On Avoiding Sufficiently Long Abelian Squares

Elyot Grant

December 3, 2010

Abstract

A finite word w is an *abelian square* if $w = xx'$ with x' a permutation of x. In 1972, Entringer, Jackson, and Schatz proved that every binary word of length k^2+6k contains an abelian square of length $\geq 2k$. We use Cartesian lattice paths to characterize abelian squares in binary sequences, and construct a binary word of length $q(q+1)$ avoiding abelian squares of length $\geq 2\sqrt{2q(q+1)}$ or greater. We thus prove that the length of the longest binary word avoiding abelian squares of length $2k$ is $\Theta(k^2)$.

1 Introduction

Let Σ be a finite alphabet. A word $w \in \Sigma^*$ is an *abelian square of order* k if $w = xx'$ with $|x| = |x'| = k$ and x' a permutation of x. In 1972, Entringer, Jackson, and Schatz proved that all infinite binary sequences contain arbitrarily large abelian squares [1]. In particular, they showed that all binary words $w \in \{0,1\}^*$ of length $k^2 + 6k$ contain an abelian square of order k or greater. In this paper, we examine $\ell(k)$, the length of the longest binary word avoiding abelian squares xx' with $|x| \geq k$.

Precise values of $\ell(k)$ have been computed for $1 \leq k \leq 10$ by Jeffrey Shallit and Narad Rampersad via a brute force search. The results are given in Section 2.

The bound $\ell(k) < k^2 + 6k$ given by Entringer, Jackson, and Schatz is not the best possible upper bound, but an improved upper bound remains unknown. A simple lower bound $\ell(k) \geq 8k - 6$ can be obtained by observing that the string $0^{2k-2}1^{2k-1}0^{2k-1}1^{2k-2}$ contains no abelian squares of order k or greater. This lower bound is tight for $2 \leq k \leq 7$, but is suboptimal for $k \geq 8$.

In this paper, we give a quadratic lower bound for $\ell(k)$, proving that $\ell(k)$ is $\Theta(k^2)$. Moreover, we provide an intuitive geometric characterization of abelian squares in a binary word by treating each character of a string as a step of a lattice path in the Cartesian plane. We use this geometric notion to construct, for all q, a word of length $q(q+1)$ containing no abelian squares of order $\geq \sqrt{2q(q+1)}$.

Many thanks go to Jeffrey Shallit for suggesting this as a problem to study as part of *CS 860: Patterns in Strings: Existence, Avoidability, Enumeration*, a course he developed and taught at the University of Waterloo.

2 Values of $\ell(k)$ for $1 \leq k \leq 10$

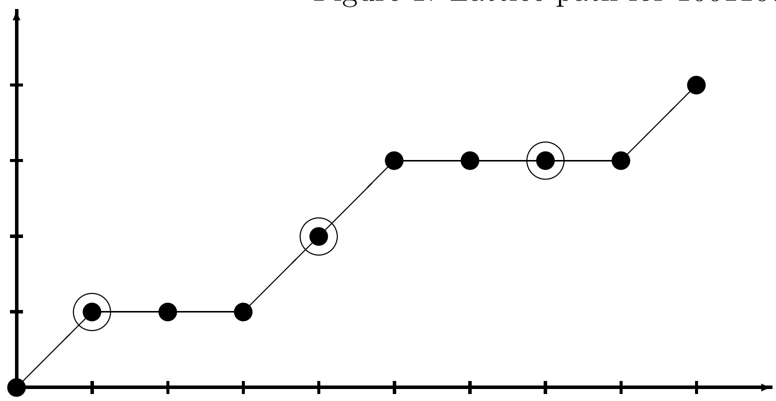
Jeffrey Shallit and Narad Rampersad have provided the values of $\ell(k)$ for $1 \leq k \leq 10$. We give them here, alongside the lexicographically least word of length $\ell(k)$ containing no abelian squares of order k or greater:

k	$\ell(k)$	
1	3	010
2	10	0011100011
3	18	000011111000001111
4	26	00000011111110000000111111
5	34	0000000011111111100000000011111111
6	42	000000000011111111111000000000001111111111
7	50	00000000000001000001100001111001111101111111111111
8	62	00000000000000010000100100011001100111011011110111111111111111
9	76	0000000000000000010000000110010000111010001110110011111110111 11111111111111
10	90	000000000000000000100000010010000011010100001111010100111101 1011111101111111111111111111
11	≥ 106	
12	≥ 124	
13	≥ 139	

3 Main Result

Given a word $w[1..t] \in \{0,1\}^*$, let $S_i = \sum_{j=1}^{i} w[j]$ be the nondecreasing sequence of *prefix sums* of w. By plotting the ordered pairs (i, S_i) for $0 \leq i \leq t$, we obtain a representation of w as a path across the Cartesian lattice, stepping east when w contains a zero, and northeast when w contains a 1. An example for the string 100110001 is shown in Figure 1.

Figure 1: Lattice path for 100110001



We note that the number of ones in $w[m..n]$ is $S_n - S_{m-1}$. Consequently, $w[i+1..i+2r]$ is an abelian square iff $S_{i+r} - S_i = S_{i+2r} - S_{i+r}$, which occurs precisely when (i, S_i), $(i+r, S_{i+r})$, and $(i+2r, S_{i+2r})$ are three equally spaced collinear points in our lattice path. In Figure 1, the three circled points indicate the presence of the subword 001100, an abelian square.

Next, we give our construction of a word of length $q(q+1)$ containing no abelian squares of order $\geq \sqrt{2q(q+1)}$. We design our word w so that its lattice path approximates a quadratic function; this ensures that three equally spaced points along the path can be collinear only if they are sufficiently close together. For $0 \leq i \leq q(q+1)$, define

$$a_i = \left\lfloor \frac{i^2}{2q(q+1)} \right\rfloor .$$

We note that if $i \leq q(q+1)$, then $i^2 - (i-1)^2 = 2i - 1 < 2q(q+1)$, and hence $a_i - a_{i-1} \in \{0, 1\}$ for all $1 \leq i \leq q(q+1)$. We can thus define a binary word $w = w[1..q(q+1)]$ by $w[i] = a_i - a_{i-1}$. We will show the following:

Theorem 1. *w contains no abelian squares xx' with $|x| \geq \sqrt{2q(q+1)}$.*

Our theorem implies that if q is an integer with $2q(q+1) \leq k^2$, then there exists a binary word of length $q(q+1)$ containing no abelian squares of order k. For a given k, the shortest such q is $\left\lfloor \frac{\sqrt{1+2k^2}-1}{2} \right\rfloor$. Consequently, we may conclude the following:

Corollary 2. $\ell(k) \geq \left(\left\lfloor \frac{\sqrt{1+2k^2}-1}{2} \right\rfloor \right) \left(\left\lfloor \frac{\sqrt{1+2k^2}-1}{2} \right\rfloor + 1 \right) > \frac{k^2}{2} - \sqrt{2k}.$

4 Proof of Theorem 1

Suppose w contains an abelian square xx' with $|x| = r$. Then there exist two adjacent blocks $w[i+1..i+r]$ and $w[i+r+1..i+2r]$ such that $|w[i+1..i+r]|_1 = |w[i+r+1..i+2r]|_1$. This implies that $a_{i+r} - a_i = a_{i+2r} - a_{i+r}$. We eliminate the floor function to bound the various a_i values above and below in the following manner:

$$\begin{aligned} \frac{i^2}{2q(q+1)} - 1 &< a_i \\ a_{i+r} &\leq \frac{(i+r)^2}{2q(q+1)} \\ \frac{(i+2r)^2}{2q(q+1)} - 1 &< a_{i+2r} \end{aligned}$$

Taking a linear combination of the above inequalities, we obtain

$$\frac{i^2}{2q(q+1)} - 1 + 2a_{i+r} + \frac{(i+2r)^2}{2q(q+1)} - 1 < a_i + 2\frac{(i+r)^2}{2q(q+1)} + a_{i+2r}$$

and we may cancel the a_i terms since $a_{i+r} - a_i = a_{i+2r} - a_{i+r}$. We simplify what remains to obtain our result:

$$\begin{aligned} \frac{i^2}{2q(q+1)} + \frac{(i+2r)^2}{2q(q+1)} - 2 &< 2\frac{(i+r)^2}{2q(q+1)} \\ i^2 + (i+2r)^2 - 4q(q+1) &< 2(i+r)^2 \\ r^2 &< 2q(q+1) \end{aligned}$$

5 Additional Remarks

One might suggest that we could improve our lower bound slightly by computing more a_i values and extending w to a longer string. Indeed, we can take $a_i = \left\lfloor \frac{i^2}{2q(q+1)} \right\rfloor$ for all i until we reach an n such that $a_{n+1} - a_n > 1$. Unfortunately, it turns out that this doesn't help us much. Taking $p = q(q+1) + \left\lceil \sqrt{2q(q+1)} \right\rceil$, we see that

$$\begin{aligned} a_p - a_{q(q+1)} &= \left\lfloor \frac{(q(q+1) + \left\lceil \sqrt{2q(q+1)} \right\rceil)^2}{2q(q+1)} \right\rfloor - \left\lfloor \frac{(q(q+1))^2}{2q(q+1)} \right\rfloor \\ &= \left\lceil \sqrt{2q(q+1)} \right\rceil + \frac{(\left\lceil \sqrt{2q(q+1)} \right\rceil)^2}{2q(q+1)} \\ &\geq p - q(q+1) + 1. \end{aligned}$$

Consequently, there must be some n with $q(q+1) \leq n < p$ such that $a_{n+1} - a_n > 1$. Thus we can extend w for at most another $\sqrt{2q(q+1)}$ symbols.

References

[1] R. C. Entringer, D. E. Jackson, J. A. Schatz. *On Nonrepetitive Sequences.* Journal of Combinatorial Theory (A) **16**, 159–164 (1974).