Characterisation of asymptotically Sturmian sequences

Alex Heinis and Rob Tijdeman

December 14, 1999

Dedicated to Kálmán Györy
on the occasion of his sixtieth birthday

Abstract. Sturmian sequences are both balanced and stiff. First we apply classical results of Morse, Hedlund and Coven to characterise the Sturmian, balanced and stiff $\mathbf{N}$-sequences in terms of Beatty $\mathbf{N}$-sequences $\{\lfloor na+b\rfloor\}_{n\in\mathbf{N}}$. Recently Nakashima, Tamura and Yasutomi showed that asymptotically balanced sequences are asymptotically stiff and conversely. We provide an explicit characterisation of asymptotically balanced $\mathbf{N}$-sequences in terms of sequences of the form $\{\lfloor na+f_n\rfloor\}_{n=1}^{\infty}$ where $\{f_n\}_{n=1}^{\infty}$ is a real sequence with $f_{n+1}-f_n\to 0$ as $n\to\infty$. We conclude with a characterisation of asymptotically Sturmian $\mathbf{N}$-sequences.

Keywords: Sturmian, balanced, stiff, Beatty, low complexity, recursive.

1 Introduction.

Let $\mathbf{N} = \{1, 2, 3, \cdots\}$. Let $S = \{s_n\}_{n\in\mathbf{N}}$ be a strictly increasing sequence of positive integers, a so-called $\mathbf{N}$-sequence. We call S a positive Beatty sequence if there exist real constants $a \geq 1$ and b such that $s_n = \lfloor an + b \rfloor$ for all n and $b < 1 \leq a + b$, or $s_n = \lceil an + b \rceil$ for all n and $b \leq 0 < a + b$. If a is rational, then every positive Beatty sequence of the form $\{\lceil an + b \rceil\}_{n\in\mathbf{N}}$ can be written as a positive Beatty sequence of the form $\{\lfloor an + b \rfloor\}_{n\in\mathbf{N}}$, but this is not true if a is irrational. The sequence S is called balanced if the number of terms of S in any two intervals $[x, x + n)$ and $[y, y + n)$ with $n, x, y \in \mathbf{N}$ differ by at most 1. A simple calculation yields that every positive Beatty sequence is balanced. Let $\sigma : \mathbf{N} \to \{0, 1\}$ be the characteristic function of S,

$$\sigma(s) = \begin{cases} 1 & \text{if } s \in S \\ 0 & \text{otherwise} \end{cases}$$

We call σ the $\mathbf{N}$-word corresponding with S. For $n \in \mathbf{N}$ denote the number of distinct vectors among $(\sigma(x), \sigma(x+1), \cdots, \sigma(x+n-1))$ for $x \in \mathbf{N}$ by the complexity number $P(n)$. The sequence S is called stiff if $P(n) \leq n + 1$ for every n. Morse and Hedlund [MH, Th. 3.5-3.7] proved that every balanced sequence is stiff and Coven and Hedlund [CH, Th. 4.09] that every periodic stiff sequence is balanced. Morse and Hedlund [MH, Th. 2.3.] further proved that every stiff sequence has a density $\alpha = \lim_{x\to\infty} x^{-1}|\{s \in S \mid 0 < s \leq x\}|$ and even that there is a $b \in \mathbf{R}$ such that $\{s_n - \alpha^{-1}n - b\}_{n\in\mathbf{N}}$ is bounded by 1. The sequence S is called almost periodic if for every $n \in \mathbf{N}$ there exists a $C \in \mathbf{N}$ such that for all $x, y \in \mathbf{N}$ there is a $z \in \mathbf{N}$ with $y \leq z < y + C$ and $\sigma(x + i) = \sigma(z + i)$ for $i = 0, 1, \cdots, n - 1$. We call S Sturmian if S is both balanced and almost periodic. (Various authors use the word 'Sturmian' in different meanings!)

The behaviour is rather different for rational and irrational density. If S is a sequence with irrational density, then the following properties are equivalent: (i) S is Sturmian, (ii) S is balanced, (iii) S is stiff, (iv) S is positive Beatty, (v) $P(n) = n + 1$ for every n (cf. [MH, Th. 7.2]). It follows that for every stiff sequence with rational density there exists an n with $P(n) \leq n$. For sequences S with rational density the following properties are equivalent: (i) $P(n) \leq n$ for some n, (ii) $\{P(n)\}_{n\in\mathbf{N}}$ is bounded, (iii) S is ultimately periodic, i.e. for suitable $c, k, n_0 \in \mathbf{N}$ we have $s_{n+k} - s_n = c$ for $n > n_0$. Obviously not all sequences with these properties are balanced or stiff and the example $00011111\cdots$ shows that balanced and stiff are not equivalent in the rational case. As far as I know, no explicit characterisation of balanced and stiff **N**-sequences in terms of Beatty sequences is available in the literature. However, it is easy to deduce the following theorem from known results. I recall the characterisation of stiff sequences with irrational density given above. We therefore restrict ourselves to stiff sequences with rational densities. Note that a stiff sequence with density 0 contains at most one term. We deduce Theorem 1 from the characterisation of stiff **Z**-sequences in [T] where it was assumed without mention that both S and its complement are bi-infinite sequences. Here we assume that S is an **N**-sequence.

Theorem 1. *Let S be a stiff **N**-sequence with rational density k/r where k and r are coprime positive integers. Then one of the following cases holds true:*
(i) (periodic case) *there exists an integer t with $k - r \leq t < k$ such that*

$$s_n = \lfloor \frac{nr+t}{k} \rfloor \quad for\ n \geq 1,$$

(ii) (skew case) *there exist integers $\delta \in \{-1, 1\}, t$ and v with $k - r - \delta \leq t < k - \delta, v > 0, vr + t \equiv 0 (\mathrm{mod}\ k)$ if $\delta = -1$, and $vr + t \equiv -1 (\mathrm{mod}\ k)$ if $\delta = 1$, such that*

$$s_n = \begin{cases} \lfloor \frac{nr+t}{k} \rfloor & for\ n > v \\ \lfloor \frac{nr+t+\delta}{k} \rfloor & for\ 1 \leq n < v+k \end{cases}$$

(iii) (Hedlund case) *there exist integers l, s, t, u, v with $0 < l \leq s, lr - ks = \pm 1, l - s \leq u < l$ and $v > 0$ such that*

$$s_n = \begin{cases} \lfloor \frac{nr+t}{k} \rfloor & \text{for } n > v \\ \lfloor \frac{ns+u}{l} \rfloor & \text{for } 1 \leq n < v+k+l \end{cases}$$

or there is some integer $m > 0$ such that $s_n = nr + m$ for $n \in \mathbf{N}$.

We conclude that S is a periodic stiff sequence if and only if S is positive Beatty. Case (i) is the only case in which S is almost periodic. Thus the Sturmian **N**-sequences are just the positive Beatty sequences.

The sequences $\{s_n\}_{n=1}^{\infty}$ given in Theorem 1 represent stiff $\mathbf{N}$-sequences provided that in (iii) the parametes l, s, t, u, v are chosen in such a way that the values of s_n coincide for $v < n < v + k + l$. This happens if and only if t and u are of the form $t = kp - rv - (\delta + 1)/2$ and $u = lp - sv + (\delta - 1)/2$ where $\delta = lr - ks$ and p is any integer.

We introduce the asymptotic analogues of the concepts defined above. We say that S is asymptotically balanced if for every $n \in \mathbf{S}$ there is an x_n such that the numbers of terms of S in any two intervals $[x, x+n)$ and $[y, y+n)$ with $x, y > x_n$ differ by at most 1. We call S asymptotically stiff if $P_{\infty}(S, n) \leq n+1$ for every n, where $P_{\infty}(S, n)$ is the cardinality of the set $\mathcal{P}_{\infty}(S, n)$ of vectors which occur infinitely often among $(\sigma(m), \sigma(m+1), \cdots, \sigma(m+n-1))$ for $m \in \mathbf{N}$. Nakashima, Tamura and Yasutomi [NTY] define *-Sturmian as the property that for every n the number of entries 1 in any two vectors from $\mathcal{P}_{\infty}(S, n)$ differ by at most 1. It is not hard to show that the notions asymptotically balanced and *-Sturmian are equivalent. Hence they [NTY, Th. 2.6] prove that asymptotically balanced and asymptotically stiff are equivalent for $\mathbf{N}$-sequences. They give two more equivalent conditions and derive some results about asymptotically balanced $\mathbf{Z}$-sequences. It turns out that asymptotically balanced $\mathbf{N}$-sequences have a density too. It is our aim to provide an explicit description of these sequences. Again there is a distinction between the rational and the irrational density case, but now the difference is more subtle.

Theorem 2. a) *The sequence S with rational density α is asymptotically balanced if and only if there exists an ultimately monotonic sequence $\{f_n\}_{n=1}^{\infty}$ with $f_{n+1} - f_n \to 0$ as $n \to \infty$ such that $s_n = \lfloor n\alpha^{-1} + f_n \rfloor$ for $n \in \mathbf{N}$.*
b) *The sequence S with irrational density α is asymptotically balanced if and only if there exists a sequence $\{f_n\}_{n=1}^{\infty}$ with $f_{n+1} - f_n \to 0$ as $n \to \infty$ such that $s_n = \lfloor n\alpha^{-1} + f_n \rfloor$ for $n \in \mathbf{N}$.*

We call S asymptotically almost periodic if for every $n \in \mathbf{N}$ there exist integers $C_n, x_n \in \mathbf{N}$ such that for all integers $x, y > x_n$ there exists a $z \in \mathbf{N}$ with $y \leq z < y + C_n$ and $\sigma(x+i) = \sigma(z+i)$ for $i = 0, 1, \cdots, n-1$. We call S asymptotically Sturmian if S is both asymptotically balanced and asymptotically almost periodic. We have the following characterisation for asymptotically Sturmian sequences.

Theorem 3. a) *If S has irrational density, then the following statements are equivalent:* (i) *S is asymptotically Sturmian,* (ii) *S is asymptotically balanced,* (iii) *S is asymptotically stiff.*
b) *If S has rational density α, then the following statements are equivalent:* (i) *S is asymptotically Sturmian,* (ii) *S is asymptotically balanced and ultimately periodic,* (iii) *there are $b \in \mathbf{R}, M \in \mathbf{N}$ such that $s_n = \lfloor \alpha^{-1} n + b \rfloor$ for $n \geq M$.*

For convenience we shall include direct proofs of some results which have been proved in a different way elsewhere.

2 Characterisation of stiff sequences.

We show first that every stiff word can be extended to a stiff $\mathbf{Z}$-word. This will enable us to deduce Theorem 1 from the corresponding result for $\mathbf{Z}$-words.

Let I be an interval of integers. A mapping $\sigma : I \to \{0,1\}$ is called an I-word. The word is called finite if I is finite. The length $|\sigma|$ of a finite I-word σ is defined as the number of integers in I. The number of integers $i \in I$ with $\sigma(i) = 1$ is called the content of σ and denoted by $c(\sigma)$. We denote the number of distinct subwords of σ of length n by $\mathcal{P}(\sigma, n)$ and its cardinality by $P(\sigma, n)$. A word σ is called balanced if $|c(A) - c(B)| \leq 1$ for every two subwords A, B of length n. Obviously every I-word induces a sequence of increasing elements of I corresponding to the values $i \in I$ with $\sigma(i) = 1$. The word σ is called balanced (stiff,...) if and only if the corresponding sequence S is balanced (stiff,...). If $\sigma_1 \sigma_2 ... \sigma_n$ is a word of n letters and $m \leq n$, then we call $\sigma_1 \cdots \sigma_m$ the first m-subword of σ and $\sigma_{n-m+1} \cdots \sigma_n$ the last.

The first lemma is the finite equivalent of the well known result that a $\mathbf{Z}$-word σ with $P(\sigma, n) \leq n$ for some n is purely periodic.

Lemma 1. *Let σ be a finite word with $P(\sigma, n) \leq n$. If both the first and the last $(n-1)$-subword of σ occur at least twice as subwords of σ, then σ is purely periodic with period $l \leq n$.*

Proof. The lemma is obvious for $n = 1$. Let $n > 1$. For $1 \leq i \leq n$ we define an injection $\mathcal{P}(\sigma, i) \hookrightarrow \mathcal{P}(\sigma, i+1)$ by mapping $w \in \mathcal{P}(\sigma, i)$ to some right extension $w* \in \mathcal{P}(\sigma, i+1)$. Since $P(\sigma,1)=2, P(\sigma,n) \leq n$ and $\mathcal{P}(\sigma,i)$ is non-decreasing in i, there is an index $i<n$ such that $P(\sigma,i)=P(\sigma,i+1)$. Then the injection above is a bijection and every $w \in \mathcal{P}(\sigma,i)$ has a unique right extension $w*$ with $* \in \{0,1\}$. Writing $[1,N]$ for the domain of σ we have $\sigma([1,i]) = \sigma([l+1,l+i])$ for some minimal $l>0$. Since right extensions of i-subwords are unique, we obtain that σ is purely periodic with period l. Then the i-subwords $\sigma([h+1,h+n])$ for $0<h\leq l$ are distinct and we have $l=P(\sigma,i)\leq P(\sigma,n)\leq n$. □

For **Z**-words σ it is well known that $P(\sigma,n)\leq n$ implies $P(\sigma,m)\leq n$ for $m \geq n$. The next lemma shows a similar result for finite stiff words.

Lemma 2. *Let σ be a finite stiff word with $P(\sigma,k)\leq k$.*
a) Then $\sigma=\beta\gamma\delta$ with $|\beta|+|\delta|<k$ such that γ has period $l\leq k-|\beta|-|\delta|$
b) Then $P(\sigma,m)\leq k$ for $m\geq k$.

Proof. a) Define $\sigma_0=\sigma$ and l_0 as the smallest l with $1\leq l\leq k$ and $P(\sigma,l)\leq l$. We define a sequence (σ_t,l_t) for $t\geq 0$ by induction. If the last (l_t-1)-word of σ_t does not occur elsewhere in σ_t, then we define σ_{t+1} as σ_t with the last letter removed. Otherwise, if the first (l_t-1)-word of σ_t does not occur elsewhere in σ_t, then we remove the first symbol of σ_t to obtain σ_{t+1}. In both cases we put $l_{t+1}=\min\{l|1\leq l\leq l_t, P(\sigma_{t+1},l)\leq l\}$. It follows from the definition of l_t that $P(\sigma_t,l_t-1)=l_t$ for every t and that $P(\sigma_{t+1},l_t-1)=P(\sigma_t,l_t-1)-1=l_t-1$. Hence $1\leq l_{t+1}<l_t$ and there is a pair (σ_{t_0},l_{t_0}), for which the process ends. Applying Lemma 1 to this pair we find that $\gamma:=\sigma_{t_0}$ has a period

$$l\leq l_{t_0}\leq l_0-t_0\leq k-t_0=k-|\beta|-|\delta|$$

where β is the subword for σ which precedes σ_{t_0} in σ and δ the subword after σ_{t_0} in σ.
b) $P(\sigma,m)\leq|\beta|+l_{t_0}+|\delta|\leq k$ by the periodicity of σ_{t_0}. □

We use Lemma 2 to show that there is no maximal finite stiff word.

Lemma 3. *Let σ be a finite stiff word. Then $\sigma 0$ or $\sigma 1$ is stiff.*

Proof. Put $n=|\sigma|$. If $P(\sigma,k)\leq k$ for all $k\leq n$, then $\sigma 1$ is stiff. Otherwise, let K be the maximal $k\leq n$ with $P(\sigma,k)=k+1$. Note that $K<n$ by

$P(\sigma, n) = 1$, and that $P(\sigma, K+1) = K+1$. By Lemma 2b) we infer that $P(\sigma, k) \leq K+1$ for $k > K$, whence $P(\sigma x, k) \leq k+1$ for $x \in \{0,1\}$ and $k > K$. If the last $(K-1)$-word in σ does not occur elsewhere as as subword of σ, then $P(\sigma', K-1) = P(\sigma, K-1) - 1 \leq K-1$ where σ' denotes σ with the last symbol removed. Then Lemma 2b) implies $P(\sigma', K) \leq K-1$, whence $P(\sigma, K) \leq K$, contradicting the choice of K. Therefore the last $(K-1)$-subword of σ appears elsewhere in σ, followed by a symbol x, say. For every $k \leq K$ we then have that the final k-word in σx occurs somewhere else in σx too. Hence $P(\sigma x, k) = P(\sigma, k) \leq k+1$. Thus σx is stiff. $\square$

Next we conclude that every stiff word can be extended to a stiff $\mathbf{Z}$-word.

Lemma 4. *Every stiff word can be extended to a stiff $\mathbf{Z}$-word.*

Proof. Suppose σ is a finite stiff word on $[1, n]$. By applying Lemma 3 inductively we find a stiff $\mathbf{N}$-word τ starting with σ. Suppose that 0τ and 1τ are both not stiff. Then $P(0\tau, i) \geq i+2$ for some i and $P(1\tau, j) \geq j+2$ for some j. Let T be a finite initial segment of τ such that $\mathcal{P}(0\tau, i) = \mathcal{P}(0T, i)$ and $\mathcal{P}(1\tau, j) = \mathcal{P}(1T, j)$. Then T is a finite stiff word admitting no stiff left extension. This contradicts the reflected form of Lemma 3. $\square$

The stiff $\mathbf{Z}$-words have been classified by Morse and Hedlund [MH] and Coven and Hedlund [CH]. A characterisation in terms of Beatty sequences has been stated in [T, Th.4], however under the unmentioned assumption that both S and its complement are bisequences (in the present notation: that the sequence corresponding to $\hat{\sigma}$ and its complement are both $\mathbf{Z}$-sequences). Here we state an unconditional classification.

Lemma 5. *Let $\hat{\sigma} = \{s_j\}_{j \in \mathbf{Z}}$ be a stiff $\mathbf{Z}$-word consisting of symbols* 0 *and* 1*. Let $\hat{S}$ be the set of $j \in \mathbf{Z}$ with $\sigma_j = 1$. If $\hat{S}$ has left and right density α, then $\hat{S}$ is balanced and of one of the following forms:*
(a) (periodic case) *there exists an integer r with $k := \alpha r \in \mathbf{N}$ and an integer t with such that $\hat{S} = \{\lfloor \frac{ir+t}{k} \rfloor\}_{i \in \mathbf{Z}}$*
(b) (irrational case) *α is irrational and there exists a $\mu \in \mathbf{R}$ such that*

$$\hat{S} = \{\lfloor i\alpha^{-1} + \mu \rfloor\}_{i \in \mathbf{Z}} \quad \text{or} \quad \hat{S} = \{\lceil i\alpha^{-1} + \mu \rceil\}_{i \in \mathbf{Z}}$$

(c) (skew case) *there exists an integer r with $k := \alpha r \in \mathbf{N}$ and an integer m such that* $\gcd(k, r) = 1$ *and*

$$\hat{S} = \{\lfloor \frac{ir}{k} \rfloor + m\}_{i<k} \cup \{\lfloor \frac{ir-1}{k} \rfloor + m\}_{i>0}$$

or

$$\hat{S} = \{\lfloor \frac{ir-1}{k} \rfloor + m\}_{i<k} \cup \{\lfloor \frac{ir}{k} \rfloor + m\}_{i>0}$$

(d) (finite case) $\hat{S} = \emptyset$ *or* $\hat{S} = \{m\}$ *for some* $m \in \mathbf{Z}$.
If $\hat{S}$ *does not have a density, then it has a left density* l/s *and a right density* k/r *with* k, l, r, s *integers subject to* $lr - ks = \pm 1$. *Then one of the following two cases holds:*
(e) (Hedlund case for positive densities) $k > 0, l > 0$ *and there is a* $m \in \mathbf{Z}$ *such that*

$$\hat{S} = \{\lfloor \frac{is}{l} \rfloor + m\}_{i<k+l} \cup \{\lfloor \frac{ir-1}{k} \rfloor + m\}_{i>0}$$

if $lr - ks = 1$ *and*

$$\hat{S} = \{\lfloor \frac{is-1}{l} \rfloor + m,\}_{i<k+l} \cup \{\lfloor \frac{ir}{k} \rfloor + m\}_{i>0}$$

if $lr - ks = -1$.
(f) (Hedlund case with some density 0) *there is some* $m \in \mathbf{Z}$ *such that*

$$\hat{S} = \{is + m\}_{i<l} \ \text{ or } \ \hat{S} = \{ir + m\}_{i>0}.$$

It follows from Lemma 5 that the almost periodic stiff **Z**-words are precisely those given by (a), (b) and by $\emptyset$ in (d), since the other words contain subwords which occur only once. Thus the Sturmian **Z**-sequences are precisely the Beatty sequences.

Proof of Theorem 1. Let S be a stiff **N**-sequence. Consider the corresponding stiff **N**-word σ. According to Lemma 4 σ is the restriction to **N** of a stiff **Z**-word $\hat{\sigma}$. We apply Lemma 5 to $\hat{\sigma}$. In case (a) the restriction S of $\hat{S}$ is a positive Beatty sequence with $s_0 < 1 \leq s_1$. This yields case (i).
Case (b) implies that S is a positive Beatty sequence with irrational density. In the former instance of case (c) put $\delta = 1$, define integers n and t by $ir - 1 + mk = rn + t$ with $k - r - 1 \leq t < k - 1$ and set $v = n - i$. Then $vr + t \equiv -1 (\bmod k)$ and s_n is as given in (ii). If $v \leq 0$, then the restriction S is periodic and covered by (i). In the latter instance of case (c) put $\delta = -1$, define n and t by $ir + mk = nr + t$ with $k - r + 1 \leq t < k + 1$ and put again $v = n - i$. The further reasoning is similar and completes case (ii). In case (d) S is not an **N**-sequence. In case (e) write $is + ml = ns + u$ in the former case and $is - 1 + ml = ns + u$ in the latter with $l - s \leq u < l$ in both cases.

Write $ir-1+mk = nr+t$ in the former case, $ir+mk = nr+t$ in the latter and $v = n-i$ in both cases. Then s_n is as given in (iii). If $v \leq 0$, then the restriction S is a positive Beatty sequence covered by (i). This covers the Hedlund case (iii) with $l > 0$. In case (f) S is finite if $\hat{S} = \{is+m\}_{i<l}$ and covered by case (i) if $\hat{S} = \{ir+m\}_{i>0}$ and $m \leq 0$. The remaining possibility yields the latter case of (iii). □

3 Characterisation of asymptotically stiff sequences.

We start with a direct proof of the equivalence for **N**-sequences of asymptotically balanced and asymptotically stiff.

Lemma 6. (Nakashima, Tamura, Yasumato, [NTY, Th. 2.6]) *An* **N**-*sequence S is asymptotically balanced if and only if it is asymptotically stiff.*

Proof. $\Longrightarrow$ Suppose S is not asymptotically stiff. Let σ be the corresponding **N**-word. Then there exists a minimal n such that $P_\infty(\sigma, n) > n+1$. Since $P_\infty(\sigma, 1) \leq 2$, we have $n > 1$. Hence $P_\infty(\sigma, n) - P_\infty(\sigma, n-1) \geq 2$. So there are at least two $(n-1)$-vectors, $(a_1^{(i)}, \cdots, a_{n-1}^{(i)}) \in \{0,1\}^{n-1}$, for $i = 1, 2$, say, such that both $(a_1^{(i)}, a_2^{(i)}, \cdots, a_{n-1}^{(i)}, 0)$ and $(a_1^{(i)}, a_2^{(i)}, \cdots, a_{n-1}^{(i)}, 1)$ are in $\mathcal{P}_\infty(\sigma, n)$. Let k be the largest index with $a_k^{(1)} \neq a_k^{(2)}$. Then both $(1, a_{k+1}^{(1)}, \cdots, a_{n-1}^{(1)}, 1)$ and $(0, a_{k+1}^{(1)}, \cdots, a_{n-1}^{(1)}, 0)$ are in $\mathcal{P}_\infty(\sigma, n-k+1)$. Thus S is not asymptotically balanced.
$\Longleftarrow$ Suppose S is asymptotically stiff, but not asymptotically balanced. Let σ be the **N**-word corresponding with S and let n be the smallest positive integer such that there exist $X, Y \in \mathcal{P}_\infty(\sigma, n)$ with $c(X) \geq c(Y) + 2$. Then $n \geq 2, X = 1v1, Y = 0w0$ and $c(X) = c(Y)+2$. If v and w are different, then $v = C\lambda D, w = C\overline{\lambda}E$ where C, D, E are words, λ is a symbol and $\overline{\lambda}$ is the other symbol. If $\lambda = 1$ then $1C1, 0C0$ are contained in $\mathcal{P}_\infty(\sigma, n)$ and contradict the minimality of n. If $\lambda = 0$ one uses $D1, E0$ to obtain a contradiction. We conclude that $X = 1v1, Y = 0v0$ for some finite word v. In particular X and Y are unique. Let N_n be a positive integer such that for every $x \geq N_n$

the vector $(\sigma(x), \cdots, \sigma(x+n-1))$ is in $\mathcal{P}_\infty(\sigma, n)$. Choose $x > y > N_n$ such that the subwords of σ with domain $[y, y+n), [x, x+n)$ contain $k-1$ and $k+1$ ones respectively and such that all intermediate n-subwords contain exactly k-ones. The words above $[y, y+n), [x, x+n)$ are Y, X and $x - y \geq n$ by the minimality of n. Define $\tau : [y, x+n) \to \{0, 1\}$ by $\tau(y) = 1, \tau(z) = \sigma(z)$ for $y < z < x+n-1, \tau(x+n-1) = 0$. Then $\tau = 1v0 \cdots 1v0$ is periodic with period n and every n-subword of τ contains exactly k ones where $k = c(Y) + 1$. We call a finite word primitive if it is not an integer power of a strictly smaller word. Observe that every finite word τ is the power of a unique primitive word π, called the root of τ, cf. [BP, Prop 3.1] for a proof. Let π be the root of τ of length d. The first d subwords of τ are all different and all contain dk/n ones. Let us suppose that $x - y > n$. Then these d subwords appear in σ too, hence $\mathcal{P}_\infty(\sigma, d) \geq d$. Since σ contains $0v0, 1v1$ infinitely often we find that $\mathcal{P}_\infty(\sigma, d)$ should also contain an element with more than dk/n ones and one with fewer than dk/n ones. Thus $P_\infty(\sigma, d) \geq d + 2$, contradicting the assumption that S is asymptotically stiff. Thus $x = y + n$. Applying the same argument with $y > x$ we find that some tail of σ equals $YXY\cdots$. This tail is periodic and asymptotically stiff. Hence it is stiff and applying [CH, Th. 4.09] we see that it is balanced. Therefore σ itself is asymptotically balanced which is a contradiction. $\square$

It is well known that if $S = \{s_n\}_{n \in \mathbf{N}}$ is balanced, then for every $h \in \mathbf{N}$ the difference $s_{n+h} - s_n$ can attain only two consecutive integers. Lemma 7 provides the corresponding property for asymptotically balanced sequences.

Lemma 7. *The sequence $S = \{s_n\}_{n \in \mathbf{N}}$ is asymptotically balanced if and only if for every h the difference $s_{n+h} - s_n$ can attain only two consecutive integers except for finitely many n.*

Proof. $\Longrightarrow$ Suppose there exists an h such that $s_{r+h} - s_r > n$ for infinitely many r and that $s_{r+h} - s_r < n$ for infinitely many r. Then there are infinitely many intervals $[x, x+n)$ with fewer than h terms from S and infinitely many intervals $[x, x+n)$ with more than h terms from S. Such an S is not asymptotically balanced.
$\Longleftarrow$ Suppose there exists an n and an h such that for infinitely many x the the interval $[x, x+n)$ contains more than h terms from S and for infinitely

many x the interval $[x, x+n)$ contains fewer than h terms from S. The former type of intervals induces infinitely many r with $s_{r+h} - s_r < n$, whereas the latter type induces infinitely many r with $s_{r+h} - s_r > n$. Thus $\{s_{r+h} - s_r\}_{r=1}^{\infty}$ is not ultimately restricted to two consecutive values. $\square$

The following lemma reveals which values can be attained.

Lemma 8. *Every asympotically balanced sequence $S = \{s_n\}_{n \in \mathbf{N}}$ has a density $\alpha > 0$. Let $h \in \mathbf{N}$. If $h\alpha \notin \mathbf{N}$, then $s_{n+h} - s_n$ equals $\lfloor h\alpha^{-1} \rfloor$ or $\lceil h\alpha^{-1} \rceil$ with the exception of only finitely many n. If $h\alpha \in \mathbf{N}$, then $s_{n+h} - s_n \in \{h\alpha^{-1}, h\alpha^{-1} - 1\}$ for all but finitely many n or $s_{n+h} - s_n \in \{h\alpha^{-1}, h\alpha^{-1} + 1\}$ for all but finitely many n. In particular, there exists an N_h such that*

$$|s_{n+h} - s_n - h\alpha^{-1}| \leq 1 \quad \textit{for} \quad n \geq N_h. \tag{2}$$

Proof. Let S be asymptotically balanced and let $h \in \mathbf{N}$. Then there exist N_h and k_h such that for $x > N_h$ the number of terms of S in the interval $[x, x+h)$ equals k_h or $k_h + 1$, by Lemma 7. Put $\beta = \limsup_{h \to \infty} k_h/h$ and $\gamma = \liminf_{h \to \infty} k_h/h$. Suppose $\epsilon := (\beta - \gamma)/3 > 0$. Choose an m with $k_m/m > \gamma + 2\epsilon$ and an $n > \epsilon^{-1}$ with $k_n/n < \gamma + \epsilon$. Consider an interval $[x, x+mn)$ with $x > \max(N_m, N_n)$. Then the number of terms of S in the interval is both at least $k_m n > (\gamma + 2\epsilon)mn$ and at most $m(k_n + 1) < ((\gamma + \epsilon)n + 1)m$. This implies $n < \epsilon^{-1}$ which is a contradiction. Thus $\beta = \gamma = \lim_{h \to \infty} k_h/h$ is the density of S.
Suppose $s_{n+h} - s_n$ equals k_h or $k_h + 1$ for all but finitely many n. Then $\alpha \geq (k_h + 1)^{-1}$. Hence $\alpha > 0$. Since the average value of $s_{n+h} - s_n$ is $h\alpha^{-1}$, we obtain from Lemma 7 that for $n \geq x_h$ the differences $s_{n+h} - s_n$ should be $\lfloor h\alpha^{-1} \rfloor$ or $\lceil h\alpha^{-1} \rceil$ if $h\alpha^{-1} \notin \mathbf{N}$ and otherwise either be in $\{h\alpha^{-1}, h\alpha^{-1} - 1\}$ or in $\{h\alpha^{-1}, h\alpha^{-1} + 1\}$. $\square$

Proof of Theorem 2a) $\Longleftarrow$ Suppose there exists an ultimately monotonically non-decreasing sequence $\{f_n\}_{n=1}^{\infty}$ with $f_{n+1} - f_n \to 0$ as $n \to \infty$ such that $s_n = \lfloor n\alpha^{-1} + f_n \rfloor$ for $n \in \mathbf{N}$. We write $s_n = n\alpha^{-1} + f_n - \delta_n$ with $0 \leq \delta_n < 1$. Let k be the numerator of α and $\epsilon = k^{-1}$. Let $h \in \mathbf{N}$. Then there is an M_h such that $0 \leq f_{n+h} - f_n < \epsilon$ for $n \geq M_h$. Hence

$$-1 < s_{n+h} - s_n - h\alpha^{-1} < 1 + \frac{1}{k}.$$

Since $k(s_{n+h}-s_n-h\alpha^{-1}) \in \mathbf{Z}$, it should be contained in $(-k,k]$. If $h\alpha^{-1} \in \mathbf{N}$, then $s_{n+h}-s_n$ equals $h\alpha^{-1}$ or $h\alpha^{-1}+1$ for $n \geq M_h$. If $h\alpha^{-1} \notin \mathbf{N}$, then $s_{n+h}-s_n$ equals $\lfloor h\alpha^{-1} \rfloor$ or $\lceil h\alpha^{-1} \rceil$ for $n \geq M_h$. Thus $\{s_{n+h}-s_n\}_{n\in\mathbf{N}}$ can attain only two consecutive integers except for finitely many n. If $\{f_n\}_{n=1}^{\infty}$ is non-increasing, the proof is similar. According to Lemma 7, S is asymptotically balanced.

$\Longrightarrow$ Suppose S is asymptotically balanced with rational density $\alpha = k/r$ with $k, r \in \mathbf{N}$ and $\gcd(k,r) = 1$. According to Lemma 8 for every $h \in \mathbf{N}$ there exists an N_h such that $s_{n+h}-s_n$ can attain only two consecutive integers t_h and t_h+1 for $n \geq N_h$ where $t_h \leq hr/k \leq t_h+1$. In particular $t_k = r-1$ or $t_k = r$. We shall assume $t_k = r$. The proof in case $t_k = r-1$ is similar. We first show that the gaps between the n for which $s_{n+k}-s_n \neq r$ grow indefinitely.

Lemma 9. *Let $\{n_i\}$ be the sequence of n such that $s_{n+k}-s_n \neq r$. Then $\{n_i\}$ is finite or $\lim_{i\to\infty}(n_{i+1}-n_i) = \infty$.*

Proof. Suppose $\{n_i\}$ is infinite and $\liminf_{i\to\infty}(n_{i+1}-n_i) < \infty$. Then there exists an $m \in \mathbf{N}$ such that $n_{i+1}-n_i = m$ occurs infinitely often. If $m \geq k$, we have on one hand $s_{n_i+k+m}-s_{n_i} \leq (k+m)\alpha^{-1}+1$ and on the other hand

$$s_{n_i+k+m}-s_{n_i} = 2(r+1)+s_{n_i+m}-s_{n_i+k} > 2r+2+(m-k)\alpha^{-1}-1.$$

Hence $r+1 > r+1$ which is a contradiction. Thus $m < k$. Then

$$\begin{aligned} s_{n_i+k}-s_{n_i+m} &= (s_{n_i+k+m}-s_{n_i+m})+(s_{n_i+k}-s_{n_i})-(s_{n_i+k+m}-s_{n_i}) \\ &\geq 2(r+1)-(k+m)\alpha^{-1}-1 = (k-m)\alpha^{-1}+1. \end{aligned}$$

Since the right-hand side is not an integer, this yields another contradiction. $\square$

Recall that $s_{n+k}-s_n \geq k\alpha^{-1}$ for $n \geq N_k$. Put $\phi_n = s_n - n\alpha^{-1}$ for every n. Then $\phi_{n+k}-\phi_n \in \{0,1\}$ for $n \geq N_k$ and the gaps between consecutive 1's increase without bound by the previous lemma. Suppose $\phi_m = \max_{m-k<n\leq m}\phi_n$ and $m-k > N_k$. By Lemma 8, s_m-s_n is either $\lfloor (m-n)\alpha^{-1} \rfloor$ or $\lceil (m-n)\alpha^{-1} \rceil$ so that $0 \leq \phi_m-\phi_n < 1$ for $m-k < n \leq m$. Let $t > m$ be the smallest integer with $\phi_t-\phi_{t-k} = 1$. Then $\phi_n = \phi_{n-k}$ for $m < n < t$ Hence, by Lemma 8,

$$0 < \phi_t-\phi_m \leq \phi_t-\phi_n < 1 \quad \text{for} \quad t-k < n < t. \tag{3}$$

For the first inequality we used that $\phi_t = \phi_{t-k} + 1 = \phi_\nu + 1$ for some $m - k < \nu \leq m$ and for the second one we used that $\phi_n = \phi_\nu$, also for some $m - k < \nu \leq m$.

Choose u such that $t - k < u < t$ and $(u - t)r \equiv 1 (\mathrm{mod}\ k)$. Then $\phi_t - \phi_u \equiv \frac{(u-t)r}{k} \equiv \frac{1}{k} (\mathrm{mod}\ 1)$. By (3) with $n = u$, we obtain $\phi_t - \phi_u = \frac{1}{k}$. It follows that $0 < \phi_t - \phi_m \leq \frac{1}{k}$. By $k\alpha^{-1} \in \mathbf{N}$, the difference equals $\frac{1}{k}$. Thus

$$\phi_t - \phi_m = \frac{1}{k}. \tag{4}$$

We shall now define $\{f_n\}_{n \in \mathbf{N}}$. Let $N > N_{2k}$. Choose the largest m such that ϕ_m is maximal for $N < m \leq N + 2k$. Since $\phi_{n+k} - \phi_n \geq 0$, we have $m \geq N + k$. Put $f_n = \phi_n$ for $n \leq m$. For $n > m$ we define f_n by induction. Suppose m is the largest index for which f_m has been defined yet and for which $f_m = \phi_m, \phi_n \leq \phi_m$ for $m - k < n \leq m$ and $s_n = \lfloor n\alpha^{-1} + f_n \rfloor$ for $n \leq m$. Put $t = \min\{n > m | \phi_n - \phi_{n-k} = 1\}$. If t does not exist, then $\phi_n - \phi_{n-k} = 0$ for all $n > m$ and we put $f_n = f_m$ for $n > m$. It follows that $\{f_n\}_{n \in \mathbf{N}}$ is ultimately monotonic, that $f_{n+1} - f_n \to 0$ as $n \to \infty$ and that $s_n = s_{n-k} + r = \lfloor (n-k)\alpha^{-1} + f_{n-k} \rfloor + r = \lfloor n\alpha^{-1} + f_n \rfloor$ for $n > m$ by induction. If, on the other hand, there exists such a t, then we put

$$f_n = f_m + \frac{n - m}{(t - m)k} \qquad \text{for } m < n \leq t. \tag{5}$$

In particular, $f_t = f_m + \frac{1}{k} = \phi_m + \frac{1}{k} = \phi_t$, by (4). By (3) we have $\phi_n \leq \phi_m < \phi_t$ for $m - k < n < t$. As shown before we have $\phi_t < \phi_n + 1$, hence

$$\phi_m - 1 < \phi_t - 1 < \phi_n \leq \phi_m \qquad \text{for } m - k < n < t.$$

Since $\phi_m = f_m$, this implies $f_m + n\alpha^{-1} - 1 < s_n \leq f_m + n\alpha^{-1}$. Hence $s_n = \lfloor n\alpha^{-1} + f_m \rfloor$. By (5) we have $0 \leq f_n - f_m < \frac{1}{k}$ for $m < n < t$, and together with $k\alpha^{-1} \in \mathbf{N}$ we find that $s_n = \lfloor n\alpha^{-1} + f_n \rfloor$ for these n. We further know $s_t = \phi_t + t\alpha^{-1} = t\alpha^{-1} + f_t$. This completes the inductive definition of $\{f_n\}_{n \in \mathbf{N}}$.
¿From the definition it is obvious that $\{f_n\}_{n \in \mathbf{N}}$ is ultimately non-decreasing. Furthermore $f_{n+1} - f_n \leq \frac{1}{(t-m)k}$ and $t - m$ tends to ∞ as $n \to \infty$ by Lemma 9. Thus $f_{n+1} - f_n \to 0$ as $n \to \infty$. □

Proof of Theorem 2b). $\Longleftarrow$ Suppose $\{f_n\}_{n\in\mathbf{N}}$ is such that $s_n = \lfloor \alpha n + f_n \rfloor$ for some $\alpha \notin \mathbf{Q}$ and $f_{n+1} - f_n \to 0$ as $n \to \infty$. Take $h \in \mathbf{N}$ and let $n \geq N_h$. Then $s_{n+h} - s_n = \alpha^{-1}h + f_{n+h} - f_n + \delta_n$ with $|\delta_n| < 1$. Put $\epsilon = |\alpha^{-1}h - \mathbf{Z}|$, the distance from $\alpha^{-1}h$ to the nearest integer. Choose M_h so large that $|f_{n+h} - f_n| < \epsilon$ for $n \geq M_h$. Then, for $n > \max(M_h, N_h)$,

$$\lfloor \alpha^{-1}h \rfloor - 1 \leq \alpha^{-1}h - 1 - \epsilon < s_{n+h} - s_n < \alpha^{-1}h + 1 + \epsilon \leq \lceil \alpha^{-1}h + 1 \rceil.$$

Thus $s_{n+h} - s_n$ can attain only the values $\lfloor \alpha^{-1}h \rfloor$ and $\lceil \alpha^{-1}h \rceil$ for $n \geq \max(M_h, N_h)$. By Lemma 7, S is asymptotically balanced.
$\Longrightarrow$ Suppose $\{s_n\}_{n\in\mathbf{N}}$ is asymptotically balanced with density $\alpha \notin \mathbf{Q}$. Then by Lemma 8, $s_{n+h} - s_n$ assumes only the values $\lfloor h\alpha^{-1} \rfloor$ and $\lceil h\alpha^{-1} \rceil$ for $n \geq N_h$. Without loss of generality we assume that $N_h \geq 2N_{h-1}$ for all h.
We give an inductive definition of $\{f_n\}_{n\in\mathbf{N}}$. Put $\phi_n = s_n - n\alpha^{-1}$ for all n and $f_n = \phi_n$ for $n \leq N_1$. Suppose m is the largest index for which f_m has been defined yet and that $s_n = \lfloor n\alpha^{-1} + f_n \rfloor$ for $n \leq m$. Then we define $t > m$ as the smallest integer such that $\phi_t > f_m$ or $\phi_t \leq f_m - 1$. If no such t exists, then we put $f_n = f_m$ for $n > m$ and have $s_n = \lfloor n\alpha^{-1} + f_n \rfloor$ for $n \in \mathbf{N}$. Since nothing remains to be proved in this case, we assume that t exists. Define h by $N_h \leq m < N_{h+1}$. Put $f_n = f_m$ for $m < n < t, f_t = \phi_t$ if $\phi_t > f_m$ and $f_t = \phi_t + 1 - \frac{1}{h}$ otherwise. Then $s_n = \lfloor n\alpha^{-1} + f_n \rfloor$ for $n \leq t$. This completes the inductive definition of $\{f_n\}_{n\in\mathbf{N}}$.

It remains to show that $f_n - f_{n-1} \to 0$ as $n \to \infty$. It suffices to show that in the above notation $f_t - f_{t-1} \to 0$ as $t \to \infty$ (since otherwise $f_n = f_{n-1}$). Fix $\epsilon > 0$ and let $\{\frac{p_n}{q_n}\}_{n\in\mathbf{N}}$ be the sequence of convergents of α^{-1}. Choose m so large that the corresponding h is at least $\frac{1}{\epsilon+1}$ and that for the integer n, defined by $q_n \leq h < q_{n+1}$, we have $|q_{n-2}\alpha - \mathbf{Z}| < \epsilon$. Observe that $t - h \geq N_h - h \geq N_h - 2^{h-1} \geq N_h - N_{h-1} \geq N_{h-1}$. We use that, by Lemma 8,

$$|\phi_t - \phi_u| < 1 \quad \text{for} \quad t - h \leq u < t. \tag{6}$$

First suppose $f_t > f_{t-1}$. Let u be maximal such that $t - h \leq u < t$ and $f_u < f_{u-1}$, if such u exist. This implies that this u is of type t and that $f_u = \phi_u + 1 - 1/(h-1)$ in view of $u \geq N_{h-1}$. Then $f_u = \phi_u + 1 - 1/(h-1) > \phi_t - \epsilon = f_t - \epsilon$ by (6). Since $f_t > f_{t-1} \geq f_u$, it follows that $0 < f_t - f_{t-1} < \epsilon$. If no such u exists, then $f_u \geq f_{u-1}$ for $t - h < u < t$. Choose $q = q_n$ or q_{n-1}

so that $q\alpha^{-1} - \lfloor q\alpha^{-1} \rfloor > 1 - \epsilon$. We have, by (6), $\phi_t - 1 < \phi_{t-q} < \phi_t$. Since $\phi_t - \phi_{t-q} \equiv -q\alpha^{-1} (\bmod 1)$, we obtain that the left-hand side is in between 0 and ϵ. Hence $0 < f_t - f_{t-q} < \epsilon$, which implies $0 < f_t - f_{t-1} < \epsilon$ by the monotonicity of f_u. We conclude that $|f_t - f_{t-1}| < \epsilon$ if $f_t > f_{t-1}$.

Now suppose $f_t < f_{t-1}$. By a similar argument we find that $f_t < f_{t-1} < f_t + \epsilon$ both in case there is a u with $t - h \le u < t$ with $f_u > f_{u-1}$ and in case there is no such u. Hence $|f_t - f_{t-1}| < \epsilon$ if $f_t < f_{t-1}$. Combining all cases we find that $|f_t - f_{t-1}| < \epsilon$ if t is sufficiently large. Thus $f_n - f_{n-1} \to 0$ as $n \to \infty$. □

Proof of Theorem 3a). The equivalence of (ii) and (iii) follows from Lemma 6. By the definition of Sturmian (i) implies (ii). We shall prove that (iii) implies (i). Suppose S is asymptotically stiff and σ is the corresponding word. Let N_j be such that σ is balanced and stiff beyond N_j for subwords of length j. Fix some $h \in \mathbf{N}$. Choose $\epsilon > 0$ so that $|j\alpha - \mathbf{Z}| > \epsilon$ for $1 \le j \le h$. Put $N = \max_{j \le 3h^2\epsilon^{-1}+2} N_j$. Choose some element from $\mathcal{P}_\infty(\sigma, h)$ and suppose it does not occur as a subword $\sigma(n)\sigma(n+1)\cdots\sigma(n+h-1)$ for $a \le n < b$ where $a > N$ and $b = a + \lceil 3h^2\epsilon^{-1} \rceil$. Then $\sigma(a)\sigma(a+1)\cdots\sigma(b)$ is a finite word τ with $P(\tau, h) \le h$ and $3h^2\epsilon^{-1} < |\tau| < 3h^2\epsilon^{-1} + 2$. According to Lemma 2a we have $\tau = \beta\gamma\delta$ with $|\beta| + |\delta| \le h$ and γ has period $l \le h - |\beta| - |\delta|$. Let d be the number of ones in such a period l. Then

$$\begin{aligned}\Big|c(\tau) - |\tau|\frac{d}{l}\Big| &\le \Big|c(\beta) - |\beta|\frac{d}{l}\Big| + \Big|c(\gamma) - |\gamma|\frac{d}{l}\Big| + \Big|c(\delta) - |\delta|\frac{d}{l}\Big| \\ &\le |\beta| + 2l + |\delta| \le 2h.\end{aligned}$$

Since S is asymptotically balanced and has some density α by Lemma 8, we have

$$\Big|c(\tau) - |\tau|\alpha\Big| \le 1, \qquad \text{since } \ a > N_{|\tau|}.$$

Hence

$$|\tau|\Big|\alpha - \frac{d}{l}\Big| \le 2h + 1.$$

Thus

$$|l\alpha - \mathbf{Z}| \le \frac{(2h+1)l}{|\tau|} < \frac{(2h+1)\epsilon}{3h} \le \epsilon$$

whereas $l \le h$. This contradiction shows that every interval $[a, b)$ with $a \ge N$ and $b - a > 3h^2\epsilon^{-1}$ contains all elements from $\mathcal{P}_\infty(\sigma, h)$. Thus S is almost periodic. □

Proof of Theorem 3b. It is clear that (iii) implies (ii) and (ii) implies (i). We prove that (i) implies (iii). By Theorem 2a) we know that there exists an ultimately monotonic sequence $\{f_n\}_{n \in \mathbf{N}}$ with $f_n - f_{n-1} \to 0$ as $n \to \infty$ such that $s_n = \lfloor n\alpha^{-1} + f_n \rfloor$ for $n \in \mathbf{N}$. Put $\alpha = k/r$ with $k, r \in \mathbf{N}$, gcd $(k, r) = 1$. If $\frac{i}{k} \leq f_n < \frac{i+1}{k}$ for $t \leq n \leq u$ and σ is the word corresponding with S, then $\sigma(n_t), \sigma(n_t+1), \cdots, \sigma(n_u)$ is a sequence with period r and with k ones in each period. By the stated properties of the sequence $\{f_n\}$, there exist arbitrarily long such intervals $[t, u]$. Since S is almost periodic, there exists an M such that for $t \geq M$ and $u - t \geq 2r$ all elements of $\mathcal{P}_\infty(\sigma, r)$ occur as subwords of the finite word $\sigma(n_t)\sigma(n_t + 1) \cdots \sigma(n_u)$. These subwords have unique right extension, as the added letter has to be the same as the left letter in order to keep the number of ones constant k. Thus $s_n = \lfloor n\alpha^{-1} + \frac{i}{k} \rfloor$ for $n \geq M$. Now take $b = i/k$. □

References

[BP] J. Berstel and D. Perrin, *Theory of codes*, Ac. Press (1985).

[CH] E.M. Coven and G.A. Hedlund, *Sequences with minimal block growth*, Math. Systems Th. **7**(1993), 138-153.

[MH] M. Morse and G.A. Hedlund, *Symbolic dynamics II: Sturmian trajectories*, Amer. J. Math. **62** (1940), 1-42.

[NTY] I. Nakashima, J. Tamura and S. Yasutomi, $*$-*Sturmian words and complexity*, preprint, 26 pp.

[R] R. Tijdeman, *Intertwinings of periodic sequences*, Indag. Math., N.S., **9** (1998), 113-122.

Mathematical Institute
Leiden University
P.O. Box 9512
2300 RA Leiden
The Netherlands
e-mail: heinis@wi.leidenuniv.nl, tijdeman@math.leidenuniv.nl