Factor complexity of D0L-systems

May 19, 2011

Basics terms & notation

$\mathcal{A} = \{1, \ldots, m\}$	an alphabet
$\mathcal{A}^*$, $\mathcal{A}^+$, $\mathcal{A}^{\mathbb{N}}$	all finite, finite non-empty, infinite words
$\mathbf{u} = (\mathbf{u}_i)_{i \geq 1}$, $\mathbf{u}_i \in \mathcal{A}$	a (left) infinite word over $\mathcal{A}$
$v = \mathbf{u}_j \mathbf{u}_{j+1} \cdots \mathbf{u}_{j+n-1}$	a factor of $\mathbf{u}$ of length n, the index j is an occurrence of v
$\mathcal{L}_n(\mathbf{u})$	a set of factors of $\mathbf{u}$ of length n
$\mathcal{L}(\mathbf{u}) = \bigcup_{n \in \mathbb{N}} \mathcal{L}_n(\mathbf{u})$	a language of $\mathbf{u}$

Basics terms & notation

$\mathcal{A} = \{1, \dots, m\}$	an alphabet
$\mathcal{A}^*$, $\mathcal{A}^+$, $\mathcal{A}^\mathbb{N}$	all finite, finite non-empty, infinite words
$\mathbf{u} = (\mathbf{u}_i)_{i \geq 1}$, $\mathbf{u}_i \in \mathcal{A}$	a (left) infinite word over $\mathcal{A}$
$v = \mathbf{u}_j \mathbf{u}_{j+1} \cdots \mathbf{u}_{j+n-1}$	a factor of $\mathbf{u}$ of length n, the index j is an occurrence of v
$\mathcal{L}_n(\mathbf{u})$	a set of factors of $\mathbf{u}$ of length n
$\mathcal{L}(\mathbf{u}) = \bigcup_{n \in \mathbb{N}} \mathcal{L}_n(\mathbf{u})$	a language of $\mathbf{u}$

A factor complexity of $\mathbf{u}$ is the function $\mathcal{C} : \mathbb{N} \to \mathbb{N}$ given by

$$\mathcal{C}(n) := \#\mathcal{L}_n(\mathbf{u}).$$

Basics terms & notation

$\mathcal{A} = \{1, \dots, m\}$	an alphabet
$\mathcal{A}^*$, $\mathcal{A}^+$, $\mathcal{A}^{\mathbb{N}}$	all finite, finite non-empty, infinite words
$\mathbf{u} = (\mathbf{u}_i)_{i \geq 1}$, $\mathbf{u}_i \in \mathcal{A}$	a (left) infinite word over $\mathcal{A}$
$v = \mathbf{u}_j \mathbf{u}_{j+1} \cdots \mathbf{u}_{j+n-1}$	a factor of $\mathbf{u}$ of length n, the index j is an occurrence of v
$\mathcal{L}_n(\mathbf{u})$	a set of factors of $\mathbf{u}$ of length n
$\mathcal{L}(\mathbf{u}) = \bigcup_{n \in \mathbb{N}} \mathcal{L}_n(\mathbf{u})$	a language of $\mathbf{u}$

A factor complexity of **u** is the function $\mathcal{C} : \mathbb{N} \to \mathbb{N}$ given by

$$\mathcal{C}(n) := \#\mathcal{L}_n(\mathbf{u}).$$

The first difference of the factor complexity of **u** is the function

$$\triangle\mathcal{C}(n) := \mathcal{C}(n+1) - \mathcal{C}(n).$$

D0L-systems

Definition

A triplet $G = (\mathcal{A}, \varphi, w)$ is called the D0L-system if $\mathcal{A}$ is an alphabet, φ a substitution on $\mathcal{A}$, and $w \in \mathcal{A}^+$ is the axiom. The language of the system $\mathcal{L}(G)$ is the set of all factors of the words $\varphi^n(w), n = 0, 1, \ldots$.
If the substitution is moreover non-erasing, then the system is called PD0L-system.

In order to keep things simple, we call a D0L-system PD0L-system with injective substitution with one-letter axiom a, such that $\varphi^\omega(a)$ is infinite.

Substitutions – classification

Let φ be a substitution defined on the alphabet $\mathcal{A}$.

Substitutions – classification

Let φ be a substitution defined on the alphabet $\mathcal{A}$.

φ is non-erasing if $\varphi(a_i) \neq \epsilon, i = 1, \ldots, m$.

Substitutions – classification

Let φ be a substitution defined on the alphabet $\mathcal{A}$.

φ is non-erasing if $\varphi(a_i) \neq \epsilon, i = 1, \ldots, m$.

φ is injective if for each $v \in \mathcal{A}^*$ there exists at most one $\bar{v}$ such that $\varphi(\bar{v}) = v$.

Substitutions – classification

Let φ be a substitution defined on the alphabet $\mathcal{A}$.

φ is non-erasing if $\varphi(a_i) \neq \epsilon, i = 1, \ldots, m$.

φ is injective if for each $v \in \mathcal{A}^*$ there exists at most one $\bar{v}$ such that $\varphi(\bar{v}) = v$.

φ is primitive if there exists $k \in \mathbb{N}$ such that for all $a, b \in \mathcal{A}$ the word $\varphi^k(a)$ contains b.

Equivalently, an incidence matrix M_φ is primitive, i.e. there exists k such that $M_\varphi^k > 0$.

Simple facts about complexity

For the factor complexity function $\mathcal{C}(n)$ of an infinite word **u** it holds that:

- $\mathcal{C}(n)$ is non-decreasing,
- $\mathcal{C}(n)$ is bounded iff **u** is eventually periodic,
- whenever $\mathcal{C}(n+1) = \mathcal{C}(n)$ for some $n \in \mathbb{N}$, then $\mathcal{C}(n)$ is bounded,
- **u** is aperiodic iff $\triangle\mathcal{C}(n) > 0$ for all $n \in \mathbb{N}$,
- $\mathcal{C}(n+k) \leq \mathcal{C}(n)\mathcal{C}(k)$ for all $n, k \in \mathbb{N}$,
- $\mathcal{C}(n) \leq (\#\mathcal{A})^n$ for all $n \in \mathbb{N}$.

Complexity of uniformly recurrent words

It was known that $\mathcal{C}(n)$ of uniformly recurrent words (inc. primitive D0L-systems) is a sublinear function, i.e.

$$\mathcal{C}(n) < an + b \quad \text{for some } a, b \in \mathbb{N}.$$

Mossé (1993) further proved that for fixed points of primitive substitutions there exists $K \in \mathbb{N}$

$$\triangle \mathcal{C}(n) < K, \text{ for all } n \in \mathbb{N}.$$

Cassaigne (1996) proved in more general context that if $\mathcal{C}(n) < an^{\alpha}$, where $a > 0$ and $1 \leq \alpha \leq \frac{3}{2}$, then

$$\triangle \mathcal{C}(n) < Kn^{3(\alpha-1)}.$$

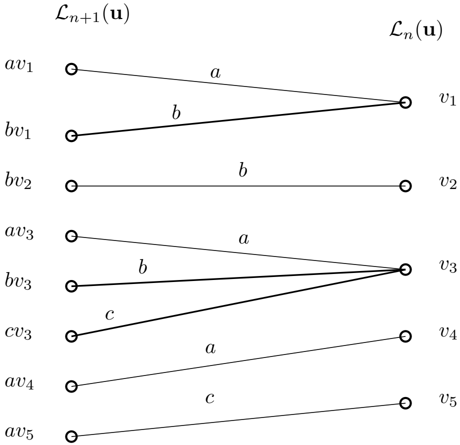
Special factors

For $v \in \mathcal{L}(\mathbf{u})$ we define the set of left extensions

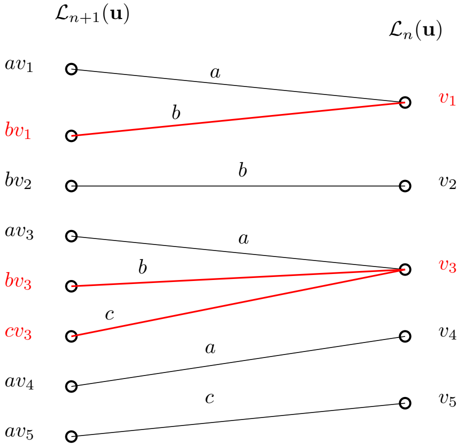
$$\mathrm{Lext}(v) := \{a \in \mathcal{A} \mid av \in \mathcal{L}(\mathbf{u})\}.$$

If $\#\mathrm{Lext}(v) > 1$, then v is said to be left special (LS) factor. Analogously are defined right special (RS) factors. If v is both LS and RS, it is called bispecial

LS factors and factor complexity



LS factors and factor complexity



$$\mathcal{C}(n+1) - \mathcal{C}(n) = 3$$

LS factors and factor complexity

For the first difference of the complexity function holds:

$$\triangle \mathcal{C}(n) := \mathcal{C}(n+1) - \mathcal{C}(n) = \sum_{\substack{v \in \mathcal{L}_n(\mathbf{u}) \\ v \text{ is LS}}} (\#\mathrm{Lext}(v) - 1).$$

LS factors and factor complexity

For the first difference of the complexity function holds:

$$\triangle\mathcal{C}(n) := \mathcal{C}(n+1) - \mathcal{C}(n) = \sum_{\substack{v\in\mathcal{L}_n(\mathbf{u}) \\ v \text{ is LS}}} (\#\mathrm{Lext}(v) - 1).$$

Complete knowledge of all LS factors along with the number of their left extensions allow us to evaluate $\mathcal{C}(n)$.

Sublinear complexity

Question 1: Which D0L-systems have sublinear complexity?

Theorem

A D0L-system has sublinear complexity if and only if the first difference of complexity is bounded.

Computing complexity

Question 2: How to find the complexity of D0L-systems / How to find the structure of special factors?

Theorem

We know how to describe (in a finite manner) all bispecial factors for so-called circular systems.

Results for φ_{TM}

$$\varphi_{TM} = \begin{cases} 0 \mapsto 01 \\ 1 \mapsto 10 \end{cases}$$

Complexity $\mathcal{C}_{TM}(n)$:

$\mathcal{C}_{TM}(1) = 2, \mathcal{C}_{TM}(2) = 4$, and, for $n \geq 3$, if $n = 2^r + q + 1, r \geq 0, 0 \leq q < 2^r$, then

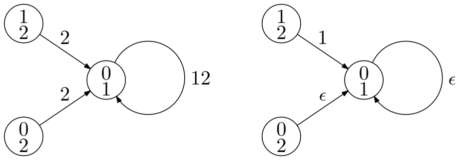
$$\mathcal{C}_{TM}(n) = \begin{cases} 6 \times 2^{r-1} + 4q & 0 \leq q \leq 2^{r-1}, \\ 2^{r+2} + 2q & 2^{r-1} < q < 2^r. \end{cases}$$

Results for φ_P

$$\varphi_P = 0 \mapsto 012, 1 \mapsto 112, 2 \mapsto 102$$

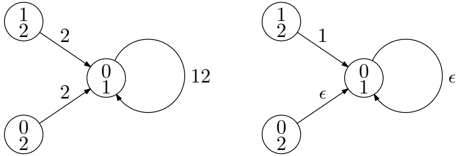
Results for φ_P

$$\varphi_P = 0 \mapsto 012, 1 \mapsto 112, 2 \mapsto 102$$



Results for φ_P

$$\varphi_P = 0 \mapsto 012, 1 \mapsto 112, 2 \mapsto 102$$



The set of all initial bispecial triples:

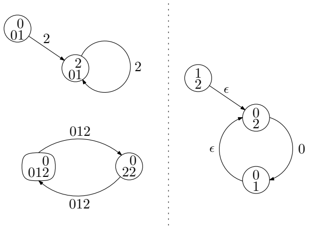
$$((0,1),\epsilon,(1,2)),\quad ((0,1),\epsilon,(0,1)),\quad ((0,1),\epsilon,(0,2)),\quad ((1,2),\epsilon,(0,2)),$$
$$((1,2),\epsilon,(1,2)),\quad ((1,2),\epsilon,(0,1)),\quad ((0,2),\epsilon,(0,1)),\quad ((0,2),\epsilon,(0,2)),$$
$$((0,2),\epsilon,(1,2)),\quad ((1,2),1,(1,2)),\quad ((0,2),1,(1,2)),\quad ((0,2),1,(0,2)),$$
$$((1,2),0,(1,2)).$$

Results for φ_S

$$\varphi_S = 0 \mapsto 0012, 1 \mapsto 2, 2 \mapsto 012$$

Results for φ_S

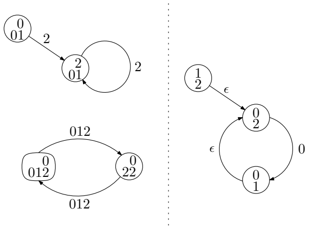
$$\varphi_S = 0 \mapsto 0012, 1 \mapsto 2, 2 \mapsto 012$$



$\mathrm{GL}_{\varphi_S}^{(\mathcal{B}_L)}$ $\mathrm{GR}_{\varphi_S}^{(\mathcal{B}_R)}$

Results for φ_S

$$\varphi_S = 0 \mapsto 0012, 1 \mapsto 2, 2 \mapsto 012$$



$GL_{\varphi_S}^{(\mathcal{B}_L)}$ $GR_{\varphi_S}^{(\mathcal{B}_R)}$

The set of all initial bispecial triples:

$$((2,01),2,(0,2)),\quad ((2,01),20,(0,1)),\quad ((0,22),012,(0,2)),$$
$$((0,22),0120,(0,1)),\quad ((0,012),012,(0,2)),\quad ((0,012),1,(0,1)).$$

An example of a non-circular substitution

$$\varphi_{NC} = 0 \mapsto 0101, 1 \mapsto 11$$

Question 3: How to find the structure of special factors for non-circular D0L-systems?

Rank zero letters

Definition

Given a D0L-system $(\mathcal{A}, \varphi, a)$, $b \in \mathcal{A}$ is of rank zero, if $|\varphi^n(b)|$ bounded. The set of all such letters is denoted by $\mathcal{A}_0$.

- *If $\mathcal{A}_0 = \emptyset$, the system is growing.*
- *If there is an infinite number of factors over $\mathcal{A}_0$, the system is pushy.*
- *Otherwise, the system is not pushy.*

Images of letters

Lemma (Salomaa, Soittola)

The sequence $|\varphi^n(x)|$ is either bounded or it grows like $n^{a_x} b_x^n$ with $b_x > 1$ and $a_x \in \mathbb{N}$.

Definition

The D0L-system $(\mathcal{A}, \varphi, a)$ is

- *quasi-uniform if all letters grows like b^n for some $b > 1$.*
- *polynomial-divergent if all letters grows like $n^{a_x} b^n$ and one a_x is nonzero.*
- *exponential-divergent if there are letters x, y growing like $n^{a_x} b_x^n$ and $n^{a_y} b_y^n$ with $1 < b_x < b_y$ and $b_z > 1$ for all z.*

Growing systems

Theorem (Pansiot)

Given a growing D0L-system $(\mathcal{A}, \varphi, a)$ with $|\varphi^n(a)|$ aperiodic. If the system is

- *quasi-uniform, $\mathcal{C}(n)$ grows like n,*
- *polynomial-divergent, $\mathcal{C}(n)$ grows like $n \log n$,*
- *exponential-divergent, $\mathcal{C}(n)$ grows like $n \log \log n$*

Not-growing systems

Theorem (Pansiot)

Given a growing D0L-system $(\mathcal{A}, \varphi, a)$ with $|\varphi^n(a)|$ aperiodic. If the system is

- *pushy, $\mathcal{C}(n)$ grows like n^2,*
- *non-pushy, then there exist an alphabet $\mathcal{B}$, a growing system $(\mathcal{B}, \varphi', b)$ and a non-erasing morphism $h : \mathcal{B}^* \to \mathcal{A}^*$ such that $\varphi^\omega(a) = h(\varphi'^\omega(b))$.*
  Moreover, the complexity is in $O(n)$, $O(n \log n)$, $O(n \log\log n)$ if $(\mathcal{B}, \varphi', b)$ is quasi-uniform, polynomial-divergent, exponential-divergent, respectively.