ON ABELIAN AND ADDITIVE COMPLEXITY IN INFINITE WORDS

Hayri Ardal  
Department of Mathematics, Simon Fraser University, Burnaby, Canada  
hardal@sfu.ca

Tom Brown  
Department of Mathematics, Simon Fraser University, Burnaby, Canada  
tbrown@sfu.ca

Veselin Jungić  
Department of Mathematics, Simon Fraser University, Burnaby, Canada  
vjungic@sfu.ca

Julian Sahasrabudhe  
Department of Mathematics, Simon Fraser University, Burnaby, Canada  
jds16@sfu.ca

Received: 8/2/11, Accepted: 2/23/12, Published: 3/1/12

Abstract

The study of the structure of infinite words having bounded abelian complexity was initiated by G. Richomme, K. Saari, and L. Q. Zamboni. 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 the aforementioned authors.

1. Introduction

This note is motivated by the question of 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 [6, 7, 8, 10]. An infinite word on a finite subset $S$ of $\mathbb{Z}$, called the alphabet, is defined as a map $\omega : \mathbb{N} \rightarrow S$ and is usually written as $\omega = x_1x_2\cdots$, with $x_i \in S$, $i \in \mathbb{N}$. For $n \in \mathbb{N}$, a factor $B$ of the infinite word $\omega$ of length $n = |B|$ is the image of a set of $n$ consecutive positive integers by $\omega$, $B = \omega^i(i, i + 1, \ldots, i + n - 1) = x_ix_{i+1}\cdots x_{i+n-1}$. The sum of the factor $B$ is $\sum B = x_i + x_{i+1} + \cdots + x_{i+n-1}$.

Recently the study of infinite words with bounded abelian complexity was initiated by G. Richomme, K. Saari, and L. Q. Zamboni [11]. The abelian complexity of a
word $\omega$ is the function defined on $\mathbb{N}$ that, for $n \in \mathbb{N}$, counts the maximum number of factors of length $n$, no two of which are permutations of one another. In particular, it is shown in [11] that if $\omega$ is an infinite word with bounded abelian complexity, then $\omega$ has $k$ adjacent factors, each two of which are permutations of one other, for all $k \geq 1$.

We define the additive complexity of a word $\omega$ on a finite subset $S$ of $\mathbb{Z}$ (in fact we allow $S$ to be a finite subset of $\mathbb{Z}^m$ for any $m \geq 1$) as the function defined on $\mathbb{N}$ that, for $n \in \mathbb{N}$, counts the number of different sums of factors of $\omega$ of length $n$. We show that if $\omega$ is an infinite word with bounded additive complexity then $\omega$ has $k$ adjacent factors with equal lengths and equal sums, for all $k \geq 1$.

The question stated above remains open, even for subsets of $\mathbb{Z}$ of size 4, although some partial results can be found in [1, 2, 6]. In [6] it is shown that if $a < b < c < d$ satisfy the Sidon equation $a + d = b + c$, then every word on $\{a, b, c, d\}$ of length 61 contains two adjacent factors with equal lengths and equal sums.

2. Additive Complexity

**Definition 1.** Let $\omega$ be an infinite word on a finite subset $S$ of $\mathbb{Z}^m$ for some $m \geq 1$. For a factor $B = x_1 x_2 \cdots x_n$ of $\omega$, $\sum B$ 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 $\omega$.

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

We say that $\omega$ has bounded additive complexity if there exists $M$ such that $|\phi_\omega(n)| \leq M$ for all $n \geq 1$.

2.1. Infinite Words on Finite Subsets of $\mathbb{Z}$

**Proposition 2.** Let $\omega$ 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 $B_1 B_2$ is a factor of $\omega$ 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 $\omega$ (not necessarily adjacent) with $|B_1| = |B_2|$, then $|\sum B_1 - \sum B_2| \leq M_2$.
3. The word $\omega$ has bounded additive complexity, that is, there exists $M_3$ such that $|\phi_\omega(n)| \leq M_3$ for all $n \geq 1$.

**Proof.** We will show that (1) $\iff$ (2) and (2) $\iff$ (3).
Clearly (2) ⇒ (1). Now assume that (1) holds, that is, if $B_1B_2$ is any factor of $\omega$ with $|B_1| = |B_2|$, it is the case that $|\sum B_1 - \sum B_2| \leq M_1$. Let $B_1$ and $B_2$ be factors of $\omega$ 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 $\omega$, where $|A_1| = |A_2|$ or $|A_1| = |A_2| + 1$.

Let $C_1 = B_1A_1$ and $C_2 = A_2B_2$. Then $|C_1| = |C_2|$ 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

$$|\sum A_2 - \sum A_1| \leq M_1 + \max\{|x| : x \in S\},$$

$$|\sum C_1 - \sum C_2| \leq M_1 + \max\{|x| : x \in S\},$$

$$|\sum B_1 - \sum B_2| \leq 2M_1 + 2\max\{|x| : x \in S\},$$

so that we can take $M_2 = 2M_1 + 2\max\{|x| : x \in S\}$. Thus (1) ⇒ (2).

Next we show that (2) ⇒ (3). Thus we assume there exists $M_2$ such that whenever $B_1$ and $B_2$ are factors of $\omega$ (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 $\phi_\omega(n) \leq \sum B_1 + M_2$, so that $|\phi_\omega(n)| \leq M_2 + 1$.

Finally, we show that (3) ⇒ (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 $\omega = x_1x_2 \cdots$ 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 the factor $x_ax_{a+1} \cdots x_b$ of $\omega$, 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 $C_i$ denote the factor $\omega[i + 1, i + n]$. Thus $C_0 = B_1$, $C_q = B_2$, and the factor $C_{i+1}$ is obtained by shifting $C_i$ one position to the right. Clearly

$$\sum C_{i+1} - \sum C_i \leq \max S - \min S.$$

Since $|C_0| = |C_1| = \cdots = |C_q| = n$, and $|\phi_\omega(n)| \leq M_3$, there can be at most $M_3$ distinct numbers in the sequence $\sum B_1 = \sum C_0, \sum C_1, \ldots, \sum C_q = \sum B_2$. Let these numbers be $\sum B_1 = d_1 < d_2 < \cdots < d_r = \sum B_2$, where $r \leq M_3$. 

Since \( \sum C_{i+1} - \sum C_i \leq \max S - \min S \), \( 0 \leq i \leq q \), it follows that \( d_{j+1} - d_j \leq \max S - \min S \), \( 0 \leq i \leq r - 1 \), and hence that

\[
\sum B_2 - \sum B_1 = (d_r - d_{r-1}) + \cdots (d_2 - d_1) \leq (M_3 - 1)(\max S - \min S).
\]

**Theorem 3.** Let \( \omega \) be an infinite word on a finite subset of \( \mathbb{Z} \). Assume that \( \omega \) has bounded additive complexity. Then \( \omega \) contains an additive \( k \)-power for every positive integer \( k \).

**Proof.** Let \( \omega = x_1x_2x_3\cdots \) be an infinite word on the finite subset \( S \) of \( \mathbb{Z} \), and assume that whenever \( B_1, B_2 \) are factors of \( \omega \) (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.)

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 \( \mathbb{N} \) and by van der Waerden’s theorem [12], for any \( k \) there are \( t, s \) such that \( f(t) = f(t + s) = f(t + 2s) = \cdots = f(t + ks) \).

Using (as before) the notation \( \omega[t + 1, t + q] = x_{t+1}x_{t+2}\cdots x_{t+q} \), we set

\[
B_i = \omega[t + (i - 1)s + 1, t + is], \quad 1 \leq i \leq k,
\]

and obtain

\[
\sum B_1 \equiv \sum B_2 \equiv \cdots \equiv \sum B_k \pmod{M_2 + 1}.
\]

Since \( B_1B_2\cdots B_k \) is a factor of \( \omega \) with \( |B_i| = |B_j| \), \( 1 \leq i < j \leq k \), we have \( |\sum B_i - \sum B_j| \leq M_2 \) and \( \sum B_i \equiv \sum B_j \pmod{M_2 + 1} \). Hence \( \sum B_i = \sum B_j \).

Thus \( \omega \) contains the additive \( k \)-power \( B_1B_2\cdots B_k \). \( \square \)

### 2.2. Infinite Words on Subsets of \( \mathbb{Z}^m \)

Let us use the notation \( (u)_j \) for the \( j \)th 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 \( B_1 \) and \( B_2 \) of an infinite word \( \omega \) on a finite subset \( S \) of \( \mathbb{Z}^m \), the notation \( |\sum B_1 - \sum B_2| \leq M_1 \) means that \( |(\sum B_1 - \sum B_2)|_j \leq M_1, 1 \leq j \leq m \).

Suppose that \( \omega \) is an infinite word on a finite subset \( S \) of \( \mathbb{Z}^m \) for some \( m \geq 1 \). The definitions of \( \phi_{\omega} \) and of the additive complexity of \( \omega \) are exactly as in Definition 1 above.

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

**Proposition 4.** Proposition 2 remains true when \( \mathbb{Z} \) is replaced by \( \mathbb{Z}^m \).
Theorem 5. Let $\omega$ be an infinite word on a finite subset of $\mathbb{Z}^m$ for some $m \geq 1$. Assume that $\omega$ has bounded additive complexity. Then $\omega$ contains an additive $k$-power for every positive integer $k$.

The following is a re-statement of Theorem 5, in terms of $m$ infinite words on $\mathbb{Z}$, rather than one infinite word on $\mathbb{Z}^m$.

Theorem 6. Let $m \in \mathbb{N}$ be given, and let $S_1, S_2, \ldots, S_m$ be finite subsets of $\mathbb{Z}$. Let $\omega_j$ be an infinite word on $S_j$ with bounded additive complexity, $1 \leq j \leq m$. Then for all $k \geq 1$, there exists a $k$-term arithmetic progression in $\mathbb{N}$, $t, t+s, t+2s, \ldots, t+ks$ such that for all $j$, $1 \leq j \leq m$,

$$\sum \omega_j[t+1, t+s] = \sum \omega_j[t+s+1, t+2s] = \cdots = \sum \omega_j[t+(k-1)s+1, t+ks].$$

Thus $\omega_1, \omega_2, \ldots, \omega_m$ have “simultaneous” additive $k$-powers for all $k \geq 1$.

3. Abelian Complexity

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

Definition 7. Let $\omega$ be an infinite word on a finite alphabet. Two factors of $\omega$ 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 $\omega$, we write $\psi(B) = (u_1, u_2, \ldots, u_t)$, where $u_i$ is the number of occurrences of the letter $a_i$ in the word $B$, $1 \leq i \leq t$. We call $\psi(B)$, a notion introduced in [9], the Parikh vector associated with $B$.

Let

$$\psi_\omega(n) = \{\psi(B) : B \text{ is a factor of } \omega \text{ of length } n\}.$$  

The function $\rho^{ab}_\omega(n)$, defined by $\rho^{ab}_\omega(n) = |\psi_\omega(n)|$, $n \geq 1$, is called the abelian complexity of $\omega$.

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

The word $B_1 B_2 \cdots B_k$ 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.)

Proposition 8. Let $\omega$ 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 $B_1 B_2$ is a factor of $\omega$ with $|B_1| = |B_2|$, then $|\psi(B_1) - \psi(B_2)| \leq M_1$.

2. There exists $M_2$ such that if $B_1$ and $B_2$ are factors of $\omega$ (not necessarily adjacent) with $|B_1| = |B_2|$, then $|\psi(B_1) - \psi(B_2)| \leq M_2$.
Theorem 3. The word \( \omega \) has bounded abelian complexity, that is, there exists \( M_3 \) such that \( \rho^{ab}_\omega(n) \leq M_3 \) 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 \( \omega \) with \( |B_1| = |B_2| \), it is the case that \( |\psi(B_1) - \psi(B_2)| \leq M_1 \). Let \( B_1 \) and \( B_2 \) be factors of \( \omega \) 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 \( \omega \), where \( |A_1| = |A_2| \) or \( |A_1| = |A_2| + 1 \).

We finish this argument exactly as in the proof of (1) \( \Rightarrow \) (2) in Proposition 2, 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 \( B_1 \) and \( B_2 \) are factors of \( \omega \) (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 \( \rho^{ab}_\omega(n) \leq (2M_2 - 1)^t \).

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

Since \( |\psi(xB) - \psi(By)| \leq 1 \) for all \( x, y \in S \), it follows that if \( \omega \) has factors \( B_1 \) and \( B_2 \) of length \( n \) where for some \( j, 1 \leq j \leq t \), \((\psi(B_1))_j = p \) and \((\psi(B_2))_j = p + q \), then \( \omega \) has factors \( C_r \) of length \( n \) with \((\psi(C_r))_j = p + r, 0 \leq r \leq q \). (This is discussed in more detail in [11].) Thus \( |\psi(B_1) - \psi(B_2)| \geq M_3 \) implies \( \rho^{ab}_\omega(n) \geq M_3 + 1 \). Since we are assuming \( \rho^{ab}_\omega(n) \leq M_3, n \geq 1 \), we conclude that \( |\psi(B_1) - \psi(B_2)| \leq M_3 - 1 \) whenever \( \omega \).

Definition 9. 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 \( j, 1 \leq j \leq m \), let \( a'_j \) be the element of \( \mathbb{Z}^m \) which has \( a_j \) in the \( j \)th 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 \( \omega \) by replacing each \( a_j \) by \( a'_j \), \( 1 \leq j \leq m \). It is convenient to visualize each \( a'_j \) as a column vector, rather than as a row vector.

Theorem 10. Referring to Definition 7, consider the following statements concerning \( \omega \) and \( \omega' \): (1) \( \omega \) has bounded abelian complexity; (2) \( \omega' \) has bounded abelian complexity; (3) \( \omega' \) has bounded additive complexity; (4) \( \omega' \) contains an additive \( k \)-power for all \( k \geq 1 \);
(5) $\omega'$ contains an abelian $k$-power for all $k \geq 1$;
(6) $\omega$ contains an abelian $k$-power for all $k \geq 1$.

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) follows from Theorem 5. 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]

$$C = 01101110010110111001001 \cdots,$$

obtained by concatenating the binary representations of 0, 1, 2, . . . . 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 word $C$, $\rho_C^\ell(n) = n + 1$ for all $n \geq 1$.)

Corollary 11. Every infinite word with bounded abelian complexity has an abelian $k$-power for every $k$.

Remark 12. To see that bounded additive complexity is indeed weaker than bounded abelian complexity, consider the following example. Let $\sigma$ be Dekking’s word, the fixed point, staring with 0, of the morphism $\theta$, where $\theta(0) = 011$ and $\theta(1) = 0001$. It is known [5] that $\sigma$ has no abelian 4th power. In $\sigma$, replace every 1 by 12, and replace every 0 by 03, obtaining the sequence $\tau$. If $\tau$ had an abelian 4th power $ABCD$, then the number of 2's in each of $A, B, C, D$ would be equal, and similarly for the number of 3’s. But then dropping the 2’s and 3’s from $ABCD$ would give an abelian 4th power in $\sigma$, a contradiction. Hence, by the preceding Corollary 1, $\tau$ does not have bounded abelian complexity. Now let a factor $B$ of $\tau$ 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 $|\phi_r(n)| \leq 4$ for all $n \geq 1$, therefore $\tau$ does have bounded additive complexity.

4. A More General Statement

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

Theorem 13. Let $S$ be a finite set, and let $S^+$ denote the free semigroup on $S$. For $t \in \mathbb{N}$, let $\mu : S^+ \to \mathbb{Z}$ be a morphism, that is, for all $B_1, B_2 \in S^+$,

$$\mu(B_1B_2) = \mu(B_1) + \mu(B_2).$$
Let $\omega$ 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 $k \geq 1$, $\omega$ contains a $k$-power modulo $\mu$, that is, $\omega$ has a factor $B_1B_2 \cdots B_k$ 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 5.
Taking $S$ to be a finite set and $\mu(B) = \psi(B) \in \mathbb{Z}^{|S|}$, we obtain Corollary 11.

Acknowledgment. The authors would like to acknowledge the IRMACS Centre at Simon Fraser University for its support.

References