Frekvence a symetrie

Ľubomíra Balková

Fryšava

květen 2011

Motivation

Theorem

Let $\mathbf{u}$ *be an infinite recurrent word and such that the frequency of any factor exists. Then*

$$\#\{\rho(e) \mid e \in \mathcal{L}_{n+1}(\mathbf{u})\} \quad \leq \quad 3\Delta\mathcal{C}(n).$$

Theorem (Balková, Pelantová)

Let $\mathbf{u}$ *be an infinite word whose language is closed under reversal and such that the frequency of any factor exists. Then*

$$\#\{\rho(e)|e \in \mathcal{L}_{n+1}(\mathbf{u})\} \quad \leq \quad 2\Delta\mathcal{C}(n) + 1.$$

Motivation

Theorem

Let $\mathbf{u}$ *be an infinite recurrent word and such that the frequency of any factor exists. Then*

$$\#\{\rho(e) \mid e \in \mathcal{L}_{n+1}(\mathbf{u})\} \quad \leq \quad 3\Delta\mathcal{C}(n).$$

Theorem (Balková, Pelantová)

Let $\mathbf{u}$ *be an infinite word whose language is closed under reversal and such that the frequency of any factor exists. Then*

$$\#\{\rho(e)|e \in \mathcal{L}_{n+1}(\mathbf{u})\} \quad \leq \quad 2\Delta\mathcal{C}(n) + 1.$$

Question: Is it possible to reduce the upper bound for infinite words invariant under more symmetries?

Motivation

Theorem

Let $\mathbf{u}$ be an infinite recurrent word and such that the frequency of any factor exists. Then

$$\#\{\rho(e) \mid e \in \mathcal{L}_{n+1}(\mathbf{u})\} \quad \leq \quad 3\Delta\mathcal{C}(n).$$

Theorem (Balková, Pelantová)

Let $\mathbf{u}$ be an infinite word whose language is closed under reversal and such that the frequency of any factor exists. Then

$$\#\{\rho(e) | e \in \mathcal{L}_{n+1}(\mathbf{u})\} \quad \leq \quad 2\Delta\mathcal{C}(n) + 1.$$

Question: Is it possible to reduce the upper bound for infinite words invariant under more symmetries?

Definitions

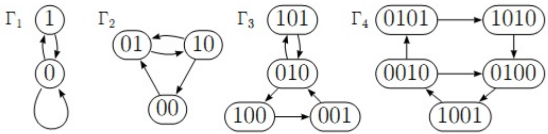
- $\mathcal{A} = \{0, 1, \ldots, k-1\}$
- $\mathbf{u} \in \mathcal{A}^{\mathbb{N}}$, $\mathbb{N} = \{0, 1, 2, \ldots\}$
- $\mathcal{L}_n(\mathbf{u})$ = set of all factors of length n of $\mathbf{u}$
- $\mathcal{C}(n) = \#\mathcal{L}_n(\mathbf{u})$ ... *complexity of* $\mathbf{u}$
- $\mathcal{P}(n) = \#\{p \in \mathcal{L}_n(\mathbf{u}) | p \text{ palindrome}\}$ ... *palindromic complexity of* $\mathbf{u}$
- if $w = v_i v_{i+1} \ldots v_{i+n-1}$, we call i an *occurrence* of w in v
- $\mathbf{u}$ is *recurrent* if every factor has ∞ many occurrences in $\mathbf{u}$
- $w \in \mathcal{L}(\mathbf{u})$ is RS if wa and $wb \in \mathcal{L}(\mathbf{u})$ for $a, b \in \mathcal{A},\ a \neq b$ (similarly LS)
- w is *special* if w is RS or LS
- w is BS if w is both RS and LS

Rauzy graphs

- *Rauzy graph* Γ_n of order n of $\mathbf{u}$ is a directed graph whose set of vertices is $\mathcal{L}_n(\mathbf{u})$ and set of edges is $\mathcal{L}_{n+1}(\mathbf{u})$

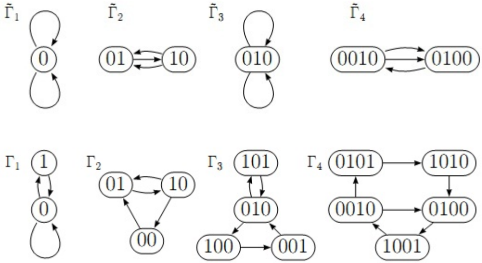
Example

$\varphi(0) = 001, \varphi(1) = 01, \quad \mathbf{u} = \varphi(\mathbf{u}) = 00100101001001010010 1\ldots$



Reduced Rauzy graphs

- a factor e is a *simple path* of order n if it starts and ends in a special factor of length n and no other factors of e are special
- *reduced Rauzy graph* $\tilde{\Gamma}_n$ of $\mathbf{u}$ (of order n) is a directed graph whose set of vertices is formed by LS and RS factors of $\mathcal{L}_n(\mathbf{u})$ and whose set of edges consists of simple paths



Factor frequency

$$\rho(w) = \lim_{|v|\to\infty, v\in\mathcal{L}(\mathbf{u})} \frac{\#\{\text{occurrences of } w \text{ in } v\}}{|v|}$$

Theorem (Kirchhoff's law)

Consider Γ_n of $\mathbf{u}$ with existing frequencies. Denote E (E') the set of edges starting (ending) in the vertex w, then

$$\sum_{e\in E}\rho(e) = \rho(w) = \sum_{e'\in E'}\rho(e').$$

Factor frequency

$$\rho(w) = \lim_{|v|\to\infty, v\in\mathcal{L}(\mathbf{u})} \frac{\#\{\text{occurrences of } w \text{ in } v\}}{|v|}$$

Theorem (Kirchhoff's law)

Consider Γ_n of $\mathbf{u}$ with existing frequencies. Denote E (E') the set of edges starting (ending) in the vertex w, then

$$\sum_{e\in E} \rho(e) = \rho(w) = \sum_{e'\in E'} \rho(e').$$

Corollary

- *If w is not RS, then $\rho(w) = \rho(wa)$.*
- $\{\rho(e) | e \text{ edge in } \Gamma_n\} = \{\rho(e) | e \text{ edge in } \tilde{\Gamma}_n\}$.

Factor frequency

$$\rho(w) = \lim_{|v|\to\infty, v\in\mathcal{L}(\mathbf{u})} \frac{\#\{\text{occurrences of } w \text{ in } v\}}{|v|}$$

Theorem (Kirchhoff's law)

Consider Γ_n of $\mathbf{u}$ with existing frequencies. Denote E (E') the set of edges starting (ending) in the vertex w, then

$$\sum_{e\in E} \rho(e) = \rho(w) = \sum_{e'\in E'} \rho(e').$$

Corollary

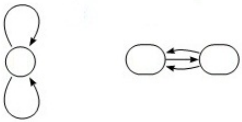
- *If w is not RS, then $\rho(w) = \rho(wa)$.*
- $\{\rho(e) | e \text{ edge in } \Gamma_n\} = \{\rho(e) | e \text{ edge in } \tilde{\Gamma}_n\}$.

Sturmian words

Definition

Let $\mathbf{u} \in \mathcal{A}^{\mathbb{N}}$ satisfy $\mathcal{C}(n) = n + 1$ for all $n \in \mathbb{N}$. Then $\mathbf{u}$ is called *Sturmian*. (Necessarily $\mathcal{A} = \{0, 1\}$.)

$\mathcal{L}(\mathbf{u})$ contains one RS and one LS factor of every length $\Rightarrow$ reduced Rauzy graphs have either two or three edges



Berthé described for every n the set of frequencies of factors of length n.

Rough upper bound

Theorem (Boshernitzan)

Let **u** *be an infinite recurrent word and such that the frequency of any factor exists. Then*

$$\#\{\rho(e) \mid e \in \mathcal{L}_{n+1}(\mathbf{u})\} \quad \leq \quad 3\Delta\mathcal{C}(n).$$

Proof

$$\#\{\rho(e) \mid e \in \mathcal{L}_{n+1}(\mathbf{u})\} \quad \leq \quad \#\{e \mid e \text{ edge in } \tilde{\Gamma}_n\}$$

$$\#\{e \mid e \text{ edge in } \tilde{\Gamma}_n\} = \sum_{w \text{ vertex in } \tilde{\Gamma}_n} \#\mathrm{Rext}(w)$$

$$\sum_{w \text{ vertex in } \tilde{\Gamma}_n} \#\mathrm{Rext}(w) = \sum_{w \text{ vertex in } \tilde{\Gamma}_n} (\#\mathrm{Rext}(w) - 1) + \sum_{w \text{ vertex in } \tilde{\Gamma}_n} 1$$

Rough upper bound

Theorem (Boshernitzan)

Let $\mathbf{u}$ *be an infinite recurrent word and such that the frequency of any factor exists. Then*

$$\#\{\rho(e) \mid e \in \mathcal{L}_{n+1}(\mathbf{u})\} \quad \leq \quad 3\Delta\mathcal{C}(n).$$

Proof

$$\#\{\rho(e) \mid e \in \mathcal{L}_{n+1}(\mathbf{u})\} \quad \leq \quad \#\{e \mid e \text{ edge in } \tilde{\Gamma}_n\}$$

$$\#\{e \mid e \text{ edge in } \tilde{\Gamma}_n\} = \sum_{w \text{ vertex in } \tilde{\Gamma}_n} \#\mathrm{Rext}(w)$$

$$\sum_{w \text{ vertex in } \tilde{\Gamma}_n} \#\mathrm{Rext}(w) = \sum_{w \text{ vertex in } \tilde{\Gamma}_n} (\#\mathrm{Rext}(w) - 1) + \sum_{w \text{ vertex in } \tilde{\Gamma}_n} 1$$

Symmetries preserving frequency

$\Psi : \mathcal{A}^* \to \mathcal{A}^*$ is called a *symmetry* if:

1. Ψ is a bijection,
2. for all $w, v \in \mathcal{A}^*$

$$\#\{\text{occurrences of } w \text{ in } v\} = \#\{\text{occurrences of } \Psi(w) \text{ in } \Psi(v)\}.$$

Proposition

Ψ is a letter permutation extended to a morphism or antimorphism.

Symmetries preserving frequency

$\Psi : \mathcal{A}^* \to \mathcal{A}^*$ is called a *symmetry* if:

1. Ψ is a bijection,
2. for all $w, v \in \mathcal{A}^*$

$$\#\{\text{occurrences of } w \text{ in } v\} = \#\{\text{occurrences of } \Psi(w) \text{ in } \Psi(v)\}.$$

Proposition

Ψ *is a letter permutation extended to a morphism or antimorphism.*

Corollary

Let $\mathcal{L}(\mathbf{u})$ be closed under a symmetry Ψ. For all w in $\mathcal{L}(\mathbf{u})$ whose frequency exists

$$\rho(w) = \rho(\Psi(w)).$$

Symmetries preserving frequency

$\Psi : \mathcal{A}^* \to \mathcal{A}^*$ is called a *symmetry* if:

1. Ψ is a bijection,
2. for all $w, v \in \mathcal{A}^*$

$$\#\{\text{occurrences of } w \text{ in } v\} = \#\{\text{occurrences of } \Psi(w) \text{ in } \Psi(v)\}.$$

Proposition

Ψ is a letter permutation extended to a morphism or antimorphism.

Corollary

Let $\mathcal{L}(\mathbf{u})$ be closed under a symmetry Ψ. For all w in $\mathcal{L}(\mathbf{u})$ whose frequency exists

$$\rho(w) = \rho(\Psi(w)).$$

Properties of symmetry

Let $\mathbf{u}$ be closed under a finite group G of symmetries containing an antimorphism. Then

1. G has an even number of elements
2. if w contains all letters, then
   1. for any distinct antimorphisms $\theta_1, \theta_2 \in G$, we have $\theta_1(w) \neq \theta_2(w)$,
   2. for any distinct morphisms $\varphi_1, \varphi_2 \in G$, we have $\varphi_1(w) \neq \varphi_2(w)$
3. if w is a θ-palindrome, then θ is an involutive antimorphism

Reduced Rauzy graphs for "symmetric" words

1. if there is an edge e in $\tilde{\Gamma}_n$
   - between vertices w and $\theta(w)$, then there are $\#G/2$ distinct edges labeled $\rho(e)$
   - otherwise, there are $\#G$ distinct edges labeled $\rho(e)$
2. if an edge e in the reduced Rauzy graph $\tilde{\Gamma}_n$ is mapped by θ onto itself, then e has a θ-palindrome of length either n or $n+1$ as its central factor

Reduced Rauzy graphs for "symmetric" words

1. if there is an edge e in $\tilde{\Gamma}_n$
   - between vertices w and $\theta(w)$, then there are $\#G/2$ distinct edges labeled $\rho(e)$
   - otherwise, there are $\#G$ distinct edges labeled $\rho(e)$
2. if an edge e in the reduced Rauzy graph $\tilde{\Gamma}_n$ is mapped by θ onto itself, then e has a θ-palindrome of length either n or $n+1$ as its central factor
3. every θ-palindrome of length $n+1$ is the central factor of an edge in $\tilde{\Gamma}_n$ mapped by θ onto itself

Reduced Rauzy graphs for "symmetric" words

1. if there is an edge e in $\tilde{\Gamma}_n$
   - between vertices w and $\theta(w)$, then there are $\#G/2$ distinct edges labeled $\rho(e)$
   - otherwise, there are $\#G$ distinct edges labeled $\rho(e)$
2. if an edge e in the reduced Rauzy graph $\tilde{\Gamma}_n$ is mapped by θ onto itself, then e has a θ-palindrome of length either n or $n+1$ as its central factor
3. every θ-palindrome of length $n+1$ is the central factor of an edge in $\tilde{\Gamma}_n$ mapped by θ onto itself
4. every θ-palindrome of length n is either the central factor of an edge mapped by θ onto itself or is a vertex

Reduced Rauzy graphs for "symmetric" words

1. if there is an edge e in $\tilde{\Gamma}_n$
   - between vertices w and $\theta(w)$, then there are $\#G/2$ distinct edges labeled $\rho(e)$
   - otherwise, there are $\#G$ distinct edges labeled $\rho(e)$
2. if an edge e in the reduced Rauzy graph $\tilde{\Gamma}_n$ is mapped by θ onto itself, then e has a θ-palindrome of length either n or $n+1$ as its central factor
3. every θ-palindrome of length $n+1$ is the central factor of an edge in $\tilde{\Gamma}_n$ mapped by θ onto itself
4. every θ-palindrome of length n is either the central factor of an edge mapped by θ onto itself or is a vertex

Factor versus θ-palindromic complexity

Theorem (Pelantová, Starosta)

Let G be a finite group of symmetries containing an antimorphism and let $\mathbf{u}$ be a uniformly recurrent infinite word whose language is invariant under all elements of G. Then there exists $N \in \mathbb{N}$ such that

$$\Delta\mathcal{C}(n) + \#G \geq \sum_{\theta \in G} \big(\mathcal{P}_\theta(n) + \mathcal{P}_\theta(n+1)\big) \quad \text{for all } n \geq N.$$

Main theorem

Theorem

Let G be a finite group of symmetries containing an antimorphism and let $\mathbf{u}$ be a uniformly recurrent infinite word whose language is invariant under all elements of G and such that the frequency of any factor exists. Then there exists $N \in \mathbb{N}$ such that

$$\#\{\rho(e) | e \in \mathcal{L}_{n+1}(\mathbf{u})\} \quad \leq \quad \frac{1}{\#G}\Big(4\Delta\mathcal{C}(n) + \#G\Big) \quad \text{for all } n \geq N.$$

Proof

Any element of G maps $\tilde{\Gamma}_n$ onto itself.

$$3\Delta\mathcal{C}(n) \ \geq \ \#\{e |\ e \text{ edge in } \tilde{\Gamma}_n\} = A + B,$$

$A =$ number of edges mapped onto themselves by some $\Psi \in G$,
$B =$ number of edges not mapped onto themselves by any $\Psi \in G$.

Proof

Any element of G maps $\tilde{\Gamma}_n$ onto itself.

$$3\Delta\mathcal{C}(n) \ \geq \ \#\{e|\ e \text{ edge in } \tilde{\Gamma}_n\} = A + B,$$

$A =$ number of edges mapped onto themselves by some $\Psi \in G$,
$B =$ number of edges not mapped onto themselves by any $\Psi \in G$.

$$A \quad \leq \quad \sum_{\theta \in G} \big(\mathcal{P}_\theta(n) + \mathcal{P}_\theta(n+1)\big) \quad \leq \quad \frac{1}{\#G}\Big(4\Delta\mathcal{C}(n) + \#G\Big).$$

Proof

Any element of G maps $\tilde{\Gamma}_n$ onto itself.

$$3\Delta\mathcal{C}(n) \;\geq\; \#\{e|\; e \text{ edge in } \tilde{\Gamma}_n\} = A + B,$$

$A =$ number of edges mapped onto themselves by some $\Psi \in G$,
$B =$ number of edges not mapped onto themselves by any $\Psi \in G$.

$$A \quad \leq \quad \sum_{\theta \in G} \big(\mathcal{P}_\theta(n) + \mathcal{P}_\theta(n+1)\big) \quad \leq \quad \frac{1}{\#G}\Big(4\Delta\mathcal{C}(n) + \#G\Big).$$

$$\#\{\rho(e)|\; e \in \mathcal{L}_{n+1}(\mathbf{u})\} \quad \leq \quad \tfrac{1}{k}A + \tfrac{1}{2k}B \quad = \quad \tfrac{1}{2k}A + \tfrac{1}{2k}(A+B),$$

where $\#G = 2k$.

Proof

Any element of G maps $\tilde{\Gamma}_n$ onto itself.

$$3\Delta\mathcal{C}(n) \ \geq \ \#\{e|\ e \text{ edge in } \tilde{\Gamma}_n\} = A + B,$$

$A =$ number of edges mapped onto themselves by some $\Psi \in G$,
$B =$ number of edges not mapped onto themselves by any $\Psi \in G$.

$$A \quad \leq \quad \sum_{\theta \in G} \big(\mathcal{P}_\theta(n) + \mathcal{P}_\theta(n+1)\big) \quad \leq \quad \frac{1}{\#G}\Big(4\Delta\mathcal{C}(n) + \#G\Big).$$

$$\#\{\rho(e)|\ e \in \mathcal{L}_{n+1}(\mathbf{u})\} \quad \leq \quad \tfrac{1}{k}A + \tfrac{1}{2k}B \quad = \quad \tfrac{1}{2k}A + \tfrac{1}{2k}(A+B),$$

where $\#G = 2k$.