Complexity of infinite sequences with zero entropy

by

CHRISTIAN MAUDUIT (Marseille) and
CARLOS GUSTAVO MOREIRA (Rio de Janeiro)

1. Introduction and notations. In the whole paper we denote by \( q \)
a fixed integer greater than or equal to 2, by \( A \) the finite alphabet \( A = \{0, 1, \ldots, q-1\} \), by \( A^* = \bigcup_{k \geq 0} A^k \) the set of finite words on the alphabet \( A \), and by \( A^\mathbb{N} \) the set of infinite words (or infinite sequences of letters) on \( A \).

For any positive integer \( n \) we denote by \( \pi_n \) the projection from \( A^\mathbb{N} \) to \( A^n \) defined by

\[
\pi_n(w) = w_1 \ldots w_n \quad \text{if} \quad w = w_1w_2\ldots \quad \text{with} \quad w_i \in A \quad \text{for any positive integer} \quad i.
\]

If \( S \) is a finite set, we denote by \( |S| \) the number of elements of \( S \).

If \( w \in A^\mathbb{N} \) we denote by \( L(w) \) the set of finite factors of \( w \):

\[
L(w) = \{ u \in A^* : \exists (u', u'') \in A^* \times A^\mathbb{N}, w = u'u'' \}
\]

and, for any nonnegative integer \( n \), we write \( L_n(w) = L(w) \cap A^n \).

If \( x \) is a real number, we denote

\[
[x] = \sup \{ n \in \mathbb{Z} : n \leq x \}, \quad \lceil x \rceil = \inf \{ n \in \mathbb{Z} : x \leq n \}.
\]

DEFINITION 1.1. The complexity function of \( w \in A^\mathbb{N} \) is defined for any nonnegative integer \( n \) by \( p_w(n) = |L_n(w)| \).

The complexity function gives information about the statistical properties of an infinite sequence of letters. In this sense, it constitutes a possible way to measure the random behaviour of the infinite sequence (see [Qu], and see [MS1] and [MS2] for connections between measure of normality and other measures of pseudorandomness).

Obviously \( 1 \leq p_w(n) \leq q^n \) for any positive integer \( n \) and it is easy to check that the sequence \( (p_w(n))_{n \in \mathbb{N}} \) is bounded if and only if \( w \) is ultimately periodic. A basic result from [CH] shows that if there exists a positive integer \( n \) such that \( p_w(n) \leq n \), then \( (p_w(n))_{n \in \mathbb{N}} \) is bounded. It follows that non-ultimately periodic sequences \( w \) with lowest complexity are such that.
$p_w(n) = n + 1$ for any positive integer $n$. Such sequences, called *sturmian sequences*, have been extensively studied since their introduction by G. A. Hedlund and M. Morse in [HM1] and [HM2] (see [Lot] Chapter 2 and [PF]).

It is interesting to notice that if $w$ represents the $q$-adic expansion (resp. the continued fraction expansion) of the irrational number $\rho \in ]0,1[$, then the combinatorial property of $w$ being a sturmian sequence implies the arithmetic property of $\rho$ being a transcendental number (see [FM] (resp. [ADQZ]) and see [AB2] (resp. [AB1]) for a generalization to the case where $w$ has a sublinear complexity).

It is easy to prove the following lemma:

**Lemma 1.2.** For any $w \in A^\mathbb{N}$ and any $(n,n') \in \mathbb{N}^2$ we have $L_{n+n'}(w) \subset L_n(w) L_n'(w)$ and so $p_w(n+n') \leq p_w(n) p_w(n')$. 

**Consequence 1.** Lemma 1.2 implies that for any $w \in A^\mathbb{N}$, the sequence $(n^{-1} \log q p_w(n))_{n \geq 1}$ converges. We denote $E(w) = \lim_{n \to \infty} n^{-1} \log q p_w(n)$.

It can be shown (see for example [Kur]) that $E(w) \log q$ is the topological entropy of the symbolic dynamical system $(X(w), T)$ where $T$ is the one-sided shift on $A^\mathbb{N}$ and $X(w) = \text{Orb}_T(w)$ is the closure of the orbit of $w$ under the action of $T$ in $A^\mathbb{N}$ ($A^\mathbb{N}$ is equipped with the product topology of the discrete topology on $A$, i.e. the topology induced for example by the distance $d(w, w') = \exp(-\min\{n \in \mathbb{N} : w_n \neq w'_n\})$).

**Consequence 2.** Another easy consequence of Lemma 1.2 is that if there exists an integer $n_0$ such that $p_w(n_0) < q^{n_0}$, then $p_w(n) = o(q^n)$.

This simple remark shows that there are necessary conditions to verify for a nondecreasing sequence of integers $(p(n))_{n \in \mathbb{N}}$ to be the complexity function of some $w \in A^\mathbb{N}$ (see for instance [Fer]). But the characterization of all complexity functions (i.e. necessary and sufficient conditions for a nondecreasing sequence of integers $(p(n))_{n \in \mathbb{N}}$ to be the complexity function of some $w \in A^\mathbb{N}$) remains an open problem.

Nevertheless, let us mention that J. Cassaigne gave a complete answer to this question in the special case where $p$ is linear ([Cas2]) and that some partial results concerning the case where $p$ is sublinear can be found in [Ale] and [Cas1].

If we weaken the question by asking only which are the possible orders of magnitude for complexity functions, the problem still remains open, but it follows from an unpublished result due to J. Goyon [Goy] that for any $k \geq 1$ and any $(\alpha_1, \ldots, \alpha_k)$ in $(1, +\infty) \times \mathbb{R}^{k-1}$, there exists $w \in A^\mathbb{N}$ such that $p_w(n)$ has order of magnitude $n^{\alpha_1} (\log n)^{\alpha_2} \cdots (\log \cdots \log n)^{\alpha_k}$ (see also [Cas2] for the case $1 < \alpha_1 < 2$).
Complexity of infinite sequences

There are many references concerning the construction of infinite sequences \( w \) with low complexity, i.e. such that \( p_w(n) = O(n^k) \) for some \( k \geq 1 \) (see [All] or [Per] for a survey concerning these constructions). But, as pointed out in [Cas3], “not many examples are known which have intermediate complexity, i.e. for which \( E(w) = 0 \) but \( \log p_w(n)/\log n \) is unbounded”.

In [Cas3] J. Cassaigne constructed a large family of infinite sequences with intermediate complexity and proved the following result:

**Theorem 1.3.** Let \( \tau : \mathbb{R}^+ \to \mathbb{R}^+ \) be a function such that:

(i) \( \lim_{t \to +\infty} \tau(t)/\log t = +\infty \),
(ii) \( \tau \) is differentiable, except possibly at 0,
(iii) \( \lim_{t \to +\infty} \tau'(t)t^a = 0 \) for some \( a > 0 \),
(iv) \( \tau' \) is decreasing.

Then there exists \( w \in \{0, 1\}^\mathbb{N} \) such that \( \log p_w(n) \sim_{n \to +\infty} \tau(n) \). Moreover, \( w \) can be taken to be uniformly recurrent.

This construction is rich enough to include examples such that \( \tau(n) = n^\alpha \) (0 < \( \alpha < 1 \)), \( \tau(n) = (\log(n + 1))^{\alpha} \) (\( \alpha > 1 \)) or \( \tau(n) = n^{\alpha+\beta} \cos(\log(n+1)\gamma) \) (\( \alpha > 0 \), \( |\beta| < \alpha \) and \( \gamma \in \mathbb{R} \)).

In the same spirit, our work provides, for any given function \( f \) satisfying some reasonable conditions, a huge set of infinite words \( w \) such that \( p_w \) is close to \( f \) (Proposition 4.8).

**2. Results**

**Definition 2.1.** We say that a function \( f \) from \( \mathbb{N} \) to \( \mathbb{R}^+ \) satisfies conditions \((C_0)\) if

(i) \( f(n + 1) > f(n) \geq n + 1 \) for any \( n \in \mathbb{N} \),
(ii) \( \exists n_0 \in \mathbb{N}, n \geq n_0 \Rightarrow f(2n) \leq f(n)^2 \) and \( f(n + 1) \leq f(1)f(n) \),
(iii) the sequence \( (n^{-1} \log_q f(n))_{n \geq 1} \) converges to zero.

**Examples 2.2.** Let us give two typical examples of functions satisfying conditions \((C_0)\). In the rest of the paper, we will apply our results to these two examples in order to help the reader understand them and to get a precise idea about the order of magnitude of our estimates.

**Example A:** For each \( \alpha \geq 1 \), the function \( f \) is defined by \( f(0) = 1 \), \( f(n) = n + q - 1 \) for \( 1 \leq n < n_0 \) and \( f(n) = n^\alpha \) for \( n \geq n_0 \), with \( n_0 = \sup(2, 1/(q^{1/\alpha} - 1)) \).

**Example B:** For each \( 0 < \alpha < 1 \), the function \( f \) is defined by \( f(0) = 1 \), \( f(n) = n + q - 1 \) for \( 1 \leq n < n_0 \) and \( f(n) = q^{n^\alpha} \) for \( n \geq n_0 \), with \( n_0 = \inf\{n \in \mathbb{N} : q^{(n+1)^\alpha} - q^{n^\alpha} \geq 1 \) and \( q^{n^\alpha} \geq n + q \} \).

Our work concerns the study of infinite sequences \( w \) the complexity function of which is bounded by a given function \( f \) satisfying conditions \((C_0)\).
More precisely, our goal is to estimate the number of words of length $n$ on the alphabet $A$ that are factors of an infinite word with a complexity function less than $f$. The sturmian case ($f(n) = n + 1$) was studied by F. Mignosi in [Mig], who proved an explicit formula conjectured by S. Dulucq and D. Gouyou-Beauchamps in [DG]: the number of words of length $n$ on the alphabet $\{0, 1\}$ that are factors of a sturmian infinite word is exactly $1 + \sum_{i=1}^{n}(n-i+1)\Phi(i)$, where $\Phi$ is the Euler function (this is asymptotically equivalent to $n^3/\pi^2$). This formula can also be found in [KLB], but it seems that the first proof appears in an earlier paper by E. Lipatov [Lip]. A geometric proof is due to J. Berstel and M. Pocchiola [BP] and a combinatorial proof was given by A. de Luca and F. Mignosi [LM] (see [Lot]). Some partial generalizations concerning the case $f(n) = kn + 1$ (for $k \geq 2$) were given by F. Mignosi and L. Zamboni [MZ]. In the case of positive entropy (i.e. $\lim_{n \to \infty} n^{-1}\log_q f(n) > 0$), some sharp estimates can be obtained by using a different method. This will be the subject of a future work.

Throughout this paper, $f$ is a function from $\mathbb{N}$ to $\mathbb{R}^+$ satisfying conditions $(C_0)$.

Set $W(f) = \{w \in A^\mathbb{N} : p_w(n) \leq f(n), \forall n \in \mathbb{N}\}$ and $L_n(f) = \bigcup_{w \in W(f)} L_n(w)$.

The aim of Sections 3 and 4 is to give upper bounds and lower bounds for $|L_n(f)|$. We will exhibit (Theorems 3.1 and 4.1), for any given function $f$ satisfying conditions $(C_0)$, functions $\varphi$ and $\psi$ of approximately the same order of magnitude such that for $n$ large enough,

$$q^{\psi(n)} \leq |L_n(f)| \leq q^{\varphi(n)}.$$

In particular, these functions $\varphi$ and $\psi$ will satisfy

$$\lim_{n \to +\infty} \frac{1}{n}\psi(n) = \lim_{n \to +\infty} \frac{1}{n}\varphi(n) = 0.$$

**3. Upper bounds for $|L_n(f)|$.** For any nonnegative integers $k$ and $N$ we have

$$L_{kN}(f) = \bigcup_{w \in W(f)} L_{kN}(w) \subset \bigcup_{w \in W(f)} L_N^k(w) \quad \text{by Lemma 1.2}.$$

But

$$\bigcup_{w \in W(f)} L_N^k(w) = \bigcup_{w \in A^\mathbb{N}} L_N^k(w) \subset \bigcup_{w \in A^\mathbb{N}} L_N^k(w)$$

$$\subset \bigcup_{S \subset A^\mathbb{N}} S^k = \bigcup_{S \subset A^\mathbb{N}} S^k,$$

where $|S| \leq f(N)$. This implies that $|L_n(f)|$ is asymptotically equivalent to $n^3/\pi^2$. This formula can also be found in [KLB], but it seems that the first proof appears in an earlier paper by E. Lipatov [Lip]. A geometric proof is due to J. Berstel and M. Pocchiola [BP] and a combinatorial proof was given by A. de Luca and F. Mignosi [LM] (see [Lot]). Some partial generalizations concerning the case $f(n) = kn + 1$ (for $k \geq 2$) were given by F. Mignosi and L. Zamboni [MZ]. In the case of positive entropy (i.e. $\lim_{n \to \infty} n^{-1}\log_q f(n) > 0$), some sharp estimates can be obtained by using a different method. This will be the subject of a future work.
so that

\[ |L_{kN}(f)| \leq \sum_{S \subseteq A_N, |S|=f(N)} f(N)^k \left( \frac{q^N}{f(N)} \right) \leq f(N)^k q^{Nf(N)}. \]

We will now choose the parameter \( k \) so as to optimize this upper bound.

Suppose that \( N \geq N_0, \) where \( f(N_0) > q, \) and take \( k = \lfloor N f(N)/\log_q f(N) \rfloor \) to obtain

\[ |L_{kN}(f)| \leq q^{2Nf(N)}. \]

It is easy to verify that if \( f \) satisfies (C0) then \( (\lfloor N f(N)/\log_q f(N) \rfloor)_{N \geq N_0} \) is nondecreasing, so \( (N \lfloor N f(N)/\log_q f(N) \rfloor)_{N \geq N_0} \) is strictly increasing.

Let \( F(N) = N \lfloor N f(N)/\log_q f(N) \rfloor \) for any integer \( N, \) and \( F^*(n) = \min \{ m \in \mathbb{N} : F(m) \geq n \} \) for any \( n \in \mathbb{N}. \)

If we still denote by \( F \) an (arbitrary) continuous and strictly increasing extension of \( F \) from \( \mathbb{R}^+ \) to \( \mathbb{R}^+ \), it follows that \( F^*(n) \leq F^{-1}(n) + 1 \) for any \( n \in \mathbb{N}. \)

Let \( F(n) = \lfloor N f(N)/\log_q f(N) \rfloor \) for any integer \( N, \) and \( F^*(n) = \min \{ m \in \mathbb{N} : F(m) \geq n \} \) for any \( n \in \mathbb{N}. \)

Given an integer \( n, \) let \( N = F^*(n). \) We have \( F(N - 1) < n \leq F(N). \)

It follows from the previous estimate that

\[ |L_n(f)| \leq |L_{F(N)}(f)| \leq q^{2Nf(N)} = q^{\varphi(n)} \]

with

\[ \varphi(n) = 2F^*(n)f(F^*(n)). \]

As \( \lim_{N \to \infty} N^{-1} \log_q f(N) = 0, \) we remark that, for any integer \( n \) such that \( F^*(n) \geq n_0 + 1, \) we have

\[ \frac{\varphi(n)}{n} \leq \frac{\varphi(F(N))}{F(N - 1)} = \frac{2Nf(N)}{F(N - 1)} \leq \frac{2qNf(N - 1)}{(N - 1) \left[ \frac{(N-1)f(N-1)}{\log_q f(N-1)} \right]} = O\left( \frac{\log_q f(N - 1)}{N - 1} \right) = o(1). \]

Finally, we have proved the following theorem:

**Theorem 3.1.** \( |L_n(f)| \leq q^{\varphi(n)} \) where \( \varphi \) is defined by (1).

**Examples 3.2.** For \( f \) defined in Example A, we have

\[ F(N) = N \left[ \frac{N^{\alpha+1}}{\alpha \log_q N} \right] = \frac{N^{\alpha+2}}{\alpha \log_q N} + O(N), \]

so that

\[ F^{-1}(n) = \left( \frac{\alpha}{\alpha + 2} \right)^{1/(\alpha+2)} n^{1/(\alpha+2)} (\log_q n)^{1/(\alpha+2)} + O(n^{1/(\alpha+2)}) \]
and
\[ f(F^*(n)) = \left( \frac{\alpha}{\alpha + 2} \right)^{\alpha/(\alpha+2)} n^{\alpha/(\alpha+2)} (\log_q n)^{\alpha/(\alpha+2)} \]
\[ + O(n^{\alpha/(\alpha+2)} (\log_q n)^{(\alpha-1)/(\alpha+2)}) \]
(since \( F^*(n) = F^{-1}(n) + O(1) \), so that
\[ \varphi(n) = 2 \left( \frac{\alpha}{\alpha + 2} \right)^{(\alpha+1)/(\alpha+2)} n^{(\alpha+1)/(\alpha+2)} (\log_q n)^{(\alpha+1)/(\alpha+2)} \]
\[ + O(n^{(\alpha+1)/(\alpha+2)} (\log_q n)^{\alpha/(\alpha+2)}). \]

For \( f \) defined in Example B, we have
\[ F(N) = N[N^{1-\alpha} q^{N^\alpha}] = N^{2-\alpha} q^{N^\alpha} + O(N), \]
so that
\[ F^{-1}(n) = (\log_q n)^{1/\alpha} - \frac{2 - \alpha}{\alpha^2} (\log_q n)^{1/\alpha - 1} \log_q \log_q n \]
\[ + O((\log_q n)^{1/\alpha - 2} (\log_q \log_q n)^2), \]
and
\[ f(F^*(n)) = n(\log_q n)^{-(2-\alpha)/\alpha} + O(n(\log_q n)^{-2/\alpha} (\log_q \log_q n)^2) \]
\[ + O(n(\log_q n)^{2-3/\alpha}) \]
\[ = (1 + o(1)) n(\log_q n)^{-(2-\alpha)/\alpha} \]
(since \( F^*(n) = F^{-1}(n) + O(1) \), so that
\[ \varphi(n) = \frac{2n}{(\log_q n)^{(1-\alpha)/\alpha}} + O\left( \frac{n(\log_q \log_q n)^2}{(\log_q n)^{1/\alpha}} \right) = \frac{(2 + o(1)) n}{(\log_q n)^{(1-\alpha)/\alpha}}. \]

4. Lower bounds for \(|\mathcal{L}_n(f)|\). The main goal of this section is to give lower bounds for \(|\mathcal{L}_n(f)|\) when \( f \) satisfies conditions \((C_0)\). To do this, we will construct, for any fixed \( \eta_0 > 0 \), a large family \( W \) of infinite words \( w \) with complexity function \( p_w \) close to \( f \) and then give lower bounds for \(|\bigcup_{w \in W} L_n(w)|\). We will end up with the following theorem:

**Theorem 4.1.** For any fixed \( \eta_0 > 0 \) there exists an integer \( N_0 \) such that for any \( n \geq N_0 \),
\[ |\mathcal{L}_n(f)| > \exp\left( \left( \frac{1}{8} - \eta_0 \right) \frac{n}{G^{-1}(4n)} \log \frac{n}{G^{-1}(4n)} \right), \]
where \( G(x) = 2xg(x) \) and \( g \) is a function satisfying conditions \((C_0)\) such that for any integer \( n \geq N_0 \),
\[ \min(G((2 + \eta_0)n \log^2 n), G((2 + \eta_0)n)^2) \leq f(n). \]
4.1. Construction of a large family $W$ of infinite words. Let
\((a_k)_{k \geq 1}\) be the sequence of integers defined by
\(a_1 = 1, a_2 = 3\) and for \(k \geq 2,
\[a_{k+1} = \left\lceil \left( 2 + \frac{1}{\log^2 a_k} \right) a_k \right\rceil,\]
and \((b_k)_{k \geq 1}\) be the sequence of integers defined by
\(b_k = a_{k+1} - 2a_k\).

**Lemma 4.2.** For any \(k \geq 3\) we have \(2^k < a_k < 2^{k+1}\).

**Proof.** An easy computation shows that \(a_3 = 9, a_4 = 20, a_5 = 43\) and
\(a_6 = 90\). As \(a_{k+1} \geq 2a_k\) for any \(k \geq 1\), it follows that \(a_k > 2^k\) for any \(k \geq 3\).

For the upper bound, we can proceed as follows:
For any \(k \geq 3\) we have
\[a_{k+1} < 2a_k + \frac{a_k}{k^2 \log^2 2} + 1\]
so that for any \(k \geq 3\),
\[\frac{a_{k+1}}{a_k} < 2 + \frac{1}{k^2 \log^2 2} + \frac{1}{2^k}.\]
It follows that for any \(k \geq 5\) we have
\[\frac{a_{k+1}/2^{k+1}}{a_k/2^k} < 1 + \frac{1}{2k^2 \log^2 2} + \frac{1}{2^k+1} < 1 + \frac{3}{2k^2} < 1 - \frac{2}{k+1},\]
so that for \(k \geq 6\) we have
\[\frac{a_{k+1}}{2^{k+1}} < \frac{90}{64} \prod_{i=6}^{k} \frac{1 - \frac{2}{i+2}}{1 - \frac{2}{i+1}} < \frac{90}{64} \prod_{i=6}^{\infty} \frac{1 - \frac{2}{i+2}}{1 - \frac{2}{i+1}} = \frac{90}{64} \cdot \frac{7}{5} = \frac{63}{32},\]
proving that \(a_k < 2^{k+1}\) for any \(k \geq 7\). \(\blacksquare\)

**Remark 4.3.** The sequence \((a_k/2^k)_{k \geq 1}\) is increasing, so Lemma 4.2 yields \(\lim_{n \to \infty} a_k/2^k = a\) with \(a \in [1, 2]\).

**Remark 4.4.** For any \(k \geq 1\) we have \(2a_k < a_{k+1} \leq 3a_k\) and for any fixed \(\eta_1 > 0\) we can easily compute explicitly \(k_1 \in \mathbb{N}\) such that for any \(k \geq k_1\) we have \(a_{k+1} < [a2^{k+1}] \leq (2 + \eta_1)a_k\).

Let \(g\) be a function satisfying conditions \((C_0)\), and \(K_0\) be a fixed large constant which will be chosen later (depending on the \(\eta_0\) of Theorem 4.1).
Define the sequence \((m_k)_{k \geq K_0}\) by \(m_{K_0} = 2\) and, for \(k \geq K_0,
\[m_{k+1} = \min\left(m_k^2, \left\lceil \frac{g([a2^{k+1}])}{m_k} \right\rceil m_k \right)\].

**Remark 4.5.** The sequence \((m_k)_{k \geq 1}\) is well defined because \(m_k \geq 2\) for any \(k \geq K_0\).

**Lemma 4.6.** There exists an integer \(K_1 \geq K_0\) such that
\[m_{k+1} = \left\lceil \frac{g([a2^{k+1}])}{m_k} \right\rceil m_k\] for any \(k \geq K_1\).
Proof. Let us first remark that, if we suppose that \( m_{k+1} = m_k^2 \) for any \( k \geq K_0 \), then it would follow, on the one hand, that

\[
m_k = m_{K_0}^{2^{k-K_0}} = \lambda^{a^{2^{k+1}}}
\]

for any \( k \geq K_0 \), with \( \lambda = 2^{1/(a^{2k_0+1})} > 1 \), and on the other hand,

\[
m_k \leq \left\lceil \frac{g([a^{2k+1}])}{m_k} \right\rceil \text{ for any } k \geq K_0,
\]

which would imply altogether that

\[
g([a^{2k+1}]) > m_k(m_k - 1) \geq \frac{1}{2} m_k^2 = \frac{1}{2} \lambda^{a^{2k+2}} > \frac{1}{2} \lambda^{a^{2k+1}},
\]

which would contradict the hypothesis \( \lim_{n \to \infty} n^{-1} \log_q g(n) = 0 \).

This proves the existence of an integer \( K_1 \) such that

\[
m_{K_1+1} = \left\lceil \frac{g([a^{2K_1+1}])}{m_{K_1}} \right\rceil m_{K_1}, \quad \text{i.e.} \quad \left\lceil \frac{g([a^{2K_1+1}])}{m_{K_1}} \right\rceil \leq m_{K_1}.
\]

It is now easy to prove by induction on \( k \) that

\[
\left\lceil \frac{g([a^{2k+1}])}{m_k} \right\rceil \leq m_k \quad \text{for any } k \geq K_1.
\]

As for any \((x, n) \in \mathbb{R} \times \mathbb{Z}\) the inequality \([x] \leq n\) is equivalent to \( x \leq n \), it is equivalent to prove that

\[
g([a^{2k+1}]) \leq m_k^2 \quad \text{for any } k \geq K_1.
\]

Indeed, the latter is true for \( k = K_1 \) and if we suppose that \( g([a^{2k+1}]) \leq m_k^2 \), i.e. \( m_{k+1} = \left\lceil g([a^{2k+1}])/m_k \right\rceil m_k \), then

\[
g([a^{2k+2}]) \leq g(2[a^{2k+1}]) \leq (g(2[a^{2k+1}]))^2 \quad \text{by condition (C_0)(ii)}
\]

\[
\leq \left( \left\lceil \frac{g([a^{2k+1}])}{m_k} \right\rceil m_k \right)^2 = m_{k+1}^2.
\]

The lemma below shows that the sequences \( (m_k)_{k \geq K_0} \) and \( (g([a^{2k}]))_{k \geq K_0} \) have the same order of magnitude:

**Lemma 4.7.**

(i) For any integer \( k \geq K_0 \) we have \( m_k \leq 2g([a^{2k}]) \).

(ii) For any integer \( k \geq K_1 + 1 \) we have \( m_k \geq g([a^{2k}]) \).

Proof. (i) The inequality is true for \( k = K_0 \), and if we suppose that \( m_k \leq 2g([a^{2k}]) \), it follows that

\[
m_{k+1} \leq \left\lceil \frac{g([a^{2k+1}])}{m_k} \right\rceil m_k \leq 2g([a^{2k+1}]),
\]

because \( g([a^{2k+1}])/m_k \geq g([a^{2k}])/m_k \geq 1/2 \) (recall that if \( x \geq 1/2 \), then \([x] \leq 2x\)).
(ii) If \( k \geq K_1 \), we have

\[
m_{k+1} = \left\lceil \frac{g([a2^{k+1}])}{m_k} \right\rceil m_k \geq g([a2^{k+1}]). \]

Starting from \( M(K_0) = \{0^{a_{K_0}}, 0^{a_{K_0}-1}\} \) we define by induction for each \( k \geq K_0 \) a set \( M(k) \) of \( m_k \) words of length \( a_k \) as follows:

If \( M(k) \) has already been constructed, we choose for each \( \alpha \in M(k) \) a set \( X(\alpha) \subset M(k) \) with \( |X(\alpha)| = m_{k+1}/m_k \). Then we set

\[
M(k+1) = \{\alpha a_k^b \beta : \alpha \in M(k), \beta \in X(\alpha)\}.
\]

We denote by \( \mathcal{M}(k) \) the union, over all possible choices of the sets \( X(\alpha) \), of the sets \( M(k) \), and by \( W \) the set of infinite words \( w \) on the alphabet \( A \) such that \( \pi_{a_k}(w) \in \mathcal{M}(k) \) for any integer \( k \geq K_0 \).

4.2. Complexity of elements of \( W \). The goal of this subsection is to show the following proposition:

**Proposition 4.8.** For any fixed \( \eta_0 > 0 \) there exists an integer \( n_0 \) such that for any \( n \geq n_0 \) and for any \( w \in W \),

\[
\frac{1}{2} g\left( \left( \frac{1}{2} - \eta_0 \right) n \right) < p_w(n) < \min(G((2 + \eta_0)n \log^2 n), G((2 + \eta_0)n)^2).
\]

**Proof.** It is easy to bound \( p_w \) from below:

If \( a_k \leq n < a_{k+1} \), we have

\[
p_w(n) \geq m_k \quad \text{by construction}
\]

\[
\geq g([a2^k]) \quad \text{by Lemma 4.7(ii)}
\]

\[
\geq g(a_k).
\]

It follows from Remark 4.4 that if \( n \geq a_{k_1} \) we have \( p_w(n) \geq \frac{1}{2} g\left( \left( \frac{1}{2} - \eta_1 \right) n \right) \).

We now have to give upper bounds for \( p_w \).

**Lemma 4.9.** Let \( \tau \) be the function defined on the interval \([e^2, +\infty)\) by \( \tau(x) = x/(\log x)^2 \). The function \( \tau \) is strictly increasing and, for any fixed \( \eta_2 > 0 \), we can explicitly compute \( n_2 \in \mathbb{N} \) such that for any \( n \geq n_2 \),

\[
(2) \quad \tau^{-1}(n) \leq (1 + \eta_2)n \log^2 n.
\]

**Proof.** The study of the derivative of \( \tau \) shows easily that \( \tau \) is strictly increasing on \([e^2, +\infty)\). The inequality (2) is thus equivalent to

\[
n \leq \tau((1 + \eta_2)n \log^2 n),
\]

which is equivalent to

\[
\left( 1 + \frac{\log(1 + \eta_2)}{\log n} + 2 \frac{\log \log n}{\log n} \right)^2 \leq 1 + \eta_2,
\]

which clearly holds for \( n \) large enough.
For any fixed \( \eta_3 > 0 \), fix \( \eta_1 \) and \( \eta_2 \) respectively in Remark 4.4 and Lemma 4.9 such that \((2 + \eta_1)(1 + \eta_2) \leq 2 + \eta_3\), and denote \( n_3 = \max(b_{k_1+1}, n_2) \).

To bound \( p_w \) from above, define, for any integer \( n \geq n_3 \), \( k_0(n) \) to be the smallest integer such that \( b_{k_0(n)} \geq n \).

**Lemma 4.10.** For any \( n \geq n_3 \), we have \([a2^{k_0(n)}]\) \( \leq (2 + \eta_3)n \log^2 n \).

**Proof.** By definition of \( k_0(n) \) we have \( b_{k_0(n)-1} < n \leq b_{k_0(n)} \), and by definition of \( (b_k)_{k \geq 1} \),

\[
\tau(a_{k_0(n)-1}) \leq b_{k_0(n)-1} < \tau(a_{k_0(n)-1}) + 1.
\]

It follows from Lemma 4.9 that \( a_{k_0(n)-1} < \tau^{-1}(n) \leq (1 + \eta_2)n \log^2 n \) and from Remark 4.4 that

\[
[a2^{k_0(n)}] \leq (2 + \eta_1)a_{k_0(n)-1} < (2 + \eta_3)n \log^2 n.
\]

Let us now use the fact that every factor of length \( n \geq n_3 \) in \( w \) must be a factor of some element of \( M(k_0(n)) \) preceded or followed by a sequence of zeros.

This means that for \( n \geq n_3 \) we have

\[
p_w(n) \leq (n - 1 + a_{k_0(n)})m_{k_0(n)} + 1 \leq (n + a_{k_0(n)})m_{k_0(n)} < 2(n + (2 + \eta_3)n \log^2 n)g((2 + \eta_3)n \log^2 n).
\]

If we now fix \( \eta_4 > 0 \) such that \( \eta_4 > \eta_3 \), there exists an integer \( n_4 \geq n_3 \) such that for \( n \geq n_4 \),

\[
p_w(n) < G((2 + \eta_4)n \log^2 n).
\]

Let us now give another upper bound for \( p_w \) that will give a better result when \( g \) is growing very fast.

Every factor of length \( n \) in \( w \) must be a factor of some element of \( M(k+1) \) (where \( a_k \leq n < a_{k+1} \)) preceded or followed by a sequence of zeros, or a factor of \( M(k+1) \) followed by \( b_r \) zeros (for some \( k+1 \leq r \leq k_0(n) \)) followed by another factor of \( M(k+1) \).

This gives the estimate (valid for \( n \geq n_3 \))

\[
p_w(n) \leq (n + a_{k+1})m_{k+1} + (k_0(n) - k)nm_{k+1}^2
\leq 4ng([a2^{k+1}]) + 4(k_0(n) - k)n \cdot g([a2^{k+1}])^2
\leq 4ng((2 + \eta_1)n) + 4\log_2((2 + \eta_3)n \log^2 n)n \cdot g((2 + \eta_1)n)^2.
\]

This shows that there exists an integer \( n_5 \geq n_3 \) such that for \( n \geq n_5 \) we have

\[
p_w(n) < G((2 + \eta_1)n)^2.
\]
To finish the proof of Proposition 4.8 it is enough, for any fixed \( \eta_0 \), to take in the previous arguments \( \eta_1 < \eta_0\), \( \eta_4 < \eta_0 \) and \( n_0 = \max (a_{k_1}, n_4, n_5) \). ■

**Remark 4.11.** The above upper bound for \( k_0(n) - k \) is a simple application of Lemmas 4.2 and 4.10. It is easy to improve it by showing that

\[
k_0(n) - k = 2 \frac{\log \log n}{\log 2} + O(1).
\]

**Corollary 4.12.** If \( g \) satisfies conditions \( (C_0) \), \( \eta_0 \) and \( n_0 \) are as in the statement of Proposition 4.8, and \( K_0 \) satisfies \( b_{K_0} > n_0 \), then \( p_w(n) \leq f(n) \) for any \( w \in W \) and \( n \geq 1 \).

**Proof.** We have two cases:

If \( n \leq b_{K_0} \), by construction a factor of size \( n \) of a word \( w \in W \) has at most one letter equal to 1 and all other letters equal to 0, so \( p_w(n) \leq n + 1 \leq f(n) \).

If \( n > b_{K_0} \), we have \( k_0(n) > K_0 \), and since \( b_{K_0} > n_0 \), it follows that \( p_w(n) < \min(G((2 + \eta_0)n \log^2 n), G((2 + \eta_0)n)^2) \leq f(n) \). ■

**4.3. Lower bounds for** \( |\bigcup_{w \in W} L_n(w)| \). For any \( k \geq K_0 \), let

(3)

\[
r(k) = \lceil \log_2 m_k \rceil.
\]

For every integer \( n \geq a_{K_0 + r(K_0)} \), let \( k \) be the unique integer satisfying

\[
a_{k-1 + r(k-1)} \leq n < a_{k + r(k)}
\]

and let \( s \) be defined by

\[
a_{k+s} \leq n < a_{k+s+1}
\]

(we have \( r(k-1) - 1 \leq s \leq r(k) - 1 \)).

We will now construct subsets of \( W \) as follows. Enumerate the set \( M(k) \) obtained in Section 4.1 as \( M(k) = \{ \alpha_1(k), \ldots, \alpha_{m_k}(k) \} \). We can assume that for \( k' \geq k \) we have \( \alpha_{j+1}(k') \in X(\alpha_j(k')) \) for each \( 1 \leq j \leq m_{k'} \) (we put \( \alpha_{m_{k'} + 1} := \alpha_1 \)) and

\[
M(k' + 1) = \{ \alpha_1(k' + 1), \ldots, \alpha_{m_{k'} + 1}(k' + 1) \}
\]

where we enumerate the elements of \( M(k' + 1) \) in such a way that

\[
\alpha_1(k' + 1) = \alpha_1(k')0^b_{k'} \alpha_2(k'),
\]

\[
\alpha_2(k' + 1) = \alpha_3(k')0^b_{k'} \alpha_4(k'),
\]

\[\vdots\]

\[
\alpha_{\lfloor (m_{k'} + 1)/2 \rfloor}(k' + 1) = \begin{cases} 
\alpha_{m_{k'} - 1}(k')0^b_{k'} \alpha_{m_{k'}}(k') & \text{for } m_{k'} \text{ even,} \\
\alpha_{m_{k'}}(k')0^b_{k'} \alpha_1(k') & \text{for } m_{k'} \text{ odd.}
\end{cases}
\]

This construction gives

\[
\alpha_1(k + s) = \alpha_1(k)0^{b_k} \alpha_2(k)0^{b_{k+1}} \ldots 0^{b_{k+1}} \alpha_{2s-1}(k)0^{b_k} \alpha_{2s}(k)
\]

where \( \alpha_1(k), \ldots, \alpha_{2s}(k) \) appear in this order as factors of length \( a_k \).
Lemma 4.13.

(i) For every integer \( n \geq a_{K_0 + r(K_0)} \) we have
\[
 n < 4a_k2^s < 4a_km_k.
\]

(ii) For every integer \( n \geq \max(a_{k_1 + r(k_1)}, a_{K_1 + 1 + r(K_1 + 1)}) \) we have
\[
 n > \frac{1}{4} G \left( \frac{a_k}{2 + \eta_1} \right).
\]

Proof. (i) We have
\[
 n < a_{k+s+1} \quad \text{by construction}
\]
\[
 < 2^{k+s+2} \quad \text{by Lemma 4.2}
\]
\[
 < 2^{s+2}a_k \quad \text{by Lemma 4.2}
\]
The second inequality results from the fact that
\[
 2^s \leq 2^{r(k)-1} = 2^{\lceil \log_2 m_k \rceil - 1} < m_k.
\]

(ii) We have, for any \( k \geq K_1 + 2 \),
\[
 n \geq a_{k-1+r(k-1)} \quad \text{from the definition of } k
\]
\[
 > 2^{k-1+r(k-1)} \quad \text{by Lemma 4.2}
\]
\[
 > \frac{1}{2} a_{k-1}2^{r(k-1)} \quad \text{by Lemma 4.2}
\]
\[
 \geq \frac{1}{2} a_{k-1}m_{k-1} \quad \text{from the definition of } r
\]
\[
 \geq \frac{1}{2} a_{k-1}g(\lceil a2^{k-1} \rceil) \quad \text{by Lemma 4.7(ii)}
\]
\[
 \geq \frac{1}{2} a_{k-1}g(a_{k-1}).
\]
It follows from Remark 4.4 that if \( k \geq \max(k_1 + 1, K_1 + 2) \), we have
\[
 n > \frac{1}{2} a_{k-1}g(a_{k-1}) > \frac{1}{4} \cdot \frac{2a_k}{2 + \eta_1} g \left( \frac{a_k}{2 + \eta_1} \right) = \frac{1}{4} G \left( \frac{a_k}{2 + \eta_1} \right). \quad \blacksquare
\]

We have \( 2^s \leq 2^{r(k)-1} < m_k \), and if we denote by \( W_0 \) the set of all infinite words obtained by this construction, we have
\[
 \left| \bigcup_{w \in W_0} L_n(w) \right| \geq A_{m_k}^{2^s} = \frac{m_k!}{(m_k - 2^s)!}.
\]
For any fixed \( \eta_5 > 0 \) there is \( k_5 \) such that for any \( k \geq k_5 \),
\[
 \frac{m_k!}{(m_k - 2^s)!} \geq ((m_k!)^{1/m_k})^{2^s} \geq (m_k/e)^{2^s} \geq m_k^{(1-\eta_5)2^s}.
\]
Then, for any $k \geq \max(k_1 + 1, K_1 + 2, k_5)$,

$$\frac{m_k!}{(m_k - 2^k)!} \geq \exp((1 - \eta_5)2^k \log m_k)$$

$$> \exp\left((1 - \eta_5)\frac{n}{4a_k} \log \frac{n}{4a_k}\right) \text{ by Lemma 4.13(i)}$$

$$> \exp\left(\frac{1 - \eta_5}{4(2 + \eta_1)} \cdot \frac{n}{G^{-1}(4n)} \log \frac{n}{4(2 + \eta_1)G^{-1}(4n)}\right) \text{ by Lemma 4.13(ii).}$$

Now for any fixed $\eta_0 > 0$ and any $\eta_1 > 0$ fixed as in Subsection 4.2 (in particular $\eta_1 < \eta_0$) choose $\eta_5$ such that $\eta_5 < 4\eta_0(2 + \eta_1) - \eta_1 / 2$. Then $1/8 - \eta_0 < (1 - \eta_5)/(4(2 + \eta_1))$ and we conclude that there exists an integer $N_0 = \max(a_{k_1 + r(k_1)}, a_{K_1 + 1 + r(K_1 + 1)}, a_{k_5 - 1 + r(k_5 - 1)}, n_0)$ such that, for any $n \geq N_0$,

$$\left| \bigcup_{w \in W} L_n(w) \right| > \left| \bigcup_{w \in W_0} L_n(w) \right| > \exp\left(\left(\frac{1}{8} - \eta_0\right)\frac{n}{G^{-1}(4n)} \log \frac{n}{G^{-1}(4n)}\right). \blacksquare$$

**Examples 4.14.** For $f$ defined in Example A, we can take, for $N \geq e^{2\alpha/(\alpha - 1)}$,

$$G(N) = \frac{N^\alpha}{(2 + \eta_0)^\alpha \log^{2\alpha} N},$$

so that

$$G^{-1}(4n) = \frac{4^{1/\alpha}(2 + \eta_0)}{\alpha^2} n^{1/\alpha} \log^2 n + O(n^{1/\alpha} \log n \log \log n).$$

If we combine this with the result obtained in Section 3, we conclude that there are positive constants $c_1(\alpha)$ and $c_2(\alpha)$ such that, for $n$ large enough,

$$\exp\left(c_1(\alpha) \frac{n^{(\alpha - 1)/\alpha}}{\log n}\right) < |\mathcal{L}_n(f)| < \exp\left(c_2(\alpha)n^{(\alpha + 1)/(\alpha + 2)}(\log n)^{(\alpha + 1)/(\alpha + 2)}\right)$$

(indeed, we can take any $c_1(\alpha) < 4^{-1/\alpha} \alpha(\alpha - 1)/16$ and any $c_2(\alpha) > 2(\log q)^{1/(\alpha + 2)}(\frac{\alpha}{\alpha + 2})^{(\alpha + 1)/(\alpha + 2)}$).

For $f$ defined in Example B, we can take, for $N \geq (2 + \eta_0)\left(\frac{2}{\alpha \log q}\right)^{1/\alpha}$,

$$G(N) = q^{N^\alpha/(2(2 + \eta_0)^\alpha)}$$

so that

$$G^{-1}(4n) = \left(\frac{2}{\log q}\right)^{1/\alpha} (2 + \eta_0)(\log n)^{1/\alpha} + O((\log n)^{(1-\alpha)/\alpha}).$$

Combining this with the result obtained in Section 3, we conclude that there are constants $c_1(\alpha)$ and $c_2(\alpha)$, $0 < c_1(\alpha) < c_2(\alpha)$, such that, for $n$
large enough,
\[
\exp\left(c_1(\alpha) \frac{n}{(\log n)^{(1-\alpha)/\alpha}}\right) < |\mathcal{L}_n(f)| < \exp\left(c_2(\alpha) \frac{n}{(\log n)^{(1-\alpha)/\alpha}}\right)
\]

(indeed, we can take any \(c_1(\alpha) < \frac{1}{16} \left((\log q)/2\right)^{1/\alpha}\) and any \(c_2(\alpha) > 2(\log q)^{1/\alpha}\)).

4.4. An open question. Our method does not work for sequences with sublinear complexity. A natural open problem is to give sharp estimates for \(|\mathcal{L}_n(f)|\) when \(f\) is a linear function.

To state more precise questions, let us give some definitions. Let \(g_0(x) = x\), \(g_1(x) = x + 1\), and, for \(k > 0\) and \(x > 0\) large, \(g_{k+1}(x) = \exp(g_k(\log x))\) and \(g_{-k}(x) = g_k^{-1}(x)\). We say that an increasing function \(f\) from \(\mathbb{R}^+\) to \(\mathbb{R}^+\) is morally polynomial if there is \(k \geq 0\) such that \(g_{-k}(x) \leq f(x) \leq g_k(f(x))\) for every \(x\) sufficiently large, and that \(f\) is morally exponential if \(\log f\) is morally polynomial. We have the following questions:

(i) Is it true that \(\ell(n) = |\mathcal{L}_n(f)|\) is morally polynomial for any linear function \(f\)?

(ii) Does there exist \(A > 0\) such that \(\ell(n) = |\mathcal{L}_n(f)|\) is morally exponential for \(f(n) = An\)?

Clearly we cannot have positive answers to both of these questions. On the other hand, it is not clear whether we will have a positive answer to one of them, since there are functions which are neither morally polynomial nor morally exponential (e.g. increasing functions \(f\) such that \(f \circ f = \exp\)). However, any logarithmico-exponential function \(f\) (in the sense of Hardy) satisfying \(x \leq f(x) \leq q^x\) for every large \(x\) is morally polynomial or morally exponential (see Section 4.1 of [Har]).

Acknowledgments. Research of C. Mauduit was partially supported by the Brazil/France Agreement in Mathematics (Proc. CNPq 60-0014/01-5 and 69-0140/03-7).

Research of C. G. Moreira was partially supported by CNPq.

References


Complexity of infinite sequences


Christian Mauduit  
Institut de Mathématiques de Luminy  
163, avenue de Luminy  
13288 Marseille Cedex 9, France  
E-mail: mauduit@iml.univ-mrs.fr

Carlos Gustavo Moreira  
Instituto de Matemática Pura e Aplicada  
Estrada Dona Castorina 110  
22460-320 Rio de Janeiro, RJ, Brasil  
E-mail: gugu@impa.br

Received on 9.2.2009  
and in revised form on 20.7.2009