How extensions impact the factor complexity of morphic images

France Gheeraert

February 28, 2024

The protagonists

Definition

The *factor complexity* of $x \in \mathcal{A}^{\mathbb{Z}}$ is the function

$$p_x(n)\colon \mathbb{N} \to \mathbb{N}, \quad n \mapsto \#\mathcal{L}(x) \cap \mathcal{A}^n.$$

Definition

A *morphism* is a monoid morphism $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$, i.e., for any $u, v \in \mathcal{A}^*$,

$$\sigma(uv) = \sigma(u)\sigma(v).$$

Every morphism is assumed to be **non-erasing**.

A classical result

How do the properties of x and σ impact the factor complexity of $\sigma(x)$?

A classical result

How do the properties of x and σ impact the factor complexity of $\sigma(x)$?

Proposition (Allouche & Shallit '03, Cassaigne & Nicolas '10)

Let $x \in \mathcal{A}^{\mathbb{Z}}$ and $\sigma \colon \mathcal{A}^* \to \mathcal{B}^*$. For all $n \geq 0$, we have

$$p_{\sigma(x)}(n) \leq \|\sigma\| \cdot p_x(n),$$

where $\|\sigma\| := \max_{a \in \mathcal{A}} |\sigma(a)|$ is the width of σ.

Sketch of proof on the blackboard.

Coverings

Definition

A *covering* of $u \in \mathcal{B}^n$ is a pair $(w, k) \in \mathcal{L}(x) \times \mathbb{Z}_{\geq 0}$ where $u = \sigma(w)_{[k+1,k+n]}$ and w is minimal, i.e.

$$k + 1 \leq |\sigma(w_1)| \quad \text{and} \quad k + n \geq \left|\sigma(w_{[1,|w|[})\right| + 1$$

The set of coverings of words of length n is denoted $C_{x,\sigma}(n)$.

Coverings

Definition

A *covering* of $u \in \mathcal{B}^n$ is a pair $(w, k) \in \mathcal{L}(x) \times \mathbb{Z}_{\geq 0}$ where $u = \sigma(w)_{[k+1,k+n]}$ and w is minimal, i.e.

$$k + 1 \leq |\sigma(w_1)| \quad \text{and} \quad k + n \geq \left|\sigma(w_{[1,|w|[})\right| + 1$$

The set of coverings of words of length n is denoted $C_{x,\sigma}(n)$.

Proposition

We have

$$p_{\sigma(x)}(n) \leq \# C_{x,\sigma}(n).$$

Coverings of letters

If (w, k) is a covering of a letter, then

Coverings of letters

If (w, k) is a covering of a letter, then

- w is a letter,

Coverings of letters

If (w, k) is a covering of a letter, then

- w is a letter,
- $0 \leq k < |\sigma(w)|$

Coverings of letters

If (w, k) is a covering of a letter, then

- w is a letter,
- $0 \leq k < |\sigma(w)|$

So

$$\#C_{x,\sigma}(1) = \sum_{a \in \mathcal{A}} |\sigma(a)|$$

Link between $C_{x,\sigma}(n)$ and $C_{x,\sigma}(n+1)$

$$\sigma: \begin{cases} 0 \mapsto 001 \\ 1 \mapsto 10 \\ 2 \mapsto 0 \end{cases}$$

$$x: \quad \ldots 2001210 \ldots$$
$$\sigma(x): \quad \ldots 0\ 001\ 001\ 10\ 0\ 10\ 001 \ldots$$

Link between $C_{x,\sigma}(n)$ and $C_{x,\sigma}(n+1)$

$$\sigma : \begin{cases} 0 \mapsto 001 \\ 1 \mapsto 10 \\ 2 \mapsto 0 \end{cases}$$

$$\begin{aligned} x : &\quad \ldots 2{\color{blue}00}1210 \ldots \\ \sigma(x) : &\quad \ldots 0\ {\color{orange}001}\ {\color{orange}0}01\ 10\ 0\ 10\ 001 \ldots \end{aligned}$$

- $(00, 0)$ is a covering of 0010

Link between $C_{x,\sigma}(n)$ and $C_{x,\sigma}(n+1)$

$$\sigma: \begin{cases} 0 \mapsto 001 \\ 1 \mapsto 10 \\ 2 \mapsto 0 \end{cases}$$

$$\begin{aligned} x: &\quad \ldots 2\,00\,1210 \ldots \\ \sigma(x): &\quad \ldots 0\ 001\ 001\ 10\ 0\ 10\ 001 \ldots \end{aligned}$$

- $(00, 0)$ is a covering of 0010 and of 00100

Link between $C_{x,\sigma}(n)$ and $C_{x,\sigma}(n+1)$

$$\sigma : \begin{cases} 0 \mapsto 001 \\ 1 \mapsto 10 \\ 2 \mapsto 0 \end{cases}$$

$$\begin{aligned} x :& \quad \dots 2001210 \dots \\ \sigma(x) :& \quad \dots 0\ 001\ 001\ 10\ 0\ 10\ 001 \dots \end{aligned}$$

- $(00, 0)$ is a covering of 0010 and of 00100
- $(121, 1)$ is a covering of 0010 but $(121, 1) \notin C_{x,\sigma}(5)$

Link between $C_{x,\sigma}(n)$ and $C_{x,\sigma}(n+1)$

$$\sigma : \begin{cases} 0 \mapsto 001 \\ 1 \mapsto 10 \\ 2 \mapsto 0 \end{cases} \qquad \begin{aligned} x : &\quad \ldots 200{\color{blue}1210}\ldots \\ \sigma(x) : &\quad \ldots 0\ 001\ 001\ 1{\color{orange}0}\ {\color{orange}0}\ {\color{orange}10}\ {\color{orange}0}01\ldots \end{aligned}$$

- $(00, 0)$ is a covering of 0010 and of 00100
- $(121, 1)$ is a covering of 0010 but $(121, 1) \notin C_{x,\sigma}(5)$
- $(1210, 1)$ is a covering of 00100 but $(1210, 1) \notin C_{x,\sigma}(4)$

Link between $\#C_{x,\sigma}(n)$ and $\#C_{x,\sigma}(n+1)$

We have

$$\#C(n+1) - \#C(n) = \sum_{(w,k)\in C(n)\setminus C(n+1)} (\#\{a : wa \in \mathcal{L}(x)\} - 1)$$

Link between $\#C_{x,\sigma}(n)$ and $\#C_{x,\sigma}(n+1)$

We have

$$\#C(n+1) - \#C(n) = \sum_{(w,k)\in C(n)\setminus C(n+1)} (\#\{a : wa \in \mathcal{L}(x)\} - 1)$$

If $(w,k) \in C(n) \setminus C(n+1)$, then $k + n = |\sigma(w)|$.

Link between $\#C_{x,\sigma}(n)$ and $\#C_{x,\sigma}(n+1)$

We have

$$\#C(n+1) - \#C(n) = \sum_{(w,k)\in C(n)\setminus C(n+1)} (\#\{a : wa \in \mathcal{L}(x)\} - 1)$$

If $(w,k) \in C(n) \setminus C(n+1)$, then $k + n = |\sigma(w)|$.

So

$$\#C(n+1) - \#C(n) = \sum_{w\in W_n} (\#\{a : wa \in \mathcal{L}(x)\} - 1)$$

where

$$W_n = \{w \in \mathcal{L}(x) : \exists\ k \text{ s.t } (w,k) \in C(n) \setminus C(n+1)\}$$

Link between $\#C_{x,\sigma}(n)$ and $\#C_{x,\sigma}(n+1)$

We have

$$\#C(n+1) - \#C(n) = \sum_{(w,k)\in C(n)\setminus C(n+1)} (\#\{a : wa \in \mathcal{L}(x)\} - 1)$$

If $(w,k) \in C(n) \setminus C(n+1)$, then $k + n = |\sigma(w)|$.

So

$$\#C(n+1) - \#C(n) = \sum_{w\in W_n} (\#\{a : wa \in \mathcal{L}(x)\} - 1)$$

where

$$\begin{aligned} W_n &= \{w \in \mathcal{L}(x) : \exists\ k \text{ s.t } (w,k) \in C(n) \setminus C(n+1)\} \\ &= \{w \in \mathcal{L}(x) : \exists\ 0 \leq k < |\sigma(w_1)| \text{ s.t } k + n = |\sigma(w)|\} \end{aligned}$$

Link between $\#C_{x,\sigma}(n)$ and $\#C_{x,\sigma}(n+1)$

We have

$$\#C(n+1) - \#C(n) = \sum_{(w,k)\in C(n)\setminus C(n+1)} (\#\{a : wa \in \mathcal{L}(x)\} - 1)$$

If $(w,k) \in C(n) \setminus C(n+1)$, then $k + n = |\sigma(w)|$.

So

$$\#C(n+1) - \#C(n) = \sum_{w\in W_n} (\#\{a : wa \in \mathcal{L}(x)\} - 1)$$

where

$$\begin{aligned} W_n &= \{w \in \mathcal{L}(x) : \exists\ k \text{ s.t } (w,k) \in C(n) \setminus C(n+1)\} \\ &= \{w \in \mathcal{L}(x) : \exists\ 0 \leq k < |\sigma(w_1)| \text{ s.t } k + n = |\sigma(w)|\} \\ &= \{w \in \mathcal{L}(x) : |\sigma(w_{[2,|w|]})| < n \leq |\sigma(w)|\} \end{aligned}$$

Extensions and complexities

The number of right, left and bi-extensions of w are respectively

$$r_x(w) = \#\{a \in \mathcal{A} : wa \in \mathcal{L}(x)\} \quad \ell_x(w) = \#\{b \in \mathcal{A} : bw \in \mathcal{L}(x)\}$$

$$b_x(w) = \#\{(a,b) \in \mathcal{A}^2 : awb \in \mathcal{L}(x)\}$$

Proposition (G. '23)

$$\#C(n+1) - \#C(n) = \sum_{w \in W_n} (r_x(w) - 1)$$

where $W_n = \{w \in \mathcal{L}(x) : |\sigma(w_{[2,|w|]})| < n \leq |\sigma(w)|\}.$

Extensions and complexities

The number of right, left and bi-extensions of w are respectively

$$r_x(w) = \#\{a \in \mathcal{A} : wa \in \mathcal{L}(x)\} \quad \ell_x(w) = \#\{b \in \mathcal{A} : bw \in \mathcal{L}(x)\}$$

$$b_x(w) = \#\{(a,b) \in \mathcal{A}^2 : awb \in \mathcal{L}(x)\}$$

Proposition (G. '23)

$$\#C(n+1) - \#C(n) = \sum_{w \in W_n} (r_x(w) - 1)$$

where $W_n = \{w \in \mathcal{L}(x) : |\sigma(w_{[2,|w|]})| < n \leq |\sigma(w)|\}$.

Proposition (Cassaigne '96)

$$p_x(n+1) - p_x(n) = \sum_{w \in \mathcal{L}_n(x)} (r_x(w) - 1) = \sum_{w \in \mathcal{L}_n(x)} (\ell_x(w) - 1)$$

The set W_n

$$W_n = \{w \in \mathcal{L}(x) : |\sigma(w_{[2,|w|]})| < n \leq |\sigma(w)|\}$$

The set W_n

$$W_n = \{w \in \mathcal{L}(x) : |\sigma(w_{[2,|w|]})| < n \leq |\sigma(w)|\}$$

- $W_n \subseteq \mathcal{L}_{\leq n}(x)$,

The set W_n

$$W_n = \{w \in \mathcal{L}(x) : |\sigma(w_{[2,|w|]})| < n \leq |\sigma(w)|\}$$

- $W_n \subseteq \mathcal{L}_{\leq n}(x)$,
- W_n is a *maximal suffix code*, i.e., any $u \in \mathcal{L}(x)$ either has a suffix in W_n or is a proper suffix of an element in W_n.

The set W_n

$$W_n = \{w \in \mathcal{L}(x) : |\sigma(w_{[2,|w|]})| < n \leq |\sigma(w)|\}$$

- $W_n \subseteq \mathcal{L}_{\leq n}(x)$,
- W_n is a *maximal suffix code*, i.e., any $u \in \mathcal{L}(x)$ either has a suffix in W_n or is a proper suffix of an element in W_n.

Proposition (G. '23)

If $W \subseteq \mathcal{L}_{\leq n}(x)$ is a maximal suffix code, then

$$\sum_{w \in W} (r_x(w) - 1) = p_x(n+1) - p_x(n) - \sum_{\substack{w \in \mathcal{L}_{<n}(x) \\ \mathrm{Suff}(w) \cap W \neq \emptyset}} m_x(w)$$

where $m_x(w) = b_x(w) - r_x(w) - \ell_x(w) + 1$.

Final result

Theorem (G. '23)

Let $x \in \mathcal{A}^{\mathbb{Z}}$ and $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$. For all $n \geq 0$, we have

$$p_{\sigma(x)}(n) \leq \# C_{x,\sigma}(n)$$

where

$$\# C_{x,\sigma}(n+1) - \# C_{x,\sigma}(n) = p_x(n+1) - p_x(n) - \sum_{\substack{w \in \mathcal{L}_{<n}(x) \\ |\sigma(w)| \geq n}} m_x(w).$$

Final result

Theorem (G. '23)

Let $x \in \mathcal{A}^{\mathbb{Z}}$ and $\sigma \colon \mathcal{A}^* \to \mathcal{B}^*$. For all $n \geq 0$, we have

$$p_{\sigma(x)}(n) \leq \# C_{x,\sigma}(n)$$

where

$$\# C_{x,\sigma}(n+1) - \# C_{x,\sigma}(n) = p_x(n+1) - p_x(n) - \sum_{\substack{w \in \mathcal{L}_{<n}(x) \\ |\sigma(w)| \geq n}} m_x(w).$$

Some observations:

- $C_{x,\sigma}(n)$ only depends on x and on the lengths $|\sigma(a)|$, $a \in \mathcal{A}$

Final result

Theorem (G. '23)

Let $x \in \mathcal{A}^{\mathbb{Z}}$ and $\sigma \colon \mathcal{A}^* \to \mathcal{B}^*$. For all $n \geq 0$, we have

$$p_{\sigma(x)}(n) \leq \# C_{x,\sigma}(n)$$

where

$$\# C_{x,\sigma}(n+1) - \# C_{x,\sigma}(n) = p_x(n+1) - p_x(n) - \sum_{\substack{w \in \mathcal{L}_{<n}(x) \\ |\sigma(w)| \geq n}} m_x(w).$$

Some observations:

- $C_{x,\sigma}(n)$ only depends on x and on the lengths $|\sigma(a)|$, $a \in \mathcal{A}$
- for any $x \in \mathcal{A}^{\mathbb{Z}}$ and any values of $|\sigma(a)|$, $a \in \mathcal{A}$, we can find a corresponding morphism σ such that $p_{\sigma(x)}(n) = \# C_{x,\sigma}(n)$

Corollary (G. '23)

For any $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$, *if* $x \in \mathcal{A}^{\mathbb{Z}}$ *is one of the following*

there exists a constant C *such that, for all* n,

$$p_{\sigma(x)}(n) \leq p_x(n) + C.$$

Corollary (G. '23)

For any $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$, if $x \in \mathcal{A}^{\mathbb{Z}}$ is one of the following

- *Sturmian or quasi-Sturmian,*

there exists a constant C such that, for all n,

$$p_{\sigma(x)}(n) \leq p_x(n) + C.$$

Corollary (G. '23)

For any $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$, if $x \in \mathcal{A}^{\mathbb{Z}}$ is one of the following

- *Sturmian or quasi-Sturmian,*
- *Arnoux-Rauzy or episturmian,*

there exists a constant C such that, for all n,

$$p_{\sigma(x)}(n) \leq p_x(n) + C.$$

Corollary (G. '23)

For any $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$, if $x \in \mathcal{A}^{\mathbb{Z}}$ is one of the following

- *Sturmian or quasi-Sturmian,*
- *Arnoux-Rauzy or episturmian,*
- *the coding of an interval exchange transformation,*

there exists a constant C such that, for all n,

$$p_{\sigma(x)}(n) \leq p_x(n) + C.$$

Corollary (G. '23)

For any $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$, if $x \in \mathcal{A}^{\mathbb{Z}}$ is one of the following

- *Sturmian or quasi-Sturmian,*
- *Arnoux-Rauzy or episturmian,*
- *the coding of an interval exchange transformation,*
- *dendric or eventually dendric,*

there exists a constant C such that, for all n,

$$p_{\sigma(x)}(n) \leq p_x(n) + C.$$

Corollary (G. '23)

For any $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$, if $x \in \mathcal{A}^{\mathbb{Z}}$ is one of the following

- *Sturmian or quasi-Sturmian,*
- *Arnoux-Rauzy or episturmian,*
- *the coding of an interval exchange transformation,*
- *dendric or eventually dendric,*
- *eventually weak (resp., strong) or neutral, i.e., for all long enough w, $m_x(w)$ is ≤ 0 (resp., ≥ 0),*

there exists a constant C such that, for all n,

$$p_{\sigma(x)}(n) \leq p_x(n) + C.$$

Corollary (G. '23)

For any $\sigma\colon \mathcal{A}^* \to \mathcal{B}^*$, if $x \in \mathcal{A}^{\mathbb{Z}}$ is one of the following

- *Sturmian or quasi-Sturmian,*
- *Arnoux-Rauzy or episturmian,*
- *the coding of an interval exchange transformation,*
- *dendric or eventually dendric,*
- *eventually weak (resp., strong) or neutral, i.e., for all long enough w, $m_x(w)$ is ≤ 0 (resp., ≥ 0),*
- *ternary Chacon word, or more generally, if for every long enough weak word, there is a strong enough word with the same Parikh vector*

there exists a constant C such that, for all n,

$$p_{\sigma(x)}(n) \leq p_x(n) + C.$$

What about the Thue-Morse word?

Proposition (G. '23)

If x is the Thue-Morse word, then for all $\sigma\colon \{0,1\}^* \to \mathcal{B}^*$ and all large enough n,

$$\#C_{x,\sigma}(n+1) - \#C_{x,\sigma}(n) \in p_x(n+1) - p_x(n) + \{2,-2,0\},$$

and the choice of 2, -2 or 0 only depends on n and $|\sigma(0)| + |\sigma(1)|$.

What about the Thue-Morse word?

Proposition (G. '23)

If x is the Thue-Morse word, then for all $\sigma\colon \{0,1\}^* \to \mathcal{B}^*$ and all large enough n,

$$\#C_{x,\sigma}(n+1) - \#C_{x,\sigma}(n) \in p_x(n+1) - p_x(n) + \{2, -2, 0\},$$

and the choice of 2, -2 or 0 only depends on n and $|\sigma(0)| + |\sigma(1)|$.

If $|\sigma(0)| + |\sigma(1)|$ is a power of 2, there exists C such that, for all n,

$$p_{\sigma(x)}(n) \leq p_x(n) + C.$$

Thank you for your attention!