A two-parameter family of an extension of Beatty sequences

Shiri Artstein-Avidan
School of Mathematical Sciences,
Tel Aviv University, Ramat Aviv, Tel Aviv 69978 ISRAEL
shiri@tau.ac.il
http://www.math.tau.ac.il/~shiri

Aviezri S. Fraenkel
Department of Computer Science and Applied Mathematics
The Weizmann Institute of Science, Rehovot 76100, Israel
fraenkel@wisdom.weizmann.ac.il
http://www.wisdom.weizmann.ac.il/~fraenkel

Vera T. Sós*
Alfred Renyi Institute of Mathematics
Hungarian Academy of Sciences, H-1053 Budapest, Hungary
sos@renyi.hu
http://www.renyi.hu/~sos

August 12, 2007

Abstract

Beatty sequences $\lfloor n\alpha + \gamma \rfloor$ are *nearly linear*, also called *balanced*, namely, the absolute value of the difference D of the number of elements in any two subwords of the same length satisfies $D \leq 1$.

*Research partially supported by the Hungarian Research Grant OTKA T042750 and OTKA T058290

For an extension of Beatty sequences, depending on two parameters $s, t \in \mathbb{Z}_{>0}$, we prove $D \leq \lfloor (s-2)/(t-1) \rfloor + 2$ $(s, t \geq 2)$, and $D \leq 2s+1$ $(s \geq 2, t = 1)$. We show that each value that is assumed, is assumed infinitely often. Under the assumption $(s-2) \leq (t-1)^2$ the first result is optimal, in that the upper bound is attained. This provides information about the gap-structure of (s,t)-sequences, which, for $s = 1$, reduce to Beatty sequences. The (s,t)-sequences were introduced in Fraenkel [9], where they were used to give a strategy for a 2-player combinatorial game on two heaps of tokens.

Keywords: Extension of Beatty sequences, sequences of differences, gap structure

1 Introduction

Denote by $\mathbb{Z}$, $\mathbb{Z}_{\geq 0}$ and $\mathbb{Z}_{>0}$ the set of integers, the set of nonnegative integers and the set of positive integers, respectively. For a subset $S \subset \mathbb{Z}_{\geq 0}$, $S \neq \mathbb{Z}_{\geq 0}$, the *minimum excluded value* of S is denoted $\operatorname{mex} S$ and defined to be the least nonnegative integer not in S. Denoting $\overline{S} = \mathbb{Z}_{\geq 0} \setminus S$, we have that [1]

$$\operatorname{mex} S = \min \overline{S}.$$

For two positive integers $s, t \in \mathbb{Z}_{>0}$, define the (s,t)-sequences $\{A_n\}$, $\{B_n\}$ by:

$$A_n = \operatorname{mex}\{A_i, B_i : 0 \leq i < n\} \quad \text{for all} \quad n \geq 0, \tag{1}$$

$$B_n = sA_n + tn \quad \text{for all} \quad n \geq 0. \tag{2}$$

Thus, $A_0 = B_0 = 0$ and $A_1 = 1$, $B_1 = s + t$. Prefixes of the two sequences, for $s = t = 2$, are displayed in Table 1.

Note that (1), (2) imply that A_n and B_n are strictly increasing sequences. Denoting $A = \bigcup_{n=1}^{\infty} A_n$ and $B = \bigcup_{n=1}^{\infty} B_n$, we have that A and B are *complementary* sets with respect to $\mathbb{Z}_{>0}$, that is, $A \cup B = \mathbb{Z}_{>0}$ (by (1)), and $A \cap B = \emptyset$. The last equality is true since if $A_m = B_n$, then $m > n > 0$ implies that A_m is the mex of a set containing $B_n = A_m$, a contradiction;

[1]The terminology "mex" was introduced in [2], and has since been used widely in the literature on the theory of combinatorial games.

Table 1: The first few entries of the $(2,2)$-sequences.

n	A_n	B_n	n	A_n	B_n
0	0	0	28	35	126
1	1	4	29	37	132
2	2	8	30	38	136
3	3	12	31	39	140
4	5	18	32	41	146
5	6	22	33	42	150
6	7	26	34	43	154
7	9	32	35	44	158
8	10	36	36	45	162
9	11	40	37	47	168
10	13	46	38	48	172
11	14	50	39	49	176
12	15	54	40	51	182
13	16	58	41	52	186
14	17	62	42	53	190
15	19	68	43	55	196
16	20	72	44	56	200
17	21	76	45	57	204
18	23	82	46	59	210
19	24	86	47	60	214
20	25	90	48	61	218
21	27	96	49	63	224
22	28	100	50	64	228
23	29	104	51	65	232
24	30	108	52	66	236
25	31	112	53	67	240
26	33	118	54	69	246
27	34	122	55	70	250

and $0 < m \le n$ is impossible since $B_n = sA_n + tn \ge sA_m + tm > A_m$. The (s,t)-sequences were introduced in Fraenkel [9], where they were used to give a strategy for a 2-player combinatorial game on two heaps of tokens.

Notation 1. For $m, n, j \in \mathbb{Z}_{\geq 0}$, let

$$D_{m,n,j} = |(A_{n+j} - A_n) - (A_{m+j} - A_m)|,$$

$$E_{m,n,j} = |(B_{n+j} - B_n) - (B_{m+j} - B_m)|.$$

Note that $D_{m,n,j}$ and $E_{m,n,j}$ are symmetric in m, n, i.e., $D_{m,n,j} = D_{n,m,j}$, $E_{m,n,j} = E_{n,m,j}$. Our main purpose is to prove the following theorem, in §2.

Theorem 1. *Let $s \geq 2$, $t \geq 2$ and assume that $(s-2) \leq (t-1)^2$. Then $D_{m,n,j} \in S_1$, where $S_1 := \{0, \ldots, q\}$, $q := \lfloor (s-2)/(t-1) \rfloor + 2$, and each of the values in S_1 is assumed infinitely often.*

Note that (2) implies $E_{m,n,j} = sD_{m,n,j}$. Thus Theorem 1 implies the following result about $E_{m,n,j}$:

Corollary 1. *Let $s \geq 2$, $t \geq 2$ and assume that $(s-2) \leq (t-1)^2$. Then $E_{m,n,j} \in sS_1 := \{0, s, 2s, \ldots, qs\}$, q as in Theorem 1. Each of the values in sS_1 is assumed infinitely often.*

For the case $s - 2 > (t-1)^2$ we can only show an upper bound, namely we have

Theorem 2. *Let $s \geq 2$, $t \geq 2$ and assume that $(s-2) > (t-1)^2$. Then $D_{m,n,j} \in S_1$, where S_1 and q are as in Theorem 1, and each of the values in S_1 which is assumed, is assumed infinitely often.*

For $s \geq 2$ and $t = 1$ we prove the following result in §3.

Theorem 3. *Let $s \geq 2$, $t = 1$. Then $D_{m,n,j} \in S_2 := \{0, \ldots, 2s+1\}$.*

The two latter results are weaker than Theorem 1, since we don't know whether the upper bounds in each of them are sharp. The corresponding corollaries applying to $E_{m,n,j}$ can be formulated for each of these cases analogously to Corollary 1.

Notice that if $t \geq s \geq 2$, then $\lfloor (s-2)/(t-1) \rfloor = 0$, so that in Theorem 1 we have that $S_1 = \{0, 1, 2\}$. However, if $s > t \geq 2$, then we may have $|S_1| > 3$.

Theorem 1 provides information about the behavior of the gap-structure of (s,t)-sequences. For $s=1$, both $\{A_m\}$ and $\{B_m\}$ are special cases of *Beatty sequences*, namely $A_n=\lfloor n\alpha\rfloor$, $B_n=\lfloor n(\alpha+t)\rfloor$, where $\alpha=(2-t+\sqrt{t^2+4})/2$ (so, for $t=1$, $\alpha=\phi$ is the golden section). A general Beatty sequence has the form $A_n=\lfloor n\alpha+\gamma\rfloor$, where $\alpha>0,\gamma$ are real numbers, $n\in\mathbb{Z}_{\geq 0}$. It is well known that for general Beatty sequences, the difference $D_{m,n,j}$ assumes only two values: $D_{m,n,j}=E_{m,n,j}\in\{0,1\}$ for all $j,m,n\in\mathbb{Z}_{\geq 0}$, where each of 0 and 1 is assumed infinitely often. In the earlier literature this property was called *nearly linear*; see Ron Graham et al. [11], Boshernitzan and Fraenkel [4], [5]. Nowadays it is called *balanced*: Berstel and Séébold [3], Tijdeman [13].

We note in passing that Theorem 1 holds also for $s=1$ and $t\geq 2$, since then $S_1=\{0,1\}$. In this case Corollary 1 coalesces with Theorem 1.

Balanced sequences have been used previously for providing a strategy for games. See Wythoff [14], Coxeter [6], Yaglom and Yaglom [15] ($s=t=1$); Fraenkel [7] ($s=1$, $t\in\mathbb{Z}_{>0}$). The subword complexity $C(n)$ of the characteristic functions of these sequences was computed in [10]. It is linear in the length n of the subword, but larger than $C(n)=n+1$, which characterizes the subword complexity of *Sturmian* sequences, the characteristic functions of Beatty sequences. The subword complexity $C(n)$ of a sequence S is the number of distinct words of length n appearing in S. See e.g. [1].

2 Proof of Theorems 1 and 2

The proofs of Theorems 1 and 2 will be separated into three steps. First we show that the number q in the statements of the theorems is, in both cases, an upper bound for $D_{m,n,j}$. Then we will show that, in both cases, for each value which is ever assumed, every value not exceeding it is assumed infinitely often. Finally we show that under the assumption $(s-2)\leq(t-1)^2$ of Theorem 1, the upper bound q is attained. We thus split the proofs into the following three parts.

Proposition 1. *Let $s\geq 2$, $t\geq 2$. Then $D_{m,n,j}\in S_1$, where $S_1:=\{0,\ldots,q\}$, $q:=\lfloor(s-2)/(t-1)\rfloor+2$.*

Proposition 2. *Let $s\geq 2$, $t\geq 2$. Then if for some d and some m,n,j, we have $D_{m,n,j}=d$, then for every $d'\leq d$ there are infinitely many m,n,j such that $D_{m,n,j}=d'$.*

Proposition 3. *Let $s \geq 2$, $t \geq 2$ and assume that $(s-2) \leq (t-1)^2$. Then for some m, n, j we have $D_{m,n,j} = \lfloor (s-2)/(t-1) \rfloor + 2$.*

We begin by introducing some notation, followed by three auxiliary results.

Notation 2. Let $n \in \mathbb{Z}_{\geq 0}$. An *$A$-gap* is $G^A = G_n^A := A_{n+1} - A_n$. A *$B$-gap* is $G^B = G_n^B := B_{n+1} - B_n$. For $i, m, n \in \mathbb{Z}_{\geq 0}$, put $\Delta_i := |D_{m,n,i+1} - D_{m,n,i}|$.

Lemma 1. *Let $s, t \in \mathbb{Z}_{>0}$. For all $n \in \mathbb{Z}_{\geq 0}$,*

$$G_n^A \in \{1, 2\}, \tag{3}$$

and each of the values $1, 2$ is assumed infinitely often. Also for all $n \in \mathbb{Z}_{>0}$,

$$\begin{aligned} G_n^B &= s + t \iff G_n^A = 1, \\ G_n^B &= 2s + t \iff G_n^A = 2, \end{aligned} \tag{4}$$

and each of the B-gaps $s + t$ and $2s + t$ is assumed infinitely often.

Proof. If there is some $G^A \geq 3$, then the complementarity of A, B implies that there is some $G^B = 1$. However, for all $n \in \mathbb{Z}_{\geq 0}$, (2) implies $G_n^B = sG_n^A + t \geq s + t \geq 2$. This proves (3); and (4) follows from (2).

If there is $N \in \mathbb{Z}_{\geq 0}$ such that $G_n^A = 1$ for all $n \geq N$, then for sufficiently large i there exists j such that $B_i = sA_i + ti = A_j$, contradicting complementarity. If there is $N \in \mathbb{Z}_{\geq 0}$ such that $G_n^A = 2$ for all $n \geq N$, then $G_n^B = 2s + t \geq 3$ for all sufficiently large n, so some positive integers are missing, again contradicting complementarity. Thus each of the values $1, 2$ in (3) is assumed infinitely often. It follows that also each of $s + t$ and $2s + t$ in (4) is assumed infinitely often. ■

Definition 1. Let $n \in \mathbb{Z}_{\geq 0}$. An *$A$-word* is a maximal run of A_i: $A_{n+1}, \ldots, A_{n+m}$, such that $G_{n+i}^A = 1$ for $i \in \{1, \ldots, m-1\}$, $G_n^A = G_{n+m}^A = 2$.

A *B-word* is the corresponding maximal run of m elements B_i: $B_{n+1}, \ldots, B_{n+m}$, which satisfies, by (3), $G_{n+i}^B = s + t$ for $i \in \{1, \ldots, m-1\}$, $G_n^B = G_{n+m}^B = 2s + t$.

The *length* of an A-word or a B-word is the number m of its elements.

An A-word of length $s + t - 1$ is a *small* A-word; An A-word of length $2s + t - 1$ is a *large* A-word.

Consider $\mathbb{Z}_{>0} = A \cup B$ as a sequence C in which the elements of A and B are sort-merged in increasing order. Note that C consists of small A-words or large ones, separated by B-singletons, since $G^A \in \{1, 2\}$, $G^B \in \{s+t, 2s+t\}$, by Lemma 1.

Lemma 2. *Let $s, t \in \mathbb{Z}_{>0}$. Following a finite prefix of small A-words, the sequence C is composed of large A-words which are separated by $s + t - 2$ or by $2s + t - 2$ small A-words. Each of these A-word lengths and separating lengths occur infinitely often, and there are no others.*

Proof. By Lemma 1, $G^B \in \{s + t, 2s + t\}$, so the complementarity of A, B implies that the A-word lengths are restricted to $\{s + t - 1, 2s + t - 1\}$.

By (4), the B-words have the same length as the A-words. A B-word of length $s + t - 1$ contains precisely $s + t - 2$ B-gaps of size $s + t$. By complementarity, each such gap constitutes an A-word of length $s+t-1$. A B-word of length $2s+t-1$ contains precisely $2s+t-2$ B-gaps of size $s+t$. Each such gap again constitutes a small A-word. Either of these two B-words is flanked on both sides by $G_B = 2s + t$, which, again by complementarity, induces A-words of length $2s + t - 1$.

By (4), the A-words of C, except the prefix, have only two possible lengths: $s + t - 1$ and $2s + t - 1$, and also the the number of consecutive B-gaps is restricted to $s + t - 2$ and $2s + t - 2$. Therefore the cases considered here are the only ones, and by Lemma 1 each occurs infinitely often. This also implies that the prefix has finite length. ■

Lemma 3. *Let $m, j \in \mathbb{Z}_{\geq 0}$, $k \in \mathbb{Z}_{>0}$. Then*

(i) $D_{m,m+k,j} = D_{m,m+j,k}$,

(ii) $D_{m,m+1,j} = D_{m,m+j,1} \in \{0, 1\}$,

(iii) $\Delta_i \in \{0, 1\}$ *for all* $i \geq 1$.

Proof. (i) We have

$$D_{m,m+k,j} = |(A_{m+k+j} - A_{m+k}) - (A_{m+j} - A_m)| = \\ |(A_{m+j+k} - A_{m+j}) - (A_{m+k} - A_m)| = D_{m,m+j,k}.$$

(ii) The equality is the special case $k = 1$ of (i). By (3), $D_{m,m+j,1} = |(A_{m+j+1} - A_{m+j}) - (A_{m+1} - A_m)| = |G^A_{m+j} - G^A_m| \in \{0, 1\}$.

(iii) By the triangle inequality (in the form $||x| - |y|| \leq |x - y|$),

$$\begin{aligned}
\Delta_i &= \big| |(A_{n+i+1} - A_n) - (A_{m+i+1} - A_m)| \\
&- |(A_{n+i} - A_n) - (A_{m+i} - A_m)| \big| \\
&\leq |(A_{n+i+1} - A_{n+i}) - (A_{m+i+1} - A_{m+i})| = D_{m+i,n+i,1} \in \{0, 1\}
\end{aligned}$$

by (ii). ■

Proof of Proposition 1. The following trivial observation will be used throughout the proof: for every $a, k \in \mathbb{Z}_{\geq 0}$, the cardinality of the half-open interval $(A_k, A_{k+a}] \subseteq C$ is

$$A_{k+a} - A_k = a + \mathrm{card}\{B \cap (A_k, A_{k+a})\}, \tag{5}$$

and for every $b, \ell \in \mathbb{Z}_{\geq 0}$, the cardinality of the half-open interval $(B_\ell, B_{\ell+b}] \subseteq C$ is

$$B_{\ell+b} - B_\ell = b + \mathrm{card}\{A \cap (B_\ell, B_{\ell+b})\}. \tag{6}$$

We show, by induction on j, that for any $m, n \in \mathbb{Z}_{\geq 0}$, we have $D_{m,n,j} \in S_1$. By Lemma 3(ii), for every m and n one has $D_{m,n,1} \in \{0, 1\} \subseteq S_1$. Let $j \geq 2$ and assume inductively that $D_{m,n,i} \in S_1$ for all $i < j$ and for all $m, n \in \mathbb{Z}_{\geq 0}$. Suppose that the assertion is false, i.e., $D_{m,n,j} = d > q$ for some $m, n \in \mathbb{Z}_{\geq 0}$. Without loss of generality, and using (5), we may assume that for some $h \geq 0$ there are h members of B in $[A_m, A_{m+j}]$ and $h + d$ $(d > q)$ members of B in $[A_n, A_{n+j}]$. So C must contain the following two subwords,

$$B_u \ldots A_m \ldots B_{u+1} \ldots B_{u+h} \ldots A_{m+j} \ldots B_{u+h+1} \ldots B_{u+h+d-1}, \tag{7}$$

$$B_v \ldots A_n \ldots B_{v+1} \ldots B_{v+h+d} \ldots A_{n+j} \ldots B_{v+h+d+1} \tag{8}$$

for suitable indices $u, v \in \mathbb{Z}_{\geq 0}$, where possibly $B_u = 0$ or $B_v = 0$. We wish to estimate $E_{v+1,u,h+d-1}$. From (6) and (7) we get

$$B_{u+h+d-1} - B_u \geq (h + d - 1) + (d - 2)(s + t - 1) + (j + 1),$$

since there are at least $s + t - 1$ members of A in $(B_{u+h+i}, B_{u+h+i+1})$ for all $i \in \{1, \ldots, d - 2\}$, and $j + 1$ members of A in (B_{u+1}, B_{u+h+1}).

Similarly (6) and (8) imply that $B_{v+h+d} - B_{v+1} \leq (h+d-1) + j - 1$, since there are at most $j-1$ members of A in (B_{v+1}, B_{v+h+d}). Hence

$$\begin{aligned} E_{v+1,u,h+d-1} &\geq (s+t-1)(d-2)+2 \qquad (9) \\ = s(d-2)+(t-1)(d-2)+2 &\geq s(d-2)+(t-1)(q-1)+2 \\ &= s(d-2)+(t-1)(\lfloor (s-2)/(t-1) \rfloor + 1)+2 \\ &> s(d-2)+(t-1)(s-2)/(t-1)+2 = s(d-1) \geq qs. \end{aligned}$$

Recall that in the interval $[A_n, A_{n+j}]$ there are $h+d$ members of B. Taking (5) and (8) into account, we can see that $h+d < j$: indeed, Lemma 1 guarantees that between each two elements of B there are at least $s+t-1 \geq 3$ elements of A. But between B_{v+1} and B_{v+h+d} there are at most $j-1$ members of A. Thus $3(h+d-1) \leq j-1$, which in turn implies $h+d < j$ (since $j \geq 2$). The induction hypothesis now implies that $E_{v+1,u,h+d-1} \leq qs$, contradicting (9). Thus also $D_{m,n,j} \in S_1$. ■

We now wish to show the second part of Theorems 1 and 2, that each of the values in S_1 is attained infinitely often. We begin by proving Proposition 2, stating that (for both Theorem 1 and Theorem 2) once a value is assumed, this value, and all of the values below it, will be assumed infinitely often.

Proof of Proposition 2. We first claim that it suffices to show that d is assumed infinitely often, and that this already implies that all $d' < d$ are also assumed infinitely often. Indeed, assume that $D_{m,n,j} = d$ for some $m, n \in \mathbb{Z}_{\geq 0}$. Without loss of generality, $m < n$. When i increases by 1, then $D_{m,m+i,j} = D_{m,m+j,i}$ changes by at most 1 (Lemma 3). Since $D_{m,m,j} = 0$, we see that as i changes from 0 to $n-m$, $D_{m,m+i,j}$ assumes all the values in $\{0, \ldots, d\}$. If d is assumed infinitely often, then so are all the values $\{0, \ldots, d\}$.

Thus, without loss of generality, we will assume that d is the largest value which is assumed. We have to show that it is assumed infinitely often. So we let d be such that $D_{m,n,j} = d$ for some $m, n, j \in \mathbb{Z}_{\geq 0}$ and $D_{m,n,j} \leq d$ for all $m, n, j \in \mathbb{Z}_{\geq 0}$. From Proposition 1 it follows that $d \leq q$.

Choose m and n such that C contains the subwords

$$B_u \ldots A_m \ldots B_{u+1} \ldots B_{u+h} \ldots A_{m+j} \ldots B_{u+h+1} \ldots B_{u+h+d-1}, \qquad (10)$$

$$B_v \ldots A_n \ldots B_{v+1} \ldots B_{v+h+d} \ldots A_{n+j} \ldots B_{v+h+d+1}, \qquad (11)$$

where, possibly, $B_u = 0$ or $B_v = 0$.

To every subword containing some terms A_i, there corresponds a subword, appearing later on in C, containing the terms B_i with the same indices as the terms A_i. In particular, corresponding to parts of the subwords (10), (11) above there exist subwords:

$$A_x B_m A_{x+1} \dots B_{m+1} \dots A_{x+i_1-1} B_{m+j} A_{x+i_1}, \tag{12}$$

$$A_{y-1} B_n A_y \dots B_{n+1} \dots A_{y+i_2} B_{n+j} A_{y+i_2+1}, \tag{13}$$

where the indices x, y are chosen so that $A_x + 2 = B_m + 1 = A_{x+1}$ and $A_{y-1} + 2 = B_n + 1 = A_y$, and the indices i_1 and i_2 are chosen so that $A_{x+i_1-1} + 2 = B_{m+j} + 1 = A_{x+i_1}$ and $A_{y+i_2} + 2 = B_{n+j} + 1 = A_{y+i_2+1}$. It suffices to show that $D_{x,y,i_1} = d$. By (2) and since $A_{m+j} - A_m = h + j$ and $A_{n+j} - A_n = h + d + j$, we have

$$B_{m+j} - B_m = sh + sj + tj = j + i_1 - 1,$$

$$B_{n+j} - B_n = sh + sj + sd + tj = j + i_2 + 1.$$

Thus, $i_2 - i_1 = sd - 2$, which implies the identity $i_2 - i_1 = F + G + H$, where

$$F = 2s + t - 1,\ G = (d-3)(s+t-1),\ H = (s-2) - (d-2)(t-1).$$

Notice that: (i) F is the length of a large A-word, (ii) G might be negative (e.g., if $s = 2$); (iii) $d \le q$ implies $H \ge 0$. To compute D_{x,y,i_1} we need to compute the cardinality of the half-open interval $I := (A_{y+i_1}, A_{y+i_2}]$, which is, by (5),

$$A_{y+i_2} - A_{y+i_1} = i_2 - i_1 + |\{B \cap (A_{y+i_1}, A_{y+i_2})\}|.$$

To estimate $|\{B \cap (A_{y+i_1}, A_{y+i_2})\}|$ suppose that I contains ≥ 2 large A-words. Since between 2 distinct large A-words there are at least $s + t - 2$ small A-words, we thus have $2F + (s+t-2)(s+t-1) \le F + G + H$, i.e., $F + (s+t-2)(s+t-1) \le G + H = s(d-2) - t - 1$. However, $s(d-2) - t - 1 \le s(q-2) - t - 1 \le s(s-2) - t - 1 < F + (s+t-2)(s+t-1)$, since $s \le s + t - 1$, $s - 2 < s + t - 2$. So there is at most one large A-word in I. If there is one, then there are at least $(d-3)$ small A-words in I, and we still have H A-terms to spare, which may also include a B-term. So I contains at least $d - 2$ A-words, the rightmost of which ends at A_{y+i_2}, since to the right of A_{y+i_2} there is a B-term by (13). The leftmost of the A-words in I doesn't begin with A_{y+i_1}, since it is outside I, nor with $A_{y+i_1}+1$, since then the first A-word of I would be incomplete. Therefore it begins with A_{y+k} for some $k > i_1$, so there is a B-term to its left which is still in I. It follows that I contains at least $d-2$ B-terms, one to the left of each A-word in I. Hence,

$$|\{B \cap (A_{y+i_1}, A_{y+i_2})\}| \geq d-2,$$

which implies $A_{y+i_2} - A_{y+i_1} \geq i_2 - i_1 + d - 2$.

In case there is no large A-word in $[A_{y+i_1}, A_{y+i_2}]$ we are only better off, and still have at least $(d-2)$ B-terms in the interval. From (12), $A_{x+i_1} - A_x = i_1 + j + 1$, and from (13), $A_{y+i_2} - A_y = i_2 + j - 1$. Therefore,

$$\begin{aligned} D_{x,y,i_1} &= |(A_{x+i_1} - A_x) - (A_{y+i_1} - A_y)| \\ &= |(A_{x+i_1} - A_x) - (A_{y+i_2} - A_y) + (A_{y+i_2} - A_{y+i_1})| \\ &= (i_1 + j + 1) - (i_2 + j - 1) + A_{y+i_2} - A_{y+i_1} \geq d, \end{aligned}$$

but as d was chosen to be the largest value assumed, we see that $D_{x,y,i_1} = d$. ■

To prove Proposition 3, we follow the same lines and notations as in the proof of Proposition 2 above, with a modification at the end.

Proof of Proposition 3. Notice that already in the proof of Proposition 2 we were quite close to proving that q is attained. Indeed, if we would have been able to show that $I = (A_{y+i_1}, A_{y+i_2}]$ contains $(d-1)$ small A-words, we could rewrite $i_2 - i_1$ as

$$i_2 - i_1 = (d-1)(s+t-1) + ((s-2) - (d-1)(t-1)).$$

Assuming $d < q$, we have $d \leq 1 + (s-2)/(t-1)$, that is, $(d-1)(t-1) \leq s-2$, so that

$$i_2 - i_1 \geq (d-1)(s+t-1).$$

By the same reasoning as in the proof of Proposition 2, we would then get that there are at least $(d-1)$ B-terms in I (one to the left of each A-word), and so, as in the Proof of Proposition 2, that

$$D_{x,y,i_1} \geq d+1,$$

contradicting the maximality of d assumed in that proof, so $d = q$. However, there is no guarantee that there are only small A-words in this interval.

Remark 1. The interval I cannot contain a large A-word and $d-2$ small A-words. Indeed, $d \geq 1$ by Lemma 1, say by choosing m, n so that $G_n^A = 1, G_m^A = 2$, hence $D_{m,n,1} = |G_n^A - G_m^A| = 1$. Thus $F + (d-2)(s+t-1) = sd + (d-1)(t-1) \geq sd > i_2 - i_1$.

The plan of the proof below is as follows: We choose d to be the largest value attained, and assume $d < q$. We then repeat the construction as in Proposition 2, that is, arrive at words of the form (10), (11), (12) and (13). By the same argument as in Proposition 2, $D_{x,y,i_1} = d$. We will then "count" the number of small A-words in $I = (A_{y+i_1}, A_{y+i_2}]$. In the case where this number is at least $(d-1)$, then, as explained above, we arrive at a contradiction and the proof is complete. In the complementary case, we will iterate the construction once more.

More precisely, we consider the subword (13) and ask what is the number of small A-words to the left of B_{n+j}. To check what are the lengths of the various A-words we have to look back at the structure of the original word (11). Let $k \in \mathbb{Z}_{\geq 0}$ denote the number of consecutive A terms to the left of A_{n+j}, that is, $B_{v+h+d} + 1 = A_{n+j-k}$. Then, in the word (13) (which is the "B-image" of (11)) we have, counting A-words from right to left, starting with the A-word which ends with A_{y+i_2}, exactly k small A-words to the right of the first large A-word. If $k \geq (d-1)$, the proof is complete. So we may assume that $k < (d-1)$, and that there are less than $(d-1)$ small A-words in I. Then we know the structure of the word (13) in more detail. We consider two cases.

(i) $k \leq d-3$. There are k small A-words to the left of B_{n+j}, followed on the left by a large A-word, and then (by the formula for $i_2 - i_1$)) we have $(d-3-k) \geq 0$ additional small A-words. The last word on the left counted up-to-now has the term $B_{n+j-(d-2)}$ to its left. To the left of that we have $H = (s-2) - (d-2)(t-1)$ extra A-terms, which do not form a full A-word by Remark 1, and then the term A_{y+i_1} (which is already outside I, since I was chosen to be half-open). Since we assume $s - 2 \leq (t-1)^2$ and $d < q$, we have $(s-2) + (t-1) \geq (s-2)t/(t-1) \geq (d-1)t$. This implies $H = (s-2) - (d-2)(t-1) \geq (d-1)$. As before, this means that there are at least $(d-1)$ B-terms in the interval $[A_{z+l_1}, A_{z+l_2})$, one to the right of each A-word. Plugging this back into the equations we get $D_{z,w,l_2} = d+1$, a contradiction.

(ii) $k = d-2$. There are $(d-2)$ small A-words to the left of B_{n+j}, the leftmost of which has $B_{n+j-(d-2)}$ to its left. This is followed on the left by

$F+G+H-(d-2)(s+t-1)=2s-2-(d-2)(t-1)$ A-terms, not enough to fill a whole large A-word, and then we have A_{y+i_1}.

We rewrite part of the subword (13) in a way emphasizing its properties discussed in cases (i) and (ii).

$$B_n A_y \dots B_{n+1} \dots B_{n+j-d+1} \dots A_{y+i_1} \dots B_{n+j-(d-2)}, \tag{14}$$

We now generate the following two subwords (15) and (16) from (14) and (12) respectively, the same way as the words (13) and (12) were generated from (10), (11):

$$A_z B_y A_{z+1} \dots B_{y+1} \dots A_{z+l_1-1} B_{y+i_1} A_{z+l_1} \dots A_{z+l_2}, \tag{15}$$

$$A_{w-1} B_x A_w \dots B_{x+1} \dots A_{w+l_2} B_{x+i_1} A_{w+l_2+1}, \tag{16}$$

where the indices z, w are chosen so that $A_z + 2 = B_y + 1 = A_{z+1}$ and $A_{w-1} + 2 = B_x + 1 = A_w$, and the indices l_1 and l_2 are chosen so that $A_{z+l_1-1} + 2 = B_{y+i_1} + 1 = A_{z+l_1}$ and $A_{w+l_2} + 2 = B_{x+i_1} + 1 = A_{w+l_2+1}$. We now repeat the reasoning of the type used in the proof of Proposition 2. We have,

$$B_{y+i_1} - B_y = s(i_1 + j - d + 1) + i_1 t = i_1 + l_1 - 1$$
$$B_{x+i_1} - B_x = s(i_1 + j + 1) + i_1 t = i_1 + l_2 + 1.$$

Thus,

$$l_2 - l_1 = sd - 2 = (s+t-1)(d-1) + ((s-2) - (d-1)(t-1)).$$

We would now like to estimate D_{z,w,l_2}. We have,

$$\begin{aligned} A_{z+l_2} - A_z &= (A_{z+l_1} - A_z) + (A_{z+l_2} - A_{z+l_1}) \\ &= i_1 + 1 + l_2 + |B \cap (A_{z+l_1}, A_{z+l_2})| \end{aligned}$$

and

$$A_{w+l_2} - A_w = l_2 + i_1 - 1.$$

We wish to show that there are at least $(d-1)$ B-terms in the interval $J := [A_{z+l_1}, A_{z+l_2})$, since then we'll have our desired contradiction: $D_{z,w,l_2} = d+1$. To this end we count the A-words in J, to the right of each of which there is a B-term.

We have exactly $2s-2-(d-2)(t-1)$ small A-words to the right of B_{y+i_1}, because this is the number of A-terms to the right of A_{y+i_1}. This number is easily checked to be at least $(d-1)$, assuming $d < q$:

$$2s-2-(d-2)(t-1) = 2s+t-3-(d-1)(t-1) \geq s+t-1 \geq d-1.$$

Thus, there are at least $(d-1)$ B-terms in the interval $[A_{z+l_1}, A_{z+l_2})$, one to the right of each A-word. Plugging this back into the equations we get $D_{z,w,l_2} = d+1$, a contradiction. ■

3 Proof of Theorem 3

Both Theorem 3 and Proposition 1 give an upper bound. In fact, their proofs are very similar. We sketch the proof of Theorem 3, elaborating only on the points that are different from that of Proposition 1.

We again use induction. Assume that the assertion is false for the smallest j in $D_{m,n,j}$. We may assume that $j \geq 2$, since $D_{m,n,1} \in S_2$ by Lemma 3(ii). Thus for some $d > q := 2s+1$ we have two words of the form (10) and (11), where now A-words have either length s or $2s$. This time we use the fact that $d-2 > 2s-1$, which implies that at least one of the A-words between B_{u+h+1} and $B_{u+h+d-1}$ is of length $2s$ (follows from Lemma 2). This in turn implies the inequality

$$B_{u+h+d-1} - B_u \geq (h+d-1) + (d-1)s + (j+1).$$

The second inequality is the same as for the case $t > 1$, $B_{v+h+d} - B_{v+1} \leq (h+d-1) + j - 1$, and combining the two we arrive at

$$(B_{u+h+d-1} - B_u) - (B_{v+h+d} - B_{v+1}) \geq s(d-1) + 2 > s(2s+1). \qquad (17)$$

However, again $h+d < j$, so the induction hypothesis guarantees that this cannot be true, and we arrive at the desired contradiction. We conclude that for all $m, n, j \in \mathbb{Z}_{\geq 0}$, $D_{m,n,j} \in S_2$. ■

The following questions remain open: (i) Is the condition $s-2 \leq (t-1)^2$ in Theorem 1 indeed necessary? We used it just once, in the proof of Proposition 3. If in that proof we would iterate the construction of (15) and (16) from (14) and (12) once more, it appears that the condition could be relaxed to $s-2 \leq t(t-1)$. (ii) Is the upper bound $2s+1$ in Theorem 3 not sharp when $t=1$?

4 Epilogue

For $n \in \mathbb{Z}_{>0}$, the *characteristic function* $\chi(n)$ of any sequence A_m is defined by

$$\chi(n) = \begin{cases} 1 & \text{if } \exists m \text{ such that } A_m = n \\ 0 & \text{otherwise.} \end{cases}$$

Let S_{2n} be any binary word of length $2n$, and $\sigma(2i)$, the sum of the elements of its prefix of length $2i$ $(1 \le i \le n)$. R. Tijdeman observed (private communication), that if $\sigma(2i) = i$ for all $i \in \mathbb{Z}_{>0}$, and we let A_k be a sequence with characteristic function S_{2n}, then:

(i) Every such sequence that contains the subwords 00, 01 and 11, satisfies $D_{m,n,1} \in \{0, 1, 2\}$, and so is not a Beatty sequence. It also satisfies $G_n^A \in \{1, 3\}$, and so it is not an (s, t)-sequence by Lemma 2. This shows that the converse of Theorem 1 does not hold.

(ii) There are $|S_{2n}| = 2^n$ such sequences. (We can always prefix such a sequence with 00, 01, or 11 if either is missing.)

We note that similar constructions (say with $\sigma(4i) = 2i$), show that also the a converse theorem in the case $t = 1$ does not hold. We also mention that Mignosi [12] has shown that $D_{k,m,j} \in \{0, 1\}$ for all $k, m, j \in \mathbb{Z}_{\ge 0}$ is satisfied only by $O(n^3)$ sequences of length n.

Explicit functions satisfying $D_{m,n,j} \le 2$; and $D_{m,n,j} = 2$ infinitely often, can be constructed using the following

Theorem 4. (See [8], Theorem 1.) *Let $n \ge 1$ and $a_0, \ldots, a_n$, $m, K, L, M \in \mathbb{Z}$. Suppose that $a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 = 0$ has a real nonzero root α. Let $A(m) = \lfloor m\alpha \rfloor$. Then*

$$A\Big(M + Lm + \sum_{i=0}^{n-2} A^i\big(Ka_{i+2}A(m)\big)\Big) = (L - Ka_1)A(m) - Ka_0 m + D,$$

where D is bounded in m, namely,

$$D = \lfloor M\alpha + (L + Ka_0\alpha^{-1})\{m\alpha\} - \theta\alpha \rfloor,$$

Table 2: Discrete chaos 1.

m	$\lfloor m\alpha \rfloor$	$\lfloor \lfloor m\alpha \rfloor \alpha \rfloor$
1	2	4
2	4	9
3	7	16
4	9	21
5	12	28
6	14	33
7	16	38
8	19	45
9	21	50
10	24	57
11	26	62
12	28	67
13	31	74
14	33	79
15	36	86
16	38	91
17	41	98
18	43	103
19	45	108
20	48	115

$$\theta = \sum_{i=1}^{n-2} \bigl(K a_{i+2} A(m) \alpha^i - A^i(K a_{i+2} A(m)) \bigr),$$

where $\{x\}$ is the fractional part of x.

Put $n = 2$, $a_2 = 1$, $a_1 = -2$, $a_0 = -1$, $K = 1$, $L = M = 0$ in (17). Then

$$\bigl\lfloor \lfloor m\alpha \rfloor \alpha \bigr\rfloor = 2\lfloor m\alpha \rfloor + m - 1$$

for $\alpha = 1 + \sqrt{2}$. Since $\lfloor m\alpha \rfloor$ satisfies $D_{k,\ell,j} \in \{0, 1\}$, the right-hand side of this identity shows that $D_{k,\ell,j} \le 2$. In fact, $D_{k,\ell,j} \in \{0, 2\}$. Theorems 1 and

Table 3: Discrete chaos 2.

m	$\lfloor m\phi \rfloor$	$\lfloor \lfloor m\phi \rfloor 2\phi \rfloor$
1	1	3
2	3	9
3	4	12
4	6	19
5	8	25
6	9	29
7	11	35
8	12	38
9	14	45
10	16	51
11	17	55
12	19	61
13	21	67
14	22	71
15	24	77
16	25	80
17	27	87
18	29	93
19	30	97
20	32	103

2 thus imply that this is not an (s,t)-sequence. The first few entries of this table are depicted in Table 2.

Putting $a_1 = -1$, $K = 2$, but retaining the other values leads to

$$\big\lfloor \lfloor m\phi \rfloor 2\phi \big\rfloor = 2\lfloor m\phi \rfloor + \lfloor (1-\sqrt{5})\{m\phi\} \rfloor,$$

where ϕ is the golden section. Note that $\lfloor (1-\sqrt{5})\{m\phi\} \rfloor \in \{-1,-2\}$ for all $m \in \mathbb{Z}_{>0}$. It can be seen that now $D_{k,\ell,j} \in \{0,1,2,3,4\}$, and each of these values is assumed infinitely often. Lemma 1 once again shows that it is not an (s,t)-sequence. See Table 3.

References

[1] J.-P. Allouche [1994], Sur la complexité des suites infinies, *Bull. Belg. Math. Soc.* **1**, 133–143.

[2] E. R. Berlekamp, J. H. Conway and R. K. Guy [2001-2004], *Winning Ways for your Mathematical Plays*, Vol. 1-4, A K Peters, Wellesley, MA, 2nd edition: vol. 1 (2001), vols. 2, 3 (2003), vol. 4 (2004); translation of 1st edition (1982) into German: *Gewinnen, Strategien für Mathematische Spiele* by G. Seiffert, Foreword by K. Jacobs, M. Reményi and Seiffert, Friedr. Vieweg & Sohn, Braunschweig (four volumes), 1985.

[3] J. Berstel and P. Séébold [1994], Morphismes de Sturm. (French) [Sturm morphisms] Journées Montoises (Mons, 1992). Bull. Belg. Math. Soc. Simon Stevin **1**, 175–189. Appeared first under the title: A characterization of Sturmian morphisms. Mathematical Foundations of Computer Science 1993 (Gdańsk, 1993), 281–290, Lecture Notes in Comput. Sci., 711, Springer, Berlin, 1993.

[4] M. Boshernitzan and A.S. Fraenkel [1981], Nonhomogeneous spectra of numbers, *Discrete Mathematics* **34**, 325–327.

[5] M. Boshernitzan and A.S. Fraenkel [1984], A linear algorithm for nonhomogeneous spectra of numbers, *J. of Algorithms* **5**, 187–198.

[6] H.S.M. Coxeter [1953], The golden section, phyllotaxis and Wythoff's game, *Scripta Math.* **19**, 135–143.

[7] A.S. Fraenkel [1982], How to beat your Wythoff games' opponent on three fronts, *Amer. Math. Monthly* **89**, 353–361.

[8] A.S. Fraenkel [1994], Iterated floor function, algebraic numbers, discrete chaos, Beatty subsequences, semigroups, *Trans. Amer. Math. Soc.* **341**, 639–664.

[9] A.S. Fraenkel [1998], Heap games, numeration systems and sequences, *Ann. Comb.* **2**, 197–210. Appeared first in: *Fun With Algorithms*, *Proceedings in Informatics* (E. Lodi, L. Pagli and N.

Santoro, eds.), Vol. 4, pp. 99–113, Carleton Scientific, University of Waterloo, Waterloo, Ont., 1999.

[10] A.S. Fraenkel, T. Seeman and J. Simpson [2001], The subword complexity of a two-parameter family of sequences, *Electron. J. Combin.* **8(2)**, R10, 19pp.
http://www.combinatorics.org/

[11] R.L. Graham, S. Lin and C.-S. Lin [1978], Spectra of numbers, *Math. Mag.* **51**, 174–176.

[12] F. Mignosi [1990], Sturmian words and ambiguous context-free languages, *Internat. J. Found. Comput. Sci.* **1**, 309–323.

[13] R. Tijdeman [2000], Fraenkel's conjecture for six sequences, *Discrete Math.* **222**, 223–234.

[14] W.A. Wythoff [1907], A modification of the game of Nim, *Nieuw Arch. Wisk.* **7**, 199–202.

[15] A.M. Yaglom and I.M. Yaglom [1967], *Challenging Mathematical Problems with Elementary Solutions* (translated by J. McCawley, Jr.; revised and edited by B. Gordon), Vol. II, Holden-Day, San Francisco, 1967.