

The Sum of Exponents of Maximal Repetitions in Standard Sturmian Words

Marcin Piątkowski

Faculty of Mathematics and Computer Science
Nicolaus Copernicus University

Prague Stringology Conference 2013



Sturmian words

Sturmian words are infinite words that has exactly $k + 1$ factors of length k for each $n \geq 0$.

Example:

ababaababababaababababaabababaababaab...

① *ababababababababababababababababababab...*

② *ababababababababababababababababababab...*

③ *ababababababababababababababababababab...*

④ *ababababababababababababababababababab...*



Standard word

- Alphabet: $\Sigma = \{a, b\}$
- Directive sequence: $\gamma = (\gamma_0, \gamma_1, \dots, \gamma_n)$,
where $\gamma_0 \geq 0$ and $\gamma_1, \dots, \gamma_n > 0$
- Recurrence:
 - $x_{-1} = b$
 - $x_0 = a$
 - $x_k = (x_{k-1})^{\gamma_{k-1}} \cdot x_{k-2}$
- $\text{Sw}(\gamma_0, \gamma_1, \dots, \gamma_n) = x_{n+1}$



Standard word example

- $\gamma = (1, 2, 1, 3, 1)$



Standard word example

- $\gamma = (1, 2, 1, 3, 1)$
- $x_{-1} = b$
- $x_0 = a$



Standard word example

- $\gamma = (1, 2, 1, 3, 1)$
- $x_{-1} = b$
- $x_0 = a$
- $x_1 = (x_0)^1 \cdot x_{-1} = a \cdot b$



Standard word example

- $\gamma = (1, 2, 1, 3, 1)$
- $x_{-1} = b$
- $x_0 = a$
- $x_1 = (x_0)^1 \cdot x_{-1} = a \cdot b$
- $x_2 = (x_1)^2 \cdot x_0 = ab \cdot ab \cdot a$



Standard word example

- $\gamma = (1, 2, 1, 3, 1)$
- $x_{-1} = b$
- $x_0 = a$
- $x_1 = (x_0)^1 \cdot x_{-1} = a \cdot b$
- $x_2 = (x_1)^2 \cdot x_0 = ab \cdot ab \cdot a$
- $x_3 = (x_2)^1 \cdot x_1 = ababa \cdot ab$



Standard word example

- $\gamma = (1, 2, 1, 3, 1)$
- $x_{-1} = b$
- $x_0 = a$
- $x_1 = (x_0)^1 \cdot x_{-1} = a \cdot b$
- $x_2 = (x_1)^2 \cdot x_0 = ab \cdot ab \cdot a$
- $x_3 = (x_2)^1 \cdot x_1 = ababa \cdot ab$
- $x_4 = (x_3)^3 \cdot x_2 = ababaab \cdot ababaab \cdot ababaab \cdot ababa$



Standard word example

- $\gamma = (1, 2, 1, 3, \mathbf{1})$
- $x_{-1} = b$
- $x_0 = a$
- $x_1 = (x_0)^1 \cdot x_{-1} = a \cdot b$
- $x_2 = (x_1)^2 \cdot x_0 = ab \cdot ab \cdot a$
- $x_3 = (x_2)^1 \cdot x_1 = ababa \cdot ab$
- $x_4 = (x_3)^3 \cdot x_2 = ababaab \cdot ababaab \cdot ababaab \cdot ababa$
- $x_5 = (x_4)^{\mathbf{1}} \cdot x_3 = ababaabababaabababaabababa \cdot ababaab$



Standard word example

- $\gamma = (1, 2, 1, 3, 1)$
- $x_{-1} = b$
- $x_0 = a$
- $x_1 = (x_0)^1 \cdot x_{-1} = a \cdot b$
- $x_2 = (x_1)^2 \cdot x_0 = ab \cdot ab \cdot a$
- $x_3 = (x_2)^1 \cdot x_1 = ababa \cdot ab$
- $x_4 = (x_3)^3 \cdot x_2 = ababaab \cdot ababaab \cdot ababaab \cdot ababa$
- $x_5 = (x_4)^1 \cdot x_3 = ababaabababaabababaabababa \cdot ababaab$
- $\text{Sw}(1, 2, 1, 3, 1) = ababaabababaabababaabababaabababaab$



Fibonacci words

- $F_{-1} = b$
- $F_0 = a$
- \vdots
- $F_{n+1} = F_n \cdot F_{n-1}$



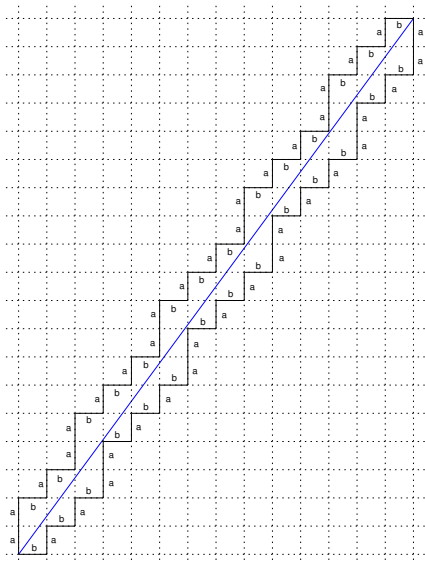
Fibonacci words

- $F_{-1} = b$
- $F_0 = a$
- \vdots
- $F_{n+1} = F_n \cdot F_{n-1}$

Observation

$$F_n = \text{Sw}(\underbrace{1, 1, \dots, 1, 1}_n)$$





Lower Christoffel word:

$bababaabababaabababaabababaabababa$

Upper Christoffel word:

$aababaabababaabababaabababaababab$

Standard words:

$ababaabababaabababaabababaabababa$ ***ab***

$ababaabababaabababaabababaabababa$ ***ba***

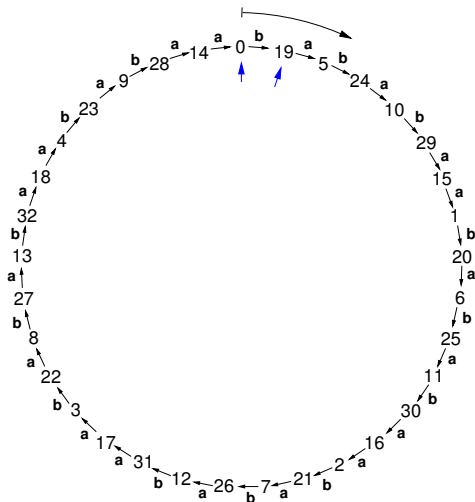


Proposition

For fixed relatively prime $p > q > 0$ there exists a unique word x such that:

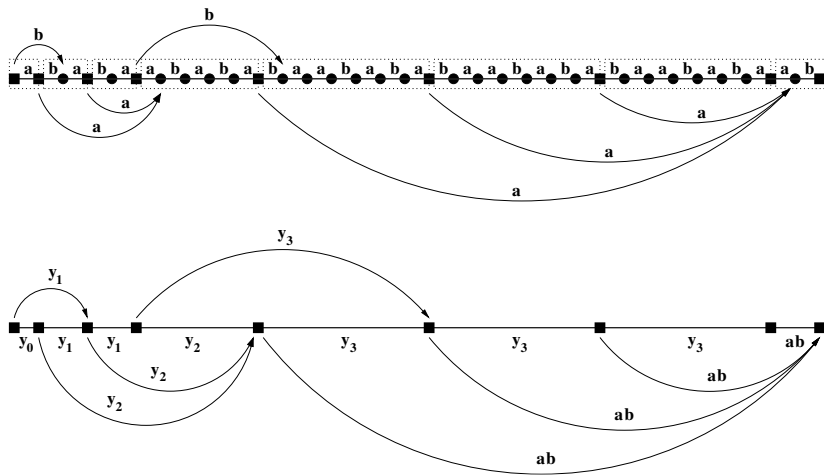
- $|x|_a = p - 1$
- $|x|_b = q - 1$
- $b \cdot x \cdot a$ – lower Christoffel word
- $a \cdot x \cdot b$ – upper Christoffel word
- $x \cdot ab$ – standard word
- $x \cdot ba$ – standard word





b a b a b a a b a b a b a a b a b a b a a b a b a a b a b a a





Notation: $y_i = (x_i)^R$



Run (maximal repetition)

ababaabababaabababaabababaababab

Details:

- starting position
- period (generator)
- exponent
- remainder
- length



Run (maximal repetition)

ababaabababaabababaabababaabababab

Details:

- starting position
- period (generator)
- exponent
- remainder
- length



Run (maximal repetition)

ababaabababaabababaabababaabababaab

Details:

- starting position
- period (generator)
- exponent
- remainder
- length



Run (maximal repetition)

ababaabababaabababaabababaabababab

Details:

- starting position
- period (generator)
- **exponent**
- remainder
- length



Run (maximal repetition)

ababaabababaabababaabababaab**aba**ab

Details:

- starting position
- period (generator)
- exponent
- remainder
- length



Run (maximal repetition)

ababaabababaabababaabababaabababaab

Details:

- starting position
- period (generator)
- exponent
- remainder
- **length**



Measures of periodicity

- 1 The number of runs
- 2 The sum of exponents of runs
- 3 The total run length



The number of runs

a b a b a a b a b a b a a b a b a b a a b a b a b a a b a b a a b

a b a b a a b a b a b a a b a b a b a b a a b a b a b a a b a b a a b

a b a b a a b a b a b a a b a b a b a a b a b a b a a b a b a a b

a b a b a a b a b a b a a b a b a b a a b a b a b a a b a b a a b

Number of runs = 19



The number of runs

Notation:

- $runs(w)$ – the number of runs in the word w
- $runs(n) = \max\{runs(w) : |w| = n\}$

Conjecture (Kolpakov, Kucherov 1999)

$$runs(n) \leq n$$



The number of runs

Theorem (Crochemore et al. 2008, Kusano et al. 2008)

$$0.944542 n \leq \text{runs}(n) \leq 1.048 n$$

Theorem (Crochemore et al. 2010)

$$0.41 n \leq \text{cubic_runs}(n) \leq 0.5 n$$



Standard words

Standard word

- $\gamma = (\gamma_0, \gamma_1, \dots, \gamma_n)$
- $w = \text{Sw}(\gamma)$

Runs (Baturo, Piatkowski, Rytter 2008)

- Compact formula for $\text{runs}(w)$.
- Algorithm for $\text{runs}(w)$ with time complexity $O(|\gamma|)$.
- $\lim_{|w| \rightarrow \infty} \text{runs}(w) = 0.8 |w|$.



Standard words

Standard word

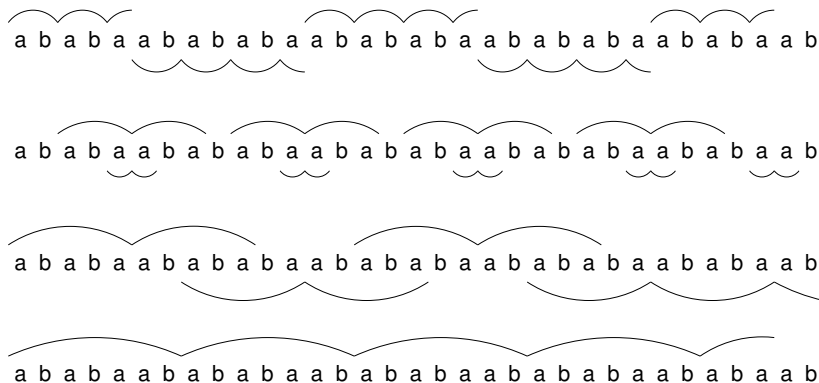
- $\gamma = (\gamma_0, \gamma_1, \dots, \gamma_n)$
- $w = \text{Sw}(\gamma)$

Cubic runs (Piatkowski, Rytter 2011)

- Compact formula for $\text{cubic_runs}(w)$.
- Algorithm for $\text{cubic_runs}(w)$ with time complexity $O(|\gamma|)$.
- $\lim_{|w| \rightarrow \infty} \text{cubic_runs}(w) \approx 0.36924841 |w|$.



The sum of exponents



$$\text{Sum of exponents} = 49 \frac{23}{70} \approx 49.3286$$



The sum of exponents

Notation:

- $se(w)$ – the sum of exponents in the word w
- $se(n) = \max\{sum_exp(w) : |w| = n\}$

Conjecture (Kolpakov, Kucherov 1999)

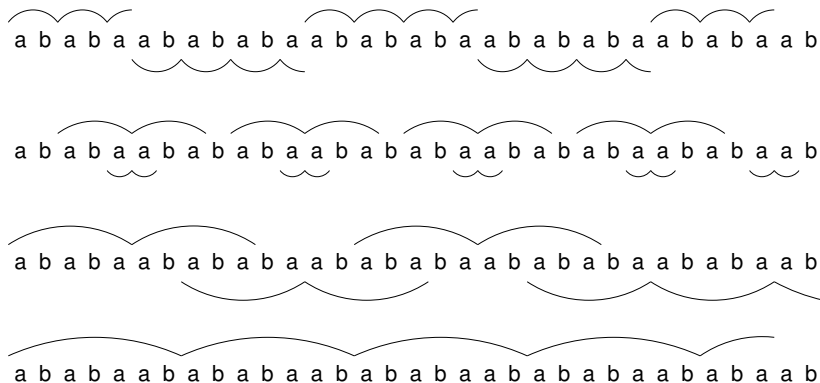
$$se(n) \leq 2n$$

Theorem (Crochemore et al. 2011)

$$2.035 n \leq se(n) \leq 4.1 n$$



The total run length



Total run length = 138



The total run length

Notation:

- $trl(w)$ – the total run length of the word w
- $trl(n) = \max\{trl(w) : |w| = n\}$

Theorem (Glen, Simpson 2013)

$$\frac{n^2}{8} < trl(n) < \frac{47n^2}{72} + 2n$$



Morphisms

$$\gamma = (\gamma_0, \gamma_1, \dots, \gamma_n) \quad h_i : \begin{cases} a \longrightarrow a^{\gamma_i} b \\ b \longrightarrow a \end{cases}, \quad \text{for } 0 \leq i \leq n$$



Morphisms

$$\gamma = (\gamma_0, \gamma_1, \dots, \gamma_n) \quad h_i : \begin{cases} a \longrightarrow a^{\gamma_i} b \\ b \longrightarrow a \end{cases}, \quad \text{for } 0 \leq i \leq n$$

Lemma

- 1 $\text{Sw}(\gamma_n) = h_n(a)$
- 2 $\text{Sw}(\gamma_i, \gamma_{i+1}, \dots, \gamma_n) = h_i(\text{Sw}(\gamma_{i+1}, \gamma_{i+2}, \dots, \gamma_n))$



Morphisms

$$\gamma = (\gamma_0, \gamma_1, \dots, \gamma_n) \quad h_i : \begin{cases} a \rightarrow a^{\gamma_i} b \\ b \rightarrow a \end{cases}, \quad \text{for } 0 \leq i \leq n$$

Lemma

- 1 $\text{Sw}(\gamma_n) = h_n(a)$
- 2 $\text{Sw}(\gamma_i, \gamma_{i+1}, \dots, \gamma_n) = h_i(\text{Sw}(\gamma_{i+1}, \gamma_{i+2}, \dots, \gamma_n))$

Observation

- 1 $\text{Sw}(\gamma_0, \gamma_1, \dots, \gamma_n) = h_0 \circ h_1 \circ \dots \circ h_n(a)$
- 2 $\text{Sw}(\gamma_{i+1}, \gamma_{i+2}, \dots, \gamma_n) = h_i^{-1}(\text{Sw}(\gamma_i, \gamma_{i+1}, \dots, \gamma_n))$



Example

$$\gamma = (1, 2, 1, 3, 1) \quad h_i : \begin{cases} a \longrightarrow a^{\gamma_i} b \\ b \longrightarrow a \end{cases} \quad \text{for } 0 \leq i \leq n$$



Example

$$\gamma = (1, 2, 1, 3, \mathbf{1}) \quad h_4 : \begin{cases} a \longrightarrow ab \\ b \longrightarrow a \end{cases}$$

$$\text{Sw}(1) = h_4(a) = ab$$



Example

$$\gamma = (1, 2, 1, \mathbf{3}, 1) \quad h_3 : \begin{cases} a \longrightarrow aaab \\ b \longrightarrow a \end{cases}$$

$$\text{Sw}(1) = h_4(a) = ab$$

$$\text{Sw}(3, 1) = h_3(\text{Sw}(1)) = aaaba$$



Example

$$\gamma = (1, 2, \mathbf{1}, 3, 1) \quad h_2 : \begin{cases} a \longrightarrow ab \\ b \longrightarrow a \end{cases}$$

$$\text{Sw}(1) = h_4(a) = ab$$

$$\text{Sw}(3, 1) = h_3(\text{Sw}(1)) = aaaba$$

$$\text{Sw}(1, 3, 1) = h_2(\text{Sw}(3, 1)) = abababaab$$



Example

$$\gamma = (1, 2, 1, 3, 1) \quad h_1 : \begin{cases} a \longrightarrow aab \\ b \longrightarrow a \end{cases}$$

$$\text{Sw}(1) = h_4(a) = ab$$

$$\text{Sw}(3, 1) = h_3(\text{Sw}(1)) = aaaba$$

$$\text{Sw}(1, 3, 1) = h_2(\text{Sw}(3, 1)) = abababaab$$

$$\text{Sw}(2, 1, 3, 1) = h_1(\text{Sw}(1, 3, 1)) = aabaaabaaabaabaaba$$



Example

$$\gamma = (1, 2, 1, 3, 1) \quad h_0 : \begin{cases} a \longrightarrow ab \\ b \longrightarrow a \end{cases}$$

$$\text{Sw}(1) = h_4(a) = ab$$

$$\text{Sw}(3, 1) = h_3(\text{Sw}(1)) = aaaba$$

$$\text{Sw}(1, 3, 1) = h_2(\text{Sw}(3, 1)) = abababaab$$

$$\text{Sw}(2, 1, 3, 1) = h_1(\text{Sw}(1, 3, 1)) = aabaaabaaabaabaaba$$

$$\text{Sw}(1, 2, 1, 3, 1) = h_0(\text{Sw}(2, 1, 3, 1)) = ababaabababaabababaabababaabaab$$

Notation: $N_\gamma = |\text{Sw}(\gamma_1, \dots, \gamma_n)|_a$



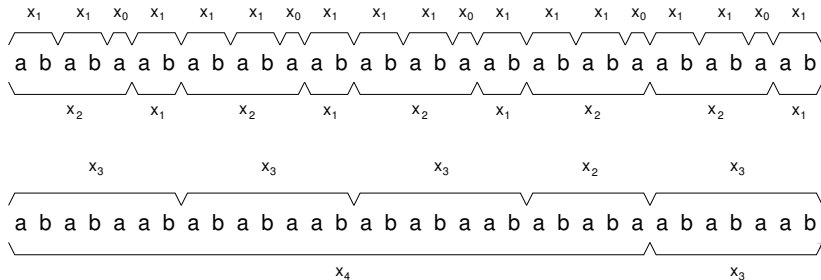
m-partition

Every standard word $Sw(\gamma_0, \dots, \gamma_n)$ can be represented in one of the forms:

$$(i) \quad x_m^{\alpha_1} x_{m-1} x_m^{\alpha_2} x_{m-1} \cdots x_m^{\alpha_s} x_{m-1} x_m$$

$$(ii) \quad x_m^{\beta_1} x_{m-1} x_m^{\beta_2} x_{m-1} \cdots x_m^{\beta_s} x_{m-1}$$

where $\alpha_k, \beta_k \in \{\gamma_m, \gamma_m + 1\}$, $0 \leq m \leq n$.



Proposition

The structure of occurrences of the block x_m (respectively x_{m-1}) in the m -partition of $\text{Sw}(\gamma_0, \dots, \gamma_n)$ corresponds to the structure of occurrences of the letter a (respectively letter b) in $\text{Sw}(\gamma_m, \dots, \gamma_n)$.

m	m -partition	$\text{Sw}(\gamma_m, \dots, \gamma_n)$
1	$ab \cdot ab \cdot a \cdot ab \cdot ab \cdot ab \cdot a \cdot ab \cdot ab \cdot ab \cdot a \cdot ab \cdot ab \cdot ab \cdot a \cdot ab \cdot ab \cdot a \cdot ab$	$aabaaabaaabaabaaba$
2	$ababa \cdot ab \cdot ababa \cdot ab \cdot ababa \cdot ab \cdot ababa \cdot ababa \cdot ab$	$abababaab$
3	$ababaab \cdot ababaab \cdot ababaab \cdot ababa \cdot ababaab$	$aaaba$
4	$ababaabababaabababaabababa \cdot ababaab$	ab



Lemma

- 1 The longest prefix of $x_{i-1} \cdot x_i$ with the period of the length $|x_i|$ is of the form $x_i \cdot \widehat{x_{i-1}}$.

Notation: $\widehat{x} - x$ with last two letters removed.



Lemma

- ① The longest prefix of $x_{i-1} \cdot x_i$ with the period of the length $|x_i|$ is of the form $x_i \cdot \widehat{x_{i-1}}$.
- ② For each occurrence of $y \cdot x_i$ in w , y is the last letter of the block x_{i-1} or x_i of the i -partition of w .

Notation: \widehat{x} – x with last two letters removed.



Lemma

- ① The longest prefix of $x_{i-1} \cdot x_i$ with the period of the length $|x_i|$ is of the form $x_i \cdot \widehat{x_{i-1}}$.
- ② For each occurrence of $y \cdot x_i$ in w , y is the last letter of the block x_{i-1} or x_i of the i -partition of w .
- ③ For every $0 \leq k \leq n$ and every $1 \leq i \leq \gamma_k$ the word $(x_k)^i x_{k-1}$ is primitive.

Notation: \widehat{x} – x with last two letters removed.



Lemma

- 1 The longest prefix of $x_{i-1} \cdot x_i$ with the period of the length $|x_i|$ is of the form $x_i \cdot \widehat{x_{i-1}}$.
- 2 For each occurrence of $y \cdot x_i$ in w , y is the last letter of the block x_{i-1} or x_i of the i -partition of w .
- 3 For every $0 \leq k \leq n$ and every $1 \leq i \leq \gamma_k$ the word $(x_k)^i x_{k-1}$ is primitive.
- 4 The period of each maximal repetition in a standard word $\text{Sw}(\gamma_0, \gamma_1, \dots, \gamma_n)$ is of the form x_i or $(x_i)^j x_{i-1}$, where $0 \leq i \leq n$, $0 < j < \gamma_i$.

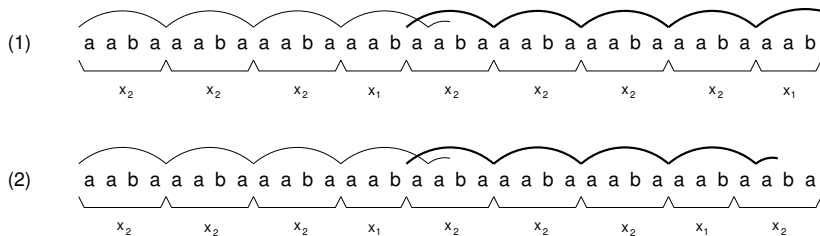
Notation: \widehat{x} – x with last two letters removed.



Computation of the sum of exponents (the total run length)

- $Sw(\gamma)$ – standard word
- $se'_i(\gamma)$ – sum of exponents of runs with periods x_i
- $se''_i(\gamma)$ – sum of exponents of runs with periods $(x_i)^j x_{i-1}$
- $trl'_i(\gamma)$ – total length of runs with periods x_i
- $trl''_i(\gamma)$ – total length of runs with periods $(x_i)^j x_{i-1}$
- $se(\gamma) = \sum_{i=0}^n se'_i(\gamma) + se''_i(\gamma)$
- $trl(\gamma) = \sum_{i=0}^n trl'_i(\gamma) + trl''_i(\gamma)$





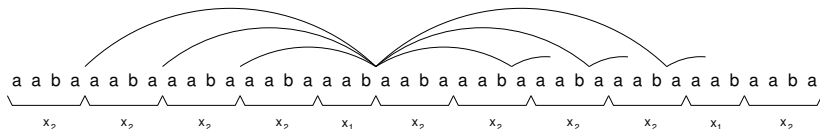
Lemma – sum of exponents

$$se'_i(\gamma) = N_\gamma(i+1) \cdot \left(\gamma_i + 1 + \frac{|x_{i-1}| - 2}{|x_i|} \right) + \left(N_\gamma(i+2) - 1 \right) + \Delta_n(i) \frac{2}{|x_i|}$$

Notation: $\Delta_n(i) = |n - i + 1| \bmod 2$

$$N_\gamma = |\text{Sw}(\gamma_i, \dots, \gamma_n)|_a$$



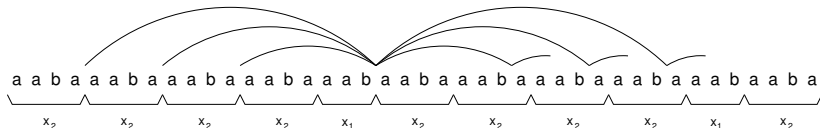
Runs with period $(x_i)^k x_{i-1}$ 

Lemma – sum of exponents

$$se''_i(\gamma) = \left(N_\gamma(i+1) - 1 \right) \cdot \sum_{k=1}^{\gamma_i-1} \left(2 + \frac{|x_i| - 2}{k \cdot |x_i| + |x_{i-1}|} \right)$$

Notation: $N_\gamma = |\text{Sw}(\gamma_1, \dots, \gamma_n)|_a$



Runs with period $(x_i)^k x_{i-1}$ 

Lemma – sum of exponents

$$se''_i(\gamma) = \left(N_\gamma(i+1) - 1 \right) \cdot \sum_{k=1}^{\gamma_i-1} \left(2 + \frac{|x_i| - 2}{k \cdot |x_i| + |x_{i-1}|} \right)$$

Lemma – total run length

$$trl''_i(\gamma) = \left(N_\gamma(i+1) - 1 \right) \cdot (\gamma_i - 1) \frac{(\gamma_i + 2)|x_i| + 2|x_i| - 4}{2}$$

Notation: $N_\gamma = |\text{Sw}(\gamma_1, \dots, \gamma_n)|_a$



Algorithm

- 1 $se \leftarrow 0$ ($trl \leftarrow 0$)
- 2 $|x_{-1}| \leftarrow 1$, $|x_0| \leftarrow 0$
- 3 $N_\gamma(n+1) \leftarrow 1$,
- 4 $N_\gamma(n+2) \leftarrow 0$
- 5 **for** $k := 0$ **to** n **do**
- 6 $|x_k| \leftarrow \gamma_k |x_{k-1}| + |x_{k-2}|$
- 7 **for** $k := n$ **downto** 0 **do**
- 8 update se (respectively trl) values depending on γ_k
- 9 update value of $|x_{k-1}|$
- 10 update value of $N_\gamma(k)$
- 11 **return** se (respectively trl)



Time complexity

- Sum of exponents – $O(\|\gamma\|)$, where $\|\gamma\| = \gamma_0 + \gamma_1 + \dots + \gamma_n$
- Total run length – $O(|\gamma|)$



Thank You
For Your Attention

