On the balance of d-bonacci word

(Extended abstract)

K. Břinda*

August 22, 2012

1 Introduction

The notion of balanced words was already introduced by Morse and Hedlund in [4] where they studied properties of Sturmian words. Its generalization, so-called c-balanced words, appeared in the work [5]. An infinite word is said to be c-balanced, if for any two its factors v and w of the same length, we have $|v|_a - |w|_a \leq c$ for any letter $a \in \mathcal{A}$. As shown by Adamczewski, any fixed point of a Pisot-type substitution is c-balanced for some c.

In this text, we use techniques of [1, 2, 3] to find an upper estimate on the smallest possible value c for the d-bonacci word, i.e., a fixed point of the substitution

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

with the incidence matrix M_φ. For a given d and the associated d-bonacci word $\mathbf{u}$, our aim is to estimate the maximal value of a so-called balance function

$$B_a(n) = \max\left\{ |v|_a - |w|_a \;\middle|\; v, w \text{ are factors of } \mathbf{u} \text{ of length } n \right\}$$

where a is from the d-letter alphabet $\mathcal{A} = \{0, 1, \ldots, d-1\}$ and $n \in \mathbb{Z}^+$.

The characteristic polynomial of M_φ is $f(x) = x^d - x^{d-1} - \ldots - 1$. Denote its roots by $\beta = \beta_1,\ \beta_2,\ \ldots,\ \beta_d$ in the way that $|\beta| \geq |\beta_2| \geq \ldots |\beta_d|$. The number β is real, strictly greater than 1 and all other roots are in absolute value strictly less than 1.

The cases which have been already investigated are:

i) $d = 2$ (Fibonacci word) – a representative of the Sturmian words (i.e., aperiodic words having the lowest possible factor complexity) which have constant balance function $B_a(n) = 1$ for both letters a;

ii) $d = 3$ (Tribonacci word) – as proven in [3], $\max_{n\in\mathbb{Z}^+} B_a(n) = 2$ for all letters.

In [3], the conjecture that the balance function of the d-bonacci word is bounded by $d-1$ was stated. We will show that it holds for all $d \in \{2, \ldots, 11\}$ and that this bound can be diminished.

*Department of Mathematics, FNSPE, Czech Technical University in Prague, Trojanova 13, 120 00 Praha 2, Czech Republic, e-mail: karel.brinda@fjfi.cvut.cz

2 Upper estimates of $B_a(n)$

Let $\mathbf{u}$ be an infinite word. By $\mathbf{u}_{[n_1,n_2)}$ we denote its factor $\mathbf{u}_{n_1}\mathbf{u}_{n_1+1}\cdots\mathbf{u}_{n_2-1}$. We define the discrepancy function in a slightly more general way than it is usual – for an arbitrary factor of the studied word instead of only for its prefixes:

$$D_a(n_1,n_2) = |\mathbf{u}_{[n_1,n_2)}|_a - \mu_a(n_2-n_1) = \underbrace{(|\mathbf{u}_{[0,n_2)}|_a - \mu_a n_2)}_{=:D_a(n_2)} - \underbrace{(|\mathbf{u}_{[0,n_1)}|_a - \mu_a n_1)}_{=:D_a(n_1)}$$

where $\mu_a = \lim_{N\to+\infty}\frac{|\mathbf{u}_{[0,N)}|_a}{N}$ is the frequency of the letter a in $\mathbf{u}$. For a given factor $\mathbf{u}_{[n_1,n_2)}$, the discrepancy function describes how much the actual count of the letter a in the factor does differ from the count expected from the letter frequency μ_a.

Since $B_a(n) = \max\left\{D_a(n_1,n_2) - D_a(\tilde{n}_1,\tilde{n}_2) \;\middle|\; n_2-n_1 = \tilde{n}_2-\tilde{n}_1\right\}$, we obtain the following estimate.

$$B_a(n) \leq \left\lfloor 2\underbrace{\left(\sup\left\{D_a(n) \;\middle|\; n\in\mathbb{Z}^+\right\} - \inf\left\{D_a(n) \;\middle|\; n\in\mathbb{Z}^+\right\}\right)}_{=:\Delta D_a}\right\rfloor.$$

Now, for all letters $a\in\mathcal{A}$, let us define row vectors

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

which enable us to express the discrepancy function of prefixes as

$$D_a(n) = f^{(a)}\cdot v$$

where v is a column vector with components $v_i = |\mathbf{u}_{[0,n)}|_i$ for $i\in\{0,\ldots,d-1\}$

As follows from [6, 7], for all finite integer sequences $(i_j)_{j=0}^N$ such that $i_N > i_{N-1} > \ldots > i_1 > i_0 \geq 0$ and the d-bonacci substitution φ, the word

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

is a prefix of the d-bonacci word and any prefix can be written in this form. Hence, the discrepancy function of every prefix of the d-bonacci word may be expressed as

$$D_a(n) = \sum_{j=0}^{N} f^{(a)}\cdot g_{(a,i_j)} = \sum_{j=0}^{i_N}\delta_j f^{(a)}\cdot g_{(a,j)}$$

for some $(\delta_0,\ldots,\delta_{i_N})\in\{0,1\}^{i_N+1}$ where

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

Remark that $g_{(a,j)}$ depends on d. The formula provides bounds on $D_a(n)$:

$$\text{sum of negative } g_{(a,i)}\text{'s} \leq D_a(n) \leq \text{sum of positive } g_{(a,i)}\text{'s}.$$

Hence, $\Delta D_a \leq \sum_{i=0}^{+\infty}|g_{(a,i)}|$.

Theorem 2.1 *The d-bonacci word is c-balanced for*

$$c = \max_{a\in\mathcal{A}}\left\lfloor 2\cdot\sum_{i=0}^{+\infty}|g_{(a,i)}|\right\rfloor.$$

	IC of $(g_{(a,i)})_{i=0}^{3}$	$a=0$	$a=1$	$a=2$	$a=3$
$g_{(a,0)}$	$(1,0,0,0)$	$+$	$-$	$-$	$-$
$g_{(a,1)}$	$(0,1,0,0)$	$-$	$+$	$-$	$-$
$g_{(a,2)}$	$(0,0,1,0)$	$-$	$-$	$+$	$-$
$g_{(a,3)}$	$(0,0,0,1)$	$-$	$-$	$-$	$+$
$g_{(a,4)}$	$(1,1,1,1)$	$+$	$-$	$-$	$-$
$g_{(a,5)}$	$(1,2,2,2)$	$-$	$+$	$-$	$-$
$g_{(a,6)}$	$(2,3,4,4)$	$-$	$-$	$+$	$-$
$g_{(a,7)}$	$(4,6,7,8)$	$-$	$-$	$-$	$+$
$g_{(a,8)}$	$(8,12,14,15)$	$+$	$+$	$-$	$-$
$g_{(a,9)}$	$(15,23,27,29)$	$-$	$+$	$+$	$-$
$g_{(a,10)}$	$(29,44,52,56)$	$-$	$-$	$+$	$+$
$g_{(a,11)}$	$(56,85,100,108)$	$+$	$-$	$-$	$+$
$g_{(a,12)}$	$(108,164,193,208)$	$+$	$+$	$-$	$-$

Table 1: 4-**bonacci** – $g_{(a,i)}$ as an integer combination of $(g_{(a,0)},\dots,g_{(a,3)})$ and its signum.

Among several different ways of expressing $g_{(a,i)}$ (e.g., using Jordan normal form of M_φ), two of them will play an essential role in the following text.

Proposition 2.2 *For a given alphabet cardinality d, a fixed letter $a\in\mathcal{A}$ and $k\geq 0$, it holds:*

i)

$$g_{(a,k)} = T_{k+d-a-1} - \frac{T_{k+d}}{\beta^{a+1}}; \tag{2}$$

ii)

$$g_{(a,k)} = \sum_{j=2}^{d}\left(\frac{1}{\beta_j^{a+1}} - \frac{1}{\beta^{a+1}}\right)\frac{\beta_j^{k+d}}{f'(\beta_j)} \tag{3}$$

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

In order to find an upper bound c, we sum up the first m members of $(|g_{(a,i)}|)_{i=0}^{+\infty}$ and estimate the rest of them:

$$\sum_{i=0}^{+\infty}|g_{(a,i)}| \leq \sum_{i=0}^{m-1}|g_{(a,i)}| + E \qquad \text{where } E \text{ is arbitrary such that} \qquad E\geq \sum_{i=m}^{+\infty}|g_{(a,i)}|.$$

The second statement of Proposition 2.2 provides setting

$$E_{(a,m)} := |\beta_2|^m \sum_{j=2}^{d}\left|\left(\frac{1}{\beta_j^{a+1}} - \frac{1}{\beta^{a+1}}\right)\frac{1}{f'(\beta_j)}\right|\frac{|\beta_j|^d}{1-|\beta_j|} = |\beta_2|^m\cdot C(d,a).$$

To conclude, we have to find m big enough to satisfy

$$\left\lfloor 2\sum_{i=0}^{m-1}\left|g_{(a,i)}\right|\right\rfloor = \left\lfloor 2\left(\sum_{i=0}^{m-1}\left|g_{(a,i)}\right| + E_{(a,m)}\right)\right\rfloor. \tag{4}$$

	$a=0$	$a=1$	$a=2$	$a=3$
$\sum_{i=0}^{12} \lvert g_{(a,i)}\rvert$ as IC	$\begin{pmatrix}123\\183\\215\\232\end{pmatrix}$	$\begin{pmatrix}39\\63\\71\\76\end{pmatrix}$	$\begin{pmatrix}-133\\-201\\-233\\-254\end{pmatrix}$	$\begin{pmatrix}-47\\-71\\-83\\-86\end{pmatrix}$
$\sum_{i=0}^{12} \lvert g_{(a,i)}\rvert$ sym.	$1664-\frac{3205}{\beta}$	$286-\frac{1057}{\beta^2}$	$\frac{3499}{\beta^3}-487$	$\frac{1209}{\beta^4}-86$
$\sum_{i=0}^{12} \lvert g_{(a,i)}\rvert$ num.	1.2778	1.5157	1.5611	1.5776
$E_{(a,13)}$	0.20054	0.22213	0.25916	0.31056
$\sum_{i=0}^{12} \lvert g_{(a,i)}\rvert + E$	1.49844	1.76006	1.84618	1.91919
$B_a(n)$ upp. est.	2	3	3	3

Table 2: **4-bonacci** – Estimates of $\sum_{i=0}^{+\infty} |g_{(a,i)}|$ and resulting upper estimates of $B_a(n)$.

$d \ \backslash \ a$	0	1	2	3	4	5	6	7	8	9	10
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

Table 3: Upper estimates of $B_a(n)$ for $d \in \{2,\dots,11\}$, $a \in \{0,\dots,d-1\}$.

3 Computation

We have derived an upper bound for ΔD_a. Since we always compute on machines working in a finite precision, we try to reduce work with non-integer numbers. Therefore, we make use of the fact that, for a fixed letter a and the alphabet cardinality d, the sequence of $g_{(a,i)}$'s satisfies the d-bonacci recurrence (follows, e.g., from (2)). Namely, $g_{(a,n+d)} = g_{(a,n+d-1)} + \dots + g_{(a,n)}$

We demonstrate the method on the 4-bonacci word. The first step is calculating[1] $sgn(g_{(a,i)})$ from (2) for all $i \in \{0,\dots,m-1\}$ (illustrated in Table 1). Then we express $\sum_{i=0}^{m-1} |g_{(a,i)}|$ as an integer combination (IC) of $(g_{(a,0)},\dots,g_{(a,d-1)})$ which can be rewritten into the form $e + \frac{f}{\beta^{a+1}}$ for some $e, f \in \mathbb{Z}$ (as also follows from (2)) and then evaluated[1] (illustrated in Table 2). The final step is verification of the equality (4).

To make our procedure reliable with respect to possible rounding errors, we replace the estimated error $E_{(a,m)}$ by a constant $E > E_{(a,m)}$. If (4) holds, it is equal to the desired upper bound of $B_a(n)$ (but it may not be optimal). In the opposite case, we must increase m and repeat the procedure.

Our obtained results for $d \in \{2,\dots,11\}$ are summarized in Table 3.

From Theorem 2.1, we have not found a derivation of an upper bound on balance for a general d, yet. It seems from Tabel 3 that it might be, e.g., $\lceil \frac{d+1}{2} \rceil$.

To find lower bounds on the constant c one needs to find among factors of d-bonacci word two factors v, w of the same length with $|w|_a - |v|_a$ big enough. Computer searching in the set of all

[1]The calculation must be performed in an environment working in enough precision, e.g., Wolfram Mathematica.

factors is very time-consuming.

Nevertheless, for a given $d \geq 4$, we know a pair of factors v, w such that $|v|_a - |w|_a = 3$ for all $a \in \{2, \ldots, d-1\}$. Therefore, we can conclude with the following theorem.

Theorem 3.1 *For $d \in \{4, 5\}$, the d-bonacci word is c-balanced with $c = 3$ and this bound cannot be improved.*

4 Acknowledgement

This work was supported by the Grant Agency of the Czech Technical University in Prague, grant No. SGS11/162/OHK4/3T/14

References

[1] B. Adamczewski, Balances for fixed points of primitive substitutions, *Theor. Comput. Sci.* **307** (2003), p. 47–75.

[2] B. Adamczewski, Symbolic discrepancy and self-similar dynamics, *Ann. de l'inst. Four.* **54 no. 7** (2004), p. 2201–2234.

[3] G. Richomme, K. Saari, L. Q. Zamboni, Balance and Abelian complexity of the Tribonacci word, *Adv. in App. Math.* **45** (2010), p. 212–231.

[4] M. Morse, G. A. Hedlund, Symbolic dynamics, *Amer. J. Math.* **60** (1938), p. 815–866.

[5] J. Cassaigne, S. Ferenczi, L. Q. Zamboni , Imbalances in Arnoux-Rauzy sequences, *Ann. Inst. Fourier*, **50(4)** (2000), p. 1265–1276.

[6] J.-M. Dumont, A. Thomas, Systemes de numeration et fonctions fractales relatifs aux substitutions, *Theor. Comput. Sci.* **65** (1989), p. 153–169.

[7] G. Rauzy, Sequences defined by iterated morphisms, in *Sequences* (Naples/Positano, 1988), p. 275–286. Springer, New York, 1990.