New results on pseudosquare avoidance

Tim Ng, Pascal Ochem, Narad Rampersad, Jeffrey Shallit

School of Computer Science
University of Waterloo
Canada

Email : shallit@uwaterloo.ca
Web : `https://cs.uwaterloo.ca/~shallit/`

Repetitions

Repetitions

A *square* is a nonempty word of the form xx, such as the English word `murmur`.

The *order* of a square xx is $|x|$, the length of x.

Antisquares

For a binary word x, we let $\overline{x}$ be the bitwise complement of x.

For example, if $x = 0110$, then $\overline{x} = 1001$.

An *antisquare* is a word of the form $x\,\overline{x}$. So 0110 is an antisquare.

Avoiding squares in binary words

- Every binary word of length 4 contains a square.
- Entringer, Jackson, and Schatz (1974) constructed an infinite binary word *avoiding* squares of order > 2.
  - We use the term *avoid* to mean "the word has no factor (= contiguous subsequence) of the given form".
- Fraenkel and Simpson (1995) constructed an infinite binary word avoiding all squares, except three: 00, 11, and 0101. There is no infinite binary word containing only two distinct squares, so their result is optimal.

Avoiding antisquares

- Remember: an antisquare is a word of the form $x\,\overline{x}$.
- The only infinite binary words avoiding all antisquares are the trivial ones: $0^\omega = 000\cdots$ and $1^\omega = 111\cdots$.
- The only infinite binary words containing exactly one antisquare are also trivial: $0\,1^\omega$ and $1\,0^\omega$.
- But every word in $\{1000, 10000\}^\omega$ has exactly two antisquares — namely 01 and 10 — and infinitely many of these are aperiodic.

Our first main result

- We consider the *simultaneous* avoidance of squares and antisquares in binary words
- We determine, for each $a, b \geq 0$, the longest binary word containing at most a squares and b antisquares (it could be infinite)

Summary of results on squares and antisquares

The following table gives the length of the longest binary word containing at most a squares and b antisquares:

a \ b	0	1	2	3	4	5	6	7	8	9	10	11	12	13	$\cdots$
0	1	2	3	3	3	3	3	3	3	3	3	3	3	3	$\cdots$
1	3	4	7	7	7	7	7	7	7	7	7	7	7	7	$\cdots$
2	5	6	11	11	11	11	12	12	12	13	15	18	18	18	$\cdots$
3	7	8	15	15	15	20	20	20	24	29	34	53	98	∞	$\cdots$
4	9	10	19	19	27	31	45	56	233	∞	∞	∞	∞	$\cdots$	
5	11	12	27	27	40	∞	∞	∞	∞	$\cdots$					
6	13	14	35	38	313	∞	$\cdots$								
7	15	16	45	∞	∞	$\cdots$									
8	17	18	147	∞	$\cdots$										
9	19	20	∞	$\cdots$											
10	21	22	∞	$\cdots$											
$\vdots$															

How these results were obtained

- For the finite results, we use breadth-first search of the tree of possibilities. Sometimes this took a while (e.g., 6 squares and 4 antisquare — longest string of length 313).
- For the infinite results corresponding to

$$(a, b) \in \{(3, 13), (4, 9), (5, 5), (7, 3), (9, 2)\}$$

  we explicitly constructed an infinite binary word with the desired properties.
- In each case this involves finding a suitable morphism $h_{a,b} : \{0, 1, 2\}^* \to \{0, 1\}^*$ and then applying that morphism to an infinite squarefree word over $\{0, 1, 2\}$.

Example of a morphism

For example, here is $h_{7,3}$:

$$0 \to 010010010000101 0000$$

$$1 \to 0100100100000 1$$

$$2 \to 010010010100 0$$

When we apply this morphism to an infinite squarefree word over $\{0, 1, 2\}$, we get only the seven squares

$$0^2, (00)^2, (01)^2, (10)^2, (001)^2, (010)^2, (100)^2$$

and only the three antisquares

$$01, 10, 1001.$$

In terms of the total number of squares plus antisquares, this achieves the minimum: ten.

Looking at these results from a broader perspective

- We tried to avoid words of the form $xh(x)$, where h is a morphism. These are called *pseudosquares*.
- More specifically, we considered the case where h is a coding (letter-to-letter map) that is a *permutation* of the underlying alphabet $\Sigma = \{0, 1\}$.
- It is impossible to avoid $xh(x)$ for all x, so we only succeeded for *sufficiently large* words x.
- We could consider larger alphabets and hope to obtain analogous results.

Avoiding pseudosquares for permutations: negative result

Theorem

For all finite alphabets Σ, and for all words w of length ≥ 10 over Σ, there exists a permutation p of Σ and a factor of w of the form $xp(x)$ where $|x| \geq 2$.

Proof idea. If there is a word avoiding $xp(x)$ for $|x| \geq 2$, then there is an *ordered* one: where the first letter is 0, the first non-0 letter is 1, the first non-$\{0, 1\}$ letter is 2, etc.

So we can use breadth-first search, but examine only ordered words, expanding the size of the alphabet each time a candidate word is extended in length.

Avoiding pseudosquares for permutations: negative result

Breadth-first search with an expanding alphabet size easily proves:

The longest ordered words avoiding $xp(x)$ are of length 9:

$$\{001000122, 001000211, 001000233, 001222022,$$

$$001222122, 001222322\}.$$

Avoiding pseudosquares for permutations: positive result

Theorem

There exists an infinite word $\mathbf{w}$ *over the binary alphabet* $\Sigma_2 = \{0, 1\}$ *that avoids* $xp(x)$ *for all permutations* p *and all* x *with* $|x| \geq 3$.

Proof.

By direct construction, using our previous results on avoiding squares and antisquares.

Further generalization: to transformations

- Now we understand avoiding pseudosquares for permutations.
- How about for *transformations* of the underlying alphabet?
- Now the space of possible pseudosquares increases in size, so it becomes harder to avoid them, and requires going to a larger set of exceptions.

Avoiding pseudosquares for transformations: negative result

Theorem

For all finite alphabets Σ, and all words w of length ≥ 31 over Σ, there exists a transformation $t : \Sigma^* \to \Sigma^*$ such that w contains a factor of the form $xt(x)$ for $|x| \geq 3$.

Proof idea. Again, use the tree traversal method while extending the alphabet size. There are 24745 ordered words of length 30 avoiding $xt(x)$ for $|x| \geq 3$; the lexicog. first is

$$000001100101001100202001101200,$$

while the lexicographically last is

$$011112233232332244343445565789.$$

Avoiding pseudosquares for permutations: positive result

Theorem

There exists an infinite word $\mathbf{w}$ *over the binary alphabet* $\Sigma_2 = \{0, 1\}$ *avoiding* $xt(x)$ *for all transformations* t *and every* x *with* $|x| \geq 4$.

Proof. By direct construction. Use the fixed point of the morphism

$$0 \to 01 \qquad 1 \to 23$$
$$2 \to 45 \qquad 3 \to 21$$
$$4 \to 23 \qquad 5 \to 42$$

followed by the coding $n \to \lfloor n/3 \rfloor$. The result can now easily be verified with `Walnut`.

Further generalization: to arbitrary morphisms

- Now we understand pseudosquares with permutations and transformations.
- Now let's generalize even further: to *arbitrary (nonerasing) morphisms*.
- We are trying to avoid $xh(x)$ for *all* nonerasing morphisms h, simultaneously!
- This leads to our second main result.

Avoiding pseudosquares for arbitrary morphisms: negative result

Theorem

No infinite word over a finite alphabet avoids all factors of the form $xh(x)$, for all nonerasing morphisms h, with $|x| \geq 4$.

Proof.

- Assume such a word $\mathbf{z}$ exists. By a result of de Luca and Varricchio, there is a uniformly recurrent word $\mathbf{y}$ whose factors are a subset of those of $\mathbf{z}$. Such a word also avoids the factors $xh(x)$.
- Suppose $\mathbf{y}$ contains a word of the form au, where a is any single letter and u does not contain a and $|u| \geq 3$.

- Since **y** is uniformly recurrent, it must also contain a factor of the form $auvu$ for some v.
- But this is an occurrence of $xh(x)$ (take $h(a) = v$ and h the identity on other letters).
- Suppose **y** contains a word of the form aaa. Clearly **y** cannot equal a^ω, so **y** contains $baaa$ for some $b \neq a$. This contradicts the previous case.
- Suppose **y** contains three consecutive distinct letters, say abc. Then the next letter has to be a from above. Repeating this reasoning on the last three letters, we see the next letter has to be b, then c, etc. So **y** contains $abcabcab$, which is an occurrence of $xh(x)$ for $h(a) = b$, $h(b) = c$, and $h(c) = a$.
- Now **y** avoids aaa, abc, and $abbc$ for distinct letters a, b, c. So it must be a binary word.

- Now suppose **y** contains both 0100 and 1011. Then 0100 is followed by 1. By recurrence **y** must contain $01001u11$ for some u, which is of the form $xh(x)$ for $h(0) = 1$ and $h(1) = u$. Without loss of generality say **y** does not contain 0100.
- Now it is easy to check that the longest binary word avoiding 000, 111, 0100, and all squares xx with $|x| \geq 4$ is of length 67. We are done.

Avoiding pseudosquares for arbitrary morphisms: positive result

Theorem

There exists an infinite binary word that avoids all factors of the form $xh(x)$ for all nonerasing binary morphisms h, with $|x| \geq 5$.

Proof. By direct construction, by applying a certain 57-uniform morphism to a squarefree word over $\{0, 1, 2\}$.

The morphism

$0 \rightarrow 101000110010100110001011001010110001010100011001011000110$

$1 \rightarrow 101000110010100110001010110001101001100010101000110101001$

$2 \rightarrow 101000110010100110001010100011010011000101011000110101001$

Related work

- Chiniforooshan, Kari, and Zhu (2013) studied avoiding words of the form $x\theta(x)$, where θ is an *antimorphic involution*.
  - This means that $\theta^2(x) = x$ and $\theta(xy) = \theta(y)\theta(x)$.
  - For binary words the only ones are $\theta(x) = x$, $\theta(x) = \overline{x}$, $\theta(x) = x^R$, and $\theta(x) = \overline{x^R}$.
- Bischoff, Currie, and Nowotka (2012) studied the avoidability of more general patterns with involution
- Rumyantsev and Ushakov (2006), Durand, Levin, and Shen (2008), and Miller (2012) studied avoiding factors of low Kolmogorov complexity. This is more general than our results, but they can't obtain explicit bounds.