RAIRO-Theor. Inf. Appl. 46 (2012) 67–76
DOI: 10.1051/ita/2011126

Available online at:
www.rairo-ita.org

THREE COMPLEXITY FUNCTIONS

Sébastien Ferenczi[1,2] and Pascal Hubert[2,3]

Abstract. For an extensive range of infinite words, and the associated symbolic dynamical systems, we compute, together with the usual language complexity function counting the finite words, the minimal and maximal complexity functions we get by replacing finite words by finite patterns, or words with holes.

Mathematics Subject Classification. 37B10, 68R15.

1. Preliminaries

Given a language L on a finite alphabet A, the *complexity function* $p_L(n)$ counts for every n the number of factors of length n of L; this is a very useful notion, both inside word combinatorics and for the study of symbolic dynamical systems, see for example the survey [7]; of particular interest are the infinite words which are determined by the complexity of their language, those words for which $p_L(n) \leq n$ for at least one n are ultimately periodic [15], while the *Sturmian words*, of complexity $n+1$ for all n, are natural codings of rotations, see [6, 16], or Chapter 6 of [17], and Section 4 below. Note that the complexity is exponential when the language has positive topological entropy, and has not been widely used for that range of languages.

To study further the combinatorial properties of infinite words, the notion of *maximal pattern complexity*, denoted by $p^*_L(n)$, was introduced in [10] and

Keywords and phrases. Infinite words, symbolic dynamical systems, complexity.

[1] Institut de Mathématiques de Luminy, CNRS UMR 6206, Case 907, 163 av. de Luminy, 13288 Marseille Cedex 9, France. ferenczi@iml.univ-mrs.fr

[2] Fédération de Recherche des Unités de Mathématiques de Marseille, CNRS FR 2291, France.

[3] Laboratoire d'Analyse, Topologie et Probabilités, CNRS UMR 6632, Case A, Faculté de Saint Jérôme, Avenue Escadrille Normandie-Niemen, 13397 Marseille Cedex 20, France. hubert@cmi.univ-mrs.fr

Article published by EDP Sciences

© EDP Sciences 2011

https://doi.org/10.1051/ita/2011126 Published online by Cambridge University Press

developped in [11]: it replaces words by *patterns*, or words with holes, see Definition 1.2 below; note that this notion is completely different from the patterns used in [3] for example. It gives an extra tool for the study of sequences of low complexity, but, as we show in Section 2 below, it becomes simply $(\#A)^n$ under the (relatively mild) condition of weak mixing for the associated symbolic dynamical system. To study a wider class of systems than the non-weakly mixing ones, here we present the dual notion of *minimal pattern complexity*, $p_{*L}(n)$, and for infinite words or symbolic dynamical systems consider together the three functions $p_{*L}(n) \leq p_L(n) \leq p_L^*(n)$; the case where these three functions are the same has been extensively studied by Kamae *et al.* [8, 12] under the name of uniform (pattern) complexity. During the refereeing process of the present paper, our three complexity function have been further studied by Kamae [9].

In the present paper, we begin by looking at the cases where the minimal pattern complexity is smallest (namely, bounded, for ultimately periodic languages) or the maximal pattern complexity is largest (the weakly mixing case mentioned above). Then we show that the minimal pattern complexity is itself exponential for a wide class of languages, those associated to mixing shifts of finite type. Then we focus on languages of zero topological entropy, where we compute the three complexity functions on several examples; in particular, we show that, contrarily to other complexity functions, the minimal pattern complexity of a general coding of a rotation of angle α by a partition in two intervals, $[0, \beta[$ and $[\beta, 1[$, depends on the value of β, even for large values of n and when β and α are rationally independent.

The authors wish to thank the referees for useful corrections, and in particular for spotting a mistake in Proposition 4.5 and giving the ideas we used to get the right result.

We recall first

Definition 1.1. A *language* L on a finite alphabet A is a set of words; the languages we consider in the present paper satisfy two extra conditions: if w is in L, all its subwords or *factors* are in L, and wb is in L for at least one letter b of A.

A language L is *uniformly recurrent* if for each w in L there exists n such that w occurs in each word of length n of L.

The language $L(u)$ of an infinite word u is the set of all its finite factors $u_r \dots u_s$.

The *symbolic dynamical system* (X_L, S) associated to a language L is the one-sided shift $S(x_0x_1x_2\dots) = x_1x_2\dots$ on the subset X_L of $A^{\mathbb{N}}$ made with the infinite words such that for every $r < s$, $x_r \dots x_s$ is in L. For a word $w = w_1 \dots w_r$ in L, the *cylinder* $[w]$ is the set $\{x \in X_L; x_0 = w_1, \dots x_{r-1} = w_r\}$.

If μ is an invariant probability measure on (X_L, S), this gives a measure-theoretic dynamical system (X_L, S, μ), and $\mathcal{L}^2(X_L, \mu)$ denotes the space of square-integrable functions on X_L for μ.

We introduce now our main notions

Definition 1.2. A *pattern* $\tau = \tau_0, \tau_1, \dots, \tau_{n-1}$ is a strictly increasing sequence of nonnegative integers, its *size* is n while its *total length* is $\tau_{n-1} - \tau_0 + 1$.

https://doi.org/10.1051/ita/2011126 Published online by Cambridge University Press

Given a language L, we define $L(\tau)$ as the set of all possible $w_{\tau_0}w_{\tau_1}\dots w_{\tau_{n-1}}$ when w varies in L. If L is generated by an infinite word $u_0u_1\dots$, we call $u[n+\tau]$ the word $u_{n+\tau_0}u_{n+\tau_1}\dots$, and $L(\tau)=L_u(\tau)$ is the set of all possible $u[n+\tau]$ when n varies.

The *minimal pattern complexity* $p_{*L}(n)$, resp. the *maximal pattern complexity* $p^*_L(n)$, is the minimum, resp. maximum of $\#L_u(\tau)$ over all patterns of size n; we denote them by $p_{*u}(n)$, resp. $p^*_u(n)$, while the complexity is $p_u(n)$, when $L=L(u)$ for an infinite word u.

2. Extremal cases

In this section we look at the languages with the smallest possible minimal pattern complexity, and at those with the largest maximal pattern complexity.

Proposition 2.1. *If there exists n such that $p_{*L}(n)\le n$, L is the union of a finite number of $L(u^j)$ where each infinite word u^j is* ultimately periodic*, i.e. there exist positive integers n_j and t_j such that $u^j_{n+t_j}=u^j_n$ for all $n>n_j$.*

Proof. If $p_{*L}(1)=1$, the result is immediate. Suppose $p_{*L}(1)>1$ and take the smallest $n>1$ satisfying the hypothesis: then there exists a pattern τ of size n such that $\#L(\tau)\le n$, while $\#L(\theta)>n-1$ for any pattern θ of size $n-1$ and in particular for $\tau'=\tau_0,\tau_1,\dots,\tau_{n-2}$. Hence $\#L(\tau)=\#L(\tau')$, which means that there is only one possible value of $w_{\tau_{n-1}}$ when the values of $w_{\tau_0},\dots,w_{\tau_{n-2}}$ are known; hence in particular every word of length bigger than the total length of τ has only one right extension, and we conclude by the reasoning of [15]. □

Of course, for languages as in Proposition 2.1, $p_{*L}(n)\le p_L(n)$, and they are both bounded in n.

We use now some dynamical notions: the elementary ones are defined in Definition 1.1 above, the more specialized ones are defined in the course of the following theorem.

Theorem 2.2. *Whenever*

- *either L is uniformly recurrent and the symbolic dynamical system (X_L,S,μ) is* weakly mixing *for some invariant measure μ (namely, there is no nonconstant eigenfunction in $\mathcal{L}^2(X_L,\mu)$ for the operator $f\to f\circ S$);*
- *or (X_L,S) is* topologically strongly mixing *(namely, for any pair of nonempty open sets E and F, $E\cap S^{-n}F$ is nonempty for all n large enough),*

then $p^*_L(n)=(\#A)^n$.

Proof. It is known, see for example [5] that, under the condition of weak mixing, for any pair of measurable sets E, F, $\mu(E\cap S^{-n}F)$ converges to $\mu(E)\mu(F)$ along a set of density 1. If L is uniformly recurrent, $\mu[w]>0$ for any word w in L, and thus for any pair of nonempty cylinders, $E\cap S^{-n}F$ has positive measure for n in

https://doi.org/10.1051/ita/2011126 Published online by Cambridge University Press

a set of density 1. Thus there exists k_1 such that $E \cap S^{-k_1}F$ has positive measure when E and F are any of the cylinders $[i]$, $i \in A$; and similarly, if we have ensured that all the sets $G = [i_0] \cap S^{-k_1}[i_1] \dots \cap S^{-k_{n-1}}[i_{n-1}]$ have positive measure, we can find a k_n such that $G \cap S^{-k_n}[i]$ has positive measure for all the G defined above and every i in A. Thus we can build inductively patterns of length n such that $L(\tau)$ contains all the possible words of length n.

The same construction of patterns works under the hypothesis of topological strong mixing, as cylinders are open sets. □

The result above would hold also under the weaker hypothesis that for any pair of nonempty open sets, $E \cap S^{-n}F$ is nonempty for n in a set of density 1; however, it does not seem to be known whether the latter condition is implied by topological weak mixing (no nonconstant continuous eigenfunction for the operator $f \to f \circ S$).

Note that Example 3 below provides systems satisfying the hypotheses of Theorem 2.2 with complexity $p_L(n)$ much less than exponential: by Remark 4.8 we can get complexity $2n + 1$ with the first hypothesis, and $3n + 1$ with the second one.

3. Examples of positive entropy

As soon as the *topological entropy* of a symbolic system is positive, as a direct consequence of the definition of topological entropy the complexity of the language is exponential, and thus also the maximal pattern complexity. We show now that under an extra condition, all complexities are exponential. The general reference for this section is [14].

Definition 3.1. A *shift of finite type* is the shift on the set $X(F)$ of all bilateral infinite sequences which do not contain any word of F, where F is a finite set of words. The associated language L is the set of all finite factors of infinite words in $X(F)$.

By replacing the original alphabet A by the alphabet A' whose letters are all words of some fixed length q in L, and coding the infinite sequence $(x_n, n \in \mathbb{Z})$ by the infinite sequence on A' $((x_{qn} \dots x_{qn+q-1}), n \in \mathbb{Z})$, we can always assume that F is made with words of length 2; the matrix of the shift is defined by $M = ((m_{i,j})), i \in A', j \in A'$ where $m_{i,j} = 0$ if ij is in F, $m_{i,j} = 1$ otherwise.

A shift of finite type is *strongly mixing* if all the entries of M^n are strictly positive for some integer n.

Theorem 3.2. *For a strongly mixing shift of finite type, $p_{*L}(n)$ is exponential.*

Proof. We show first that there exists a number m such that, for any words $v \in L$, $w \in L$ there exists a word $x \in L$ of length m with $vxw \in L$: we look first at the subshift on the alphabet A'; because of the property of M^n, for every $a' \in A'$, $b' \in A'$, there exists a word w' of length $n - 1$ on A' such that $a'w'b'$ is in the language of the shift on A'; but then so is $u'w'v'$ for every word u' ending with a'

https://doi.org/10.1051/ita/2011126 Published online by Cambridge University Press

and v' beginning with b', as the constraints are only on pairs of consecutive letters. Coming back to the alphabet A, for any word u ending with a suffix u' of length q, any word v beginning with a prefix v' of length q, the standard decoding gives a word w of length $m = qn$ such that uwv is in L.

Now we take N much bigger than m, and a pattern τ of size N and total length $l \geq N$. We cut the integer interval $[\tau_0, \tau_N]$ into m parts of length (roughly) $\frac{l}{m}$; in at least $\frac{N}{m}$ of these parts, denoted by $J_0, \dots J_s$, there is some τ_i; we select a subpattern $(\tau_{i_1}, \dots \tau_{i_r})$ such that τ_{i_j} is in J_{2j}, $1 \leq j \leq r = \frac{s}{2}$. As $\tau_{i_{j+1}} - \tau_{i_j} > m$, the language of this subpattern contains all possible words of length r, and thus $p_{*L}(N) \geq ((\#A)^{\frac{1}{2m}})^N$. □

In general, few results are known in positive entropy; it would be interesting to find systems of positive entropy for which the minimal pattern complexity is not exponential, or for which the three complexity functions are exponential but all different.

4. Examples of entropy zero

For the systems of zero topological entropy, the three complexity functions can take many kind of values, and thus provide a useful tool for distinguishing between systems.

Example 1: rotation infinite words.

Definition 4.1. Let α irrational, β, x be given, R be the rotation of angle α on the torus, $I_0 = [0, \beta[$, $I_1 = [\beta, 1[$ and let u_n be 0 for the n such that $R^n x$ falls into I_0, 1 otherwise.

The maximal pattern complexity of all these words is $p^*(n) = 2n$ for all n [10]. The complexity, however, is less well-known, except in the elementary case of the *Sturmian words*, that is when $\beta = \alpha$ or $\beta = 1-\alpha$, and $p_u(n) = n+1$ for all n. In the general case, as far as we know, the full computations have been made only in [1]. It is proved there that $p_u(n)$ is the number of atoms of the partition $\vee_{i=0}^{n-1}(I - i\alpha)$, where I is the partition $\{I_0, I_1\}$, and we recall that, for partitions $P^1, \dots P^k$ with r atoms P_t^i, $\vee_{i=1}^k P^i$ is the partition made with all the nonempty $\cap_{i=1}^k P_{t_i}^i$, $1 \leq i \leq k$, $1 \leq t_i \leq r$. Then $p_u(n+1) - p_u(n)$ takes values 1, 2 and 3, with $p_u(n) = 2n$ for n large enough if $\beta \neq p\alpha + q$ for any integers (p, q), as has been known at least since [18], and $p_u(n) = n + p$ for n large enough if $\beta = p\alpha + q$ for integers p and q.

Proposition 4.2. *If $\beta = p\alpha + q$ for some integers (p, q), $p_{*u}(n) = n + 1$ for all n.*

Proof. As is noticed in [10], for any pattern $\tau = \tau_0, \tau_1, \dots, \tau_{n-1}$, the complexity $\#L_u(\tau)$ is the number of atoms of the partition $J^n = \vee_{i=0}^{n-1}(I - \tau_i\alpha)$, where I is the partition $\{I_0, I_1\}$; for $0 \leq i \leq n-1$ to build J^{i+1} from J^i we add on the circle the points $-\tau_{i+1}\alpha$ and $\beta - \tau_{i+1}\alpha$. Now, when $p > 0$, we consider the pattern τ defined by $\tau_i = ip$ for $0 \leq i \leq n-1$, and count the number of atoms of the partitions

https://doi.org/10.1051/ita/2011126 Published online by Cambridge University Press

$J^n = \vee_{i=0}^{n-1}(I - ip\alpha)$: to build J^{i+1} from J^i we add on the circle only one point, $-(i+1)p\alpha$ (modulo 1), thus by induction J^n is the partition of the circle by the points $ip\alpha$ (modulo 1), $-n+1 \leq i \leq 1$, thus $\#L_u(\tau) = n+1$ and $p_{\star u}(n) \leq n+1$ for all n, but it cannot be strictly smaller because of Proposition 2.1; a similar reasoning holds for $p < 0$ with the pattern defined by $\tau_{n-1-i} = (n-1+i)p$, $0 \leq i \leq n-1$. □

Theorem 4.3. *Let $\frac{1}{2} < \beta < 1$, and $\beta \neq p\alpha + q$ for any integers (p,q); let $c = \left[\frac{\beta}{1-\beta}\right]'$, $[x]'$ being the integer part of x if x is not an integer and $x-1$ if x is an integer.*

Then $p_{*u}(n) = n+1$ for $1 \leq n \leq c+1$, $p_{*u}(n) = 2n - c$ for $c+2 \leq n$.

Proof. Again, as in [10], for any pattern τ the complexity $\#L_u(\tau)$ is the number of atoms of the partition $\vee_{i=0}^{n-1}(I - \tau_i\alpha)$; if the atoms are connected, this partition is the partition of the circle by the points $-\tau_i\alpha$ and $\beta - \tau_i\alpha$; at each stage, there are two new points $-\tau_i\alpha$ and $\beta - \tau_i\alpha$, which cannot be any of the $-\tau_j\alpha$ and $\beta - \tau_j\alpha$ for $j < i$; the only atom which may be not connected is $H_n = \cap_{i=0}^{n-1}(I_0 - \tau_i\alpha)$. The growth of $q(n) = \#L_u(\tau_0, \dots \tau_{n-1})$ will depend on how many of the points $-\tau_n\alpha$ and $-\tau_n\alpha + \beta$ fall into H_n.

Namely, we start from $\tau_0 = 0$.

Now, we have $\beta > 1-\beta$; because $-n\alpha$ (modulo 1), $n \in \mathbb{N}$, is dense in $[0,1[$, we can choose τ_1 such that $-\tau_1\alpha$ is in $]\beta-\delta, \beta[$ for a fixed δ small enough; then $-\tau_1\alpha+\beta$ cannot be at β to the right of $-\tau_1\alpha$ as there is not enough space, so it must be at $1-\beta$ to the left of $-\tau_1\alpha$, hence in $]0,\beta[$. Thus both $-\tau_1\alpha$ and $-\tau_1\alpha+\beta$ are in $H_1 = [0,\beta[$, $q(2) - q(1) = 1$, and $H_2 = [0, -\tau_1\alpha + \beta[$ contains the interval $[0, 2\beta - 1 - \delta]$.

Then, if $\beta \geq 2(1-\beta)$ we choose τ_2 such that $-\tau_2\alpha$ is in $]2\beta - 1 - 2\delta, 2\beta - 1 - \delta[$; by the same reasoning both $-\tau_2\alpha$ and $-\tau_2\alpha + \beta$ are in H_2, $q(3) - q(2) = 1$, and $H_3 = [0, -\tau_2\alpha + \beta[$ contains the interval $[0, 3\beta - 2 - 2\delta]$.

In the same way, from stage $i-1$ we get that $H_i = [0, -\tau_{i-1}\alpha + \beta[$ contains the interval $[0, i\beta - i + 1 - (i-1)\delta]$; if $\beta \geq i(1-\beta)$ we choose τ_i such that $-\tau_i\alpha$ is in $]i\beta - i + 1 - i\delta, i\beta - i + 1 - (i-1)\delta[$ and we get $q(i+1) - q(i) = 1$; this is possible for $i = 1, \dots, c$, where c is as above.

Then $H_{c+1} = [0, -\tau_c\alpha + \beta[$ is included in the interval $[0, c\beta - c + 1[$ whose length is (by definition of c) at most $1 - \beta$ (and thus *a fortiori* smaller than β), and $H_n \subset H_{c+1}$ for $n \geq c+1$; thus for these n we cannot find τ_n such that both $-\tau_n\alpha$ and $-\tau_n\alpha + \beta$ are in H_n, as this two points are separated by β on one side, $1-\beta$ on the other side; the best we can do is to find τ_n such that either $-\tau_n\alpha$ or $-\tau_n\alpha + \beta$ are in H_n; that is always possible by density of the $-n\alpha$ (modulo 1), $n \in \mathbb{N}$, and ensures $q(n+1) - q(n) = 2$. Thus, by construction, the pattern $\tau_0, \dots \tau_{n-1}$ realizes the minimal pattern complexity of the sequence, with the claimed formulas. □

https://doi.org/10.1051/ita/2011126 Published online by Cambridge University Press

The above result holds also for $\beta < \frac{1}{2}$ with $c = \left[\frac{1-\beta}{\beta}\right]'$, and for $\beta = \frac{1}{2}$ with $c = 0$.

Corollary 4.4. *If $\beta = \frac{1}{2}$ we get a family of infinite words such that $p_{*u}(n) = p_u^*(n) = p_u(n)$ for all n.*

Infinite words with this property have been extensively studied in [8, 12], and subsequent papers.

Example 2: doubled Sturmian words.

Let v be any Sturmian word and let u be deduced from v by the substitution $0 \to 00$, $1 \to 11$.

Proposition 4.5. *$p_{*u}(n) = n + 1$;*
$p_u^*(n) \leq 4n$ with $\lim_{n\to+\infty} \frac{p_u^*(n)}{4n} = 1$;
$p_u(n) = n + 3$ for n large enough.

Proof. Take the pattern $0, 2, \ldots, 2n-2$; then $L_u(\tau)$ is just $L_v(0, 1, \ldots, n-1)$, hence $p_{*u}(n) \leq p_v(n) \leq n+1$, hence $p_{*u}(n) = n+1$ in view of Proposition 2.1, as u is not ultimately periodic.

Fix now a positive integer r, and take a pattern τ of size $n = pr + q$ with p large and $0 \leq q \leq r-1$, with $\tau_j = 2t_j$ if $j \neq ir+1$, $0 \leq i \leq p$, $\tau_j = 2t_j + 1$ if $j = ir+1$, $0 \leq i \leq p$, and t_{j+1} much larger than t_j. We have $L(\tau) = \{u[2k+\tau], k \in \mathbb{N}\} \cup \{u[2k+1+\tau], k \in \mathbb{N}\}$.

As $u_{2i} = u_{2i+1} = v_i$, we have $\{u[2k+\tau], k \in \mathbb{N}\} = \{v[k+\phi], k \in \mathbb{N}\}$ and $\{u[2k+1+\tau], k \in \mathbb{N}\} = \{v[k+\psi], k \in \mathbb{N}\}$, where $\phi_j = t_j$ if $j \neq ir+1$, $0 \leq i \leq p$, $\phi_j = t_j$ if $j = ir+1$, $0 \leq i \leq p$, and $\psi_j = t_j$ if $j \neq ir+1$, $0 \leq i \leq p$, $\psi_j = t_j + 1$ if $j = ir+1$, $0 \leq i \leq p$.

If p is large enough, we can choose the t_{ri}, $0 \leq i \leq p$, such that for any $k \in \mathbb{N}$, there exists i such that $v_{k+t_{ri}} \neq v_{k+t_{ri}+1}$; this will be true if there is a $t_{ri}\alpha$ (mod. 1) in every interval of the circle of length at least L for some fixed L, and this can be realized by density of the $m\alpha$. This ensures that the sets $\{u[2k+\tau], k \in \mathbb{N}\}$ and $\{u[2k+1+\tau], k \in \mathbb{N}\}$ described above are disjoint. But, as v is a rotation word and, by the same reasoning as in [10] or Theorem 4.3, both $\#\{v[k+\phi], k \in \mathbb{N}\}$ and $\#\{v[k+\psi], k \in \mathbb{N}\}$ are at least $2n - 2p$, thus $\frac{p_u^*(n)}{4n} \geq 1 - \frac{1}{r}$ for every $r \geq 1$ and n larger than some $N(r)$.

And for any pattern τ of size m, $p_u(\tau) \leq \#\{u[2k+\tau], k \in \mathbb{N}\} + \#\{u[2k+1+\tau], k \in \mathbb{N}\}$. In the same way as above, we can define two increasing (though maybe not strictly) sequences of integers ϕ and ψ, of size (defined as for patterns) at most m, such that $\{u[2k+\tau], k \in \mathbb{N}\} = \{v[k+\phi], i \in \mathbb{N}\}$ and $\{u[2k+1+\tau], k \in \mathbb{N}\} = \{v[k+\psi], k \in \mathbb{N}\}$, thus $p_u^*(n) \leq 4n$ by [10].

As v is a Sturmian word, it does not contain 00 or 11 as a factor; suppose for example 11 is forbidden. Take $n = 2p$: the possible factors of length n of u are either $x_1x_1 \ldots x_px_p$ where $x_1 \ldots x_p$ is factor of u, or $x_0x_1x_1 \ldots x_{p-1}x_{p-1}x_p$ where $x_0 \ldots x_p$ is factor of u, and as soon as we see 01 or 10 in a factor of v we are able

https://doi.org/10.1051/ita/2011126 Published online by Cambridge University Press

to say of which of these two forms it is; hence these $2p+3$ factors are all different as soon as the word 0^p is forbidden in u, which will be true for p large enough; a similar reasoning applies for $n = 2p+1$. □

Thus u has the same lower pattern complexity as the Sturmian words, but not the same upper pattern complexity. Note that for example when v is the Fibonacci word then $p_u(n) = n+3$ for all $n \geq 4$.

Example 3: interval exchange transformations.

Definition 4.6. A k-interval exchange transformation T with probability vector $(\alpha_1, \alpha_2, \ldots, \alpha_k)$, and permutation π is defined by

$$T = x + \sum_{\pi(j)<\pi(i)} \alpha_j - \sum_{j<i} \alpha_j,$$

when x is in the interval

$$\left[\sum_{j<i} \alpha_j, \sum_{j\leq i} \alpha_j\right[,$$

T satisfies the infinite distinct orbit condition or i.d.o.c. if the $k-1$ negative orbits of the discontinuities of T are infinite disjoint sets.

Let T be a k-interval exchange transformation satisfying the i.d.o.c condition and let x by a given point. Let $I_1, \ldots I_k$ be the intervals between discontinuities, and $u_n = j$ whenever $T^n x$ falls into I_j.

Proposition 4.7. $p_{*u}(n) = p_u(n) = (k-1)n+1$.

Proof. For any pattern τ the complexity $\#L_u(\tau)$ is the number of atoms of the partition $\vee_{i=0}^{n-1} T^{-\tau_i} I$, and this partition has at least as many atoms as the partition of the interval $[0,1[$ by the points $T^{-\tau_i} d$, $d \in D$, $0 \leq i \leq n-1$, where D is the set made with 0, 1, and the discontinuities of T. The i.d.o.c. condition implies that, for any value of the τ_i, we add exactly $k-1$ points at each stage, and thus the number of elements of this last partition is $(k-1)n+1$. □

Remark 4.8. Though it uses a very heavy measure-theoretic and geometric machinery [2], we recall that if $k \geq 3$, for every "nontrivial" permutation π and almost all (for the Lebesgue measure) probability vectors, k-interval exchange transformations are weakly mixing for one invariant measure, while the i.d.o.c. condition implies minimality [13], thus by Theorem 2.2 their maximal pattern complexity is k^n.

This last result could also be deduced from topological strong mixing for a residual set of 4-interval exchange transformations, see [4].

https://doi.org/10.1051/ita/2011126 Published online by Cambridge University Press

Example 4: lacunary words.

In [10], the authors define an infinite word by putting $u_n = 1$ on an infinite sequence $k_1, \ldots, k_p, \ldots$, with $k_1 = 0$ and $k_{p+1} \geq 2k_p$ for all p, and $u_n = 0$ otherwise. They show that $p_u^*(n) = 2n$.

Proposition 4.9. $p_{*u}(n) = n + 1$.

Proof. Along any pattern of length n with $\tau_{k+1} - \tau_k$ large enough, we see only the words $00\ldots0$ and all the possible $\ldots010\ldots$ for every position of 1, hence the result in view of Proposition 2.1. □

Thus we get non-Sturmian examples with the same lower and upper pattern complexities as the Sturmian words. Note that the complexity depends on the actual k_p [10]. $p_u(n+1) - p_u(n)$ is either 2 (for isolated values of p) or 1.

Example 5: doubled lacunary words.

Let v be the infinite word defined in Example 4, and let u be deduced from v by the substitution $0 \to 00$, $1 \to 11$.

Proposition 4.10. $p_{*u}(n) = n + 1$, $p_u^*(n)$ *is not* $2n$.

Proof. $p_{*u}(n)$ is given by the same reasoning as in Example 4.

As for the maximal pattern complexity: take a pattern $l_0, l_0 + 1,\ l_1, l_1 + 1, \ldots l_{q-1}, l_{q-1} + 1$ such that $l_{i+1} - l_i = 2(k_{p'} - k_p)$ for some p and $p' > p$ which are both large enough and increase fast as i increases. Along this pattern we see all the possible words of length $2q$ with respectively no 1, one 1 and 0 elsewhere, one 11 at even places and 0 elsewhere, one 1111 at even places and 0 elsewhere, and one 101 and 0 elsewhere. Thus $p_u^*(2q) \geq 6q - 2$. □

This provides another example with the same lower pattern complexity as the Sturmian words, but not the same upper pattern complexity.

References

[1] P. Alesssandri, *Codages de rotations et basses complexités*. Université Aix-Marseille II, Ph.D. thesis (1996).

[2] A. Avila and G. Forni, Weak mixing for interval exchange maps and translation flows, *Ann. Math. (2)* **165** (2007) 637–664.

[3] G. Castiglione, A. Restivo and S. Salemi, Patterns in words and languages. *Discrete Appl. Math.* **144** (2004) 237–246.

[4] J. Chaika, Topological mixing for some residual sets of interval exchange transformations. Preprint (2011).

[5] I.P. Cornfeld, S.V. Fomin and Y.G. Sinai, *Ergodic theory. Translated from the Russian by A.B. Sosinski*, *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, New York **245** (1982) x+486.

[6] E.M. Coven and G.A. Hedlund, Sequences with minimal block growth. *Math. Syst. Theory* **7** (1973) 138–153.

https://doi.org/10.1051/ita/2011126 Published online by Cambridge University Press

[7] S. Ferenczi, Complexity of sequences and dynamical systems. Combinatorics and number theory (Tiruchirappalli, 1996). *Discrete Math.* **206** (1999) 145–154.
[8] T. Kamae, Uniform sets and complexity. *Discrete Math.* **309** (2009) 3738−3747.
[9] T. Kamae, Behavior of various complexity functions. Preprint (2011).
[10] T. Kamae and L. Zamboni, Sequence entropy and the maximal pattern complexity of infinite words. *Ergod. Theory Dyn. Syst.* **22** (2002) 1191–1199.
[11] T. Kamae and L. Zamboni, Maximal pattern complexity for discrete systems. *Ergod. Theory Dyn. Syst.* **22** (2002) 1201–1214.
[12] T. Kamae, H. Rao, B. Tan and Y.-M. Xue, Super-stationary set, subword problem and the complexity. *Discrete Math.* **309** (2009) 4417–4427.
[13] M.S. Keane, Non-ergodic interval exchange transformations. *Israël J. Math.* **26** (1977), 188–196.
[14] D. Lind and B. Marcus, *An introduction to symbolic dynamics and coding.* Cambridge University Press, Cambridge (1995) xvi+495.
[15] M. Morse and G.A. Hedlund, Symbolic dynamics. *Amer. J. Math.* **60** (1938) 815–866.
[16] M. Morse and G.A. Hedlund, Symbolic dynamics II. Sturmian trajectories. *Amer. J. Math.* **62** (1940) 1−42.
[17] N. Pytheas-Fogg, *Substitutions in dynamics, arithmetics and combinatorics. Lect. Notes Math.* **1794**, edited by V. Berthé, S. Ferenczi, C. Mauduit and A. Siegel. Springer-Verlag, Berlin (2002).
[18] G. Rote, Sequences with subword complexity $2n$. *J. Number Theory* **46** (1994) 196–213.

Communicated by G. Richomme.
Received November 2, 2010. Accepted October 10, 2011.

https://doi.org/10.1051/ita/2011126 Published online by Cambridge University Press