On abelian and additive complexity in infinite words

Hayri Ardal, Tom Brown, Veselin Jungić, Julian Sahasrabudhe

July 2011

Abstract

The study of the structure of infinite words having *bounded abelian complexity* was initiated by G. Richomme, K. Saari, and L. Q. Zamboni [10]. In this note we define *bounded additive complexity* for infinite words over a finite subset of $\mathbb{Z}^m$. We provide an alternative proof of one of the results of [10].

1 Introduction

Recently the study of infinite words with *bounded abelian complexity* was initiated by G. Richomme, K. Saari, and L. Q. Zamboni [10]. (See also [3] and the references in [3] and [10].) In particular, it is shown (in [10]) that if ω is an infinite word with bounded abelian complexity, then ω has *abelian k-factors* for all $k \geq 1$. (All these terms are defined below.)

In this note we define *bounded additive complexity*, and we show in particular that if ω is an infinite word (whose alphabet is a finite subset S of $\mathbb{Z}^m$ for some $m \geq 1$) with bounded additive complexity, then ω has *additive k-factors* for all $k \geq 1$. As we shall see, this provides an alternative proof of the just-mentioned result concerning abelian k-factors.

We are motivated by the following question. In [6], [7], [8], and [9], it is asked whether or not there exists an infinite word on a finite subset of $\mathbb{Z}$ in which there do not exist two adjacent factors with equal lengths and equal sums. (The *sum* of the factor $x_1x_2\dots x_n$ is $x_1 + x_2 + \cdots + x_n$.) This question remains open, although some partial results can be found in [1], [2], [6].

2 Additive complexity

2.1 Infinite words on finite subsets of $\mathbb{Z}$

Definition 2.1. Let ω be an infinite word on a finite subset S of $\mathbb{Z}$. For a factor $B = x_1x_2\dots x_n$ of ω, $\sum B$ denotes the sum $x_1 + x_2 + \cdots + x_n$. Let

$$\phi_\omega(n) = \{\sum B : \ B \text{ is a factor of } \omega \text{ with length } n\}.$$

The function $|\phi_\omega|$ (where $|\phi_\omega|(n) = |\phi_\omega(n)|, n \geq 1$) is called the *additive complexity* of the word ω.

If $B_1B_2\cdots B_k$ is a factor of ω such that $|B_1| = |B_2| = \cdots = |B_k|$ and $\sum B_1 = \sum B_2 = \cdots = \sum B_k$, we call $B_1B_2\cdots B_k$ an *additive k-power.*

We say that ω has *bounded additive complexity* if any one (and hence all) of the three conditions in the following proposition (Proposition 2.1) hold.

Proposition 2.1. *Let ω be an infinite word on the alphabet S, where S is a finite subset of $\mathbb{Z}$. Then the following three statements are equivalent.*

1. There exists M_1 such that if B_1B_2 is a factor of ω with $|B_1| = |B_2|$, then $|\sum B_1 - \sum B_2| \leq M_1$.

2. There exists M_2 such that if B_1, B_2 are factors of ω (not necessarily adjacent) with $|B_1| = |B_2|$, then $|\sum B_1 - \sum B_2| \leq M_2$.

3. There exists M_3 such that $|\phi_\omega(n)| \leq M_3$ for all $n \geq 1$.

Proof. We will show that $1 \Leftrightarrow 2$ and $2 \Leftrightarrow 3$.

Clearly $2 \Rightarrow 1$. Now assume that 1 holds, that is, if B_1B_2 is any factor of ω with $|B_1| = |B_2|$, it is the case that $|\sum B_1 - \sum B_2| \leq M_1$. Now let B_1 and B_2 be factors of ω with $|B_1| = |B_2|$, and assume that B_1 and B_2 are non-adjacent, with B_1 to the left of B_2.

Thus, assume that

$$B_1A_1A_2B_2$$

is a factor of ω, where

$$|A_1| = |A_2| \text{ or } |A_1| = |A_2| + 1.$$

Let

$$C_1 = B_1A_1, C_2 = A_2B_2.$$

Then

$$|C_1| = |C_2| \text{ or } |C_1| = |C_2| + 1.$$

Now

$$\sum C_1 - \sum C_2 = (\sum B_1 + \sum A_1) - (\sum A_2 + \sum B_2),$$

or

$$\sum B_1 - \sum B_2 = (\sum C_1 - \sum C_2) + (\sum A_2 - \sum A_1).$$

Therefore, since A_1, A_2 and C_1, C_2 are adjacent, we have

$$|\sum A_2 - \sum A_1| \leq M_1 + \max S, \quad |\sum C_1 - \sum C_2| \leq M_1 + \max S,$$

and

$$|\sum B_1 - \sum B_2| \leq 2M_1 + 2\max S,$$

so that we can take $M_2 = 2M_1 + 2\max S$. Thus $1 \Rightarrow 2$.

Next we show that $2 \Rightarrow 3$. Thus we assume there exists M_2 such that whenever B_1, B_2 are factors of ω (not necessarily adjacent) with $|B_1| = |B_2|$, it is the case that $|\sum B_1 - \sum B_2| \leq M_2$.

Let n be given, and let $\sum B_1 = \min \phi_\omega(n)$. Then for any B_2 with $|B_2| = n$, we have $\sum B_2 = \sum B_1 + (\sum B_2 - \sum B_1)$. Therefore $\sum B_2 \leq \sum B_1 + M_2$. This means that $\phi_\omega(n) \subset [\sum B_1, \sum B_1 + M_2]$, so that $|\phi_\omega(n)| \leq M_2 + 1$.

Finally, we show that $3 \Rightarrow 2$. We assume there exists M_3 such that $|\phi_\omega(n)| \leq M_3$ for all $n \geq 1$. Suppose that B_1 and B_2 are factors of ω such that $|B_1| = |B_2| = n$ and $\sum B_1 = \min \phi_\omega(n)$, $\sum B_2 = \max \phi_\omega(n)$. To simplify the notation, for all $a \leq b$ let $\omega[a, b]$ denote $x_a x_{a+1} \ldots x_b$, and let us assume that $B_1 = \omega[1, n], B_2 = \omega[q+1, q+n]$, where $q > 1$.

For each $i, 0 \leq i \leq q$, let b_i denote the factor $\omega[i+1, i+n]$. Thus $B_1 = b_0, B_2 = b_q$, and the factor b_{i+1} is obtained by shifting b_i one position to the right. Clearly

$$\sum b_{i+1} - \sum b_i \leq \max S - \min S.$$

Since $|b_0| = |b_1| = \cdots = |b_q| = n$, and $|\phi_\omega(n)| \leq M_3$, there can be at most M_3 distinct numbers in the sequence $\sum B_1 = \sum b_0, \sum b_1, \ldots, \sum b_q = \sum B_2$. Let these numbers be

$$\sum B_1 = c_1 < c_2 < \cdots < c_r = \sum B_2,$$

where $r \leq M_3$.

Since $\sum b_{i+1} - \sum b_i \leq \max S - \min S,\ 0 \leq i \leq q$, it follows that $c_{j+1} - c_j \leq \max S - \min S,\ 0 \leq i \leq r-1$, and hence that

$$|\sum B_1 - \sum B_2| \leq (M_3 - 1)(\max S - \min S).$$

□

Theorem 2.2. *Let ω be an infinite word on a finite subset of $\mathbb{Z}$. Assume that ω has bounded additive complexity. Then ω contains an additive k-power for every positive integer k.*

Proof. Let $\omega = x_1 x_2 x_3 \cdots$ be an infinite word on the finite subset S of $\mathbb{Z}$, and assume that whenever B_1, B_2 are factors of ω (not necessarily adjacent) with $|B_1| = |B_2|$, then $|\sum B_1 - \sum B_2| \leq M_2$. (This is from part 2 of Proposition 2.1.)

Define the function f from $\mathbb{N}$ to $\{0, 1, 2, \ldots, M_2\}$ by

$$f(n) = x_1 + x_2 + x_3 + \cdots + x_n \pmod{M_2 + 1}, \quad n \geq 1.$$

This is a finite coloring of $\mathbb{N}$; by van der Waerden's theorem, for any k there are t, s such that

$$f(t) = f(t+s) = f(t+2s) = \cdots f(t+ks).$$

Setting
$$B_i = \omega[t + (i-1)s + 1, t + is], \quad 1 \le i \le k,$$
we have
$$\sum B_1 \equiv \sum B_2 \equiv \cdots \equiv \sum B_k \pmod{M_2 + 1}.$$

Since $B_1 B_2 \cdots B_k$ is a factor of ω with $|B_i| = |B_j|, 1 \le i < j \le k$, we have $|\sum B_i - \sum B_j| \le M_2$ and $\sum B_i \equiv \sum B_j \pmod{M_2+1}$, hence $\sum B_i = \sum B_j$.

Thus $|B_1| = |B_2| = \cdots = |B_k|$ and $\sum B_1 = \sum B_2 = \cdots = \sum B_k$, and ω contains the additive k-power $B_1 B_2 \cdots B_k$.

□

2.2 Infinite words on subsets of $\mathbb{Z}^m$

Let us use the notation $(u)_j$ for the jth coordinate of $u \in \mathbb{Z}^m$. That is, if $u = (u_1, \ldots, u_m)$ then $(u)_j = u_j$. Also, $|u| = |(u_1, \ldots, u_m)|$ denotes the vector $(|u_1|, \ldots, |u_m)|)$. In other words, $(|u|)_j = |(u)_j|$.

For factors B_1, B_2 of an infinite word ω on a finite subset S of $\mathbb{Z}^m$, the notation $|\sum B_1 - \sum B_2| \le M_1$ means that $(|\sum B_1 - \sum B_2|)_j \le M_1$, $1 \le j \le m$.

Now we suppose that ω is an infinite word on a finite subset S of $\mathbb{Z}^m$ for some $m \ge 1$. The definition of ϕ_ω and the additive complexity of ω is exactly as in Definition 1.1 above. The function
$$\phi_\omega(n) = \{\sum B : \ B \text{ is a factor of } \omega \text{ with length } n\}$$
is called the *additive complexity* of the word ω.

By working with the coordinates $(B_1)_j, (|\sum B_1 - \sum B_2|)_j$, we easily obtain the following results.

Proposition 2.3. *Proposition 2.1 remains true when $\mathbb{Z}$ is replaced by $\mathbb{Z}^m$.*

Theorem 2.4. *Let ω be an infinite word on a finite subset of $\mathbb{Z}^m$ for some $m \ge 1$. Assume that ω has bounded additive complexity. Then ω contains an additive k-power for every positive integer k.*

The following is a re-statement of Theorem 2.4, in terms of m infinite words on $\mathbb{Z}$, rather than one infinite word on $\mathbb{Z}^m$.

Theorem 2.5. *Let $m \in \mathbb{N}$ be given, and let $S_1, S_2, \ldots, S_m$ be finite subsets of $\mathbb{Z}$. Let ω_j be an infinite word on S_j with bounded additive complexity, $1 \le j \le m$. Then for all $k \ge 1$, there exists a k-term arithmetic progression in $\mathbb{N}, t, t+s, t+2s, \ldots, t+ks$ such that for all $j, 1 \le j \le m$,*
$$\sum \omega_j[t+1, t+s] = \sum \omega_j[t+s+1, t+2s] = \cdots = \sum \omega_j[t+(k-1)s+1, t+ks].$$
Thus $\omega_1, \omega_2, \cdots, \omega_m$ have "simultaneous" additive k-powers for all $k \ge 1$.

3 Abelian complexity

Definition 3.1. Let ω be an infinite word on a finite alphabet. Two factors of ω are called *abelian equivalent* if one is a permutation of the other. If the alphabet is $A = \{a_1, a_2, \ldots, a_t\}$, and the finite word B is a factor of ω, we write $\psi(B) = (u_1, u_2, \ldots, u_t)$, where u_i is the number of occurrences of the letter i in the word $B, 1 \leq i \leq t$. We call $\psi(B)$ the *Parikh vector* associated with B.

Let $\psi_\omega(n) = \{\psi(B) : B \text{ is a factor of } \omega, |B| = n\}$. The function ρ_ω^{ab}, defined by $\rho_\omega^{ab}(n) = |\psi_\omega(n)|, n \geq 1$, is called the *abelian complexity* of ω.

Thus $\rho_\omega^{ab}(n)$ is the largest number of factors of ω of length n, no two of which are abelian equivalent. If there exists M such that $\rho_\omega^{ab}(n) \leq M$ for all $n \geq 1$, then ω is said to have *bounded abelian complexity.*

The word $B_1B_2 \cdots B_k$ is called an *abelian k-power* if $B_1, B_2, \ldots, B_k$ are pairwise abelian equivalent. (Being abelian equivalent, they all have the same length.)

Recall that we are using the notation $|(u_1, u_2, \ldots, u_t)| \leq M$ to denote $|u_i| \leq M, 1 \leq i \leq t$.

Proposition 3.1. *Let ω be an infinite word on a t-element alphabet S. Then the following three statements are equivalent.*

1. There exists M_1 such that if B_1B_2 is a factor of ω with $|B_1| = |B_2|$, then $|\psi(B_1) - \psi(B_2)| \leq M_1$.

2. There exists M_2 such that if B_1, B_2 are factors of ω (not necessarily adjacent) with $|B_1| = |B_2|$, then $|\psi(B_1) - \psi(B_2)| \leq M_2$.

3. There exists M_3 such that such that $\rho_\omega^{ab}(n) \leq M_3$ for all $n \geq 1$.

Proof. We show that $1 \Leftrightarrow 2$ and $2 \Leftrightarrow 3$.

Clearly $2 \Rightarrow 1$. Now assume that 1 holds, that is, if B_1B_2 is any factor of ω with $|B_1| = |B_2|$, it is the case that $|\psi(B_1) - \psi(B_2)| \leq M_1$. Now let B_1 and B_2 be factors of ω with $|B_1| = |B_2|$, and assume that B_1 and B_2 are non-adjacent, with B_1 to the left of B_2.

Thus, assume that

$$B_1A_1A_2B_2$$

is a factor of ω, where

$$|A_1| = |A_2| \text{ or } |A_1| = |A_2| + 1.$$

Now we proceed exactly as in the proof of $1 \Rightarrow 2$ in Proposition 2.1, noting that $|\psi(A_1) - \psi(A_2)| \leq M_1 + 1$.

Next we show that $2 \Rightarrow 3$. Thus we assume there exists M_2 such that whenever B_1, B_2 are factors of ω (not necessarily adjacent) with $|B_1| = |B_2|$, it is the case that $|\psi(B_1) - \psi(B_2)| \leq M_2$.

Let n be given, and let $B_1 \in \psi_\omega(n)$. Then for any B_2 with $|B_2| = n$, we have $\psi(B_2) = \psi(B_1) + (\psi(B_2) - \psi(B_1))$. Therefore $|\psi(B_2)| \leq |\psi(B_1)| + M_2$. (This inequality is component-wise, that is, $(|\psi(B_2)|)_j \leq (|\psi(B_1)|)_j + M_2, 1 \leq j \leq t$.)

Therefore there are at most $2M_2 - 1$ choices for each component of B_2, and hence $\rho^{ab}_\omega(n) \le (2M_2 - 1)^t$.

Finally, we show that $3 \Rightarrow 2$. We assume there exists M_3 such that $\rho^{ab}_\omega(n) \le M_3$ for all $n \ge 1$.

Since $|\psi(xB) - \psi(By)| \le 1$ for all $x, y \in S$, it follows that if ω has factors B_1, B_2 of length n where for some $j, 1 \le j \le t, (\psi(B_1))_j = p$ and $(\psi(B_2))_j = p + q$, then ω has factors C_r of length n with $(\psi(C_r))_j = p + r, 0 \le r \le q$. (This is discussed in more detail in [10].) Thus $|\psi(B_1) - \psi(B_2)| \ge M_3$ implies $\rho^{ab}_\omega(n) \ge M_3 + 1$. Since we are assuming $\rho^{ab}_\omega(n) \le M_3, n \ge 1$, we conclude that $|\psi(B_1) - \psi(B_2)| \le M_3 - 1$ whenever $|B_1| = |B_2|$. Hence $|\psi(B_1) - \psi(B_2)| \le M_3 - 1$ whenever $|B_1| = |B_2|$.

$\square$

Remark 3.1. To see that bounded sum complexity is indeed weaker than bounded abelian complexity, consider the following example. Let $\sigma = x_1x_2x_3\cdots$ be the binary sequence constructed by Dekking [2] which has no abelian 4th power. In σ, replace every 1 by 12, and replace every 0 by 03, obtaining the sequence τ. If τ had an abelian 4th power $ABCD$, then the number of 2s in each of A, B, C, D are equal, and similarly for the number of 3s. But then dropping the 2s and 3s from $ABCD$ would give an abelian 4th power in σ, a contradiction. Hence τ does not have bounded abelian complexity. Now let a factor B of τ be given. By shifting B to the right or left, we see, by examining cases, that if $|B|$ is even then $\sum B = \frac{3}{2}|B| + s$, where $s \in \{-1, 0, 1\}$. If $|B|$ is odd, then $\sum B = \frac{3}{2}(|B| - 1) + s$, where $s \in \{0, 1, 2, 3\}$. Hence $|\phi_\tau(n)| \le 4$ for all $n \ge 1$, and τ does have bounded sum complexity.

Definition 3.2. Let $S = \{a_1, a_2, \ldots, a_m\}$ be a subset of $\mathbb{Z}$, and let $\omega = x_1x_2x_3\cdots$ be an infinite word on the alphabet S. For each $j, 1 \le j \le m$, let a'_j be the element of $\mathbb{Z}^m$ which has a_j in the in the jth coordinate and $0's$ elsewhere. Let $\omega' = x'_1x'_2x'_3\cdots$ be the word on the subset S' of $\mathbb{Z}^m, S' = \{a'_1, a'_2, \ldots, a'_m\}$, obtained from ω by replacing each a_j by a'_j, $1 \le j \le m$. It is convenient to visualize each a'_j as a column vector, rather than as a row vector.

Theorem 3.2. *Referring to Definition 2.2, consider the following statements concerning ω and ω':*

1. *ω has bounded abelian complexity.*
2. *ω' has bounded abelian complexity.*
3. *ω' has bounded additive complexity.*
4. *ω' contains an additive k-power for all $k \ge 1$.*
5. *ω' contains an abelian k-power or all $k \ge 1$,*
6. *ω contains an abelian k-power for all $k \ge 1$*

Then $1 \Leftrightarrow 2 \Leftrightarrow 3$, $4 \Leftrightarrow 5 \Leftrightarrow 6$, $3 \Rightarrow 4$, *and* $4 \not\Rightarrow 3$

Proof. Clearly $1 \Leftrightarrow 2$ and $5 \Leftrightarrow 6$.

The linear independence of S' over $\mathbb{Z}$ implies that $2 \Leftrightarrow 3$ and $4 \Leftrightarrow 5$.

The implication $3 \Rightarrow 4$ is a special case of the second part of Theorem 2.4.

To see that $4 \not\Rightarrow 3$, note that if $4 \Rightarrow 3$ then $6 \Rightarrow 1$, which is shown to be false by the Champernowne word [4]

$$C = 0110111001011101111000 1001 \cdots,$$

obtained by concatenating the binary representations of $0, 1, 2, \ldots$. This word has arbitrarily long strings of 1's (and 0's), hence satisfies condition 6; but C does not satisfy condition 1. (Clearly for the sequence C, $\rho_C^{ab}(n) = n + 1$ for all $n \geq 1$.)

$\square$

Corollary. *Every infinite word with bounded abelian complexity has an abelian k-power for every k.*

4 A more general statement

One can cast the arguments above into a more general form, and prove (we leave the details to the reader) the following statement.

Theorem 4.1. *Let S be a finite set, and let S^+ denote the free semigroup on S. For $t \in \mathbb{N}$, let*

$$\mu : S^+ \to \mathbb{Z}^t$$

be a morphism, that is, for all $B_1, B_2 \in S^+$,

$$\mu(B_1 B_2) = \mu(B_1) + \mu(B_2).$$

Let ω be an infinite word on S. Assume further that there exists $M \in \mathbb{N}$ such that

$$|B_1| = |B_2| \Rightarrow ||\mu(B_1) - \mu(B_2)|| \leq M,$$

where $|| \cdot ||$ denotes Euclidean distance in $\mathbb{Z}^t$. Then for all $k \geq 1$, ω contains a k-power modulo μ, that is, ω has a factor $B_1 B_2 \cdots B_k$ with

$$|B_1| = |B_2| = \cdots = |B_k|, \quad \mu(B_1) = \mu(B_2) = \cdots = \mu(B_k).$$

Thus taking S to be a finite subset of $\mathbb{Z}^m$, and $\mu(B) = \sum B \in \mathbb{Z}^m$, we obtain Theorem 2.4.

Taking S to be a finite set and $\mu(B) = \psi(B) \in \mathbb{Z}^{|S|}$, we obtain the Corollary to Theorem 3.2

References

[1] Yu-Hin Au, Aaron Robertson, and Jeffrey Shallit, Van der Waerden's Theorem and Avoidability in Words, *Integers* **11** #A6 (electronic), 2011.

[2] Julien Cassaigne, James D. Currie, Luke Schaeffer, and Jeffrey Shallit, Avoiding Three Consecutive Blocks of the Same Size and Same Sum, arXiv:1106.5204.

[3] Julien Cassaigne, Gwénaël Richomme, Kalle Saari, Luca Q. Zamboni, Avoiding Abelian powers in binary words with bounded Abelian complexity, arXiv:1005.2514.

[4] David G. Champernowne, The Construction of Decimals Normal in the Scale of Ten, *J. London Math. Soc.* (1933) s1-8 (4), 254-260.

[5] F. M. Dekking, Strongly non-repetitive sequences and progression-free sets, *J. Combin. Theory, Series A* **27**, 181-185, 1979.

[6] Allen R. Freedman, Sequences on Sets of Four Numbers, to appear in INTEGERS.

[7] Jaroslaw Grytczuk, Thue type problems for graphs, points, and numbers, *Discrete Math.* **308**, 4419-4429, 2008.

[8] L. Halbeisen and N. Hungerbühler, An application of van der Waerden's theorem in additive number theory, *Integers* **0** # A7 (electronic), 2000.

[9] G. Pirillo and S. Varricchio, On uniformly repetitive semigroups, *Semigroup Forum* **49**, 125-129, 1994.

[10] Gwénaël Richomme, Kalle Saari, Luca Q. Zamboni, Abelian Complexity in Minimal Subshifts, arXiv:0911.2914.

Department of Mathematics, Simon Fraser University, Burnaby, BC, Canada, V5A 1S6
hardal@sfu.ca
tbrown@sfu.ca
vjungic@sfu.ca
jds16@sfu.ca