CDMTCS Research Report Series

Quasiperiods, Subword Complexity and the Smallest Pisot Number

Ronny Polley
itCampus Software- und Systemhaus GmbH, Leipzig

Ludwig Staiger
Martin-Luther-Universität Halle-Wittenberg

CDMTCS-479
March 2015, revised September 2016

Centre for Discrete Mathematics and Theoretical Computer Science

Quasiperiods, Subword Complexity and the smallest Pisot Number

Ronny Polley [*]
IT Sonix custom development GmbH
Georgiring 3, D-04103 Leipzig, Germany
and
Ludwig Staiger [†]
Martin-Luther-Universität Halle-Wittenberg
Institut für Informatik
von-Seckendorff-Platz 1, D–06099 Halle (Saale), Germany

Abstract

A quasiperiod of a finite or infinite string/word is a word whose occurrences cover every part of the string. A word or an infinite string is referred to as quasiperiodic if it has a quasiperiod. It is obvious that a quasiperiodic infinite string cannot have every word as a subword (factor). Therefore, the question arises how large the set of subwords of a quasiperiodic infinite string can be [Mar04].

Here we show that on the one hand the maximal subword complexity of quasiperiodic infinite strings and on the other hand the size of the sets of maximally complex quasiperiodic infinite strings both are intimately related to the smallest Pisot number t_P (also known as *plastic constant*).

We provide an exact estimate on the maximal subword complexity for quasiperiodic infinite words.

Keywords: quasiperiodic ω-words, subword complexity, Hausdorff measure

[*] email: r.polley@itsonix.eu
[†] email: ludwig.staiger@informatik.uni-halle.de

Contents

In his tutorial [Mar04] Solomon Marcus discussed some open questions on quasiperiodic infinite words. Soon after its publication Levé and Richomme gave answers on some of the open problems (see [LR04]). In connection with Marcus' Question 2 they presented a quasiperiodic infinite word (with quasiperiod aba) of exponential subword complexity, and they posed the new question of what is the maximal complexity of a quasiperiodic infinite word.

In a recent paper [PS10] we estimated the maximal asymptotic (in the sense of [Sta12]) subword complexity of quasiperiodic infinite words. More precisely, it is shown in [PS10] that every quasiperiodic infinite word ξ has at most $f(\xi,n) \le O(1)\cdot t_P^n$ factors (subwords) of length n, where t_P is the smallest Pisot number, that is, the unique positive root of the polynomial $t^3 - t - 1$. Moreover, the general construction of Section 5 in [Sta93] yields quasiperiodic infinite words achieving this bound. In fact, also Levé's and Richomme's [LR04] example meets this asymptotic upper bound $O(1)\cdot t_P^n$.

Surprisingly, it turned out in [PS10] that there are also infinite words meeting this bound having $aabaa$—a different word—as quasiperiod. Moreover, it was shown that all other quasiperiods yield infinite words asymptotically below this bound.

The aim of this paper is to compare these two maximal quasiperiods *aba* and *aabaa* in order to obtain an answer as to which one of them yields infinite words of greater complexity. Here we compare the quasiperiods *aba* and *aabaa* in two respects.

1. Which one of the words *aba* or *aabaa* generates the larger set (ω-language) of infinite words having q as quasiperiod, and

2. which one of the words *aba* or *aabaa* generates an ω-word ξ_q having a maximal subword function $f(\xi_q, n)$?

As a measure of ω-languages in Item 1 we use the Hausdorff dimension and Hausdorff measure of a subset of the Cantor space of infinite words (ω-words). We obtain that, when neglecting the fixed prefix q of quasiperiodic ω-words having the quasiperiod q the sets of ω-words having quasiperiod *aba* or *aabaa* have the same Hausdorff dimension $\log t_P$ and the same Hausdorff measure t_p, both values showing the close connection to the smallest Pisot number.

A difference for these quasiperiods appears when we consider the subword function $f(\xi, n)$. It turns out that *aabaa* is the quasiperiod having the maximally achievable subword complexity for quasiperiodic ω-words.

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 $L \subseteq X^*$ and $B \subseteq X^* \cup X^\omega$. For a language L let $L^* := \bigcup_{i \in \mathbb{N}} L^i$, and by $L^\omega := \{w_1 \cdots w_i \cdots : w_i \in L \setminus \{e\}\}$ we denote the set of infinite strings formed by concatenating words in L. Furthermore $|w|$ is the *length* 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 $w \sqsubseteq \eta$.

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 $L \subseteq X^*$ is *regular* provided it is accepted by a finite automaton. An equivalent condition is that its set of left derivatives $\{L/w : w \in X^*\}$ is finite.

The sets of infixes of B or η are $\mathbf{infix}(B) := \bigcup_{w\in X^*} \mathbf{pref}(B/w)$ and $\mathbf{infix}(\eta) := \bigcup_{w\in X^*} \mathbf{pref}(\{\eta\}/w)$, respectively. Similarly $\mathbf{suff}(B) := \bigcup_{w\in X^*} B/w$ is the set of suffixes of elements of B. In the sequel we assume the reader to be familiar with basic facts of language theory.

2 Quasiperiodicity

2.1 General properties

A finite or infinite word $\eta \in X^* \cup X^\omega$ is referred to as *quasiperiodic* with quasiperiod $q \in X^* \setminus \{e\}$ provided for every $j < |\eta| \in \mathbb{N} \cup \{\infty\}$ there is a prefix $u_j \sqsubseteq \eta$ of length $j - |q| < |u_j| \leq j$ such that $u_j \cdot q \sqsubseteq \eta$, that is, for every $w \sqsubseteq \eta$ the relation $u_{|w|} \sqsubseteq w \sqsubset u_{|w|} \cdot q$ is valid (cf. [LR04, Mar04, Mou00]).

Next we introduce the finite language P_q ($\mathcal{L}(q)$ in [Mou00]) which generates the set of quasiperiodic ω-words having quasiperiod q. We set

$$P_q := \{v : e \sqsubset v \sqsubseteq q \sqsubset v \cdot q\} = \{v : \exists w (w \sqsubset q \wedge v \cdot w = q)\}. \tag{1}$$

Corollary 4 in [PS10] yields the following characterisation of ω-words having quasiperiod $q \in X^* \setminus \{e\}$.

$$\xi \text{ has quasiperiod } q \text{ if and only if } \mathbf{pref}(\xi) \subseteq \mathbf{pref}(P_q^\omega) \tag{2}$$

We list some further properties of the set of quasiperiodic ω-words which will be useful in the sequel.

Proposition 1

$$P_q^\omega = \{\xi : \mathbf{pref}(\xi) \subseteq \mathbf{pref}(P_q^\omega)\} \tag{3}$$

There is a $V \subseteq \mathbf{infix}(P_q^{|q|})$ *such that*

$$P_q^\omega = q \cdot V^\omega. \tag{4}$$

Proof. Since P_q is finite, Eq. (3) follows from Eq. (2).

For the proof of the second identity observe that every word in $P_q^{|q|}$ starts with the quasiperiod q. Then the assertion follows from the identity $P_q^\omega = (P_q^k)^\omega$, $k \geq 1$, and the rotation property $(W_1 \cdot W_2)^\omega = W_1 \cdot (W_2 \cdot W_1)^\omega$ where $W_1, W_2 \subseteq X^*$. ❑

Eq. (3) shows that the set of quasiperiodic ω-words P_q^ω belongs to the class of ω-languages $F \subseteq X^\omega$ satisfying the property $F = \{\xi : \xi \in X^\omega \wedge \mathbf{pref}(\xi) \subseteq \mathbf{pref}(F)\}$. Those ω-languages are the ones *closed* in the CANTOR topology of the set X^ω (see [Sta97, Tho90]). Moreover, if $\mathbf{pref}(F)$ is a regular language, these ω-languages F can be accepted by every deterministic finite automaton $\mathcal{B} = (X, S, s_0, \delta, S_F)$ accepting $\mathbf{pref}(F)$ (see Fig. 1).

$$\xi \in F \longleftrightarrow \forall w(w \in \mathbf{pref}(\xi) \to \delta(s_0, w) \in S_F)\,. \tag{5}$$

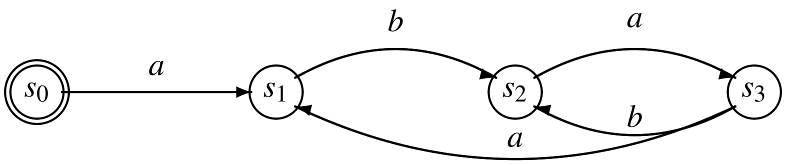

Figure 1: Deterministic partial automaton accepting P_{aba}^ω, all states being final.

The proof of Eq. (4) yields a rather large language V. Using the rotation property successively starting with P_q yields a more concise version of V. We exhibit this for the set $P_{aabaa} = \{aab, aaba, aabaa\}$.

Example 2 $$\begin{aligned} P_{aabaa}^\omega &= \{aab, aaba, aabaa\}^\omega = \big(aab\cdot\{e, a, aa\}\big)^\omega \\ &= aab\cdot\{aab, aaab, aaaab\}^\omega \\ &= aabaa\cdot\{baa, abaa, aabaa\}^\omega \end{aligned}$$

Observe that the resulting language $\{baa, abaa, aabaa\}$ is prefix-free. Moreover $\{baa, abaa, aabaa\} = \mathcal{R}(aabaa)$ in terms of [Mou00].

In the same way one obtains $P_{aba}^\omega = aba\cdot\{ba, aba\}^\omega$. □

3 Hausdorff Dimension and Hausdorff Measure

3.1 General properties

First, we shall briefly describe the basic formulae needed for the definition of Hausdorff measure and Hausdorff dimension of a subset of X^ω. For more background see the textbooks [Edg08, Fal90] or Section 1 in [MS94].

In the setting of languages and ω-languages this can be read as follows (see [MS94, Sta93]). For $F \subseteq X^\omega$, $r = |X| \geq 2$ and $0 \leq \alpha \leq 1$ the equation

$$\mathbb{L}_\alpha(F) := \lim_{l\to\infty} \inf\Big\{\sum_{w\in W} r^{-\alpha\cdot|w|} : F \subseteq W\cdot X^\omega \wedge \forall w(w \in W \to |w| \geq l)\Big\} \tag{6}$$

defines the α-dimensional metric outer measure on X^ω. The measure $\mathbb{L}_\alpha$ satisfies the following properties (see [MS94, Sta93, Sta15]).

Proposition 3 *Let $F \subseteq X^\omega$, $V \subseteq X^*$ and $\alpha \in [0,1]$.*

1. *If $\mathbb{L}_\alpha(F) < \infty$, then $\mathbb{L}_{\alpha+\varepsilon}(F) = 0$, for all $\varepsilon > 0$.*
2. *It holds the scaling property $\mathbb{L}_\alpha(w \cdot F) = r^{-\alpha \cdot |w|} \cdot \mathbb{L}_\alpha(F)$.*

Then the *Hausdorff dimension* of F is defined as

$$\dim F := \sup\{\alpha : \alpha = 0 \vee \mathbb{L}_\alpha(F) = \infty\} = \inf\{\alpha : \mathbb{L}_\alpha(F) = 0\}.$$

It should be mentioned that $\dim$ is countably stable and invariant under scaling, that is, for $F_i \subseteq X^\omega$ we have

$$\dim \bigcup_{i \in \mathbb{N}} F_i = \sup\{\dim F_i : i \in \mathbb{N}\} \quad \text{and} \quad \dim w \cdot F_0 = \dim F_0 . \tag{7}$$

If $\alpha = \dim F$ then we call $\mathbb{L}_\alpha(F)$ the *Hausdorff measure* of F.

We have the following relation between a language of finite words V and the Hausdorff dimension of its ω-power V^ω.

Proposition 4 ([Sta93, Eq. (6.2)]) *Let $V \subseteq X^*$ and V^ω be non-empty. Then $\dim V^\omega = \limsup_{n\to\infty} \frac{1}{n} \log_r |V^* \cap X^n|$.*

3.2 The Hausdorff measure of P_{aba}^ω and P_{aabaa}^ω

In Section 4.1 in [PS10] the value $\limsup_{n\to\infty} \frac{1}{n} \log_r |P_w^* \cap X^n|$, for $w \in \{aba, aabaa\}$, was found as $\log_r t_P$ where t_P is the smallest Pisot number, that is, the (unique) positive root of the polynomial $t^3 - t - 1$. Thus Proposition 4 shows that the Hausdorff dimension of P_{aba}^ω and P_{aabaa}^ω is $\log_r t_P$.

In what concerns the Hausdorff measure of P_{aba}^ω and P_{aabaa}^ω we consider the identities $P_{aabaa}^\omega = aabaa \cdot \{baa, abaa, aabaa\}^\omega$ and $P_{aba}^\omega = aba \cdot \{ba, aba\}^\omega$ derived in Example 2.

Here the languages $\{baa, abaa, aabaa\}$ and $\{ba, aba\}$ are prefix-free. Therefore we can apply Theorem 4 of [Sta05] which gives the following general formula for the Hausdorff measure of V^ω for prefix-free languages $V \subseteq X^*$.

Theorem 5 ([Sta05, Theorem 4]) *Let $V \subseteq X^*$ be prefix-free and $\alpha = \dim V^\omega$. Then*

$$\mathbb{L}_\alpha(V^\omega) = \begin{cases} 0 & , \text{ if } \sum_{u\in V} r^{-\alpha\cdot|u|} < 1, \text{ and} \\ \inf\{\big(\sum_{wv\in V} r^{-\alpha\cdot|v|}\big)^{-1} : w \in \mathbf{pref}(V)\}, & \text{if } \sum_{u\in V} r^{-\alpha\cdot|u|} = 1. \end{cases}$$

Observe that $\sum_{v\in V} r^{-\alpha\cdot|v|} > 1$ implies $\alpha < \dim V^\omega$ (see e.g. Proposition 3 in [Sta05]).

Example 6 For $V = \{ba, aba\}$ we have

$w\in \mathbf{pref}(V)$	$w\in V\cup\{e\}$	a	ab	b
$\sum_{wv\in V} r^{-\alpha\cdot\|v\|}$	1	t_P^{-2}	t_P^{-1}	t_P^{-1}

and for $V = \{baa, abaa, aabaa\}$

$w\in \mathbf{pref}(V)$	$w\in V\cup\{e\}$	a	aa	aab	$aaba$	b	ba
$\sum_{wv\in V} r^{-\alpha\cdot\|v\|}$	1	$(t_P^{-3}+t_P^{-4})$	t_P^{-3}	t_P^{-2}	t_P^{-1}	t_P^{-2}	t_P^{-1}

Since $t_P^{-3}+t_P^{-4} = t_P^{-1} < 1$, we obtain

$$\mathbb{L}_\alpha(\{ba,aba\}^\omega) = \mathbb{L}_\alpha(\{baa,abaa,aabaa\}^\omega) = 1$$

and in view of the scaling property Proposition 3(2), $\mathbb{L}_\alpha(P_{aba}) = t_P^{-3}$ and $\mathbb{L}_\alpha(P_{aabaa}) = t_P^{-5}$. □

3.3 The Hausdorff measure of $\mathbf{suff}(P^\omega_{aba})$ and $\mathbf{suff}(P^\omega_{aabaa})$

The estimates $\mathbb{L}_\alpha(P_{aabaa}) = t_P^{-5}$ and $\mathbb{L}_\alpha(P_{aba}) = t_P^{-3}$, however, do not seem to represent the 'real' size of the sets P^ω_{aba} and P^ω_{aabaa}: All ω-words in P^ω_{aba} start with aba and all ω-words in P^ω_{aabaa} start with the longer word $aabaa$. Thus, in view of the scaling property Proposition 3(2), these prefixes contribute the factors t_P^{-3} and t_P^{-5}, respectively, to the Hausdorff measure, whereas according to Example 6 the tails $\{ba,aba\}^\omega$ and $\{baa,abaa,aabaa\}^\omega$ both have Hausdorff measure 1.

In order to eliminate this influence of the scaling down of the Hausdorff measures we consider instead the sets $\mathbf{suff}(P^\omega_q)$ of all tails (suffixes) of ω-words in P^ω_q. In view of Eq. (7) we have

$$\dim \mathbf{suff}(P^\omega_q) = \dim \bigcup_{w\in X^*} P^\omega_q/w = \dim P^\omega_q .$$

The following proposition enables us to derive a representation of $\mathbf{suff}(P^\omega_q)$ suitable for calculating its Hausdorff measure.

Proposition 7 *Let the ω-language $F \subseteq X^\omega$ satisfy the condition $F = \{\xi : \xi \in X^\omega \wedge \mathbf{pref}(\xi) \subseteq \mathbf{pref}(F)\}$ and let its set of left derivatives $\{F/w : w \in X^*\}$ be finite. Then $\{\mathbf{suff}(F)/w : w \in X^*\}$ is also finite and $\mathbf{suff}(F) = \{\xi : \xi \in X^\omega \wedge \mathbf{pref}(\xi) \subseteq \mathbf{infix}(F)\}$.*

Both of the assumptions in Proposition 7 are essential. First consider that $\{E/w : w\in X^*\}$ be finite. Here the ω-language $E = \{a^\omega\}\cup\bigcup_{n\in\mathbb{N}} a^n b\{a,b\}^n\cdot a^\omega$ has infinitely many left derivatives, for $E/a^n b \neq E/a^m b$ unless $n = m$.

We have $\mathbf{infix}(E) = \{a,b\}^*$ but $\mathbf{suff}(E) \neq \{a,b\}^\omega$. Next $\{a,b\}^* \cdot a^\omega \subseteq \{a,b\}^\omega$ has $F/w = F/v$ for $w,v \in \{a,b\}^*$ but F does not satisfy the condition $F = \{\xi : \xi \in X^\omega \wedge \mathbf{pref}(\xi) \subseteq \mathbf{pref}(F)\}$. Here we have $\mathbf{suff}(F) = F \neq \{a,b\}^\omega = \{\xi : \mathbf{pref}(\xi) \in \{a,b\}^*\}$.

Proof of Proposition 7. Let $V \subseteq X^*$ be finite such that $\{F/w : w \in X^*\} = \{F/w : w \in V\}$. Then in view of $\{\mathbf{suff}(F)/w : w \in X^*\} \subseteq \{\bigcup_{v \in V'} F/v : V' \subseteq V\}$ the set of left derivatives of $\mathbf{suff}(F)$ is obviously finite.

Concerning the second assertion, the inclusion "$\subseteq$" follows from $\mathbf{pref}(\xi/w) \subseteq \mathbf{infix}(F)$ whenever $w \in X^*$ and $\xi \in F$.

Let now $\mathbf{pref}(\zeta) \subseteq \mathbf{infix}(F)$. Then for every $v \in \mathbf{pref}(\zeta)$ there is a w_v with $v \in \mathbf{pref}(F/w_v)$. Since the set $\{F/w : w \in X^*\}$ is finite, there is a $w \in \mathbf{pref}(F)$ such that the language $W_{\zeta,w} := \{v : v \in \mathbf{pref}(\zeta) \wedge F/w_v = F/w\}$ is infinite. Then $\mathbf{pref}(W_{\zeta,w}) = \mathbf{pref}(\zeta)$ and from the assumption that $F = \{\xi : \mathbf{pref}(\xi) \subseteq \mathbf{pref}(F)\}$ we obtain $w \cdot \zeta \in F$. ❏

Proposition 7 yields $\mathbf{suff}(P_q^\omega) = \{\xi : \xi \in X^\omega \wedge \mathbf{pref}(\xi) \subseteq \mathbf{infix}(P_q^\omega)\}$. In Table 1 the partial automata $\mathcal{B}_{aba} = (\{a,b\}, \{s_0,s_1,s_2,s_3\}, s_0, \delta_{aba})$ and $\mathcal{B}_{aabaa} = (\{a,b\}, \{z_0,z_1,z_2,z_3,z_4,z_5,z_6\}, z_0, \delta_{aabaa})$ accepting the languages $\mathbf{infix}(P_{aba}^\omega)$ and $\mathbf{infix}(P_{aabaa}^\omega)$, respectively, are given.

$\mathcal{B}_{aba}$	s_0	s_1	s_2	s_3
a	s_1	s_2		s_1
b	s_3	s_3	s_3	

$\mathcal{B}_{aabaa}$	z_0	z_1	z_2	z_3	z_4	z_5	z_6
a	z_1	z_2	z_3	z_4		z_6	z_2
b	z_5	z_5	z_5	z_5	z_5		

Table 1: Partial automata $\mathcal{B}_{aba}$ and $\mathcal{B}_{aabaa}$ accepting the languages $\mathbf{infix}(P_{aba}^\omega)$ and $\mathbf{infix}(P_{aabaa}^\omega)$, respectively

In view of Proposition 7 these automata accept also the ω-languages $\mathbf{suff}(P_{aba}^\omega)$ and $\mathbf{suff}(P_{aabaa}^\omega)$. Moreover, replacing the initial states s_0 by s_1 and z_0 by z_2 the modified automata accept the ω-languages $\{ba,aba\}^\omega$ and $\{baa,abaa,aabaa\}^\omega$, respectively. Consequently, the automaton $\mathcal{B}_{aba}$ accepts the ω-language $\{a,ba\} \cdot \{ba,aba\}^\omega$ and the automaton $\mathcal{B}_{aabaa}$ accepts $\{aa,baa,abaa\} \cdot \{baa,abaa,aabaa\}^\omega$.

Using $\mathbb{L}_\alpha(\{ba,aba\}^\omega) = \mathbb{L}_\alpha(\{baa,abaa,aabaa\}^\omega) = 1, \alpha = \log_r t_P$, the scaling property Proposition 3, the facts that the unions are disjoint and $\mathbb{L}_\alpha$ is an outer measure we obtain $\mathbb{L}_\alpha(\mathbf{suff}(P_{aba}^\omega)) = t_P^{-1} + t_P^{-2} = t_P$ and also $\mathbb{L}_\alpha(\mathbf{suff}(P_{aabaa}^\omega)) = t_P^{-2} + t_P^{-3} + t_P^{-4} = t_P$.

In summary, if we neglect the influence of the prefixes, with respect to Hausdorff measure and Hausdorff dimension both maximal quasiperiods have the same behaviour. Our results support also the close connection between the smallest Pisot number t_P and the sets of quasiperiodic ω-words of largest complexity P_{aba}^{ω} and P_{aabaa}^{ω}.

4 Subword Complexity

4.1 The subword complexity of quasiperiodic ω-words

In this section we recall some results from [PS10] on the subword complexity function $f(\xi,n)$ for quasiperiodic ω-words. If $\xi \in X^{\omega}$ is quasiperiodic with quasiperiod q then Eq. (2) shows $\mathbf{infix}(\xi) \subseteq \mathbf{infix}(P_q^{\omega})$. Thus

$$f(\xi,n) \leq |\mathbf{infix}(P_q^{\omega}) \cap X^n| \text{ for } \xi \in P_q^{\omega}. \tag{8}$$

Similarly to the proof of Proposition 5.5 in [Sta93] let $\xi_q := \prod_{v \in P_q^* \setminus \{e\}} v$ where the order of the factors $v \in P_q^* \setminus \{e\}$ is an arbitrary but fixed well-order, e.g. the length-lexicographical order.

This implies $\mathbf{infix}(\xi_q) = \mathbf{infix}(P_q^{\omega})$. Consequently, the tight upper bound on the subword complexity of quasiperiodic ω-words having a certain quasiperiod q is $f_q(n) := |\mathbf{infix}(P_q^{\omega}) \cap X^n|$.

The following facts are known from the theory of formal power series (cf. [BP85, SS78]). As $\mathbf{infix}(P_q^{\omega})$ is a regular language the power series $\sum_{n\in\mathbb{N}} f_q(n)\cdot t^n$ is a rational series and, therefore, f_q satisfies a recurrence relation

$$f_q(n+k) = \textstyle\sum_{i=0}^{k-1} m_i \cdot f_q(n+i) \tag{9}$$

with integer coefficients $m_i \in \mathbb{Z}$. Thus $f_q(n) = \sum_{i=0}^{k'-1} g_i(n)\cdot\lambda_i^n$ where $k' \leq k$, λ_i are pairwise distinct roots of the polynomial $\chi_q(t) = t^n - \sum_{i=0}^{k-1} m_i \cdot t^i$ and g_i are polynomials of degree not larger than k.

The growth of $f_q(n)$ mainly depends on the (positive) root λ_q of largest modulus among the λ_i and the corresponding polynomial g_i. Using Corollary 4 in [Sta85] (see also Eq. (8) in [PS10]) one can show—without explicitly inspecting the polynomials $\chi_q(t)$—that the polynomial g_i corresponding to the maximal root λ_q is constant.

Lemma 8 ([PS10, Lemma 16]) *Let $q \in X^* \setminus \{e\}$. Then there are constants $c_{q,1}, c_{q,2} > 0$ and a $\lambda_q \geq 1$ such that*

$$c_{q,1}\cdot\lambda_q^n \leq |\mathbf{infix}(P_q^*) \cap X^n| \leq c_{q,2}\cdot\lambda_q^n.$$

The quasiperiods *aba* and *aabaa* yield the largest value of λ_q among all quasiperiods.

Lemma 9 ([PS10, Lemma 18]) *Let X be an arbitrary alphabet containing at least the two letters a,b. Then the maximal value λ_q is obtained for $q = aba$ or $q = aabaa$.*
This value is $\lambda_{aba} = \lambda_{aabaa} = t_P$ where t_P is the positive root of the polynomial $t^3 - t - 1$.

Remark 10 The bound in Lemma 9 is independent of the size of the alphabet X.

4.2 Quasiperiods of maximal subword complexity

We have seen that the quasiperiods aba and $aabaa$ yield quasiperiodic ω-words of maximal asymptotic subword complexity. In this section we investigate which one of these two quasiperiods $q \in \{aba, aabaa\}$ yields ω-words $\xi_q \in \{a,b\}^\omega$ of larger subword complexity $f(\xi_q, n) = f_q(n)$.

From the deterministic automata $\mathcal{B}_{aba}$ and $\mathcal{B}_{aabaa}$ (see Table 1) accepting the languages $\mathbf{infix}(P^*_{aba})$ and $\mathbf{infix}(P^*_{aabaa})$, respectively, we obtain the adjacency matrices

$$\mathcal{A}_{aba} = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 0 & \mathbf{0} & \mathbf{1} & \mathbf{1} \\ 0 & \mathbf{1} & \mathbf{0} & \mathbf{0} \\ 0 & \mathbf{0} & \mathbf{1} & \mathbf{0} \end{pmatrix} \quad \text{and} \quad \mathcal{A}_{aabaa} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & \mathbf{0} & \mathbf{1} & \mathbf{0} & \mathbf{1} & \mathbf{0} \\ 0 & 0 & \mathbf{0} & \mathbf{0} & \mathbf{1} & \mathbf{1} & \mathbf{0} \\ 0 & 0 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{1} & \mathbf{0} \\ 0 & 0 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{1} \\ 0 & 0 & \mathbf{1} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix}. \tag{10}$$

Remark 11 Here the boldface entries are irreducible square submatrices which correspond to the automata accepting the ω-languages $\{ba, aba\}^\omega$ and $\{baa, abaa, aabaa\}^\omega$, respectively, obtained from $\mathcal{B}_{aba}$ and $\mathcal{B}_{aabaa}$ by replacing the initial states s_0 by s_1 and z_0 by z_2.

Then we obtain $f_q(n) = |\mathbf{infix}(P_q^\omega \cap X^n)|$ as the vector-matrix-vector product $f_q(n) = (1,0,\ldots,0) \cdot \mathcal{A}_q^n \cdot (1,\ldots,1)^\perp$ (see Chapter II.9 in [SS78] or [MS94]).

The characteristic polynomials of $\mathcal{A}_{aba}$ and $\mathcal{A}_{aabaa}$ are

$$\begin{aligned} \chi_{aba}(t) &= t \cdot (t^3 - t - 1) \text{ and} \\ \chi_{aabaa}(t) &= t^2 \cdot (t^3 - t - 1) \cdot (t^2 + 1) = t^7 + t^4 + t^3 + t^2. \end{aligned} \tag{11}$$

Thus the sequence $\left(f_{aba}(n)\right)_{n \in \mathbb{N}}$ satisfies the recurrence relation

$$f_{aba}(n+4) = f_{aba}(n+2) + f_{aba}(n+1)$$

and $\big(f_{aabaa}(n)\big)_{n\in\mathbb{N}}$ satisfies

$$f_{aabaa}(n+7) = f_{aabaa}(n+4) + f_{aabaa}(n+3) + f_{aabaa}(n+2)\,.$$

Since the polynomial $\chi_{aba}(t)$ divides $\chi_{aabaa}(t)$, both sequences $\big(f_{aba}(n)\big)_{n\in\mathbb{N}}$ and $\big(f_{aabaa}(n)\big)_{n\in\mathbb{N}}$ satisfy the recurrence relation

$$f_q(n+7) = f_q(n+4) + f_q(n+3) + f_q(n+2)\,.$$

The initial values are $(1,2,3,4,5,7,9)$ for $q = aba$ (see also [LR04]) and $(1,2,3,4,6,8,10)$ for $q = aabaa$. This gives already evidence that the sequence $\big(f_{aabaa}(n)\big)_{n\in\mathbb{N}}$ grows faster than $\big(f_{aba}(n)\big)_{n\in\mathbb{N}}$.

In order to calculate the growth of $(f_q(n))_{n\in\mathbb{N}}$ where $q \in \{aba, aabaa\}$ more accurately, we use standard methods of recurrent relations (see [GKP94] or Chapter 3 of [Hal67])

The non-zero roots of the polynomials $\chi_{aba}(t)$ and $\chi_{aabaa}(t)$ are the roots t_P, t_1, t_2 of $t^3 - t - 1$ and, for $\chi_{aabaa}(t)$ additionally, i and $-\mathrm{i}$ where $\mathrm{i} = \sqrt{-1}$ is the imaginary unit. The roots t_P, t_1, t_2 satisfy the relations

$$\begin{aligned} t_P + t_1 + t_2 &= 0, \\ t_P \cdot t_1 \cdot t_2 &= 1 \\ t_P &> 1 \text{ and} \\ |t_1| = |t_2| &< 1\,. \end{aligned}$$

Since both characteristic polynomials have only simple non-zero roots, $f_{aba}(n)$ and $f_{aabaa}(n)$ satisfy the following identities (cf. [BR88, GKP94, SS78]).

$$\begin{aligned} f_{aba}(n) &= \gamma_1 \cdot t_P^n + \gamma_2 \cdot t_1^n + \gamma_3 \cdot t_2^n,\ n \geq 1 \text{ and} && (12) \\ f_{aabaa}(n) &= \gamma'_1 \cdot t_P^n + \gamma'_2 \cdot t_1^n + \gamma'_3 \cdot t_2^n + \gamma'_4 \cdot \mathrm{i}^n + \gamma'_5 \cdot (-\mathrm{i})^n,\ n \geq 2. && (13) \end{aligned}$$

For the function $f_{aabaa}(n)$ the following initial conditions hold.

$$\begin{aligned} f_{aabaa}(2) &= 3 = \gamma'_1 \cdot t_P^2 + \gamma'_2 \cdot t_1^2 + \gamma'_3 \cdot t_2^2 + \gamma'_4 \cdot \mathrm{i}^2 + \gamma'_5 \cdot (-\mathrm{i})^2 \\ f_{aabaa}(3) &= 4 = \gamma'_1 \cdot t_P^3 + \gamma'_2 \cdot t_1^3 + \gamma'_3 \cdot t_2^3 + \gamma'_4 \cdot \mathrm{i}^3 + \gamma'_5 \cdot (-\mathrm{i})^3 \\ f_{aabaa}(4) &= 6 = \gamma'_1 \cdot t_P^4 + \gamma'_2 \cdot t_1^4 + \gamma'_3 \cdot t_2^4 + \gamma'_4 \cdot \mathrm{i}^4 + \gamma'_5 \cdot (-\mathrm{i})^4 \qquad (14) \\ f_{aabaa}(5) &= 8 = \gamma'_1 \cdot t_P^5 + \gamma'_2 \cdot t_1^5 + \gamma'_3 \cdot t_2^5 + \gamma'_4 \cdot \mathrm{i}^5 + \gamma'_5 \cdot (-\mathrm{i})^5 \\ f_{aabaa}(6) &= 10 = \gamma'_1 \cdot t_P^6 + \gamma'_2 \cdot t_1^6 + \gamma'_3 \cdot t_2^6 + \gamma'_4 \cdot \mathrm{i}^6 + \gamma'_5 \cdot (-\mathrm{i})^6 \end{aligned}$$

Then $f_{aabaa}(5) - f_{aabaa}(3) - f_{aabaa}(2) = 1$ and $f_{aabaa}(6) - f_{aabaa}(4) - f_{aabaa}(3) = 0$ in view of $t^3 = t + 1$ for $t \in \{t_P, t_1, t_2\}$ imply

$$\begin{aligned} 2 \cdot \mathrm{i} \cdot (\gamma'_4 - \gamma'_5) + (\gamma'_4 + \gamma'_5) &= 1\,, \text{ and} \\ \mathrm{i} \cdot (\gamma'_4 - \gamma'_5) - 2 \cdot (\gamma'_4 + \gamma'_5) &= 0 \end{aligned} \qquad (15)$$

which in turn yields $\gamma'_4+\gamma'_5=\frac{1}{5}$ and $\gamma'_4-\gamma'_5=-\frac{2\cdot i}{5}$. Thus we may reduce the numbers of equations in Eq. (14) to three.

$$\begin{array}{rcccl} f_{aabaa}(2) & = & 3 & = & \gamma'_1\cdot t_P^2+\gamma'_2\cdot t_1^2+\gamma'_3\cdot t_2^2-1/5 \\ f_{aabaa}(3) & = & 4 & = & \gamma'_1\cdot t_P^3+\gamma'_2\cdot t_1^3+\gamma'_3\cdot t_2^3-2/5 \\ f_{aabaa}(4) & = & 6 & = & \gamma'_1\cdot t_P^4+\gamma'_2\cdot t_1^4+\gamma'_3\cdot t_2^4+1/5 \end{array} \tag{16}$$

And for $f_{aba}(n)$ we obtain the following three equations from the initial conditions.

$$\begin{array}{rcccl} f_{aba}(1) & = & 2 & = & \gamma_1\cdot t_P+\gamma_2\cdot t_1+\gamma_3\cdot t_2 \\ f_{aba}(2) & = & 3 & = & \gamma_1\cdot t_P^2+\gamma_2\cdot t_1^2+\gamma_3\cdot t_2^2 \\ f_{aba}(3) & = & 4 & = & \gamma_1\cdot t_P^3+\gamma_2\cdot t_1^3+\gamma_3\cdot t_2^3 \end{array} \tag{17}$$

To solve these for the values of γ_1 and γ'_1, respectively, we use Cramer's rule.

$$\gamma_1=\begin{vmatrix} 2 & t_1 & t_2 \\ 3 & t_1^2 & t_2^2 \\ 4 & t_1^3 & t_2^3 \end{vmatrix}\cdot\begin{vmatrix} t_P & t_1 & t_2 \\ t_P^2 & t_1^2 & t_2^2 \\ t_P^3 & t_1^3 & t_2^3 \end{vmatrix}^{-1} \text{ and } \gamma'_1=\begin{vmatrix} \frac{16}{5} & t_1^2 & t_2^2 \\ \frac{22}{5} & t_1^3 & t_2^3 \\ \frac{29}{5} & t_1^4 & t_2^4 \end{vmatrix}\cdot\begin{vmatrix} t_P^2 & t_1^2 & t_2^2 \\ t_P^3 & t_1^3 & t_2^3 \\ t_P^4 & t_1^4 & t_2^4 \end{vmatrix}^{-1}$$

The following auxiliary consideration alleviates the calculation of the determinants. Here we use the identities $t_1+t_2=-t_P$, $t_1\cdot t_2=t_P^{-1}$, which hold for the roots t_P,t_1,t_2 of t^3-t-1.

$$\begin{aligned} \begin{vmatrix} x & 1 & 1 \\ y & t_1 & t_2 \\ z & t_1^2 & t_2^2 \end{vmatrix} &= (t_2-t_1)\cdot\begin{vmatrix} x & 1 & 0 \\ y & t_1 & 1 \\ z & t_1^2 & t_2+t_1 \end{vmatrix}=(t_2-t_1)\cdot\begin{vmatrix} x & 1 & 0 \\ y & 0 & 1 \\ z & -t_1\cdot t_2 & t_2+t_1 \end{vmatrix} \\ &= (t_2-t_1)\cdot\frac{y\cdot t_P^2+z\cdot t_P+x}{t_P} \\ &= (t_2-t_1)\cdot\bigl(x\cdot t_P^2+y\cdot t_P+(z-x)\bigr) \end{aligned} \tag{18}$$

Extracting common factors, applying the auxiliary Eq. (18) and reducing to lowest terms we obtain from Eq. (4.2)

$$\gamma_1 = \frac{2\cdot t_P^2+3\cdot t_P+2}{2\cdot t_P+3}\approx 1{,}6787356, \text{ and} \tag{19}$$

$$\gamma'_1 = \frac{13\cdot t_P^2+16\cdot t_P+9}{5\cdot(2\cdot t_P+3)}\approx 1{,}876608 \tag{20}$$

Since $|t_1|=|t_2|<1$ we have $|\gamma_2\cdot t_1^n+\gamma_3\cdot t_2^n|<\frac{1}{2}$ and $|\gamma'_2\cdot t_1^n+\gamma'_3\cdot t_2^n\pm|\frac{2}{5}||<\frac{1}{2}$ for sufficiently large $n\in\mathbb{N}$. Taking $|\gamma'_4+\gamma'_5|\leq\frac{2}{5}$ into account, Eqs. (12) and (13) show that for these $n\in\mathbb{N}$ the values of $f_{aba}(n)$ and $f_{aabaa}(n)$ are the integers closest to $\gamma_1\cdot t_P^n$ and $\gamma'_1\cdot t_P^n$, respectively.

Acknowledgement

The authors thank the referees for critical and helpful comments that improved the paper.

References

[BP85] Jean Berstel and Dominique Perrin. *Theory of codes*, volume 117 of *Pure and Applied Mathematics*. Academic Press Inc., Orlando, FL, 1985.

[BR88] Jean Berstel and Christophe Reutenauer. *Rational series and their languages*, volume 12 of *EATCS Monographs on Theoretical Computer Science*. Springer-Verlag, Berlin, 1988.

[Edg08] Gerald Edgar. *Measure, topology, and fractal geometry*. Undergraduate Texts in Mathematics. Springer-Verlag, New York, second edition, 2008.

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

[GKP94] Ronald L. Graham, Donald E. Knuth, and Oren Patashnik. *Concrete mathematics*. Addison-Wesley Publishing Company, Reading, MA, second edition, 1994.

[Hal67] Marshall Hall, Jr. *Combinatorial theory*. Blaisdell Publishing Co., Waltham, Mass., 1967.

[LR04] Florence Levé and Gwénaël Richomme. Quasiperiodic infinite words: Some answers (column: Formal language theory). *Bulletin of the EATCS*, 84:128–138, 2004.

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

[Mou00] Laurent Mouchard. Normal forms of quasiperiodic strings. *Theoret. Comput. Sci.*, 249(2):313–324, 2000.

[MS94] Wolfgang Merzenich and Ludwig Staiger. Fractals, dimension, and formal languages. *RAIRO Inform. Théor. Appl.*, 28(3-4):361–386, 1994.

[PS10] Ronny Polley and Ludwig Staiger. The maximal subword complexity of quasiperiodic infinite words. In Ian McQuillan and Giovanni Pighizzini, editors, *DCFS*, volume 31 of *Electr. Proc. Theor. Comput. Sci. (EPTCS)*, pages 169–176, 2010.

[SS78] Arto Salomaa and Matti Soittola. *Automata-theoretic aspects of formal power series*. Springer-Verlag, New York, 1978.

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

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

[Sta97] Ludwig Staiger. ω-languages. In Grzegorz Rozenberg and Arto Salomaa, editors, *Handbook of Formal Languages*, volume 3, pages 339–387. Springer-Verlag, Berlin, 1997.

[Sta05] Ludwig Staiger. Infinite iterated function systems in Cantor space and the Hausdorff measure of ω-power languages. *Internat. J. Found. Comput. Sci.*, 16(4):787–802, 2005.

[Sta12] Ludwig Staiger. Asymptotic subword complexity. In Henning Bordihn, Martin Kutrib, and Bianca Truthe, editors, *Languages Alive*, volume 7300 of *Lecture Notes in Computer Science*, pages 236–245. Springer-Verlag, Heidelberg, 2012.

[Sta15] Ludwig Staiger. On the Hausdorff measure of regular ω-languages in Cantor space. *Discrete Mathematics & Theoretical Computer Science*, 17(1):357–368, 2015.

[Tho90] Wolfgang Thomas. Automata on infinite objects. In Jan van Leeuwen, editor, *Handbook of Theoretical Computer Science, Vol. B*, pages 133–191. Elsevier, Amsterdam, 1990.