On the balance of d-bonacci word

Karel Břinda
karel.brinda@fjfi.cvut.cz

joint work with Edita Pelantová and Ondřej Turek

Theoretical Informatics Group
Faculty of Nuclear Sciences and Physical Engineering
Czech Technical University in Prague

September 27, 2012

Introduction

Notation

$\mathcal{A} = \{0, \dots, d-1\}$	alphabet
$\mathbf{u} = \mathbf{u}_0\mathbf{u}_1\mathbf{u}_2\cdots$	infinite word
v, w	finite words
$\mathbf{u}_{[n,n+k)} = \mathbf{u}_n\mathbf{u}_{n+1}\cdots\mathbf{u}_{n+k-1}$	factor of the infinite word $\mathbf{u}$
$\mathcal{L}_n(\mathbf{u})$	the set of factors of $\mathbf{u}$ of length n
$\mu_a = \lim_{n\to+\infty} \frac{\lvert\mathbf{u}_{[0,n)}\rvert_a}{n}$	frequency of the letter $a \in \mathcal{A}$ in $\mathbf{u}$
$\mu = (\mu_a)_{a\in\mathcal{A}}$	vector of frequencies
$\Psi(v) = \begin{pmatrix} \lvert v\rvert_0 \\ \vdots \\ \lvert v\rvert_{d-1} \end{pmatrix}$	Parikh vector of v

d-bonacci word

d-bonacci word is the unique fixed point of the substitution

$$\varphi : \begin{cases} 0 & \to & 01 \\ 1 & \to & 02 \\ & \vdots & \\ (d-2) & \to & 0(d-1) \\ (d-1) & \to & 0 \end{cases}$$

d-bonacci word

d-bonacci word is the unique fixed point of the substitution

$$\varphi : \begin{cases} 0 & \to & 01 \\ 1 & \to & 02 \\ & \vdots & \\ (d-2) & \to & 0(d-1) \\ (d-1) & \to & 0 \end{cases}$$

Fibonacci word $=$ 2-bonacci word

Tribonacci word $=$ 3-bonacci word

d-bonacci word

d-bonacci word is the unique fixed point of the substitution

$$\varphi : \begin{cases} 0 & \to & 01 \\ 1 & \to & 02 \\ & \vdots & \\ (d-2) & \to & 0(d-1) \\ (d-1) & \to & 0 \end{cases}$$

Fibonacci word $=$ 2-bonacci word

Tribonacci word $=$ 3-bonacci word

Example (5-bonacci)

$$0102010301020104010201030102010010201030102010401020\cdots$$

Balance function

Definition

Balance function of an infinite word $\mathbf{u}$ is the function defined as

$$B_a(n) = \max_{v,w \in \mathcal{L}_n(\mathbf{u})} \{|v|_a - |w|_a\}.$$

A word $\mathbf{u}$ is said to be: **c-balanced** for some c if $B_a(n)$ is bounded by the constant c for all letters $a \in \mathcal{A}$;

Balance function

Definition

Balance function of an infinite word $\mathbf{u}$ is the function defined as

$$B_a(n) = \max_{v,w \in \mathcal{L}_n(\mathbf{u})} \{|v|_a - |w|_a\}.$$

A word $\mathbf{u}$ is said to be: **c-balanced** for some c if $B_a(n)$ is bounded by the constant c for all letters $a \in \mathcal{A}$;

Our aim

Our aim is to **estimate maximum of the balance function $B_a(n)$ of the d-bonacci word** in dependence on the letter a and the constant d.

Balance function

Definition

Balance function of an infinite word $\mathbf{u}$ is the function defined as

$$B_a(n) = \max_{v,w \in \mathcal{L}_n(\mathbf{u})} \{|v|_a - |w|_a\}.$$

A word $\mathbf{u}$ is said to be: **c-balanced** for some c if $B_a(n)$ is bounded by the constant c for all letters $a \in \mathcal{A}$;

Our aim

Our aim is to **estimate maximum of the balance function $B_a(n)$ of the d-bonacci word** in dependence on the letter a and the constant d.

Problem: It is almost impossible to compute it directly from definition. For higher d, number of factors which should be directly compared is very high and the procedure would be too time-consuming.

Already known facts

Theorem

The Fibonacci word is 1*-balanced.*

Already known facts

Theorem

The Fibonacci word is 1*-balanced.*

Theorem (Richomme, Saari, Zamboni 2010)

The Tribonacci word is 2*-balanced.*

Already known facts

Theorem

The Fibonacci word is 1*-balanced.*

Theorem (Richomme, Saari, Zamboni 2010)

The Tribonacci word is 2*-balanced.*

Conjecture (Richomme, Saari, Zamboni 2010)

The d*-bonacci word is* $(d-1)$*-balanced.*

Already known facts

Theorem

The Fibonacci word is 1-balanced.

Theorem (Richomme, Saari, Zamboni 2010)

The Tribonacci word is 2-balanced.

Conjecture (Richomme, Saari, Zamboni 2010)

The d-bonacci word is $(d-1)$-balanced.

Our result:
For $d \leq 12$ the d-bonacci word is $(d-1)$-balanced and this bound can be diminished.

The main idea

Discrepancy function

Definition (discrepancy function of a prefix)

$$D_a(n) = |\mathbf{u}_{[0,n)}|_a - \mu_a n$$

Discrepancy function

Definition (discrepancy function of a prefix)

$$D_a(n) = |\mathbf{u}_{[0,n)}|_a - \mu_a n$$

Then:

$$B_a(n) = \max_{n=n_2-n_1=\tilde{n}_2-\tilde{n}_1} \big(D_a(n_2) - D_a(n_1) - D_a(\tilde{n}_2) + D_a(\tilde{n}_1)\big)$$

Discrepancy function

Definition (discrepancy function of a prefix)

$$D_a(n) = |\mathbf{u}_{[0,n)}|_a - \mu_a n$$

Then:

$$B_a(n) = \max_{n=n_2-n_1=\tilde{n}_2-\tilde{n}_1} \left(D_a(n_2) - D_a(n_1) - D_a(\tilde{n}_2) + D_a(\tilde{n}_1)\right)$$

Hence:

$$B_a(n) \leq \left\lfloor 2\left(\sup_{n\in\mathbb{Z}^+} D_a(n) - \inf_{n\in\mathbb{Z}^+} D_a(n)\right)\right\rfloor .$$

Our method

Follows from ideas in the articles

- B. Adamczewski. Balances for fixed points of primitive substitutions. *Theor. Comput. Sci.* **307**, p. 47–75, 2003.
- B. Adamczewski. Symbolic discrepancy and self-similar dynamics. *Ann. de l'inst. Four.* **54**, p. 2201–2234, 2004.
- G. Richomme, K. Saari, and L. Q. Zamboni. Balance and Abelian complexity of the Tribonacci word. *Adv. in App. Math.* **45**, p. 212–231, 2010.

We express the discrepancy function as a sum which we estimate numerically.

For all letters $a \in \mathcal{A}$, we define the row vectors

$$f^{(a)} = (0, \ldots, 0, \underbrace{1}_{a^{\text{th}} \text{ entry}}, 0, \ldots 0) - (\mu_a, \ldots, \mu_a)$$

For all letters $a \in \mathcal{A}$, we define the row vectors

$$f^{(a)} = (0, \ldots, 0, \underbrace{1}_{a^{\text{th}} \text{ entry}}, 0, \ldots 0) - (\mu_a, \ldots, \mu_a)$$

Then:

$$D_a(n) = f^{(a)} \cdot \Psi\left(|\mathbf{u}_{[0,n)}|\right)$$

For all letters $a \in \mathcal{A}$, we define the row vectors

$$f^{(a)} = (0, \ldots, 0, \underbrace{1}_{a^{\text{th}} \text{ entry}}, 0, \ldots 0) - (\mu_a, \ldots, \mu_a)$$

Then:

$$D_a(n) = f^{(a)} \cdot \Psi\left(|\mathbf{u}_{[0,n)}|\right)$$

Lemma (J.-M. Dumont and A. Thomas, 1989; G. Rauzy, 1990)

For all finite integer sequences $(i_j)_{j=0}^{N}$ *such that* $i_N > i_{N-1} > \ldots > i_1 > i_0 \geq 0$, *the word*

$$\varphi^{i_N}(0)\varphi^{i_{N-1}}(0) \cdots \varphi^{i_1}(0)\varphi^{i_0}(0)$$

is a prefix of the d*-bonacci word and any prefix can be written in this form.*

Proposition

$$D_a(n) = \sum_{j=0}^{i_N} \delta_j g_{(a,j)} \quad \textit{where:} \qquad \delta_j \in \{0,1\}$$

$$g_{(a,j)} = f^{(a)} \cdot M_\varphi^j \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Proposition

$$D_a(n) = \sum_{j=0}^{i_N} \delta_j g_{(a,j)} \quad \textit{where:} \quad \delta_j \in \{0,1\}$$

$$g_{(a,j)} = f^{(a)} \cdot M_\varphi^j \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Theorem

The d-bonacci word is c-balanced for $c = \max_{a\in\mathcal{A}} \left\lfloor 2 \cdot \sum_{i=0}^{+\infty} |g_{(a,i)}| \right\rfloor$.

Upper estimates

Upper estimates

Way of computing: $$\sum_{i=0}^{+\infty} |g_{(a,i)}| \leq \underbrace{\sum_{i=0}^{m-1} |g_{(a,i)}|}_{\text{i)}} + \underbrace{E}_{\text{ii)}}$$

Upper estimates

Way of computing: $$\sum_{i=0}^{+\infty} |g_{(a,i)}| \leq \underbrace{\sum_{i=0}^{m-1} |g_{(a,i)}|}_{\text{i)}} + \underbrace{E}_{\text{ii)}}$$

i) We use

$$g_{(a,j)} = T_{j+d-a-1} - \frac{T_{j+d}}{\beta^{a+1}}$$

where T_n is defined by the d-bonacci recurrence $T_n = \sum_{i=n-d}^{n-1} T_i$ with the initial conditions $T_0 = T_1 = \ldots = T_{d-2} = 0, \; T_{d-1} = 1$.

Upper estimates

Way of computing: $$\sum_{i=0}^{+\infty} |g_{(a,i)}| \leq \underbrace{\sum_{i=0}^{m-1} |g_{(a,i)}|}_{\text{i)}} + \underbrace{E}_{\text{ii)}}$$

ii) We use

$$g_{(a,j)} = \sum_{k=2}^{d} \left(\frac{1}{\beta_k^{a+1}} - \frac{1}{\beta^{a+1}} \right) \frac{\beta_k^{j+d}}{f'(\beta_k)}$$

where $f(x) = x^d - x^{d-1} - \ldots - 1$ is the characteristic polynomial of the incidence matrix of φ and $\beta, \beta_2, \ldots, \beta_d$ its roots such that β is the dominant root.

Obtained upper estimates of the optimal c

Example

d	a	i_{max}	$\sum_{i=0}^{i_{max}} \lvert g_{(a,i)} \rvert$	$E_{(a,i_{max}+1)}$	$B_a(n)$ **upp. est.**
8	7	103	2.45375	0.0410838	4

Obtained upper estimates of the optimal c

Example

d	a	i_{max}	$\sum_{i=0}^{i_{max}} \lvert g_{(a,i)} \rvert$	$E_{(a,i_{max}+1)}$	$B_a(n)$ **upp. est.**
8	7	103	2.45375	0.0410838	4

$d \setminus a$	0	1	2	3	4	5	6	7	8	9	10	11
2	1	1	×	×	×	×	×	×	×	×	×	×
3	2	2	2	×	×	×	×	×	×	×	×	×
4	2	3	3	3	×	×	×	×	×	×	×	×
5	2	3	3	3	3	×	×	×	×	×	×	×
6	3	3	4	4	4	4	×	×	×	×	×	×
7	3	4	4	4	4	4	4	×	×	×	×	×
8	3	4	4	4	4	4	4	4	×	×	×	×
9	3	4	5	5	5	5	5	5	5	×	×	×
10	3	5	5	5	5	5	5	5	5	5	×	×
11	4	5	5	6	6	6	6	6	6	6	6	×
12	4	5	6	6	6	6	6	6	6	6	6	6

Lower estimates

Lower estimates

The upper estimate provides us tips for factors rich in the letter a. Then we compare them with their neighbors.

Lower estimates

The upper estimate provides us tips for factors rich in the letter a. Then we compare them with their neighbors.

Example

d	a	i_{max}	n	n_1	n_2	$B_a(n)$ **low. est.**
8	7	77	14116643186476334729665	124459569654115865416319	138576212840592200145984	4

(We compare factors $\mathbf{u}_{[n_1,n_1+n)}$ and $\mathbf{u}_{[n_2,n_2+n)}$.)

Obtained lower estimates of the optimal c

$d \setminus a$	0	1	2	3	4	5	6	7	8	9	10	11
2	1	1	×	×	×	×	×	×	×	×	×	×
3	2	2	2	×	×	×	×	×	×	×	×	×
4	2	2	2	2	×	×	×	×	×	×	×	×
5	2	3	3	3	3	×	×	×	×	×	×	×
6	2	3	3	3	3	3	×	×	×	×	×	×
7	2	3	3	3	3	3	3	×	×	×	×	×
8	3	3	4	4	4	4	4	4	×	×	×	×
9	3	4	4	4	4	4	4	4	4	×	×	×
10	3	4	4	4	4	4	4	4	4	4	×	×
11	3	4	4	4	4	4	4	4	4	4	4	×
12	3	4	5	5	5	5	5	5	5	5	5	5

Lower estimates of the optimal c

Besides numerical lower estimates, we can use more:

Lower estimates of the optimal c

Besides numerical lower estimates, we can use more:

Proposition

For $d \geq 5$, there exist factors v and w of the same length such that $|v|_a - |w|_a = 3$ for all $a \in \{2, \ldots, d-1\}$.

Lower estimates of the optimal c

Besides numerical lower estimates, we can use more:

Proposition

For $d \geq 5$, there exist factors v and w of the same length such that $|v|_a - |w|_a = 3$ for all $a \in \{2, \ldots, d-1\}$.

Proposition (obtained by direct comparing of factors)

For $d = 4$ and all letters $a \in \{1, 2, 3\}$, there exist factors v and w of the same length such that

$$|v|_a - |w|_a = 3.$$

Conclusion

Obtained estimates of the optimal c

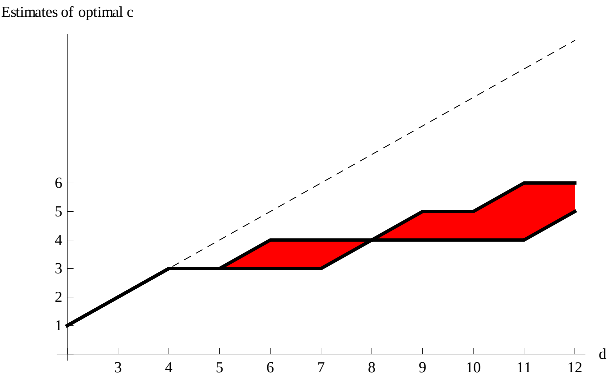

Figure: Estimates of the optimal constant c

Obtained estimates of the optimal c

Theorem

2-bonacci word is 1-balanced.
3-bonacci word is 2-balanced.
4-bonacci word is 3-balanced.
5-bonacci word is 3-balanced.
8-bonacci word is 4-balanced.
These constants cannot be improved.

Connection with Abelian complexity

Abelian complexity is defined as

$$\mathcal{AC}(n) = \#\{\Psi(w) \mid w \text{ is a factor of } \mathbf{u} \text{ of length } n\}$$

- Algorithm for computing provided in
  - O. Turek. Abelian complexity and Abelian co-decomposition. Preprint: *arXiv:1201.2109v1 [math.CO]*
- Estimate on bounds of $\mathcal{AC}(n)$ can be found from c by a formula provided in
  - S. Widmer. Topics in Word Complexity. Doctoral thesis, *University of Lyon*, 2010.

Open questions

1) Does our method provide by the upper estimate the optimal c?

2) If the answer is yes, does there exist an analytical expression of c in dependence on d?

Thank you for your attention!