\{1,3\}}$ in linear time is provided.



Our transducer is a finite state machine using one tape, and two heads used for reading and writing on it. The "next state" function labels the transitions between two states by (u, v): in a given state, the transducer reads u and write v, and moves to the next state.

The next table describes how the transducer is used to compute $m_{\{1,3\}}$.

Read	Write	Prefix of $m_{\{1,3\}}$
ε	11	11
11	1	111
1	3	1113
3	111	1113111
1	3	11131113
1	1	111311131
1	3	1113111313
3	111	1113111313111
1	3	1131113131113
...	...	...

In the next section, we show that for same parity alphabets the situation becomes simpler, a rather surprising fact.

4 Extremal words over odd alphabets

In this section, we assume that the letters of $\Sigma = \{a, b\}$, are both odd integers and such that $a < b$. We start by a useful lemma.

Lemma 11 *For all $u \in \Sigma^+, \Phi^{-1}(u)$ is a palindrome of odd length.*

Proof. Let $w = \Phi^{-1}(u)$. We proceed by induction on the length of u. If $n = |u| = 1$ then $w = \beta \in \Sigma$, which is a palindrome. If $n = 2$ then $u = \alpha\beta$, with $\alpha, \beta \in \{a, b\}$. Then $\Phi^{-1}(u) = w = \alpha^\beta$ is palindromic. Since a and b are odd, it follows that w has odd length. Assume now that the statement is true for every u such that $|u| \leq k$. Let $u' \in \Sigma^k$ and $w = \Phi^{-1}(u')$ is a palindrome of odd length. Let $|w| = 2j+1$. We then can write $w = w' \cdot w[j+1] \cdot \widetilde{w'}$, $w' \in \Sigma^*$ and

$$\Delta_\alpha^{-1}(w) = \Delta_\alpha^{-1}(w' \cdot w[j+1] \cdot \widetilde{w'}),$$

for $\alpha \in \Sigma$. There are two cases to consider: if $|w'|$ is odd, then

$$\Delta_\alpha^{-1}(w) = \Delta_\alpha^{-1}(w') \cdot \Delta_{\overline{\alpha}}^{-1}(w[j+1]) \cdot \Delta_\alpha^{-1}(\widetilde{w'}) = \Delta_\alpha^{-1}(w') \cdot \Delta_{\overline{\alpha}}^{-1}(w[j+1]) \cdot \widetilde{\Delta_\alpha^{-1}(w')},$$

and if $|w'|$ is even then

$$\Delta_\alpha^{-1}(w) = \Delta_\alpha^{-1}(w') \cdot \Delta_\alpha^{-1}(w[j+1]) \cdot \Delta_{\overline{\alpha}}^{-1}(\widetilde{w'}) = \Delta_\alpha^{-1}(w') \cdot \Delta_\alpha^{-1}(w[j+1]) \cdot \widetilde{\Delta_\alpha^{-1}(w')}.$$

The last equalities hold because of Property (3) of Section 2. In both cases each factor is a palindrome of odd length so that $\Delta_{\alpha}^{-1}(w)$ is palindromic too. We conclude by using the fact that $\Delta_{\alpha}^{-1}(w)$ are exactly the words $\Phi^{-1}(u)$ with $|u| = k+1$. $\square$

We state now a fundamental result, showing that for odd alphabets the situation is much simpler than for the alphabet $\{1, 2\}$.

Theorem 12 $\Phi(m_{\{a,b\}}) = (ab)^{\omega}$.

Proof. We proceed by induction on the length of the prefixes of $u = \Phi(m_{\{a,b\}})$. Note first that $m_{\{a,b\}}$ starts with a, the smallest letter. One easily checks that $\Phi^{-1}(ab) = a^b \prec a^a b \cdot w = \Phi^{-1}(aax)$, for any $x \in \Sigma, w \in \Sigma^*$. Assume now that $\Phi^{-1}((ab)^k)$ is minimal, for every $k \leq n$. Figure a) shows that since a and b are odd, the prefix defined by the vertical word $(ab)^n$ starts and ends with a. The same argument holds for each line, alternating a and b.

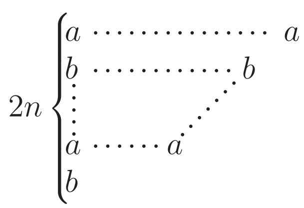


Figure a)

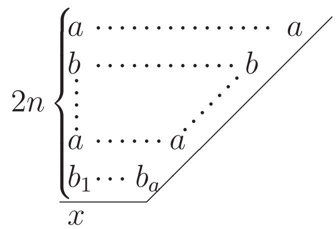


Figure b)

Let x be the $(2n+1)$-th letter of $\Phi(m_{\{a,b\}})$. We can then deduce from the value of x the next letter for every line. Either $x = a$ or $x = b$, the $2n$-th line starts with at least a occurences of the letter b. Since a, b are odd, each line starts and ends with the same letter, still alternating. This is shown in Figure b). The subscripts in the figure count the number of letters. For instance, $b_1 \cdots b_a$ means that there are a consecutive b's.

If $x = a$, then $\Delta_b^{-1}(a) = b^a$ and the $2n$-th line has the prefix $b^a x = b^a a$. If $x = b$, then $\Delta_b^{-1}(b) = b^a b^{b-a}$ and then the $2n$-th line starts by $b^a x = b^a b$. In both cases, that means that the $2n$-th line starts with $b^a x$. This is shown in Figure c).

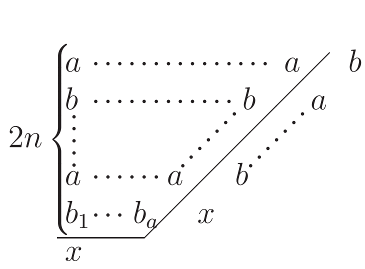


Figure c)

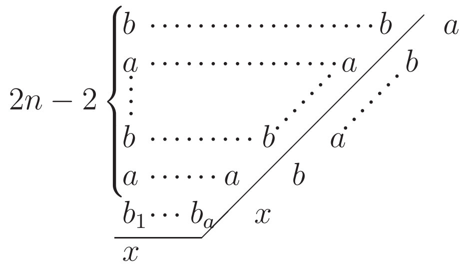


Figure d)

By the Glueing Lemma, $\Phi^{-1}((ab)^n x) = \Phi^{-1}((ab)^n a) \cdot \Phi^{-1}((ba)^{n-1}bx)s$, for some $s \in \Sigma^*$. Then, we deduce that the letter x is the one that makes $\Phi^{-1}((ba)^{n-1}bx)$ minimal. In Figure d), we consider $\Phi^{-1}((ba)^{n-1}bx)$. The letter x is the one that makes $\Phi^{-1}((ab)^{n-1}x)$ minimal. By the induction hypothesis, we get $x = a$. It follows that if $\Phi^{-1}((ab)^n)$ is minimal, then $\Phi^{-1}((ab)^n a)$ is so.

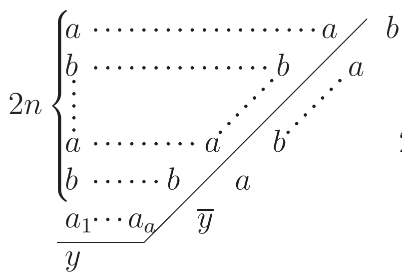


Figure e)

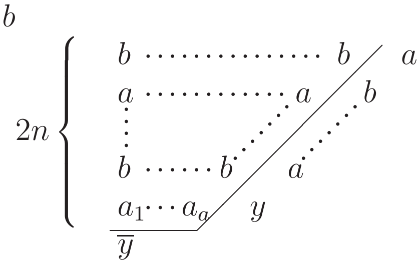


Figure f)

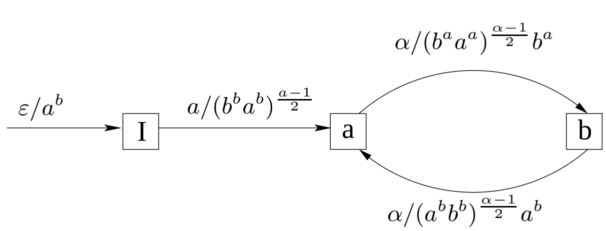
Figure e) shows that the next letter y is the one rending $\Phi^{-1}((ba)^n\overline{y})$ minimal. Figure f) describes that situation: $\overline{y}$ is such that $\Phi^{-1}((ab)^{n-1}ay)$ is minimal. By the induction hypothesis, we obtain $y = b$ and the conclusion follows. □

Using the equality $\Delta(m_{\{a,b\}}) = \Delta(M_{\{a,b\}})$, we get free the computation of Φ for the maximal word:

Corollary 13 $\Phi(M_{\{a,b\}}) = b(ba)^\omega$.

The periodicity of $\Phi(m_{\{a,b\}})$ yields a linear time algorithm generating the minimal (therefore the maximal) infinite smooth word for odd alphabets:

Corollary 14 *Let $\alpha \in \Sigma = \{a, b\}$. The following transducer computes $m_{\{a,b\}}$.*



Permuting the letters a and b in the transducer above yields directly the transducer for the maximal smooth word.

Two long standing conjectures of Dekking [8] concern, on one hand the closure of the set $F(K)$ of factors of the Kolakoski word by reversal and complementation, and on the other hand the recurrence of K. Dekking also showed that closure of $F(K)$ by complementation would imply the recurrence property. These conjectures were stated for every infinite smooth word over $\{1,2\}$ in [5]. Although the existence of arbitrarily long palindromes in smooth words on $\{1,2\}$ remains an unsolved conjecture, their existence would imply the recurrence property, a fact that was first observed in [4].

Corollary 9 (iii) implies that $F(m_{\{1,3\}})$ is not closed by complementation. However, for odd alphabets, the peculiar palindromic structure of smooth words (see Lemma 11) is powerful to establish the next result.

Proposition 15 *For every infinite smooth word w, the set $F(w)$ is closed under reversal and w is recurrent.*

Proof. Let f be a finite factor of w. Then $w = ufv$ for some $u, f \in \Sigma^*$ and $v \in \Sigma^\omega$. Since every smooth word w has, by Lemma 11, arbitrarily long palindromic prefixes, there exists a palindromic prefix p of w starting with uf, hence containing $\widetilde{uf}$ and the result follows. For the recurrence property one extra step is necessary. Since p contains both f and $\widetilde{f}$, any longer palindromic prefix q contains necessarily the same two occurrences of f and $\widetilde{f}$. As p is both a prefix and a suffix of q, p and consequently f occurs twice in q. □

4.1 Lyndon factorizations

We take now a closer look to the minimal words and start with a negative result.

Lemma 16 *If $a \neq 1$, then $m_{\{a,b\}} \notin \mathcal{L}_\infty$.*

Proof. Computing $\Phi^{-1}((ab)^2)$, we get $w_1 = b$, $w_2 = a^b$, $w_3 = (b^a a^a)^{\frac{b-1}{2}} b^a$ and the prefix of $m_{\{a,b\}}$:

$$w_4 = \Phi^{-1}((ab)^2) = [(a^b b^b)^{\frac{a-1}{2}} \mathbf{a^b} (\mathbf{b^a a^a})^{\frac{\mathbf{a-1}}{\mathbf{2}}} \mathbf{b^a}]^{\frac{\mathbf{b-1}}{\mathbf{2}}} (\mathbf{a^b b^b})^{\frac{\mathbf{a-1}}{\mathbf{2}}} \mathbf{a^b}.$$

Therefore, we can write $m_{\{a,b\}} = a^b b^b s$, with $s \in \Sigma^\omega$. A suffix of $m_{\{a,b\}}$ is $a^b b^a a s'$, with $s' \in \Sigma^\omega$. Then $a^b b^a a s' \prec a^b b^b s$, and hence, $m_{\{a,b\}} \notin \mathcal{L}_\infty$. □

Example. The word $m_{\{3,5\}} = 3333355555\mathbf{33333555333555333335555}\cdots$ has $s = 33333555333\cdots$ as a smaller suffix, then $m_{\{3,5\}} \notin \mathcal{L}_\infty$.

In Lemma 16, we assumed $a \neq 1$ to ensure that the word was starting with $a^b b^b$. In the case $a = 1$, the situation is different and we establish that $m_{\{1,b\}} \in \mathcal{L}_\infty$. Before proving this fact, some technical results are required about the prefixes of smooth words. For $k \geq 1$ we set

$$w_{2k} = \Phi^{-1}((1b)^k) \quad \text{and} \quad w_{2k-1} = \Phi^{-1}(b(1b)^k). \tag{4}$$

Proposition 17 *Let $\Sigma = \{1, b\}$. Then the following conditions hold:*

(i) $w_n = (w_{n-2} \cdot w_{n-3})^{\frac{b-1}{2}} \cdot w_{n-2}$, *for all* $n \geq 4$*;*

(ii) $w_{2k}w_{2k-1}$, $w_{2k}w_{2k+1} \in \mathcal{L}$, *for all* $k \geq 1$*;*

(iii) $w_{2k-2}w_{2k-1} \succeq w_{2k}$ *and* $w_{2k} \notin \mathrm{Pref}(w_{2k-2}w_{2k-1})$, *for all* $k \geq 2$.

Proof. We proceed by induction. (i) Direct computation yields $w_1 = b$, $w_2 = 1^b$, $w_3 = (b1)^{\frac{b-1}{2}}b$ and $w_4 = (1^b b)^{\frac{b-1}{2}}1^b$. Since $w_4 = (1^b \cdot b)^{\frac{b-1}{2}} \cdot 1^b = (w_2 \cdot w_1)^{\frac{b-1}{2}} \cdot w_2$, the claim is true for $n = 4$. Assume now that $w_m = (w_{m-2}w_{m-3})^{\frac{b-1}{2}} w_{m-2}$, for all $m \leq n$. Then, since the function Δ^{-1} distributes nicely because all w_i are palindromic of odd length by Lemma 11, we have:

$$\begin{aligned}
w_{n+1} &= \Delta_\alpha^{-1}(w_n), \\
&= \Delta_\alpha^{-1}\left((w_{n-2}w_{n-3})^{\frac{b-1}{2}} w_{n-2}\right), \\
&= \Delta_\alpha^{-1}\left((w_{n-2}w_{n-3})^{\frac{b-1}{2}}\right)\Delta_\alpha^{-1}(w_{n-2}), \\
&= \left(\Delta_\alpha^{-1}(w_{n-2})\Delta_{\overline{\alpha}}^{-1}(w_{n-3})\right)^{\frac{b-1}{2}} \Delta_\alpha^{-1}(w_{n-2}), \\
&= (w_{n-1}w_{n-2})^{\frac{b-1}{2}} w_{n-1},
\end{aligned}$$

with $\alpha = b$ if n even, $\alpha = 1$ otherwise.

(ii) From formulas (4), it follows that $w_2w_1 = 1^b b$, $w_2w_3 = 1^b(b1)^{\frac{b-1}{2}}b \in \mathcal{L}$, so that the claim is true for $k = 1$. Assume now that $w_{2k}w_{2k-1}$, $w_{2k}w_{2k+1} \in \mathcal{L}$ for every $k \leq n$.

1. $w_{2n+2}w_{2n+1} = (w_{2n}w_{2n-1})^{\frac{b-1}{2}} \cdot w_{2n}w_{2n+1}$, by (i). Then, using the induction hypothesis, $w_{2n}w_{2n-1}$, $w_{2n}w_{2n+1} \in \mathcal{L}$, so that $w_{2n+2}w_{2n+1} = u^{\frac{b-1}{2}}v$, where u, $v \in \mathcal{L}$ with $u \in \mathrm{Pref}(v)$ implies $u \prec v$. Now Corollary 4 applies, which concludes.

2. $w_{2n+2}w_{2n+3} = w_{2n+2}w_{2n+1} \cdot (w_{2n}w_{2n+1})^{\frac{b-1}{2}}$, by (i). Then, using (i) and the induction hypothesis, we deduce that $w_{2n+2}w_{2n+1}$, $w_{2n}w_{2n+1} \in \mathcal{L}$. Then, $w_{2n+2}w_{2n+3} = uv^{\frac{b-1}{2}}$, where $u, v \in \mathcal{L}$, $v \in \mathrm{Suff}(u)$ implies $u \prec v$. Again Corollary 4 permits to conclude.

(iii) For $k = 2$, $w_2w_3 = 1^b(b1)^{\frac{b-1}{2}}b = 1^b b1bs$ and $w_4 = (1^b b)^{\frac{b-1}{2}}1^b = 1^b b11s'$, $s, s' \in \Sigma^*$. Thus, the lemma is verified for $k = 2$. Assume now that it is true for all $k \leq n$. Then,

$$w_{2n}w_{2n+1} = w_{2n}w_{2n-1}(w_{2n-2}w_{2n-1})^{\frac{b-1}{2}}$$

and

$$w_{2n+2} = w_{2n}(w_{2n-1}w_{2n})^{\frac{b-1}{2}}.$$

Since $w_{2n-2}w_{2n-1} \succeq w_{2n}$, the conclusion follows. $\square$

Example. Let $\Sigma = \{1, 5\}$. Then $\Phi(m_{\{1,5\}}) = (15)^\omega$ and $w_1 = 5$, $w_2 = 11111$, $w_3 = 51515$. Proposition 17 (i) gives

$$w_4 = (w_2w_1)^{\frac{5-1}{2}}w_2 = w_2w_1w_2w_1w_2 = 11111511111511111.$$

Observe that $w_2w_1 = 111115 \in \mathcal{L}$ and $w_2w_3 = 1111151515 \in \mathcal{L}$.

Proposition 18 *Let $\Sigma = \{1, b\}$ and let L_n be the Lyndon factorization of w_n defined in (4). Then for $n \geq 4$:*

$$L_n = \begin{cases} \left(\bigodot_{i=1}^{\frac{b-1}{2}} w_{n-2}w_{n-3}\right) \cdot L_{n-2}, & \text{if } n \text{ even;} \\ L_{n-2} \cdot \left(\bigodot_{i=1}^{\frac{b-1}{2}} w_{n-3}w_{n-2}\right), & \text{if } n \text{ odd,} \end{cases}$$

where the dots separate the different Lyndon words of the factorization, as described in (2).

Proof. (By induction on n) Direct computations yield $w_1 = b$, $w_2 = 1^b$, $w_3 = (b1)^{\frac{b-1}{2}}b$, $w_4 = (1^b b)^{\frac{b-1}{2}}1^b$, $w_5 = ((b1)^{\frac{b-1}{2}}b1^b)^{\frac{b-1}{2}}(b1)^{\frac{b-1}{2}}b$ and the corresponding Lyndon factorizations are:

$$L_1 = b, \quad L_2 = \bigodot_{i=1}^{b} 1, \quad L_3 = b\bigodot_{i=1}^{\frac{b-1}{2}}(1b), \quad L_4 = \bigodot_{i=1}^{\frac{b-1}{2}}(1^b b)\bigodot_{i=1}^{b} 1$$

and

$$L_5 = b\bigodot_{i=1}^{\frac{b-1}{2}}(1b)\bigodot_{i=1}^{\frac{b-1}{2}}\left(1^b(b1)^{\frac{b-1}{2}}b\right).$$

As $L_4 = \left(\bigodot_{i=1}^{\frac{b-1}{2}} w_2w_1\right) \cdot L_2$ and $L_5 = L_3 \cdot \left(\bigodot_{i=1}^{\frac{b-1}{2}} w_2w_3\right)$, the Lyndon factorization is verified for $n = 4, 5$.

Assume now that the equality holds for every $m \leq n$. Using Proposition 17 we have for claims :

(i) if n even: $w_{n+1} = (w_{n-1}w_{n-2})^{\frac{b-1}{2}} w_{n-1}$; since $w_{n-1}w_{n-2} \in \mathcal{L}$ with w_{n-1} as a proper prefix, we deduce the Lyndon factorization L_{n+1}.

(ii) if n odd: $w_{n+1} = w_{n-1}(w_{n-2}w_{n-1})^{\frac{b-1}{2}}$, and $w_{n-2}w_{n-1} \in \mathcal{L}$ with w_{n-1} as a proper suffix. It follows that $w_{n-1} \succeq w_{n-2}w_{n-1}$ and that the last factor of L_{n-1}, $w_{n-4}w_{n-3}$, is greater than $w_{n-2}w_{n-1}$, since $w_{n-4}w_{n-3} \succeq w_{n-2}$ and $w_{n-2} \notin \mathrm{Pref}(w_{n-4}w_{n-3})$. The conclusion follows. □

We are now in a position to state the main result about the Lyndon factorization of the minimal infinite smooth word $m_{\{1,b\}}$.

Theorem 19 *Let* $\Sigma = \{1, b\}$. *Then :*

(i) $m_{\{1,b\}} \in \mathcal{L}_\infty$;

(ii) *the Lyndon factorization of* $\Delta(m_{\{1,b\}})$ *is an infinite sequence of finite Lyndon words.*

Proof. It suffices to take the limit as $n \to \infty$ of the statements in Proposition 18. □

4.2 Letter frequencies

The Dekking conjecture about the frequency of 1's in the Kolakoski word still holds, but is solved for the minimal word on $\Sigma = \{1, b\}$.

Theorem 20 *Let* $\Sigma = \{1, b\}$. *Then the frequency of* b*'s in* m_Σ *is*

$$d_b(m_\Sigma) = \frac{1}{\sqrt{2b-1}+1}. \tag{5}$$

Proof. By Theorem 12 and Proposition 17 i), w_{2n} is a prefix of $m_{\{1,b\}}$ for all $n \geq 1$ and we have the following recursive definition of $m_{\{1,b\}}$:

$$\begin{aligned} w_1 &= b;\ w_2 = 1^b;\ w_3 = (b1)^{\frac{b-1}{2}} b; \\ w_k &= (w_{k-2}w_{k-3})^{\frac{b-1}{2}} w_{k-2}; \\ m_{\{1,b\}} &= \lim_{n\to\infty} w_{2n}. \end{aligned}$$

Putting $f_n = |w_n|_b$ and $g_n = |w_n|_1$, the recursive definition of w_n yields the following recursive definitions for the number of occurences f_n and g_n:

$$f_n = \frac{b-1}{2}(f_{n-2} + f_{n-3}) + f_{n-2} = \frac{b+1}{2} f_{n-2} + \frac{b-1}{2} f_{n-3}, \tag{6}$$

with the initial conditions $f_1 = 1$, $f_2 = 0$, $f_3 = \frac{b+1}{2}$, and

$$g_n = \frac{b+1}{2} g_{n-2} + \frac{b-1}{2} g_{n-3}, \tag{7}$$

with the initial conditions $g_1 = 0$, $g_2 = b$ and $g_3 = \frac{b-1}{2}$. To complete this proof, it suffices to solve the recurrences.

Equation (6): the characteristic polynomial associated to the recurrence f_n is

$$z^3 - \frac{b+1}{2} z - \frac{b-1}{2} = 0,$$

which can be written as

$$(z+1)\left(z - \frac{1+\sqrt{2b-1}}{2}\right)\left(z - \frac{1-\sqrt{2b-1}}{2}\right) = 0.$$

It follows that $f_n = c_1(-1)^n + c_2(\frac{1+\sqrt{2b-1}}{2})^n + c_3(\frac{1-\sqrt{2b-1}}{2})^n$, with $c_1, c_2, c_3 \in \mathbb{R}$, except for $b = 5$ since the roots of the polynomial are $-1, -1$ and 2 and then, $f_n = c_1(-1)^n + c_2 n(-1)^n + c_3(2)^n$. This case will be done later.

Using the initial conditions, we find

$$c_1 = \frac{2}{b-5}, \quad c_2 = \frac{b+\sqrt{2b-1}}{\sqrt{2b-1}(1+b+2\sqrt{2b-1})}, \quad c_3 = -\frac{b-2+\sqrt{2b-1}}{\sqrt{2b-1}(b-5)}.$$

We then have a closed formula for f_n.

Equation (7): proceeding in the same way, we find for $b \neq 5$:

$$g_n = c_1'(-1)^n + c_2'\left(\frac{1+\sqrt{2b-1}}{2}\right)^n + c_3'\left(\frac{1-\sqrt{2b-1}}{2}\right)^n,$$

with

$$c_1' = -\frac{b+1}{b-5}, \quad c_2' = \frac{b\sqrt{2b-1}+2b-1}{\sqrt{2b-1}(1+b+2\sqrt{2b-1})},$$

and

$$c_3' = \frac{2b-1+\sqrt{2b-1}(b-2)}{\sqrt{2b-1}(b-5)}.$$

Now, the frequency of b's is given by

$$\lim_{n\to\infty} \frac{f_{2n}}{f_{2n}+g_{2n}} = \frac{1}{\sqrt{2b-1}+1}.$$

For $b = 5$, using the initial conditions, we find $c_1 = \frac{-2}{9}$, $c_2 = \frac{-1}{3}$ and $c_3 = \frac{2}{9}$, and $c'_1 = \frac{1}{3}$, $c'_2 = 1$ and $c'_3 = \frac{2}{3}$. Then, $\lim_{n\to\infty} \frac{f_{2n}}{f_{2n}+g_{2n}} = \frac{1}{4}$, which is equal to $\frac{1}{\sqrt{2\cdot 5-1}+1}$. $\square$

5 Extremal words over even alphabets

In this section, we assume that the letters of $\Sigma = \{a, b\}$ are both even integers and such that $a < b$. Let start by a useful lemma.

Lemma 21 *If $w \in \Sigma^+$ then for all $\alpha \in \Sigma$, $|\Delta_\alpha^{-1}(w)|$ has even length.*

Proof. Let $|w| = n$. Applying Δ_α^{-1} to w yields:

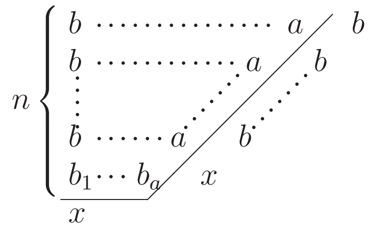
$$\begin{aligned}
\Delta_\alpha^{-1}(w) &= \Delta_\alpha^{-1}(w[1]w[2]\cdots w[n]) \\
&= \Delta_\alpha^{-1}(w[1])\Delta_{\overline{\alpha}}^{-1}(w[2])\cdots\Delta_\beta^{-1}(w[n]) \\
&= \alpha^{w[1]}\overline{\alpha}^{w[2]}\cdots\beta^{w[n]}
\end{aligned}$$

where $\beta = \alpha$ if n is odd and $\beta = \overline{\alpha}$ if n is even. Since $|\Delta_\alpha^{-1}(w)| = \sum_{i=1}^{n} w[i]$ the result follows. $\square$

As for odd alphabets, any extremal word w over even alphabets is characterized by the periodicity of $\Phi(w)$:

Theorem 22 $\Phi(M_{\{a,b\}}) = b^\omega$.

Proof. We proceed by induction on the length of the prefixes of $u = \Phi(M_{\{a,b\}})$. First, $M_{\{a,b\}}$ starts with the prefix b^b and $\Phi(M_{\{a,b\}})[1] = b$. One easily checks that $\Phi^{-1}(bb) \succeq \Phi^{-1}(bas)$ for any $s \in \Sigma$: indeed, $\Phi^{-1}(bb) = b^b$ and $\Phi^{-1}(bas)$ begins with $b^a a$. Assume now that $\Phi^{-1}(b^k)$ is maximal, for every $k \leq n$. Set $v = \Delta_b^{-1}(x)$. It follows that if $\Phi(M_{\{a,b\}})[n+1] = x$ then $v = \Delta_b^{-1}(x) = b^x$ and consequently $v[x+1] = a$. We have the following situation

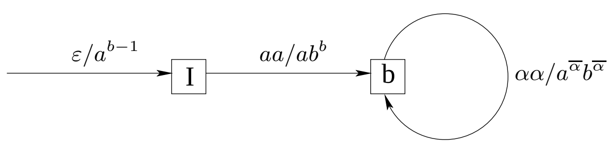


where each prefix is of even length by Lemma 21, and therefore ends with the letter a. Next, using the Glueing Lemma (see Lemma 7), the letter x should be the one that makes the word $\Phi^{-1}(b^{n-1}x)$ the greatest. By induction hypothesis, it follows that $x = b$. $\square$

The equality $\Delta(m_{\{a,b\}}) = \Delta(M_{\{a,b\}})$ yields :

Corollary 23 $\Phi(m_{\{a,b\}}) = ab^{\omega}$.

Therefore, $M_{\{a,b\}} = \Delta(m_{\{a,b\}})$ and hence is the generalized Kolakoski word $K_{(b,a)}$. This last property yields a linear time algorithm generating prefixes of the minimal (hence the maximal by simply permuting the letters) infinite smooth word for an even alphabet, represented by the following transducer, where $\alpha \in \{a, b\}$.



This transducer has two cycles (one for each letter) with same base state, and therefore any infinite path runs through these two cycles. Since an equal number of a's and b's are written in each cycle, the frequency of both letters is $\frac{1}{2}$. This again is a surprising fact: for the well-known Kolakoski word $K_{(1,2)}$ it is still a challenging conjecture. Indeed, the best known bound is 0.50084 and is due to Chvátal [7], who designed an ingenious procedure for computing an approximation of the frequency.

The analogue of Lemma 11, showing the palindromic structure of the prefixes of smooth words on odd alphabets, is given now for even alphabets, where prefixes are repetitions.

Lemma 24 *For all $u \in \Sigma^{\geq 2}$, there exists $p \in \Sigma^{2m}$ such that $\Phi^{-1}(u) = p^{\frac{u[k]}{2}}$.*

Proof. Let $w = \Phi^{-1}(u)$. We proceed by induction on the length of u. If $|u| = 2$, then $u = \alpha\beta$, $\alpha, \beta \in \Sigma$ and $w = \alpha^{\beta} = (\alpha\alpha)^{\frac{\beta}{2}}$, thus $p = \alpha\alpha$. Assume now that the statement holds for every u such that $|u| \leq k$. Let $v \in \Sigma^*$ be such that $|v| = k{+}1$. Then $\Phi^{-1}(v) = \Delta^{-1}_{v[1]}(\Phi^{-1}(v[2..k{+}1]))$, and by induction hypothesis, we have for an even length p

$$v = \Delta^{-1}_{v[1]}(p^{\frac{u[k+1]}{2}}),$$

which may be written as $v = (\Delta^{-1}_{v[1]}(p))^{\frac{u[k+1]}{2}}$. Then Lemma 11 applies and the conclusion follows. $\square$

This property may be used to show that extremal words are recurrent by adapting the proof provided in the case of odd alphabets. In fact the recurrence property holds for all infinite smooth words including the generalized Kolakoski words $K_{(b,a)}$.

Theorem 25 *Smooth words are recurrent.*

Proof. Let $u \in \Sigma^\omega$ and $w = \Phi^{-1}(u)$. Let $f \in F(w)$. Let n be an index such that $p = \Phi^{-1}(u[1..n])$ contains f as a factor. Let $q = \Phi^{-1}(u[1..(n+2)])$ and set $\alpha = u[n+1]$ and $\beta = u[n+2]$. By definition

$$\Delta^n(q) = \Delta^{-2}_{\alpha,u[n]}(\beta) = \Delta^{-1}_{u[n]}(\alpha\alpha) \cdot x$$

where $\Delta^{-1}_{u[n]}(\alpha\alpha)$ ends with the letter $\overline{u[n]}$ and $x \in \Sigma^*$. Let q' be the prefix of q such that $\Delta^{n+1}(q') = \alpha\alpha$. Then $w = q'w'$ for some word w', and by using the Glueing Lemma, we have for every k such that $0 \leq k \leq n$

$$\Delta^k(w) = \Delta^k(q') \cdot \Delta^k(w')$$

where $\Delta^k(q')$ starts with $u[k]$ and ends with $\overline{u[k]}$, by using the length parity of Lemma 21. It follows that $\Delta^k(w')$ starts with $u[k]$, and therefore, w' contains necessarily another occurrence of p, hence of f. $\square$

On the other hand, we have:

Proposition 26 *$F(m_{\{a,b\}})$ and $F(M_{\{a,b\}})$ are not closed under reversal and under complementation.*

Proof. Consider the prefix $p = (b^b a^b)^{b/2}$ of $M_{\{a,b\}}$ and assume that its reversal $\widetilde{p} = (a^b b^b)^{b/2} \in F(M_{\{a,b\}})$. Then, $\Delta(\widetilde{p}) = b^b$, would be a factor in $\Delta(M_{\{a,b\}})$ coding $\widetilde{p}$ in $M_{\{a,b\}}$. By Lemma 21, any factor a^a, a^b, b^a, b^b in $\Delta(M_{\{a,b\}})$ codes a factor in $M_{\{a,b\}}$ starting by b and ending by a. Contradiction. For the non closure under complementation, it suffices to observe that $\widetilde{p} = \overline{p}$. $\square$

We establish now that minimal smooth words over even alphabets are infinite Lyndon words. Some technical lemmas are required.

Lemma 27 *Let* $w_n = \Phi^{-1}(b^n)$. *Then,* $w_n = (v_1^{b/2} v_2^{b/2})^{b/2}$, $v_2 \prec v_1$, $v_2 \notin \mathrm{Pref}(v_1)$ *and* $|v_1|$, $|v_2|$ *are even, for* $n \geq 3$.

Proof. We proceed by induction on n. By direct computation, we have

$$w_1 = b, \quad w_2 = b^b \quad \text{and} \quad w_3 = (b^b a^b)^{b/2} = ((bb)^{b/2}(aa)^{b/2})^{b/2}.$$

Since $aa \prec bb$, $aa \notin \mathrm{Pref}(bb)$, $|aa|$ and $|bb|$ even, the property is verified for $n = 3$. Assume now that the statement is true for all $k \leq n$. Then,

$$w_{n+1} = \Delta_b^{-1}(w_n) = \Delta_b^{-1}((v_1^{b/2} v_2^{b/2})^{b/2}) = [(\Delta_b^{-1}(v_1))^{b/2}(\Delta_b^{-1}(v_2))^{b/2}]^{b/2},$$

with $|\Delta_b^{-1}(v_1)|$ and $|\Delta_b^{-1}(v_2)|$ even, by Lemma 21. As w_n is prefix of w_{n+1}, $\Delta_b^{-1}(v_2) \prec \Delta_b^{-1}(v_1)$ and $\Delta_b^{-1}(v_2) \notin \mathrm{Pref}(\Delta_b^{-1}(v_1))$. $\square$

Notation. As $w_n = (v_1^{b/2} v_2^{b/2})^{b/2}$ for all $n \geq 3$, $\underline{w_n}$ denotes the word $(v_1^{a/2} v_2^{a/2})^{b/2}$.

Lemma 28 *Let* $w_n = \Phi^{-1}(b^n)$. *For* $n \geq 4$,

(i) $w_n = (w_{n-1}\underline{w_{n-1}})^{b/2}$;

(ii) $\underline{w_{n-1}} \prec w_{n-1}$, $\underline{w_{n-1}} \notin \mathrm{Pref}(w_{n-1})$ *and* $|w_{n-1}|$, $|\underline{w_{n-1}}|$ *are even;*

(iii) $\overline{u}_1, \overline{u}_2 \in \mathcal{L}$, *where* $w_n = u_1^{b/2}$ *and* $\underline{w_n} = u_2^{b/2}$.

Proof. We proceed by induction on n. By direct computation, we get

$$w_1 = b, \quad w_2 = b^b, \quad w_3 = (b^b a^b)^{b/2}, \quad w_4 = ((b^b a^b)^{b/2}(b^a a^a)^{b/2})^{b/2}.$$

Then, $\underline{w_3} = \underline{(b^b a^b)^{b/2}} = (b^a a^a)^{b/2}$ and

$$w_4 = ((b^b a^b)^{b/2} \cdot (b^a a^a)^{b/2})^{b/2} = (w_3 \cdot \underline{w_3})^{b/2}.$$

Thus, i) is verified for $n = 4$. Since $\underline{w_3} = (b^a a^a)^{b/2} \prec (b^b a^b)^{b/2} = w_3$, $\underline{w_3} \notin \mathrm{Pref}(w_3)$, $|w_3| = b^2$ and $|\underline{w_3}| = ab$, ii) is also verified. Finally, $w_4 = (u_1)^{b/2}$, $u_1 = (b^b a^b)^{b/2}(b^a a^a)^{b/2}$, $\underline{w_4} = (u_2)^{b/2}$, $u_2 = (b^b a^b)^{a/2}(b^a a^a)^{a/2}$ and $\overline{u_1}, \overline{u_2} \in \mathcal{L}$. Assume now the 3 statements true for all $k \leq n$.

(i) $w_{n+1} = \Phi^{-1}(b^{n+1}) = \Delta_b^{-1}(\Phi^{-1}(b^n)) = \Delta_b^{-1}((w_{n-1}\underline{w_{n-1}})^{b/2})$. Since $|w_{n-1}|$ and $|\underline{w_{n-1}}|$ are even by hypothesis, we get

$$w_{n+1} = \Delta_b^{-1}((w_{n-1}\underline{w_{n-1}})^{b/2}) = [\Delta_b^{-1}(w_{n-1})\Delta_b^{-1}(\underline{w_{n-1}})]^{b/2}.$$

Let $w_{n-1} = (v_3^{b/2} v_4^{b/2})^{b/2}$. Then, $\underline{w_{n-1}} = (v_3^{a/2} v_4^{a/2})^{b/2}$ and

$$w_n = \Delta_b^{-1}(w_{n-1}) = [(\Delta_b^{-1}(v_3))^{b/2} (\Delta_b^{-1}(v_4))^{b/2}]^{b/2}$$

and

$$\Delta_b^{-1}(\underline{w_{n-1}}) = [(\Delta_b^{-1}(v_3))^{a/2} (\Delta_b^{-1}(v_4))^{a/2}]^{b/2} = \underline{w_n}.$$

Thus, $w_{n+1} = (w_n \underline{w_n})^{b/2}$.

(ii) By Lemma 27 and i), $w_n = (v_1^{b/2} v_2^{b/2})^{b/2}$ and $\underline{w_n} = (v_1^{a/2} v_2^{a/2})^{b/2}$ with $v_1^{b/2} = w_{n-1}$ and $v_2^{b/2} = \underline{w_{n-1}}$. By hypothesis, $\underline{w_{n-1}} \prec w_{n-1}$ and $\underline{w_{n-1}} \notin \mathrm{Pref}(w_{n-1})$. Since $v_2 \notin \mathrm{Pref}(v_1)$ and $v_2 \prec v_1$, we have:

$$\underline{w_n} = v_1^{a/2} v_2^{a/2} s \prec v_1^{a/2} v_1^{(b-a)/2} s' = w_n, \;\; s, s' \in \Sigma^*.$$

We also have that $\underline{w_n} \notin \mathrm{Pref}(w_n)$ and their lengths are respectively $ab(v_1 + v_2)/2$ and $b^2(v_1 + v_2)/2$, which are even.

(iii) Using i) and ii), we get $w_{n+1} = (w_n \underline{w_n})^{b/2} = u_3^{b/2}$, with $\underline{w_n} \prec w_n$. Then, $\overline{u_3} = \overline{w_n \underline{w_n}} = \overline{w_n} \cdot \overline{\underline{w_n}}$. By (1), we get $\underline{w_n} \prec w_n \iff \overline{w_n} \prec \overline{\underline{w_n}}$. By hypothesis, $w_n = u_1^{b/2}, \underline{w_n} = u_2^{b/2}$, with $\overline{u_1}, \overline{u_2} \in \mathcal{L}$. Using Corollary 4, we get that $\overline{u_3} \in \mathcal{L}$. Consider now $\overline{u_4}$ satisfaying $\underline{w_{n+1}} = u_4^{b/2}$. Using Lemma 27, we know that $u_3 = v_1^{b/2} v_2^{b/2}$, and then, that $u_4 = v_1^{a/2} v_2^{a/2}$, with $v_2 \prec v_1$. Hence, $\overline{u_4} = \overline{v_1}^{a/2} \overline{v_2}^{a/2}$, with $\overline{v_1} \prec \overline{v_2}$, and using Corollary 4, the conclusion follows. □

Theorem 29 $m_{\{a,b\}} \in \mathcal{L}_\infty$.

Proof. By Theorem 22, $\Phi(M_{\{a,b\}}) = b^\omega$. Let $w_n = \Phi^{-1}(b^n)$. Then,

$$M_{\{a,b\}} = \lim_{n \to \infty} w_n.$$

From Lemma 28, we know that $w_n = (w_{n-1} \underline{w_{n-1}})^{b/2}$, with $\overline{w_{n-1} \underline{w_{n-1}}} \in \mathcal{L}$. Since

$$m_{\{a,b\}} = \overline{M_{\{a,b\}}} = \lim_{n \to \infty} (\overline{w_n}),$$

and $|w_n| < |w_{n+1}|$, the conclusion follows. □

6 Concluding remarks

The frequency of letters in an infinite smooth word over $\{1, 2\}$ is a still unsolved conjecture. Nevertheless for even alphabets this frequency is 0.5 for the extremal words. For odd alphabets of the type $\{1, b\}$, the inductive formulas in Proposition 17 enable us to compute the frequency for extremal words.

Moreover, the work presented here raises a number of questions. It is quite surprising that for alphabets of same parity, some of the Dekking conjectures are rather easy to prove: recurrence, frequency for extremal words on even alphabets, closure by reversal for odd alphabets. The frequency problem remains open for odd alphabets, as well as all the conjectures for the alphabet $\{1,2\}$, an instance of a different parities alphabet. The results presented here beg for an investigation of smooth words on different parities: study of the extremal words, combinatorial properties, Lyndon factorizations, closure properties, and so on. In another direction it would be interesting to compute the complexity function $P(n)$ in the way Weakley did for the alphabet $\{1,2\}$. The case of larger k-letter alphabets is also challenging.

Acknowledgements. This paper is an extended and enhanced version of a paper presented in Rennes (France) during the 11-th conference JM'06 [12]. We are also grateful to the anonymous referee for his valuable comments that improved paper's readability.

References

[1] A. Bergeron-Brlek, S. Brlek, A. Lacasse, X. Provençal, Patterns in Smooth Tilings, Proc. WORDS'03, 4-th International Conference on Words Turku (Finland) September 10–13 2003 TUCS General Publication No 27, August (2003) pp. 370–381.

[2] V. Berthé, S. Brlek, P. Choquette, Smooth words over arbitrary alphabets, Theoret. Comput. Sci. 341 (2005) 293–310.

[3] S. Brlek, Enumeration of factors in the Thue-Morse word, Discrete Appl. Math. 24 (1989) 83–96.

[4] S. Brlek, A. Ladouceur, A note on differentiable palindromes, Theoret. Comput. Sci. 302 (2003) 167–178.

[5] S. Brlek, S. Dulucq, A. Ladouceur, L. Vuillon, Combinatorial properties of smooth infinite words, Theoret. Comput. Sci. 352 (2006) 306–317.

[6] A. Carpi, On repeated factors in C^∞-words, Inform. Process. Lett. 52 (1994) 289–294.

[7] V. Chvátal, Notes on the Kolakoski Sequence, DIMACS Techn. Rep. 93-84, February 1994.

[8] F. M. Dekking, On the structure of self generating sequences, Séminaire de théorie des nombres de Bordeaux, exposé 31 (1980–1981).

[9] D. Jamet, G. Paquin, Discrete surfaces and infinite smooth words, FPSAC'05 - 17th annual International conference on Formal Power Series and Algebraic Combinatorics, Taormina (Sicile) June 20–25, 2005.

[10] W. Kolakoski, Self Generating Runs, Problem 5304, Amer. Math. Monthly 72 (1965) 674. Solution: Amer. Math. Monthly 73 (1966) 681–682.

[11] M. Lothaire, Combinatorics on words, Cambridge University Press, Cambridge, 1983.

[12] G. Paquin, S. Brlek, D. Jamet, Extremal generalized smooth words, JM'06 - 11ième Journées Montoises d'Informatique Théorique, Rennes (France) August 30th - September 2, 2006.

[13] G. Paquin, G. Melançon, S. Brlek, Properties of the extremal infinite smooth words, J. of Automata, Languages and Combinatorics, to appear.

[14] N. Pytheas-Fogg, Substitutions in Dynamics, Arithmetics and Combinatorics, LNM 1794, Springer, Berlin, 2002.

[15] R. Siromoney, L. Matthew, V. R. Dare, K. G. Subramanian, Infinite lyndon words, Inform. Process. Lett. 50 (1994) 101–104.

[16] W. D. Weakley, On the number of C^∞-words of each length, J. Combinatorial Theory Series A 51 no. 1 (1989) 55–62.