Recognizable morphisms

Revekka KYRIAKOGLOU

February 20, 2018

UNIVERSITÉ
PARIS-EST
MARNE-LA-VALLÉE

Gaspard - Monge
Laboratoire d'informatique

Outline

Definition

A morphism $\varphi : A^* \to A^*$ is said to be primitive if there is an integer $n \geq 1$ such that for any $a, b \in A$, a occurs in $\varphi^n(b)$. A morphism φ is called non-erasing if $\varphi(a)$ is a non-empty word for every $a \in A$.

Definition

If $a \in A$ is such that the word $\varphi(a)$ begins with the letter a and if $|\varphi^n(a)|$ tends to infinity with n, there is a unique infinite word denoted $\varphi^\omega(a) = \lim_{n\to\infty} f^n(a)$ which has all words $\varphi^n(a)$ as prefixes. This word is called a fixed point of the morphism φ.

- The language of a fixed point $\mathbf{u}$ is the set of all its factors and we symbolize it as $F(\mathbf{u})$. This language is called substitutive.

Example (Fibonacci word)

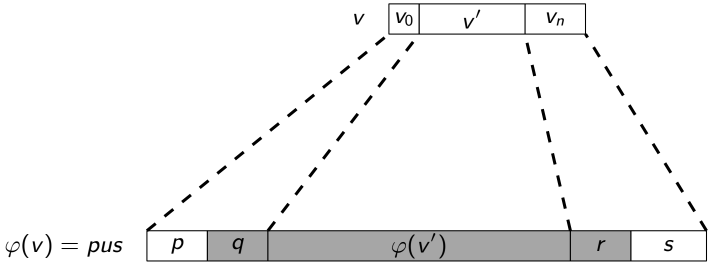
$A = \{a, b\}$

$$\begin{aligned} \varphi : A^* &\to A^*, \\ a &\mapsto ab, \\ b &\mapsto a. \end{aligned}$$

The Fibonacci word is the fixed point of φ, i.e.
$\varphi^\omega(a) = abaababaabaababaababaaba\ldots$.

Definition

Let $\varphi : A \to A^*$ be a non-erasing substitution and $\mathbf{u} \in A^{\mathbb{N}}$ its φ-fixed point. A triplet (p, v, s) where $p, s \in A^*$, $v = v_0 v' v_n \in A^+$, $v' = v_1 \dots v_{n-1}$, $(v_i)_{0 \leq i \leq n} \in A$ is an external interpretation of a word $u \in F(\mathbf{u})$ if $\varphi(v) = pus$, where p is prefix of $\varphi(v_0)$ and q is suffix of $\varphi(v_n)$, while the triplet (q, v', r) is an internal interpretation of the word u.



Definition

A morphism φ is called prefix (resp. suffix) if φ is non-erasing and $\forall a, b \in A$ if $\varphi(a)$ is prefix (resp. suffix) of $\varphi(b)$, then $\varphi(a) = \varphi(b)$. A morphism φ is called bifix if φ is simultaneously prefix and suffix.

Definition

A morphism φ is called prefix (resp. suffix) if φ is non-erasing and $\forall a, b \in A$ if $\varphi(a)$ is prefix (resp. suffix) of $\varphi(b)$, then $\varphi(a) = \varphi(b)$. A morphism φ is called bifix if φ is simultaneously prefix and suffix.

Example

The Fibonacci morphism φ_F is suffix but not prefix since $\varphi_F(1)$ is a prefix of $\varphi_F(0)$.

Definition

Let $w \in A^*$. We say that (w_1, w_2) is a synchronization point of w for the morphism φ, if $w = w_1 w_2$ and

$$\forall v_1 \in A^*, \forall v_2 \in A^*, v_1 w v_2 \in \varphi(A^*) \Rightarrow v_1 w_1 \in \varphi(A^*) \text{ and } w_2 v_2 \in \varphi(A)^*$$

and we denote this fact by $w = w_1 | w_2$.

Definition

Let $w \in A^*$. We say that (w_1, w_2) is a synchronization point of w for the morphism φ, if $w = w_1 w_2$ and

$$\forall v_1 \in A^*, \forall v_2 \in A^*, v_1 w v_2 \in \varphi(A^*) \Rightarrow v_1 w_1 \in \varphi(A^*) \text{ and } w_2 v_2 \in \varphi(A)^*$$

and we denote this fact by $w = w_1 | w_2$.

Definition

Let $w \in A^*$. We say that (w_1, w_2) is a stroke of w for the morphism φ, if $w = w_1 w_2$ and

$$\exists v_1, v_2 \in A^*, v_1 w v_2 \in \varphi(A^*) \text{ such that } v_1 w_1 \in \varphi(A^*) \text{ and } w_2 v_2 \in \varphi(A)^*.$$

Definition

Let $w \in A^*$. We say that (w_1, w_2) is a synchronization point of w for the morphism φ, if $w = w_1 w_2$ and

$$\forall v_1 \in A^*, \forall v_2 \in A^*, v_1 w v_2 \in \varphi(A^*) \Rightarrow v_1 w_1 \in \varphi(A^*) \text{ and } w_2 v_2 \in \varphi(A)^*$$

and we denote this fact by $w = w_1 | w_2$.

Definition

Let $w \in A^*$. We say that (w_1, w_2) is a stroke of w for the morphism φ, if $w = w_1 w_2$ and

$$\exists v_1, v_2 \in A^*, v_1 w v_2 \in \varphi(A^*) \text{ such that } v_1 w_1 \in \varphi(A^*) \text{ and } w_2 v_2 \in \varphi(A)^*.$$

Example

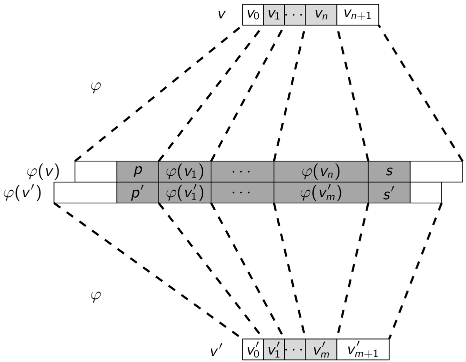
Let φ_M be the Thue-Morse morphism with $\varphi_M(a) = ab$, $\varphi_M(b) = ba$ and fixed point $\mathbf{u} = \varphi_M^\omega(a)$. We know that the word $aba \in F(\mathbf{u})$ contains one stroke (either at position 1 or 2), but none of them is a synchronization point.

Outline

1. Basic definitions
2. The many different definitions of recognizability
3. Results on the non primitive case

Definition (Martin)

Let $\varphi : A \to A^*$ be a primitive substitution with a fixed point $\mathbf{u}$. Let $u \in F(\mathbf{u})$ such that $|u| > |\varphi(a)|, \forall a \in A$. We say that u is determined if there is only one internal interpretation (q, v, r) of u.



Example (Thue-Morse morphism)

Let $\varphi_M : \{a, b\} \to \{a, b\}^*$ with $\varphi_M(a) = ab$, $\varphi_M(b) = ba$ and fixed point $\mathbf{u} = \varphi_M^\omega(a)$.
All $u \in F(\mathbf{u})$, $|u| \geq 5$ are determined:
All such words have as factor the word *aa* or *bb*, otherwise they should have been factors of $(ab)^*$.
Having as factor *aa* or *bb* means that there is a cutting point between the *a*'s and the *b*'s, since there is no letter in the alphabet that under the morphism φ_M has an image with factor *aa* or *bb*.
Also, since φ_M is bifix, if $u \in F(\mathbf{u})$ is large enough ($|u| \geq 5$) in order to have a cutting point, the decoding of the word is unique and consequently the internal interpretation of u is also unique.

Martin states the following (Lemma 1):
If φ is primitive, there exists integer t so that any factor of length t of $\mathbf{u}$ is determined.

Martin states the following (Lemma 1):
If φ is primitive, there exists integer t so that any factor of length t of $\mathbf{u}$ is determined.

Example

Let φ_F be the Fibonacci morphism that is primitive. The prefixes of the Fibonacci word are left-special and $F(\mathbf{u})$ is closed by reversal. Thus, the reversal of any prefix of the Fibonacci word is right special and there are right special words of every length.
Let $u \in F(\mathbf{u})$ be right special, then it can be extended on the right by a and b. Also, since all the words in the language $F(\mathbf{u})$ cannot have two consecutive b's, the last letter of u has to be a. Similarly, $u = wba$ since otherwise it would had three consecutive a's (that implies that the preimage of this factor has two consecutive b's).
following the same preocedure we have that $u = w'aba$ that it has two different internal interpretations (w', ab, ϵ) and (w', a, a).

Definition

Let $\varphi : A \to A^*$ be a morphism with a fixed point $\mathbf{u}$. We say that φ is strongly circular if there is integer t such that any factor of $\mathbf{u}$ of length more than t is determined.

Definition

Let $\varphi : A \to A^*$ be a morphism with a fixed point $\mathbf{u}$. We say that φ is strongly circular if there is integer t such that any factor of $\mathbf{u}$ of length more than t is determined.

Example

The Morse morphism φ_M is strongly-circular for $t = 5$.
The Fibonacci morphism φ_F is not strongly-circular.

Definition (B. Host)

Let $\varphi : A \to A^*$ be a primitive substitution with a fixed point $\mathbf{u}$. We say that φ is right recognizable if there is an integer $L > 0$ such that if $\mathbf{u}_{[i,i+L]} = \mathbf{u}_{[j,j+L]}$ and $i \in E$, then $j \in E$.

Similarly we define the left recognizability.
Let us define the set E_k as following,

$$E_k = \{0\} \cup \{|\varphi^k(\mathbf{u}_{[0,p-1]})|; p > 0\}.$$

We are going to use the notation E instead of E_1.

Definition (B. Host)

Let $\varphi : A \to A^*$ be a primitive substitution with a fixed point $\mathbf{u}$. We say that φ is right recognizable if there is an integer $L > 0$ such that if $\mathbf{u}_{[i,i+L]} = \mathbf{u}_{[j,j+L]}$ and $i \in E$, then $j \in E$.

Similarly we define the left recognizability.
Let us define the set E_k as following,

$$E_k = \{0\} \cup \{|\varphi^k(\mathbf{u}_{[0,p-1]})|; p > 0\}.$$

We are going to use the notation E instead of E_1.

Example

The Fibonacci morphism φ_F is right recognizable with $L = 1$.

$$F(\mathbf{u}) \cap A^2 = \{00, 01, 10\}$$

For the words $01, 00$ we know that the letter 0 is always the beginning of an image of a letter, hence $i \in E$.
For the word 10 we are sure that $i \notin E$ for all i such that $\mathbf{u}_{[i,i+1]} = 10$.

Definition (Mossé)

Let $\varphi : A \to A^*$ be a substitution with a fixed point $\mathbf{u}$. We say that φ is two-sided recognizable if there is an integer $L > 0$ such that if $\mathbf{u}_{[i-L,i+L]} = \mathbf{u}_{[j-L,j+L]}$ and $i \in E_1$, then $j \in E_1$.

Definition (Mossé)

Let $\varphi : A \to A^*$ be a substitution with a fixed point $\mathbf{u}$. We say that φ is two-sided recognizable if there is an integer $L > 0$ such that if $\mathbf{u}_{[i-L,i+L]} = \mathbf{u}_{[j-L,j+L]}$ and $i \in E_1$, then $j \in E_1$.

Example (Fibonacci morphism)

Every $w \in F(\mathbf{u})$ cannot have as factor two consecutive 1's nor more than two consecutive 0's.
For a factor 00, there is a synchronization point between the two 0's and after each appearence of 1 there is a synchronization point.
Any $|w| > 3$ ($L = 1$) has at least one synchronization point (cause it includes 1's or 00) and is uniquely decomposed. Thus, it is two-sided recognizable.

Definition (Mossé)

Let $\varphi : A \to A^*$ be a substitution with a fixed point $\mathbf{u}$. We say that φ is two-sided recognizable if there is an integer $L > 0$ such that if $\mathbf{u}_{[i-L,i+L]} = \mathbf{u}_{[j-L,j+L]}$ and $i \in E_1$, then $j \in E_1$.

Example (Fibonacci morphism)

Every $w \in F(\mathbf{u})$ cannot have as factor two consecutive 1's nor more than two consecutive 0's.
For a factor 00, there is a synchronization point between the two 0's and after each appearence of 1 there is a synchronization point.
Any $|w| > 3$ ($L = 1$) has at least one synchronization point (cause it includes 1's or 00) and is uniquely decomposed. Thus, it is two-sided recognizable.

Theorem (Mossé)

Every primitive substitution that admits an aperiodic fixed point is two-sided recognizable.

The one-sided recognizability is stronger than the two-sided.

Example (Fibonacci)

From previous example we have that this morphism is two-sided recognizable. However, this morphism is not left recognizable.
The word 0 is right special since it can be extended on the left with 0 and 1. We know that the image of it extended by the longest common prefix of the images of its extensions, $\varphi(0)0$, is also bispecial, etc. Hence, it cannot be left recognizable despite the fact that it is two-sided recognizable.

Proposition

Every suffix morphism that is two-sided recognizable is also right recognizable.

Similarly, every prefix morphism that is two-sided recognizable is also left recognizable.

Example (Fibonacci)

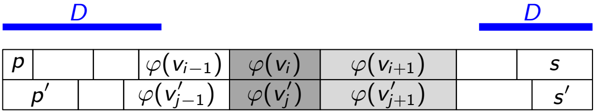
The morphism φ_F is suffix. Also, φ_F is two-sided recognizable. With use of the above proposition we expect φ_F to be right recognizable.

Definition (F. Mignosi and P. Séébold)

We say that φ has synchronizing delay $D > 0$ if $\forall u \in F(\mathbf{u})$ such that,

$$u = p\varphi(v_1)\ldots\varphi(v_n)s = p'\varphi(v'_1)\ldots\varphi(v'_m)s' \text{ where } v_i, v'_i \in A,$$

whenever $|p\varphi(v_1)\ldots\varphi(v_{i-1})| > D$ and $|\varphi(v_{i+1})\ldots\varphi(v_n)s| > D$, there exists a number $j \in \mathbb{N}$ such that $p\varphi(v_1)\ldots\varphi(v_i) = p'\varphi(v'_1)\ldots\varphi(v'_j)$ and $v_i = v'_j$.

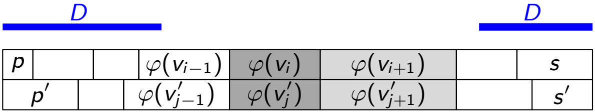


Definition (F. Mignosi and P. Séébold)

We say that φ has synchronizing delay $D > 0$ if $\forall u \in F(\mathbf{u})$ such that,

$$u = p\varphi(v_1)\dots\varphi(v_n)s = p'\varphi(v_1')\dots\varphi(v_m')s' \text{ where } v_i, v_i' \in A,$$

whenever $|p\varphi(v_1)\dots\varphi(v_{i-1})| > D$ and $|\varphi(v_{i+1})\dots\varphi(v_n)s| > D$, there exists a number $j \in \mathbb{N}$ such that $p\varphi(v_1)\dots\varphi(v_i) = p'\varphi(v_1')\dots\varphi(v_j')$ and $v_i = v_j'$.



Proposition

If a morphism φ is bifix and has finite synchronizing delay then it is strongly circular.

Definition (P. Kůrka)

Let $\varphi : A \to A^*$ be substitution with a φ-fixed point $\mathbf{u}$. Define $f : \mathbb{N} \to \mathbb{N}$ by $f(n) = |\varphi(\mathbf{u}_{[0,n)})|$. Then φ is strongly two-sided recognizable if there exists a context length $l > 0$ such that for any $u \in F(\mathbf{u})$ of length at least $2l$, there exist i, j with $0 \leq i \leq l$, $|u| - l \leq j \leq |u|$ and unique $v \in F(\mathbf{u})$, such that $\mathbf{u}_{[i,j)} = \varphi(v)$ and whenever $\mathbf{u}_{[m,m+|u|)} = u$, then there exist i', j' with $f(i') = m + i$, $f(j') = m + j$ and $\mathbf{u}_{[i',j')} = v$.

The constant l is going to be called constant of recognizability of φ.

Definition (P. Kůrka)

Let $\varphi : A \to A^*$ be substitution with a φ-fixed point $\mathbf{u}$. Define $f : \mathbb{N} \to \mathbb{N}$ by $f(n) = |\varphi(\mathbf{u}_{[0,n)})|$. Then φ is strongly two-sided recognizable if there exists a context length $l > 0$ such that for any $u \in F(\mathbf{u})$ of length at least $2l$, there exist i, j with $0 \leq i \leq l$, $|u| - l \leq j \leq |u|$ and unique $v \in F(\mathbf{u})$, such that $\mathbf{u}_{[i,j)} = \varphi(v)$ and whenever $\mathbf{u}_{[m,m+|u|)} = u$, then there exist i', j' with $f(i') = m + i$, $f(j') = m + j$ and $\mathbf{u}_{[i',j')} = v$.

The constant l is going to be called constant of recognizability of φ.

Example (Fibonacci morphism)

Let φ_F with fixed point $\mathbf{u}$. This morphism is strongly two-sided recognizable with constant of recognizability $l = 2$.

$$F(\mathbf{u}) \cap A^4 = \{0010, 0100, 0101, 1001, 1010\}$$

They can be written as $0\varphi_F(0)0$, $01\varphi_F(1)0$, $\varepsilon\varphi_F(0)01$, $1\varphi_F(1)01$, $1\varphi_F(0)0$, no matter their position in $\mathbf{u}$.

Definition (P. Kůrka)

Let $\varphi : A \to A^*$ be substitution with a φ-fixed point $\mathbf{u}$. Define $f : \mathbb{N} \to \mathbb{N}$ by $f(n) = |\varphi(\mathbf{u}_{[0,n)})|$. Then φ is strongly two-sided recognizable if there exists a context length $l > 0$ such that for any $u \in F(\mathbf{u})$ of length at least $2l$, there exist i, j with $0 \leq i \leq l$, $|u| - l \leq j \leq |u|$ and unique $v \in F(\mathbf{u})$, such that $\mathbf{u}_{[i,j)} = \varphi(v)$ and whenever $\mathbf{u}_{[m,m+|u|)} = u$, then there exist i', j' with $f(i') = m + i$, $f(j') = m + j$ and $\mathbf{u}_{[i',j')} = v$.

The constant l is going to be called constant of recognizability of φ.

Example (Fibonacci morphism)

Let φ_F with fixed point $\mathbf{u}$. This morphism is strongly two-sided recognizable with constant of recognizability $l = 2$.

$$F(\mathbf{u}) \cap A^4 = \{0010, 0100, 0101, 1001, 1010\}$$

They can be written as $0\varphi_F(0)0$, $01\varphi_F(1)0$, $\varepsilon\varphi_F(0)01$, $1\varphi_F(1)01$, $1\varphi_F(0)0$, no matter their position in $\mathbf{u}$.

Theorem (Mossé)

Let $\varphi : A \to A^*$ be a primitive substitution with a non-periodic fixed point $\mathbf{u}$. Then the substitution φ is strongly two-sided recognizable.

Definition

Let $\varphi : A \to A^*$ be a non-erasing substitution and $\mathbf{u} \in A^{\mathbb{N}}$ its φ-fixed point $\mathbf{u}$. We say that two external interpretations (p, v, s) and (p', v', s') of a word $u \in F(\mathbf{u})$ are synchronized at position k if there exist nonnegative indices i and j such that

$$\varphi(v_1 \dots v_i) = pu_1 \dots u_k \text{ and } \varphi(v'_1 \dots v'_j) = p' u_1 \dots u_k$$

with $v = v_1 \dots v_n$, $v' = v'_1 \dots v'_m$ and $u = u_1 \dots u_l$.
We say that a word $u \in F(\mathbf{u})$ has a synchronizing point at position k with $0 \leq k \leq |u|$ if all its external interpretations are pairwise synchronized at position k.

Definition

Let $\varphi : A \to A^*$ be a non-erasing substitution and $\mathbf{u} \in A^{\mathbb{N}}$ its φ-fixed point $\mathbf{u}$. We say that two external interpretations (p, v, s) and (p', v', s') of a word $u \in F(\mathbf{u})$ are synchronized at position k if there exist nonnegative indices i and j such that

$$\varphi(v_1 \dots v_i) = pu_1 \dots u_k \text{ and } \varphi(v'_1 \dots v'_j) = p' u_1 \dots u_k$$

with $v = v_1 \dots v_n$, $v' = v'_1 \dots v'_m$ and $u = u_1 \dots u_l$.
We say that a word $u \in F(\mathbf{u})$ has a synchronizing point at position k with $0 \leq k \leq |u|$ if all its external interpretations are pairwise synchronized at position k.

Definition (K. Klouda and Š. Starosta)

Let $\varphi : A \to A^*$ be a non-erasing substitution and $\mathbf{u} \in A^{\mathbb{N}}$ its φ-fixed point. The morphism φ is called weakly circular if there is a constant $D > 0$ such that any v that belongs in the language $F(\mathbf{u})$ longer than $2D$ has a synchronizing point.

Definition (K. Klouda and Š Starosta)

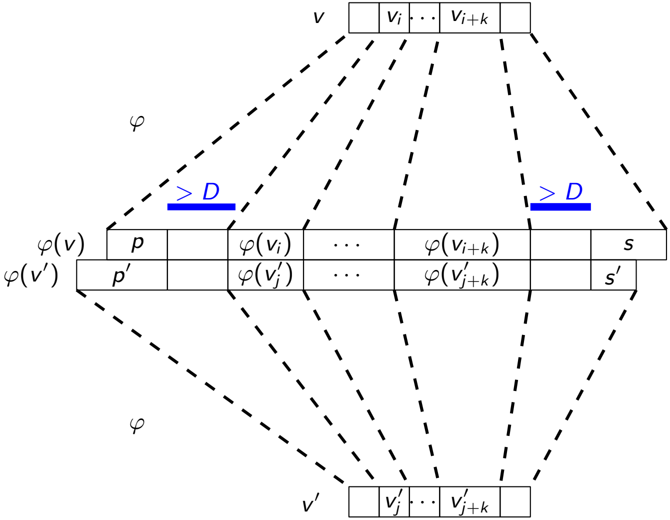
Let $\varphi : A \to A^*$ be a non-erasing substitution and $\mathbf{u} \in A^{\mathbb{N}}$ its φ-fixed point.
Let (p, v, s) and (p', v', s') be two external interpretations of $u \in F(\mathbf{u}) \setminus \epsilon$.
We say that φ is circular with delay $D > 0$ if whenever

$$|\varphi(v_1 \ldots v_i)| - |p| > D \text{ and } |\varphi(v_{i+1} \ldots v_n)| - |s| > D$$

for some $1 \leq i \leq n$, then there is $1 \leq j \leq m$ such that

$$|\varphi(v_1 \ldots v_{i-1})| - |p| = |\varphi(v'_1 \ldots v'_{j-1})| - |p'|$$

and $v_i = v'_j$.


v
v_i
$\cdots$
v_{i+k}
φ
$> D$
$> D$
$\varphi(v)$
p
$\varphi(v_i)$
$\cdots$
$\varphi(v_{i+k})$
s
$\varphi(v')$
p'
$\varphi(v'_j)$
$\cdots$
$\varphi(v'_{j+k})$
s'
φ
v'
v'_j
$\cdots$
v'_{j+k}

Circularity implies weak circularity but the converse is not always true.

Proposition

If $\varphi : A \to A^*$ is an injective substitution that is weakly circular then it is also circular.

Circularity implies weak circularity but the converse is not always true.

Proposition

If $\varphi : A \to A^*$ *is an injective substitution that is weakly circular then it is also circular.*

The above is not true for the non-injective case.

Circularity implies weak circularity but the converse is not always true.

Proposition

If $\varphi : A \to A^*$ is an injective substitution that is weakly circular then it is also circular.

The above is not true for the non-injective case.

Example

Let us suppose the non-injective morphism $\varphi : \{0,1,2\} \to \{0,1,2\}^*$ with $\varphi(0) = 0120$, $\varphi(1) = 12$, $\varphi(2) = 12$ and fixed point $\mathbf{u} = \varphi^\omega(0)$. The language $F(\mathbf{u})$ is not circular since for every $m \in \mathbb{N}$ the word $(12)^{2m}$ has two different preimages $(12)^m$ and $(21)^m$. The corresponding external interpretations however, have synchronizing points for $m > 1$ at the positions $2k$ for $0 \leq k \leq m$. Moreover, the language $F(\mathbf{u})$ is weakly circular since every $v \in F(\mathbf{u})$ with length more than $D = 3$ has a synchronization point.

Theorem

Let $\varphi : A \to A^*$ be a primitive substitution with a φ-fixed point $\mathbf{u}$ which is not periodic. Then, the morphism φ is,

1. *two-sided recognizable,*
2. *circular,*
3. *weakly circular.*

Outline

Theorem

Let $\varphi : A \to A^*$ *be a substitution with* φ*-fixed point* $\mathbf{u}$*. The substitution* φ *is circular if and only if it is strongly two-sided recognizable.*

Theorem

Let $\varphi : A \rightarrow A^*$ be a substitution with φ-fixed point $\mathbf{u}$. If the substitution φ is two-sided recognizable then it is also weakly circular.

Theorem

Let $\varphi : A \to A^*$ be a substitution with φ-fixed point $\mathbf{u}$. If the substitution φ is two-sided recognizable then it is also weakly circular.

If the morphism φ is injective, weakly circularity is equivalent to two-sided recognizability, this is not true for the non-injective case as shown by the following example.

Theorem

Let $\varphi : A \to A^*$ be a substitution with φ-fixed point $\mathbf{u}$. If the substitution φ is two-sided recognizable then it is also weakly circular.

If the morphism φ is injective, weakly circularity is equivalent to two-sided recognizability, this is not true for the non-injective case as shown by the following example.

Example

Consider the non-injective morphism $\varphi : a \to abc, b \to cab, c \to c$ with fixed point $\mathbf{u} = \varphi^\omega(a)$ (period $p = 4$). φ is weakly circular for $D = 3$, since all $v \in F(\mathbf{u})$ with $|v| > 2D$ have cc as factor (there is synchronizing point between the two c).
However φ is not two-sided recognizable.
For $L = 2$ we can find $v = \mathbf{u}_{[i-2,i+2]} = ccabc \in F(\mathbf{u})$ that has two different decodings $\varepsilon\varphi(c)\varphi(c)\varphi(a)b$ and $a\varphi(b)\varphi(c)\varphi(c)$ and for the first one $\mathbf{u}_i \in E$ while for the second one $\mathbf{u}_j \notin E$. For every $L' > L$, the word v can be extended in both sides in such a way that $w = \mathbf{u}_{[i'-L',i'+L']} = \mathbf{u}_{[i'-L',i'+L-1]} v \mathbf{u}_{[i'+L+1,i'+L']} \in F(\mathbf{u})$ and for w there are always two decodings such that $\mathbf{u}'_i \in E$ and $\mathbf{u}'_j \notin E$.


THANK YOU
FOR YOUR
ATTENTION