STURMIAN WORDS AND CONSTANT ADDITIVE COMPLEXITY

Julian Sahasrabudhe [1]
Department of Mathematics, University of Memphis, Memphis, Tennessee
jdshsrbd@memphis.edu

Received: 6/10/14, Accepted: 5/27/15, Published: 6/5/15

Abstract

Resolving a question of Banero, we show that for every integer $K > 1$, there exists a word with additive complexity identically K. This result is perhaps surprising in light of the rather strong restriction on the existence of words with constant *abelian* complexity, given in the work of Currie and Rampersad. To prove our result we generalize the notion of a sturmian word. We also pose some questions regarding the existence and structure of words with fixed additive complexity.

1. Introduction

The study of words with bounded abelian complexity was initiated by Richomme, Saari and Zamboni [9]. They conjectured that, for a positive integer K, recurrent words of constant abelian complexity identically K exist if and only if $K \leq 4$. This conjecture was proved by Currie and Rampersad [5]. Banero [3] asked for what K there exist words of fixed *additive* complexity, and gave constructions of words with fixed odd complexity. We show that, in sharp contrast to the case of abelian complexity, there exist recurrent words with additive complexity exactly K for every K.

2. Definitions and Results

An *infinite word* w, will be understood to be an infinite sequence w_n, $n \in \mathbb{N}$, with $w_n \in A$ where A is a finite set. We will say w is a word *over* the *alphabet* A. In this paper, we will only consider words over alphabets A of integers.

A *block* b of a word w is a sequence of consecutive elements of w, that is $b = w_k \cdots w_{k'}$ for $k \leq k'$. For a block $b = w_k \cdots w_{k'}$ of w, we define $\sum b := \sum_{i=k}^{k'} w_i$.

[1] This paper was written while the author was at Simon Fraser University

Finally, define a word w to be *recurrent* if every block that occurs in w occurs infinitely many times.

The *abelian complexity* of an infinite word was defined in Currie and Rampersad [5] as follows. Suppose that w is an infinite word on alphabet $A = \{a_1, \ldots, a_n\}$ and that b is a block of w. Define the *composition vector* of b to be the vector $C(b) \in \mathbb{Z}^n$ such that the ith coordinate counts the number of occurrences of the symbol a_i in the block b. We now define the *abelian complexity* of w, denoted $\phi^{ab}(w; n)$, as

$$\phi^{ab}(w; n) = |\{C(b) : b \text{ is a block of } w \text{ of length } n\}| .$$

Richomme, Saari and Zamboni [9] asked the following question: for what K is there an infinite, word with abelian complexity exactly K? That is, for what K is there a word with $\phi^{ab}(w; n) = K$, for all $n \in \mathbb{N}$? This question was resolved by Currie and Rampersad [5] with the following result.

Theorem 1. *There exists a recurrent word with abelian complexity exactly $K \in \mathbb{N}$ if and only if $K \leq 3$.*

To study a Ramsey-type problem posed by Halbeisen and Hungerbühler [6], Ardal, Brown, Jungić and Sahasrabudhe [1] defined the *additive complexity* of a infinite word. The *additive complexity* of an infinite word w over a finite set of integers is the function $\phi^+(w, n)$ that counts the number of distinct sums obtained by summing n consecutive symbols of w. More precisely, write $w = w_1 w_2 \cdots$ then define

$$\phi^+(w; n) = \left| \left\{ \sum_{i=l}^{l+n-1} w_i : l \in \mathbb{N} \right\} \right| .$$

Banero [3] asked for what K is there a recurrent word with constant additive complexity exactly K. In this note we answer this question. Somewhat surprisingly, the situation is quite different from that of *abelian* complexity.

Theorem 2. *For every $K \in \mathbb{N}, K > 1$, there is a recurrent word of additive complexity exactly K.*

We note that the restriction to recurrent words is needed to make the problem interesting. In particular, the rather dull word

$$w = 012 \cdots (K-2)(K-1)(K-1)(K-1) \cdots$$

has additive (and abelian) complexity exactly K.

3. d-Dimensional Sturmian Words

We first recall the definition of sturmian words. Let $\alpha > 0$ be an irrational real number and let δ be an arbitrary real. Define the sturmian word $w(\alpha;\delta) = w_1w_2\cdots$ as the sequence $w_n = \lfloor \alpha(n+1)+\delta \rfloor - \lfloor \alpha n + \delta \rfloor$ for $n = 1, 2, \ldots$. In what follows, we shall always take $\delta = 0$ and write $w(\alpha) = w(\alpha, 0)$. Sturmian words have been well studied and are known by several equivalent definitions. We refer the reader to [7] ,[8], and the references therein, for further information.

The construction we give can be viewed as a generalization of sturmian words. We define these words in such a way that their connection to additive complexity is clear. Suppose that one slides an interval of length $l \notin \mathbb{Z}$ along the real numbers. Observe that this interval will cover either $\lfloor l \rfloor$ or $\lfloor l \rfloor + 1$ integer points. In a similar way, we may imagine d distinct intervals of lengths $l_1, \ldots, l_d \notin \mathbb{Z}$ independently sliding on the real numbers. We can make a similar observation regarding the number of integer points covered. For emphasis, we make this observation precise.

Observation 1. *If $l_1, \ldots, l_d \notin \mathbb{Z}$ and $t_1, \ldots, t_d \in \mathbb{R}$ then*

$$\sum_{i=1}^{d} |[t_i, t_i + l_i] \cap \mathbb{Z}| = \sum_{i=1}^{d} \lfloor l_i \rfloor + a,$$

where $a \in \{0, \ldots, d\}$. Moreover, for each $a \in \{0, \ldots, d\}$ one can find values $t'_1, \ldots, t'_d \in \mathbb{R}$ such that $\sum_i |[t'_i, t'_i + l_i] \cap \mathbb{Z}| = \sum_i \lfloor l_i \rfloor + a$.

To prove Theorem 2, we aim to "discretize" the above observation in an appropriate way. The following basic result of Kronecker will allow us to do just this.

Theorem 3. *Let $\alpha_1, \ldots, \alpha_d \in \mathbb{R}$ be such that $1, \alpha_1, \ldots, \alpha_d$ are linearly independent over $\mathbb{Q}$. Then $\{(\{n\alpha_1\}, \ldots, \{n\alpha_d\})\}_{n\in\mathbb{N}}$ is dense in $(\mathbb{R}/\mathbb{Z})^d$, where $\{x\}$ denotes the fractional part of x.*

We are now in a position to give our construction. Given $\alpha_1, \ldots, \alpha_d \in \mathbb{R}$ with $1, \alpha_1, \ldots, \alpha_d$ linearly independent over $\mathbb{Q}$, we define a d-dimensional sturmian word $w = w(\alpha_1, \ldots, \alpha_d)$ $(= w_1w_2, \cdots)$ by

$$w_n = \sum_{i=1}^{d} |[n\alpha_i, (n+1)\alpha_i) \cap \mathbb{Z})| = \sum_{i=1}^{d} \lfloor (n+1)\alpha_i \rfloor - \sum_{i=1}^{d} \lfloor n\alpha_i \rfloor.$$

We now turn to analyse the additive complexity of d-dimensional sturmian words.

Lemma 1. *For $d > 1$, let w be a d-dimensional sturmian word. Then $\phi^+(w; n) = d + 1$.*

Proof. Let w be a d-dimensional sturmian word and let $b(n,k) = w_n \cdots w_{n+k}$ be a block of w. Notice that

$$\sum b(n,k) = \sum_{i=1}^{d} |[n\alpha_i, (n+k+1)\alpha_i) \cap \mathbb{Z})| \,. \tag{1}$$

Observation 1 tells us that this sum takes at most $d+1$ different values. So $\phi^+(w;n) \leq d+1$. To prove the converse, fix some value $a \in \{0,\ldots,d\}$. From Observation 1 again, we know that there exists $(t_1,\ldots,t_d) \in \mathbb{R}^d$ so that

$$\sum_i |[t_i, t_i + (k+1)\alpha_i) \cap \mathbb{Z})| = \sum_i \lfloor l_i \rfloor + a. \tag{2}$$

Also notice that shifting the intervals integer amounts preserves the equality in (2). That is, if $M_1,\ldots,M_d \in \mathbb{Z}$ we have

$$\sum_i |[t_i + M_i, t_i + (k+1)\alpha_i + M_i) \cap \mathbb{Z})| = \sum_i \lfloor l_i \rfloor + a.$$

We observe that the equality at (2) is also preserved by small perturbations. In other words, there exist $\epsilon_1,\ldots,\epsilon_d > 0$ so that for $0 \leq \delta_i < \epsilon_i$ we have that

$$\sum_i |[t_i - \delta_i, t_i + (k+1)\alpha_i - \delta_i) \cap \mathbb{Z})| = \sum_i \lfloor l_i \rfloor + a.$$

Now, by Kronecker's theorem, we know that there exists an n and $M_1,\ldots,M_d \in \mathbb{Z}$ so that $t_i + M_i - \epsilon_i < n\alpha_i < t_i + M_i$, for all $i = 1,\ldots,d$.

Thus there is an n with

$$\sum b(n,k) = \sum_{i=1}^{d} |[n\alpha_i, (n+k+1)\alpha_i)] \cap \mathbb{Z}| = \sum_i |[t_i - \delta_i, t_i + (k+1)\alpha_i - \delta_i) \cap \mathbb{Z})|$$

$$= \sum_i \lfloor l_i \rfloor + a \ ,$$

where $0 \leq \delta_i < \epsilon_i$. This shows that $\phi^+(w;n) \geq d+1$ as desired. □

4. Uniform Recurrence

In the previous section we established that $\phi^+(n;w) = d+1$ where w is a d-dimensional sturmian word. To complete the proof of Theorem 2 we must show that w is recurrent. We shall in fact prove more, namely that d-dimensional sturmian words are *uniformly* recurrent.

Fix an infinite word $w = w_1w_2\cdots$ and for $n, l \geq 1$ set $b(n,l) = w_n \cdots w_{n+l-1}$. We define the *distance* between two blocks $b(n,l)$ and $b(n',l')$ to be $|n-n'|$. An infinite word w is said to be *uniformly recurrent* if every block b that occurs in w, occurs infinitely often and the gap between consecutive occurrences of b is bounded by a constant depending only on b. More formally, for a block b occurring in w, define $R(b,w)$ to be the supremum over all distances of consecutive occurrences of the block b in w. A word is said to *uniformly recurrent* if $R(b,w) < \infty$ for all blocks b appearing in w.

In the discussion that follows, we will just consider the case $d = 1$. Let $0 < \alpha < 1$ be an irrational and we consider constructing the word $w = w(\alpha)$. As we have described above, $w(\alpha)_n$ is simply the number of integer points intersected by the interval $[n\alpha, (n+1)\alpha)$. Thus, we imagine moving a interval while leaving the integer lattice fixed. Alternatively, one can construct $w(\alpha)$ by fixing the interval $[0, \alpha)$ and moving the integer lattice in the opposite direction in increments of size α. In symbols we have

$$w_n = |[0,\alpha) \cap (\mathbb{Z} - n\alpha)\,|.$$

In what follows, this seems to be the natural way of thinking about $w(\alpha)$.
Given infinite words w, w' on an alphabet A, we define the product word $w \times w'$ to be the infinite word on $A \times A$ defined by

$$(w \times w')_n = (w_n, w'_n).$$

It is immediate that this operation $\times$ on words is associative so we may speak of words of the form $w_1 \times \cdots \times w_n$ without ambiguity. To show that $w(\alpha_1, \ldots, \alpha_n)$ is uniformly recurrent we shall show that the product word $w(\alpha_1) \times \cdots \times w(\alpha_d)$ is uniformly recurrent.

In the proof of the lemma that follows, we shall need the following variant of an old and basic lemma of Dirichlet. Given $0 < \delta < 1$ and $x \in (\mathbb{R}/\mathbb{Z})^N$ define a δ-box of x to be a set $B_\delta \subseteq (\mathbb{R}/\mathbb{Z})^N$, with $x \in B_\delta$ and of the form $B_\delta = I_1 \times \cdots \times I_N$ where each of the I_i are intervals (open, closed, half closed or half-open) of length δ.

Lemma 2. *Let $\alpha_1, \ldots, \alpha_N$ be real numbers and let $0 < \delta < 1$ and $k \in \mathbb{N}$. Then for any δ-box B_δ of 0 there exists $n \in \mathbb{N}$ such that $k < n < (\frac{1}{\delta})^N + k$ and $(\{n\alpha_1\}, \ldots, \{n\alpha_N\}) \in B_\delta$.*

Lemma 3. *Let $\alpha_1, \ldots, \alpha_d \in \mathbb{R}$ and let $w(\alpha_1), \ldots, w(\alpha_d)$ be 1-dimensional sturmian words. Then $w(\alpha_1) \times \cdots \times w(\alpha_d)$ is uniformly recurrent.*

Proof. Set $w = w(\alpha_1) \times \cdots \times w(\alpha_d)$ and fix a block $b = b(n_0, l) = w_{n_0} \cdots w_{n_0+l-1}$ of ω. We may assume that $0 < \alpha_i < 1$ for $i = 1, \ldots, d$. We need to show that $R(b, w)$ (the maximum distance between consecutive occurrences of b) exists, for all blocks b occurring in w. To see this we will define a transformation T on the space $X = (\mathbb{R}/\mathbb{Z})^{ld}$ and a point $x \in X$. The block $b(n, l)$ will be determined by what region of X that the point $T^n(x)$ lies in. In what follows we will often think of the space $(\mathbb{R}/\mathbb{Z})^{ld}$ as $(\mathbb{R}/\mathbb{Z})^l \times \cdots \times (\mathbb{R}/\mathbb{Z})^l$.
We start by defining the maps $T_i : (\mathbb{R}/\mathbb{Z})^l \to (\mathbb{R}/\mathbb{Z})^l$ for $i = 1, \ldots, d$ by

$$T_i(x_1, \ldots, x_l) = (x_1 + \alpha_i, x_2 + \alpha_i, \ldots, \ldots, x_l + \alpha_i)$$

We then define $T = T_1 \times \cdots \times T_d$ and take

$$x = (0, \alpha_1, 2\alpha_1, \ldots, (l-1)\alpha_1, 0, \alpha_2, 2\alpha_2 \ldots, (l-1)\alpha_2, \ldots, 0, \alpha_d, 2\alpha_d \ldots, (l-1)\alpha_d)$$

Now for each $i = 1, \ldots, d$, define an i-region of $(\mathbb{R}/\mathbb{Z})^l$ to be a set of the form $R_i = I_1 \times \cdots \times I_l$ where $I_j \in \{[0, \alpha_i), [\alpha_i, 1)\}$ for $j = 1, \ldots, l$. Notice that the collection of i-regions forms a partition of $(\mathbb{R}/\mathbb{Z})^l$. Now define a partition $\mathcal{P}$ of X as

$$\mathcal{P} = \{A_1 \times \cdots \times A_d : A_i \text{ } is \text{ } an \text{ } i-region\}$$

The crucial observation is that $b(n, l)$ is determined entirely by the region of $\mathcal{P}$ that $T^n(x)$ falls into. So set $I \in \mathcal{P}$ to be the region for which $T^{n_0}(x)$ falls into. The problem of bounding the recurrence time of $b(n_0, l)$ is reduced to bounding the number of applications of T to the point $T^{n_0}(x)$ that are required to return to I.
Now, to bound this return time, choose a δ-box B_δ for some $\delta > 0$ so that $T^{n_0}(x) \in B_\delta \subseteq I$. We now shift this box to the origin, i.e. put

$$B'_\delta = B_\delta - T^{n_0}(x).$$

By Lemma 2, for $k \in \mathbb{N}$ we have an integer n_k satisfying $k < n_k \leq (\frac{1}{\delta})^{ld} + k$ with

$$T^{n_k}(0) \in B'_\delta.$$

Thus $T^{n_k}(0) + T^{n_0}(x) \in D_\delta$. Recalling the definition of T, we see that this implies that $T^{n_k+n_0}(x) \in B_\delta$. This means that $b(n_k + n_0, l)$ is identical to $b(n_0, l)$. Hence we can conclude that $R(b) \leq 2(\frac{1}{\delta})^{ld}$ and that $w(\alpha_1) \times \cdots \times w(\alpha_d)$ is uniformly recurrent.

□

Lemma 4. *Let $\alpha_1, \ldots, \alpha_d, 1 \in \mathbb{R}$ be linearly independent over $\mathbb{Q}$ and let $w(\alpha_1, \ldots, \alpha_d)$ be the corresponding d-dimensional sturmian word. The word $w(\alpha_1, \ldots, \alpha_d)$ is uniformly recurrent.*

Proof. Simply notice that we may obtain the d-dimensional sturmian word by summing the coordinates of the product word $w(\alpha_1) \times w(\alpha_2) \times \cdots \times w(\alpha_d)$, which we know to be uniformly recurrent. $\square$

5. Some Further Questions

Banero [3] pointed out that one can actually construct recurrent words of fixed *odd* complexity with a different construction. Indeed, enumerate all words of finite length on the alphabet $\{0, \ldots, n\}$, and form an infinite word w by concatenating all of the finite words in the order specified by the enumeration. Then apply the morphism defined by

$$0 \to (0)(2n) \quad 1 \to (1)(2n-1) \quad \cdots \quad n \to (n)(n).$$

It is easy to check that the resulting word has additive complexity exactly $2n + 1$. This construction suggests that recurrent words with fixed additive complexity are much more abundant than the class of generalized sturmian words. Can a construction of this form be generalized to give constructions of words with fixed *even* complexity?

Before we ask our second question, we make an observation. Above we defined the nth symbol of $w(\alpha_1, \ldots, \alpha_n)$ as

$$\sum_{i=1}^{d} \lfloor (n+1)\alpha_i \rfloor - \sum_{i=1}^{d} \lfloor n\alpha_i \rfloor.$$

However, if we define $w = (w_n)_n$ by

$$w_n = \sum_{i=1}^{d} \lfloor (n+1)\alpha_i + \delta_i \rfloor - \sum_{i=1}^{d} \lfloor n\alpha_i + \delta_i \rfloor,$$

where $\delta_1, \ldots, \delta_n \in \mathbb{R}$ then one can show, by modifying our proof of Theorem 2, that w is also uniformly recurrent with additive complexity exactly $d+1$. Are there any uniformly recurrent words *not* of this type that have fixed additive complexity? The fact that there are no other such words for $n = 1$ is (in a slightly different form) the well known theorem of Hedlund and Morse. See [8] for the original proof and [7] for a modern presentation of this result.

Acknowledgements I would like to thank Tom Brown and Veselin Jungić for introducing me to this circle of problems and Béla Bollobás for comments on late drafts of the paper. I would also like to thank the anonymous referee.

References

[1] H. Ardal, T. Brown, V. Jungić, J. Sahasrabudhe, On additive and Abelian complexity in infinite Words, *Integers* **12** (2012).

[2] G. Banero, On additive complexity of infinite words, ArXiv, http://arxiv.org/abs/1209.4691 (2012).

[3] G. Banero, *Private communication*, (2012).

[4] E.M. Coven, G.A. Hedlund, Sequences with minimal block growth, *Math. Systems Theory*, **7** (1973), 138-153.

[5] J. Currie, M. Rampersad, Recurrent words with constant Abelian complexity, *Advances in Applied Mathematics* **47** (2011), 116-124.

[6] L. Halbeisen and N. Hungerbühler, An application of van der Waerdens theorem in additive number theory, *Integers* **0**, (2000).

[7] M. Lothaire, Algebraic Combinatorics on Words, Cambridge: Cambridge University Press. (2002).

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

[9] G. Richomme, K. Saari, L. Zamboni, Abelian complexity in minimal subshifts, *J. London Math. Soc.*, **83** (2011), 79-95.