



# Transition Property for Cube-Free Words

Elena A. Petrova<sup>1</sup> · Arseny M. Shur<sup>1</sup>

Published online: 21 April 2020

© Springer Science+Business Media, LLC, part of Springer Nature 2020

## Abstract

We study cube-free words over arbitrary non-unary finite alphabets and prove the following structural property: for every pair  $(u, v)$  of  $d$ -ary cube-free words, if  $u$  can be infinitely extended to the right and  $v$  can be infinitely extended to the left respecting the cube-freeness property, then there exists a “transition” word  $w$  over the same alphabet such that  $uvw$  is cube free. The crucial case is the case of the binary alphabet, analyzed in the central part of the paper. The obtained “transition property”, together with the developed technique, allowed us to solve cube-free versions of three old open problems by Restivo and Salemi. Besides, it has some further implications for combinatorics on words; e.g., it implies the existence of infinite cube-free words of very big subword (factor) complexity.

**Keywords** Power-free word · Cube-free word · Extendable word · Transition property

## 1 Introduction

The concept of power-freeness is in the center of combinatorics on words. This concept expresses the restriction on repeated blocks (factors) inside a word: an

---

This article belongs to the Topical Collection: *Special Issue on Computer Science Symposium in Russia (2019)*

Guest Editor: Gregory Kucherov

---

E.A. Petrova — Supported by the Russian Science Foundation, grant 18-71-00043.

This paper extends the conference paper [15] with full proofs and other details omitted in [15] due to space constraints.

---

✉ Elena A. Petrova  
elena.petrova@urfu.ru

Arseny M. Shur  
arseny.shur@urfu.ru

<sup>1</sup> Ural Federal University, Ekaterinburg, Russia

$\alpha$ -power-free word contains no block which consecutively occurs in it  $\alpha$  or more times. For example, the block *an* in the word *banana* is considered as having  $5/2$  consecutive occurrences; thus the word *banana* is 3-power-free (*cube-free*) but not  $(5/2)$ -power-free or 2-power-free (*square-free*); the block *mag* in the word *magma* occurs consecutively  $5/3$  times; the word *magma* is square-free but not  $(5/3)$ -power-free. Power-free words and languages are studied in lots of papers starting with the seminal works by Thue [20, 21], who proved, in particular, the infiniteness of the sets of binary cube-free words and ternary square-free words. However, many phenomena related to power-freeness are still not understood.

One group of problems about power-free words concerns their structure and extendability. In 1985, Restivo and Salemi presented [16] a list of five problems, originally considered only for ternary square-free words and binary overlap-free words, but equally important for every power-free language. Suppose that a finite alphabet  $\Sigma$  is fixed and we study  $\alpha$ -power-free words over  $\Sigma$ . Here are the problems.

*Problem 1* Given an  $\alpha$ -power-free word  $u$ , decide whether there are infinitely many  $\alpha$ -power-free words having (a) the prefix  $u$ ; or (b) the suffix  $u$ ; or (c) the form  $uvw$ , where  $v$  and  $w$  have equal length. (Such words  $u$  are called, respectively, *right extendable*, *left extendable*, and *two-sided extendable*.)

*Problem 2* Given an  $\alpha$ -power-free word  $u$ , construct explicitly an  $\alpha$ -power-free infinite word having  $u$  as prefix, provided that  $u$  is right extendable.

*Problem 3* Given an integer  $k \geq 0$ , does there exist an  $\alpha$ -power-free word  $u$  with the properties (i) there exists a word  $v$  of length  $k$  such that  $uv$  is  $\alpha$ -power free and (ii) for every word  $v'$  of bigger length,  $uv'$  is not  $\alpha$ -power free?

*Problem 4* Given two  $\alpha$ -power-free words  $u$  and  $v$ , decide whether there is a “transition” from  $u$  to  $v$  (i.e., does there exist a word  $w$  such that  $uwv$  is  $\alpha$ -power free).

*Problem 5* Given two  $\alpha$ -power-free words  $u$  and  $v$ , find explicitly a transition word  $w$ , if it exists.

These natural problems appear to be rather hard. Only for Problem 1a,b there is a sort of a general solution: a backtracking decision procedure exists for all  $k$ -power-free languages, where  $k \geq 2$  is an integer [3, 4]. In a number of cases, the parameters of backtracking were found by computer search, so it is not clear whether this technique can be extended for  $\alpha$ -power-free words with rational  $\alpha$ . The decision procedure also gives no clue to Problem 2.

There is a particular case of binary overlap-free words, for which all problems are solved in [1, 16] (more efficient solutions were given in [2]). These words have a regular structure deeply related to the famous Thue-Morse word, and it seems that all natural algorithmic problems for them are solved. For example, the asymptotic order of growth for the binary overlap-free language is computed exactly [6, 7], and even the word problem in the corresponding syntactic monoid has a linear-time solution [19]. Most of the results can be extended, with additional technicalities, to binary

$\alpha$ -power-free words for any  $\alpha \leq 7/3$ , because the structure of these words is essentially the same as of overlap-free words (see, e.g., [8]). However, the situation changes completely if we go beyond the polynomial-size language of binary  $(7/3)$ -power-free words. In the exponential-size  $\alpha$ -power-free languages<sup>1</sup> the diversity of words is much bigger, so it becomes harder to find a universal decision procedure. The only results on Problems 1–5 apart from those mentioned above are the positive answers to Problem 3 (including its two-sided analog) for the two classical test cases: for ternary square-free words [12] and for binary cube-free words [11].

In this paper, we study cube-free words over arbitrary alphabets. Still, the crucial case is the one of the binary alphabet; the central part of the paper is the proof of the following transition property of binary cube-free words.

**Theorem 1** *For every pair  $(u, v)$  of binary cube-free words such that  $u$  is right extendable and  $v$  is left extendable, there exists a binary word  $w$  such that  $uwv$  is cube free.*

After proving Theorem 1 in Section 3, we use it and its proof to derive further results. In Section 4 we prove the transition property for arbitrary alphabets (Theorem 3), while in Section 5 we use this property to solve the cube-free case of Restivo–Salemi Problems 2, 4, and 5. Thus, all Restivo–Salemi problems for binary cube-free words are solved; this is the first fully solved case since the original publication of the problems. For cube-free words over bigger alphabets, only Problem 3 is not yet solved.

We finish the introduction with two remarks. First, the result of Theorem 1 and Theorem 3 (see Section 4) was conjectured in a slightly weaker form for all infinite power-free languages [18, Conj 1]. This conjecture is related to the properties of finite automata recognizing some approximations of power-free languages and was supported by extensive numerical studies. The transition words can be naturally interpreted as transitions in those automata and the transition property forces the automata to be strongly connected. Second, recently it was shown [17, Thm 39] that the transition property implies the existence of infinite  $\alpha$ -power-free words of very big subword complexity. Namely, Theorems 1 and 3 imply that for every  $d \geq 2$  there exists a  $d$ -ary cube-free infinite word which contains *all* two-sided extendable  $d$ -ary cube-free finite words as factors.

## 2 Preliminaries

### 2.1 Notation and Definitions

By default, we study words over finite alphabets  $\Sigma_d$  of cardinality  $d \geq 2$ , writing  $\Sigma_d = \{a, b, c_1, \dots, c_{d-2}\}$  (mostly we work with  $\Sigma_2 = \{a, b\}$ ). Standard notions

<sup>1</sup>For  $\alpha > 7/3$ , the language of binary  $\alpha$ -power-free words has exponential size [8]. *The exponential conjecture* says that for  $k \geq 3$  all infinite power-free languages over  $k$  letters have exponential size. This conjecture is proved for  $k \leq 10$  [9, 10] and odd  $k$  up to 101 [22].

of factor, prefix, and suffix are used. The set of all finite (nonempty finite, infinite) words over an alphabet  $\Sigma$  is denoted by  $\Sigma^*$  (resp.,  $\Sigma^+$ ,  $\Sigma^\infty$ ). We write  $[i..j]$  for the range  $i, i+1, \dots, j$  of positive integers; the notation  $w[i..j]$  stands for the factor of the word  $w$  occupying this range as well as for the particular occurrence of this factor in  $w$  at position  $i$ . Note that  $w[i..i] = w[i]$  is just the  $i$ th letter of  $w$ . Let  $w[i_1..j_1]$  and  $w[i_2..j_2]$  be two factors of  $w$ . If the ranges  $[i_1..j_1]$  and  $[i_2..j_2]$  have a nonempty intersection, their intersection and union are also ranges; we refer to the factors of  $w$ , occupying these ranges, as the *intersection* and the *union* of  $w[i_1..j_1]$  and  $w[i_2..j_2]$ . The word  $\overleftarrow{w} = w[n] \dots w[1]$  is called the *reversal* of the word  $w$  of length  $n$ .

We write  $\lambda$  for the empty word and  $|w|$  for the length of a word  $w$  (infinite words have length  $\infty$ ). A word  $w$  has *period*  $p < |w|$  if  $w[1..|w|-p] = w[p+1..|w|]$ ; the prefix  $w[1..p]$  of  $w$  is the *root* of this period of  $w$ . One of the most useful properties of periodic words is the following.

**Lemma 1** (Fine, Wilf [5]) *If a word  $u$  has periods  $p$  and  $q$  and  $|u| \geq p + q - \text{gcd}(p, q)$  then  $u$  has period  $\text{gcd}(p, q)$ .*

A *cube* is a nonempty word of the form  $uuu$ , also written as  $u^3$ ; we refer to  $u$ ,  $|u|$  as the root and the period of this cube. A word is *cube-free* (*overlap-free*) if it has no cubes as factors (resp., no factors of the form  $cwcwc$ , where  $c$  is a letter). There exist binary overlap-free (and thus cube-free) infinite words [21].

Let  $\Sigma_d$  be fixed. A word  $w \in \Sigma_d^* \cup \Sigma_d^\infty$  is called a *right context* of a cube-free word  $u \in \Sigma_d^*$  if  $uw$  is cube free; we call  $u$  *right extendable* if it has an infinite right context (or, equivalently, infinitely many finite right contexts). Left contexts and left extendability are defined in a symmetric way.

The *Thue–Morse morphism*  $\theta$  is defined over  $\Sigma_2^+$  by the rules  $\theta(a) = ab, \theta(b) = ba$ . The fixed points of  $\theta$  are the infinite *Thue–Morse word*

$$\mathbf{T} = abbabaabbaababbabaababbaabbabaab \dots$$

and its complement, obtained from  $\mathbf{T}$  by exchanging  $a$ 's and  $b$ 's. We refer to the factors of  $\mathbf{T}$  as *Thue–Morse factors*. The word  $\mathbf{T}$ , first introduced by Thue in [21] and rediscovered many times, possesses a huge number of nice properties; we need just a few. The Thue–Morse word is overlap free, *uniformly recurrent* (every Thue–Morse factor occurs in  $\mathbf{T}$  infinitely many times with a bounded gap), and *closed under reversals* ( $u$  is a Thue–Morse factor iff  $\overleftarrow{u}$  is).

### 2.2 Uniform Words and Markers

We call a word  $w \in \Sigma_2^*$  *uniform* if  $w = c\theta(u)d$  for some  $c, d \in \{a, b, \lambda\}, u \in \Sigma_2^*$ ; a uniform word with  $d = \lambda$  is *right aligned*. Similarly, a uniform infinite word has the form  $c\theta(\mathbf{u})$  for  $c \in \{a, b, \lambda\}, \mathbf{u} \in \Sigma_2^\infty$ . Such “almost”  $\theta$ -images play a crucial role in further considerations. Note that all factors and suffixes of  $\mathbf{T}$  are uniform. The following observation is well known.

**Observation 1** A word  $u \in \Sigma_2$  is uniform iff all occurrences of factors of the form  $cc$  in  $u$ , where  $c \in \Sigma_2$ , are at positions of the same parity.

Thus the word is non-uniform iff it has the factors of the form  $cc$  occurring in positions of different parity. The following observation is straightforward.

**Observation 2** A cube-free word  $u \in \Sigma_2$  is non-uniform iff it contains at least one of the factors  $aabaa$ ,  $aababaa$ ,  $bbabb$ ,  $bbababb$ .

All right (resp., left) contexts of the word  $ababa$  begin (resp., end) with  $a$ , so  $ababa$  occurs in a cube-free word only as a prefix/suffix or inside the non-uniform factor  $aababaa$  (the same argument applies to  $babab$ ). This allows us to view binary cube-free words as sequences of uniform factors separated by markers  $aabaa$ ,  $ababa$ ,  $babab$ , and  $bbabb$ , which break uniformity.

The importance of markers for the analysis of cube-free words is demonstrated by the following theorem, proved in Section 3.2.

**Theorem 2** Every right-extendable cube-free word  $u \in \Sigma_2$  has an infinite right context with finitely many markers.

### 3 Proof of the Transition Property for Binary Words

The proof of Theorem 1 consists of two stages. In the first stage we show that its result is implied by Theorem 2. In the second stage we prove Theorem 2. All words in this section are over  $\Sigma_2$  if the converse is not stated explicitly.

#### 3.1 Reduction to Theorem 2

**Lemma 2** Suppose that cube-free words  $u$  and  $v$  have right contexts which are Thue-Morse factors of length  $2|u|$  and  $2|v|$  respectively. Then there exists a word  $w$  such that  $uw\overleftarrow{v}$  is cube free.

*Proof* Let  $u_1$  and  $v_1$  be the mentioned contexts of  $u$  and  $v$  respectively. Since  $\mathbf{T}$  is recurrent and closed under reversals, there exists a Thue-Morse factor  $w = u_1 w_1 \overleftarrow{v_1}$  for some  $w_1 \neq \lambda$ . Assume to the contrary that  $uw\overleftarrow{v}$  contains a cube of period  $p$ . If this cube intersects  $u$  and  $\overleftarrow{v}$  simultaneously, then  $p$  is a period of  $w$  and, since  $w$  is overlap free,  $2p \geq |w|$ ; by definition,  $|w| > 2(|u| + |v|)$ , implying  $|uw\overleftarrow{v}| < 3p$ . This is a contradiction, because a word cannot be shorter than its factor. Now w.l.o.g. assume that the cube intersects  $u$ . Then it must contain the whole  $u_1$ . Hence  $p$  is a period of  $u_1$  and thus  $p \geq |u_1|/2 \geq |u|$  because  $u_1$  is overlap free. Further, the overlap-freeness of  $w$  means that  $u$  contains at least the whole period of the cube, implying  $p \leq |u|$ . So we have  $|u| = p$  and the cube is a prefix of  $uw\overleftarrow{v}$ . But in this

case  $|u_1| \geq 2p$  and the cube is contained in the cube-free word  $uu_1$ , resulting in a contradiction.  $\square$

We say that a cube-free word  $u$  is *T-extendable* if it has a right context of the form  $w\mathbf{T}[n..\infty]$  for some  $w \in \Sigma^*$ ,  $n \geq 1$ . By Lemma 2, if the words  $u$  and  $\overleftarrow{v}$  are T-extendable, there is a word  $w$  such that  $uwv$  is cube-free. We analyze T-extendability in Lemmas 3–5.

**Lemma 3** *If a uniform cube-free word  $u$  has a right context of length 3, or is right aligned and has a right context of length 2, then  $u$  is T-extendable.*

*Proof* We assume  $|u| \geq 5$ ; otherwise,  $u$  is a factor of  $\mathbf{T}$  and there is nothing to prove. Consider two cases.

*Case 1:*  $u = c\theta(v) = cv_1 \dots v_n$ , where  $c \in \{\lambda, a, b\}$ ,  $v_i \in \{ab, ba\}$  for  $i \in [1..n]$ . W.l.o.g.,  $v_n = ab$ .

*Claim 1.* At least one of the words  $ua, ubab$  is cube free.

Assume that  $ua$  is not cube free and thus has a suffix  $y^3$  of period  $p$ . Then  $u$  has the suffix  $u'$  of length  $3p - 1$  and period  $p$ . If  $u' \neq u$ , by the cube-freeness of  $u$  one has  $u[|u|-3p+1] \neq u[|u|-2p+1]$ ,  $u[|u|-3p+2] = u[|u|-2p+2]$ . Note that  $u[|u|-2p+1..|u|-2p+2] = v_{n-p+1} \in \{ab, ba\}$ . Then  $u[|u|-3p+1] = u[|u|-3p+2]$ , so these two letters do not form a block  $v_i$ ; then  $p$  must be odd. Consider the suffix  $xx = v_{n-p+1} \dots v_n$  of length  $2p$  of  $u'$ . Both prefix and suffix of  $x$  of length  $p - 1$  are concatenations of blocks  $ab, ba$ . Hence  $x$  consists of alternating letters. Then  $x[1] = x[p]$ . On the other hand,  $v_{n-\lfloor p/2 \rfloor} = x[p]x[1] \in \{ab, ba\}$ . This contradiction proves that  $u' = u$  and then  $|u| = 3p - 1$ . Further, since  $ua$  and  $ubb$  end with cubes, the only length-2 right context of  $u$  is  $ba$ , so  $uba = c\theta(vb)$  is cube-free. If  $ubab$  ends with a cube of period  $p'$ , we repeat the above argument for  $uba$  to obtain  $|uba| = 3p' - 1$ . Hence  $3p - 1 = 3p' - 3$ , which is impossible since the periods are integers. So  $ubab$  is cube free and Claim 1 holds.

*Claim 2.* At least one of the words  $uaa, ubabb$  is cube free.

Assume that  $ua$  is cube free. If  $uaa$  has a suffix  $y^3$  of period  $p$ , then  $y$  ends with  $aa$  (recall that  $u$  ends with  $ab$ ). On the other hand, the leftmost  $y$  in the suffix  $y^3$  of  $uaa$  ends with  $v_{n-p+1} \in \{ab, ba\}$ . Thus  $uaa$  cannot have a cube as a suffix and hence is cube free. The same argument works for  $ubabb$  if  $ubab$  is cube free. The reference to Claim 1 concludes the proof.

Assuming that  $uaa$  is cube free, we show that the word

$$\mathbf{v} = u\mathbf{T}[6..\infty] = cv_1 \dots v_{n-1} ab aabbaababba \dots$$

is cube free and then  $u$  is T-extendable. Depending on  $v_{n-1}$ ,  $\mathbf{v}$  has the non-uniform factor  $aabaa$  or  $aababaa$ , and this factor, denoted by  $x$ ,

has a unique occurrence in  $\mathbf{v}$  because  $u$  and  $\mathbf{T}$  are uniform. Note that  $ab\mathbf{T}[6..\infty] = \mathbf{T}[4..\infty]$ , and both words  $a\mathbf{T}[4..\infty]$  and  $aab\mathbf{T}[4..\infty]$  are cube free. So if a cube  $y^3$  is a factor of  $\mathbf{v}$ , then Claim 2 implies that  $x$  is a factor of  $y^3$ . If  $x$  is a factor of  $y^2$ , then  $x$  occurs in  $y^3$  at least twice, which is not the case. So  $x = y'y''$ , where  $y'$  and  $y''$  are nonempty suffix and nonempty prefix of  $y$  respectively. Then  $y = aba$  if  $x = aabaa$ , and  $y = ababa$  if  $x = aababaa$ . In both cases a direct check shows that  $y^3$  is not a factor of  $\mathbf{v}$ . So we proved that  $\mathbf{v}$  is cube free. Assuming that  $ubabb$  is cube free, we use the same argument for another suffix of  $\mathbf{T}$ :

$$\mathbf{v} = u\mathbf{T}[20..\infty] = cv_1 \cdots v_{n-1}ab babbaabbaba \cdots$$

(here  $\mathbf{v}$  contains a unique occurrence of  $bbabb$ ). Note that if  $v_n = ba$ , then we can take  $\mathbf{T}[20..\infty]$  (resp.,  $\mathbf{T}[6..\infty]$ ) as the extension of  $u$  if  $ubb$  (resp.,  $uabaa$ ) is cube free.

Case 2. W.l.o.g., the last letter of  $u$  is  $a$ ;  $u = c\theta(v)a = cv_1 \dots v_n a, c \in \{\lambda, a, b\}, v_i \in \{ab, ba\}$  for  $i \in [1..n]$ .

Case 2.1:  $v_n = ab$ . The word  $u\mathbf{T}[7..\infty] = cv_1 \cdots v_n\mathbf{T}[6..\infty]$  is cube free as in Case 1.

Case 2.2:  $v_n = ba$ . Since  $ua$  ends with  $a^3$ ,  $ub$  is cube free, right aligned and has a right context of length 2. Then  $ub$  is  $\mathbf{T}$ -extendable by Case 1, and so is  $u$ . □

**Lemma 4** *If a cube-free word  $u$  has a uniform right context  $w$  such that  $|w| \geq 2|u| + 3$  and  $w$  has no prefix  $ababa$  or  $babab$ , then  $u$  is  $\mathbf{T}$ -extendable.*

*Proof* Let  $\hat{w}$  be the right aligned prefix of  $w$  of length  $2|u|$  or  $2|u|+1$ . We will prove that  $u\hat{w}$  is  $\mathbf{T}$ -extendable, which implies the result immediately, because any prefix of a  $\mathbf{T}$ -extendable word is  $\mathbf{T}$ -extendable by definition. Suppose  $u\hat{w}$  is non-uniform (otherwise, it is  $\mathbf{T}$ -extendable by Lemma 3). Then it contains markers, and all of them begin in  $u$ , because  $\hat{w}$  is uniform and has no marker  $ababa/babab$  as a prefix. Let  $z$  be the rightmost marker in  $u\hat{w}$ . W.l.o.g., the first letter of  $z$  is  $a$  and we can write  $u = u'au''$ , where  $au''w$  begins with this distinguished occurrence of  $z$ . The words  $u'a$  and  $u''w$  satisfy the condition of the lemma:  $u'a$  is cube-free,  $u''w$  is its uniform right context of length  $\geq 2|u'a| + 3$  and  $u''w$  has no prefix  $ababa/babab$ . So for the rest of the proof we rename  $u'a$  as  $u$  and  $u''\hat{w}$  as  $\hat{w}$ . This renaming retains the value of the word  $u\hat{w}$ ,  $\mathbf{T}$ -extendability of which we want to prove; still,  $\hat{w}$  is right aligned and  $|\hat{w}| \geq 2|u|$ . Since  $\hat{w}$  has a right context of length at least 2, it is  $\mathbf{T}$ -extendable by Lemma 3, and, moreover, there is a suffix  $\mathbf{v}$  of  $\mathbf{T}$  such that  $\hat{w}\mathbf{v}$  is cube-free (one possible choice of  $\mathbf{v}$  is given in the proof of Lemma 3. If  $\hat{w}$  is a factor of  $\mathbf{T}$ , we choose  $\mathbf{v}$  such that  $\hat{w}\mathbf{v}$  is a suffix of  $\mathbf{T}$ ; otherwise,  $\mathbf{v}$  is chosen as in the proof of Lemma 3, Case 1.

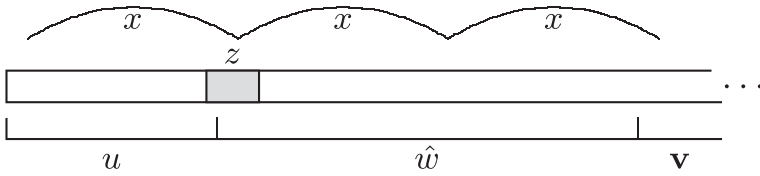


Fig. 1 A cube in the word  $u\hat{w}v$

Assume to the contrary that  $u\hat{w}v$  contains a cube  $x^3$ ; it starts in  $u$  and ends in  $v$ , thus containing the distinguished occurrence of  $z$ : If the marker  $z$  occurs in  $x^3$  only once, its first (resp., last) letter belongs to the leftmost (resp., rightmost)  $x$ , implying  $|x| \leq 3$ . There are, up to symmetry, two cases corresponding to two types of markers:

$$\underbrace{\dots aba}_{u} \underbrace{aba}_{\hat{w}v} \dots \quad \text{or} \quad \dots \underbrace{aba}_{u} \underbrace{bab}_{\hat{w}v} \dots$$

In the first case  $\hat{w}$  is a prefix of  $abaab$  and  $u$  has a suffix  $aba$ , so the condition  $|\hat{w}| \geq 2|u|$  is violated; the second case contradicts the choice of  $v$  (here  $\hat{w}$  is a prefix of  $baba$  and thus a factor of  $\mathbf{T}$ , so  $v$  should be chosen such that  $wv$  is a suffix of  $T$ ). Therefore,  $z$  must have two occurrences in  $x^3$  at distance  $|x|$ . If  $z$  occurs in  $x^3$  to the left of the distinguished occurrence, then  $|x| < |u|$  (see Fig. 1) and hence  $x^3$  cannot end in  $v$ , because  $|\hat{w}| \geq 2|u|$ . Otherwise,  $x^3$  contains exactly two occurrences of  $z$ : the distinguished one and another one on the border of  $\hat{w}$  and  $v$ . Then  $x$  does not contain  $z$ , implying  $|x| \leq |u| + 3$ . On the other hand,  $|x| \geq |\hat{w}| - 3$  as the distance between the occurrences of  $z$ . Now the inequality  $|w| \geq 2|u|$  implies  $|u| \leq 6$  and  $|x| \leq 9$ . This leaves, up to symmetry, the following options for  $x^3$ :

$$\begin{aligned} |x| = 5 : & \text{ abaa } \mathbf{babaabab} \mathbf{aab} \dots & |x| = 9 : & \text{ aabbaa } \mathbf{babaabbaabab} \mathbf{aabbaabab} \dots \\ |x| = 7 : & \text{ abbba } \mathbf{abaabbaaba} \mathbf{abbaab} \dots & & \text{ ababba } \mathbf{abaababbaaba} \mathbf{ababbaaba} \dots \end{aligned}$$

The factor between the marginal letters of markers contains  $\hat{w}$ ; in first three cases, this factor occurs in  $\mathbf{T}$  but  $\hat{w}v$  is not a suffix of  $\mathbf{T}$ , contradicting the choice of  $v$ . In the last case  $\hat{w}$  is not a factor of  $\mathbf{T}$ , so  $v$  is chosen as in the proof of Lemma 3; hence the marker on the border of  $\hat{w}$  and  $v$  must be followed by  $bb$ , not  $ba$ . This contradiction finishes the proof. □

Some right-extendable words have no long uniform right contexts, as Fig. 2 shows. However, a weaker property is enough for our purposes.



Fig. 2 A right-extendable word of length 76 having no long uniform right contexts: all its infinite right contexts begin with the marker  $aabaa$



**Lemma 5** Every cube-free word having an infinite right context with finitely many markers is  $T$ -extendable.

*Proof* Let  $u$  be the word and  $w$  be its context from the conditions of the lemma. The finiteness of the number of markers allows us to write  $w = w_1v$ , where  $v$  is uniform. Then  $uw_1$  has an infinite uniform right context, and hence is  $T$ -extendable by Lemma 4. Then  $u$  is  $T$ -extendable as well.  $\square$

Thus if Theorem 2 holds, then Lemmas 5 and 2 imply Theorem 1.

### 3.2 Proof of Theorem 2

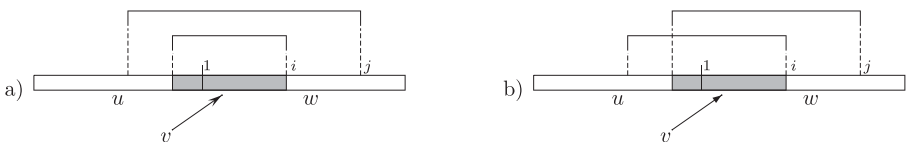
We prove Theorem 2 by reductio ad absurdum; to obtain a contradiction, we use the following lemma on cube-free words over an arbitrary alphabet. (For Theorem 2, bounding  $k$  in Lemma 6 by any function of  $n$  would be sufficient; however, better bounds can be useful for the algorithmic applications.)

**Lemma 6** Let  $u$  be a cube-free word of length  $n$  over an arbitrary fixed alphabet and let  $u$  have a length- $k$  right context  $w$  with the following property: for each  $i = 1, \dots, k$ , there exists an integer  $p_i \geq 2$  such that the suffix of length  $3p_i - 2$  of the word  $u \cdot w[1..i]$  has period  $p_i$  and, moreover,  $p_i \neq p_{i+1}$ . Then  $k = O(\log n)$ ; more precisely,  $k \leq \max\{1, 8.13 \log n - 15.64\}$ .

*Proof* In the proof we can assume  $k \geq 2$ . Let  $1 \leq i < j \leq k$ ,  $p = p_i$ ,  $q = p_j$ ,  $l = j - i$ , and let  $v$  be the intersection of the periodic suffixes of  $u \cdot w[1..i]$  and  $u \cdot w[1..j]$  (see Fig. 3a and b). If  $|v| \geq p + q - \gcd(p, q)$ , then  $v$  has the period  $\gcd(p, q)$  by the Fine–Wilf property (Lemma 1). If  $p \neq q$ , this means that the root of the longer periodic suffix is an integer power of a shorter word; thus  $uw$  contains a cube, which is impossible. If  $p = q$ , then we are in the situation shown in Fig. 3b, and the union of two suffixes has period  $p$  and the length  $3p - 2 + l$ . Since  $p = q$ , one has  $j \neq i + 1$  by conditions of the lemma. Then  $l \geq 2$  and we again obtain a cube. Thus we conclude that

$$|v| \leq p + q - \gcd(p, q) - 1. \tag{1}$$

The case in Fig. 3b corresponds to  $|v| = 3q - 2 - l$ . Comparing this condition to (1), we get  $q \leq \frac{p+l}{2}$  (but  $q = p$  only if  $l \geq 2p - 1$  and  $q = p/2$  only if  $l \geq p/2 - 1$ ). Similarly, the case in Fig. 3a corresponds to  $|v| = 3p - 2$  and we get  $q > 2p$  from (1). Thus, all possible values of the pair  $(l, q)$  are outside the red area in Fig. 4.



**Fig. 3** The mutual location of periodic factors in the word  $uw$  (Lemma 6)

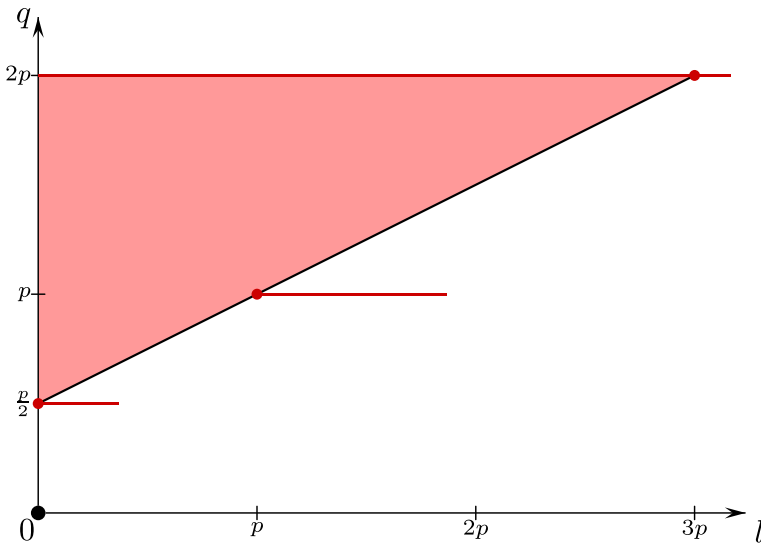


Fig. 4 The restrictions on periods of periodic factors in the word  $uw$  (Lemma 6)

Now we estimate how many elements of the sequence  $\{p_1, \dots, p_k\}$  can belong to the range  $[p..2p]$  for some fixed  $p \geq 2$ . This is an analog of [13, Lemmas 4,5] and [14, Lemma 9]. Let  $i_0 < i_1 < \dots < i_s$  be the list of all positions such that the periodic suffix of  $u \cdot w[1..i_j]$  has the period from the range  $[p..2p]$ ; let  $q_0, \dots, q_s$  denote these periods. Then Fig. 4 gives us the lower bound for the distance  $l_j = i_{j+1} - i_j$  between consecutive positions from the list (we take  $l_j$  as  $l$ ,  $q_j$  as  $p$ , and  $q_{j+1}$  as  $q$  for Fig. 4):

$$l_j \geq 2q_{j+1} - q_j; \quad l_j \geq 2q_{j+1} - 1 \text{ if } q_{j+1} = q_j; \quad l_j \geq q_{j+1} - 1 \text{ if } q_{j+1} = q_j/2. \quad (2)$$

The densest packing of the numbers  $i_j$ , satisfying the restrictions (2), is achieved for  $q_0 = 2p - 1, q_1 = q_3 = q_5 = \dots = p, q_2 = q_4 = q_6 = \dots = p + 1$ : one can take  $i_0 = 1, i_1 = 2$ , and  $i_{2j} = i_{2j-1} + p + 2, i_{2j+1} = i_{2j} + p - 1$  for all subsequent positions. Since  $i_s \leq k$ , we have  $\lceil \frac{s-1}{2} \rceil \cdot (p + 2) + \lfloor \frac{s-1}{2} \rfloor \cdot (p - 1) + 2 \leq k$ . To avoid rounding, we use the inequality  $\frac{(s-1)(2p+1)}{2} \leq \lceil \frac{s-1}{2} \rceil \cdot (p + 2) + \lfloor \frac{s-1}{2} \rfloor \cdot (p - 1)$ . Now we have  $\frac{(s-1)(2p+1)}{2} \leq k - 2$  and obtain the upper bound for the number of periods from the range  $[p..2p]$ :

$$s+1 \leq \frac{2(k-2)}{2p+1} + 2. \quad (3)$$

Since  $3p - 2 \leq |uw| = n + k$ , the maximum possible value of  $p$  is  $\lfloor \frac{n+k+2}{3} \rfloor$ . We partition all possible periods into  $r$  ranges of the form  $[p..2p]$ :

$$[2..4], [5..10], [11..22], \dots, [3 \cdot 2^{r-2} - 1..3 \cdot 2^{r-1} - 2], [3 \cdot 2^{r-1} - 1.. \lfloor \frac{n+k+2}{3} \rfloor].$$

The number of ranges thus satisfies  $r \leq \log \frac{n+k+2}{9} + 1$ . The sum of the upper bounds (3) for all ranges is at least  $k$ ; observing that the number  $2p+1$  in (3) is the first period from the range next to  $[p..2p]$ , we can write

$$k \leq 2(k - 2) \cdot \sum_{i=1}^r \frac{1}{3 \cdot 2^i - 1} + 2r. \tag{4}$$

The sum in (4) is bounded by  $\frac{1}{5} + \frac{1}{11} + \frac{1}{23} + \frac{1}{47} \cdot \sum_{i=0}^{\infty} \frac{1}{2^i} < 0.377$ ; substituting this value and the upper bound for  $r$ , we get

$$0.246(k - 2) \leq 2 \log \frac{n+k+2}{9}. \tag{5}$$

For  $k \geq n - 1$ , (5) implies  $0.246(k - 2) \leq 2 \log \frac{2k+3}{9}$ , but this inequality fails for  $k \geq 2$ . So  $k \leq n - 2$  and we replace (5) with the inequality  $0.246(k - 2) \leq 2 \log \frac{2n}{9}$ , which can be finally transformed to the required bound on  $k$ .  $\square$

*Proof of Theorem 2* For the sake of contradiction, assume that all infinite right contexts of some right-extendable cube-free word  $u$  contain infinitely many markers. W.l.o.g. we can assume that  $u$  ends with a marker (if not, choose a prefix  $v$  of an infinite right context of  $u$  such that  $uv$  ends with a marker, and replace  $u$  with the word  $uv$  having the same property of right contexts). Let  $z$  be the marker which is a suffix of  $u$ . For example, if  $u$  is the word written in the “trunk” of the tree in Fig. 2, then  $z = bbabb$ .

By our assumption,  $u$  has no infinite uniform right contexts. Thus  $u$  has finitely many uniform right contexts (in Fig. 2 such contexts are  $\lambda$ ,  $a$ ,  $aa$ ,  $aab$ ,  $aaba$ , and  $aabb$ ). Other uniform words, being appended to  $u$ , produce cubes; in Fig. 2, appending a word beginning with  $ab$  (resp.,  $aabab$ ,  $aabba$ ) gives the cube  $(bab)^3$  (resp.,  $(babbaabab)^3$ ,  $(babbaababbabbaababbabbaabba)^3$ ). These three cubes contain markers and illustrate three types of cubes with respect to the occurrences of markers ( $x$  below denotes the root of the cube):

- *mini*:  $x^2$  contains no markers,  $x^3$  contains a marker (example:  $x = bab$ ,  $x^3 = babbabbab$ );
- *midi*:  $x$  contains no markers,  $x^2$  contains a marker (example:  $x = babbaabab$ ,  $x^2 = babbaababbabbaabab$ );
- *maxi*:  $x$  contains markers (example:  $x = babbaababbabbaababbabbaabba$ );

Note that mini cubes are exactly those having the root  $x \in \{ab, ba, aba, bab\}$ . Further, the intersection of two markers in a cube-free word is either empty or one-letter; this fact implies that each midi cube contains exactly two markers.

We call a cube-free word  $w$  a *semi-context* of  $u$  if  $uw$  ends with a cube but  $u \cdot w[1..|w|-1]$  is cube free. Next we show the following fact.

*Claim.* The word  $u$  has two distinct uniform semi-contexts  $w_1$  and  $w_2$  such that  $uw_1$  and  $uw_2$  end with midi or maxi cubes.

*Proof of Claim* We need a case analysis. W.l.o.g.,  $z$  begins with  $a$ . Let  $\mathbf{w}$  be an infinite right context of  $u$ .

Case 1:  $z = aabaa$ . We have  $u = \dots baabaa$ , so  $w$  cannot begin with  $a$  or  $baa$  because  $uw$  is cube-free. So  $w$  begins with  $bba$ ,  $babba$ , or  $baba$ . In the first case, some prefixes of words

$$u\mathbf{T}[2..\infty] = \dots baabaa bba baababba \dots, \tag{6}$$

$$u\mathbf{T}[22..\infty] = \dots baabaa bba abbabaab \dots \tag{7}$$

must end with cubes. The longest common prefix  $ubba$  of these words is cube free as a prefix of  $uw$ , so these cubes are different, contain the marker  $z$ , and are not mini. Hence we can take some prefixes of  $\mathbf{T}[2..\infty]$  and  $\mathbf{T}[22..\infty]$  as the semi-contexts required in the claim. If  $w$  begins with  $babba$ , the same result is obtained with prefixes of the words

$$u\mathbf{T}[12..\infty] = \dots baabaa babba baabbaab \dots, \tag{8}$$

$$u\mathbf{T}[20..\infty] = \dots baabaa babba abbabaab \dots \tag{9}$$

Finally, if  $w$  begins with  $baba$ , we can take one word from each pair (say,  $u\mathbf{T}[2..\infty]$  and  $u\mathbf{T}[12..\infty]$ ). Their longest common prefix is the cube-free word  $ub$ , so the cubes given by the corresponding semi-contexts are distinct.

Case 2:  $z = ababa$ . Here  $w = ab \dots$ . Taking the pair of words

$$u\mathbf{T}[7..\infty] = \dots ababa ab baababba \dots, \tag{10}$$

$$u\mathbf{T}[19..\infty] = \dots ababa ab abbaabba \dots, \tag{11}$$

we achieve the same result as in Case 1: some prefixes of these words end with midi or maxi cubes, and these cubes are distinct because the common prefix  $uab$  of the presented words is cube-free. Thus, the claim is proved.  $\square$

Now take the uniform semi-contexts  $w_1, w_2$  given by Claim such that  $uw_1$  and  $uw_2$  end with cubes  $x_1^3$  of period  $p_1$  and  $x_2^3$  of period  $p_2$  respectively. Let  $w$  be the longest common prefix of  $w_1$  and  $w_2$ ; w.l.o.g.,  $w_1 = aww'_1, w_2 = bbw'_2$ . In both  $x_1^3$  and  $x_2^3$ , the suffix  $z$  of  $u$  is the rightmost marker and hence matches an earlier occurrence of the same marker in  $u$ . These occurrences are different, because  $z$  is followed by  $wa$  in  $x_1^3$  and by  $wb$  in  $x_2^3$ . In particular,  $p_1 \neq p_2$ . W.l.o.g.,  $p_1 > p_2$ ; then  $x_1$  contains  $z$  and so  $x_1^3$  is maxi (see Fig. 5).

Let  $z_1 = z, z_2, \dots, z_m$  be all markers in  $u$ , right to left. We factorize  $u$  as  $u = y_m \dots y_2 y_1$ , where  $y_i$  begins with the first letter after  $z_{i+1}$  ( $y_m$  is a prefix of  $u$ ) and ends with the last letter of  $z_i$  (even if  $z_{i+1}$  and  $z_i$  overlap); see Fig. 6 for the example. Assume that  $z_{j+1}$  matches  $z_1$  in the maxi cube  $x_1^3$  (note that  $j \geq 2$ , because a marker

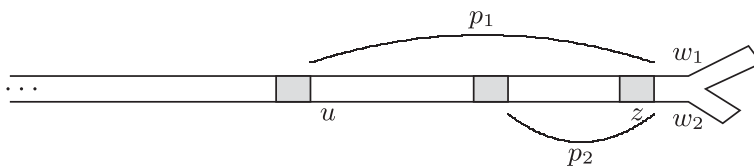
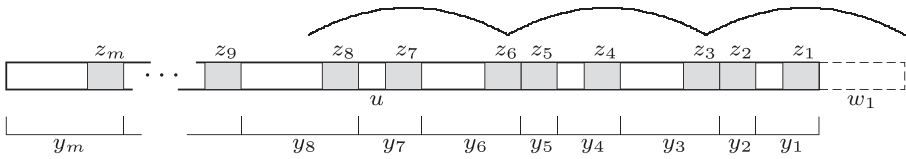


Fig. 5 Semi-contexts  $w_1$  and  $w_2$  of the word  $u$ : periods of cubes and corresponding markers. Three grey factors are markers equal to  $z$ , other markers are not shown



**Fig. 6** Marker-based factorization of the word  $u$ . Markers are grey, arcs indicate the cube after appending  $w_1$  to  $u$

with a smaller number matches  $z_1$  in  $x_2^3$ ;  $j = 3$  in Fig. 6). Then  $z_{j-1}, \dots, z_1$  are in the rightmost occurrence of  $x_1$ , and  $z_j$  is either also in this occurrence or on the border between the middle and the rightmost occurrences (in Fig. 6, the latter case is shown). Depending on this,  $x_1^3$  contains either  $3j$  or  $3j - 1$  markers. Further, we see that  $w_1$  is a prefix of  $y_j$ ,  $y_1 = y_{j+1}, \dots, y_{2j-2} = y_{3j-2}$ .

Let us extend  $u$  to the right by a context  $y_0$  such that  $uy_0$  is right extendable,  $y_0$  ends with a marker  $z_0$ , and all proper prefixes of  $y_0$  are uniform. Applying all the above argument to  $uy_0$  and its factorization  $y_m \cdots y_1 y_0$ , we get another maxi cube (say,  $x_0^3$ ) and the corresponding set of equalities between  $y_i$ 's. Note that  $y_0 \neq y_j$ : as was mentioned in the previous paragraph,  $y_j$  has the prefix  $w_1$ , while  $y_0$  cannot have this prefix because  $uw_1$  contains a cube.

Let us iterate the procedure of appending a context  $k$  times, getting a right-extendable word  $uy = y_m \cdots y_1 y_0 \cdots y_{1-k}$  as the result (according to our assumption on  $u$ , the number  $k$  can be arbitrarily big). Now consider the finite alphabet  $\Gamma = \{y_m, \dots, y_1, y_0, \dots, y_{1-k}\}$  and let  $U = y_m \cdots y_2 y_1$ ,  $Y = y_0 \cdots y_{1-k}$  be words over  $\Gamma$ . They are cube free and  $Y$  is a length- $k$  right context of  $U$ . Each word  $U \cdot Y[1..i]$  ends with a suffix having some period  $p_i$  and length  $3p_i - 2$  or  $3p_i - 1$ . In addition,  $p_i \neq p_{i+1}$ , because  $y_{-i} \neq y_{p_i-i}$ . So all conditions of Lemma 6 are satisfied, and we apply it to get an upper bound on  $k$ . The existence of this bound contradicts our assumption that all infinite right contexts of  $u$  have infinitely many markers. The theorem is proved.  $\square$

### 4 Transition Property for Big Alphabets

Here we extend the results of the previous section to arbitrary finite alphabets, following the main idea for the binary alphabet.

**Theorem 3** *For every  $d \geq 3$  and every pair  $(u, v)$  of cube-free words from  $\Sigma_d^*$  such that  $u$  is right extendable and  $v$  is left extendable, there exists a word  $w \in \Sigma_d^*$  such that  $uwv$  is cube free.*

As in the binary case, we use an auxiliary theorem about the existence of an infinite context with finitely many markers (but the markers are different now).

**Theorem 4** *Let  $d \geq 3$ . Every right-extendable cube-free word  $u \in \Sigma_d^*$  has an infinite right context with finitely many occurrences of all letters except for  $a$  and  $b$ .*

*Proof* We follow the main idea of the proof of Theorem 2 and use the same notation. The difference, which actually simplifies the argument, is that the role of markers is now played by the *c*-letters  $c_1, \dots, c_{d-2}$ . (One simplification is the fact that any cube containing a marker is a maxi cube in the terminology of Theorem 2.) Aiming at a contradiction, assume that all infinite right contexts of some right-extendable cube-free word  $u$  contain infinitely many *c*-letters. W.l.o.g. we can assume that  $u$  ends with a *c*-letter; we denote this letter by  $z$ . As in the proof of Theorem 2, we let  $z_1 = z, z_2, \dots, z_m$  be all markers in  $u$ , right to left, and factorize  $u$  as  $u = y_m \cdots y_2 y_1$ , where  $y_i$  begins with the first letter after  $z_{i+1}$  and ends with the *c*-letter  $z_i$ .

By our assumption,  $u$  contains finitely many contexts from  $\{a, b\}^*$ ; then it has two semi-contexts  $w_1 = uaw'_1, w_2 = wbw'_2 \in \{a, b\}^+$  (each of the words  $w, w_1, w_2$  may be empty). Let  $x_1^3, x_2^3$  be suffixes of  $uw_1$  and  $uw_2$  respectively, and w.l.o.g.  $|x_1| = p_1 > p_2 = |x_2|$ . The suffix  $zw_1$  of  $x_1$  matches some earlier occurrence of  $zw_1$  in  $u$ ; same for the suffix  $zw_2$  of  $x_2$ . As in the proof of Theorem 2 we see that  $z_{j+1} = z_1$  for some  $j \geq 2$ ,  $w_1$  is a prefix of  $y_j$ , and the equalities  $y_1 = y_{j+1}, \dots, y_{2j-1} = y_{3j-1}$  hold.

Next we extend  $u$  to the right by a context  $y_0$  such that  $uy_0$  is right extendable,  $y_0$  ends with a marker  $z_0$ , and all proper prefixes of  $y_0$  are over  $\{a, b\}$ . Applying all the above argument to  $uy_0$  and its factorization  $y_m \cdots y_1 y_0$ , we get another cube  $x_0^3$  and the corresponding set of equalities between  $y_i$ 's. Again,  $y_0 \neq y_j$ , since  $y_j$  has the prefix  $w_1$ , while  $y_0$  has not. After iterating the procedure of appending a context  $k$  times, we obtain a right-extendable word  $uy = y_m \cdots y_1 y_0 \cdots y_{1-k}$  and consider the words  $U = y_m \cdots y_2 y_1, Y = y_0 \cdots y_{1-k}$  over the alphabet  $\Gamma = \{y_m, \dots, y_1, y_0, \dots, y_{1-k}\}$ . They are cube free and  $Y$  is a length- $k$  right context of  $U$ . Each word  $UY[1..i]$  ends with a suffix having some period  $p_i$  and length  $3p_i - 1$ . In addition,  $p_i \neq p_{i+1}$ , because  $y_{-i} \neq y_{p_i-i}$ . So we can apply Lemma 6 to get an upper bound on  $k$ . The existence of this bound contradicts our assumption that all infinite right contexts of  $u$  have infinitely many *c*-letters. Hence  $u$  has an infinite context with finitely many *c*-letters, as required. □

*Proof of Theorem 3* By Theorem 4, the word  $u \in \Sigma_d^*$  has an infinite right context with finitely many *c*-letters. First we note that we can choose such a context containing a *c*-letter (if a context  $\mathbf{w}$  is over  $\Sigma_2$ , one can get another context of  $u$  replacing, say, the letter  $\mathbf{w}[|u|]$  with  $c_1$ ). So we can write this context  $\mathbf{w}$  as  $x\mathbf{u}_1$ , where  $x$  ends with a *c*-letter and  $\mathbf{u}_1 \in \Sigma_2^\infty$ . Let  $u_1$  be the prefix of  $\mathbf{u}_1$  of length  $\lceil |ux|/2 \rceil$ . In the same way, we take a right context  $\overleftarrow{y} \overleftarrow{v}_1$  of  $\overleftarrow{v}$  and the prefix  $\overleftarrow{v}_1 \in \{a, b\}^*$  of  $\overleftarrow{v}_1$  of length  $\lceil |yv|/2 \rceil$ . Then the binary words  $u_1, v_1$  are cube free,  $u_1$  is right extendable, and  $v_1$  is left extendable. Applying Theorem 1, we take a binary transition word  $w_1$  such that  $u_1 w_1 v_1$  is cube free. Then  $s = uxu_1 w_1 v_1 yv$  is cube free. Indeed,  $x$  ends with a *c*-letter,  $y$  begins with a *c*-letter, and these *c*-letters are separated by a cube-free word over  $\Sigma_2$ . Hence a cube in  $s$ , if any, must contain one of these *c*-letters. But the lower bounds on  $|u_1|$  and  $|v_1|$  imply that this *c*-letter cannot match another *c*-letter to produce a cube (recall that  $uxu_1$  and  $v_1 yv$  are cube free). Thus  $s$  is cube-free and we obtain a transition word  $xu_1 w_1 v_1 y$  for the pair  $(u, v)$ . □

## 5 Solving the Restivo-Salemi Problems and Future Work

To give the solutions to the Restivo–Salemi Problems 2, 4, and 5, recall the solution to Problem 1a [4]: a  $d$ -ary  $\alpha$ -power-free word  $u$  is right extendable iff it has a right context of length  $f_{\alpha,d}(|u|)$  for some computable function  $f_{\alpha,d}$ . Algorithm 1 below solves Problem 4.

The natural next step is to find an *efficient* algorithm for Problem 4. The function  $f_{3,d}(n)$  is sublinear, but the search space is still of size  $2^{n^{\Omega(1)}}$ . The possible way to a polynomial-time solution is to strengthen the connection with Lemma 6 to show that it is sufficient to process the contexts of length  $O(\log n)$ , where  $n = \max\{|u|, |v|\}$ .

For Problem 2, the first step is the reduction to the binary case. Let  $u \in \Sigma_d, d \geq 3$ , be a right-extendable cube-free word; we write  $u = u'cu''$ , where  $c$  is the rightmost  $c$ -letter in  $u$ . We check all cube-free words  $w \in \Sigma_2^*$  such that  $u''w$  is right extendable and  $|u''w| = \lceil |u|/2 \rceil$  for being right contexts of  $u$ . If  $w$  is a right context of  $u$ , then any (binary) right context of  $u''w$  is a right context of  $uw$ , so the problem is reduced to binary words. If no word  $w$  suits, we take the shortest right context of  $u$  of the form  $vc$ , where  $v \in \Sigma_2^*, c \in \{c_1, \dots, c_{d-2}\}$  such that the word  $u_1 = uvc$  is right extendable; such a context can be found in finite time because  $|v| < |u|/2$ . Then we replace  $u$  by  $u_1$  and repeat the search of long binary right contexts. By Theorem 4, we will succeed after a finite number of iterations, and Lemma 6 gives the upper bound on the maximum number  $k$  of iterations depending on  $|u|$ . Thus we end this step getting a word  $y\hat{u}$  such that  $uy\hat{u}$  is cube free,  $\hat{u} \in \Sigma_2^*$ , and all binary right contexts of  $\hat{u}$  are right contexts of  $uy\hat{u}$ . If  $u$  is binary, we skip this step setting  $\hat{u} = u$ .

---

**Algorithm 1** Deciding the existence of a transition word for cube-free words  $u, v \in \Sigma_d$ .

---

- For all words  $w \in \Sigma_d$  such that  $|w| \leq f_{3,d}(|u|)$  check whether  $w$  is a right context of  $u$
  - If a context  $w$  with the suffix  $v$  is found, return “yes”
  - If no context of length  $f_{3,d}(|u|)$  is found, return “no”
  - Else ▷  $u$  is right extendable
    - For all words  $w \in \Sigma_d$  such that  $|w| \leq f_{3,d}(|v|)$  check whether  $w$  is a left context of  $v$
    - If a context  $w$  with the prefix  $u$  is found, return “yes”
    - If no context of length  $f_{3,d}(|v|)$  is found, return “no”
    - Else return “yes” ▷  $v$  is left extendable; apply Theorem 1 or 3
- 

On the second step we further reduce the problem to uniform words. We act as in the first step, using Theorem 2 and Lemma 4. Namely, we check for uniform contexts and if  $\hat{u}$  has no uniform context  $w$  of length  $2|\hat{u}| + 3$  such that  $\hat{u}w$  is right extendable and  $w$  has no prefix  $ababa/babab$ , we append the shortest context  $v$  ending with a marker, repeating the search for  $\hat{u}_1 = \hat{u}v$ . Theorem 2 guarantees that we will find the required uniform context in at most  $k$  iterations, where  $k$  is as in Lemma 6. Thus at this step we build a right context  $\hat{y}\hat{w}$  of  $\hat{u}$  such that  $\hat{u}\hat{y}\hat{w}$  is right extendable,  $|\hat{w}| \geq 2|\hat{u}\hat{y}| + 3$  and  $\hat{w}$  is uniform and has no prefix  $ababa/babab$ .

Finally we choose, as described in Lemma 4, a suffix  $\mathbf{T}[r..\infty]$  of  $\mathbf{T}$  which is a right context of  $\hat{w}$ : if  $\hat{w} = \mathbf{T}[i..j]$  for some  $i, j$ , then we take  $r = j + 1$ , otherwise the choice is performed according to Case 1 in the proof of Lemma 3. Now Lemma 4 guarantees that  $\hat{u}\hat{y}\hat{w}\mathbf{T}[r..\infty]$  is cube free. Thus the infinite right context of the original word  $u$  is given by the finite word  $Y = y\hat{u}\hat{y}\hat{w}$  and the number  $r$ . The above description is summarized below as Algorithm 2.

Again, the natural direction of the future work is to make Algorithm 2 efficient.

Finally we approach Problem 5. We first run Algorithm 1, which can provide us with an example of a transition word if  $u$  or  $\overleftarrow{v}$  is not right extendable. If both  $u, \overleftarrow{v}$  are right extendable, we run for each of them Algorithm 2, getting  $Y_1, Y_2, r_1, r_2$  such that  $uY_1\mathbf{T}[r_1..\infty]$  and  $\overleftarrow{\mathbf{T}[\infty..r_2]}Y_2v$  are cube free. It remains to use Lemma 2: take big enough  $r'_1, r'_2$  and find a word  $w$  such that  $\mathbf{T}[r_1..r'_1]w\overleftarrow{\mathbf{T}[r'_2..r_2]}$  is a factor of  $\mathbf{T}$  and a transition word for the pair  $(uY_1, Y_2v)$ ; the uniform recurrence of  $\mathbf{T}$  ensures that the word  $w$  can be found in finite time. Thus  $Y_1\mathbf{T}[r_1..r'_1]w\overleftarrow{\mathbf{T}[r'_2..r_2]}Y_2$  is the transition word for the pair  $(u, v)$ , so Problem 5 is solved.

---

**Algorithm 2** Finding an infinite right context of a right-extendable cube-free word  $u \in \Sigma_d$ .

---

- $U \leftarrow u, Y \leftarrow \lambda$
  - If  $U$  is not binary ▷ first step
    - While  $U$  has no long binary right context
      - Find the shortest right context  $vc$  ending with a  $c$ -letter
      - $U \leftarrow Uvc, Y \leftarrow Yvc$
    - $\hat{u} \leftarrow$  long binary right context of  $U, Y \leftarrow Y\hat{u}$
  - Else  $\hat{u} = u$
  - While  $\hat{u}$  has no long uniform right context without prefix  $ababa/babab$  ▷ second step
    - Find the shortest right context  $v$  such that  $\hat{u}v$  ends with a marker
    - $\hat{u} \leftarrow \hat{u}v, Y \leftarrow Yv$
  - $\hat{w} \leftarrow$  long uniform right context of  $\hat{u}$  without prefix  $ababa/babab, Y \leftarrow Y\hat{w}$
  - Find  $r$  such that  $\mathbf{T}[r..\infty]$  is a right context of  $\hat{w}$  ▷ final step
  - return  $Y, r$
- 

Once again, it is clear that some steps of the above solution can be significantly sped up, so it would be nice to finally get a polynomial-time algorithm for Problem 5 (and thus for Problems 1, 2, 4 as well). From the experimental study we learned that if a length- $n$  cube-free word is not right extendable, then likely not only all its right contexts have the length  $O(\log n)$ , but the number of such contexts is  $O(\log n)$ . The proof of this fact would lead to a linear-time solution of Problem 1.

Another obvious continuation of the current research is the study of the same problems for other power-free languages. One line is to use Thue-Morse words to solve Problems 2, 4, and 5 for other binary power-free languages. For example, we



are able to extend the results of Section 3.1 to  $\alpha$ -power-free binary words for any  $\alpha \in (5/2, 3]$ , changing only some constants. Another line is to obtain similar results for ternary square-free words, in the absence of such a strong tool as Thue-Morse words.

## References

1. Carpi, A.: On the centers of the set of weakly square-free words on a two-letter alphabet. *Inform. Process. Lett.* **19**, 187–190 (1984)
2. Carpi, A.: Overlap-free words and finite automata. *Theoret. Comput. Sci.* **115**, 243–260 (1993)
3. Currie, J.D.: On the structure and extendibility of  $k$ -power free words. *European J. Combinatorics* **16**, 111–124 (1995)
4. Currie, J.D., Shelton, R.O.: The set of  $k$ -power free words over  $\sigma$  is empty or perfect. *European J. Combinatorics* **24**, 573–580 (2003)
5. Fine, N.J., Wilf, H.S.: Uniqueness theorems for periodic functions. *Proc. Amer. Math. Soc.* **16**, 109–114 (1965)
6. Guglielmi, N., Protasov, V.: Exact computation of joint spectral characteristics of linear operators. *Found. Comput. Math.* **13**, 37–97 (2013)
7. Jungers, R.M., Protasov, V.Y., Blondel, V.D.: Overlap-free words and spectra of matrices. *Theoret. Comput. Sci.* **410**, 3670–3684 (2009)
8. Karhumäki, J., Shallit, J.: Polynomial versus exponential growth in repetition-free binary words. *J. Combin. Theory, Ser. A* **104**, 335–347 (2004)
9. Kolpakov, R., Rao, M.: On the number of Dejean words over alphabets of 5, 6, 7, 8, 9 and 10 letters. *Theoret. Comput. Sci.* **412**, 6507–6516 (2011)
10. Ochem, P.: A generator of morphisms for infinite words. *RAIRO Inform. Théor App.* **40**, 427–441 (2006)
11. Petrova, E.A., Shur, A.M.: Constructing premaximal binary cube-free words of any level. *Internat. J. Found. Comp. Sci.* **23**(8), 1595–1609 (2012)
12. Petrova, E.A., Shur, A.M.: Constructing premaximal ternary square-free words of any level. In: *Proc. 37Th Internat. Conf. on Mathematical Foundations of Computer Science. MFCS 2012, LNCS*, vol. 7464, pp. 752–763 (2012)
13. Petrova, E.A., Shur, A.M.: On the tree of ternary square-free words. In: *Combinatorics on Words - 10Th Internat. Conf. WORDS 2015, Proceedings, LNCS*, vol. 9304, pp. 223–236. Springer, Berlin (2015)
14. Petrova, E.A., Shur, A.M.: On the Tree of Binary Cube-Free Words Developments in Language Theory - 21St International Conference, DLT 2017, Proceedings, Lecture Notes in Computer Science, Vol. 10396, Pp. 296–307. Springer (2017)
15. Petrova, E.A., Shur, A.M.: Transition property for cube-free words. In: *Computer Science - Theory and Applications - 14Th International Computer Science Symposium in Russia, CSR 2019, Proceedings, Lecture Notes in Computer Science*, vol. 11532, pp. 311–324. Springer, Berlin (2019)
16. Restivo, A., Salemi, S.: Some decision results on non-repetitive words. In: Apostolico, A., Galil, Z. (eds.) *Combinatorial algorithms on words, NATO ASI series*, vol. F12, pp. 289–295. Springer, Berlin (1985)
17. Shallit, J., Shur, A.: Subword complexity and pattern avoidance. *Theoret. Comput. Sci.* <https://doi.org/10.1016/j.tcs.2018.09.010> (2018)
18. Shur, A.M.: Two-sided bounds for the growth rates of power-free languages. In: *Proc. 13Th Int. Conf. on Developments in Language Theory. DLT 2009, LNCS*, vol. 5583, pp. 466–477. Springer, Berlin (2009)
19. Shur, A.M.: Deciding context equivalence of binary overlap-free words in linear time. *Semigroup Forum* **84**, 447–471 (2012)

20. Thue, A.: Über unendliche Zeichenreihen. *Norske Vid. Selsk. Skr. Mat. Nat. Kl.* **7**, 1–22 (1906)
21. Thue, A.: Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen. *Norske Vid. Selsk. Skr. Mat. Nat. Kl.* **1**, 1–67 (1912)
22. Tunev, I.N., Shur, A.M.: On Two stronger versions of Dejean's conjecture. In: *Proc. 37Th Internat. Conf. on Mathematical Foundations of Computer Science. MFCS 2012, LNCS*, vol. 7464, pp. 801–813. Springer, Switzerland (2012)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.