Theoretical Computer Science 412 (2011) 5649–5655

Contents lists available at SciVerse ScienceDirect

Theoretical Computer Science

journal homepage: www.elsevier.com/locate/tcs

On Brlek–Reutenauer conjecture

L'. Balková *, E. Pelantová, Š. Starosta

Department of Mathematics FNSPE, Czech Technical University in Prague, Trojanova 13, 120 00 Praha 2, Czech Republic

ARTICLE INFO

Article history:
Received 26 January 2011
Received in revised form 6 June 2011
Accepted 22 June 2011
Communicated by D. Perrin

Keywords:
Defect
Rich word
Full word
Infinite word
Factor complexity
Palindromic complexity

ABSTRACT

Brlek and Reutenauer conjectured that any infinite word $\mathbf{u}$ with language closed under reversal satisfies the equality $2D(\mathbf{u}) = \sum_{n=0}^{+\infty} T_{\mathbf{u}}(n)$ in which $D(\mathbf{u})$ denotes the defect of $\mathbf{u}$ and $T_{\mathbf{u}}(n)$ denotes $\mathcal{C}_{\mathbf{u}}(n+1) - \mathcal{C}_{\mathbf{u}}(n) + 2 - \mathcal{P}_{\mathbf{u}}(n+1) - \mathcal{P}_{\mathbf{u}}(n)$, where $\mathcal{C}_{\mathbf{u}}$ and $\mathcal{P}_{\mathbf{u}}$ are the factor and palindromic complexity of $\mathbf{u}$, respectively. Brlek and Reutenauer verified their conjecture for periodic infinite words. Using their result, we prove the conjecture for uniformly recurrent words. Moreover, we summarize results and some open problems related to defects, which may be useful for the proof of the Brlek–Reutenauer conjecture in full generality.

1. Introduction

There have recently been quite a lot of papers devoted to palindromes in infinite words. Droubay, Justin, and Pirillo determined in [6] the upper bound on the number of distinct palindromes occurring in a finite word — a finite word w contains at most $|w| + 1$ different palindromes, where $|w|$ denotes the length of w. The difference between the utmost number $|w| + 1$ and the actual number of palindromes in w is called the defect of w and it is usually denoted by $D(w)$. An infinite word $\mathbf{u}$ whose factors all have zero defect was baptized rich or full. In [1], Baláži, et al. proved for infinite words with language closed under reversal an inequality relating the palindromic and factor complexity of an infinite word $\mathbf{u}$ denoted $\mathcal{P}_{\mathbf{u}}$ and $\mathcal{C}_{\mathbf{u}}$, respectively. For such infinite words, it holds

$$\mathcal{C}_{\mathbf{u}}(n+1) - \mathcal{C}_{\mathbf{u}}(n) + 2 - \mathcal{P}_{\mathbf{u}}(n) - \mathcal{P}_{\mathbf{u}}(n+1) \geq 0 \quad \text{for all } n \in \mathbb{N}. \tag{1}$$

In [5], Bucci, et al. showed that rich words with language closed under reversal can be characterized by the equality in (1). Brlek, et al. in [3] defined the defect $D(\mathbf{u})$ of an infinite word $\mathbf{u}$ as the maximum defects of all its factors and they studied its value for periodic words.

Recently, in [2], the authors of this paper have proved that for a uniformly recurrent word $\mathbf{u}$, its defect $D(\mathbf{u})$ is finite if and only if the equality in (1) is attained for all but a finite number of indices n.

Despite the fact that numerous researchers study palindromes, only recently Brlek and Reutenauer have noticed that the value of defect is closely tied with the expression on the left-hand side of (1) — let us denote it by $T_{\mathbf{u}}(n)$. They have shown that for periodic infinite words with language closed under reversal, it holds $2D(\mathbf{u}) = \sum_{n=0}^{+\infty} T_{\mathbf{u}}(n)$. Their conjecture says that the same equation holds for all infinite words with language closed under reversal.

In this paper, using the result of Brlek and Reutenauer for periodic words, we will prove that the Brlek–Reutenauer conjecture is true for uniformly recurrent words and in the last chapter we will discuss some aspects concerning the conjecture for infinite words that are not uniformly recurrent.

* Corresponding author. Tel.: +420 224358544.
E-mail addresses: lubomira.balkova@fjfi.cvut.cz, l.balkova@centrum.cz (L'. Balková).

doi:10.1016/j.tcs.2011.06.031

2. Preliminaries

By $\mathcal{A}$ we denote a finite set of symbols called *letters*; the set $\mathcal{A}$ is therefore called an *alphabet*. A finite string $w = w_0w_1 \ldots w_{n-1}$ of letters from $\mathcal{A}$ is said to be a *finite word*, its length is denoted by $|w| = n$. Finite words over $\mathcal{A}$ together with the operation of concatenation and the empty word ϵ as the neutral element form a free monoid $\mathcal{A}^*$. The map

$$w = w_0w_1 \ldots w_{n-1} \mapsto \overline{w} = w_{n-1}w_{n-2} \ldots w_0$$

is a bijection on $\mathcal{A}^*$, the word $\overline{w}$ is called the *reversal* or the *mirror image* of w. A word w which coincides with its mirror image is a *palindrome*.

Under an *infinite word* we understand an infinite string $\mathbf{u} = u_0u_1u_2 \ldots$ of letters from $\mathcal{A}$. A finite word w is a *factor* of a word v (finite or infinite) if there exist words p and s such that $v = pws$. If $p = \epsilon$, then w is said to be a *prefix* of v, if $s = \epsilon$, then w is a *suffix* of v.

The *language* $\mathcal{L}(\mathbf{u})$ of an infinite word $\mathbf{u}$ is the set of all its factors. Factors of $\mathbf{u}$ of length n form the set denoted by $\mathcal{L}_n(\mathbf{u})$. We say that the language $\mathcal{L}(\mathbf{u})$ is *closed under reversal* if $\mathcal{L}(\mathbf{u})$ contains with every factor w, also its reversal $\overline{w}$.

For any factor $w \in \mathcal{L}(\mathbf{u})$, there exists an index i such that w is a prefix of the infinite word $u_iu_{i+1}u_{i+2} \ldots$. Such an index is called an *occurrence* of w in $\mathbf{u}$. If each factor of $\mathbf{u}$ has infinitely many occurrences in $\mathbf{u}$, the infinite word $\mathbf{u}$ is said to be *recurrent*. It is easy to see that if the language of $\mathbf{u}$ is closed under reversal, then $\mathbf{u}$ is recurrent (a proof can be found in [7]). For a recurrent infinite word $\mathbf{u}$, we may define the notion of a *complete return word* of any $w \in \mathcal{L}(\mathbf{u})$. It is a factor $v \in \mathcal{L}(\mathbf{u})$ such that w is a prefix and a suffix of v and w occurs in v exactly twice. Under a *return word* of a factor w is usually understood a word $q \in \mathcal{L}(\mathbf{u})$ such that qw is a complete return word of w. If any factor $w \in \mathcal{L}(\mathbf{u})$ has only finitely many return words, then the infinite word $\mathbf{u}$ is called *uniformly recurrent*. If $\mathbf{u}$ is a uniformly recurrent word, we can find for any $n \in \mathbb{N}$ a number R such that any factor of $\mathbf{u}$ which is longer than R already contains all factors of $\mathbf{u}$ of length n.

The *factor complexity* of an infinite word $\mathbf{u}$ is the mapping $\mathcal{C}_{\mathbf{u}} : \mathbb{N} \mapsto \mathbb{N}$ defined by the prescription $\mathcal{C}_{\mathbf{u}}(n) := \#\mathcal{L}_n(\mathbf{u})$. To determine the first difference of the factor complexity, one has to count the possible extensions of factors of length n. A *right extension* of $w \in \mathcal{L}(\mathbf{u})$ is any letter $a \in \mathcal{A}$ such that $wa \in \mathcal{L}(\mathbf{u})$. Of course, any factor of $\mathbf{u}$ has at least one right extension. A factor w is called *right special* if w has at least two right extensions. Similarly, one can define a *left extension* and a *left special* factor. We will deal mainly with recurrent infinite words $\mathbf{u}$. In such a case, any factor of $\mathbf{u}$ has at least one left extension.

The *defect* $D(w)$ of a finite word w is the difference between the utmost number of distinct palindromes $|w| + 1$ and the actual number of distinct palindromes contained in w. Finite words with zero defects – called *rich* or *full* words – can be viewed as the most saturated by palindromes. This definition may be extended to infinite words as follows.

Definition 2.1. *An infinite word $\mathbf{u} = u_0u_1u_2 \ldots$ is called rich or full, if for any index $n \in \mathbb{N}$, the prefix $u_0u_1u_2 \ldots u_{n-1}$ of length n contains exactly $n + 1$ different palindromes.*

Let us remark that not only all prefixes of rich words are rich, but also all factors are rich. A result from [6] provides us with a handful tool which helps to evaluate the defect of a factor.

Proposition 2.2 ([6]). *A finite or infinite word $\mathbf{u}$ is rich if and only if the longest palindromic suffix of w occurs exactly once in w for any prefix w of $\mathbf{u}$.*

In accordance with the terminology introduced in [6], the factor with a unique occurrence in another factor is called *unioccurrent*. From the proof of the previous proposition directly follows the next corollary.

Corollary 2.3. *The defect $D(w)$ of a finite word w is equal to the number of prefixes w' of w, for which the longest palindromic suffix of w' is not unioccurrent in w'. In other words, if b is a letter and w a finite word, then $D(wb) = D(w) + \delta$, where $\delta = 0$ if the longest palindromic suffix of wb occurs exactly once in wb and $\delta = 1$ otherwise.*

This corollary implies that $D(v) \geq D(w)$ whenever w is a factor of v. It enables to give a reasonable definition of the defect of an infinite word (see [3]).

Definition 2.4. *The defect of an infinite word $\mathbf{u}$ is the number (finite or infinite)*

$$D(\mathbf{u}) = \sup\{D(w) \mid w \text{ is a prefix of } \mathbf{u}\}.$$

Let us point out several facts concerning defects that are easy to prove:

(1) If we consider all factors of a finite or an infinite word $\mathbf{u}$, we obtain the same defect, i.e.,

$$D(\mathbf{u}) = \sup\{D(w) \mid w \in \mathcal{L}(\mathbf{u})\}.$$

(2) Any infinite word with finite defect contains infinitely many palindromes.
(3) Infinite words with zero defect correspond exactly to rich words.

Periodic words with finite defect have been studied in [3] and in [7]. It holds that the defect of an infinite periodic word with the minimal period w is finite if and only if $w = pq$, where both p and q are palindromes. Words with finite defect have been studied in [2] and [7].

The number of palindromes of a fixed length occurring in an infinite word is measured by the so called *palindromic complexity* $\mathcal{P}_{\mathbf{u}}$, the mapping which assigns to any non-negative integer n the number

$$\mathcal{P}_{\mathbf{u}}(n) := \#\{w \in \mathcal{L}_n(u) \mid w \text{ is a palindrome}\}\,.$$

Denote by

$$T_{\mathbf{u}}(n) = \mathcal{C}_{\mathbf{u}}(n+1) - \mathcal{C}_{\mathbf{u}}(n) + 2 - \mathcal{P}_{\mathbf{u}}(n+1) - \mathcal{P}_{\mathbf{u}}(n).$$

The following proposition is proved in [1] for uniformly recurrent words, however the uniform recurrence is not needed in the proof, thus it holds for any infinite word with language closed under reversal.

Proposition 2.5 (*[1]*)**.** *Let* $\mathbf{u}$ *be an infinite word with language closed under reversal. Then*

$$T_{\mathbf{u}}(n) \geq 0, \tag{2}$$

for all $n \in \mathbb{N}$.

It is shown in [5] that this bound can be used for a characterization of rich words as well. The following proposition states this fact.

Proposition 2.6 (*[5]*)**.** *An infinite word* $\mathbf{u}$ *with language closed under reversal is rich if and only if the equality in* (2) *holds for all* $n \in \mathbb{N}$.

Let $\mathbf{u}$ be an infinite word with language closed under reversal. Using the proof of Proposition 2.5, those $n \in \mathbb{N}$ for which $T_{\mathbf{u}}(n) = 0$ can be characterized in the graph language.

An *n-simple path* e is a factor of $\mathbf{u}$ of length at least $n+1$ such that the only special (right or left) factors of length n occurring in e are its prefix and suffix of length n. If w is the prefix of e of length n and v is the suffix of e of length n, we say that the n-simple path e starts in w and ends in v. We will denote by $G_n(\mathbf{u})$ an undirected graph whose set of vertices is formed by unordered pairs $(w, \overline{w})$ such that $w \in \mathcal{L}_n(\mathbf{u})$ is right or left special. We connect two vertices $(w, \overline{w})$ and $(v, \overline{v})$ by an unordered pair $(e, \overline{e})$ if e or $\overline{e}$ is an n-simple path starting in w or $\overline{w}$ and ending in v or $\overline{v}$. Note that the graph $G_n(\mathbf{u})$ may have multiple edges and loops.

Remark 2.7. *Let us point out that if* $\mathcal{L}_n(\mathbf{u})$ *contains no special factor, then* $G_n(\mathbf{u})$ *is an empty graph. In this case the word* $\mathbf{u}$ *is periodic, i.e., there exists a primitive word* w *such that* $\mathbf{u} = w^{\omega}$ *and* $|w| \leq n$. *As proved in [3], since the language of* $\mathbf{u}$ *is closed under reversal, the word* w *is a product of two palindromes. It is easy to see that* $\mathcal{C}_{\mathbf{u}}(n+1) = \mathcal{C}_{\mathbf{u}}(n)$ *and* $2 = \mathcal{P}_{\mathbf{u}}(n+1) + \mathcal{P}_{\mathbf{u}}(n)$. *Therefore* $T_{\mathbf{u}}(n) = 0$.

Lemma 2.8. *Let* $\mathbf{u}$ *be an infinite word with language closed under reversal,* $n \in \mathbb{N}$. *Then* $T_{\mathbf{u}}(n) = 0$ *if and only if both of the following conditions are met:*

(1) *The graph obtained from* $G_n(\mathbf{u})$ *by removing loops is a tree;*
(2) *Any n-simple path forming a loop in the graph* $G_n(\mathbf{u})$ *is a palindrome.*

Proof. It is a direct consequence of the proof of Theorem 1.2 in [1] (recalled in this paper as Proposition 2.5). □

Corollary 2.9. *Let* $\mathbf{u}$ *and* $\mathbf{v}$ *be infinite words with language closed under reversal and* $n \in \mathbb{N}$.

$$\mathcal{L}_{n+1}(\mathbf{v}) \subset \mathcal{L}_{n+1}(\mathbf{u}) \quad \textit{and} \quad T_{\mathbf{u}}(n) = 0 \Longrightarrow T_{\mathbf{v}}(n) = 0.$$

Proof. Our assumptions imply that $G_n(\mathbf{v})$ is a subgraph of $G_n(\mathbf{u})$ and $G_n(\mathbf{u})$ meets both conditions in the previous lemma. These conditions are hereditary, i.e., any connected subgraph inherits these conditions as well. □

3. Brlek–Reutenauer conjecture

Brlek and Reutenauer gave in [4] a conjecture relating the defect and the factor and palindromic complexity of infinite words with language closed under reversal.

Conjecture 3.1 (*Brlek–Reutenauer Conjecture*)**.** *Let* $\mathbf{u}$ *be an infinite word with language closed under reversal. Then*

$$2D(\mathbf{u}) = \sum_{n=0}^{+\infty} T_{\mathbf{u}}(n)\,. \tag{3}$$

It is known from [5] that Conjecture 3.1 holds for rich words.

Theorem 3.2. *Let* $\mathbf{u}$ *be a rich infinite word with the language closed under reversal. Then* (3) *holds.*

Brlek and Reutenauer provided in [4] a result for periodic words.

Theorem 3.3. *Let* $\mathbf{u}$ *be a periodic infinite word. Then* (3) *holds.*

In the sequel, we will prove the following theorem.

Theorem 3.4. *Let* $\mathbf{u}$ *be an infinite word with the language closed under reversal. If* $\mathbf{u}$ *satisfies two assumptions:*

(1) *Both* $D(\mathbf{u})$ *and* $\sum_{n=0}^{+\infty} T_{\mathbf{u}}(n)$ *are finite.*
(2) *For any* $M \in \mathbb{N}$ *there exists a factor* $w \in \mathcal{L}(\mathbf{u})$ *such that*
   - w *contains all factors of* $\mathbf{u}$ *of length* M,
   - ww *is a factor of* $\mathbf{u}$.

Then (3) *holds.*

In order to prove Theorem 3.4, we need to put together several claims. Let us first describe the main ideas of the proof. The assumptions of Theorem 3.4 enable us to construct a periodic word $\mathbf{v}$ with language closed under reversal such that

- $D(\mathbf{u}) = D(\mathbf{v})$ and
- $T_{\mathbf{v}}(n) = T_{\mathbf{u}}(n)$ for all $n \in \mathbb{N}$.

Theorem 3.3 applied to the periodic word $\mathbf{v}$ then concludes the proof.

Let us construct a suitable periodic word. As $D(\mathbf{u})$ is finite, there exists a factor $f \in \mathcal{L}(\mathbf{u})$ such that $D(\mathbf{u}) = D(f)$. Let us denote its length by $H = |f|$. According to the inequality (2), the finiteness of $\sum_{n=0}^{+\infty} T_{\mathbf{u}}(n)$ implies that there exists an integer $N \in \mathbb{N}$ such that $T_{\mathbf{u}}(n) = 0$ for all $n \geq N$. Let us put

$$M = \max\{N, H\}. \tag{4}$$

By Assumption (2), there exists a factor w containing all elements of $\mathcal{L}_M(\mathbf{u})$. Let us define

$$\mathbf{v} = w^{\omega}.$$

Claim 3.5. *The word* w *is a concatenation of two palindromes, in particular, the periodic word* w^{ω} *has the language closed under reversal.*

Proof. Since the factor w contains the factor f and the square ww belongs to $\mathcal{L}(\mathbf{u})$, we have $D(f) \leq D(w) \leq D(ww) \leq D(\mathbf{u})$. As the factor f was chosen to satisfy $D(\mathbf{u}) = D(f)$, we may conclude that

$$D(f) = D(w) = D(ww) = D(\mathbf{u}). \tag{5}$$

The factor ww is longer than the factor f and has the same defect as f. Let us denote by p the longest palindromic suffix of $ww \in \mathcal{L}(\mathbf{u})$. According to Corollary 2.3, the palindrome p occurs in ww exactly once and therefore $|p| > |w|$. There exists a proper prefix w' of w such that $ww = w'p$. Let us denote by w'' the suffix of w for which $w = w'w''$. It means that $p = w''w'w''$. As p is a palindrome, we have $\overline{w''} = w''$ and $\overline{w'} = w'$. Hence the word w is a concatenation of two palindromes. $\square$

Claim 3.6. $D(\mathbf{v}) = D(\mathbf{u})$.

Proof. We will use Theorem 6 from [3]. It implies that if w is a product of two palindromes, then $D(w^{\omega}) = D(ww)$. This together with (5) concludes the proof. $\square$

Claim 3.7. $T_{\mathbf{v}}(n) = T_{\mathbf{u}}(n)$ *for all* $n \in \mathbb{N}$.

Proof. Let us first consider $n \leq M - 1$, where M is the constant given by (4). Since w contains all elements of $\mathcal{L}_M(\mathbf{u})$, we have $M \leq |w|$. Since $ww \in \mathcal{L}(\mathbf{u})$, we also have $\mathcal{L}_M(\mathbf{v}) = \mathcal{L}_M(\mathbf{u})$. It implies

$$\mathcal{C}_{\mathbf{u}}(n) = \mathcal{C}_{\mathbf{v}}(n) \quad \text{and} \quad \mathcal{P}_{\mathbf{u}}(n) = \mathcal{P}_{\mathbf{v}}(n) \quad \text{for all } n \leq M.$$

It gives the statement of the claim for all $n \leq M - 1$.

Now we will consider $|w| > n \geq M$. According to the definition of $N \leq M$, it holds that $T_{\mathbf{u}}(n) = 0$. Since $\mathcal{L}_{n+1}(\mathbf{v}) \subset \mathcal{L}_{n+1}(\mathbf{u})$, Corollary 2.9 gives $T_{\mathbf{v}}(n) = 0$ as well.

Finally, we consider $n \geq |w| \geq M$. Since n is longer than or equal to the period of $\mathbf{v}$ and since w is a product of two palindromes, we have $\mathcal{C}_{\mathbf{v}}(n+1) = \mathcal{C}_{\mathbf{v}}(n)$ and $\mathcal{P}_{\mathbf{v}}(n+1) + \mathcal{P}_{\mathbf{v}}(n) = 2$. It implies $T_{\mathbf{v}}(n) = 0$. The value $T_{\mathbf{u}}(n)$ is zero as well, according to the fact that $N \leq M$. $\square$

Proof of Theorem 3.4. It suffices to put together Claims 3.5–3.7 and to realize that Conjecture 3.1 was already proved for periodic words, here stated as Theorem 3.3. $\square$

4. The Brlek–Reutenauer conjecture holds for uniformly recurrent words

In this section we will show that either both sides in the Brlek–Reutenauer equality (3) are infinite or both assumptions of Theorem 3.4 are satisfied for uniformly recurrent words, which results in the main theorem of this paper.

Theorem 4.1. *If* $\mathbf{u}$ *is a uniformly recurrent infinite word with the language closed under reversal, then* (3) *holds.*

In order to prove Theorem 4.1, we will make use of several equivalent characterizations of infinite words with finite defect.

Theorem 4.2. *Let* $\mathbf{u}$ *be a uniformly recurrent infinite word with language closed under reversal. Then the following statements are equivalent.*

(1) *The defect of* $\mathbf{u}$ *is finite.*
(2) *There exists an integer* K *such that any complete return word of a palindrome of length at least* K *is a palindrome as well.*
(3) *There exists an integer* H *such that the longest palindromic suffix of any factor* w *of length* $|w| \geq H$ *occurs in* w *exactly once.*
(4) *There exists an integer* N *such that*

$$T_{\mathbf{u}}(n) = 0 \quad \text{for all } n \geq N.$$

Proof. (1) and (2) are equivalent by Theorem 4.8 from [7]. It follows from the definition of $D(\mathbf{u})$ that (1) and (3) are equivalent. The equivalence of (1) and (4) was stated as Theorem 4.1 in [2]. □

Corollary 4.3. *Let* $\mathbf{u}$ *be a uniformly recurrent infinite word with language closed under reversal. Then*

$$D(\mathbf{u}) \text{ is finite} \Longleftrightarrow \sum_{n=0}^{+\infty} T_{\mathbf{u}}(n) \text{ is finite}.$$

Thanks to Corollary 4.3, we can focus on uniformly recurrent words $\mathbf{u}$ with finite defect. An important role in the proof of Theorem 4.1 is the presence of squares in $\mathbf{u}$.

Lemma 4.4. *Let* $\mathbf{u}$ *be a uniformly recurrent infinite word with finite defect and with language closed under reversal. Then the set*

$$\{w \in \mathcal{A}^* \mid ww \in \mathcal{L}(\mathbf{u})\}$$

is infinite.

Proof. We shall prove that for any $L \in \mathbb{N}$ there exists a factor w such that $ww \in \mathcal{L}(\mathbf{u})$ and $|w| > L$. Without loss of generality take $L > K$, where K is the constant from the statement (3) of Theorem 4.2. Then any complete return word of a palindrome which is longer than L is a palindrome as well. This implies that $\mathbf{u}$ has infinitely many palindromes. Thus there exists an infinite palindromic branch, i.e., a both-sided infinite word $\dots v_3v_2v_1v_0v_1v_2v_3\dots$, where $v_i \in \mathcal{A}$ for $i = 1, 2, 3, \dots$ and $v_0 \in \mathcal{A} \cup \{\epsilon\}$ such that $v_kv_{k-1}\dots v_0\dots v_{k-1}v_k \in \mathcal{L}(\mathbf{u})$ for any $k \in \mathbb{N}$. Consider a palindrome $q = v_kv_{k-1}\dots v_0\dots v_{k-1}v_k$ where $|q| > 3L$. Since $\mathbf{u}$ is uniformly recurrent, there exists an index $i > k$ such that the factor $f = v_iv_{i-1}\dots v_{k+2}v_{k+1}$ is a return word of q. The factor fq is a complete return word of the palindrome q and therefore fq is a palindrome.

At first suppose that the return word f is longer then $|q|$. In this case, $f = qp$ for some palindrome p. Hence the palindromic branch has as its central factor the word $qpqpq$. We can put $w = qp$.

Now suppose that the return word f satisfies $|f| \leq |q|$. In this case there exists an integer $j \geq 2$ and a factor y such that $fq = f^jy$ and $|y| < |f|$. If we put $w = f^i$, with $i = \lfloor \frac{j}{2} \rfloor$, then $ww \in \mathcal{L}(\mathbf{u})$ and $|w| > \frac{1}{3}|q| \geq L$. □

Proof of Theorem 4.1. By Corollary 4.3, the equality $2D(\mathbf{u}) = \sum_{n=0}^{+\infty} T_{\mathbf{u}}(n)$ holds as soon as one of the sides is infinite. Assume that $D(\mathbf{u}) < +\infty$ and $\sum_{n=0}^{+\infty} T_{\mathbf{u}}(n) < +\infty$. Let $M \in \mathbb{N}$ be an arbitrary integer. As $\mathbf{u}$ is uniformly recurrent, there exists an integer R such that any factor longer than R contains all factors of $\mathbf{u}$ of length at most M. According to Lemma 4.4, the set of squares occurring in $\mathbf{u}$ is infinite, thus there exists a factor w longer than R such that ww belongs to the language of $\mathbf{u}$. Its length guarantees that w contains all elements of $\mathcal{L}_M(\mathbf{u})$.

Consequently, Assumptions (1) and (2) of Theorem 3.4 are met and the equality $2D(\mathbf{u}) = \sum_{n=0}^{+\infty} T_{\mathbf{u}}(n)$ follows. □

5. Open problems

In this section, we summarize which statements concerning defects are known for infinite words which are not necessarily uniformly recurrent.

Let us transform the Brlek–Reutenauer conjecture into a more general question: "For which infinite words $\mathbf{u}$ does the equality

$$2D(\mathbf{u}) = \sum_{n=0}^{+\infty} T_{\mathbf{u}}(n) \tag{6}$$

hold?"

In our summary of properties related to the above question, let us first recall Proposition 4.6 from [7] which applies in full generality.

Proposition 5.1. *Let* $\mathbf{u}$ *be an infinite word.*
$D(\mathbf{u}) \geq \#\big\{\{v, \overline{v}\} \;\big|\; v \neq \overline{v}$ *and* v *or* $\overline{v}$ *is a complete return word in* $\mathbf{u}$ *of a palindrome* $w\big\}$.

In [7], the set $\{v, \overline{v}\}$ is called an *oddity*.

Observation 5.2. *If an infinite word* $\mathbf{u}$ *contains finitely many distinct palindromes, then the equality* (6) *holds.*

Proof. It follows from the definition that $D(\mathbf{u}) = +\infty$. Since $\mathcal{P}_{\mathbf{u}}(n) = 0$ for n large enough and $\mathcal{C}_{\mathbf{u}}$ is non-decreasing, we have $T_{\mathbf{u}}(n) \geq 2$ for such indices n. Consequently, $\sum_{n=0}^{+\infty} T_{\mathbf{u}}(n) = +\infty$. □

Observation 5.3. *Let* $\mathbf{u}$ *be a periodic word. Then the equality* (6) *holds.*

Proof. Theorem 3.3 states this fact for infinite words with language closed under reversal. In [3] it is shown that periodic words whose language is not closed under reversal contain only finitely many palindromes. Thus, the previous observation implies that the equality is reached for such words, too. □

Observation 5.4. *Let* $\mathbf{u}$ *be a uniformly recurrent word. Then the equality* (6) *holds.*

Proof. Theorem 4.1 states this fact for infinite words with language closed under reversal. It is well known for uniformly recurrent words whose language is not closed under reversal that it contains only a finite number of palindromes. In such a case, both sides of (6) are infinite. □

From now on, let us limit our considerations to infinite words containing infinitely many palindromes in their language.

Observation 5.5. *The equality* (6) *does not hold in general for infinite words which are not recurrent.*

Proof. The word $\mathbf{u} = ab^{\omega}$ is rich, i.e., $D(\mathbf{u}) = 0$, however $\sum_{n=0}^{+\infty} T_{\mathbf{u}}(n) = -1$. □

Problem 1. It is an open problem whether the equality (6) holds for recurrent words whose language is not closed under reversal and contains infinitely many palindromes. We have examples for which the equality holds and we have so far no example refuting the equality (6).

Example 5.6. *Let* $\mathbf{u}$ *be an infinite ternary word satisfying* $\mathbf{u} = \lim_{n\to+\infty} u_n$, *where* $u_0 = a$ *and* $u_{n+1} = u_n b^{n+1} c^{n+1} u_n$. *The word* $\mathbf{u}$ *is recurrent, however not closed under reversal (it does not contain the factor cb). On one hand,* $D(\mathbf{u}) = +\infty$ *because* b^k *has non-palindromic complete return words for any* $k \geq 1$, *thus the number of oddities is infinite. On the other hand, since the only left extension of a is c and the only right extension of a is b, it is readily seen that the only palindromes of length greater than* 1 *are of the form* b^n *and* c^n, *thus* $\mathcal{P}_{\mathbf{u}}(n) = 2$ *for all* $n \geq 2$. *It is also easy to show that* c^n, b^n, *and* $b^{n-1}c$ *are distinct left special factors of length* $n \geq 2$, *therefore* $\mathcal{C}_{\mathbf{u}}(n+1) - \mathcal{C}_{\mathbf{u}}(n) \geq 3$ *for all* $n \geq 2$. *This implies that* $\sum_{n=0}^{+\infty} T_{\mathbf{u}}(n) = +\infty$.

In the sequel, let us consider infinite words whose language is closed under reversal and contains infinitely many palindromes. In this case, the sum from the left-hand side of (3) exists by Proposition 2.5, i.e., it is either a nonnegative number or $+\infty$.

Any rich word with language closed under reversal satisfies (6) by Theorem 3.2. For instance, the Rote word $\mathbf{u}$ - the fixed point of the morphism φ defined by $\varphi(0) = 001$ and $\varphi(1) = 111$, i.e., $\mathbf{u} = \varphi(\mathbf{u})$ - is rich because it satisfies $T_{\mathbf{u}}(n) = 0$ for all $n \in \mathbb{N}$, which is not difficult to show. Therefore, the Rote word is an example of an infinite word which is not uniformly recurrent (it contains blocks of ones of any length) satisfying the equality (6). We have, of course, no counterexample which would refute the Brlek–Reutenauer conjecture.

There exist several equivalent characterizations of words with finite defect.

Theorem 5.7. *Let* $\mathbf{u}$ *be an infinite word with language closed under reversal and containing infinitely many palindromes. Then the following statements are equivalent.*

(1) *The defect of* $\mathbf{u}$ *is finite.*
(2) $\mathbf{u}$ *has only finitely many oddities.*
(3) *There exists an integer* H *such that the longest palindromic suffix of any factor* w *of length* $|w| \geq H$ *occurs in* w *exactly once.*

Proof. (1) and (3) are equivalent by the definition of defect. (1) implies (2) by Proposition 5.1. The implication (2) ⇒ (1) was proved as Proposition 4.8 in [7] for uniformly recurrent words. However, we will show that the proof works for words with language closed under reversal and containing infinitely many palindromes too.

Assume that $D(\mathbf{u}) = +\infty$ and the number of oddities is finite.

A finite number of oddities means that only finitely many palindromes can have non-palindromic complete return words. Let the longest such palindrome be of length K.

Since the number of palindromes is infinite, there exist infinitely many non-defective positions. Denote by $u^{(n)}$ the prefix of $\mathbf{u}$ of length n. Then n is a non-defective position if $D(u^{(n-1)}) = D(u^{(n)})$ (such positions correspond to the first occurrences of palindromes).

There exists an integer H such that the prefix of $\mathbf{u}$ of length H contains all palindromes of length lower than $K+3$. Hence, if $n > H$ is a non-defective position, then the longest palindromic suffix of $u^{(n)}$ is of length greater than $K+2$.

Since both the number of defective and non-defective positions is infinite, we can find an index $k > H$ such that k is a defective and $k+1$ a non-defective position. The longest palindromic suffix p of $u^{(k)}$ occurs at least twice in $u^{(k)}$, thus $u^{(k)}$ ends in a non-palindromic complete return word of p. Since $k+1$ is a non-defective position, it can be easily shown by contradiction that the longest palindromic suffix of $u^{(k+1)}$ is of length equal to or lower than $|p| + 2 \leq K + 2$.

This is a contradiction with the fact that non-defective positions greater than H have their longest palindromic suffix longer than $K+2$. $\square$

For words with language closed under reversal, some implications remain valid. The first one is Proposition 4.3 and the second one is Proposition 4.5 from [2].

Proposition 5.8. *Let* $\mathbf{u}$ *be an infinite word with language closed under reversal. Suppose that there exists an integer* N *such that for all* $n \geq N$ *the equality* $T_{\mathbf{u}}(n) = 0$ *holds. Then the complete return words of any palindromic factor of length* $n \geq N$ *are palindromes.*

Proposition 5.9. *Let* $\mathbf{u}$ *be an infinite word with language closed under reversal. If there exists an integer* H *such that for any factor* $f \in \mathcal{L}(\mathbf{u})$ *of length* $|f| \geq H$*, the longest palindromic suffix of* f *is unioccurrent in* f*. Then* $T_{\mathbf{u}}(n) = 0$ *for any* $n \geq H$*.*

The last proposition together with Theorem 5.7 results in the following corollary.

Corollary 5.10. *Let* $\mathbf{u}$ *be an infinite word with language closed under reversal. Then we have*

$$D(\mathbf{u}) < +\infty \Rightarrow \sum_{n=0}^{+\infty} T_{\mathbf{u}}(n) < +\infty.$$

It is an open question whether the implications in the previous propositions can be reversed.

Problem 2. Let $\mathbf{u}$ be an infinite word with language closed under reversal and containing infinitely many palindromes. Assume that there exists an integer K such that all palindromes of length equal to or greater than K have palindromic complete return words. Does there exist an integer N such that $T_{\mathbf{u}}(n) = 0$ for any $n \geq N$?

Problem 3. Let $\mathbf{u}$ be an infinite word with language closed under reversal and containing infinitely many palindromes. Suppose that there exists an integer N such that for all $n \geq N$ the equality $T_{\mathbf{u}}(n) = 0$ holds. Does there exist also an integer H such that for any factor $f \in \mathcal{L}(\mathbf{u})$ of length $|f| \geq H$ the longest palindromic suffix of f is unioccurrent in f?

We have seen that in the proof of the validity of the Brlek–Reutenauer conjecture for uniformly recurrent words, an important role was played by the presence of long squares in such words. This leads to the last open problem.

Problem 4. Find other classes of infinite words containing for any L a factor w such that $|w| > L$ and ww belongs to the language.

Acknowledgements

We acknowledge financial support by the Czech Science Foundation grant 201/09/0584, by the grants MSM6840770039 and LC06002 of the Ministry of Education, Youth, and Sports of the Czech Republic, and by the Grant Agency of the Czech Technical University in Prague, grant No. SGS11/162/OHK4/3T/14.

References

[1] P. Baláži, Z. Masáková, E. Pelantová, Factor versus palindromic complexity of uniformly recurrent infinite words, Theoret. Comput. Sci. 380 (2007) 266–275.
[2] L' Balková, E. Pelantová, Š. Starosta, Infinite words with finite defect, Adv. Appl. Math. (2011) doi:10.1016/j.aam.2010.11.006, the original publication is available at http://www.sciencedirect.com/science/journal/01968858.
[3] S. Brlek, S. Hamel, M. Nivat, C. Reutenauer, On the palindromic complexity of infinite words, in: J. Berstel, J. Karhumäki, D. Perrin (Eds.), Combinatorics on Words with Applications, Int. J. Found. Comput. Sci. 15 (2) (2004), 293–306.
[4] S. Brlek, C. Reutenauer, Complexity and palindromic defect of infinite words, Theoret. Comput. Sci. 412 (4–5) (2011) 493–497.
[5] M. Bucci, A. De Luca, A. Glen, L.Q. Zamboni, A connection between palindromic and factor complexity using return words, Adv. in Appl. Math 42 (2009) 60–74.
[6] X. Droubay, J. Justin, G. Pirillo, Episturmian words and some constructions of de Luca and Rauzy, Theoret. Comput. Sci. 255 (2001) 539–553.
[7] A. Glen, J. Justin, S. Widmer, L.Q. Zamboni, Palindromic richness, European J. Combin 30 (2009) 510–531.