Theoretical Computer Science 445 (2012) 63–74

Contents lists available at SciVerse ScienceDirect

Theoretical Computer Science

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

Bispecial factors in circular non-pushy D0L languages

Karel Klouda *

Faculty of Information Technology, Czech Technical University in Prague, Thákurova 9, 160 00, Prague 6, Czech Republic

ARTICLE INFO

Article history:
Received 7 January 2012
Accepted 3 May 2012
Communicated by J. Karhumaki

Keywords:
Bispecial factors
Circular D0L systems
Non-pushy D0L systems
Critical exponent

ABSTRACT

We study bispecial factors in fixed points of morphisms. In particular, we propose a simple method of finding all bispecial words of non-pushy circular D0L-systems. This method can be formulated as an algorithm. Moreover, we prove that non-pushy circular D0L-systems are exactly those with finite critical exponents.

1. Introduction

Bispecial factors proved to be a powerful tool for better understanding of complexity of aperiodic sequences of symbols from a finite set. One of the most studied families of such sequences are fixed points of morphisms. In this paper we present a method of how to describe the structure of all bispecial factors in a given fixed point.

The method we describe here can be partially spotted in results of several authors: It is a sort of inverse of the algorithm by Cassaigne from paper [1] which is concerned by pattern avoidably. A very similar approach was used in [2] by Avgustinovich and Frid and in [3] by Frid to describe bispecial factors of biprefix circular morphisms and marked uniform morphisms, respectively. Actually, the fact that all bispecial factors in a fixed point can be generated as elements of some easily constructed sequences was noticed in many papers where factor complexity was computed, see, e.g., [4,5]. In this paper we formalize this approach and prove that it works for a very wide class of morphisms, namely non-pushy circular morphisms. Moreover, it seems that the assumptions we need for proofs can be weakened or even omitted and the main theorems remain true.

The paper is organized as follows. In the next section we introduce necessary notation and notions and also explain the importance of bispecial factors. Since it is easier to explain the main result using examples than to formulate it as a theorem, we do so in Section 3. Section 4 contains proofs of the crucial theorems and in Section 5 we explain how to use our results to identify immediately all infinite special branches. In Section 6 we prove that non-pushy circular morphisms are exactly those whose fixed points have finite critical exponents.

2. Preliminaries

Let $\mathcal{A} = \{0, 1, \ldots, n-1\}$, $n \geq 2$, be a finite alphabet of n letters; if needed, we denote this particular n-letter alphabet as $\mathcal{A}_n$. An *infinite word* over the alphabet $\mathcal{A}$ is a sequence $\mathbf{u} = (u_i)_{i\geq 1}$ where $u_i \in \mathcal{A}$ for all $i \geq 1$. If $v = u_j u_{j+1} \cdots u_{j+n-1}$,

* Tel.: +420 774552735.
E-mail address: karel.klouda@fit.cvut.cz.
URL: http://www.kloudak.eu/.

0304-3975/$ – see front matter © 2012 Published by Elsevier B.V.
doi:10.1016/j.tcs.2012.05.007

for some $j, n \geq 1$, then v is said to be a *factor* of $\mathbf{u}$ of length n, the empty word ϵ is the factor of length 0. The set of all finite words over $\mathcal{A}$ is the free monoid $\mathcal{A}^*$, the set of nonempty finite words is denoted by $\mathcal{A}^+ = \mathcal{A}^* \setminus \{\epsilon\}$.

A map $\varphi : \mathcal{A}^* \to \mathcal{A}^*$ is called a *morphism* if $\varphi(wv) = \varphi(w)\varphi(v)$ for every $w, v \in \mathcal{A}^*$. Any morphism φ is uniquely determined by the set of images of letters $\varphi(a)$, for all $a \in \mathcal{A}$. If all these images are nonempty words, the morphism is called *non-erasing*. A famous example of a morphism is the Thue-Morse morphism φ_{TM} defined by

$$\varphi_{\mathrm{TM}}(0) = 01,$$
$$\varphi_{\mathrm{TM}}(1) = 10.$$

This paper studies infinite *fixed points* of morphisms: an infinite word w is a fixed point of a morphism φ if $\varphi(w) = w$. If $\varphi^{\ell}(w) = w$ for some positive ℓ, w is a *periodic point* of φ. The fixed point of φ_{TM} beginning in the letter 0 is the infinite word

$$\mathbf{u}_{\mathrm{TM}} = \lim_{n\to\infty} \varphi_{\mathrm{TM}}^n(0) = \varphi_{\mathrm{TM}}^{\omega}(0) = 0110100110\cdots, \tag{1}$$

which is called the *Thue-Morse word*.

An infinite word $\mathbf{u}$ is *aperiodic* if it is not *eventually periodic*, i.e., there are no finite words v and w such that $\mathbf{u} = vwwww\cdots = vw^{\omega}$. If a word $u = vw$, then v is a *prefix* of u and w is its *suffix*. In this case we put $(v)^{-1}u = w$ and $u(w)^{-1} = v$. Given a morphism φ on $\mathcal{A}$, if $\varphi(a)$ is not a suffix of $\varphi(b)$ for any distinct $a, b \in \mathcal{A}$, then φ is said to be *suffix-free*. *Prefix-free* morphisms are defined analogously.

The *language of a fixed point* $\mathbf{u}$ is the set of all its factors and is denoted by $\mathcal{L}(\mathbf{u})$. When speaking about a morphism, we usually mean a morphism together with its particular infinite fixed point. But a morphism can have more than one fixed point and not all of them must have the same language (this is true if the morphism is primitive). For instance, consider the morphism $0 \mapsto 010$, $1 \mapsto 11$: it has two fixed points, one aperiodic starting in 0 and one periodic starting in 1. Therefore, instead of speaking only about a morphism we will always speak about a morphism and its particular infinite fixed point. A well-established way of how to do so is to treat a morphism and its fixed point as a D0L-system (see, e.g., [6,7]).

Definition 1. A triplet $G = (\mathcal{A}, \varphi, w)$ is called a D0L-system, where $\mathcal{A}$ is an alphabet, φ a morphism on $\mathcal{A}$, and $w \in \mathcal{A}^+$ is an *axiom*. The language of G denoted by $\mathcal{L}(G)$ is the set of all factors of the words $\varphi^n(w)$, for all $n = 0, 1, \ldots$

If φ is non-erasing, then the system is called a PD0L-system.

In what follows, when referring to a D0L-system, we always mean a PD0L-system. In fact, for any D0L-system, it is possible to construct its elementary (not simplifiable) version which is a PD0L-system with an injective morphism [8,9].

Clearly, if $\varphi(a) = av$ for some $a \in \mathcal{A}$, $v \in \mathcal{A}^+$, and if φ is non-erasing, then the language of the D0L-system $(\mathcal{A}, \varphi, a)$ is the language of the infinite fixed point $\varphi^{\omega}(a)$.

There are several tools which help us to study the structure of the language of D0L-systems. We mention here two basic ones: the factor complexity and critical exponent. The factor complexity of a language is the function $C(n)$ which counts the number of factors of length n. An effective method to obtain the factor complexity is using special factors.

Definition 2. Let w be a factor of the language $\mathcal{L}(G)$ of a D0L-system G over $\mathcal{A}$. The set of *left extensions* of w is defined as

$$\mathrm{Lext}(w) = \{a \in \mathcal{A}\colon\ aw \in \mathcal{L}(G)\}.$$

If $\#\mathrm{Lext}(w) \geq 2$, then w is said to be a *left special (LS) factor* of $\mathcal{L}(G)$.

Analogously we define the set of *right extensions* $\mathrm{Rext}(w)$ and a *right special (RS) factor*. If w is both left and right special, then it is called *bispecial (BS)*.

The connection between special factors and factor complexity is described in [5]; the complete knowledge of LS, RS, or BS factors enables one to determine the factor complexity.

The critical exponent is related to the repetitions in the language. Let w be a finite and nonempty word. Any finite prefix v of $w^{\omega} = www\cdots$ is a *power* of w. We denote this by $v = w^r$, in words v is r-*power* of w, where $r = \frac{|v|}{|w|}$. Further, we define the *index* of w in a language $\mathcal{L}(G)$ of a D0L-system G as

$$\mathrm{ind}(w, G) = \sup\left\{r \in \mathbb{Q}\colon\ w^r \in \mathcal{L}(G)\right\}.$$

And finally, the *critical exponent* of the language $\mathcal{L}(G)$ is the number

$$\sup\{\mathrm{ind}(w)\colon\ w \in \mathcal{L}(G)\}.$$

More details about the critical exponent can be found, e.g., in [10]. Examples of how knowledge of BS factors can help to compute the critical exponent of a fixed point of a morphism are in [11,12].

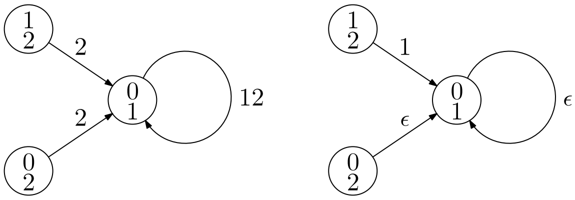

Fig. 1. The graphs defining the f-image for the morphism φ_E.

3. Explaining the main result

Since the main result of this paper is a tool rather than a theorem, we demonstrate it using example morphisms. The tool has two ingredients. The first one is a mapping that maps a BS factor to another one and so, applied repetitively, it generates sequences of BS factors. This mapping is defined by two directed labeled graphs. The other ingredient is a finite set of BS factors such that the sequences generated from them by the mapping cover all BS factors in a given fixed point.

Let us consider the morphism φ_E defined by $0 \mapsto 012$, $1 \mapsto 112$, $2 \mapsto 102$ and the corresponding D0L-system $(\mathcal{A}_3, \varphi_E, 0)$ with the fixed point $\mathbf{u}_E$. The factor 2112 is LS and has left extensions 0 and 1. If we apply the morphism φ_E on this structure – both on the factor and its two extensions, then the resulting factor $\varphi_E(2212)$ is no more LS since the respective extensions $\varphi_E(0) = 012$ and $\varphi_E(1) = 112$ end in the same letter 2. In order to obtain an LS factor, we have to cut off the longest common suffix of the new extensions, here it is 12, and append it to the beginning of the φ_E-image of the factor: the result is the LS factor $12\varphi_E(2212)$ with left extensions 0 and 1. We can proceed in the same manner and obtain another LS factor $12\varphi_E(12)\varphi_E^2(2112)$ again with the same extensions 0 and 1. Clearly, the same process works for RS factors and right extensions.

Let us formalize what we did in the previous paragraph. Instead of BS factors we use a slightly different notion of *BS triplets* $((a, b), v, (c, d))$, where v is a BS factor and (a, b) and (c, d) are unordered pairs of its left and right extensions, respectively. We assume that either avc and bvd or avd and bvc are factors. Thus, $((0, 1), 2112, (0, 1))$ is a BS triplet in $\mathbf{u}_E$. In terms of the previous paragraph, we can get another BS triplet from this one, namely $((0, 1), 12\varphi_E(2112), (0, 1))$; we call this BS triplet the *f-image* of $((0, 1), 2112, (0, 1))$. The fact that left extensions $(0, 1)$ result again in extensions $(0, 1)$ with prepending of 12 can be represented as a directed edge from vertex $(0, 1)$ to vertex $(0, 1)$ with label 12. The edge corresponding to the right extensions starts in $(0, 1)$, ends again in $(0, 1)$ and is labeled by the empty word. Applying this idea on all possible pairs of left and right extensions gives us two directed labeled graphs depicted in Fig. 1. We call these graphs *graph of left and right prolongations*. With these graphs in hand, it is easy to generate infinitely many BS triplets from a given starting one.

The other ingredient of our method bears on the fact that all BS triplets can be generated by taking repetitively f-image of only finitely many *initial* BS triplets. Initial BS triplets are those which are not f-images of another BS triplet. For instance, $((0, 1), 2112, (0, 1))$ is not initial as it is the f-image of the BS triplet $((0, 1), 1, (0, 2))$ which is initial. Later we show how to find all the initial factors for a given fixed point. For the case of $\mathbf{u}_E$, we have eight initial BS triplets:

$$\begin{array}{cccc} ((0,1),121,(0,1)), & ((0,1),12,(0,1)), & ((0,1),21,(0,1)), & ((0,1),2,(0,1)), \\ ((1,2),1,(1,2)), & ((0,2),1,(1,2)), & ((0,2),1,(0,2)), & ((1,2),0,(1,2)). \end{array}$$

The vertices of the graphs from Fig. 1 are just all pairs of distinct letters, but the situation is not that simple for all morphisms. In fact, it happens for graphs of left (right) prolongations only if the respective morphism is suffix-free (prefix-free). This case when the morphism is both prefix- and suffix-free has been already solved in [2], where not only describe the authors all BS factors, but they also give a formula for the factor complexity.

Let us consider the morphism φ_S defined by $0 \mapsto 0012$, $1 \mapsto 2$, $2 \mapsto 012$ and the corresponding D0L-system $(\mathcal{A}_3, \varphi_S, 0)$. Clearly, the morphism is not suffix-free. Let v be a LS factor with left extensions $(1, 2)$. If we apply the morphism as above, we have a problem: the longest common suffix of factors $\varphi_S(1) = 2$ and $\varphi_S(2) = 012$ is 2 and so we do not know what are the left extensions of the factor $2\varphi_S(v)$. A solution is to consider left extensions longer than one letter, in such a case we say left prolongation instead of left extension. Clearly, the factor 1 is always preceded by 0. Hence, let us consider left extensions $(01, 2)$. Now, $\varphi_S(01) = 00122$ is no more a suffix of $\varphi_2 = 012$ and so we know the new left extensions: again $(01, 2)$. In this way we can construct a complete graph defining the respective f-image, the result is in Fig. 3 (the notation will be explained later).

To prove that a proper *finite* set of pairs of left and right extensions of arbitrary length always exists is not trivial. It is not simple even to describe the properties such sets should possess so that they define a correct f-image. We call such sets *left* and *right forky sets*, see Definition 20. It may happen that a finite forky set does not exist and so our method fails. Therefore we will have to put some restriction on the D0L-systems considered: we will assume that the systems are *circular* and *non-pushy*. These notions are explained in the following section.

Finally, we can now state the main result of this paper: Given a circular non-pushy D0L-system with an aperiodic fixed point, there exist finite left and right forky sets defining two directed graphs and a finite set of initial BS triplets such that the corresponding f-image applied repetitively on the initial BS-triplets generates all BS factors.

4. Forky sets and initial factors

4.1. Circular and non-pushy D0L-systems

Any factor of a fixed point of a morphism φ can be decomposed into (possibly incomplete) φ-images of letters. For example, in the case of φ_E we have: 01210 is a factor of $\varphi_E(0)\varphi_E(2)$, i.e., 01210 is composed by φ_E-images of 0 and 2. We denote this using bars, i.e., 012|10. The decomposition may not be unique. For instance 210 is always decomposed as 2|10 but we do not know whether 2 is a suffix of $\varphi_E(0)$ or $\varphi_E(1)$ (it cannot be a suffix of $\varphi_E(2)$ since 22 is not a factor of $\mathbf{u}_E$). In the case of the factor 1, we do not even know where to place the bar if not at all.

A factor can have more than one decomposition; however, if there is a common bar for all these decompositions, this bar is called a *synchronizing point*. Coming back to our example, 210 has a synchronizing point between 2 and 10, formally we say that (2, 10) is a synchronizing point of 210.

Definition 3 (*Cassaigne [1]*)**.** Let φ be a morphism with a fixed point $\mathbf{u}$, φ injective on $\mathcal{L}(u)$, and let w be a factor of $\mathbf{u}$. An ordered pair of factors (w_1, w_2) is called a *synchronizing point* of w if $w = w_1 w_2$ and

$$\forall v_1, v_2 \in \mathcal{A}^*, \big(v_1 w v_2 \in \varphi(\mathcal{L}(\mathbf{u})) \Rightarrow v_1 w_1 \in \varphi(\mathcal{L}(\mathbf{u})) \text{ and } v_2 w_2 \in \varphi(\mathcal{L}(\mathbf{u}))\big).$$

We denote this by $w = w_1|_s w_2$.

Definition 4. A D0L-system $G = (\mathcal{A}, \varphi, w)$ is *circular* if φ is injective on $\mathcal{L}(G)$ and if there exists $D \in \mathbb{N}$ such that any $v \in \mathcal{L}(G)$ such that $|v| \geq D$ has at least one synchronizing point. The integer D is called a *synchronizing delay*.

Some examples of both circular and non-circular D0L-systems follow.

Example 5. The system $G = (\mathcal{A}_2, \varphi_{TM}, 0)$ is circular with a synchronizing delay 4. It is clear that any $w \in \mathcal{L}(G)$ containing 00 or 11 has the synchronizing point $w = \cdots 0|_s 0 \cdots$ or $w = \cdots 1|_s 1 \cdots$. To see that, let us consider a word w of length 4 not containing these two factors. Without loss of generality, assume that w begins in 1, then $w = 1010$. This word can be decomposed into $\varphi_{TM}(0)$ and $\varphi_{TM}(1)$ in exactly two ways: |10|10| and 1|01|0. But the latter one is not admissible since it arises as the φ_{TM}-image of 000 which is not an element of $\mathcal{L}(G)$.

Example 6. The system $G = (\mathcal{A}_2, \varphi, 0)$, where $\varphi(0) = 01, \varphi(1) = 11$, is not circular. Indeed, for all n the word 1^n has no synchronizing point since it can be decomposed as $|11|11|\cdots$ and $1|11|11|\cdots$.

This example is very simple since the respective infinite fixed point $011111\cdots$ is eventually periodic. However, there are also aperiodic non-circular systems.

Example 7. The system $G = (\mathcal{A}_3, \varphi, 0)$, where $\varphi(0) = 010, \varphi(1) = 22, \varphi(2) = 11$, is not circular. The argument is the same as in the previous example, since the words 1^n are for all $n \in \mathbb{N}$ elements of $\mathcal{L}(G)$. However, the infinite word $\varphi^\omega(0) = 0102201011110102\cdots$ is aperiodic.

One can notice that the languages in the both non-circular examples contain an arbitrary power of 1. It is not just a coincidence but a general rule.

Theorem 8 (*Mignosi and Séébold [13]*)**.** *If a* D0L*-system is k-power-free (i.e.,* $\mathcal{L}(G)$ *does not contain the k-power of any word) for some* $k \geq 1$*, then it is circular.*

Thus, non-circular fixed points must have infinite critical exponent, in fact, they must contain an unbounded power of some word.

Theorem 9 (*Ehrenfeucht and Rozenberg [14]*)**.** *Given a* D0L*-system* $G = (\mathcal{A}, \varphi, w)$*, if* $\mathcal{L}(G)$ *contains a k-power for all* $k \in \mathbb{N}$*, then G is* strongly repetitive*, i.e., there exists a nonempty* $v \in \mathcal{L}(G)$ *such that* $v^\ell \in \mathcal{L}(G)$ *for all* $\ell \in \mathbb{N}$*.*

We see that non-circular systems have very special properties. Furthermore, the morphism of a non-circular system cannot be even primitive. A morphism φ over $\mathcal{A}$ is primitive if there is $k \in \mathbb{N}$ such that $\varphi^k(a)$ contains b for all $a, b \in \mathcal{A}$.

Theorem 10 (*Mossé [15]*)**.** *Any* D0L*-system* $G = (\mathcal{A}, \varphi, a)$ *with* φ *injective on* $\mathcal{G}$ *and primitive is circular.*[1]

No matter how non-circular systems seem to be bizarre, there is no known finite algorithm which would decide whether a given general D0L-system is circular or not. Of course, if the respective morphism is primitive, it is easy to prove it in finite steps. Later on we also prove that if the system is non-pushy, then the circularity is equivalent to repetitiveness which is decidable.

Example 11. An example of non-primitive but circular morphism is the one given by $0 \mapsto 0010$, $1 \mapsto 1$. This is the Chacon morphism [16] and 5 is its synchronizing delay.

---

[1] In the article [15] the circular systems are called “recognizable”.

4.2. Non-pushy D0L-systems

The following two definitions and the lemma are taken from [14].

Definition 12. Let $G = (\mathcal{A}, \varphi, w)$ be a D0L-system. A letter $b \in \mathcal{A}$ has *rank zero* if $\mathcal{L}(G_b)$, where $G_b = (\mathcal{A}, \varphi, b)$, is finite.

Definition 13. A D0L-system $G = (\mathcal{A}, \varphi, w)$ is *pushy* if for all $n \in \mathbb{N}$ there exists $v \in \mathcal{L}(G)$ of length n which is composed of letters that have rank zero; otherwise G is *non-pushy*. If G is non-pushy, then $q(G)$ denotes

$$q(G) = \max\{|v| : v \in \mathcal{L}(G) \text{ is composed of letters that have rank zero}\}.$$

Lemma 14. 1. *It is decidable whether or not an arbitrary* D0L-*system is pushy.*
2. *If G is pushy, then $\mathcal{L}(G)$ is strongly repetitive (see Theorem 9).*
3. *If G is non-pushy, then $q(G)$ is effectively computable.*
4. *It is decidable whether or not an arbitrary* D0L-*system is strongly repetitive.*

Corollary 15 (*Krieger [10]*)**.** *Let $G = (\mathcal{A}, \varphi, a)$, for a letter $a \in \mathcal{A}$, be a non-pushy* D0L-*system and let $\mathbf{u} = \varphi^{\omega}(a)$ be an infinite fixed point of φ. There exists a non-erasing morphism φ' and an effectively computable $C \in \mathbb{N}$ such that $\mathbf{u} = (\varphi')^{\omega}(a)$ and for all $v \in \mathcal{L}(G)$ with $|\varphi'(v)| = |v|$ we have $|v| < C$.*

This means that for any word v of length at least C we have $|\varphi(v)| \geq |v|+1$. More generally, if $|v| \geq KC$, then $|\varphi(v)| \geq |v|+K$. We use this in the proof of the main theorem of the following subsection.

In the sequel, we always suppose that the $G = (\mathcal{A}, \varphi, a)$ is such that φ' can be taken equal to φ (in fact φ' is just a power of φ, see the proof of the corollary in [10]). Since the language is the same, this assumption is without loss of generality.

4.3. Forky sets

Our aim is to define properly the notion of f-image introduced in Section 3. As explained, we need to have two directed labeled graphs defined on unordered pairs of left and right prolongations. Since left and right extensions usually refer to letters and the vertices of our graphs might be pairs of words, we give the following definition.

Definition 16. Let $\mathbf{u}$ be an infinite word and w its factor. The set of *left prolongations* of w is the set

$$\mathrm{Lpro}(w) = \{v \in \mathcal{A}^{+} : vw \in \mathcal{L}(\mathbf{u})\}.$$

In an analogous way we define the set of *right prolongations* $\mathrm{Rpro}(w)$.

The sets $\mathrm{Lpro}(w)$ and $\mathrm{Rpro}(w)$ are, in general, infinite.

Our aim is to specify a suitable finite sets $\mathcal{B}_L$ and $\mathcal{B}_R$ of (unordered) pairs of left and right prolongations such that it allows to define correctly an f-image of all triplets $((w_1, w_2), v, (w_3, w_4))$, where v is a BS factor, (w_1, w_2) a pair of its left and (w_3, w_4) a pair of its right prolongations from $\mathcal{B}_L$ and $\mathcal{B}_R$, respectively. The f-image defined by the sets $\mathcal{B}_L$ and $\mathcal{B}_R$ are to be defined as a BS triplet $((w_1', w_2'), v', (w_3', w_4'))$, where (w_1', w_2') and (w_3', w_4') are again in $\mathcal{B}_L$ and $\mathcal{B}_R$ and the factor $v' = f_L(w_1', w_2')\varphi(v)f_R(w_3', w_4')$ is BS. The mappings f_L and f_R are defined as follows.

Definition 17. Let φ be a morphism over $\mathcal{A}$ and let (v_1, v_2) be an unordered pair of words from $\mathcal{A}^{+}$. We define

$$f_L(v_1, v_2) = \text{the longest common suffix of } \varphi(v_1) \text{ and } \varphi(v_2),$$
$$f_R(v_1, v_2) = \text{the longest common prefix of } \varphi(v_1) \text{ and } \varphi(v_2).$$

The purpose of the following definitions is just to describe "good" choices of $\mathcal{B}_L$ and $\mathcal{B}_R$.

Definition 18. Let (w_1, w_2) and (v_1, v_2) be unordered pairs of words. We say that

(i) (w_1, w_2) is a *prefix* (*suffix*) of (v_1, v_2) if either w_1 is a prefix (suffix) of v_1 and w_2 of v_2, or w_1 is a prefix (suffix) of v_2 and w_2 of v_1;

(ii) (w_1, w_2) and (v_1, v_2) are *L-aligned* if

$$(v_1 = uw_1 \text{ or } w_1 = uv_1) \quad \text{and} \quad (v_2 = u'w_2 \text{ or } w_2 = u'v_2)$$

or

$$(v_1 = uw_2 \text{ or } w_2 = uv_1) \quad \text{and} \quad (v_2 = u'w_1 \text{ or } w_1 = u'v_2)$$

for some words u, u'.

Analogously, we define pairs which are *R-aligned*.

Example 19. The pairs (01, 0) and (001, 10) are L-aligned, while (01, 0) and (011, 10) are not L-aligned. Schematically, the notion of L-aligned pairs of words is depicted in Fig. 2.

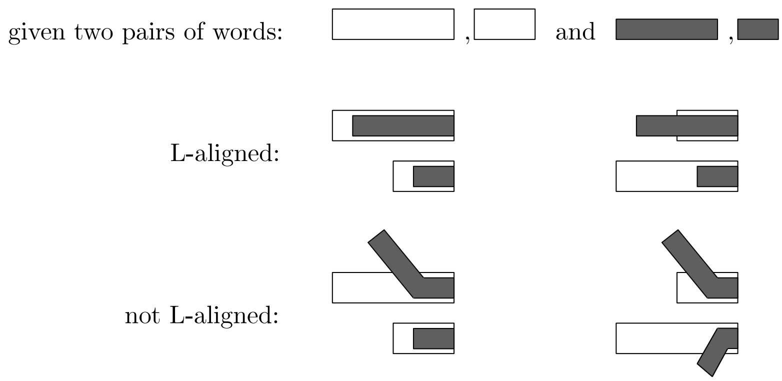

Fig. 2. L-aligned and not L-aligned pairs of words.

Definition 20. Let φ be a morphism with a fixed point $\mathbf{u}$. A finite set $\mathcal{B}_L$ of unordered pairs (w_1, w_2) of nonempty factors of $\mathbf{u}$ is called *L-forky* if all the following conditions are satisfied:

(i) the last letters of w_1 and w_2 are different for all $(w_1, w_2) \in \mathcal{B}_L$,
(ii) no distinct pairs (w_1, w_2) and (w'_1, w'_2) from $\mathcal{B}_L$ are L-aligned,
(iii) for any $v_1, v_2 \in \mathcal{L}(\mathbf{u}) \setminus \{\epsilon\}$ with distinct last letters there exists $(w_1, w_2) \in \mathcal{B}_L$ such that (w_1, w_2) and (v_1, v_2) are L-aligned,
(iv) for any $(w_1, w_2) \in \mathcal{B}_L$ there exists $(w'_1, w'_2) \in \mathcal{B}_L$ such that

$$(w'_1 f_L(w_1, w_2), w'_2 f_L(w_1, w_2))$$

is a suffix of $(\varphi(w_1), \varphi(w_2))$.

Analogously we define an *R-forky* set.

Since the definition may look a bit intricate, we now comment on all the conditions. Condition (i) says that w_1 and w_2 are left prolongations of LS factors (note that all pairs of words $w_1, w_2 \in \mathcal{L}(\mathbf{u})$ are prolongations of the empty word, i.e., elements of $\mathrm{Lpro}(\epsilon)$). Condition (ii) is required to avoid redundancy in $\mathcal{B}_L$. Condition (iii) ensures that any two left prolongations of any LS factor are included in $\mathcal{B}_L$ in the following sense: if we prolong or shorten them in a certain way we obtain a pair from $\mathcal{B}_L$. And, finally, Condition (iv) is there because of the definition of the f-image: we want to be able to apply it repetitively. Note that due to (ii) and (iii) the pair (w'_1, w'_2) from (iv) is uniquely given. Note also that if (i) is satisfied and the words from all the pairs of $\mathcal{B}_L$ are of the same length, then (ii) and (iii) are satisfied automatically.

Example 21. Consider the morphism φ_S from Section 3 defined by $0 \mapsto 0012$, $1 \mapsto 2$, $2 \mapsto 012$. This morphism is injective and primitive and so, by Theorem 10, the respective D0L-system is circular. One can easily prove that 3 is a synchronizing delay (note that all factors containing 2 has a synchronizing point $\cdots 2|\cdots$.)

Since φ_S is prefix-free, we get that the set

$$\mathcal{B}_R = \{(0, 1), (0, 2), (1, 2)\}$$

is R-forky. However, this set is not L-forky: Condition (iv) is not satisfied for any pair since $\varphi_S(1)$ is a suffix of $\varphi_S(2)$ which is a suffix of $\varphi_S(0)$. To remedy this, we consider left prolongations one letter longer which are ending in 1 and 2. Since the list of all factors of length 2 reads

00, 01, 12, 20, 22

the new pairs are

$$(0, 01), (0, 12), (0, 22), (2, 01). \tag{2}$$

For these pairs conditions (i)–(iii) are again satisfied. But (iv) is not satisfied for $(0, 12)$ since $f_L(0, 12) = 012$ and $(\varphi_S(0)(012)^{-1}, \varphi_S(12)(012)^{-1}) = (0, 2)$ has no suffix in list (2). Hence, we have to prolong 12 again. There is only one possibility, namely 012. The resulting set

$$\mathcal{B}_L = \{(0, 01), (0, 012), (0, 22), (2, 01)\}$$

is then L-forky since now we get $(\varphi_S(0)(012)^{-1}, \varphi_S(012)(012)^{-1}) = (0, 00122)$ with a suffix $(0, 22) \in \mathcal{B}_L$.

Theorem 22. *Let φ be a morphism on $\mathcal{A}$ with a fixed point $\mathbf{u} = \varphi^\omega(a)$. If $(\mathcal{A}, \varphi, a)$ is circular non-pushy system, then it has L-forky and R-forky sets.*

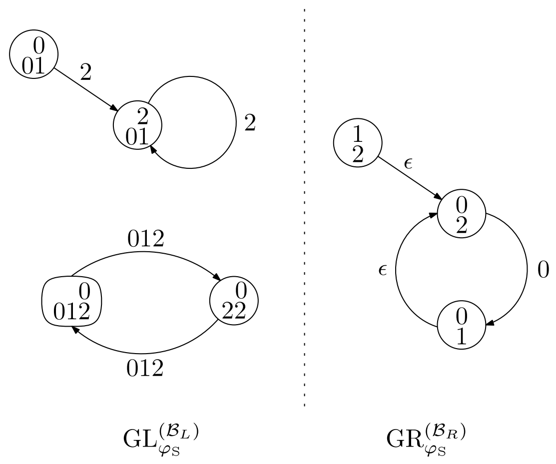

Fig. 3. The graphs $\mathrm{GL}^{(\mathcal{B}_L)}$ and $\mathrm{GR}^{(\mathcal{B}_R)}$ for the morphism φ_S.

Proof. Set $M = DC$, where D is a synchronizing delay and C is the constant from Corollary 15. Define

$$\mathcal{B}_L = \{(w_1, w_2)\colon\ w_1, w_2 \in \mathcal{L}(\mathbf{u}), |w_1| = |w_2| = M,\ \text{and the last letters of } w_1 \text{ and } w_2 \text{ are distinct}\}. \tag{3}$$

We claim that $\mathcal{B}_L$ is L-forky. Conditions (i)–(iii) from Definition 20 are trivially fulfilled. It remains to prove (iv).

It is clear that we must have $|f_L(w_1, w_2)| \le D$ for any (w_1, w_2) since $f_L(w_1, w_2)$ without the last letter does not have any synchronizing point. Now, it suffices to realize that for any w of length M we have $|\varphi(w)| \ge |w| + D = M + D$ and so $(\varphi(w_1)(f_L(w_1, w_2))^{-1}, \varphi(w_2)(f_L(w_1, w_2))^{-1})$ has a suffix in $\mathcal{B}_L$. The proof of existence of an R-forky set is perfectly the same. $\square$

This proof does not give us a general guideline how to construct L-forky and R-forky sets since, as we have recalled earlier, we do not have an efficient algorithm computing a synchronizing delay for a general morphism. Moreover, the L-forky set constructed in the proof is usually too huge. As in the case of our example morphism φ_S from Example 21 (for this morphism $C = 2$ and $D = 3$), there usually exists a smaller L-forky set.

Remark 23. The techniques of the proof of Theorem 22 are the same as those of the proof of Lemma 11 in [10]. This Lemma, however, is concerned with the notion of critical exponent.

Definition 24. Let φ be a morphism with a fixed point $\mathbf{u}$ and let $\mathcal{B}_L$ be an L-forky set. We define the directed labeled *graph of left prolongations* $\mathrm{GL}_{\varphi}^{(\mathcal{B}_L)}$ as follows:

(i) the set of vertices is $\mathcal{B}_L$,
(ii) there is an edge from (w_1, w_2) to (w_3, w_4) if $(w_3 f_L(w_1, w_2), w_4 f_L(w_1, w_2))$ is a suffix of $(\varphi(w_1), \varphi(w_2))$. The label of this edge is $f_L(w_1, w_2)$.

In the same manner we define the *graph of right prolongations* $\mathrm{GR}_{\varphi}^{(\mathcal{B}_R)}$.

As a straightforward consequence of the definition of forky sets (especially of Condition (iv)) we have the following property of the graphs.

Lemma 25. *Each vertex in a graph of left and right prolongations has its out-degree equal to one.*

Consequently, any path long enough in the graph ends in a cycle and any component contains exactly one cycle.

Example 26. The graphs $\mathrm{GL}_{\varphi_S}^{(\mathcal{B}_L)}$ and $\mathrm{GR}_{\varphi_S}^{(\mathcal{B}_R)}$ for φ_S and for the sets $\mathcal{B}_L$ and $\mathcal{B}_R$ from Example 21 are in Fig. 3.

Definition 27. Let φ be a morphism on $\mathcal{A}$ with a fixed point $\mathbf{u}$ and let $\mathcal{B}_L$ and $\mathcal{B}_R$ be L-forky and R-forky sets, respectively. A triplet $((w_1, w_2), v, (w_3, w_4))$ is called a *bispecial (BS) triplet* in $\mathbf{u}$ if $(w_1, w_2) \in \mathcal{B}_L$, $(w_3, w_4) \in \mathcal{B}_R$ and $w_1 v w_3, w_2 v w_4 \in \mathcal{L}(\mathbf{u})$ or $w_1 v w_4, w_2 v w_3 \in \mathcal{L}(\mathbf{u})$.

Lemma 28. *Let φ be a morphism with a fixed point $\mathbf{u}$, let $\mathcal{B}_L$ be an L-forky and $\mathcal{B}_R$ an R-forky set and let $\mathcal{T} = ((w_1, w_2), v, (w_3, w_4))$ be a bispecial triplet of $\mathbf{u}$. If we denote by*

(i) *$g_L(w_1, w_2)$ the end of the edge of $\mathrm{GL}_{\varphi}^{(\mathcal{B}_L)}$ starting in (w_1, w_2),*
(ii) *$g_R(w_3, w_4)$ the end of the edge of $\mathrm{GR}_{\varphi}^{(\mathcal{B}_R)}$ starting in (w_3, w_4),*

then

$$\mathcal{T}' = (g_L(w_1, w_2), f_L(w_1, w_2)\varphi(v)f_R(w_3, w_4), g_R(w_3, w_4))$$

is also a bispecial triplet of $\mathbf{u}$.

Definition 29. Denote $\mathcal{B} = (\mathcal{B}_L, \mathcal{B}_R)$. The bispecial triplet $\mathcal{T}'$ from the previous lemma is called the $f_{\mathcal{B}}$*-image* of a bispecial triplet $\mathcal{T} = ((w_1, w_2), v, (w_3, w_4))$.

Example 30. Consider again the morphism φ_S. If $\mathcal{B} = (\mathcal{B}_L, \mathcal{B}_R)$, where $\mathcal{B}_L$ and $\mathcal{B}_R$ are those from Example 21, then $((0, 012), 0, (0, 1))$ is a bispecial triplet since both 001 and 01200 are factors. Its $f_{\mathcal{B}}$-image reads $((0, 22), 0120012, (0, 2))$ for we have $g_L(0, 012) = (0, 22)$, $f_L(0, 012) = 012$, $g_R(0, 1) = (0, 2)$, and $f_R(0, 1) = \epsilon$.

Condition (iv) from Definition 20 of forky sets allows us to get a compact formula for the $(f_{\mathcal{B}})^n$-image, i.e., $f_{\mathcal{B}}$-image applied repetitively n times.

Lemma 31. *Let φ be a morphism and let $((w_1, w_2), v, (w_3, w_4))$ be a bispecial triplet for some forky sets $\mathcal{B} = (\mathcal{B}_L, \mathcal{B}_R)$. Then for all $n \in \mathbb{N}$ it holds that its $(f_{\mathcal{B}})^n$-image equals*

$$(g_L^n(w_1, w_2), f_L(\varphi^{n-1}(w_1), \varphi^{n-1}(w_2))\varphi^n(v)f_R(\varphi^{n-1}(w_3), \varphi^{n-1}(w_4)), g_R^n(w_3, w_4)).$$

4.4. Initial BS factors

From the previous subsection, we know how to get a sequence of BS factors from some starting one: we just apply $f_{\mathcal{B}}$-image repetitively. The goal of the present subsection is to prove that for circular systems there exists a finite set of initial BS factors (triplets) such that any other BS factor is an $(f_{\mathcal{B}})^n$-image of one of them.

Definition 32. Let φ be a morphism injective on $\mathcal{L}(\mathbf{u})$, where $\mathbf{u}$ is its fixed point, let $\mathcal{B}_L$ and $\mathcal{B}_R$ be L- and R-forky sets, and $\mathcal{T} = ((w_1, w_2), v, (w_3, w_4))$ a bispecial triplet. Assume, without loss of generality, that $w_1 v w_3, w_2 v w_4 \in \mathcal{L}(\mathbf{u})$. An ordered pair of factors (v_1, v_2) is called a *BS-synchronizing point* of $\mathcal{T}$ if $v = v_1 v_2$ and

$$\forall u_1, u_2, u_3, u_4 \in \mathcal{A}^*, (u_1 w_1 v w_3 u_3, u_2 w_2 v w_4 u_4 \in \varphi(\mathcal{L}(\mathbf{u})) \Rightarrow u_1 w_2 v_1, u_2 w_2 v_1, v_2 w_3 u_3, v_2 w_4 u_4 \in \varphi(\mathcal{L}(\mathbf{u})).$$

We denote this by $v = v_1|_{bs} v_2$.

The notion of the BS-synchronizing point is weaker than the one of synchronizing point: it holds that if $v = v_1|_s v_2$, then $v = v_1|_{bs} v_2$. It follows from the following example that the converse is not true.

Example 33. Given a morphism $0 \mapsto 010$, $1 \mapsto 210$, $2 \mapsto 220$, the factor 0 has no synchronizing point. On the other hand, if we take it as the bispecial triplet $((1, 2), 0, (0, 2))$ it has the BS-synchronizing point $0|_{bs}$.

Definition 34. Let φ be a morphism with a fixed point $\mathbf{u}$ which is injective on $\mathcal{L}(\mathbf{u})$. A bispecial triplet $\mathcal{T} = ((w_1, w_2), v, (w_3, w_4))$ is said to be *initial* if it does not have any BS-synchronizing point.

Definition 35. Let $\mathcal{T} = ((w_1, w_2), v, (w_3, w_4))$ be a bispecial triplet which is not initial and let $(v_1, v_2), (v_3, v_4), \ldots, (v_{2m-1}, v_{2m})$ be all its BS-synchronizing points such that $|v_1| < |v_3| < \cdots < |v_{2m-1}|$. The factor v_1 is said to be the *non-synchronized prefix*, v_{2m} the *non-synchronized suffix* and the factor $(v_1)^{-1} v (v_{2m})^{-1}$ is called the *synchronized factor* of $\mathcal{T}$.

Now we state the main theorem of this section.

Theorem 36. *Let $(\mathcal{A}, \varphi, a)$ be a circular non-pushy* D0L*-system, $\mathcal{B}_L$ and $\mathcal{B}_R$ its L-forky and R-forky set, respectively, and let $\mathbf{u} = \varphi^\omega(a)$ be infinite. There exists a finite set $\mathcal{I}$ of bispecial triplets such that for any bispecial factor v there exist a bispecial triplet $\mathcal{T} \in \mathcal{I}$ and $n \in \mathbb{N}$ such that $((w_1, w_2), v, (w_3, w_4)) = (f_{\mathcal{B}})^n(\mathcal{T})$ for some $(w_1, w_2) \in \mathcal{B}_L$ and $(w_3, w_4) \in \mathcal{B}_R$.*

Proof. Let $\mathcal{I}$ be the set of initial bispecial triplets. The finiteness of $\mathcal{I}$ is a direct consequence of the definition of circularity: elements of $\mathcal{I}$ cannot be longer than the synchronizing delay. The rest of the statement follows from the fact that any non-initial triplet has at least one $f_{\mathcal{B}}$-preimage.

To prove this, it suffices to realize that the synchronized factor of $((w_1, w_2), v, (w_3, w_4))$ has a unique φ-preimage v' (possibly the empty word) and that the non-synchronized prefix (resp. suffix) must be equal to $f_L(w_1', w_2')$ for some $(w_1', w_2') \in \mathcal{B}_L$ (resp. to $f_R(w_3', w_4')$ for some $(w_3', w_4') \in \mathcal{B}_R$). □

Example 37. We now find the set $\mathcal{I}$ for $\mathbf{u}_S$ with $\mathcal{B} = (\mathcal{B}_L, \mathcal{B}_R)$, where $\mathcal{B}_L$ and $\mathcal{B}_R$ are from Example 21, the graphs of prolongations are in Fig. 3. The only BS factors without synchronizing points are ϵ and 0 since any other factor contains the letter 2 (and hence it has a synchronizing point) or is not BS. It remains to find all corresponding triplets:

$$\begin{array}{cccc} ((0, 01), \epsilon, (1, 2)), & ((0, 01), \epsilon, (0, 2)), & ((0, 012), \epsilon, (0, 1)), & ((0, 012), \epsilon, (0, 2)), \\ ((0, 012), \epsilon, (1, 2)), & ((0, 22), \epsilon, (0, 1)), & ((2, 01), \epsilon, (0, 2)), & ((0, 012), 0, (0, 1)). \end{array}$$

Since we are usually interested in nonempty BS factors, we can replace the bispecial triplets containing ϵ with their $f_{\mathcal{B}}$-images and get:

$$\begin{array}{ccc} ((2, 01), 2, (0, 2)), & ((2, 01), 20, (0, 1)), & ((0, 22), 012, (0, 2)), \\ ((0, 22), 0120, (0, 1)), & ((0, 012), 012, (0, 2)), & ((0, 012), 1, (0, 1)). \end{array}$$

There are only 6 bispecial triplets since $((0, 012), \epsilon, (0, 1))$ and $((0, 012), \epsilon, (1, 2))$ have the same $f_{\mathcal{B}}$-image and so do $((0, 01), \epsilon, (0, 2))$ and $((2, 01), \epsilon, (0, 2))$.

5. Infinite special branches

In the preceding section we have described a tool allowing us to find all BS factors. It requires some effort to construct the graphs GL and GR and to find all initial bispecial triplets, but it can be done by an algorithm. However, even if we have all these necessities in hand, it may be still a long way to the complete knowledge of the structure of all BS factors. Nevertheless, there is a class of special factors which can be identified directly from the graphs GL and GL, namely, the prefixes (or suffixes) of the so-called infinite LS (or RS) branches.

Definition 38. An infinite word $\mathbf{w}$ is an *infinite LS branch* of an infinite word $\mathbf{u}$ if each prefix of $\mathbf{w}$ is a LS factor of $\mathbf{u}$. We put

$$\mathrm{Lext}(\mathbf{w}) = \bigcap_{v \text{ prefix of } \mathbf{w}} \mathrm{Lext}(v).$$

Infinite RS branches are defined in the same manner, only that they are infinite to the right.

Here are some (almost) obvious statements on infinite special branches in an infinite word:

Proposition 39. *Let $\mathbf{u}$ be an infinite word.*

(i) *If $\mathbf{u}$ is eventually periodic, then there is no infinite LS branch of $\mathbf{u}$,*
(ii) *if $\mathbf{u}$ is aperiodic, then there exists at least one infinite LS branch of $\mathbf{u}$,*
(iii) *if $\mathbf{u}$ is a fixed point of a primitive morphism, then the number of infinite LS branches is bounded.*

Proof. Item (i) is obvious, (iii) is a direct consequence of the fact that the first difference of complexity is bounded [17]. The proof of item (ii) is due to the famous König's infinity lemma [18] applied on sets $V_1, V_2, \ldots$, where the set V_k comprises all LS factors of length k and where $v_1 \in V_i$ is connected by an edge with $v_2 \in V_{i+1}$ if v_1 is prefix of v_2. $\square$

Imagine now that we have an L-forky set $\mathcal{B}_L$ and an infinite LS branch $\mathbf{w}$. There must exist $(v_1, v_2) \in \mathcal{B}_L$ such that $v_1 w$ and $v_2 w$ are factors for any prefix w of $\mathbf{w}$. Such a pair is called an infinite LS pair.

Definition 40. Let (v_1, v_2) be an element of an L-forky set corresponding to a fixed point $\mathbf{u}$ of a morphism φ. The ordered pair $((v_1, v_2), \mathbf{w})$ is called an *infinite LS pair* if for any prefix w of $\mathbf{w}$ the words $v_1 w$ and $v_2 w$ are factors of $\mathbf{u}$.

Further, we define the *$f_{\mathcal{B}_L}$-image* of an infinite LS pair $((v_1, v_2), \mathbf{w})$ as the infinite LS pair $((v_1', v_2'), \mathbf{w}')$, where $(v_1', v_2') = g_L(v_1, v_2)$ and $\mathbf{w}' = f_L(v_1, v_2)\varphi(\mathbf{w})$.

Having the $f_{\mathcal{B}_L}$-image of an infinite LS branch, we are again interested in its $f_{\mathcal{B}_L}$-preimage.

Lemma 41. *Let $(\mathcal{A}, \varphi, a)$, $a \in \mathcal{A}$, be a circular* D0L-*system with an infinite fixed point $\mathbf{u} = \varphi^\omega(a)$. If $\mathcal{B}_L$ be its L-forky set, then any infinite LS pair is the $f_{\mathcal{B}_L}$-image of a unique infinite LS pair.*

Proof. Let $((v_1, v_2), \mathbf{w})$ be an infinite LS pair and let D be a synchronizing delay of φ. Then any prefix of w of length at least D has the same left-most synchronizing point $(w_1, (w_1)^{-1}w)$. Since such w is LS, then w_1 must be a label of an edge in $\mathrm{GL}_\varphi^{(\mathcal{B}_L)}$ whose end-vertex is (v_1, v_2) and starting one in (v_1', v_2'). The infinite word $(w_1)^{-1}\mathbf{w}$ must have a unique φ-preimage $\mathbf{w}'$. $\square$

Since any infinite LS pair $((v_1, v_2), \mathbf{w})$ has an $f_{\mathcal{B}_L}$-preimage, the in-degree of the vertex (v_1, v_2) in the graph of left prolongations $\mathrm{GL}_\varphi^{(\mathcal{B}_L)}$ must be at least one.

Corollary 42. *Let $(\mathcal{A}, \varphi, a)$ be a circular* D0L-*system with a fixed point $\mathbf{u}$ and an L-forky set $\mathcal{B}_L$. If $((v_1, v_2), \mathbf{w})$ is an infinite LS pair then (v_1, v_2) is a vertex of a cycle in $\mathrm{GL}_\varphi^{(\mathcal{B}_L)}$.*

We know that the number of infinite LS pairs in a fixed point of a primitive morphism is finite (see Proposition 39); the following proposition says that this is true even if we weaken the assumption from primitive to circular and non-pushy. The proof of the proposition will, moreover, give us a simple method of how to find all these infinite LS pairs.

Theorem 43. *Let $(\mathcal{A}, \varphi, a)$ be a circular* D0L-*system such that $\varphi^\omega(a) = \mathbf{u}$ is an infinite fixed point. If there exists an L-forky set for this system, then there is only a finite number of infinite LS pairs.*

Proof. Denote the forky set by $\mathcal{B}_L$. Let $((v_1, v_2), \mathbf{w})$ be an infinite LS pair and let (v_1, v_2) be a vertex of a cycle in $\mathrm{GL}_\varphi^{(\mathcal{B}_L)}$. If we denote the length of the cycle by k, then it is labeled by words $f_L(v_1, v_2), f_L(g_L(v_1, v_2)), \ldots, f_L(g_L^{k-1}(v_1, v_2))$ where $f_L(v_1, v_2)$ is the label of the edge starting in (v_1, v_2) (see Fig. 4). We distinguish two cases:
(a) At least one of the labels of the cycle is not the empty word. Applying k times Lemma 41 we can find the infinite LS pair $((v_1, v_2), \mathbf{w}')$ such that w is the $(f_{\mathcal{B}_L})^k$-image of $((v_1, v_2), \mathbf{w}')$, i.e.,

$$\mathbf{w} = \underbrace{f_L(g_L^{k-1}(v_1, v_2)) \cdots \varphi^{k-2}(f_L(g_L(v_1, v_2))\varphi^{k-1}(f_L(v_1, v_2))}_{\text{denoted by } s} \varphi^k(\mathbf{w}') = s\varphi^k(\mathbf{w}').$$

Since $\mathbf{w}'$ can be expressed again as $\mathbf{w}' = s\varphi^k(\mathbf{w}'')$ for some infinite LS pair $((v_1, v_2), \mathbf{w}'')$, we have

$$\mathbf{w} = s\varphi^k(s)\varphi^{2k}(s)\varphi^{3k}(\mathbf{w}'').$$

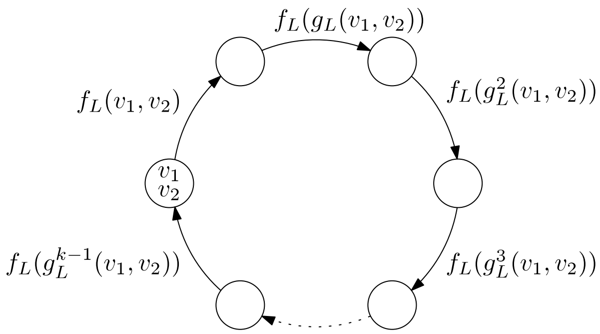

Fig. 4. The notation from the proof of Theorem 43.

Continuing in this construction one can prove that $s\varphi^k(s)\cdots\varphi^{nk}(s)$ is a prefix of $\mathbf{w}$ for all $n \in \mathbb{N}$. Therefore, we get

$$\mathbf{w} = s\varphi^k(s)\varphi^{2k}(s)\varphi^{3k}(s)\cdots.$$

We have just shown that exactly one infinite LS pair corresponds to each vertex of the cycle.

(b) Now assume that all the labels of the cycle are empty words. In such a case the $f_{\mathcal{B}_L}$-image coincides with the φ-image, meaning that the $(f_{\mathcal{B}_L})^j$-image of $((v_1, v_2), \mathbf{w})$ is $(g_L^j(v_1, v_2), \varphi^j(\mathbf{w}))$ for all $j = 1, 2, \cdots$. We want to prove that $\mathbf{w}$ must be a periodic point of φ. Consider the directed graph whose vertices are the first letters of $\varphi(b)$, $b \in \mathcal{A}$, and there is an edge from b to c if c is the first letter of $\varphi(b)$. Clearly, the first letter of $\mathbf{w}$, say b, must be again a vertex of a cycle in this graph. Let ℓ be the length of this cycle. For reasons analogous to those above the $(f_{\mathcal{B}_L})^{j\ell}$-image and $(f_{\mathcal{B}_L})^{j\ell}$-preimage of $\mathbf{w}$ must also begin in b. Therefore, $\mathbf{w}$ contains the factor $\varphi^{j\ell}(b)$ as a prefix for all $j = 1, 2, \ldots$ and this implies that $\mathbf{w} = (\varphi^\ell)^\omega(b)$, i.e., $\mathbf{w}$ is a periodic point of φ.

Since the number of vertices of $\mathrm{GL}_\varphi^{(\mathcal{B}_L)}$ and of periodic points is finite, the number of infinite LS pairs must be finite as well. $\square$

The previous proof is also a proof of the following corollary which gives us a method of how to find all infinite LS branches.

Corollary 44. *Let $(\mathcal{A}, \varphi, a)$, $a \in \mathcal{A}$, be a circular* D0L*-system, $\mathbf{u} = \varphi^\omega(a)$ infinite with $\mathcal{B}_L$ an L-forky set and let $((v_1, v_2), \mathbf{w})$ be an infinite LS pair. Then either $\mathbf{w}$ is a* periodic point *of φ, i.e.,*

$$\mathbf{w} = \varphi^\ell(\mathbf{w}) \quad \text{for some } \ell \geq 1, \tag{4}$$

and (v_1, v_2) is a vertex of a cycle in $\mathrm{GL}_\varphi^{(\mathcal{B}_L)}$ labeled by ϵ only, or $\mathbf{w} = s\varphi^\ell(s)\varphi^{2\ell}(s)\cdots$ is the unique solution of the equation

$$\mathbf{w} = s\varphi^\ell(\mathbf{w}), \tag{5}$$

where (v_1, v_2) is a vertex of a cycle in GL_φ containing at least one edge with a non-empty label, ℓ is the length of this cycle and

$$s = f_L(g_L^{\ell-1}(v_1, v_2))\cdots\varphi^{\ell-2}(f_L(g_L(v_1, v_2))\varphi^{\ell-1}(f_L(v_1, v_2)). \tag{6}$$

We demonstrate this method on an example morphism.

Example 45. We consider the morphism

$$\varphi_P : 1 \mapsto 1211,\ 2 \mapsto 311,\ 3 \mapsto 2412,\ 4 \mapsto 435,\ 5 \mapsto 534 \tag{7}$$

with $\mathbf{u} = \varphi_P^\omega(1)$. This morphism is suffix- and prefix-free and so the set of all unordered pairs of distinct letters is L-forky. The graph of left prolongations is in Fig. 5. The morphism φ_P has five periodic points

$$\varphi_P^\omega(1),\ \varphi_P^\omega(4),\ \varphi_P^\omega(5),\ (\varphi_P^2)^\omega(2),\ (\varphi_P^2)^\omega(3).$$

It is easy to show that

$$\mathrm{Lext}(1) = \{1, 2, 3, 4, 5\},\ \mathrm{Lext}(2) = \{1, 4, 5\},\ \mathrm{Lext}(3) = \{1, 4, 5\},$$
$$\mathrm{Lext}(4) = \{1, 2, 3\},\ \mathrm{Lext}(5) = \{1, 2, 3\}.$$

Looking at the graph of left prolongations depicted in Fig. 5, we see that $\varphi_P^\omega(4)$ and $\varphi_P^\omega(5)$ are not infinite LS branches as none of the vertices $(1, 2)$, $(2, 3)$ and $(1, 3)$ is a vertex of a cycle labeled by ϵ only. Hence, only $\varphi_P^\omega(1)$, $(\varphi_P^2)^\omega(2)$, $(\varphi_P^2)^\omega(3)$ are infinite LS branches with left extensions 1, 4, 5.

As for infinite LS branches corresponding to Eq. (5), in the case of our example, there is only one cycle which is not labeled by the empty word: the cycle between vertices $(1, 2)$ and $(2, 3)$. Since the length of the cycle determines the number of equations, there are two equations corresponding to this cycle, namely

$$\mathbf{w} = \varphi_P(11)\varphi_P^2(\mathbf{w}) \quad \text{and} \quad \mathbf{w} = 11\varphi_P^2(\mathbf{w}).$$

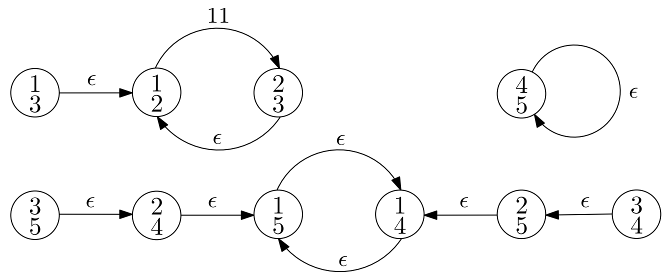

Fig. 5. The graph $GL_{\varphi_P}^{(\mathcal{B}_1)}$ for morphism defined by (7), $\mathcal{B}_1$ is the set of all unordered pairs of letters.

They give us two infinite LS branches

$$\varphi_P(11)\varphi_P^3(11)\varphi_P^5(11)\cdots,$$
$$11\varphi_P^2(11)\varphi_P^4(11)\cdots,$$

the former having left extensions 1 and 2 and the latter 2 and 3.

6. Assumptions and a connection with the critical exponent

The method of how to generate all BS factors of a given D0L-system we have described above bears on two facts: There exist L- and R-forky sets and the number of initial BS triplets is finite. The former was proved for circular and non-pushy systems and the latter for circular ones only. Are these assumptions necessary or can they be weakened?

Let us take a morphism $0 \mapsto 001$, $1 \mapsto 1$, which is pushy and circular, and its fixed point $\mathbf{u}$ starting in 0. If we try to construct an L-forky set as it has been defined in this paper, we will find out that it is not possible. A natural candidate for vertices of the graph of left prolongations are pairs $(0, 1^n)$ with $n \in \mathbb{N}$, but any finite set of such pairs does not satisfy the property (iv) of Definition 20. So it seems that to assume the morphism being non-pushy is inevitable for existence of forky sets. However, if we relax the definition and enable the pairs of factors to be infinitely long, we can find something like an L-forky set even for this morphism: Define a directed graph of left prolongations such that it has only one vertex $(0, \cdots 111)$ and one loop on this vertex with label 1 and a directed graph of right prolongations with one vertex $(0, 1)$ and a loop on it with empty label, then all BS factors in $\mathbf{u}$ are the f-images of BS-triplet $((0, \cdots 111), 0, (0, 1))$, namely $0, 1\varphi(0), 1\varphi(1)\varphi^2(0), \ldots$

Now, consider a morphism $0 \mapsto 001$, $1 \mapsto 11$ and its fixed point $\mathbf{u}$ starting in 0. This morphism is non-pushy and non-circular. In this case, L- and R-forky sets exists, we can simply take $\{(0, 1)\}$ and there is only one (nonempty) initial *BS*-triplet with no *BS-synchronizing* point $((0, 1), 0, (0, 1))$. Its f-images again reed $0, 1\varphi(0), 1\varphi(1)\varphi^2(0), \ldots$. In fact, to prove that the set of initial BS-triplets is finite, we need to know only that there is not infinite number of BS-triplets without BS-synchronizing point and it seems to be true even for non-circular morphisms.

All considered morphisms for which our method does not work (or is not proved to work) have an infinite critical exponent. The following theorem says this is not a misleading observation but a general rule.

Theorem 46. *Let $G = (\mathcal{A}, \varphi, w)$ be a* D0L*-system. Then the critical exponent of $\mathcal{L}(G)$ is finite if and only if G is circular and non-pushy.*

Proof. ($\Rightarrow$): Circularity follows from Theorem 8. G being pushy is in contradiction with Lemma 14, thus, it is non-pushy.
($\Leftarrow$): Suppose the critical exponent of $\mathcal{L}(G)$ is infinite and that G is circular and non-pushy. According to Theorem 9, there exists a non-empty factor $v \in \mathcal{L}(G)$ such that for all $n \in \mathbb{N}$, $v^n \in \mathcal{L}(G)$. Take the shortest factor v having such a property. Since G is circular, there exists a finite synchronizing delay D. Take $N \in \mathbb{N}$ such that $|v^N| \geq D$. Then v^N contains a synchronizing point, i.e., $v^N = v_1|_s v_2$. It is clear that v^{N+1} contains at least two synchronizing points, i.e., $v^{N+1} = v_1|_s v_2 v = v v_1|_s v_2$. In general, v^{N+k} contains $k+1$ synchronizing points at fixed distances equal to $|v|$. Since φ is injective, it implies that there exists a unique $z \in \mathcal{L}(G)$ such that $v^{N+k} = p\varphi(z^k)s$ (for some factors p and s) and $z^k \in \mathcal{L}(G)$ for all $k \geq 0$. According to the choice of v, it is clear that $|\varphi(z)| = |z| = |v|$. Denote by $\mathcal{L}_1(z)$ the set of letters occurring in z. It is clear that $\varphi(\mathcal{L}_1(z)) = \mathcal{L}_1(v)$ and $\forall a \in \mathcal{L}_1(z)$ we have $|\varphi(a)| = 1$.

We can now repeat the process: take the factor z to play the role of factor v. Thus, we can find an infinite sequence of factors $z_0 = z, z_1, z_2 \ldots$ such that $\varphi(\mathcal{L}_1(z_{k+1})) = \mathcal{L}_1(z_k)$ and $|z_k| = |z|$ for all $k \geq 0$. Since $\mathcal{A}$ is finite, it is clear that there exists integers $m \neq \ell$ such that $\mathcal{L}_1(z_m) = \mathcal{L}_1(z_\ell)$. This implies that for all k the factor z_k is composed of letters of rank zero. This is a contradiction with G being non-pushy. □

7. Conclusion

The tool we have introduced in this paper enables to construct an algorithm which can find all BS factors in a given circular non-pushy D0L-system so that it produces the graphs of prolongations and the set of initial BS factors—its slightly simplified version was implemented by Štěpán Starosta using the open source mathematical software SAGE [19]. The sketch of the algorithm is as follows:

1. Decide whether the input D0L-system is strongly repetitive using the algorithm from [14]. If it is, then by the previous theorem and Theorem 9 the D0L-system is non-circular or pushy and our method does not work. If it is not, proceed with the next two steps.
2. Construct L- and R-forky sets: the details of the construction are a bit technical but the basic idea is the same we used in Example 21.
3. Find all initial BS triplets without any BS-synchronizing point. The fact that the system is circular ensures the algorithm stops after a finite number of steps.

Acknowledgments

We acknowledge financial support by the Czech Science Foundation grant 201/09/0584, and by the grants MSM6840770039 and LC06002 of the Ministry of Education, Youth, and Sports of the Czech Republic. The author also thanks Štěpán Starosta and Edita Pelantová for their suggestions and the fruitful discussions.

References

[1] J. Cassaigne, An algorithm to test if a given circular HD0L-language avoids a pattern, in: IFIP World Computer Congress'94, North-Holland, 1994, pp. 459–464.
[2] S.V. Avgustinovich, A.E. Frid, On bispecial words and subword complexity of D0L sequences, in: Proceedings of SETA'98, in: DMTCS Series, Springer, 1999, pp. 191–204.
[3] A. Frid, On uniform D0L words, in: STACS 98, in: Lect. Notes Comp. Sci., vol. 1373, Springer, Berlin, Heidelberg, 1998, pp. 544–554.
[4] A. De Luca, S. Varricchio, On the factors of the Thue-Morse word on three symbols, Inform. Process. Lett. 27 (6) (1988) 281–285. http://dx.doi.org/10.1016/0020-0190(88)90214-1.
[5] J. Cassaigne, Complexité et facteurs spéciaux, Bull. Belg. Math. Soc. Simon Stevin 4 (1997) 67–88.
[6] G. Rozenberg, A. Salomaa (Eds.), The Book of L, Springer-Verlag New York, Inc., New York, NY, USA, 1986.
[7] G. Rozenberg, A. Salomaa, Mathematical Theory of L Systems, Academic Press, Inc., Orlando, FL, USA, 1980.
[8] A. Ehrenfeucht, G. Rozenberg, Elementary homomorphisms and a solution of the D0L sequence equivalence problem, Theoret. Comput. Sci. 7 (2) (1978) 169–183. http://dx.doi.org/10.1016/0304-3975(78)90047-6.
[9] P. Séébold, Sequences generated by infinitely iterated morphisms, Discrete Appl. Math. 11 (3) (1985) 255–264. http://dx.doi.org/10.1016/0166-218X(85)90077-0.
[10] D. Krieger, On critical exponents in fixed points of non-erasing morphisms, Theoret. Comput. Sci. 376 (1–2) (2007) 70–88. http://dx.doi.org/10.1016/j.tcs.2007.01.020.
[11] L. Balková, K. Klouda, E. Pelantová, Repetitions in beta-integers, Lett. Math. Phys. 87 (2009) 181–195.
[12] L. Balková, K. Klouda, E. Pelantová, Critical exponent of infinite words associated with non-simple Parry numbers, Integers 11B.
[13] F. Mignosi, P. Séébold, If a D0L language is k-power free then it is circular, in: ICALP'93: Proceedings of the 20th International Colloquium on Automata, Languages and Programming, Springer-Verlag, London, UK, 1993, pp. 507–518.
[14] A. Ehrenfeucht, G. Rozenberg, Repetitions of subwords in D0L languages, Inform. Comput. 59 (1983) 13–35.
[15] B. Mossé, Notions de reconnaissabilité pour les substitutions et complexité des suites automatiques, Bull. Soc. Math. France 124 (1996) 101–108.
[16] R.V. Chacon, Weakly mixing transformations which are not strongly mixing, Proc. Amer. Math. Soc. 22 (3) (1969) 559–562.
[17] J. Cassaigne, Special factors of sequences with linear subword complexity, in: Developments in Language Theory, World Scientific, 1996, pp. 25–34.
[18] D. König, Theorie der endlichen und unendlichen Graphen, Akademische Verlagsgesellschaft, Leipzig, 1936.
[19] W. Stein, et al. Sage Mathematics Software (Version 4.7.2), The Sage Development Team, 2011. http://www.sagemath.org.