CDMTCS Research Report Series

Quasiperiods of Infinite Words

Ludwig Staiger
Martin-Luther-Universität
Halle-Wittenberg

CDMTCS-509
July 2017

Centre for Discrete Mathematics and
Theoretical Computer Science

Quasiperiods of Infinite Words

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 is a word whose occurrences cover every part of the string. An infinite string is referred to as quasiperiodic if it has a quasiperiod.

We present a characterisation of the set of infinite strings having a certain word q as quasiperiod via a finite language P_q consisting of prefixes of the quasiperiod q. It turns out its star root $\sqrt[*]{P_q}$ is a suffix code having a bounded delay of decipherability.

This allows us to calculate the maximal subword (or factor) complexity of quasiperiodic infinite strings having quasiperiod q and further to derive that maximally complex quasiperiodic infinite strings have quasiperiods aba or $aabaa$.

Contents

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

1 Introduction

Around 2000 Solomon Marcus presented some tutorials dealing with language-theoretic properties of infinite words [MP94, Mar02, Mar04]. One topic of interest was their *subword complexity* (or *factor complexity* [CN10]). Besides the asymptotic behaviour of the factor complexity, also known as their topological entropy [CN10, Section 4.2.2] Marcus was also interested in the behaviour of the complexity function $f(\xi,n)$ assigning to a natural number $n \in \mathbb{N}$ the number of subwords of the infinite word (ω-word) ξ.

In his tutorial [Mar04] Solomon Marcus provided some initial facts on quasiperiodic infinite words. Here he was also concerned with recurrences in ω-words and their influence to subword complexity. A well-known fact established by Grillenberger is that the asymptotic subword complexity (or topological entropy) of an almost periodic (or uniformly recurrent) ω-word can be arbitrarily close (but not equal) to the maximal subword complexity (see [CN10, Theorem 4.4.4]).

In [Mar04] Marcus posed several questions on the complexity of quasiperiodic infinite words. The papers [LR04, LR07] studied in more detail quasiperiodic infinite words generated by morphisms and their relation to Sturmian words. Their results concern mainly infinite words of low complexity. This fits into the line pursued in the tutorial [BK03] or the book [AS03] where also mainly infinite words of low (polynomial) complexity were considered. Some results on high (exponential) subword complexity were derived in [Sta12] or concerning the relation between subword and Kolmogorov complexities in [Sta93, Section 5].

The investigations of the present paper are related to Question 2 in Marcus' tutorial [Mar04] and to the question posed in [LR04] of finding the maximally possible complexity functions for those words. As complexity here and in the cited above papers one considers the (subword) complexity function $f(\xi,n)$.

As a final result we deduce that the maximally possible complexity functions for quasiperiodic infinite words ξ are bounded from above by a function of the form $f(\xi,n) \le c \cdot t_P^n, n \ge n_\xi$, where n_ξ is a number depending on ξ and t_P is the smallest Pisot-Vijayaraghavan number, that is, the unique real root t_P of the cubic polynomial $x^3 - x - 1$, which is approximately equal to $t_P \approx 1.324718$. We show also that this bound is tight, that is, there are ω-words ξ having $f(\xi,n) \approx c \cdot t_P^n$. Moreover, we estimate the quasiperiods for which this bound can be achieved.

The paper is organised as follows. After introducing some notation we derive in Section 3 a characterisation of quasiperiodic words and ω-words having a certain quasiperiod q. Moreover, we use the finite basis sets P_q and its dual R_q ($\mathcal{L}(q)$ and $\mathcal{R}(q)$ in [Mou00]) from which the sets of quasiperiodic words or ω-words having quasiperiod q can be constructed. In Section 4 it is then proved that the star root of P_q is a suffix code having a bounded delay of decipherability and, dually, the star root of R_q is a prefix code.

This much prerequisites allow us, in Section 5, to estimate the number of subwords of the language Q_q of all quasiperiodic words having quasiperiod q. It turns out that $c_{q,1} \cdot \lambda_q^n \leq f(Q_q,n) \leq c_{q,2} \cdot \lambda_q^n$ where $f(Q_q,n)$ is the number of subwords of length n of words in Q_q and $1 \leq \lambda_q \leq t_P$ depends on q. We construct, for every quasiperiod q, a quasiperiodic ω-word ξ_q with quasiperiod q whose subword complexity $f(\xi_q,n)$ is maximal. Finally, we estimate the quasiperiods q for which the subword complexity of Q_q is maximal.

Some of the results of this paper were presented at the conference "Workshop on Descriptional Complexity of Formal Systems 2010" [PS10].

2 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. In the sequel we assume the reader to be familiar with basic facts of language theory.

A word $w \in X^* \setminus \{e\}$ is called *primitive* if $w = v^n$ implies $n = 1$, that is, w is not the power of a shorter word. The following facts are known (e.g. [BP85, Shy01])

Claim 1 *Every word $w \in X^* \setminus \{e\}$ has a unique representation $w = v^n$ where v is primitive.*

Claim 2 *If $w \cdot v = v \cdot w,\ w,v \in X^*$ the w,v are powers of a common (primitive) word.*

As usual a language $L \subseteq X^*$ is called a *code* provided $w_1 \cdots w_l = v_1 \cdots v_k$ for $w_1,\dots,w_l,\ v_1,\dots,v_k \in L$ implies $l = k$ and $w_i = v_i$. A code L is said to be a *prefix code* (*suffix code*) provided no codeword is a prefix (suffix) of another codeword.

3 Quasiperidicity

3.1 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. Informally, η has quasiperiod q if every position of η occurs within some occurrence of q in η [AFI91, Mou00].

Let for $q \in X^* \setminus \{e\}$, Q_q be the set of quasiperiodic words with quasiperiod q. Then $\{q\}^* \subseteq Q_q = Q_q^*$ and $Q_q \setminus \{e\} \subseteq X^* \cdot q \cap q \cdot X^*$. In order to describe the set of quasiperiodic strings having a certain quasiperiod $q \in X^* \setminus \{e\}$ the following definition is helpful.

Definition 1 A family $(w_i)_{i=1}^{\ell}$, $\ell \in \mathbb{N} \cup \{\infty\}$, of words $w_i \in X^* \cdot q$ is referred to as a *q-chain* provided $w_1 = q$, $w_i \sqsubset w_{i+1}$ and $|w_{i+1}| - |w_i| \leq |q|$.

It holds the following.

Lemma 1

(1) *$w \in Q_q \setminus \{e\}$ if and only if there is a q-chain $\left(w_i\right)_{i=1}^{\ell}$ such that $w_\ell = w$.*

(2) *An ω-word $\xi \in X^\omega$ is quasiperiodic with quasiperiod q if and only if there is a q-chain $\left(w_i\right)_{i=1}^{\infty}$ such that $w_i \sqsubset \xi$.*

Proof. It suffices to show how a family $\left(u_j\right)_{j=0}^{|\eta|-1}$ can be converted to a q-chain $\left(w_i\right)_{i=1}^{\ell}$ and vice versa.

Consider $\eta \in X^* \cup X^\omega$ and let $\left(u_j\right)_{j=0}^{|\eta|-1}$ be a family such that $u_j \cdot q \sqsubseteq \eta$ and $j - |q| < |u_j| \leq j$ for $j < |\eta|$.

Define $w_1 := q$ and $w_{i+1} := u_{|w_i|} \cdot q$ as long as $|w_i| < |\eta|$. Then $w_i \sqsubseteq \eta$ and $|w_i| < |w_{i+1}| = |u_{|w_i|} \cdot q| \leq |w_i| + |q|$. Thus $\left(w_i\right)_{i=1}^{\ell}$ is a q-chain with $w_i \sqsubseteq \eta$.

Conversely, let $\left(w_i\right)_{i=1}^{\ell}$ be a q-chain such that $w_i \sqsubseteq \eta$ and set

$$u_j := \max_{\sqsubseteq}\left\{w' : \exists i (w' \cdot q = w_i \wedge |w'| \leq j)\right\}, \text{ for } j < |\eta| .$$

By definition, $u_j \cdot q \sqsubseteq \eta$ and $|u_j| \leq j$. Assume $|u_j| \leq j - |q|$ and $u_j \cdot q = w_i$. Then $|w_i| \leq j < |\eta|$. Consequently, in the q-chain there is a successor w_{i+1}, $|w_{i+1}| \leq |w_i| + |q| \leq j + |q|$. Let $w_{i+1} = w'' \cdot q$. Then $u_j \sqsubset w''$ and $|w''| \leq j$ which contradicts the maximality of u_j. ❑

Lemma 1 yields the following consequences.

Corollary 1 *Let $u \in \mathbf{pref}(Q_q)$. Then there are words $w, w' \in Q_q$ such that $w \sqsubseteq u \sqsubseteq w'$ and $|u| - |w|, |w'| - |u| \leq |q|$.*

Corollary 2 *Let $\xi \in X^\omega$. Then the following are equivalent.*

(1) *ξ is quasiperiodic with quasiperiod q.*

(2) *$\mathbf{pref}(\xi) \cap Q_q$ is infinite.*

(3) *$\mathbf{pref}(\xi) \subseteq \mathbf{pref}(Q_q)$.*

3.2 Finite generators for quasiperiodic words

In this part we introduce finite languages P_q and R_q which generate the set of quasiperiodic words as well as the set of quasiperiodic ω-words having quasiperiod q.

We set

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

Then we have the following properties.

Proposition 1

$$\begin{aligned} Q_q &= P_q^* \cdot q \cup \{e\} \subseteq P_q^*, & (2)\\ \mathbf{pref}(Q_q) &= \mathbf{pref}(P_q^*) = P_q^* \cdot \mathbf{pref}(q) & (3) \end{aligned}$$

Proof. In order to prove $Q_q \subseteq P_q^* \cdot q \cup \{e\}$ we show that $w_i \in P_q^* \cdot q$ for every q-chain $(w_i)_{i=1}^{\ell}$. This is certainly true for $w_1 = q$. Now proceed by induction on i. Let $w_i = w_i' \cdot q \in P_q^* \cdot q$ and $w_{i+1} = w_{i+1}' \cdot q$. Then $w_i' \cdot v_i = w_{i+1}'$. Now from $w_i \sqsubset w_{i+1}$ we obtain $e \sqsubset v_i \sqsubseteq q \sqsubset v_i \cdot q$, that is, $v_i \in P_q$.

Conversely, let $v_i \in P_q$ and consider $v_1 \cdots v_\ell \cdot q$. Since $q \sqsubseteq v_i \cdot q$ the family $(v_1 \cdots v_j \cdot q)_{j=0}^{\ell}$ is a q-chain. This shows $P_q^* \cdot q \cup \{e\} \subseteq Q_q$.

Eq. (3) is an immediate consequence of Eq. (2). ❏

Proposition 1 implies the following characterisation of ω-words having quasiperiod q.

$$\{\xi : \xi \in X^\omega \wedge \xi \text{ has quasiperiod } q\} = P_q^\omega \tag{4}$$

Proof. Since P_q is finite, $P_q^\omega = \{\xi : \xi \in X^\omega \wedge \mathbf{pref}(\xi) \subseteq \mathbf{pref}(P_q^*)\}$. ❏

A dual generator of Q_q is obtained by the right-to-left duality of reading words using the suffix relation $\le_s$ instead of the prefix relation $\sqsubseteq$.

$$R_q := \{v : e <_s v \le_s q <_s v \cdot q\}. \tag{5}$$

Analogously to Proposition 1 we obtain

Proposition 2

$$\begin{aligned} Q_q &= q \cdot R_q^* \cup \{e\} \subseteq R_q^*, & (6)\\ \mathbf{pref}(Q_q) &= \mathbf{pref}(q) \cup q \cdot \mathbf{pref}(R_q^*) & (7) \end{aligned}$$

The proof is similar to the proof of Proposition 1 using the reversed version of q-chain. A slight difference appears with an analogy to Eq. (4).

$$\{\xi : \xi \in X^\omega \wedge \xi \text{ has quasiperiod } q\} = q \cdot R_q^\omega \tag{8}$$

An alternative derivation of the languages P_q and R_q can be found in Definition 2 of [Mou00]. Here the borders, that is, prefixes which are simultaneously suffixes of the quasiperiod q, are used:

$$\begin{aligned} P_q &= \{v : \exists w(w \sqsubset q \wedge w <_s q \wedge q = v \cdot w)\}, \text{ and}\\ R_q &= \{v : \exists w(w \sqsubset q \wedge w <_s q \wedge q = w \cdot v)\}. \end{aligned}$$

In the subsequent sections we focus on the investigation of P_q due to the left-to-right direction of ω-words.

3.3 Combinatorial properties of P_q

We investigate basic properties of P_q using simple facts from combinatorics on words (see [BP85, Shy01]).

Proposition 3 *$v \in P_q$ if and only if $|v| \leq |q|$ and there is a prefix $\bar{v} \sqsubset v$ such that $q = v^k \cdot \bar{v}$ for $k = \lfloor |q|/|v| \rfloor$.*

Proof. Sufficiency is clear. Let now $v \in P_q$. Then $v \sqsubseteq q \sqsubset v \cdot q$. This implies $v^l \sqsubseteq q \sqsubset v^l \cdot q$ as long as $l \leq k$ and, finally, $q \sqsubset v^{k+1}$. ❑

Corollary 3 *$v \in P_q$ if and only if $|v| \leq |q|$ and there is a $k' \in \mathbb{N}$ such that $q \sqsubseteq v^{k'}$.*

Now set $q_0 := \min_{\sqsubseteq} P_q$. Then in view of Proposition 3 and Corollary 3 we have the following.

$$q = q_0^k \cdot \bar{q} \text{ for } k = \lfloor |q|/|q_0| \rfloor \text{ and some } \bar{q} \sqsubset q_0 . \tag{9}$$

Corollary 4 *The word q_0 is primitive.*

Proof. Assume $q_0 = q_1^l$ for some $l > 1$. Then $\bar{q} = q_1^j \cdot \bar{q}_1$ where $\bar{q}_1 \sqsubset q_1$, and, consequently, $q \sqsubset q_1^{k \cdot l + j + 1}$ contradicting the fact that q_0 is the shortest word in P_q. ❑

Proposition 4 *Let $q \in X^*, q \neq e$, $q_0 = \min_{\sqsubseteq} P_q$, $q = q_0^k \cdot \bar{q}$ and $v \in P_q^* \setminus \{e\}$.*

(1) *If $w \sqsubseteq q$ then $v \cdot w \sqsubseteq q$ or $q \sqsubseteq v \cdot w$.*

(2) *If $w \cdot v \sqsubseteq q$ then $w \in \{q_0\}^*$.*

(3) *If $|v| \leq |q| - |q_0|$ then $v = q_0^m$ for some $m \in \mathbb{N}$.*

Proof. The first assertion follows from $q \sqsubset v \cdot q$ and $v \cdot w \sqsubseteq v \cdot q$ by induction.

Since $q_0 \sqsubseteq v$, it suffices to prove the second assertion for q_0. First one observes that, $w \sqsubseteq q$ and $|w| \leq |q| - |q_0|$. Thus $w \sqsubseteq q_0^{k-1} \cdot \bar{q}$. Therefore, we have $w \cdot q_0 \sqsubseteq q$ and $q_0 \cdot w \sqsubseteq q$ which implies $w \cdot q_0 = q_0 \cdot w$ and, according to Claim 2, w and q_0 are powers of a common word. The assertion follows because q_0 is primitive.

The third assertion follows from the second one as $v\cdot q_0 \sqsubseteq q$ for $v \in P_q^*$ with $|v| \le |q| - |q_0|$. ❑

Next we investigate the relation between a quasiperiod $q = q_0^k \cdot \bar{q}$ where $q_0 = \min_{\sqsubseteq} P_q$ and $\bar{q} \sqsubset q_0$ and its *shortening* $\hat{q} := q_0 \cdot \bar{q}$. Since $q \in Q_{\hat{q}}$, we have $Q_{\hat{q}} \supseteq Q_q$.

We continue with a relation between P_q and $P_{\hat{q}}$. It is obvious that $q_0^i \in P_q$ for every $i = 1,\ldots,k$. Then Proposition 4.(3) shows that[1]

$$\begin{aligned} \{q_0^i : i = 1,\ldots,k\} &\subseteq P_q \\ &\subseteq \{q_0^i : i = 1,\ldots,k-1\} \cup \{v' : v' \sqsubseteq q \wedge |v'| > |q| - |q_0|\}. \end{aligned} \tag{10}$$

Lemma 2 *Let $q \in X^*, q \neq e$, $q_0 = \min_{\sqsubseteq} P_q$, and $q = q_0^k \cdot \bar{q}$ and $\hat{q} = q_0 \cdot \bar{q}$ the shortening of q. Then*

$$P_q = \{q_0^i : i = 1,\ldots,k-1\} \cup \{q_0^{k-1} \cdot v : v \in P_{\hat{q}}\}.$$

Proof. Let $v \in P_{\hat{q}}$, that is, $v \sqsubseteq q_0\bar{q} \sqsubset v \cdot q_0\bar{q}$. Then $q_0^{k-1} \cdot v \sqsubseteq q_0^k \cdot \bar{q} \sqsubset q_0^{k-1} \cdot v \cdot q_0\bar{q} \sqsubset q_0^{k-1} \cdot v \cdot q_0^k \cdot \bar{q}$, that is, $q_0^{k-1} \cdot v \in P_q$.

Conversely, let $v' \in P_q$ and $v' \notin \{q_0^i : i = 1,\ldots,k-1\}$. Then, according to Proposition 4.(3) there is a unique $v \neq e$ such that $v' = q_0^{k-1} \cdot v$. Now $v' = q_0^{k-1} \cdot v \sqsubseteq q = q_0^k \cdot \bar{q} \sqsubset v' \cdot q = q_0^{k-1} \cdot v \cdot q_0^k \cdot \bar{q}$ implies $v \sqsubseteq q_0 \cdot \bar{q} \sqsubset v \cdot q_0^k \cdot \bar{q}$. Since $|v| \le |q_0 \cdot \bar{q}|$ and $q_0 \cdot \bar{q} \sqsubseteq q_0^k \cdot \bar{q}$, we have $v \sqsubseteq q_0 \cdot \bar{q} \sqsubset v \cdot q_0 \cdot \bar{q}$. ❑

As a particular result we obtain from Lemma 2 and Eq. (10) that $P_{q_0\bar{q}} \subseteq \{v : \bar{q} \sqsubset v \sqsubseteq q_0\bar{q}\}$. This result can be generalised as follows.

Lemma 3 *If q is primitive, $\bar{q} \sqsubset q$ and $v \in P_{q\bar{q}}$ then $\bar{q} \sqsubset v$.*

Proof. Assume $v \sqsubseteq \bar{q}$ and $v \in P_{q\bar{q}}$. Then $v \sqsubseteq q\bar{q} \sqsubset vq\bar{q}$ implies $vq \sqsubseteq q\bar{q}$. On the other hand $qv \sqsubseteq q\bar{q}$. Thus $qv = vq$, and $|v| < |q|$ contradicts the fact that q is primitive. ❑

3.4 Primitivity and Superprimitivity

In this section we consider the inclusion relations between the languages $P_q, q \neq e$. These languages are generators for the set of quasiperiodic words Q_q in the sense of Eq. (2). As we can see from Lemma 2

[1]Observe that $q_0^k \sqsubseteq q$ and $|q_0^k| > |q| - |q_0|$.

and Eq. (2) the language P_q is not always the smallest one which generates Q_q. In order to obtain the smallest one we consider the star root of languages. Define now the *star-root* of a language $L \subseteq X^*$ as usual as the smallest language L' satisfying $(L')^* = L^*$:

$$\sqrt[*]{L} := \big(L \setminus \{e\}\big) \setminus \big(L \setminus \{e\}\big)^2 \cdot L^*$$

From Lemma 2 we obtain immediately.

Lemma 4 *Let $q \in X^*, q \neq e$ and $q_0 = \min_{\sqsubseteq} P_q$. Then $P_q = \sqrt[*]{P_q}$ if and only if $|q_0| > |q|/2$.*

Proof. It is obvious that $q_0 \in \sqrt[*]{P_q}$ and $q_0^m \neq \sqrt[*]{P_q}$ if $m \geq 2$. It suffices to show that $v \in P_q \setminus \{q_0\}^*$ belongs to $\sqrt[*]{P_q}$. To this end observe that in view of Proposition 4 (3), for $v' \in P_q$, the product $v \cdot v'$ is longer than q. Thus $v \in \sqrt[*]{P_q}$. ❏

Cast into the language of borders, it holds $\sqrt[*]{P_q} = P_q$ if and only if the longest proper border of q has length $< |q|/2$.

Corollary 5

$$\sqrt[*]{P_q} = \big(P_q \setminus \{q_0\}^*\big) \cup \{q_0\}$$

Analogously to the primitivity of words in [AFI91, Mou00] a word was referred to as *superprimitive* if it is not covered by a shorter one. This leads to the following definition.

Definition 2 (superprimitive) A non-empty word $q \in X^* \setminus \{e\}$ is *superprimitive* if and only in Q_q is maximal w.r.t. "$\subseteq$" in the family $\{Q_q : q \in X^* \setminus \{e\}\}$.

The next proposition relates Lemma 4 to superprimitivity.

Proposition 5 *If $q \in X^* \setminus \{e\}$ is superprimitive then $|\min_{\sqsubseteq} P_q| > |q|/2$, and if $|\min_{\sqsubseteq} P_q| > |q|/2$ then q is primitive.*

Proof. If $q_0 = \min_{\sqsubseteq} P_q$ and $|q_0| \leq |q|/2$ then $q = q_0^k \cdot \bar{q}$ for some $\bar{q} \sqsubset q_0$. Thus $q \in Q_{q_0\bar{q}}$ and $q_0\bar{q} \notin Q_q$.

As $q = q'^m$ with $m > 1$ implies $|q_0| \leq |q'| \leq |q|/2$, the other assertion follows. ❏

The converse of Proposition 5 is not valid.

Example 1 Let $q = abaabaabababaab$. Then $P_q = \{abaabaabab, abaabaababa, q\}$, and $|\min_{\sqsubseteq} P_q| = 8 > 13/2$ but as $abaabaababaab \in Q_{abaab}$ the word q is not superprimitive.

The word $q = ababa$ is primitive but $q_0 = ab$ has $|q_0| \leq |q|/2$. ❏

Lemma 5 *Let q' be the longest word in $P_q \setminus \{q\}$. Then $P_{q'} \supseteq P_q \setminus \{q\}$.*

Moreover, if $q = q_0^k$ for $q_0 = \min_{\sqsubseteq} P_q$ and some $k \geq 2$ then $P_{q'}^* \supseteq P_q^*$.

Proof. Let $v \in P_q \setminus \{q\}$. Then $e \sqsubset v \sqsubset q \sqsubset v \cdot q$. Since $0 < |v| \leq |q'|$ and $q' \sqsubset q$, we obtain the required relation $v \sqsubset q' \sqsubset v \cdot q'$.

If $q = q_0^k$ then $P_{q'} \supseteq P_q \setminus \{q_0^k\}$ and $q_0 \in P_{q'}$. ❏

In Lemma 5 equality as well as proper inclusion are possible.

Example 2 Let $q = abaaba$. Then $P_q = \{aba, abaab, q\}$ and $P_{abaab} = \{aba, abaab\} = P_q \setminus \{q\}$. ❏

Example 3 Let $q = abaaabaa$. Then $P_{q_2} = \{abaa, abaaaba, q\}$ and $P_{abaaaba} = \{abaa, abaaab, abaaaba\} \supset P_q \setminus \{q\}$. ❏

In contrast to the fact that the word $q_0 = \min_{\sqsubseteq} P_q$ is always primitive, it need not satisfy $|\min_{\sqsubseteq} P_{q_0}| > |q_0|/2$ let alone be superprimitive..

Example 4 $q = aabaaabaaaa$ has $P_q = \{aabaaabaa, q\}$, that is $q_0 = aabaaabaa$ which, in turn has $P_{q_0} = \{aaba, aabaaaba, q_0\}$ with $|aaba| = 4 < |q_0|/2$. ❏

4 P_q and R_q as Codes

In this section we investigate in more detail the properties of the star root of P_q. It turns out that $\sqrt[*]{P_q}$ is a suffix code which, additionally, has a bounded delay of decipherability. This delay is closely related to the largest power of q_0 being a prefix of q.

According to [BP85, Sta86, BWZ90, FRS07] a subset $C \subseteq X^*$ is a code of a *delay of decipherability* $m \in \mathbb{N}$ if and only if for all $v, v', w_1, \ldots, w_m \in C$ and $u \in C^*$ the relation $v \cdot w_1 \cdots w_m \sqsubseteq v' \cdot u$ implies $v = v'$. Observe that $C \subseteq X^* \setminus \{e\}$ is a prefix code if and only if C has delay 0.

First we show that $\sqrt[*]{P_q}$ is a suffix code. This generalises Proposition 7 of [Mou00].

Proposition 6 *$\sqrt[*]{P_q}$ is a suffix code.*

Proof. Assume $u = w \cdot v$ for some $u, v \in \sqrt[*]{P_q}$, $u \neq v$. Then $u \sqsubseteq q$ and Proposition 4 (2) proves $w \in \{q_0\}^* \setminus \{e\}$. Consequently, $|v| \leq |q| - |q_0|$. Now Proposition 4 (3) implies $v \in \{q_0\}^*$ and hence $u \in \{q_0\}^*$. Since $u, v \in P_q$, we obtain $u = v = q_0$ contradicting $u \neq v$. ❑

Using the duality of P_q and R_q one shows in an analogous manner that R_q is a prefix code.

We conclude this part by investigating the delay of decipherability of $\sqrt[*]{P_q}$. We prove that the delay depends on the relation between the quasiperiod q and the minimal w.r.t. $\sqsubseteq$ word $q_0 \in P_q$.

Theorem 1 *Let $q \in X^* \setminus \{e\}$, $q_0 = \min_{\sqsubseteq} P_q$, $q_0^m \sqsubset q \sqsubseteq q_0^{m+1}$ and $|\sqrt[*]{P_q}| > 1$. Then $\sqrt[*]{P_q}$ is a code having a delay of decipherability of m or $m+1$.*

Proof. We have $q_0, q \in \sqrt[*]{P_q}$ if $q \sqsubset q_0^{m+1}$ or, as $|\sqrt[*]{P_q}| > 1$, in view of Proposition 4 (3) we have $q_0, q' \in \sqrt[*]{P_q}$ where $q_0^m \sqsubset q' \sqsubset q_0^{m+1}$. In both cases, $q_0 \cdot q_0^{m-1} \sqsubset q'$ for $q_0 \in \sqrt[*]{P_q}$ and some $q' \in \sqrt[*]{P_q}$ implies that the delay of decipherability is at least m.

Next we show that it cannot exceed $m+1$. Assume $v \cdot w_1 \cdots w_{m+1} \sqsubseteq v' \cdot u$ for $v, v', w_1, \ldots, w_{m+1} \in \sqrt[*]{P_q}$ and $u \in P_q^*$. From Proposition 4 (1) we obtain $u \sqsubseteq q$ or $q \sqsubseteq u$ and, since $|w_i| \geq |q_0|$, also $q \sqsubseteq w_1 \cdots w_{m+1}$. Moreover, $v_1, v_2 \in P_q$ implies $|v_1| + |q| \geq |v_2| + |q_0|$.

If $v \sqsubset v'$, in view of the inequality $|v| + |q| \geq |v'| + |q_0|$ our assumption yields $v' \cdot q_0 \sqsubseteq v \cdot q$. Therefore, $w \cdot q_0 \sqsubseteq q$ for the word $w \neq e$ with $v \cdot w = v'$ and, according to Proposition 4 (2) $w \in \{q_0\}^*$. This contradicts the fact that $\sqrt[*]{P_q}$ is a suffix code.

If $v' \sqsubset v$, then $|u| > |w_1 \cdots w_{m+1}| \geq |q|$, and via $|v'| + |q| \geq |v| + |q_0|$ we obtain $v \cdot q_0 \sqsubseteq v' \cdot q$ from our assumption. This yields the same contradiction as in the case when $v \sqsubset v'$. ❑

Thus, if $q_0^m \sqsubset q \sqsubseteq q_0^{m+1}$ and $|\sqrt[*]{P_q}| > 1$ the code $\sqrt[*]{P_q}$ may have a minimum delay of decipherability of m or $m+1$. We provide examples that both cases are possible.

Example 5 Let $q := aabaaaaba$. Then $q_0 = aabaa$, $m = 1$ and $\sqrt[*]{P_q} = P_q = \{q_0, aabaaaab, q\}$ which is a code having a delay of decipherability 2.

Indeed $aabaaaabaa = q_0 \cdot q_0 \sqsubseteq q \cdot q_0$ or

$aabaaaabaa = q_0 \cdot q_0 \sqsubseteq aabaaaab \cdot q_0$. ❑

Moreover, in Example 5, $q \cdot q_0 \notin Q_q$. Thus our example shows also that $q \cdot P_q^*$ need not be contained in Q_q.

Example 6 Let $q := aba$. Then $m = 1$ and $P_q = \{ab, aba\}$ is a code having a delay of decipherability 1. ❑

5 Subword Complexity

In this section we investigate upper bounds on the the subword complexity function $f(\xi, n)$ for quasiperiodic ω-words. If $\xi \in X^\omega$ is quasiperiodic with quasiperiod q then Proposition 3 and Corollary 3 show $\mathbf{infix}(\xi) \subseteq \mathbf{infix}(P_q^*)$. Thus

$$f(\xi, n) \leq |\mathbf{infix}(P_q^*) \cap X^n| \text{ for } \xi \in P_q^\omega. \tag{11}$$

Similar to [Sta93, Proposition 5.5] let $\xi_q := \prod_{v \in P_q^* \setminus \{e\}} v$. This implies $\mathbf{infix}(\xi_q) = \mathbf{infix}(P_q^*)$. Consequently, the tight upper bound on the subword complexity of quasiperiodic ω-words having a certain quasiperiod q is $f_q(n) := |\mathbf{infix}(P_q^*) \cap X^n|$. Observe that in view of Propositions 1 and 2 the identity

$$\mathbf{infix}(P_q^*) = \mathbf{infix}(R_q^*) = \mathbf{infix}(Q_q) \tag{12}$$

holds.

The asymptotic upper bound on the subword complexity $f_q(n)$ is obtained from

$$\lambda_q = \limsup_{n \to \infty} \sqrt[n]{|\mathbf{infix}(P_q^*) \cap X^n|}, \tag{13}$$

that is, for large n, $f_q(n) \leq \hat{\lambda}^n$ whenever $\hat{\lambda} > \lambda_q$.

The following facts are known from the theory of formal power series (cf. [BR88, 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

$$f_q(n+k) = \sum_{i=0}^{k-1} a_i \cdot f_q(n+i)$$

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

In the subsequent parts we estimate values characterising the exponential growth of the family $\left(|\mathbf{infix}(P_q^*) \cap X^n|\right)_{n \in \mathbb{N}}$. This growth mainly depends on the root of largest modulus among the t_i and the corresponding polynomial g_i.

First we show that, independently of the quasiperiod q the polynomial g_i is constant. Then we show that, for every quasiperiod q, a root of largest modulus is always positive and we estimate those quasiperiods for which this root is maximal.

In the remainder of this section we use, without explicit reference, known results from the theory of formal power series, in particular about generating functions of languages and codes which can be found in the literature, e.g. in [BP85, BR88] or [SS78].

5.1 The subword complexity of a regular star language

The language P_q^* is a regular star-language of special shape. Here we show that, generally, the number of subwords of regular star-languages grows only exponentially without a polynomial factor. We start with some easily derived relations between the number of words in a regular language and the number of its subwords.

Lemma 6 *If $L \subseteq X^*$ is a regular language then there is a $k \in \mathbb{N}$ such that*

$$|L \cap X^n| \quad \leq \quad |\mathbf{infix}(L) \cap X^n| \quad \leq \quad \sum_{i=0}^{k} |L \cap X^{n+i}| \tag{14}$$

If the finite automaton accepting L has k states then for every $w \in \mathbf{infix}(L)$ there are words u, v of length $\leq k$ such that $u \cdot w \cdot v \in L$. Thus as a suitable k one may choose twice the number of states of an automaton accepting the language $L \subseteq X^*$.

A first consequence of Lemma 6 is that the identity

$$\limsup_{n \to \infty} \sqrt[n]{|L \cap X^n|} = \limsup_{n \to \infty} \sqrt[n]{|\mathbf{infix}(L) \cap X^n|} \tag{15}$$

holds for regular languages $L \subseteq X^*$.

In order to derive the announced exponential growth we use Corollary 4 of [Sta85] which shows that for every regular language $L \subseteq X^*$ there are constants $c_1, c_2 > 0$ and a $\lambda \geq 1$ such that

$$c_1 \cdot \lambda^n \leq |\mathbf{pref}(L^*) \cap X^n| \leq c_2 \cdot \lambda^n . \tag{16}$$

A consequence of Lemma 6 is that Eq. (16) holds also (with a different constant c_2) for $\mathbf{infix}(L^*)$.

5.2 The subword complexity of Q_q

In this part we estimate the value λ_q of Eq. (13). In view of Eqs. (12) and (16) the value λ_q satisfies the inequality $c_1 \cdot \lambda_q^n \leq |\mathbf{infix}(P_q^*) \cap X^n| \leq c_2 \cdot \lambda_q^n$.

As P_q^* is a regular language Eqs. (13) and (15) show that

$$\lambda_q = \limsup_{n\to\infty} \sqrt[n]{|P_q^* \cap X^n|}$$

which is the inverse of the convergence radius $\mathsf{rad}\,\mathfrak{s}_q^*$ of the power series $\mathfrak{s}_q^*(t) := \sum_{n\in\mathbb{N}} |P_q^* \cap X^n| \cdot t^n$. The series $\mathfrak{s}_q^*$ is also known as the structure generating function of the language P_q^*.

Since $\sqrt[*]{P_q}$ is a code, we have $\mathfrak{s}_q^*(t) = \frac{1}{1-\mathfrak{s}_q(t)}$ where $\mathfrak{s}_q(t) := \sum_{v\in \sqrt[*]{P_q}} t^{|v|}$ is the structure generating function of the finite language $\sqrt[*]{P_q}$. As $\mathfrak{s}_q^*$ has non-negative coefficients Pringsheim's theorem shows that $\mathsf{rad}\,\mathfrak{s}_q^* = \lambda_q^{-1}$ is a singular point of $\mathfrak{s}_q^*$. Thus λ_q^{-1} is the smallest root of $1-\mathfrak{s}_q(t)$. Hence λ_q is the largest positive root of the polynomial $\mathfrak{p}_q(t) := t^{|q|} - \sum_{v\in \sqrt[*]{P_q}} t^{|q|-|v|}$.

Remark 1 *If the length of $q_0 = \min_{\sqsubseteq} \sqrt[*]{P_q}$ does not divide $|q|$ then $\mathfrak{p}_q(t)$ is the reversed polynomial of $1-\mathfrak{s}_q(t)$, that is, has as roots exactly the the inverses of the roots of $1-\mathfrak{s}_q(t)$.*

If $|q_0|$ divides $|q|$ then $q \notin \sqrt[*]{P_q}$ (cf. Lemma 4) and $\mathfrak{p}_q(t)$ has additionally the root 0 with multiplicity $|q|-|q'|$ where q' is the longest word in $\sqrt[*]{P_q}$.

Summarising our observations we obtain the following.

Lemma 7 *Let $q \in X^* \setminus \{e\}$. Then there are constants $c_{q,1}, c_{q,2} > 0$ such that the structure function of the language $\mathbf{infix}(P_q^*)$ satisfies*

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

where λ_q is the largest (positive) root of the polynomial $\mathfrak{p}_q(t)$.

Remark 2 *One could prove Lemma 7 by showing that, for each polynomial $\mathfrak{p}_q(t)$, its largest (positive) root has multiplicity 1. Referring to Corollary 4 of [Sta85] (see Eq. (16)) we avoided these more detailed considerations of a particular class of polynomials.*

5.3 Quasiperiods of maximum subword complexity

In this concluding part we are looking for those quasiperiods q which yield the largest value of λ_q among all quasiperiods thus answering

Question 2 of [Mar04]. All polynomials $\mathfrak{p}_q(t)$ are of the form $p(t) = t^n - \sum_{i\in M} t^i$ where $\emptyset \neq M \subseteq \{0,\ldots,n-1\}$.

We start with a general property of those polynomials.

Proposition 7 *Suppose* $p(t) = t^n - \sum_{i\in M} t^i$ *where* $\emptyset \neq M \subseteq \{0,\ldots,n-1\}$. *Then*

(1) $p(0) \leq 0$, $p(1) \leq 0$, $p(2) > 0$ *and* $p(t') < 0$ *for* $0 < t' < 1$.

(2) *If* $p(t') \geq 0$ *for some* $t' > 0$ *then* $p(t) > 0$ *for* $t > t'$.

(3) *Let* $t_{\max}$ *be the largest positive root of* $p(t)$. *If* $p(t') = 0$ *then* $|t'| \leq t_{\max}$.

Proof. The first assertion is obvious.

For the proof of the second one, let $t = (1+\varepsilon)\cdot t'$ where $\varepsilon > 0$ and observe that $p((1+\varepsilon)\cdot t') > (1+\varepsilon)^n \cdot p(t')$.

The first assertion shows that $1 \leq t_{\max} < 2$. Then third assertion follows via $p(|t'|) = |t'|^n - \sum_{i\in M} |t'|^i \geq |t'|^n - |\sum_{i\in M} t'^i| = 0$ from the second one. ❏

This yields the following fundamental property.

Corollary 6 *If* $t_{\max}$ *is the largest positive root of a polynomial* $p(t) = t^n - \sum_{i\in M} t^i$ *with* $\emptyset \neq M \subseteq \{0,\ldots,n-1\}$ *then* $t_{\max} \in [1,2)$, *and* $p(t') \leq 0$ *if and only if* $t' \leq t_{\max}$, *for* $1 \leq t' < 2$.

Recall that $\mathbf{infix}(P_q^*) = \mathbf{infix}(Q_q)$. Moreover, $Q_q \subseteq Q_{\hat{q}}$ for some shorter quasiperiod $\hat{q}$ whenever q is not superprimitive. As Proposition 5 shows the latter is always the case if q_0 is not longer that $|q|/2$.

For quasiperiods q where q_0 is not longer that $|q|/2$ we have the following property. Consider the successive shortenings (see Section 3.3) $q^{(i)}$ of the quasiperiod q, that is $q^{(0)} := q$ and $q^{(i+1)} := \widehat{q^{(i)}}$. This sequence trivially ends at least after $|q|$ steps with a shortening $\widetilde{q} = q^{(n)}$ for which $|\min_{\sqsubseteq} P_{\widetilde{q}}| > |\widetilde{q}|/2$. Moreover $Q_{q^{(1)}} \subseteq \ldots \subseteq Q_{\widetilde{q}}$ and its predecessor $q^{(n-1)}$ has $|q_0^{(n-1)}| \leq |q^{(n-1)}|/2$. In this situation we have the following.

Proposition 8 *Let* $q \in X^* \setminus \{e\}$ *be such that* $q = q_0^k \cdot \bar{q}$ *where* $\bar{q} \sqsubset q_0$ *and* $k \geq 2$. *If* $|\min_{\sqsubseteq} P_{\hat{q}}| > |\hat{q}|/2$ *for* $\hat{q} := q_0 \cdot \bar{q}$ *then* $\lambda_{\hat{q}} > \lambda_q$ *or* $P_q^* = \{q_0\}^*$.

Proof. Lemma 4 shows that $P_{\hat{q}} = \sqrt[*]{P_{\hat{q}}}$. Then $\mathfrak{p}_{\hat{q}}(t) = t^{|\hat{q}|} - \sum_{v\in P_{\hat{q}}} t^{|\hat{q}|-|v|} = t^{|\hat{q}|} - t^{(|\hat{q}|-|q_0|)} - \sum_{v\in P_{\hat{q}}\setminus\{q_0\}} t^{(|\hat{q}|-|v|)}$.

Via Lemma 2 and Corollary 5 we obtain the following relation between $\sqrt[*]{P_{\hat{q}}}$ and $\sqrt[*]{P_q}$

$$\sqrt[*]{P_q} = \{q_0\} \cup \{q_0^{k-1} \cdot v : v \in P_{\hat{q}} \setminus \{q_0\}\}.$$

If $P_{\hat{q}} = \{q_0\}$ then $\sqrt[*]{P_q} = \{q_0\}$ and, consequently $P_q^* = \{q_0\}^*$.

Let $P_{\hat{q}} \supset \{q_0\}$. This yields $\mathfrak{p}_q(t) = t^{|q|} - t^{(|q|-|q_0|)} - \sum_{v \in P_{\hat{q}} \setminus \{q_0\}} t^{(|q|-|q_0^{k-1}v|)}$. Since $\lambda_{\hat{q}}$ is a root of $\mathfrak{p}_{\hat{q}}(t)$ we have, in view of $q = q_0^{k-1} \cdot \hat{q}$,

$$\begin{aligned} 0 &= \lambda_{\hat{q}}^{k-1} \cdot \mathfrak{p}_{\hat{q}}(\lambda_{\hat{q}}) = \lambda_{\hat{q}}^{|q|} - \lambda_{\hat{q}}^{(|q|-|q_0|)} - \sum_{v \in P_{\hat{q}} \setminus \{q_0\}} \lambda_{\hat{q}}^{(|q|-|v|)} \\ &< \lambda_{\hat{q}}^{|q|} - \lambda_{\hat{q}}^{(|q|-|q_0^{k-1}|)} - \sum_{v \in P_{\hat{q}} \setminus \{q_0\}} \lambda_{\hat{q}}^{(|q|-|q_0^{k-1}v|)} = \mathfrak{p}_q(\lambda_{\hat{q}}). \end{aligned}$$

The assertion $\lambda_{\hat{q}} > \lambda_q$ follows with Corollary 6. ❑

Thus every quasiperiod q having $|q_0|$ not longer than $|q|/2$ has $\lambda_q = 1$ or $\lambda_q < \lambda_{\hat{q}}$, and we may confine the subsequent considerations to estimate quasiperiods yielding maximal subword complexity to quasiperiods q satisfying $|q_0| > |q|/2$. In this case the corresponding polynomials $\mathfrak{p}_q(t)$ are of the form $t^n - \sum_{i \in M} t^i$ where $\emptyset \neq M \subseteq \{0, \ldots, \lfloor \frac{n-1}{2} \rfloor\}$.

Next we consider the positive roots of these polynomials. Define $p_n(t) := t^n - \sum_{i=0}^{\lfloor \frac{n-1}{2} \rfloor} t^i$.

Corollary 7 *For every $n \geq 1$ the polynomial $p_n(t)$ has the largest positive root among all polynomials $p(t) = t^n - \sum_{i \in M} t^i$ with $\emptyset \neq M \subseteq \{j : j \leq \frac{n-1}{2}\}$.*

Proof. This follows from $t'^n - \sum_{i=0}^{\lfloor \frac{n-1}{2} \rfloor} t'^i \leq p(t')$ when $1 \leq t' < 2$ and Corollary 6. ❑

Corollary 7 allows us to restrict the further considerations to the polynomials $p_n(t)$.

Observe that $p_{2n+1}(t) = t^{2n+1} - \sum_{i=0}^n t^i$ and $p_{2n+2}(t) = t^{2n+2} - \sum_{i=0}^n t^i$.

Remark 3 *It holds*

$$\begin{aligned} \mathfrak{p}_{a^n b a^n}(t) &= p_{2n+1}(t), \text{ and} \\ \mathfrak{p}_{a^n b a^{n+1}}(t) &= p_{2n+2}(t). \end{aligned}$$

In particular, $\mathfrak{p}_{ba}(t) = t^2 - 1$ and $\mathfrak{p}_b(t) = t - 1$. So for all degrees ≥ 1 there are polynomials of the form $\mathfrak{p}_q(t)$.

In view of Remark 3 and Lemma 7 in the sequel the positive root $t_{\max}$ of $p_i(t)$ is denoted by λ_i. The roots λ_i can be ordered as follows.

Proposition 9 *Let λ_i be as above. Then*

(1) $\lambda_{2n-1} > \lambda_{2n+1}$ *for* $n \geq 3$, *and*

(2) $\lambda_{2n+1} > \lambda_{2n}$ *for* $n \geq 1$.

Proof. We have

$$t^{n-2} \cdot p_{2n+1}(t) - (t^n + 1) \cdot p_{2n-1}(t) = \sum_{i=0}^{n-3} t^i \text{ for } n \geq 3\,. \tag{17}$$

Then $\lambda_{2n-1}^{n-2} \cdot p_{2n+1}(\lambda_{2n-1}) = \sum_{i=0}^{n-3} \lambda_{2n-1}^i > 0$ and Corollary 6 yields the first assertion.

The second follows in a similar way from the identity $t \cdot p_{2n}(t) - 1 = p_{2n+1}(t)$. ❑

The polynomials $p_1(t)$ and $p_2(t)$ have $\lambda_1 = \lambda_2 = 1$.

If $n = 2$ the identity Eq. (17) obtains as $p_5(t) = (t^2+1) \cdot p_3(t)$, that is, $\lambda_3 = \lambda_5$. Together with the inequalities of Proposition 9 this yields another proof of Lemma 18 in [PS10].

Lemma 8 *The polynomials $t^3 - t - 1$ and $t^5 - t^2 - t - 1 = (t^2+1)\cdot(t^3 - t - 1)$ have the largest positive roots among all polynomials $\mathfrak{p}_q(t)$, $q \in X^* \setminus \{e\}$.*

The ω-words $\xi_{aba} = \prod_{v \in P^*_{aba} \setminus \{e\}} v$ and $\xi_{aabaa} = \prod_{w \in P^*_{aabaa} \setminus \{e\}} w$ are quasiperiodic ω-words having maximum subword complexity.

We conclude with two remarks.

Remark 4

(1) *The positive root t_P of $\mathfrak{p}_{aba}(t)$ (or of $\mathfrak{p}_{aabaa}(t)$) is known as the smallest Pisot-Vijayaraghavan number, that is, a positive root > 1 of an irreducible polynomial (here $t^3 - t - 1$) with integer coefficients all of whose conjugates have modulus smaller than 1.*

(2) *In [PS16] several connections between the ω-languages P^ω_{aba}, P^ω_{aabaa} and the smallest Pisot number t_P are derived. In particular, it was shown that, for sufficiently large n, we have $f_{aba}(n) = \mathrm{INT}\big(\frac{2\cdot t_P^2 + 3\cdot t_P + 2}{2\cdot t_P + 3} \cdot t_P^n\big)$ and $f_{aabaa}(n) = \mathrm{INT}\big(\frac{13\cdot t_P^2 + 16\cdot t_P + 9}{5\cdot(2\cdot t_P + 3)} \cdot t_P^n\big)$ where $\mathrm{INT}(\alpha)$ is the integer closest to the real number α.*

Here the coefficient $\frac{13\cdot t_P^2+16\cdot t_P+9}{5\cdot(2\cdot t_P+3)}$ *for* $aabaa$ *is larger than the one for* aba*. This shows that the subword complexity of* ξ_{aabaa} *exceeds the one of* ξ_{aba}.

Acknowledgement

My first acquaintance with Solomon Marcus was when I read his book *Teoretiko-množestvennyje modeli jazykov* [Mar70] in early 1970s. But it

Figure 1: Participants at DMTCS'01, Constanţa

was not until 2001 when I met him at the conference "Discrete Mathematics and Theoretical Computer Science 2001" organised by the "Ovidius" University of Constanţa, Romania. Here his talk "Languages, Infinite Words and their Interaction", the content of which appeared in his papers [MP94, Mar02] and [Mar04] drew my attention to the subject of the present paper.

Two years later my joint paper with Solomon Marcus [CMS03] was published. In 2004 I met Solomon Marcus again at the small conference on " Automata and Formal Languages " in Caputh organised by the Computer Science group of the Potsdam University. There he presented the talk " Contextual Grammars as a Bridge Between the Analytical and the Generative Approach to Natural Languages ".

Figure 2: Professor Marcus presenting his talk in Caputh, September 2004

References

[AFI91] Alberto Apostolico, Martin Farach, and Costas S. Iliopoulos. Optimal superprimitivity testing for strings. *Inform. Process. Lett.*, 39(1):17–20, 1991.

[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.

[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.

[BWZ90] Véronique Bruyère, Li Min Wang, and Liang Zhang. On completion of codes with finite deciphering delay. *European J. Combin.*, 11(6):513–521, 1990.

[CMS03] Cristian S. Calude, Solomon Marcus, and Ludwig Staiger. A topological characterization of random sequences. *Inform.*

Process. Lett., 88(5):245–250, 2003. Reprinted in S. Marcus. *Words and Languages Everywhere*, Polimetrica, Milano, 2007, pp. 357–364.

[CN10] Julien Cassaigne and François Nicolas. Factor complexity. In *Combinatorics, automata and number theory*, volume 135 of *Encyclopedia Math. Appl.*, pages 163–247. Cambridge Univ. Press, Cambridge, 2010.

[FRS07] Henning Fernau, Klaus Reinhardt, and Ludwig Staiger. Decidability of code properties. *Theor. Inform. Appl.*, 41(3):243–259, 2007.

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

[LR07] Florence Levé and Gwénaël Richomme. Quasiperiodic Sturmian words and morphisms. *Theor. Comput. Sci.*, 372(1):15–25, 2007.

[Mar70] Solomon Markus. *Set-theoretic Models of Languages*. Izdat. "Nauka", Moscow, 1970. [In Russian], Translated from the English by M. V. Arapov, Edited and with an appendix by Ju. A. Šreĭder.

[Mar02] Solomon Marcus. Bridging two hierarchies of infinite words. *J. UCS*, 8(2):292–296, 2002.

[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. Combinatorial algorithms (Perth, 1998).

[MP94] Solomon Marcus and Gheorghe Păun. Infinite (almost periodic) words, formal languages and dynamical systems. *Bulletin of the EATCS*, 54:224–231, 1994.

[PS10] Ronny Polley and Ludwig Staiger. The maximal subword complexity of quasiperiodic infinite words. In Ian McQuillan and Giovanni Pighizzini, editors, *Proceedings Twelfth Annual Workshop on Descriptional Complexity of Formal Systems*, volume 31 of *Electr. Proc. Theor. Comput. Sci. (EPTCS)*, pages 169–176. Open Publishing Association, 2010.

[PS16] Ronny Polley and Ludwig Staiger. Quasiperiods, subword complexity and the smallest Pisot number. *J. Autom. Lang. Comb.*, 21(1-2):93–106, 2016.

[Shy01] Huei-Jan Shyr. *Free Monoids and Languages*. Hon Min Book Company, Taichung, third edition, 2001.

[SS78] Arto Salomaa and Matti Soittola. *Automata-theoretic aspects of formal power series*. Texts and Monographs in Computer Science. 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.

[Sta86] Ludwig Staiger. On infinitary finite length codes. *RAIRO Inform. Théor. Appl.*, 20(4):483–494, 1986.

[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-Verlag, Heidelberg, 2012.