APPROXIMATIONS OF ADDITIVE SQUARES IN INFINITE WORDS

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

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

Abstract

We show that every infinite word ω on a finite subset of $\mathbb{Z}$ must contain arbitrarily large factors B_1B_2 which are "close" to being *additive squares*. We also show that for all $k > 1$, ω must contain a factor $U_1U_2\cdots U_k$ where $U_1, U_2, \cdots, U_k$ all have the same *average*.

1. Introduction

If S is a finite subset of $\mathbb{Z}$ and $\omega \in S^{\mathbb{N}}$, we write $\omega = x_1x_2x_3\cdots$. For any (finite) factor $B = x_ix_{i+1}\cdots x_{i+n}$ of ω, we write $|B|$ for the *length* of B (here $|B| = n+1$), and we write

$$\sum B = x_i + x_{i+1} + \cdots + x_{i+n}.$$

If B_1B_2 is a factor of ω with

$$|B_1| = |B_2| \text{ and } \sum B_1 = \sum B_2,$$

we say that B_1B_2 is an *additive square* contained in ω. For example, if $\omega = 2135126\cdots$ (a word on the alphabet $S = \{1,2,3,4,5,6\}$), then ω contains the additive square B_1B_2, where $B_1 = 135, B_2 = 126$, with $|B_1| = |B_2| = 3$ and $\sum B_1 = \sum B_2 = 9$.

A celebrated result of Keränen [11] (see also [12]) is that there exist infinite words ω on an alphabet of 4 symbols which contain no *abelian square*, that is, ω contains no factor B_1B_2 where B_1, B_2 are permutations of one another. (For early background, see [3].)

After Keränen's result, it was natural to consider the question of whether an infinite word ω on 4 (or more) integers must contain an *additive square*.

Freedman [8] showed that if $a, b, c, d \in \mathbb{Z}$ (or more generally if a, b, c, d belong to any field of characteristic 0) and $a + d = b + c$, then every word of length 61 on $\{a, b, c, d\}$ contains an additive square.

Cassaigne, Currie, Schaeffer, and Shallit [4] showed that there is an infinite word ω on the alphabet $\{0, 1, 3, 4\}$ which contains no *additive cube*, that is, ω contains no factor $B_1B_2B_3$ such that $|B_1| = |B_2| = |B_3|$ and $\sum B_1 = \sum B_2 = \sum B_3$.

A few remarks follow.

For each $k \geq 1$, let $g(1, 2, \cdots, k)$ denote the length of a longest word on $\{1, 2, \cdots, k\}$ which does *not* contain an additive square. (We allow $g(1, 2, \cdots, k) = \infty$.) Then the following three statements are equivalent:

1. For all $k \geq 1$, $g(1, 2, \cdots, k) < \infty$.

2. For all $k \geq 1$, and all infinite words ω on $\{1, 2, \cdots, k\}$, ω contains arbitrarily large additive squares.

3. Let $x_1 < x_2 < x_3 \cdots$ be any sequence of positive integers such that, for some M, $0 < x_{i+1} - x_i < M$ for all $i \geq 1$. Then there exist $i < j < k$ such that both $\{i, j, k\}$ and $\{x_i, x_j, x_k\}$ are arithmetic progressions. (Statement 3 is equivalent to statement 1 via van der Waerden's theorem on arithmetic progressions [16].)

Finally, denote by $g(a, b, c, d)$ the length of a longest word on $\{a, b, c, d\}$ which does not contain an additive square. Then the statement $\lim_{n\to\infty} g(1, n, n^2, n^3) = \infty$ is equivalent (by standard combinatorial arguments) to the result of Keränen stated above.

The question concerning the presence of additive squares seems to have appeared in print for the first time in a paper by Pirillo and Varricchio [14]. Other related material can be found in [1, 2, 4, 5, 8, 10, 13, 14, 15].

In this note we show that for every finite subset S of $\mathbb{Z}$ there is a constant C (which depends only on S) such that every infinite word ω on S contains arbitrarily long factors UV such that

$$|U| = |V| \text{ and } |\sum U - \sum V| \leq C.$$

We also show that for every infinite word ω on a finite subset of $\mathbb{Z}$ there must exist, for every $k > 1$, a factor $B_1 B_2 \cdots B_k$ of ω such that $B_1, B_2, \cdots, B_k$ all have the same *average*. Here, the *average* of a factor B is $\frac{1}{|B|} \sum B$.

2. Adjacent Equal Length Blocks with Nearly Equal Sums

Here we exploit the fact that if U, V are words on a 2-element subset of $\mathbb{Z}$, then UV is an *additive* square ($|U| = |V|$ and $\sum U = \sum V$) if and only if UV is an *abelian* square (U and V are permutations of one another).

Theorem 1. *For every finite subset S of $\mathbb{Z}$ there exists a constant C (depending only on S) such that every infinite word ω on S contains arbitrarily long factors UV such that*

$$|U| = |V| \text{ and } |\sum U - \sum V| \leq C.$$

Proof. First assume that S is a finite subset of $\mathbb{N}$, and let $\omega = x_1 x_2 x_3 \cdots$ be an infinite word on S. Let 1^{x_i} denote a string of 1s of length x_i (e.g., $1^4 = 1111$), and let ω' be the binary word $1^{x_1}01^{x_2}01^{x_3}0\cdots$, which we write for convenience as $x_10x_20x_30\cdots$. By a theorem of Entringer, Jackson, and Schatz [7] the word ω' contains arbitrarily large abelian squares $U'V'$, and hence arbitrarily large factors $U'V'$ with $|U'| = |V'|$ and $\sum U' = \sum V'$. Re-numbering the indices for convenience, such a square $U'V'$, since each of U' and V' must contain the same number, say k, of 0s, has the form

$$U' = \alpha_2 0 x_2 0 x_3 0 \cdots 0 x_k 0 \alpha_3, V' = \alpha_4 x_{k+2} 0 \cdots 0 x_{2k} 0 \alpha_5,$$

where $\alpha_1 + \alpha_2 = x_1, \alpha_3 + \alpha_4 = x_{k+1}, \alpha_5 + \alpha_6 = x_{2k+1}$. (All the α_i are non-negative integers.) Since $U'V'$ is an additive square,

$$\alpha_2 + \sum_{i=2}^{k} x_i + \alpha_3 = \alpha_4 + \sum_{i=k+2}^{2k} x_i + \alpha_5,$$

or (using $\alpha_1 + \alpha_2 = x_1$ and $\alpha_3 + \alpha_4 = x_{k+1}$)

$$|\sum_{i=1}^{k} x_i - \sum_{i=k+1}^{2k+1} x_i| = |\alpha_1 - 2\alpha_3 + \alpha_5| \leq 2 \max S.$$

With $U = x_1 x_2 \cdots x_k, V = x_{k+1} x_{k+2} \cdots x_{2k}$, we have the factor UV of ω with

$$|U| = |V| \;\text{ and }\; |\sum U - \sum V| \leq 2 \max S.$$

When S is a finite subset of $\mathbb{Z}$ which contains non-positive integers, translate S to the right by $|\min S| + 1$ and apply the above argument, to get arbitrarily large factors UV such that $|U| = |V|$ and $|\sum U - \sum V| \leq 2(|\min S| + 1 + \max S)$. $\square$

3. Adjacent Factors with Equal Averages

Theorem 2. *For any finite subset S of $\mathbb{Z}$, any infinite word ω on S, and any $k > 1$, there exists a factor $U_1U_2 \cdots U_k$ with*

$$\frac{1}{|U_1|}\sum U_1 = \frac{1}{|U_2|}\sum U_2 = \cdots = \frac{1}{|U_k|}\sum U_k.$$

Proof. Let $\omega = x_1x_2x_3\cdots$ be a given infinite word on the set of integers $S = \{s_1, s_2, \cdots, s_t\}$. Consider the infinite sequence of points in the plane $P_i = (i, x_1 + x_2 + \cdots + x_i), i \geq 1$. Since $P_{i+1} - P_i = (1, x_{i+1}) \in \{(1, s_j) : 1 \leq j \leq t\}$, a theorem of Gerver and Ramsey [9] asserts that the set $\{P_i : i \geq 1\}$ contains, for any given $k > 1$, $k + 1$ collinear points $P_{i_1}P_{i_2}\cdots P_{i_{k+1}}$. For $1 \leq j \leq k$, let $U_j = x_{i_j+1}x_{i_j+2}\cdots x_{i_{j+1}}$. The slope of the line segment joining P_{i_j} and $P_{i_{j+1}}$ is $\frac{1}{|U_j|}\sum U_j$. Since this slope is the same for each choice of j, we have

$$\frac{1}{|U_1|}\sum U_1 = \frac{1}{|U_2|}\sum U_2 = \cdots = \frac{1}{|U_k|}\sum U_k. \qquad \square$$

Acknowledgments. The author would like to thank Allen Freedman for mentioning, many years ago, the idea of "equal averages," Veso Jungić, Hayri Ardal, and Julian Sahasrabudhe for helpful conversations, and the IRMACS Centre at Simon Fraser University for its support.

Remark. The author has learned that Theorem 2 was proved independently by Jeffrey Shallit.

References

[1] Hayri Ardal, Tom Brown, Veselin Jungíc, Julian Sahasrabudhe, On abelian and additive complexity in infinite words, *ArXiv:1107.4654v1* [math.CO] 23 Jul 2011.

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

[3] T. C. Brown, Is there a sequence on four symbols in which no two adjacent segments are permutations of one other? *American Math. Monthly* **78**, 886–888, 1971.

[4] Julien Cassaigne, James D. Currie, Luke Schaeffer, and Jeffrey Shallit, Avoiding three consecutive blocks of the same size and same sum, arXiv:1106.5204.

[5] Julien Cassaigne, Gwénaël Richomme, Kalle Saari, Luca Q. Zamboni, Avoiding abelian powers in binary words with bounded abelian complexity, *Int. J. Found. Comput. Sci.* **22**(4), 905–920, 2011.

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

[7] R. C. Entringer, D. E. Jackson, and J. A. Schatz, On non-repetitive sequences, *J. Combinatorial Theory Ser. A* **16**, 159–164, 1974.

[8] Allen R. Freedman, Sequences on sets of four numbers, to appear.

[9] Joseph L. Gerver and L. Thomas Ramsey, On certain sequences of lattice points, *Pacific J. of Math.* **83**, 357–363, 1979.

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

[11] V. Keränen, Abelian squares are avoidable on 4 letters, in W. Kuich, editor, *Proc. 19th Int'l Conf. on Automata, Languages, and Programming (ICALP)*, Vol. 623 of *Lecture Notes in Computer Science*, pp. 41–52. Springer-Verlag, 1992.

[12] V. Keränen, A powerful abelian square-free substitution over 4 letters, *Theoretical Computer Science* **410**, 3893–3900, 2009.

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

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

[15] Gwénaël Richomme, Kalle Saari, Luca Q. Zamboni, Abelian complexity in minimal subshifts, arXiv:0911.2914.

[16] B. L. van der Waerden, Beweis einer Baudetschen Vermutung, *Nieuw Archief voor Wiskunde* **15**, 212–216, 1927.