Introduction to Abelian complexity

Karel Břinda

FNSPE CTU in Prague

May 18, 2010

Notation

- $\mathcal{A}$ - alphabet
- $\mathcal{A}^*$ - set of all finite words
- $\mathcal{A}^{\mathbb{N}}$ - set of all right infinite words
- $\mathcal{A}^{\mathbb{Z}}$ - set of all biinfinite words
- v, w - finite words
- $\mathbf{u}$ - infinite word
- $|v|$ - length of the word v
- $|v|_a$ - number of occurrences of letter a in the word v
- $\mathcal{L}_n(\mathbf{u})$ - set of all factors of length n of word $\mathbf{u}$

Basic definitions

Definition

LET $\mathcal{C}(n) = \#\mathcal{L}_n(\mathbf{u})$
THEN $\mathcal{C}(n)$ is called Subword complexity function (Factor complexity function)

Basic definitions

Definition

LET $\mathcal{C}(n) = \#\mathcal{L}_n(\mathbf{u})$
THEN $\mathcal{C}(n)$ is called Subword complexity function (Factor complexity function)

Definition

LET $v, w \in \mathcal{A}^*$, $(\forall a \in \mathcal{A})(|v|_a = |w|_a)$
THEN v and w are said to be Abelian equivalent (denoted by $\sim_{ab}$)

Basic definitions

Definition

LET $\mathcal{C}(n) = \#\mathcal{L}_n(\mathbf{u})$
THEN $\mathcal{C}(n)$ is called Subword complexity function (Factor complexity function)

Definition

LET $v, w \in \mathcal{A}^*$, $(\forall a \in \mathcal{A})(|v|_a = |w|_a)$
THEN v and w are said to be Abelian equivalent (denoted by $\sim_{ab}$)

Definition

LET $\mathcal{C}_{ab}(n) = \#(\mathcal{L}_n(\mathbf{u})/ \sim_{ab})$
THEN $\mathcal{C}_{ab}(n)$ is called Abelian complexity

Basic definitions

Definition

LET $\mathcal{A} = \{0, 1, ..., k-1\}, v \in \mathcal{A}^*$
THEN $\Psi(v) = (|v|_0, |v|_1, ..., |v|_{k-1})$ is a Parikh vector associated to v

Basic definitions

Definition

LET $\mathcal{A} = \{0, 1, ..., k-1\}, v \in \mathcal{A}^*$
THEN $\Psi(v) = (|v|_0, |v|_1, ..., |v|_{k-1})$ is a Parikh vector associated to v

It is seen that

$$\mathcal{C}_{ab}(n) = \#\{\Psi(v) | v \in \mathcal{L}_n(\mathbf{u})\}$$

Example

$\mathbf{u} = 001000100010\ldots$

Factors of length 3: $\mathcal{L}_3 = \{001, 010, 100, 000\}$, $\mathcal{C}(3) = 4$
Parikh vectors: $(2,1), (2,1), (2,1), (3,0)$, $\mathcal{C}_{ab}(3) = 2$

Factors of length 4: $\mathcal{L}_4 = \{0010, 0100, 1000, 0001\}$, $\mathcal{C}(4) = 4$
Parikh vectors: $(3,1), (3,1), (3,1), (3,1)$, $\mathcal{C}_{ab}(3) = 1$

Extremal values

Lemma

LET $\mathbf{u} \in \mathcal{A}^{\mathbb{N}} \cup \mathcal{A}^{\mathbb{Z}}$
THEN $\mathbf{u}$ *is periodic of period* $p \Leftrightarrow \mathcal{C}_{ab}(p) = 1$

Extremal values

Lemma

LET $\mathbf{u} \in \mathcal{A}^{\mathbb{N}} \cup \mathcal{A}^{\mathbb{Z}}$
THEN $\mathbf{u}$ *is periodic of period* $p \Leftrightarrow \mathcal{C}_{ab}(p) = 1$

Proof.

$\Rightarrow$**:** it follows from $(\forall v, w \in \mathcal{L}_p(\mathbf{u}))(\forall a \in \mathcal{A})(|v|_a = |w|_a)$

Extremal values

Lemma

LET $\mathbf{u} \in \mathcal{A}^{\mathbb{N}} \cup \mathcal{A}^{\mathbb{Z}}$
THEN $\mathbf{u}$ *is periodic of period* $p \Leftrightarrow \mathcal{C}_{ab}(p) = 1$

Proof.

$\Rightarrow$: it follows from $(\forall v, w \in \mathcal{L}_p(\mathbf{u}))(\forall a \in \mathcal{A})(|v|_a = |w|_a)$
$\Leftarrow$: by contradiction ($\mathbf{u}$ not periodic of period $p \wedge \mathcal{C}_{ab}(p) = 1$) Q.E.D.

Extremal values

Theorem

LET $\mathbf{u} \in \mathcal{A}^{\mathbb{N}}, \#\mathcal{A} = k, n \geq 0$
THEN

1. $1 \leq \mathcal{C}_{ab}(n) \leq \binom{n+k-1}{k-1}$
2. *Abelian complexity is bounded by* $\mathcal{O}(n^{k-1})$

Extremal values

Theorem

LET $\mathbf{u} \in \mathcal{A}^{\mathbb{N}}, \#\mathcal{A} = k, n \geq 0$
THEN

1. $1 \leq \mathcal{C}_{ab}(n) \leq \binom{n+k-1}{k-1}$
2. *Abelian complexity is bounded by* $\mathcal{O}(n^{k-1})$

Proof.

1. Using generating functions:
$$\underbrace{(1 + x + x^2 + \ldots) \cdots (1 + x + x^2 + \ldots)}_{k} = (\ldots + \varpi \cdot x^n + \ldots),\ \varpi = ?$$

Extremal values

Theorem

LET $\mathbf{u} \in \mathcal{A}^{\mathbb{N}}, \#\mathcal{A} = k, n \geq 0$
THEN

1. $1 \leq \mathcal{C}_{ab}(n) \leq \binom{n+k-1}{k-1}$
2. *Abelian complexity is bounded by* $\mathcal{O}(n^{k-1})$

Proof.

1. Using generating functions:
$$\underbrace{(1 + x + x^2 + \dots) \cdots (1 + x + x^2 + \dots)}_{k} = (\dots + \varpi \cdot x^n + \dots),\ \varpi = ?$$
2. $\binom{n+k-1}{k-1} = \frac{(n+k-1)!}{(k-1)!n!} = \frac{(n+k-1)(n+k-2)\cdots(n+1)}{(k-1)!} \sim \mathcal{O}(n^{k-1})$

Q.E.D.

Links with balance properties

Definition

LET $\mathbf{u} \in \mathcal{A}^{\mathbb{N}} \cup \mathcal{A}^{\mathbb{Z}}$,
$(\exists \mathcal{C} \in \mathbb{N})(\forall n)(\forall v, w \in \mathcal{L}_n(\mathbf{u}))(\forall a \in \mathcal{A})(\big||v|_a - |w|_a\big| \leq \mathcal{C})$
THEN $\mathbf{u}$ is $\mathcal{C}$-balanced

Links with balance properties

Definition

LET $\mathbf{u} \in \mathcal{A}^{\mathbb{N}} \cup \mathcal{A}^{\mathbb{Z}}$,
$(\exists \mathcal{C} \in \mathbb{N})(\forall n)(\forall v, w \in \mathcal{L}_n(\mathbf{u}))(\forall a \in \mathcal{A})\left(\left||v|_a - |w|_a\right| \leq \mathcal{C}\right)$
THEN $\mathbf{u}$ is $\mathcal{C}$-balanced

Theorem

LET $\mathbf{u} \in \mathcal{A}^{\mathbb{N}} \cup \mathcal{A}^{\mathbb{Z}}$
THEN $\mathcal{C}_{ab}(n)$ *is bounded* $\Leftrightarrow$ $\mathbf{u}$ *is* $\mathcal{C}$*-balanced for some positive integer* $\mathcal{C}$

Links with balance properties

Theorem

LET $\mathbf{u} \in \mathcal{A}^{\mathbb{N}}$ *be aperiodic,* $\mathcal{A} = \{0, 1\}$
THEN $\mathbf{u}$ *is balanced (it means 1-balanced)* $\Leftrightarrow (\forall n \geq 1)(\mathcal{C}_{ab}(n) = 2)$

Proof.

Let $n \in \mathbb{N}$; $v \in \mathcal{L}_n(\mathbf{u})$ then $\Psi(v) = (|v|_0, |v|_1)$, where $|v|_0 + |v|_1 = n$
⇒: There can be max. one another Parikh vector:
$(|v|_0 - 1, |v|_1 + 1)$ **XOR** $(|v|_0 + 1, |v|_1 - 1)$
$\Rightarrow \mathcal{C}_{ab}(n) \leq 2$
$(\forall n \geq 1)(\exists v \in \mathcal{L}_{n-1}(\mathbf{u}))(v0 \in \mathcal{L}(\mathbf{u}) \wedge v1 \in \mathcal{L}(\mathbf{u})) \Rightarrow \mathcal{C}_{ab}(n) \geq 2$
Q.E.D.

Remark: $\Leftrightarrow$ $\mathbf{u}$ is Sturmian

Links with balance properties

Theorem

LET u $\in \mathcal{A}^{\mathbb{N}}$ *be aperiodic,* $\mathcal{A} = \{0, 1\}$
THEN u *is balanced (it means 1-balanced)* $\Leftrightarrow (\forall n \geq 1)(\mathcal{C}_{ab}(n) = 2)$

Proof.

Let $n \in \mathbb{N}$; $v \in \mathcal{L}_n(\mathbf{u})$ then $\Psi(v) = (|v|_0, |v|_1)$, where $|v|_0 + |v|_1 = n$
$\Leftarrow$**:** For every $n \geq 1$, there are 2 different Parikh vectors of the form $(a+1, b), (a, b+1)$, where $a + b + 1 = n$.

Q.E.D.

Remark: $\Leftrightarrow$ **u** is Sturmian

Question of Gérard Rauzy

Is there any aperiodic 3-letter word with $\mathcal{C}_{ab} = 3$?

Question of Gérard Rauzy - answer 1

Is there any aperiodic 3-letter word with $\mathcal{C}_{ab} = 3$? :

Example

LET $\mathcal{A} = \{0, 1, 2\}$, **u** Sturmian word over alphabet $\{0, 1\}$

Question of Gérard Rauzy - answer 1

Is there any aperiodic 3-letter word with $\mathcal{C}_{ab} = 3$? :

Example

LET $\mathcal{A} = \{0, 1, 2\}$, $\mathbf{u}$ Sturmian word over alphabet $\{0, 1\}$
Word: $2\mathbf{u}$

Question of Gérard Rauzy - answer 2

Is there any aperiodic 3-letter word with $\mathcal{C}_{ab} = 3$? :-

Theorem

LET u *be aperiodic, uniformly recurrent and balanced,* $\#\mathcal{A} = 3$
THEN $(\forall n \geq 1)(\mathcal{C}_{ab}(n) = 3)$

Question of Gérard Rauzy - answer 3

Is there any aperiodic 3-letter word with $\mathcal{C}_{ab} = 3$? :-)

Theorem

LET u' $\in \{0,1\}^{\mathbb{N}}$ *be aperiodic,* $\mathbf{u} = f(\mathbf{u'})$ *where* f *is the morphism*
$0 \to 012, 1 \to 021$
THEN $(\forall n \geq 1)(\mathcal{C}_{ab}(n) = 3)$

Open question

Is there any recurrent infinite word with $\mathcal{C}_{ab} = 4$?

Thank you for your attention!