CDMTCS
Research
Report
Series

Quasiperiods, Subword Complexity and the smallest Pisot Number

Ronny Polley[1] **and Ludwig Staiger**[2]
[1] itCampus Software- und Systemhaus GmbH, Leipzig
[2] Martin-Luther-Universität Halle-Wittenberg

CDMTCS-479
March 2015

Centre for Discrete Mathematics and
Theoretical Computer Science

Quasiperiods, Subword Complexity and the smallest Pisot Number

Ronny Polley
itCampus Software- und Systemhaus GmbH
D-04229 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 word or an infinite string is a word which covers 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.

*email: `ludwig.staiger@informatik.uni-halle.de`

Contents

1 Notation **3**

2 Quasiperiodicity **4**

2.1 General properties 4

3 Hausdorff Dimension and Hausdorff Measure **5**

3.1 General properties 5

3.2 The Hausdorff measure of P_{aba}^{ω} and P_{aabaa}^{ω} 6

4 Subword Complexity **8**

4.1 The subword complexity of quasiperiodic ω-words 8

4.2 Quasiperiods of maximal subword complexity 9

A Calculating the Constants **11**

In his tutorial [Mar04] Solomon Marcus discussed some open questions on quasiperiodic infinite words. Soon after its publication Levé and Richomme [LR04] gave answers on some of the open problems. 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 $f(\xi, n) \leq O(1) \cdot t_P^n$ 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 [Sta93, Section 5] yields quasiperiodic infinite words achieving this bound. In fact, also Levé's and Richomme's [LR04] example meets this upper bound.

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 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 this quasiperiod q, for both words, 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 constant in the bound on $f(\xi, n)$. It turns out that the bounding constants c_{aba} and c_{aabaa} satisfy $c_{aba} < c_{aabaa}$, thus *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 $u_{|w|} \sqsubset w \sqsubseteq u_{|w|} \cdot q$ is valid (cf. [LR04, Mar04]).

Next we introduce the finite language P_q which generates the set of quasiperiodic ω-words having quasiperiod q. We set

$$P_q := \{v : e \sqsubset v \sqsubseteq q \sqsubset v \cdot q\}. \tag{1}$$

Corollary 4 of [PS10] yields the following characterisation of ω-words having quasiperiod.

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

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

Proposition 2.1

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

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

$$P_q^\omega = q \cdot (W \cdot q)^\omega. \tag{4}$$

Proof. Eq. (3) is Eq. (4) of [PS10].

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 $P_q^\omega = (P_q^k)^\omega$, $k \geq 1$, Eq. (3) and the rotation property $(V \cdot W)^\omega = V \cdot (W \cdot V)^\omega$. ❑

Proposition 2.2 *If* $F = \{\xi : \xi \in X^\omega \wedge \mathbf{pref}(\xi) \subseteq \mathbf{pref}(F)\}$ *and its set of left derivatives* $\{F/w : w \in X^*\}$ *is finite then* $\{\mathbf{suff}(F)/w : w \in X^*\}$ *is also finite and* $\mathbf{suff}(F) = \{\xi : \xi \in X^\omega \wedge \mathbf{pref}(\xi) \subseteq \mathbf{infix}(F)\}$.

The assumption that $\{F/w : w \in X^*\}$ be finite is essential, consider e.g. $F = \{a^\omega\} \cup \bigcup_{n \in \mathbb{N}} a^n b\{a,b\}^n \cdot a^\omega$. Here $\mathbf{infix}(F) = \{a,b\}^*$ but $\mathbf{suff}(F) \neq \{a,b\}^\omega$.

Proof. Let $V \subseteq X^*$ be finite such that $\{F/w : w \in X^*\} = \{F/w : w \in V\}$. Then $\{\mathbf{suff}(F)/w : w \in X^*\} \subseteq \{\bigcup_{v \in V'} F/v : V' \subseteq V\}$ 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 are w_v and ζ_v such that $v \cdot \zeta_v \in F/w_v$. Since the set $\{F/w : w \in X^*\}$ is finite, there are infinitely many $v \in \mathbf{pref}(\zeta)$ such that $F/w_v = F/w$ for some $w \in \mathbf{pref}(F)$. Consider the infinite set $W_{\zeta,w} := \{v : v \in \mathbf{pref}(\zeta) \wedge F/w_v = F/w\}$. Then $\mathbf{pref}(W_{\zeta,w}) = \mathbf{pref}(\zeta)$ and, since $F = \{\xi : \mathbf{pref}(\xi) \subseteq \mathbf{pref}(F)\}$, we obtain $w \cdot \zeta \in F$. ❏

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 and motivation see Section 1 of [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

$$\mathbb{L}_\alpha(F) := \lim_{l\to\infty} \inf\Big\{ \sum_{w\in W} r^{-\alpha\cdot|w|} : F \subseteq W \cdot X^\omega \wedge \forall w(w \in W \Rightarrow |w| \geq l)\Big\} \tag{5}$$

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

Proposition 3.1 *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

$$\dim \textstyle\bigcup_{i\in\mathbb{N}} F_i = \sup\{\dim F_i : i \in \mathbb{N}\} \quad \text{and} \quad \dim w \cdot F_0 = \dim F_0. \tag{6}$$

We have the following relations between languages of finite words and the Hausdorff dimension (cf. [MS94, Sta93]).

Proposition 3.2 1. *Let $V \subseteq X^*$. Then* $\dim V^\omega = \limsup_{n\to\infty} \frac{\log_{|X|} |V \cap X^n|}{n}$.

2. *If* $F = \{\xi : \mathbf{pref}(\xi) \subseteq \mathbf{pref}(F)\}$ *and* $\mathbf{pref}(F)$ *is a regular language then* $\dim F = \lim_{n\to\infty} \frac{\log_{|X|} |\mathbf{pref}(F)\cap X^n|}{n}$.

From Eq. (2) and Proposition 2.1 we see that the sets of ω-words having quasiperiod q have a special shape. With respect to Hausdorff measure ω-languages having this shape satisfy the following properties.

Lemma 3.3 *Let* $V \subseteq X^*$ *and* $\dim V^\omega = \alpha$. *Then*

1. $\mathbb{L}_\alpha(V^\omega) \le 1$, *and*
2. *if* V *is a regular language then* $\mathbb{L}_\alpha(V^\omega) > 0$.
3. *If* $F = \{\xi : \xi \in X^\omega \wedge \mathbf{pref}(\xi) \subseteq \mathbf{pref}(F)\}$, $M_F = \{F/w : w \in \mathbf{pref}(F)\}$ *is finite and* $\alpha = \dim F$ *then* $\mathbb{L}_\alpha(F) \ge |X|^{-\alpha\cdot(|M_F|-1)}$.

3.2 The Hausdorff measure of P^ω_{aba} and P^ω_{aabaa}

The Hausdorff dimension of P^ω_{aba} and P^ω_{aabaa} can be easily estimated from Proposition 3.2 and [PS10, Section 4] as $\log_{|X|} t_P$ where t_P the smallest Pisot number, that is, the (single) positive root of the polynomial $t^3 - t - 1$.

What concerns the Hausdorff measure of P^ω_{aba} consider the (partial) automaton in Fig. 1 accepting $\mathbf{pref}(P^\omega_{aba})$. It has four states z_1, z_2, z_3, z_4 and which correspond to the left derivatives $\mathbf{pref}(P^\omega_{aba})$, $\mathbf{pref}(P^\omega_{aba})/a$, $\mathbf{pref}(P^\omega_{aba})/ab$ and $\mathbf{pref}(P^\omega_{aba})/aba$, respectively.

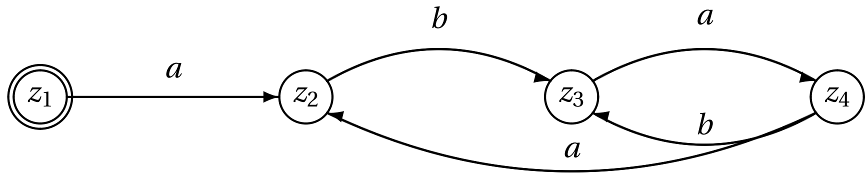

Figure 1: Deterministic automaton accepting P^ω_{aba}

In view of Eq. (3) the identity $\mathbf{pref}(P^\omega_{aba})/w = \mathbf{pref}(P^\omega_{aba})/v$ holds if and only if $P^\omega_{aba}/w = P^\omega_{aba}/v$, and thus from Lemma 3.3.3 we infer $\mathbb{L}_\alpha(P^\omega_{aba}) \ge t_P^{-3}$. On the other hand, Eq. (4), Proposition 3.1.2 and Lemma 3.3.1 imply $\mathbb{L}_\alpha(P^\omega_{aabaa}) \le t_P^{-5}$.

This estimate, however, does 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 Proposition 3.1.2, these prefixes contribute the factors t_P^{-3} and t_P^{-5}, respectively, to the Hausdorff measure.

In order to eliminate the influence of the prefixes we consider instead the sets $\mathbf{suff}(P_q^\omega)$ of all tails (suffixes) of ω-words in P_q^ω. In view of Eq. (6) we have $\dim \mathbf{suff}(F) = \dim \bigcup_{w\in X^*} F/w = \dim F$, that is $\dim \mathbf{suff}(P_q^\omega) = \dim P_q^\omega$. The Hausdorff measures $\mathbb{L}_\alpha(\mathbf{suff}(P_{aba}^\omega))$ and $\mathbb{L}_\alpha(\mathbf{suff}(P_{aabaa}^\omega))$, $\alpha = \log_{|X|} t_P$, are obtained using the procedure of [MS94, Section 3]:

To this end we consider for $F = \mathbf{suff}(P_q^\omega)$ the adjacency matrix $\mathscr{A}_q$: Let $\{F/w : w \in \mathbf{pref}(F)\} = \{F_1 = F, F_2, \ldots, F_k\}$ (without repetitions) and $\mathscr{A}_q = (a_{i,j})_{i,j=1}^k$ where $a_{i,j} := |\{x : x \in X \wedge F_i/x = F_j\}|$.

In view of Proposition 2.2 this adjacency matrix is the adjacency matrix of the minimal partial automaton $\mathscr{B}_q$ accepting the language $\mathbf{infix}(P_q^\omega)$. The automata $\mathscr{B}_{aba}$ and $\mathscr{B}_{aabaa}$ are depicted in Table 1.

$\mathscr{B}_{aba}$	W	W/aa	W/b	W/a
a	W/a		W/a	W/aa
b	W/b	W/b		W/b

$\mathscr{B}_{aabaa}$	V	V/a	V/a^4	V/b	V/ba	V/aa	V/a^3
a	V/a	V/aa		V/ba	V/aa	V/a^3	V/a^4
b	V/b	V/b	V/b			V/b	V/b

Table 1: Automata $\mathscr{B}_{aba}$ and $\mathscr{B}_{aabaa}$ accepting $W = \mathbf{infix}(P_{aba}^*)$ and $V = \mathbf{infix}(P_{aabaa}^*)$, respectively

We obtain the adjacency matrices

$$\mathscr{A}_{aba} = \begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & \mathbf{0} & \mathbf{1} & \mathbf{0} \\ 0 & \mathbf{0} & \mathbf{0} & \mathbf{1} \\ 0 & \mathbf{1} & \mathbf{1} & \mathbf{0} \end{pmatrix} \quad \text{and} \quad \mathscr{A}_{aabaa} = \begin{pmatrix} 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & \mathbf{0} & \mathbf{1} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ 0 & 0 & \mathbf{0} & \mathbf{0} & \mathbf{1} & \mathbf{0} & \mathbf{0} \\ 0 & 0 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{1} & \mathbf{0} \\ 0 & 0 & \mathbf{0} & \mathbf{1} & \mathbf{0} & \mathbf{0} & \mathbf{1} \\ 0 & 0 & \mathbf{1} & \mathbf{1} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix} \tag{7}$$

Here the boldface entries correspond to the terminal strongly connected components (irreducible square submatrices) (cf. [MS94, Section 3]). These submatrices have the Frobenius eigenvalue (eigenvalue of maximum modulus) $t_P > 1$.

Then, for $q \in \{aba, aabaa\}$, the value $\mathbb{L}_\alpha(\mathbf{suff}(P_q^\omega))$ is the topmost entry of a non-negative eigenvector $\vec{a}_q$ of $\mathscr{A}_q$ corresponding to the eigenvalue $|X|^\alpha = t_P$ having the maximum entry 1 at positions corresponding to the terminal strongly connected component. This yields

$$\vec{a}_{aba}^\top = (t_p, t_p^{-2}, t_p^1, 1) \quad \text{and} \quad \vec{a}_{aabaa}^\top = (t_p, 1 + t_p^{-2}, t_p^{-3}, t_p^{-2}, t_p^1, 1, t_p^{-1}) \tag{8}$$

and the value $\mathbb{L}_\alpha(\mathbf{suff}(P_q^\omega)) = t_P$, for both $q = aba$ and $q = aabaa$.

As a consequence, if we neglect the influence of the prefixes, with respect to Hausdorff measure both maximal quasiperiods have the same behaviour. It is interesting to state that in view of $\dim \mathbf{suff}(P_q^\omega) = \log_{|X|} t_p$ and $\mathbb{L}_\alpha(\mathbf{suff}(P_q^\omega)) = t_P$, $q \in \{aba, aabaa\}$, these values are closely related to the smallest Pisot number.

4 Subword Complexity

4.1 The subword complexity of quasiperiodic ω-words

In this section we investigate upper bounds 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 $\mathbf{infix}(\xi) \subseteq \mathbf{infix}(P_q^*)$. Thus

$$f(\xi, n) \le |\mathbf{infix}(P_q^*) \cap X^n| \text{ for } \xi \in P_q^\omega . \tag{9}$$

Similarly to the proof of Proposition 5.5 of [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 $\mathbf{infix}(\xi) = \mathbf{infix}(P_q^*)$. Consequently, the tight upper bound on the subword complexity of quasiperiodic ω-words having a certain quasiperiod q is $f_q(n) := |\mathbf{infix}(P_q^*) \cap X^n|$.

The following facts are known from the theory of formal power series (cf. [BP85, SS78]). As $\mathbf{infix}(P_q^*)$ 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

$$f_q(n+k) = \textstyle\sum_{i=0}^{k-1} m_i \cdot f_q(n+i) \tag{10}$$

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 $k' \le k$, λ_i are pairwise distinct roots of the polynomial $\chi_q(t) = t^n - \sum_{i=0}^{k-1} a_i \cdot t^i$ 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 of [Sta85] (see also [PS10, Eq. (8)]) 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 4.1 ([PS10, Lemma 16]) *Let $q \in X^* \setminus \{e\}$. Then there are constants $c_{q,1}, c_{q,2} > 0$ and a $\lambda_q \ge 1$ such that*

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

Next we are looking for those quasiperiods q which yield the largest value of λ_q among all quasiperiods.

Lemma 4.2 ([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 $aabaa$.*
This value is $\lambda_{aba} = \lambda_{aabaa} = t_P$ where t_P is the positive root of the polynomial $t^3 - t - 1$.

Remark 4.3 The bound in Lemma 4.2 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 yields ω-words $\xi \in \{a, b\}^\omega$ of larger subword complexity $f(\xi, n)$, that is, forces the larger constant $c_{q,2}$ where $q \in \{aba, aabaa\}$ in the upper bound of Lemma 4.1.

From the deterministic automata $\mathscr{B}_{aba}$ and $\mathscr{B}_{aabaa}$ (see Table 1) accepting the languages $\mathbf{infix}(P^*_{aba})$ and $\mathbf{infix}(P^*_{aabaa})$, respectively, we obtain the adjacency matrices $\mathscr{A}_{aba}$ and $\mathscr{A}_{aabaa}$ of Eq. (7) and their characteristic polynomials

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

So both sequences $(|\mathbf{infix}(P^*_{aba}) \cap X^n|)_{n\in\mathbb{N}}$ and $(|\mathbf{infix}(P^*_{aabaa}) \cap X^n|)_{n\in\mathbb{N}}$ satisfy the recurrence relation $f_q(n+7) = f_q(n+4) + f_q(n+3) + f_q(n+2)$ with the initial values $(1, 2, 3, 4, 5, 7, 9)$ for $q = aba$ (see also [LR04]) and $(1, 2, 3, 4, , 6, 8, 10)$ for $q = aabaa$ which shows already that the growth of $(|\mathbf{infix}(P^*_{aabaa}) \cap X^n|)_{n\in\mathbb{N}}$ is the larger one.

In order to calculate the growth of $(f_q(n))_{n\in\mathbb{N}}$ $q \in \{aba, aabaa\}$ more accurately, we observe the following. The characteristic polynomials χ_{aba} and χ_{aabaa} have as root of maximal modulus the smallest Pisot number $t_P > 1$. The other roots satisfy $|t| < 1$ or, additionally, $t = \pm\sqrt{-1}$ in case of χ_{aabaa}. Thus $t_P > 1$ determines the growth of $(f_q(n))_{n\in\mathbb{N}}$.

Using the standard methods of recurrent relations one obtains for a quasiperiodic ω-word ξ with quasiperiod aba the largest achievable subword complexity $f(\xi, n) = \mathrm{INT}(\frac{2t_P^2+3t_P+2}{2t_P+3} \cdot t_P^n)$, for large n, where $\mathrm{INT}(\alpha)$ is the integer closest to the real α.

Similarly, for a quasiperiodic ω-word η with quasiperiod *aabaa* the largest achievable subword complexity satisfies $f(\eta,n) = \mathrm{INT}(\frac{13t_P^2+16t_P+9}{10t_P+15}\cdot t_P^n)$, for large n. Observe that for the constants it holds $\frac{2t_P^2+3t_P+2}{2t_P+3} < \frac{13t_P^2+16t_P+9}{10t_P+15}$.

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.

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

[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.

[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. Texts and Monographs in Computer Science.

[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.

[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, Heidelberg, 2012.

A Calculating the Constants

In the appendix we give a derivation of how to calculate the constants $\frac{2t_P^2+3t_P+2}{2t_P+3}$ and $\frac{13t_P^2+16t_P+9}{10t_P+15}$ in the expansion of $f(\xi,n)$ and $f(\eta,n)$. To this end we start from the deterministic automata $\mathscr{B}_{aba}$ and $\mathscr{B}_{aabaa}$ accepting $W = \mathbf{infix}(P_{aba}^*)$ and $V = \mathbf{infix}(P_{aabaa}^*)$, respectively (see Table 1).

From the automata $\mathscr{B}_{aba}$ and $\mathscr{B}_{aabaa}$ we calculate the adjacency matrices $\mathscr{A}_{aba}$ and $\mathscr{A}_{aabaa}$ (see Eq. (7)). For these we have $f_{aba}(n) := \mathbf{infix}(W) \cap X^n = \mathfrak{e}_0 \cdot \mathscr{A}_{aba}^n \cdot \bar{\mathfrak{e}}$ and $f_{aabaa}(n) := \mathbf{infix}(V) \cap X^n = \mathfrak{e}_0 \cdot \mathscr{A}_{aabaa}^n \cdot \bar{\mathfrak{e}}$ where $\mathfrak{e}_0$ is the row vector $(1,0,\ldots,0)$ and $\bar{\mathfrak{e}}$ is the all ones column vector, both being chosen of appropriate length.

Then $f_{aba}(n)$ and $f_{aabaa}(n)$ fulfil recurrence relation $f_{aba}(n+4) = \sum_{j=0}^{3} \chi_j \cdot f_{aba}(n+j)$ and $f_{aabaa}(n+7) = \sum_{j=0}^{6} \chi'_j \cdot f_{aabaa}(n+j)$ where $t^4 - \sum_{j=0}^{3} \chi_j \cdot t^j = \chi_{aba}(t)$ and $t^7 - \sum_{j=0}^{6} \chi'_j \cdot t^j = \chi_{aabaa}(t)$ are the characteristic polynomials of the matrices $\mathscr{A}_{aba}$ and $\mathscr{A}_{aabaa}$, respectively (cf. Eq. (11)).

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, $\mathfrak{i}$ and $-\mathfrak{i}$ where $\mathfrak{i} = \sqrt{-1}$ is the imaginary unit. The roots t_P, t_1, t_2 satisfy the relations $t_P + t_1 + t_2 = 0$, $t_P \cdot t_1 \cdot t_2 = 1$, $t_P > 1$ and $|t_1| = |t_2| < 1$.

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]).

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

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

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

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

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

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

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

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

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

To solve these for values of γ_1 and γ'_1, respectively, we use Cramer's rule. To this end we consider the following determinant and 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 $t^3 - t - 1$.

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

Applying Cramer's rule to Eqs. (17) and (16) yields

$$\gamma_1 = \frac{2 \cdot t_P^2 + 3 \cdot t_P + 2}{2 \cdot t_P^2 + 3} = \frac{10 \cdot t_P^2 + 15 \cdot t_P + 10}{5 \cdot (2 \cdot t_P^2 + 3)} \approx 1{,}6787356,\text{ and} \tag{18}$$

$$\gamma'_1 = \frac{13 \cdot t_P^2 + 16 \cdot t_P + 9}{5 \cdot (2 \cdot t_P^2 + 3)} \approx 1{,}876608 \tag{19}$$

Since $|t_1| = |t_2| < 1$ we have $|\gamma_2 \cdot t_1^n + \gamma_3 \cdot t_2^n| < 1$ and $|\gamma'_2 \cdot t_1^n + \gamma'_3 \cdot t_2^n \pm |\frac{2}{5}|| < 1$ for sufficiently large $n \in \mathbb{N}$. Thus, 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.