The constant of recognizability is computable for primitive morphisms

Fabien Durand
Laboratoire Amiénois de Mathématiques Fondamentales et Appliquées,
CNRS-UMR 7352,
Université de Picardie Jules Verne,
33 rue Saint Leu, 80000 Amiens,
France
fabien.durand@u-picardie.fr

Julien Leroy
Institut de mathématique,
Université de Liège
Allée de la découverte 12 (B37)
4000 Liège,
Belgium
J.Leroy@ulg.ac.be

Abstract
Mossé proved that primitive morphisms are recognizable. In this paper we give a computable upper bound for the constant of recognizability of such a morphism. This bound can be expressed only using the cardinality of the alphabet and the length of the longest image under the morphism of a letter.

1 Introduction
Infinite words, i.e., infinite sequences of symbols from a finite set, usually called alphabet, form a classical object of study. They have an important representation power: they provide a natural way to code elements of an infinite set using finitely many symbols, e.g., the coding
of an orbit in a discrete dynamical system or the characteristic sequence of a set of integers. A rich family of infinite words, with a simple algorithmic description, is made of the words obtained by iterating a morphism $\sigma : A^* \to A^*$ [2], where $A^*$ is the free monoid generated by the finite alphabet $A$.

If $\sigma$ is prolongable on some letter $a \in A$, that is, if $\sigma(a) = au$ for some non-empty word $u$ and $\lim_{n \to +\infty} |\sigma^n(a)| = +\infty$, then $\sigma^n(a)$ converges to an infinite word $x = \sigma^\omega(a) \in A^\mathbb{N}$ that is a fixed point of $\sigma$. Two-sided fixed points are similarly defined as infinite words of the form $\sigma^\omega(a \cdot b) \in A^\mathbb{Z}$, where $\sigma(a) = ua$ and $\sigma(b) = bv$ with $u, v \in A^+$ and $\lim_{n \to +\infty} |\sigma^n(a)| = \lim_{n \to +\infty} |\sigma^n(b)| = +\infty$. Such a fixed point is said to be admissible if $ab$ occurs in $\sigma^n(c)$ for some $n \in \mathbb{N}$ and some $c \in A$. When the morphism is primitive, i.e., there exists $k \in \mathbb{N}$ such that $b$ occurs in $\sigma^k(c)$ for all $b, c \in A$, then $x$ is uniformly recurrent: any finite word that occurs in $x$ occurs infinitely many times in it and with bounded gaps [15]. The converse almost holds true: if $x = \sigma^\omega(a)$ is uniformly recurrent, then there exist a primitive morphism $\varphi : B^* \to B^*$, a letter $b \in B$ and a morphism $\psi : B^* \to A^*$ such that $x = \psi(\varphi^\omega(b))$ [3]. We let $\mathcal{L}(x)$ denote the set of factors of $x$, i.e., $\mathcal{L}(x) = \{u \in A^* \mid \exists p \in A^*, w \in A^\mathbb{N} : x = puw\}$ (with a similar definition of two-sided fixed points).

Recognizability is a central notion when dealing with fixed point of morphisms. It is linked to existence of long powers $u^k$ in $\mathcal{L}(x)$ [12]. An infinite word $x \in A^\mathbb{Z}$ is said to be $k$-power-free if there is no non-empty word $u$ such that $u^k$ belongs to $\mathcal{L}(x)$. We refer, for example, to [4, 7, 1, 6]. It roughly means that any long enough finite word that occurs in $\sigma^\omega(a)$ has a unique pre-image under $\sigma$, except for a prefix and a suffix of bounded length which is called the constant of recognizability. A fundamental result concerning recognizability is due to Mossé who proved that aperiodic primitive morphisms (i.e., primitive morphisms with aperiodic fixed points) are recognizable [13, 14]. In this paper, we present a detailed proof of this result. This allows us to give a bound on the constant of recognizability.

## 2 Recognizability

Given a morphism $\sigma : A^* \to A^*$, we respectively define $|\sigma|$ and $\langle \sigma \rangle$ by

$$|\sigma| = \max_{a \in A} |\sigma(a)|, \quad \text{and}, \quad \langle \sigma \rangle = \min_{a \in A} |\sigma(a)|.$$

Assuming that $\sigma$ has an admissible fixed point $x \in A^\mathbb{Z}$, for all $p \in \mathbb{N}$, we let $f_x^{(p)}$ denote the function

$$f_x^{(p)} : \mathbb{Z} \to \mathbb{Z}, i \mapsto f_x^{(p)}(i) = \begin{cases} |\sigma^p(x[i,0])| & \text{if } i > 0, \\ 0 & \text{if } i = 0, \\ |\sigma^p(x[i,0])| & \text{if } i < 0. \end{cases}$$

We set $E(x, \sigma^p) = f_x^{(p)}(\mathbb{Z})$. When it is clear from the context, we simply write $f^{(p)}$ instead of $f_x^{(p)}$. 

2
Given two integers $i, j$ with $i \leq j$, we let $x_{[i,j]}$ and $x_{[i,j]}$ respectively denote the factors $x_ix_{i+1}\cdots x_j$ and $x_ix_{i+1}\cdots x_{j-1}$ (with $x_{[i,i]} = \varepsilon$, where $\varepsilon$ is the empty word, i.e., the neutral element of $A^*$).

\textbf{Definition 1.} We say that $\sigma$ is \textit{recognizable} on $x$ if there exists some constant $L > 0$ such that for all $i, m \in \mathbb{Z}$,

$$(x_{[m-L,m+L]} = x_{[f^{(1)}(i)-L,f^{(1)}(i)+L]}) \Rightarrow (\exists j \in \mathbb{Z})((m = f^{(1)}(j)) \land (x_i = x_j)).$$

The smallest $L$ satisfying this condition is called the \textit{constant of recognizability} of $\sigma$ for $x$. When $\sigma$ is recognizable on all its admissible fixed points, we say that it is recognizable and its constant of recognizability is the greatest one.

\textbf{Lemma 2.} If $\sigma : A^* \to A^*$ is recognizable on the admissible fixed point $x \in A^\mathbb{Z}$ and if $L$ is the constant of recognizability of $\sigma$ for $x$, then for all $k > 0$, $x$ is also an admissible fixed point of $\sigma^k$ and $\sigma^k$ is recognizable on $x$ and its constant of recognizability for $x$ is at most $L|\sigma|^k-1$.

\textbf{Proof.} The result holds by induction on $k > 0$. The infinite word $x$ is obviously an admissible fixed point of $\sigma^k$. With $L' = L|\sigma|^k-1$, let us show that for all $i \in \mathbb{Z}$, the word

$$x_{[f^{(k)}(i)-L',f^{(k)}(i)+L']}$$

uniquely determines the letter $x_i$.

By recognizability, the word $x_{[f^{(k)}(i)-L',f^{(k)}(i)+L']} \Rightarrow x_{[f^{(k)}(i)-L',f^{(k)}(i)+L']} \Rightarrow x_{[f^{(k)}(i)-L'',f^{(k)}(i)+L'']}$, where $L'' = L|\sigma|^k-1$. \hfill \Box

\textbf{Theorem 3.} Let $\sigma : A^* \to A^*$ be an aperiodic primitive morphism and let $x \in \mathbb{Z}$ be an admissible fixed point of $\sigma$.

1. [13] There exists $M > 0$ such that, for all $i, m \in \mathbb{Z}$,

$$x_{[f^{(1)}(i)-M,f^{(1)}(i)+M]} = x_{[m-M,m+M]} \Rightarrow m \in E(x,\sigma).$$

2. [14] There exists $L > 0$ such that, for all $i, j \in \mathbb{Z}$,

$$x_{[f^{(1)}(i)-L,f^{(1)}(i)+L]} = x_{[f^{(1)}(j)-L,f^{(1)}(j)+L]} \Rightarrow x_i = x_j.$$
By a careful reading of the proofs of Mossé’s results, we can improve it as follows. The proof is given in Section 3. For an infinite word \( x \in A^\mathbb{Z} \), we let \( p_x : \mathbb{N} \rightarrow \mathbb{N} \) denote the complexity function of \( x \) defined by \( p_x(n) = |\mathcal{L}_n(x)| \) where \( \mathcal{L}_n(x) = (\mathcal{L}(x) \cap A^n) \).

**Theorem 4.** Let \( \sigma : A^* \rightarrow A^* \) be a morphism with an admissible fixed point \( x \in A^\mathbb{Z} \). If \( x \) is \( k \)-power-free and if there is some constant \( N \) such that for all \( n \in \mathbb{N} \), \( |\sigma^n| \leq N(\sigma^n) \), then \( \sigma \) is recognizable on \( x \) and its constant of recognizability for \( x \) is at most \( R|\sigma^dQ| + |\sigma^d| \), where

- \( R = \lceil N^2(k + 1) + 2N \rceil \); 
- \( Q = 1 + p_x(R) \left( \sum_{\frac{N}{2} \leq i \leq RN + 2} p_x(i) \right) \); 
- \( d \in \{1, 2, \ldots, \#A\} \) is such that for any words \( u, v \in \mathcal{L}(x) \), we have 
  \[ \sigma^{d-1}(u) \neq \sigma^{d-1}(v) \Rightarrow \forall n, \sigma^n(u) \neq \sigma^n(v). \]

Then, we give some computable bounds for \( N, R, k, Q \) and \( d \) in the case of primitive morphisms. These bounds are not sharp but can be expressed only using the cardinality of the alphabet and the maximal length \( |\sigma| \). The proof is given in Section 4.

**Theorem 5.** Any aperiodic primitive morphism \( \sigma : A^* \rightarrow A^* \) that admits a fixed point \( x \in A^\mathbb{Z} \) is recognizable on \( x \) and the constant of recognizability for \( x \) is at most 

\[
2|\sigma|^{6(\#A)^2 + 6(\#A)|\sigma|^{28(\#A)^2} + |\sigma|^{(\#A)}).
\]

The bound given in the previous theorem is far from being sharp. When the morphism \( \sigma \) is injective on \( \mathcal{L}(x) \) (which is decidable, see [5]), we can take \( d = 1 \) in Theorem 4 and the computation in the proof of Theorem 5 gives the bound 

\[
2|\sigma|^{6(\#A)^2 + 6|\sigma|^{28(\#A)^2} + |\sigma|}.
\]

The notion of recognizability is also known as circularity in the terminology of D0L-systems [9]. Assume that \( \sigma : A^* \rightarrow A^* \) is non-erasing and that \( a \in A \) is a letter such that the language \( \text{Fac}(\sigma, a) \) defined as the set of factors occurring in \( \sigma^n(a) \) for some \( n \) is infinite. Given a word \( u = u_1 \cdots u_{|u|} \in \mathcal{L}(a) \), we say that a triplet \((p, v, s)\) is an interpretation of \( u \) if \( \sigma(v) = \text{pus} \). Two interpretations \((p, v, s), (p', v', s')\) are said to be synchronized at position \( k \) if there exist \( i, j \) such that \( 1 \leq i \leq |v|, 1 \leq j \leq |v'| \) and 

\[
\sigma(v_1 \cdots v_i) = pu_1 \cdots u_k \quad \text{and} \quad \sigma(v'_1 \cdots v'_j) = p'u_1 \cdots u_k.
\]

The word \( u \) has a synchronizing point (at position \( k \)) if all its interpretations are synchronized (at position \( k \)). The pair \((\sigma, a)\) is said to be circular if \( \sigma \) is injective on \( \text{Fac}(\sigma, a) \) and if there is a constant \( C \), called the synchronizing delay of \( \sigma \), such that any word of length at least \( C \) has a synchronizing point. Thus, despite some considerations about whether we deal with fixed points or languages, recognizability and circularity are roughly the same notion and the synchronizing delay \( C \) is associated with the constant of recognizability \( L \) through the equation \( C = 2L + 1 \). Using the terminology of D0L-systems, Klouda and Medková obtained the following result which greatly improves our bounds, but for restricted cases.
Theorem 6 ([10]). If \#A = 2 and if \((\sigma, a)\) is circular with \(\sigma : A^* \rightarrow A^*\) a \(k\)-uniform morphism for some \(k \geq 2\), then the synchronizing delay \(C\) of \((\sigma, a)\) is bounded as follows:

1. \(C \leq 8\) if \(k = 2\),
2. \(C \leq k^2 + 3k - 4\) if \(k\) is an odd prime number,
3. \(C \leq k^2 \left(\frac{k}{d} - 1\right) + 5k - 4\) otherwise,

where \(d\) is the least divisor of \(k\) greater than 1.

3 Proof of Theorem 4

Like in Mossé’s original proof, the proof of Theorem 4 goes in two steps.

As a first step, we express the constant \(M\) of Theorem 3 in terms of the constants \(N, R, k\) and \(Q\) of Theorem 4. This is done in Proposition 8 with a proof following the lines of the proof of [11, Proposition 4.35]. The difference is that we take care of all the needed bounds to express the constant of recognizability.

As a second step, we show that the constant \(L\) of Theorem 3 can be taken equal to \(M' + |\sigma^d|\), where \(d\) is as defined in Theorem 4 and \(M'\) is such that for all \(i, m \in \mathbb{Z}\),

\[
x[f^{(d)}(i) - M', f^{(d)}(i) + M'] = x|m - M, m + M| \implies m \in E(x, \sigma^d).
\]

We first start with the following lemma.

Lemma 7. Let \(\sigma : A^* \rightarrow A^*\) be a non-erasing morphism, \(u \in A^*\) be a word and \(n\) be a positive integer. If \(v = v_0 \cdots v_{t+1} \in A^*\) is a word of length \(t + 2\) such that \(\sigma^n(v[1, t])\) is a factor of \(\sigma^n(u)\), and \(\sigma^n(u)\) is a factor of \(\sigma^n(v)\), then

\[
\frac{|\sigma^n|}{|\sigma^n|} |u| - 2 \leq t \leq \frac{|\sigma^n|}{\langle \sigma^n \rangle} |u|.
\]

Proof. Indeed, since \(\sigma^n(v[1, t])\) is a factor of \(\sigma^n(u)\) we have \(t \langle \sigma^n \rangle \leq |\sigma^n(v[1, t])| \leq |\sigma^n(u)| \leq |u|/\langle \sigma^n \rangle\). Hence \(t \leq |u|/\langle \sigma^n \rangle\). Similarly, since \(\sigma^n(u)\) is a factor of \(\sigma^n(v)\), we have \(|u| \leq (t + 2)|\sigma^n|/\langle \sigma^n \rangle\). We thus have

\[
|u| \frac{\langle \sigma^n \rangle}{\sigma^n} - 2 \leq t \leq |u| \frac{\sigma^n}{\sigma^n}.
\]

\[
\square
\]

Proposition 8. Let \(\sigma : A^* \rightarrow A^*\) be a morphism with an admissible fixed point \(x \in A^\mathbb{Z}\). Assuming that \(x\) is \(k\)-power-free and that there is some constant \(N\) such that for all \(n \in \mathbb{N}\), \(|\sigma^n| \leq N\langle \sigma^n \rangle\), we consider the constants
• \( R = \lceil N^2(k+1) + 2N \rceil \);

• \( Q = 1 + p_x(R) \left( \sum_{\# \leq RN+2} p_x(i) \right) \).

The constant \( M = R|\sigma^Q| \) is such that for all \( i,m \in \mathbb{Z} \),

\[
x_{[f^{(1)}(i)-M, f^{(1)}(i)+M]} = x_{[m-M, m+M]} \implies m \in E(x, \sigma).
\]  

(1)

Proof. We follow the lines of the proof of Theorem 3 that is in [11]. Obviously, if \( l \) satisfies (1), then so does \( l' \) whenever \( l' \geq l \). Let us show that such an \( l \), with \( R|\sigma^Q| \), exists.

We proceed by contradiction, assuming that for all \( l \), there exist \( i,j \) such that \( x_{[i-l,i+l]} = x_{[j-l,j+l]} \) with \( i \in E(x, \sigma) \) and \( j \notin E(x, \sigma) \). For any integer \( p \) such that \( 0 < p \leq Q \), we consider the integer \( l_p = R|\sigma^p| \). Let \( i_p \) and \( j_p \) be some integers such that

\[
x_{[i_p-l_p, i_p+l_p]} = x_{[j_p-l_p, j_p+l_p]} \text{ with } i_p \in E(x, \sigma) \text{ and } j_p \notin E(x, \sigma).
\]

We let \( r_p \) and \( s_p \) denote the smallest integers such that

\[
\text{Card} ([i_p-r_p, i_p] \cap E(x, \sigma^p)) = \left\lceil \frac{R}{2} \right\rceil \quad \text{and} \quad \text{Card} ([i_p, i_p+s_p] \cap E(x, \sigma^p)) = \left\lfloor \frac{R}{2} \right\rfloor + 1.
\]

There is an integer \( i'_p \) such that

\[
f^{(p)}(i'_p) = i_p - r_p \quad \text{and} \quad f^{(p)}(i'_p + R) = i_p + s_p.
\]

We set

\[
u_p = x_{[i'_p, i'_p + R]}.
\]

We have \( \sigma^p(u_p) = x_{[i_p-r_p, i_p+s_p]} \).

Notice that any interval of length \( l_p \) contains at least \( R-1 \) elements of \( E(x, \sigma^p) \). We thus have \( i_p - l_p \leq i_p - r_p \leq i_p + s_p \leq i_p + l_p \). Consequently we also have

\[
x_{[j_p-l_p, j_p+s_p]} = \sigma^p(u_p).
\]

(2)

However \( j_p - r_p \) does not need to belong to \( E(x, \sigma^p) \). Let \( j'_p \) and \( t_p \) denote the unique integers such that

\[
f^{(p)}(j'_p) < j_p - r_p \leq f^{(p)}(j'_p + 1);
\]

\[
f^{(p)}(j'_p + t_p + 1) \leq j_p + s_p < f^{(p)}(j'_p + t_p + 2).
\]

(3)

Consequently \( \sigma^p(x_{[j'_p+1, j'_p+t_p+1]}) \) is a factor of \( \sigma^p(u_p) \) and \( \sigma^p(u_p) \) is a factor of \( \sigma^p(x_{[j'_p+1, j'_p+t_p+1]}) \).

By Lemma 7, we have

\[
R \left\langle \frac{\sigma^p}{|\sigma^p|} \right\rangle - 2 \leq t_p \leq R \left\langle \frac{\sigma^p}{|\sigma^p|} \right\rangle.
\]

(4)
Hence

\[ \frac{R}{N} - 2 \leq t_p \leq RN. \]

Let \( v_p = x[j'_p, j'_p + t_p + 1] \). The number of possible pairs of words \((u_p, v_p)\) is at most

\[ p_x(R) \left( \sum_{R / N - 2 \leq i \leq RN + 2} p_x(i) \right) < Q. \]

Therefore, there exist \( p \) and \( q \) in \([1, Q]\) such that \( p < q \) and \((u_p, v_p) = (u_q, v_q)\). In particular we also have \( t_p = t_q \). We write

\[ t = t_p, \quad u = u_p, \quad v = v_p, \quad \tilde{v} = x[j'_p + 1, j'_p + t]. \]

Using the above notation we recall that we have

\[ u = x[i'_p, i'_p + R], \quad v = x[j'_p, j'_p + t + 1]. \]

Let \( A_p, B_p, A_q \) and \( B_q \) be the words

\[ A_p = x[j_p - r_p, f(j'_p + 1)], \quad B_p = x[f(j'_p + t + 1), j_p + s_p], \quad A_q = x[j_q - r_q, f(j'_q + 1)], \quad B_q = x[f(j'_q + t + 1), j_q + s_q]. \]

We thus have

\[ x[j_p - r_p, j_p + s_p| = A_p \sigma^p(\tilde{v}) B_p \quad \text{and} \quad x[j_q - r_q, j_q + s_q| = A_q \sigma^q(\tilde{v}) B_q. \]

with, using (3),

\[ \max\{ |A_p|, |B_p| \} \leq |\sigma^p| \quad \text{and} \quad \max\{ |A_q|, |B_q| \} \leq |\sigma^q|. \]

From (2) and (7), we obtain

\[ \sigma^{q-p}(A_p) \sigma^q(\tilde{v}) \sigma^{q-p}(B_p) = A_q \sigma^q(\tilde{v}) B_q. \]

We claim that

\[ A_q = \sigma^{q-p}(A_p) \quad \text{(and hence \( B_q = \sigma^{q-p}(B_p) \)).} \]

If not, the word \( \sigma^q(\tilde{v}) \) has a prefix which is a power \( w^r \) with \( r = \left\lfloor \frac{|\sigma^q(\tilde{v})|}{||A_q| - |\sigma^{q-p}(A_p)||} \right\rfloor \). Since, using (4) and (8),

\[ |\sigma^q(\tilde{v})| \geq t(\sigma^q) \geq \left( \frac{R}{N} - 2 \right) \langle \sigma^q \rangle \quad \text{and} \quad ||A_q| - |\sigma^{q-p}(A_p)|| \leq |\sigma^q|, \]

...
we deduce from the choice of $R$ that $r \geq k + 1$, which contradicts the definition of $k$. We thus have $A_q = \sigma^{q-p}(A_p)$ and $B_q = \sigma^{q-p}(B_p)$.

We now show that

$$[j_q - r_q, j_q + s_q] \cap E(x, \sigma) = ([i_q - r_q, i_q + s_q] \cap E(x, \sigma)) - (i_q - j_q). \quad (10)$$

This will contradict the fact that $i_q$ belongs to $E(x, \sigma)$ and $j_q$ does not.

By (6), we have

$$\sigma^p(v) = x[\sigma^p(j_q), \sigma^p(j_q + t + 2)] = x[\sigma^p(j_q), \sigma^p(j_q + t + 2)].$$

Since $\sigma^p(u)$ is a factor of $\sigma^p(v)$, we deduce from (3) that there exists $m_q \in \mathbb{Z}$ such that

$$f^{(p)}(j_q') < m_q - r_p < m_q + s_p < f^{(p)}(j_q' + t + 2) \quad (11)$$

and

$$x_{[m_q - r_p, m_q + s_p]} = \sigma^p(u) = A_p \sigma^p(\tilde{\nu}) B_p.$$

By applying $\sigma^{q-p}$, we obtain

$$x_{[f^{(q-p)}(m_q - r_p), f^{(q-p)}(m_q + s_p)]} = A_q \sigma^q(\tilde{\nu}) B_q,$$

and, from (11),

$$f^{(q)}(j_q') < f^{(q-p)}(m_q - r_p) < f^{(q-p)}(m_q + s_p) < f^{(q)}(j_q' + t + 2).$$

As we also have

$$x_{[j_q - r_q, j_q + s_q]} = A_q \sigma^q(\tilde{\nu}) B_q$$

with, by (3),

$$f^{(q)}(j_q') < j_q - r_q \leq f^{(q)}(j_q' + 1) \leq f^{(q)}(j_q' + t + 1) \leq j_q + s_q < f^{(q)}(j_q' + t + 2),$$

we apply the same argument as to show (9) and get $j_q - r_q = f^{(q-p)}(m_q - r_p)$ (hence $j_q + s_q = f^{(q-p)}(m_q + s_p)$). We thus get that $j_q - r_q$ belongs to $E(x, \sigma^{q-p}) \subset E(x, \sigma)$. Since we also have

$$x_{[f^{(1)}(j_q - r_q), f^{(1)}(j_q + s_q)]} = \sigma^{q-p-1}(x_{[m_q - r_p, m_q + s_p]}) = \sigma^{q-p-1}(A_p \sigma^p(\tilde{\nu}) B_p),$$

$$x_{[f^{(1)}(i_q - r_q), f^{(1)}(i_q + s_q)]} = \sigma^{q-p-1}(x_{[i_q', i_q' + r_q]}) = \sigma^{q-p-1}(A_p \sigma^p(\tilde{\nu}) B_p),$$

we get

$$x_{[f^{(1)}(j_q - r_q), f^{(1)}(j_q + s_q)]} = x_{[f^{(1)}(i_q - r_q), f^{(1)}(i_q + s_q)]}$$

with $j_q - r_q, i_q - r_q$ belonging to $E(x, \sigma)$. By applying $\sigma$ to these two word, we thus obtain (10), which ends the proof. \qed
In Proposition 8, we compute a constant such that any long enough word can be cut into words in \(\sigma(A)\) in a unique way except for a prefix and a suffix of bounded length. However, it does not give information on the letters in \(A\) that the words in \(\sigma(A)\) come from. A key argument in Mossé’s original proof is to prove the existence of an integer \(d\) such that for all \(a, b \in A\), if \(\sigma^n(a) = \sigma^n(b)\) for some \(n\), then \(\sigma^d(a) = \sigma^d(b)\). We then prove that the constant \(L\) of Theorem 3 can be taken equal to \(M + |\sigma^{d+1}|\), where \(M\) is the constant of Proposition 8 associated with \(\sigma^{d+1}\). Theorem 9 below ensures that we can take \(d = \#A - 1\), which ends the proof of Theorem 4.

**Theorem 9** ([5, Theorem 3]). Let \(\sigma : A^* \to A^*\) be a morphism such that \(\sigma(A) \neq \{\varepsilon\}\). For any words \(u, v \in A^*\), we have

\[
\sigma^{\#A-1}(u) \neq \sigma^{\#A-1}(v) \Rightarrow \forall n, \sigma^n(u) \neq \sigma^n(v).
\]

We give the proof of Mossé’s second step result for the sake of completeness.

**Proposition 10.** Let \(\sigma : A^* \to A^*\) be a morphism with an admissible fixed point \(x \in A^\mathbb{Z}\). Let \(d \in \{1, 2, \ldots, \#A\}\) be such that for any words \(u, v \in \mathcal{L}(x)\),

\[
\sigma^{d-1}(u) \neq \sigma^{d-1}(v) \Rightarrow \forall n, \sigma^n(u) \neq \sigma^n(v).
\]

If \(M\) is a constant such that for all \(i, m \in \mathbb{Z}\),

\[
x[f^d(i) - M, f^d(i) + M] = x[m - M, m + M] \implies m \in E(x, \sigma^d),
\]

then \(\sigma\) is recognizable on \(x\) and its constant of recognizability for \(x\) is at most \(M + |\sigma^d|\).

**Proof.** Let \(i, m \in \mathbb{Z}\) such that

\[
x[f^{(1)}(i) - M - |\sigma^d|, f^{(1)}(i) + M + |\sigma^d|] = x[m - M - |\sigma^d|, m + M + |\sigma^d|].
\]

By definition of \(M\), there exists \(j \in \mathbb{Z}\) such that \(m = f^{(1)}(j)\). Our goal is to show that \(x_i = x_j\).

There exists \(k \in \mathbb{Z}\) such that

\[
f^{(1)}(i) - |\sigma^d| < f^{(d)}(k) \leq f^{(1)}(i) < f^{(d)}(k + 1) \leq f^{(1)}(i) + |\sigma^d|.
\]

In particular, this implies that \(f^{(d-1)}(k) \leq i < f^{(d)}(k + 1)\).

Consider \(c = f^{(1)}(i) - f^{(d)}(k)\) and \(d = f^{(d)}(k + 1) - f^{(1)}(i)\). We have

\[
x[f^{(d)}(k) - M, f^{(d)}(k) + M] = x[f^{(1)}(j) - c - M, f^{(1)}(j) - c + M];
\]

\[
x[f^{(d)}(k+1) - M, f^{(d)}(k+1) + M] = x[f^{(1)}(j) - d - M, f^{(1)}(j) + d + M].
\]

By definition of \(M\), there exists \(l \in \mathbb{Z}\) such that

\[
f^{(d)}(l) = f^{(1)}(j) - c \quad \text{and} \quad f^{(d)}(l + 1) = f^{(1)}(j) + d.
\]
We thus have \( f^{(d-1)}(l) \leq j < f^{(d-1)}(l + 1) \), and,

\[
x[f^{(d)}(k), f^{(d)}(k + 1)] = x[f^{(d)}(l), f^{(d)}(l + 1)].
\]

Hence \( \sigma^d(x_k) = \sigma^d(x_l) \). By definition of \( d \), we also have \( \sigma^{d-1}(x_k) = \sigma^{d-1}(x_l) \). Hence

\[
x[f^{(d-1)}(k), f^{(d-1)}(k + 1)] = x[f^{(d-1)}(l), f^{(d-1)}(l + 1)].
\]

Since we have \( f^{(1)}(i) - f^{(d)}(k) = f^{(1)}(j) - f^{(d)}(l) \), we also have \( i - f^{(d-1)}(k) = j - f^{(d-1)}(l) \). Hence \( x_i = x_j \).

\section{Proof of Theorem 5}

In this section, we show that the constants appearing in Theorem 4 can all be bounded by some computable constants. In all what follows, we assume that \( \sigma : A^* \to A^* \) is a primitive morphism. By taking a power of \( \sigma \) if needed, we can assume that it has an admissible fixed point \( x \in A^\mathbb{Z} \). Furthermore, we have \( \mathcal{L}(x) = \mathcal{L}(y) \) for all admissible fixed points \( y \) of \( \sigma \). We let \( \mathcal{L}(\sigma) \) denote this set. The constants appearing in Theorem 4 are thus the same whatever the admissible fixed point we consider and the morphism is recognizable.

With the morphism \( \sigma \), one associates its incidence matrix \( M_\sigma \) defined by \( (M_\sigma)_{a,b} = |\sigma(b)|_a \), where \( |u|_a \) denotes the number of occurrences of the letter \( a \) in the word \( u \).

\begin{lemma} [\cite{8}] \end{lemma}

A \( d \times d \) matrix \( M \) is primitive if, and only if, there is an integer \( k \leq d^2 - 2d + 2 \) such that \( M^k \) contains only positive entries.

Given an infinite word \( x \in A^\mathbb{Z} \) and a word \( u \in \mathcal{L}(x) \), a return word to \( u \) in \( x \) is a word \( r \) such that \( ru \) belongs to \( \mathcal{L}(x) \), \( u \) is a prefix of \( ru \) and \( ru \) contains exactly two occurrences of \( u \). The infinite word \( x \) is linearly recurrent if it is recurrent (all words in \( \mathcal{L}(x) \) appear infinitely many times in \( x \)) and there exists some constant \( K \) such that for all \( u \in \mathcal{L}(x) \), any return word to \( u \) has length at most \( K|u| \). The set of return words to \( u \) in \( x \) is denoted \( R_{x,u} \).

The next two results give bounds on the constants appearing in Theorem 4.

\begin{theorem} [\cite{4}] \end{theorem}

If \( x \in A^\mathbb{Z} \) is a aperiodic and linearly recurrent sequence (with constant \( K \)), then \( x \) is \((K + 1)\)-power-free and \( p_x(n) \leq Kn \) for all \( n \).

\begin{proposition} [\cite{3}] \end{proposition}

Let \( \sigma : A^* \to A^* \) be an aperiodic primitive morphism and \( x \) be one of its admissible fixed points. Then we have

\[
|\sigma^n| \leq |\sigma|^{(\#A)^2} < |\sigma|^{(\#A)^2}
\]

for all \( n \) and \( x \) is linearly recurrent for some constant

\[
K_\sigma < |\sigma|^{4(\#A)^2}.
\]

10
Proof. Durand [3] showed that the constant of linear recurrence $K_\sigma$ is at most equal to $RN|\sigma|$, where

- $N$ is a constant such that $|\sigma^n| \leq N(\sigma^n)$ for all $n$;
- $R$ is the maximal length of a return word to a word of length 2 in $L(\sigma)$.

We only prove here that $N \leq |\sigma|^{|A|^2}$ and $R \leq 2|\sigma|^{|A|^2}$. The constant of linear recurrence is thus at most $2|\sigma|^{1+3(#A)^2} < |\sigma|^{4(#A)^2}$.

Let us write $d = \#A$. By Lemma 11, the matrix $M_{d^2}$ contains only positive entries. For all $n \geq 0$ and all $a \in A$, we have $|\sigma^{n+d^2}(a)| = \sum_{b \in A} |\sigma^{d^2(a)}(a)| |\sigma^n(b)| \geq |\sigma^n|$. Since this is true for all $a$, we get $|\sigma^n| \leq (\sigma^{n+d^2}) \leq |\sigma^2| |(\sigma^n)|$, so $N \leq |\sigma^d|$.

Let $a \in A$ such that $\sigma$ is prolongable on $a$. Thus for all $n$, any word that occurs in $\sigma^n(a)$ also occurs in $\sigma^{n+1}(a)$. Let us show that for all $n > d^2$, any word $u \in L(\sigma)$ of length 2 occurs in $\sigma^n(a)$. For all $n$, the words of length 2 that occur in $\sigma^{n+1}(a)$ occur in images under $\sigma$ of the words of length 2 that occur in $\sigma^n(a)$. As any word occurring in $\sigma^n(a)$ also occurs in $\sigma^{n+1}(a)$, the words of length 2 that occurs in $\sigma^n(a)$ are those that occur in $\sigma^n(a)$ together with those occurring in the images under $\sigma$ of these words. Thus, if there is a word of length 2 that does not occur in $\sigma^n(a)$, there is a sequence $(u_1, u_2, \ldots, u_n)$ of words of length 2 in $L(\sigma)$ such that for all $i \leq n$, $u_i$ occurs in $\sigma^i(a)$ and does not occur in $\sigma^{i-1}(a)$. Hence all words $u_1, \ldots, u_n$ are distinct. For $n > d^2$, this is a contradiction since there are at most $d^2$ words of length 2 on the alphabet $A$. Thus, for any letter $b \in A$, all words $u \in L(\sigma)$ of length 2 occur in $\sigma^{2d^2}(b)$. We deduce that $R \leq 2|\sigma|^{2d^2}$.

\begin{proof}[Proof of Theorem 5] We just have to make the computation. Using Theorem 12, Proposition 13 and the notation of Theorem 4, we can take $d = \#A$ and we successively have

\begin{align*}
N & \leq 1 + K_\sigma \leq |\sigma|^{4d^2}, \\
R & = \lfloor N^2(k+1) + 2N \rfloor \leq |\sigma|^{2d^2}(|\sigma|^{4d^2} + 1) + 2|\sigma|^{d^2} \leq 2|\sigma|^{6d^2}, \\
Q & = 1 + p_\sigma(R), \quad \left( \sum_{\frac{-N}{d} \leq i \leq RN+2} p_\sigma(i) \right) \leq K_\sigma 2|\sigma|^{6d^2} \left( \sum_{0 \leq i \leq 2+2|\sigma|^{d^2}} i K_\sigma \right) \leq 6|\sigma|^{28d^2}
\end{align*}

We finally get that the constant of recognizability of $\sigma$ is at most

\[ 2|\sigma|^{|d^2|} |\sigma|^{6d|\sigma|^{28d^2}} + |\sigma|^d = 2|\sigma|^{6d^2+6d|\sigma|^{28d^2} + |\sigma|^d}. \]
\end{proof}

References

[1] V. Canterini and A. Siegel, Geometric representation of substitutions of pisot type, *Trans. Amer. Math. Soc.* 353 (2001), 5121–5144.
[2] C. Choffrut and J. Karhumäki, Combinatorics of words. In *Handbook of formal languages, Vol. 1*, pp. 329–438. Springer, Berlin, 1997.

[3] F. Durand, A characterization of substitutive sequences using return words, *Discrete Math.* **179** (1998), 89–101.

[4] F. Durand, B. Host, and C. Skau, Substitutive dynamical systems, Bratteli diagrams and dimension groups, *Ergod. Th. & Dynam. Sys.* **19** (1999), 953–993.

[5] A. Ehrenfeucht and G. Rozenberg, Simplifications of homomorphisms, *Inform. and Control* **38**(3) (1978), 298–309.

[6] N. P. Fogg, *Substitutions in dynamics, arithmetics and combinatorics*, Vol. 1794 of *Lecture Notes in Mathematics*, Springer-Verlag, Berlin, 2002. Edited by V. Berthé, S. Ferenczi, C. Mauduit and A. Siegel.

[7] C. Holton and L. Q. Zamboni, Descendants of primitive substitutions, *Theory Comput. Systems* **32** (1999), 133–157.

[8] R. A. Horn and C. R. Johnson, *Matrix analysis*, Cambridge University Press, Cambridge, 1990. Corrected reprint of the 1985 original.

[9] L. Kari, G. Rozenberg, and A. Salomaa, *L* systems. In *Handbook of formal languages, Vol. 1*, pp. 253–328. Springer, Berlin, 1997.

[10] K. Klouda and K. Medková, Synchronizing delay for binary uniform morphisms, *Theoret. Comput. Sci.* **615** (2016), 12–22.

[11] P. Kürka, *Topological and symbolic dynamics*, Vol. 11 of *Cours Spécialisés [Specialized Courses]*, Société Mathématique de France, Paris, 2003.

[12] F. Mignosi and P. Séébold, If a DOL language is $k$-power free then it is circular. In *Automata, languages and programming (Lund, 1993)*, Vol. 700 of *Lecture Notes in Comput. Sci.*., pp. 507–518. Springer, Berlin, 1993.

[13] B. Mossé, Puissances de mots et reconnaissabilité des points fixes d’une substitution, *Theoret. Comput. Sci.* **99** (1992), 327–334.

[14] B. Mossé, Reconnaissabilité des substitutions et complexité des suites automatiques, *Bull. Soc. Math. France* **124** (1996), 329–346.

[15] M. Queffélec, *Substitution dynamical systems—spectral analysis. Second Edition*, Vol. 1294 of *Lecture Notes in Mathematics*, Springer-Verlag, Berlin, 2010.