CDMTCS
Research
Report
Series

Asymptotic Subword Complexity

Ludwig Staiger
Martin-Luther-Universität
Halle-Wittenberg

CDMTCS-412
November 2011

Centre for Discrete Mathematics and
Theoretical Computer Science

Asymptotic Subword Complexity

Ludwig Staiger [*]

Martin-Luther-Universität Halle-Wittenberg
Institut für Informatik
von-Seckendorff-Platz 1, D–06099 Halle (Saale), Germany

Abstract

The subword complexity of an infinite word ξ is a function $f(\xi, n)$ returning the number of finite subwords (factors, infixes) of length n of ξ. In the present paper we investigate infinite words for which the set of subwords occurring infinitely often is a regular language. Among these infinite words we characterise those which are eventually recurrent.

Furthermore, we derive some results comparing the asymptotics of $f(\xi, n)$ to the information content of sets of finite or infinite words related to ξ. Finally we give a simplified proof of Theorem 6 of [Sta98].

[*] email: staiger@informatik.uni-halle.de

Contents

Following [Mar04] the subword complexity of an infinite word ξ is a function $f(\xi,n)$ returning the number of finite subwords (factors, infixes) of ξ having length n. It was mainly investigated for infinite words of low complexity (see [BK03, Mar04] or the book [AS03]). However [Mar04, Question 2] asked for the general complexity of quasiperiodic infinite words. An answer on their maximally possible complexity was given in [PS10] showing that this complexity satisfies $f(\xi,n) \leq_{\mathrm{ae}} c \cdot t_P^n$ where t_P is the smallest Pisot number. Moreover, for quasiperiodic infinite words with maximal subword complexity the set of factors form a regular language.

The aim of our paper is to investigate in more detail those infinite words whose set of factors occurring infinitely often is a regular language. Therefore, in contrast to [BK03] and [AS03] we are mainly interested in infinite words ξ whose subword complexity $f(\xi,n)$ is not bounded by a subexponential function.

In the case of exponentially growing subword complexity the results of [Sta93] and [Sta98] show a close connection between the growth of $f(\xi, n)$ and the Hausdorff dimension of regular ω-languages containing the infinite word ξ. Using this connection we prove that every infinite word having a regular subword language satisfies the condition $f(\xi, n) \approx c \cdot t_\xi^n$ for a suitable real number t_ξ.

As a consequence we obtain a simplified proof of Theorem 6 of [Sta98]. This theorem states, roughly speaking, that finite automata cannot distinguish one-sided eventually recurrent infinite words having the same set of infinitely often occurring factors provided this set of factors is a regular language. A more general result for two-sided infinite words had been obtained earlier [Sem84, PS86].

After introducing some necessary notation in Section 2 we derive some basic facts on infinite words having a regular language of infinitely often occurring factors. Moreover, the concept of asymptotic subword complexity of infinite words is introduced. This concept proves to be useful in the following.

The entropy of languages known from [CM58, Kui70, HPS92] is closely related to asymptotic subword complexity. In Section 3 we derive some elementary properties an also some results relating the entropy of languages to the Hausdorff dimension of ω-languages are presented (cf. also [Sta89, Sta93]). These facts are used to derive our results in the last section. Here we give a characterisation of eventually recurrent infinite words having a regular language of infinitely often occurring subwords. From this characterisation several conditions necessary or sufficient for an infinite word to be eventually recurrent are obtained. Finally, we give a simple proof of Theorem 6 of [Sta98].

The previous proof in [Sta98] uses considerations involving Hausdorff measure. In the present paper we circumvent these measure-theoretic considerations confining to language-theoretic results only, although we make implicitly use of the close connection between the entropy of languages and Hausdorff dimension.

1 Notation

In this section we introduce the notation used throughout the paper. By $\mathbb{N} = \{0, 1, 2, \ldots\}$ we denote the set of natural numbers. Let X be an alphabet of cardinality $|X| = r \geq 2$. By X^* we denote the set of finite words on X, including the *empty word* e, and X^ω is the set of infinite strings (ω-words) over X. Subsets of X^* will be referred to as *languages* and subsets of X^ω as ω-*languages*.

For $w \in X^*$ and $\eta \in X^* \cup X^\omega$ let $w \cdot \eta$ be their *concatenation*. This concatenation product extends in an obvious way to subsets $W \subseteq X^*$ and $B \subseteq X^* \cup X^\omega$. For a language W let $W^* := \bigcup_{i \in \mathbb{N}} W^i$, and let $W^\omega := \{w_1 \cdots w_i \cdots : w_i \in W \setminus \{e\}\}$ the set of infinite strings formed by concatenating words in W.

We denote by $B/w := \{\eta : w \cdot \eta \in B\}$ the *left derivative* of the set $B \subseteq X^* \cup X^\omega$. As usual a language $W \subseteq X^*$ is *regular* provided it is accepted by a finite automaton. An equivalent condition is that its set of left derivatives $\{W/w : w \in X^*\}$ is finite. In the sequel we assume the reader to be familiar with basic facts of language theory.

Furthermore $|w|$ is the *length*[1] of the word $w \in X^*$ and $\mathbf{pref}(B)$ is the set of all finite prefixes of strings in $B \subseteq X^* \cup X^\omega$. We shall abbreviate $w \in \mathbf{pref}(\eta)$ $(\eta \in X^* \cup X^\omega)$ by $w \sqsubseteq \eta$.

$\mathbf{T}(B) := \bigcup_{w \in X^*} \mathbf{pref}(B/w)$ is set of infixes (factors) of words in $B \subseteq X^* \cup X^\omega$, and for an infinite word $\xi \in X^\omega$ its sets of factors occurring infinitely often is $\mathbf{T}_\infty(\xi) := \bigcap_{w \sqsubset \xi} \mathbf{T}(\xi/w)$.

As usual a language $V \subseteq X^*$ is called a *code* provided $w_1 \cdots w_l = v_1 \cdots v_k$ for $w_1, \ldots, w_l, v_1, \ldots, v_k \in V$ implies $l = k$ and $w_i = v_i$. A code V is said to be a *prefix code* provided $v \sqsubseteq w$ implies $v = w$ for $v, w \in V$.

[1] Since there is no danger of confusion, the length $|w|$ of a word $w \in X^*$ is denoted in the same way as the cardinality $|M|$ of a set M.

2 The Languages of Subwords

In this part, we consider, for an infinite word $\xi \in X^\omega$, the languages of subwords $\mathbf{T}(\xi)$ and of subwords occurring infinitely often $\mathbf{T}_\infty(\xi)$, respectively.

For the tails (suffixes) of ξ we have the following obvious inclusion.

$$\mathbf{T}(\xi/w) \supseteq \mathbf{T}(\xi/v) \text{ whenever } w \sqsubseteq v \tag{1}$$

Thus the family $\big(\mathbf{T}(\xi/w)\big)_{w\sqsubset\xi}$ is an infinite decreasing chain of languages, and the infinite intersection $\mathbf{T}_\infty(\xi) := \bigcap_{w\sqsubset\xi} \mathbf{T}(\xi/w)$ consists of all subwords occurring infinitely often in ξ.

It depends on the ω-word ξ whether the chain in Eq. (1) is stationary or not. If the family $\big(\mathbf{T}(\xi/v)\big)_{v\sqsubset\xi}$ is stationary, that is, there is a prefix $v \sqsubset \xi$ such that $\mathbf{T}(\xi/v) = \mathbf{T}_\infty(\xi)$, we will refer to the ω-word $\xi \in X^\omega$ as *eventually recurrent*[2] (see [Tho05]).

Next we consider the case when one of the languages $\mathbf{T}(\xi/w)$ is a regular language. To this end we derive the following relation between $\mathbf{T}(\xi)/v$ and $\mathbf{T}(\xi/v)$.

Lemma 1 *Let $v \sqsubset \xi$. Then $\mathbf{T}(\xi)/v \subseteq \mathbf{T}(\xi/v) = \mathbf{T}(\mathbf{T}(\xi)/v)$.*

Proof. If $u \in \mathbf{T}(\xi)/v$ then $vu \in \mathbf{T}(\xi)$ and thus there is a w such that $wvu \sqsubset \xi$. Since $v \sqsubset \xi$, we have also $v \sqsubseteq wv$. Consequently, $wv = v\bar{w}$ for some $\bar{w}$, and we obtain $v\bar{w}u \sqsubset \xi$, that is, $u \in \mathbf{T}(\xi/v)$.

$\mathbf{T}(\xi)/v \subseteq \mathbf{T}(\xi/v)$ implies $\mathbf{T}(\mathbf{T}(\xi)/v) \subseteq \mathbf{T}(\xi/v)$, so it suffices to show $\mathbf{T}(\xi/v) \subseteq \mathbf{T}(\mathbf{T}(\xi)/v)$. Let $u \in \mathbf{T}(\xi/v)$. Then there is a $\bar{w} \in X^*$ such that $v\bar{w}u \sqsubset \xi$. Consequently, $\bar{w}u \in \mathbf{T}(\xi)/v$, whence $u \in \mathbf{T}(\mathbf{T}(\xi)/v)$. ❏

As in [Sta98] we refer to an ω-word $\xi \in X^\omega$ as *infix-regular* provided there is a prefix $w \sqsubset \xi$ such that $\mathbf{T}(\xi/w)$ is a regular language. The following lemma yields a connection between infix-regular ω-words and eventually recurrent ω-words.

[2] An ω-word ξ is referred to as *recurrent* iff $\mathbf{T}_\infty(\xi) = \mathbf{T}(\xi)$. This resembles the notion of recurrence for $\mathbb{Z}$-words as considered in [Sem84, PS86].

Lemma 2 *An ω-word $\xi \in X^\omega$ is a infix-regular ω-word if and only if ξ is eventually recurrent and $\mathbf{T}_\infty(\xi)$ is a regular language.*

Proof. Let $\xi \in X^\omega$ be infix-regular. Then in Lemma 5 of [Sta98] it is shown that there is a $w' \sqsubset \xi$ such that $\mathbf{T}(\xi/w')$ is a regular language and $\mathbf{T}(\xi/w') = \mathbf{T}_\infty(\xi)$.

The other direction is follows from the definition and the fact that $\mathbf{T}_\infty(\xi)$ is a regular language. ❑

Corollary 1 *If $\mathbf{T}(\xi)$ is regular then $\mathbf{T}_\infty(\xi)$ is also regular.*

It should be noted that not every ω-word ξ for which $\mathbf{T}_\infty(\xi)$ is a regular language is eventually recurrent. The following example shows that $\mathbf{T}_\infty(\xi)$ might be regular, although none of the sets $\mathbf{T}(\xi/w)$, $w \sqsubset \xi$, is regular.

Example 1 Consider $\xi_0 := \prod_{i=1}^{\infty} a^i \cdot b$. Then $\mathbf{T}_\infty(\xi_0) = a^* \cup a^* \cdot b \cdot a^*$, but, for every $w \sqsubset \xi_0$, the intersection $\mathbf{T}(\xi_0/w) \cap b \cdot a^* \cdot b \cdot a^* \cdot b$ is a non-regular language of the form $\{b \cdot a^i \cdot b \cdot a^{i+1} \cdot b : i \in \mathbb{N} \wedge i \geq c_w\}$, hence $\mathbf{T}(\xi_0/w)$ is also non-regular. ❑

2.1 Subword Complexity and Asymptotic Subword Complexity of ω-words

The *subword complexity* of an infinite word ξ is the function $f(\xi,n) := |\mathbf{T}(\xi) \cap X^n|$. In this section we focus on the growth of the function $f(\xi,n)$, in particular, on the real number λ_ξ for which $\lim_{n\to\infty} \left(\frac{f(\xi,n)}{\lambda_\xi+\varepsilon}\right)^n = 0$ and $\lim_{n\to\infty} \left(\frac{f(\xi,n)}{\lambda_\xi-\varepsilon}\right)^n = \infty$. [3].

First observe the following simple property for eventually recurrent ω-words.

Lemma 3 *If $\xi \in X^\omega$ and $\mathbf{T}(\xi/w_0) = \mathbf{T}_\infty(\xi)$ then $f(\xi,n) \leq |w_0| + |\mathbf{T}_\infty(\xi) \cap X^n|$.*

---

[3] We have to express this fact in the complicated manner because the growth of $f(\xi,n)$ need not behave like $c \cdot \lambda_\xi^n$.

Proof. This follows from the fact that every infix of length n of ξ is an infix of ξ/w_0 or an infix of the length $|w_0| + n - 1$ prefix of ξ. ❑

Along with the subword complexity we consider the *asymptotic subword complexity* $\tau(\xi)$ of an ω-word ξ. This quantity is defined as the logarithm of the real number λ_ξ.

$$\tau(\xi) := \lim_{n\to\infty} \frac{\log_{|X|} f(\xi, n)}{n}$$

Definition 1 (Asymptotic subword complexity)

Since $f(\xi, n+m) \le f(\xi, n) \cdot f(\xi, m)$, the limit in Definition 1 exists and equals $\tau(\xi) = \inf\left\{\frac{\log_{|X|} f(\xi,n)}{n} : n \in \mathbb{N}\right\}$. Moreover, we have the following relation between $f(\xi, n)$ and $|\mathbf{T}_\infty(\xi) \cap X^n|$ (see [Sta93, Eq. (5.2)]).

$$\tau(\xi) = \lim_{n\to\infty} \frac{\log_{|X|} |\mathbf{T}_\infty(\xi) \cap X^n|}{n} \tag{2}$$

3 The Entropy of Languages

Closely related with the asymptotic subword complexity is the concept of the entropy of languages introduced in [CM58]. Let $W \subseteq X^*$. Then the quantity

$$\mathsf{H}_W := \limsup_{n\to\infty} \frac{\log_{|X|} \max\{1, |W \cap X^n|\}}{n} \tag{3}$$

is referred to as the *entropy* of the language W. Eq. (3) strongly resembles Eq. (2). Since the limit need not exist, we use the limit superior instead, and the additional 1 in the numerator is added to ensure that $\mathsf{H}_W = 0$ for finite languages W. For more details on the entropy of languages see also [Kui70, HPS92, Sta05].

3.1 The entropy of regular languages

Next we derive some properties of the entropy of regular languages (cf. also [Eil74, Sta93]).

We start with some easily derived relations between the number of words in a regular language and the number of its subwords.

Lemma 4 *If $W \subseteq X^*$ is a regular language then there is a $k \in \mathbb{N}$ such that*

$$|W \cap X^n| \leq |\mathbf{T}(W) \cap X^n| \leq \tfrac{k}{2} \cdot \textstyle\sum_{i=0}^{k} |W \cap X^{n+i}|\,.$$

As a suitable k one may choose twice the number of states of an automaton accepting the language $W \subseteq X^*$.

A first consequence of Lemma 4 is the following.

Corollary 2 *Let $W \subseteq X^*$ be a non-empty regular language. Then* $\mathsf{H}_{\mathbf{T}(W)} = \mathsf{H}_{\mathbf{pref}(W)} = \mathsf{H}_W$.

Corollary 4 of [Sta85] shows a more precise bound for the number of words in regular star languages $W^* \subseteq X^*$.

Lemma 5 *For every regular language $W \subseteq X^*$ there are constants $c_1, c_2 > 0$ and a λ, $0 \leq \lambda \leq |X|$, such that*

$$c_1 \cdot \lambda^n \leq |\mathbf{pref}(W^*) \cap X^n| \leq c_2 \cdot \lambda^n\,.$$

A consequence of Lemma 4 is that $|\mathbf{T}(W) \cap X^n| \leq k \cdot |\mathbf{pref}(W) \cap X^{n+k}|$. Thus Lemma 5 holds also (with constant $k \cdot c_2 \cdot |X|^k$ instead of c_2) for $\mathbf{T}(W^*)$.

In order to obtain a relation between H_W and H_{W^*} we consider, for a language $W \subseteq X^*$, the generating function $\mathsf{S}_W(t) := \sum_{i \in \mathbb{N}} |W \cap X^i| \cdot t^i$. It is well-known (cf. [Kui70]) that $\mathsf{H}_W = -\log_{|X|} \sup\{t : 0 \leq t \leq 1 \wedge \mathsf{S}_W(t) < \infty\}$. Moreover, for regular languages W, the function $\mathsf{S}_W(t)$ is a rational function [CM58, Eil74], that is, in particular, if $W \neq \emptyset$ there is always a value $\mathbf{t}_1 < |X|^{-\mathsf{H}_W}$ such that $\mathsf{S}_W(\mathbf{t}_1) = 1$.

For codes $V \subseteq X^*$ we have $\mathsf{S}_{V^*}(t) = (1 - \mathsf{S}_V(t))^{-1}$, and consequently, $\mathsf{H}_{V^*} = -\log_{|X|} \mathbf{t}_1$ whenever $\mathbf{t}_1 < |X|^{-\mathsf{H}_V}$. Thus we have the following.

Lemma 6 *Let $\emptyset \neq V \subseteq X^*$ be a regular language and simultaneously a code. Then $\mathsf{H}_{V^*} > \mathsf{H}_V$.*

Proposition 1 *If V is a regular code, $v \in V$ and $W = V \setminus \{v\}$ then* $\mathsf{H}_{W^*} < \mathsf{H}_{V^*}$.

Proof. Since V is regular, there is a value $\mathbf{t}_1$ such that $\mathsf{S}_V(\mathbf{t}_1) = 1$, that is, $\mathsf{H}_{V^*} = -\log_{|X|} \mathbf{t}_1$.

We use the inequality $\mathsf{S}_W(t) < \mathsf{S}_V(t)$ which holds for $0 \leq t < |X|^{-\mathsf{H}_V}$ and the fact that W is also a regular code. Then the value $\mathbf{t}'_1$ for which $\mathsf{S}_W(\mathbf{t}'_1) = 1$ satisfies $\mathbf{t}_1 < \mathbf{t}'_1$, and the assertion follows. ❑

We conclude this part with the following connection between the asymptotic subword complexity $\tau(\xi)$ and the entropy of regular languages containing $\mathbf{pref}(\xi)$.

Theorem 1 $\tau(\xi) = \inf\{\mathsf{H}_W : W \textit{ is regular} \wedge \mathbf{pref}(\xi) \subseteq \mathbf{pref}(W)\}$

Proof. The inequality $\tau(\xi) \leq \mathsf{H}_W$ follows from $\tau(\xi) = \mathsf{H}_{\mathbf{T}(\xi)}$, $\mathbf{T}(\xi) \subseteq \mathbf{T}(W)$ and Corollary 2.

Since $\tau(\xi) = \inf\{\frac{\log_{|X|} f(\xi,n)}{n} : n \in \mathbb{N}\}$, the relations $\mathbf{pref}(\xi) \subseteq \mathbf{pref}((\mathbf{T}(\xi) \cap X^n)^*)$, for $n > 0$, and $\mathsf{H}_{(\mathbf{T}(\xi)\cap X^n)^*} = \frac{\log_{|X|} f(\xi,n)}{n}$ show the other inequality. ❑

3.2 Entropy of languages and Hausdorff dimension

In the next sections we will see that the asymptotic subword complexity of an ω-word ξ is closely related to the Hausdorff dimension of certain ω-languages containing ξ. To this end we derive here some properties of the entropy of languages and the Hausdorff dimension of related ω-languages.

The usual definition of Hausdorff dimension (see e.g. [Fal90, Sta93]) is based on measure theoretical notions. Here we avoid this and refer instead to a characterisation via the entropy of languages given in Eq. (3.11) of [Sta93].

Definition 2 Let $F \subseteq X^\omega$. Then

$$\dim_{\mathrm{H}} F := \inf\{\mathsf{H}_W : W \subseteq X^* \wedge F \subseteq \{\xi : |\mathbf{pref}(\xi) \cap W| = \infty\}\}$$

is referred to as the *Hausdorff dimension* of the set F.

We mention the following well-known stability property of the Hausdorff dimension.

$$\dim_{\mathrm{H}} \bigcup_{i\in\mathbb{N}} F_i = \sup\{\dim_{\mathrm{H}} F_i : i \in \mathbb{N}\} \tag{4}$$

In what follows we shall use Eq. (4) mainly to show that $F' \subseteq F$ implies $\dim_{\mathrm{H}} F' \leq \dim_{\mathrm{H}} F$ or that $\dim_{\mathrm{H}} W \cdot F = \dim_{\mathrm{H}} F$ when $W \neq \emptyset$.

Next we consider the *limit* (or *adherence*) $\boldsymbol{ls}\, W := \{\xi : \mathbf{pref}(\xi) \subseteq \mathbf{pref}(W)\} \subseteq X^\omega$ of a language $W \subseteq X^*$.

For languages of the form $\mathbf{T}(V)$ the language itself and its limit $\boldsymbol{ls}\,\mathbf{T}(V)$ satisfy $\mathbf{pref}(\boldsymbol{ls}\,\mathbf{T}(V)) = \mathbf{T}(V)$, $\mathbf{T}(V) \supseteq \mathbf{T}(V)/v$ and $\boldsymbol{ls}\,\mathbf{T}(V) \supseteq (\boldsymbol{ls}\,\mathbf{T}(V))/v$, for $v \in X^*$. Then one can apply Theorem 6 of [Sta89] and obtains

$$\dim_{\mathrm{H}} \boldsymbol{ls}\,\mathbf{T}(V) = \mathsf{H}_{\mathbf{T}(V)}\,. \tag{5}$$

In view of Corollary 2 our Eq. (5) implies $\dim_{\mathrm{H}} \boldsymbol{ls}\, W \leq \mathsf{H}_W$ for regular languages $W \subseteq X^*$. Furthermore, the Hausdorff dimension of the ω-power V^ω equals the entropy of V^* (see Eq. (6.2) of [Sta93]).

$$\dim_{\mathrm{H}} V^\omega = \mathsf{H}_{V^*} \tag{6}$$

Now Corollary 2, Eqs. (5), (6) and Lemma 6 yield the following.

Corollary 3 *Let* $V \subseteq X^*$ *be a regular language. Then* $\dim_{\mathrm{H}} \boldsymbol{ls}\, V \leq \dim_{\mathrm{H}} V^\omega$*, and if, moreover,* V *is a code then* $\dim_{\mathrm{H}} \boldsymbol{ls}\, V < \dim_{\mathrm{H}} V^\omega$*.*

4 Maximum Subword Complexity in Regular ω-languages

In this section we derive the announced above results on eventually recurrent ω-words having a regular language of infinitely often occurring subwords. To this end we investigate the relations between the asymptotic subword complexity $\tau(\xi)$ of an ω-word ξ and its containment in ω-languages of a special shape. Here we consider the class of regular ω-languages (see [Sta97a, Tho90]), that is, the class of ω-languages accepted by finite automata. This class of regular ω-languages is closely related to regular languages.

As usual an ω-language $F \subseteq X^\omega$ is referred to as *regular* provided there are an $n \in \mathbb{N}$ and regular languages $W_i, V_i \subseteq X^*$ such that

$$F = \bigcup_{i=1}^{n} W_i \cdot V_i^\omega .$$

Here the languages V_i can be chosen to be prefix codes (see [Cho74]). We mention still that the class of regular ω-languages is closed under Boolean operations (see [Sta97a, Tho90]).

In the sequel we need the identity

$$\boldsymbol{ls}\, V^* = V^\omega \cup V^* \cdot \boldsymbol{ls}\, V \text{ for } V \subseteq X^* \tag{7}$$

which can be found in [Sta97b] and the fact that $\boldsymbol{ls}\, V$ is a regular ω-language whenever V is a regular language (see [Sta93, Sta97a]).

Then the following relation between the asymptotic subword complexity and the Hausdorff dimension of regular ω-languages can be proved.

$$\tau(\xi) = \inf\{\dim_{\mathrm{H}} F : F \subseteq X^\omega \wedge F \text{ is regular } \wedge \xi \in F\} \tag{8}$$

Proof. Since $\xi \in \boldsymbol{ls}\, W$ if and only if $\mathbf{pref}(\xi) \subseteq \mathbf{pref}(W)$ and $\boldsymbol{ls}\, W$ is regular provided W is regular, the inequality "$\geq$" follows from Theorem 1 and Eq. (5), and the reverse inequality is Proposition 5.4 of [Sta93]. ❑

We proceed with a relation between $\mathbf{T}_\infty(\xi)$ and an ω-power V^ω containing a tail of ξ.

Lemma 7 *1. If $\xi \in w \cdot V^\omega$ for some $w \in X^*$ then $\mathbf{T}_\infty(\xi) \subseteq \mathbf{T}(V^*) \subseteq \mathbf{T}(V) \cdot V^* \cdot \mathbf{T}(V)$.*

2. If η is eventually recurrent then there is a $w \in X^*$ such that $\eta \in w \cdot \boldsymbol{ls}\, \mathbf{T}_\infty(\eta)$.

Proof. The first assertion is immediate.

Since η is eventually recurrent, $\mathbf{T}_\infty(\eta) = \mathbf{T}(\eta/w)$ for some $w \sqsubset \eta$. Thus $\{\eta\} = \boldsymbol{ls}\, w \cdot \mathbf{pref}(\eta/w) \subseteq w \cdot \boldsymbol{ls}\, \mathbf{T}_\infty(\eta)$. ❑

This yields an obvious upper bound on $\tau(\xi)$ when $\xi \in w \cdot V^\omega$.

Corollary 4 *If $\xi \in w \cdot V^\omega$ then $\tau(\xi) \leq \mathsf{H}_{\mathbf{T}(V^*)}$.*

For regular codes $V \subseteq X^*$ we have a stronger property.

Theorem 2 *Let $V \subseteq X^*$ be a regular code, $\xi \in w \cdot V^\omega$ for some $w \in X^*$ and $\tau(\xi) = \mathsf{H}_{V^*}$. Then $V^* \subseteq \mathbf{T}_\infty(\xi) = \mathbf{T}(V^*)$.*

Proof. The inclusion $\mathbf{T}_\infty(\xi) \subseteq \mathbf{T}(V^*)$ is Lemma 7.1, and together with $V^* \subseteq \mathbf{T}_\infty(\xi)$ it implies $\mathbf{T}_\infty(\xi) = \mathbf{T}(V^*)$. Thus, it remains to show $V^* \subseteq \mathbf{T}_\infty(\xi)$.

Assume the contrary, that is, there is a $v_0 \in V^*$ such that $v_0 \notin \mathbf{T}_\infty(\xi)$. Since, for $n > 0$, $V^\omega = (V^n)^\omega$ and V^n is also a regular code whenever V is a regular code, we may assume $v_0 \in V$. Set $W := V \setminus \{v_0\}$.

Then $\xi \in w \cdot W^\omega$, and according to Corollary 4 and Proposition 1 we have $\tau(\xi) \leq \mathsf{H}_{W^*} < \mathsf{H}_{V^*}$. This contradicts our assumption. ❑

4.1 Eventually recurrent ω-words with regular $\mathbf{T}_\infty(\xi)$

Theorem 2 allows us to derive conditions necessary or sufficient for an ω-word ξ with a regular language $\mathbf{T}_\infty(\xi)$ to be eventually recurrent.

The first condition is a sufficient one.

Theorem 3 *Let $F \subseteq X^\omega$ be a regular ω-language. If $\xi \in F$ and $\tau(\xi) = \dim_{\mathrm{H}} F$ then ξ is eventually recurrent and $\mathbf{T}_\infty(\xi)$ is a regular language.*

Proof. Since F is regular and $\xi \in F$ there are a word $w \in X^*$ and a regular prefix code $V \subseteq X^*$ such that $\xi \in w \cdot V^\omega \subseteq F$. Corollaries 4 and 2 and Eq. (6) show that $\tau(\xi) \leq \mathsf{H}_{V^*} = \dim_{\mathrm{H}} V^\omega \leq \dim_{\mathrm{H}} F$.

Now the assertion follows with Theorem 2. ❑

The next two conditions are necessary ones.

Lemma 8 *If ξ is eventually recurrent and $\mathbf{T}_\infty(\xi)$ is a regular language then there is a regular prefix code $V \subseteq X^*$ such that $\mathbf{T}_\infty(\xi) = \mathbf{T}(V^*)$.*

Proof. Lemma 7.2 shows $\xi \in w \cdot \mathbf{ls}\, \mathbf{T}_\infty(\xi)$ for a suitable $w \sqsubset \xi$. By assumption, the ω-language $w \cdot \mathbf{ls}\, \mathbf{T}_\infty(\xi) = \mathbf{ls}\,(w \cdot \mathbf{T}_\infty(\xi))$ is regular. Thus there is a regular prefix code such that $\xi \in w' \cdot V^\omega \subseteq \mathbf{ls}\,(w \cdot \mathbf{T}_\infty(\xi))$ and according to Theorem 2 we have $\mathbf{T}_\infty(\xi) = \mathbf{T}(V^*)$. ❑

Together with Lemmata 5 and 3 we obtain the following.

Corollary 5 *If ξ is eventually recurrent and $\mathbf{T}_\infty(\xi)$ is a regular language then there are constants $c_1, c_2 > 0$ such that*

$$c_1 \cdot |X|^{\tau(\xi)\cdot n} \leq |\mathbf{T}_\infty(\xi) \cap X^n| \leq |\mathbf{T}(\xi) \cap X^n| \leq c_2 \cdot |X|^{\tau(\xi)\cdot n}.$$

The conditions in Lemma 8 and Corollary 5 are, however, not sufficient as will be seen in the subsequent example. To this end we derive a relation between $\mathbf{T}(\xi)$ and $\mathbf{T}_\infty(\xi)$.

Lemma 9 *Let $M_\xi := \mathrm{Min}_{\mathrm{infix}}(\mathbf{T}(\xi) \setminus \mathbf{T}_\infty(\xi))$ the set of minima w.r.t. to the infix relation of $\mathbf{T}(\xi) \setminus \mathbf{T}_\infty(\xi)$. If every $w \in M_\xi$ occurs only once as a factor in ξ then $|\mathbf{T}(\xi) \cap X^n| \leq |\mathbf{T}_\infty(\xi) \cap X^n| + \sum_{w \in M_\xi} \max\{0, n - |w| + 1\}$.*

Proof. If $v \in \mathbf{T}(\xi) \setminus \mathbf{T}_\infty(\xi)$ then some $w \in M_\xi$ is a subword of v. Since w occurs only once as a factor in ξ, v is one of the $|v| - |w| + 1$ factors of length $|v|$ of ξ containing w. ❑

Example 2 Let $V := (aa)^* \cdot ab$. Then $\mathsf{H}_{V^*} = \frac{1}{2}$. We use an enumeration $\{v_i : i \in \mathbb{N}\}$ of $V^* \setminus \{e\}$ and set $\xi_1 := \prod_{i \in \mathbb{N}} v_i a^{2i} b$. Then $\mathbf{T}_\infty(\xi_1) = \mathbf{T}(V^*)$, $M_{\xi_1} = b(aa)^*b$ and every word of M_{ξ_1} occurs only once as a factor in ξ_1.

Using Lemma 9 we calculate $|\mathbf{T}(\xi_1) \cap X^n| \leq |\mathbf{T}_\infty(\xi_1) \cap X^n| + n^2$, and thus the inequality of Corollary 5 is satisfied although every $\mathbf{T}(\xi_1/w) \setminus \mathbf{T}_\infty(\xi_1)$, $w \sqsubset \xi_1$, contains infinitely many words from $b(aa)^*b$. ❑

It should be mentioned that the ω-word ξ_0 from Example 1 satisfies $\mathbf{T}_\infty(\xi_0) = a^*ba^* \cup a^*$, whence $|\mathbf{T}_\infty(\xi_0) \cap X^n| = n+1$ and $\tau(\xi_0) = 0$. Thus Corollary 5 yields another proof that ξ_0 is not eventually recurrent.

4.2 A new proof of Theorem 6 of [Sta98]

Theorem 2 and Lemma 7 allow us to simplify the proof of Theorem 6 in [Sta98]. We start with an auxiliary lemma.

Lemma 10 *Let $F \subseteq X^\omega$ be regular, $\xi \in F$ and $\tau(\xi) = \dim_{\mathrm{H}} F$. If η is eventually recurrent and $\mathbf{T}_\infty(\xi) = \mathbf{T}_\infty(\eta)$ then there are $u, u' \in X^*$ such that $u' \cdot (\eta/u) \in F$.*

Proof. First Theorem 3 shows that ξ is eventually recurrent and $\mathbf{T}_\infty(\xi)$ is a regular language. Thus, for a suitable $w \sqsubset \xi$, $F \cap w \cdot \mathbf{ls}\,\mathbf{T}_\infty(\xi)$ is a regular language containing ξ. Consequently, there are a $u' \sqsubset \xi$ and a regular prefix code $V \subseteq X^*$ such that $\xi \in u' \cdot V^\omega \subseteq F \cap w \cdot \mathbf{ls}\,\mathbf{T}_\infty(\xi)$. Now, it suffices to prove $\eta \in X^* \cdot V^\omega$. Then $\eta \in u \cdot V^\omega$ and, consequently, $u' \cdot (\eta/u) \in u' \cdot V^\omega \subseteq F$.

To this end observe that in view of $\mathsf{H}_{V^*} = \dim_{\mathrm{H}} V^\omega \geq \tau(\xi) = \dim_{\mathrm{H}} F$ Theorem 2 and Lemma 7.2 imply $\mathbf{T}(V^*) = \mathbf{T}_\infty(\xi) = \mathbf{T}_\infty(\eta)$ and $\eta \in v \cdot \mathbf{ls}\,\mathbf{T}(V^*)$ for a suitable $v \sqsubset \eta$. From $\mathbf{T}(V^*) \subseteq \mathbf{T}(V) \cdot V^* \cdot \mathbf{T}(V)$ and Eq. (7) we obtain $\mathbf{ls}\,\mathbf{T}(V^*) \subseteq \mathbf{T}(V) \cdot V^* \cdot \mathbf{ls}\,\mathbf{T}(V) \cup \mathbf{T}(V) \cdot V^\omega$. Since V is a regular prefix code, in view of Corollary 3 we have $\dim_{\mathrm{H}} \mathbf{ls}\,\mathbf{T}(V) < \dim_{\mathrm{H}} V^\omega = \tau(\eta)$. This shows $\eta \in v \cdot \mathbf{T}(V) \cdot V^\omega$. ❑

Now we can drop the assumption that $\xi \in F$ but have to ensure that ξ is eventually recurrent and $\mathbf{T}_\infty(\xi)$ is regular.

Theorem 4 *Let $F \subseteq X^\omega$ be regular, ξ, η be eventually recurrent and $\mathbf{T}_\infty(\xi) = \mathbf{T}_\infty(\eta)$ be a regular language.*

If $\xi \in F$ then there are $u, u' \in X^*$ such that $u' \sqsubset \xi$ and $u' \cdot (\eta/u) \in F$.

Proof. Since ξ is eventually recurrent and $\mathbf{T}_\infty(\xi)$ is regular there is a $u' \sqsubset \xi$ such that $\xi \in u' \cdot \mathbf{ls}\,\mathbf{T}_\infty(\xi)$ and $\mathbf{ls}\,\mathbf{T}_\infty(\xi)$ is a regular ω-

language. Moreover, $\tau(\xi) = \dim_{\mathrm{H}} \textbf{\textit{ls}}\,\mathrm{T}_\infty(\xi)$. Now apply Lemma 10 to the ω-language $F \cap u' \cdot \textbf{\textit{ls}}\,\mathrm{T}_\infty(\xi)$. ❑

Our Example 2 shows that the assumption that η be eventually recurrent cannot be dropped in Theorem 4 and Lemma 10. Take e.g. $F := \left((aa)^* \cdot ab\right)^\omega$, $\xi := \prod_{i\in\mathbb{N}} v_i$ and $\eta := \xi_1$.

References

[AS03] Jean-Paul Allouche and Jeffrey Shallit. *Automatic sequences*. Cambridge University Press, Cambridge, 2003. Theory, applications, generalizations.

[BK03] Jean Berstel and Juhani Karhumäki. Combinatorics on words: a tutorial. *Bulletin of the EATCS*, 79:178–228, 2003.

[Cho74] Yaacov Choueka. Theories of automata on ω-tapes: a simplified approach. *J. Comput. System Sci.*, 8:117–141, 1974.

[CM58] Noam Chomsky and George A. Miller. Finite state languages. *Information and Control*, 1:91–112, 1958.

[Eil74] Samuel Eilenberg. *Automata, languages, and machines. Vol. A*. Academic Press [A subsidiary of Harcourt Brace Jovanovich, Publishers], New York, 1974. Pure and Applied Mathematics, Vol. 58.

[Fal90] Kenneth Falconer. *Fractal geometry*. John Wiley & Sons Ltd., Chichester, 1990.

[HPS92] Georges Hansel, Dominique Perrin, and Imre Simon. Compression and entropy. In A. Finkel and M. Jantzen, editors, *STACS 92 (Cachan, 1992)*, volume 577 of *Lecture Notes in Computer Science*, pages 515–528, Berlin, 1992. Springer-Verlag.

[Kui70] Werner Kuich. On the entropy of context-free languages. *Information and Control*, 16:173–200, 1970.

[Mar04] Solomon Marcus. Quasiperiodic infinite words (column: Formal language theory). *Bulletin of the EATCS*, 82:170–174, 2004.

[PS86] Dominique Perrin and Paul E. Schupp. Automata on the integers, recurrence distinguishability, and the equivalence and decidability of monadic theories. In *Proceedings, Symposium on Logic in Computer Science*, pages 301–304, Cambridge, Massachusetts, June 16–18 1986. IEEE Computer Society.

[PS10] Ronny Polley and Ludwig Staiger. The maximal subword complexity of quasiperiodic infinite words. In *Electronic Proceedings in Theoretical Computer Science*, volume 31, pages 169–176, 2010.

[Sem84] Aleksei L. Semenov. Decidability of monadic theories. In Michal P. Chytil, editor, *Mathematical foundations of computer science, 1984 (Prague, 1984)*, volume 176 of *Lecture Notes in Computer Science*, pages 162–175, Berlin, 1984. Springer-Verlag.

[Sta85] Ludwig Staiger. The entropy of finite-state ω-languages. *Problems Control Inform. Theory/Problemy Upravlen. Teor. Inform.*, 14(5):383–392, 1985.

[Sta89] Ludwig Staiger. Combinatorial properties of the Hausdorff dimension. *J. Statist. Plann. Inference*, 23(1):95–100, 1989.

[Sta93] Ludwig Staiger. Kolmogorov complexity and Hausdorff dimension. *Inform. and Comput.*, 103(2):159–194, 1993.

[Sta97a] Ludwig Staiger. ω-languages. In Grzegorz Rozenberg and Arto Salomaa, editors, *Handbook of Formal Languages*, volume 3, pages 339–387. Springer-Verlag, Berlin, 1997. Beyond words.

[Sta97b] Ludwig Staiger. On ω-power languages. In Gheorghe Păun and Arto Salomaa, editors, *New trends in formal languages*, volume 1218 of *Lecture Notes in Computer Science*, pages 377–394. Springer-Verlag, Berlin, 1997. Control, cooperation, and combinatorics.

[Sta98] Ludwig Staiger. Rich ω-words and monadic second-order arithmetic. In Mogens Nielsen and Wolfgang Thomas, editors, *Computer science logic (Aarhus, 1997)*, volume 1414 of *Lecture Notes in Computer Science*, pages 478–490. Springer-Verlag, Berlin, 1998. Selected papers from the 11th International Workshop (CSL '97) held at the 6th Annual Conference of the European Association for Computer Science Logic (EACSL) at the University of Aarhus, Aarhus, August 23–29, 1997.

[Sta05] Ludwig Staiger. The entropy of Łukasiewicz-languages. *Theor. Inform. Appl.*, 39(4):621–639, 2005.

[Tho90] Wolfgang Thomas. Automata on infinite objects. In Jan van Leeuwen, editor, *Handbook of theoretical computer science, Vol. B*, pages 133–191. Elsevier Science Publishers B.V., Amsterdam, 1990. Formal models and semantics.

[Tho05] Klaus Thomsen. Languages of finite words occurring infinitely many times in an infinite word. *Theor. Inform. Appl.*, 39(4):641–650, 2005.