AVOIDING 2-BINOMIAL SQUARES AND CUBES

MICHAEL RAO, MICHEL RIGO, AND PAVEL SALIMOV†

Abstract. Two finite words $u, v$ are 2-binomially equivalent if, for all words $x$ of length at most 2, the number of occurrences of $x$ as a (scattered) subword of $u$ is equal to the number of occurrences of $x$ in $v$. This notion is a refinement of the usual abelian equivalence. A 2-binomial square is a word $uv$ where $u$ and $v$ are 2-binomially equivalent.

In this paper, considering pure morphic words, we prove that 2-binomial squares (resp. cubes) are avoidable over a 3-letter (resp. 2-letter) alphabet. The sizes of the alphabets are optimal.

1. Introduction

A square (resp. cube) is a non-empty word of the form $xx$ (resp. $xxx$). Since the work of Thue, it is well-known that there exists an infinite squarefree word over a ternary alphabet, and an infinite cubefree word over a binary alphabet [11, 12]. A main direction of research in combinatorics on words is about the avoidance of a pattern, and the size of the alphabet is a parameter of the problem.

A possible and widely studied generalization of squarefreeness is to consider an abelian framework. A non-empty word is an abelian square (resp. abelian cube) if it is of the form $xy$ (resp. $xyz$) where $y$ is a permutation of $x$ (resp. $y$ and $z$ are permutations of $x$). Erdős raised the question whether abelian squares can be avoided by an infinite word over an alphabet of size 4 [2]. Keränen answered positively to this question, with a pure morphic word [7]. Moreover Dekking has previously obtained an infinite word over a 3-letter alphabet that avoids abelian cubes, and an infinite binary word that avoids abelian 4-powers [1]. (Note that in all these results, the size of the alphabet is optimal.)

In this paper, we are dealing with another generalization of squarefreeness and cubefreeness. We consider the 2-binomial equivalence which is a refinement of the abelian equivalence, i.e., if two words $x$ and $y$ are 2-binomially equivalent, then $x$ is a permutation of $y$ (but in general, the converse does not hold, see Example 1 below). This equivalence relation is defined thanks to the binomial coefficient \( \binom{u}{v} \) of two words $u$ and $v$ which is the number of times $v$ occurs as a subsequence of $u$ (meaning as a “scattered” subword). For more on these binomial coefficients, see for instance [8, Chap. 6]. Based on this classical notion, the $m$-binomial equivalence of two words has been recently introduced [10].

Definition 1. Let $m \in \mathbb{N} \cup \{+\infty\}$ and $u, v$ be two words over the alphabet $A$. We let $A^{\leq m}$ denote the set of words of length at most $m$ over $A$. We say that $u$ and $v$ are $m$-binomially equivalent if

$\binom{u}{x} = \binom{v}{x}, \forall x \in A^{\leq m}.$

We simply write $u \sim_m v$ if $u$ and $v$ are $m$-binomially equivalent. The word $u$ is obtained as a permutation of the letters in $v$ if and only if $u \sim_1 v$. In that case, we say that $u$ and $v$ are abelian equivalent and we write instead $u \sim_{ab} v$. Note that if $u \sim_{k+1} v$, then $u \sim_k v$, for all $k \geq 1$.

Example 1. The four words 0101110, 0110101, 1001101 and 1010011 are 2-binomially equivalent. Let $u$ be any of these four words. We have

$\binom{u}{0} = 3, \binom{u}{1} = 4, \binom{u}{00} = 3, \binom{u}{01} = 7, \binom{u}{10} = 5, \binom{u}{11} = 6.$

†The third author is supported by the Russian President’s grant no. MK-4075.2012.1 and Russian Foundation for Basic Research grants no. 12-01-00089 and no. 11-01-00997 and by a University of Liège post-doctoral grant.
For instance, the word 0001111 is abelian equivalent to 0101110 but these two words are not 2-binomially equivalent. Let \( a \) be a letter. It is clear that \( \binom{n}{u_0} a \) and \( \binom{n}{a} \) carry the same information, i.e., \( \binom{n}{u_0} = \binom{n}{a} \) where \( |u|_a \) is the number of occurrences of \( a \) in \( u \).

A 2-binomial square (resp. 2-binomial cube) is a non-empty word of the form \( xy \) where \( x \sim y \) (resp. \( x \sim y \sim z \)). Squares are avoidable over a 3-letter alphabet and abelian squares are avoidable over a 4-letter alphabet. Since 2-binomial equivalence lies between abelian equivalence and equality, the question is to determine whether or not 2-binomial squares are avoidable over a 3-letter alphabet. We answer positively to this question in Section 2. The fixed point of the morphism \( g : 0 \rightarrow 012, 1 \rightarrow 02, 2 \rightarrow 1 \) avoids 2-binomial squares.

In a similar way, cubes are avoidable over a 2-letter alphabet and abelian squares are avoidable over a 3-letter alphabet. The question is to determine whether or not 2-binomial cubes are avoidable over a 2-letter alphabet. We also answer positively to this question in Section 3. The fixed point of the morphism \( h : 0 \rightarrow 001, 1 \rightarrow 011 \) avoids 2-binomial cubes.

**Remark 1.** The \( m \)-binomial equivalence is not the only way to refine the abelian equivalence. Recently, a notion of \( m \)-abelian equivalence has been introduced \[6\]. To define this equivalence, one counts the number \( |u|_x \) of occurrences in \( u \) of all factors \( x \) of length up to \( m \) (it is meant factors made of consecutive letters). That is \( u \) and \( v \) are \( m \)-abelian equivalent if \( |u|_x = |v|_x \) for all \( x \in A^{\leq m} \). In that context, the results on avoidance are quite different. Over a 3-letter alphabet \( 2 \)-abelian squares are unavoidable: the longest ternary word which is \( 2 \)-abelian squarefree has length 537 \[4\], and pure morphic words cannot avoid \( k \)-abelian-squares for every \( k \) \[4\]. On the other hand, it has been shown that there exists a 3-abelian squarefree morphic word over a 3-letter alphabet \[5\]. Moreover \( 2 \)-abelian-cubes can be avoided over a binary alphabet by a morphic word \[9\].

The number of occurrences of a letter \( a \) in a word \( u \) will be denoted either by \( \binom{u}{a} \) or \( |u|_a \). Let \( A = \{0, 1, \ldots, k\} \) be an alphabet. The Parikh map is an application \( \Psi : A^* \rightarrow \mathbb{N}^k+1 \) such that \( \Psi(u) = (|u|_0, \ldots, |u|_k)^T \). Note that we will deal with column vectors (when multiplying a square matrix with a column vector on its right). In particular, two words are abelian equivalent if and only if they have the same Parikh vector. The mirror of the word \( u = u_1u_2\cdots u_k \) is denoted by \( \tilde{u} = u_k \cdots u_2u_1 \).

### 2. Avoiding 2-binomial squares over a 3-letter alphabet

Let \( A = \{0, 1, 2\} \) be a 3-letter alphabet. Let \( g : A^* \rightarrow A^* \) be the morphism defined by

\[
g : \begin{cases}
0 & \mapsto 012 \\
1 & \mapsto 02 \\
2 & \mapsto 1
\end{cases}
\text{ and thus, } g^2 : \begin{cases}
0 & \mapsto 012021 \\
1 & \mapsto 0121 \\
2 & \mapsto 02.
\end{cases}
\]

It is prolongable on 0: \( g(0) \) has 0 as a prefix. Hence the limit \( x = \lim_{n \to +\infty} g^n(0) \) is a well-defined infinite word

\[x = g^\omega(0) = 0120210121020120121\cdots\]

which is a fixed point of \( g \). Since the original work of Thue, this word \( x \) is well-known to avoid (usual) squares. It is sometimes referred to as the ternary Thue–Morse word. We will make use of the fact that \( X = \{012, 021\} \) is a prefix-code and thus an \( \omega \)-code: Any finite word in \( X^* \) (resp. infinite word in \( X^\omega \)) has a unique factorization as a product of elements of \( X \). Let us make an obvious but useful observation.

**Observation 1.** The factorization of \( x \) in terms of the elements in \( X \) permits to write \( x \) as

\[x = 0\alpha_1 2\alpha_2 0\alpha_3 2\alpha_4 0\alpha_5 2\alpha_6 0\cdots\]

where, for all \( i \geq 1 \), \( \alpha_i \in \{\varepsilon, 1\} \). That is, the image of \( x \) by the morphism \( e : 0 \mapsto 0, 1 \mapsto \varepsilon, 2 \mapsto 2 \) (which erases all the 1’s) is \( e(x) = (02)^\omega \).

The next property is well known. For example, it comes from the fact that the image of the ternary Thue–Morse word by the morphism \( 0 \mapsto 011, 1 \mapsto 01, 2 \mapsto 0 \) is the Thue–Morse word. However, for the sake of completeness, we give a direct proof here.
Lemma 1. A word u is a factor occurring in x if and only if \( \tilde{u} \) is a factor occurring in x.

Proof. We define the morphism \( \tilde{g} : A^* \to A^* \) by considering the mirror images of the images of the letters by g,

\[
\tilde{g} : \begin{cases}
0 \mapsto \text{210} \\
1 \mapsto \text{20} \\
2 \mapsto \text{1}
\end{cases}
\]

and thus, \( \tilde{g}^2 : \begin{cases}
0 \mapsto \text{120210} \\
1 \mapsto \text{1210} \\
2 \mapsto \text{20}.
\end{cases} \]

Note that \( \tilde{g} \) is not prolongable on any letter. But the morphism \( \tilde{g}^2 \) is prolongable on the letter 1. We consider the infinite word

\[
\text{y} = (\tilde{g}^2)^\omega(1) = 1210201210120210201202101210 \cdots.
\]

If \( v \in A^* \) is a non-empty word ending with \( a \in A \), i.e., \( v = u a \) for some word \( u \in A^* \), we denote by \( va^{-1} \) the word obtained by removing the suffix \( a \) from \( v \). So \( va^{-1} = u \).

For every words \( r \) and \( s \) we have \( r = g^2(s) \Leftrightarrow \tilde{r} = \tilde{g}^2(\tilde{s}) \). Obviously, \( u \) is a factor occurring in \( x \) if and only if \( \tilde{u} \) is a factor occurring in \( y \).

On the other hand, \( \tilde{g}^2 \) is a cyclic shift of \( g^2 \), since \( g^2(u) = 0\tilde{g}^2(\tilde{u}) \) for every \( a \in \{0, 1, 2\} \). Thus \( u \) is a factor occurring in \( x \) if and only if \( \tilde{u} \) is a factor occurring in \( y \). To summarize, \( u \) is a factor occurring in \( y \) if and only if \( u \) is a factor occurring in \( x \), and \( u \) is a factor occurring in \( y \) if and only if \( \tilde{u} \) is a factor occurring in \( \tilde{x} \). This concludes the proof. \( \square \)

We will be dealing with 2-binomial squares so, in particular, with abelian squares. The next lemma permit to "desubstitute", meaning that we are looking for the inverse image of a factor under the considered morphism.

Lemma 2. Let \( u, v \in A^* \) be two abelian equivalent non-empty words such that \( uv \) is a factor occurring in \( x \). There exists \( u', v' \in A^* \) such that \( u'v' \) is a factor of \( x \), and either:

1. \( u = g(u') \) and \( v = g(v') \);
2. or, \( \tilde{u} = g(v') \) and \( \tilde{v} = g(u') \).

Proof. We will make an extensive use of Observation 1. Note that \( u \) and \( v \) must contain at least one 0 or one 2. Obviously \( e(uv) \) is an abelian square of \( (02)^i \), thus either \( e(u) = e(v) = (02)^i \) or \( e(u) = e(v) = (20)^i \) for an \( i > 0 \).

If \( e(u) = e(v) = (02)^i \), then we have \( u = a \cdot 2 \cdot b \) and \( v = c \cdot 0 \cdot 2 \cdot d \) with \( a, b, c, d \in \{\varepsilon, 1\} \). In this case, we deduce that \( u \) and \( v \) belongs to \( A^* \). Otherwise stated, since \( uv \) is a factor of \( x \), there exists a factor \( u'v' \) in \( x \) such that \( g(u') = u \) and \( g(v') = v \).

Otherwise we have \( e(u) = e(v) = (20)^i \). Thanks to Lemma 1 \( \tilde{v}u \) is a factor occurring in \( x \), and \( e(\tilde{u}) = e(\tilde{v}) = (02)^i \). Thus we are reduced to the previous case, and there is a factor \( u', v' \) in \( x \) such that \( g(u') = \tilde{v} \) and \( g(v') = \tilde{u} \). \( \square \)

Let \( u \) be a word. We set

\[
\lambda_u := \begin{bmatrix} u \\ 01 \\ u \\ 12 \end{bmatrix}.
\]

When we use the desubstitution provided by the previous lemma, the shorter factors \( u' \) and \( v' \) derived from \( u \) and \( v \) keep properties from their ancestors.

Lemma 3. Let \( u, v \in A^* \) be two abelian equivalent non-empty words such that \( uv \) is a factor occurring in \( x \). Let \( u', v' \) be given by Lemma 2. If \( \lambda_u = \mu_v \), then \( u' \) and \( v' \) are abelian equivalent and \( \lambda_{u'} = \mu_{v'} \).

Proof. If we are in the second situation described by Lemma 2 then \( \tilde{v}u \) is also a factor occurring in \( x \). Obviously \( v \) and \( \tilde{u} \) are also abelian equivalent, \( \lambda_\tilde{v} = \lambda_\tilde{u} \) and the case is reduced to the first situation.

Assume now w.l.o.g. that we are in the first situation, that is \( u = g(u') \) and \( v = g(v') \). First observe that we have, for all \( a, b \in A \), \( a \neq b \),

\[
(1) \quad \begin{bmatrix} u' \end{bmatrix}_{ab} = \begin{bmatrix} u' |a| + |u'| |b| \\ 2 \\ 2 \\ 2 \end{bmatrix} - \begin{bmatrix} |u'| |a| \\ 2 \\ 2 \end{bmatrix} - \begin{bmatrix} |u'| |b| \\ 2 \end{bmatrix} - \begin{bmatrix} u' \end{bmatrix}_{ba}.
\]
Since \( u = g(u') \), we derive that
\[
\begin{pmatrix}
u \\
01
\end{pmatrix} = |u'|_0 + \begin{pmatrix}u' \\
00
\end{pmatrix} + \begin{pmatrix}u' \\
02
\end{pmatrix} + \begin{pmatrix}u' \\
12
\end{pmatrix} + \left( |u'|_0 + |u'|_1 \right) \frac{2}{2} - \left( |u'|_0 + |u'|_1 \right) \frac{2}{2} - \left( |u'|_0 + |u'|_1 \right) \frac{2}{2} - \left( |u'|_0 + |u'|_1 \right) \frac{2}{2} - \left( |u'|_0 + |u'|_1 \right) \frac{2}{2}
\]
\[
\begin{pmatrix}
u \\
12
\end{pmatrix} = |u'|_0 + \begin{pmatrix}u' \\
00
\end{pmatrix} + \begin{pmatrix}u' \\
02
\end{pmatrix} + \begin{pmatrix}u' \\
12
\end{pmatrix} + \left( |u'|_0 + |u'|_2 \right) \frac{2}{2} - \left( |u'|_0 + |u'|_2 \right) \frac{2}{2} - \left( |u'|_0 + |u'|_2 \right) \frac{2}{2} - \left( |u'|_0 + |u'|_2 \right) \frac{2}{2} - \left( |u'|_0 + |u'|_2 \right) \frac{2}{2}
\]

Hence
\[
\lambda_u = 2 \left[ \begin{pmatrix}u' \\
02
\end{pmatrix} - \begin{pmatrix}u' \\
01
\end{pmatrix} + \begin{pmatrix}u' \\
12
\end{pmatrix} - \left( |u'|_0 + |u'|_2 \right) \frac{2}{2} - \left( |u'|_0 + |u'|_2 \right) \frac{2}{2} - \left( |u'|_0 + |u'|_2 \right) \frac{2}{2} - \left( |u'|_0 + |u'|_2 \right) \frac{2}{2} - \left( |u'|_0 + |u'|_2 \right) \frac{2}{2}
\]

Similar relations holds for \( v \).

Since \( u' \) and \( v' \) occur in \( x \), from Observation 1, we get
\[
(2) \quad ||v'|_0 - |v'|_2| \leq 1 \text{ and } ||v'|_0 - |v'|_2| \leq 1.
\]

Since \( u \sim_{ab} v \), we have \(|u|_1 = |v|_1\). Hence, from the definition of \( g \), \( |u'|_0 + |u'|_2 = |v'|_0 + |v'|_2 \). In the same way, \(|u|_2 = |v|_2\) implies that \( |u'|_0 + |u'|_1 = |v'|_0 + |v'|_1 \) or equivalently, \( |u'|_1 - |v'|_1 = |v'|_0 - |u'|_0 \).

From the above relation and (2), we get
\[
||v'|_0 - |u'|_0 + |u'|_2| - |v'|_2| \leq 2 \text{ and } ||u'|_0 + |u'|_2 - |v'|_0 + |v'|_2| \leq 2.
\]

Hence the difference of the following two Parikh vectors can only take three values
\[
\Psi(u') - \Psi(v') \in \left\{ \begin{pmatrix}0 \\
0
\end{pmatrix}, \begin{pmatrix}1 \\
0
\end{pmatrix}, \begin{pmatrix}-1 \\
1
\end{pmatrix}, \begin{pmatrix}-1 \\
1
\end{pmatrix} \right\}.
\]

To prove that \( u' \) and \( v' \) are abelian equivalent, we will rule out the last two possibilities.

By assumption, \( \lambda_u = \lambda_v \). So this relation also holds modulo 2. Hence
\[
\begin{pmatrix}0 \\
0
\end{pmatrix} = \begin{pmatrix}|u'|_0 + |u'|_1 \\
2
\end{pmatrix} - \begin{pmatrix}|u'|_0 + |u'|_1 \\
2
\end{pmatrix} - \begin{pmatrix}|u'|_0 + |v'|_0 \\
2
\end{pmatrix} - \begin{pmatrix}|u'|_0 + |v'|_0 \\
2
\end{pmatrix} (mod 2).
\]

Assume that we have
\[
\Psi(u') - \Psi(v') = \begin{pmatrix}1 \\
-1
\end{pmatrix}, \text{i.e.,} \quad |u'|_0 + |u'|_1 = |v'|_0 + |v'|_1, \quad |u'|_0 + |u'|_2 = |v'|_0 + |v'|_2, \quad |u'|_1 + |u'|_2 = |v'|_1 + |v'|_2 - 2.
\]

This leads to a contradiction because then
\[
\begin{pmatrix}|u'|_1 + |u'|_2 \\
2
\end{pmatrix} \neq \begin{pmatrix}|v'|_1 + |v'|_2 \\
2
\end{pmatrix} (mod 2).
\]

Indeed, it is easily seen that \( \begin{pmatrix}4n \\
2
\end{pmatrix} \equiv 0 \) (mod 2), \( \begin{pmatrix}4n+1 \\
2
\end{pmatrix} \equiv 0 \) (mod 2), \( \begin{pmatrix}4n+2 \\
2
\end{pmatrix} \equiv 1 \) (mod 2) and \( \begin{pmatrix}4(n+3) \\
2
\end{pmatrix} \equiv 1 \) (mod 2).

The case \( \Psi(u') - \Psi(v') = \begin{pmatrix}-1 \\
1
\end{pmatrix} \) is handled similarly. So we can assume now that \( \Psi(u') = \Psi(v') \), that is \( u' \sim_{ab} v' \). It remains to prove that \( \lambda_{u'} = \lambda_{v'} \). By assumption \( \lambda_u = \lambda_v \), and from the above formula describing \( \lambda_u \) (resp. \( \lambda_v \)) we get
\[
\begin{pmatrix}u' \\
02
\end{pmatrix} - \begin{pmatrix}u' \\
01
\end{pmatrix} + \begin{pmatrix}u' \\
12
\end{pmatrix} = \begin{pmatrix}v' \\
02
\end{pmatrix} - \begin{pmatrix}v' \\
01
\end{pmatrix} + \begin{pmatrix}v' \\
12
\end{pmatrix}.
\]

To conclude that \( \lambda_{u'} = \lambda_{v'} \), we should simply show that \( \begin{pmatrix}u' \\
02
\end{pmatrix} = \begin{pmatrix}v' \\
02
\end{pmatrix} \). But \( u'v' \) is a factor occurring in \( x \) (from Observation 1 when discarding the 1’s with just alternate 0’s and 2’s) and \( u' \sim_{ab} v' \).

This concludes the proof.

\textbf{Theorem 1.} The word \( x = g^2(0) = 0120210121021021020120121 \cdots \) avoids 2-binomial squares.
Lemma 4. Assume to the contrary that \( x \) contains a 2-binomial square \( uv \) where \( u \) and \( v \) are 2-binomially equivalent. In particular, \( u \) and \( v \) are abelian equivalent and moreover \( \lambda_u = \lambda_v \). We can therefore apply iteratively Lemma 2 and the above lemma to words of decreasing lengths and get finally a repetition \( uu \) with \( u \in A \) in \( x \). But \( x \) does not contain any such factor. \( \square \)

Remark 2. The fixed point of \( g \) is 2-binomial-square free, but \( g \) is not 2-binomial-square-free, that is the image of a 2-binomial-square-free word may contain a 2-binomial-square (e.g., \( g(010) = 01202012 \) contains the square \( 2020 \)).

3. Avoiding 2-binomial cubes over a 2-letter alphabet

Consider the morphism \( h : 0 \mapsto 001 \) and \( h : 1 \mapsto 011 \). In this section, we show that \( h \) is 2-binomial-cube-free, that is for every 2-binomial-cube free binary word \( w \), \( h(w) \) is 2-binomial-cube-free. As a direct corollary, we get that the fixed point of \( h \),

\[
\mathbf{z} = h^{\infty}(0) = 001001011001001011001011011011 \ldots
\]

avoids 2-binomial cubes.

Let \( u \) be a word over \( \{0, 1\} \). The extended Parikh vector of \( u \) is

\[
\Psi_2(u) = \left( |u|_0, |u|_1, \begin{pmatrix} u \\ 00 \\ u \\ 01 \\ u \\ 10 \\ u \\ 11 \end{pmatrix}^{\top} \right).
\]

Observe that two words \( u \) and \( v \) are 2-binomially equivalent if and only if \( \Psi_2(u) = \Psi_2(v) \).

Consider the matrix \( M_h \) given by

\[
M_h = \begin{pmatrix}
2 & 1 & 0 & 0 & 0 & 0 \\
1 & 2 & 0 & 0 & 0 & 0 \\
1 & 0 & 4 & 2 & 2 & 1 \\
2 & 2 & 2 & 4 & 1 & 2 \\
0 & 0 & 2 & 1 & 4 & 2 \\
0 & 1 & 1 & 2 & 2 & 1
\end{pmatrix}.
\]

One can check that \( M_h \) is invertible. We will make use of the following observations:

**Proposition 2.** For every \( u \in \{0, 1\}^* \),

\[
\Psi_2(h(u)) = M_h \Psi_2(u).
\]

**Proposition 3.** Let \( u = 1x \) and \( u' = x1 \) be two words over \( \{0, 1\} \). We have \( |u|_0 = |u'|_0 \), \( |u|_1 = |u'|_1 \),

\[
\begin{pmatrix} u \\ 00 \\ u \\ 11 \end{pmatrix} = \begin{pmatrix} u' \\ 00 \\ u' \\ 11 \end{pmatrix}, \quad \begin{pmatrix} u' \\ 01 \\ u' \\ 10 \end{pmatrix} = \begin{pmatrix} u \\ 01 \\ u \\ 10 \end{pmatrix} + |u|_0, \quad \begin{pmatrix} u' \\ 11 \end{pmatrix} = \begin{pmatrix} u \\ 10 \end{pmatrix} - |u|_0.
\]

In particular, if \( 1x \sim_2 1y \), then \( x1 \sim_2 y1 \). Similar relations hold for \( 0x \) and \( x0 \). In particular, if \( x0 \sim_2 y0 \), then \( 0x \sim_2 0y \).

Let \( x, y \in \{0, 1\} \). We set \( \delta_{x,y} = 1 \), if \( x = y \); and \( \delta_{x,y} = 0 \), otherwise.

**Lemma 4.** Let \( p' \), \( q' \) and \( r' \) be binary words, and let \( a, b \in \{0, 1\} \). Let \( p = h(p')0 \), \( q = a1 h(q')0 b \) and \( r = 1 h(r') \). Then either \( p \not\sim_2 q \) or \( p \not\sim_2 r \).

**Proof.** Assume, for the sake of contradiction, that \( p \sim_2 q \sim_2 r \). Then \( |p'| = |q'| + 1 = |r'| = n \). The following relations can mostly be derived from the coefficients of \( M_h \) (we also have to take into account the extra suffix 0 of \( p \), respectively the extra prefix 1 in \( r \)):

\[
\begin{align*}
\begin{pmatrix} p \\ 01 \end{pmatrix} &= 2 \begin{pmatrix} p' \\ 0 \end{pmatrix} + 2 \begin{pmatrix} p' \\ 1 \end{pmatrix} + 2 \begin{pmatrix} p' \\ 00 \end{pmatrix} + 4 \begin{pmatrix} p' \\ 01 \end{pmatrix} + \begin{pmatrix} p' \\ 10 \end{pmatrix} + 2 \begin{pmatrix} p' \\ 11 \end{pmatrix}, \\
\begin{pmatrix} p \\ 10 \end{pmatrix} &= 2 \begin{pmatrix} p' \\ 0 \end{pmatrix} + 2 \begin{pmatrix} p' \\ 1 \end{pmatrix} + 2 \begin{pmatrix} p' \\ 00 \end{pmatrix} + \begin{pmatrix} p' \\ 01 \end{pmatrix} + 4 \begin{pmatrix} p' \\ 10 \end{pmatrix} + 2 \begin{pmatrix} p' \\ 11 \end{pmatrix}, \\
\Rightarrow \begin{pmatrix} p \\ 01 \end{pmatrix} - \begin{pmatrix} p \\ 10 \end{pmatrix} &= 3 \begin{pmatrix} p' \\ 01 \end{pmatrix} - 3 \begin{pmatrix} p' \\ 10 \end{pmatrix};
\end{align*}
\]
We derive that
\[
\binom{r}{01} = 2\binom{r'}{0} + 2\binom{r'}{1} + 2\binom{r'}{00} + 4\binom{r'}{01} + \binom{r'}{10} + 2\binom{r'}{11},
\]
\[
\binom{r}{10} = 2\binom{r'}{0} + \binom{r'}{1} + 2\binom{r'}{00} + 4\binom{r'}{01} + 4\binom{r'}{10} + 2\binom{r'}{11},
\]
\[
\Rightarrow \binom{r}{01} - \binom{r}{10} = \left(\binom{r'}{1} + 3\binom{r'}{01} - 3\binom{r'}{10}\right).
\]

We also get the following relations:
\[
\binom{q}{01} = 2\binom{q'}{0} + 2\binom{q'}{01} + 4\binom{q'}{01} + 2\binom{q'}{11} + \delta_{a,0}\left[1 + \binom{q'}{0} + 2\binom{q'}{1} + \delta_{b,1}\right] + \binom{q'}{0} + \binom{q'}{1},
\]
\[
\binom{q}{10} = 3\binom{q'}{0} + 2\binom{q'}{1} + 2\binom{q'}{00} + 4\binom{q'}{01} + 2\binom{q'}{11} + 1
+ \delta_{a,1}\left[1 + \delta_{b,0} + 2\binom{q'}{0} + \binom{q'}{1} + \delta_{b,0}\left[1 + \binom{q'}{0} + 2\binom{q'}{1}\right]\right]
\]
\[
= \left(6 - 2\delta_{a,0} - \delta_{b,1}\right)\binom{q'}{0} + \left(6 - \delta_{a,0} - 2\delta_{b,1}\right)\binom{q'}{1} + 4 - 2\delta_{a,0} - 2\delta_{b,1} + \delta_{a,0}\delta_{b,1}
+ 2\binom{q'}{00} + 4\binom{q'}{01} + 2\binom{q'}{11}.
\]

Where for the last equality, we have used the fact that \(\delta_{a,1} = 1 - \delta_{a,0}\) and \(\delta_{b,0} = 1 - \delta_{b,1}\). Finally, we obtain
\[
\binom{q}{01} - \binom{q}{10} = \left(-4 + 3\delta_{a,0} + 3\delta_{b,1}\right)\left[\binom{q'}{0} + \binom{q'}{1}\right] + 3\left(\binom{q'}{01} - 3\binom{q'}{10}\right) = 4 + 3\delta_{a,0} + 3\delta_{b,1}.
\]

Since \(p \sim q \sim r\), we have \(\binom{q}{0} - \binom{q}{1} = \binom{q}{0} - \binom{q}{1}\). In particular, these equalities modulo 3 give
\[
(3) \quad \binom{p'}{0} \equiv \binom{r'}{1} \equiv 2\left[\binom{q'}{0} + \binom{q'}{1} + 1\right] \equiv 2n \pmod{3}.
\]

Now, we take into account the fact that \(p\) and \(r\) are abelian equivalent to get a contradiction. Since \(p = h(p')0\) and \(r = 1h(r')\), we get
\[
\binom{|p|_0}{|p|_1} = \binom{2}{1} \binom{|p'|_0}{|p'|_1} + \binom{1}{0} \binom{|r|_0}{|r|_1} = \binom{2}{1} \binom{|r'|_0}{|r'|_1} + \binom{1}{0}.
\]

Hence, we obtain
\[
\binom{|p|_0 - |r|_0}{|p|_1 - |r|_1} = \binom{0}{0} = \binom{2}{1} \binom{|p'|_0 - |r'|_0}{|p'|_1 - |r'|_1} + \binom{1}{0}.
\]

We derive that \(|p'|_0 - |r'|_0| = -1\) and \(|p'|_1 - |r'|_1| = 1\). Recalling that \(|p'|_0 + |p'|_1 = n. If we subtract the last two equalities, we get \(|p'|_0 + |r'|_1 = n - 1. From (3), we know that \(|p'|_0 \equiv |r'|_1 \pmod{3}\). Hence \(2|p'|_0 \equiv n - 1 \pmod{3}\) and thus
\[
|p'|_0 \equiv 2n - 2 \pmod{3}.
\]

This contradicts the fact again given by (3) that \(|p'|_0 \equiv 2n \pmod{3}\).

Similarly, one get the following lemma.

**Lemma 5.** Let \(p', q'\) and \(r'\) be binary words, and let \(a, b \in \{0, 1\}\). Let \(p = h(p')0 a, q = 1 h(q')0\) and \(r = b h(r')\). Then either \(p \not\sim q\) or \(p \not\sim r\).
Similarly, the form of $H$. Hence, we get $p$.

Assume, for the sake of contradiction, that $p \sim q \sim r$. Then $|p'| = |q'| = |r'| = n$. Taking into account the special form of $p$ and $q$, we get

$$
\begin{align*}
\left( \begin{array}{c}
p_1 \\
p_0
\end{array} \right) &= 2 \left( \begin{array}{c}
p_1' \\
p_0'
\end{array} \right) + 2 \left( \begin{array}{c}
p_1'' \\
p_0''
\end{array} \right) + 4 \left( \begin{array}{c}
p_0' \\
p_1
\end{array} \right) + 2 \left( \begin{array}{c}
p_1' \\
p_0
\end{array} \right) + \delta_{a,1} \left( 1 + 2 \left( \begin{array}{c}
p_0' \\
p_1
\end{array} \right) + \left( \begin{array}{c}
p_1' \\
p_0
\end{array} \right) \right), \\
\left( \begin{array}{c}
p_0 \\
p_1
\end{array} \right) &= \left( \begin{array}{c}
p_1' \\
p_0'
\end{array} \right) + 2 \left( \begin{array}{c}
p_1'' \\
p_0''
\end{array} \right) + 2 \left( \begin{array}{c}
p_0' \\
p_1
\end{array} \right) + 4 \left( \begin{array}{c}
p_0'' \\
p_1'
\end{array} \right) + \delta_{a,0} \left( \left( \begin{array}{c}
p_1' \\
p_0
\end{array} \right) + 2 \left( \begin{array}{c}
p_1' \\
p_0
\end{array} \right) \right), \\
\left( \begin{array}{c}
q_1 \\
q_0
\end{array} \right) &= 2 \left( \begin{array}{c}
q_1' \\
q_0'
\end{array} \right) + 2 \left( \begin{array}{c}
q_1'' \\
q_0''
\end{array} \right) + 2 \left( \begin{array}{c}
q_0' \\
q_1
\end{array} \right) + 4 \left( \begin{array}{c}
q_0'' \\
q_1'
\end{array} \right) + \left( \begin{array}{c}
q_1' \\
q_0
\end{array} \right) + 2 \left( \begin{array}{c}
q_1' \\
q_0
\end{array} \right), \\
\left( \begin{array}{c}
q_0 \\
q_1
\end{array} \right) &= 3 \left( \begin{array}{c}
q_1' \\
q_0'
\end{array} \right) + 3 \left( \begin{array}{c}
q_1'' \\
q_0''
\end{array} \right) + 2 \left( \begin{array}{c}
q_0' \\
q_1
\end{array} \right) + 4 \left( \begin{array}{c}
q_0'' \\
q_1'
\end{array} \right) + 2 \left( \begin{array}{c}
q_1' \\
q_0
\end{array} \right) + 1.
\end{align*}
$$
Hence, we get

$$
\begin{align*}
\left( \begin{array}{c}
p_1 \\
p_0
\end{array} \right) - \left( \begin{array}{c}
p_1 \\
p_0
\end{array} \right) &= -2 \left( \begin{array}{c}
p_1' \\
p_0'
\end{array} \right) + 3 \left( \begin{array}{c}
p_0' \\
p_1
\end{array} \right) - 3 \left( \begin{array}{c}
p_1' \\
p_0
\end{array} \right) + \delta_{a,1} \left( 1 + 3 \left( \begin{array}{c}
p_0' \\
p_1
\end{array} \right) + 3 \left( \begin{array}{c}
p_1' \\
p_0
\end{array} \right) \right), \\
\left( \begin{array}{c}
q_1 \\
q_0
\end{array} \right) - \left( \begin{array}{c}
q_1 \\
q_0
\end{array} \right) &= - \left( \begin{array}{c}
q_1' \\
q_0'
\end{array} \right) - \left( \begin{array}{c}
q_1'' \\
q_0''
\end{array} \right) + 3 \left( \begin{array}{c}
q_0' \\
q_1
\end{array} \right) - 3 \left( \begin{array}{c}
q_1' \\
q_0
\end{array} \right) - 1.
\end{align*}
$$
Since, $p \sim q$, the last two relations evaluated modulo 3 give

$$
|p'|_1 + \delta_{a,1} \equiv 2n + 2 \pmod{3}. \tag{4}$$

Similarly, the form of $r$ gives the following relations

$$
\begin{align*}
\left( \begin{array}{c}
r_1 \\
r_0
\end{array} \right) &= 2 \left( \begin{array}{c}
r_1' \\
r_0'
\end{array} \right) + 2 \left( \begin{array}{c}
r_1'' \\
r_0''
\end{array} \right) + 4 \left( \begin{array}{c}
r_0' \\
r_1
\end{array} \right) + \left( \begin{array}{c}
r_1' \\
r_0
\end{array} \right) + 2 \left( \begin{array}{c}
r_1' \\
r_0
\end{array} \right) + \delta_{b,0} \left( 1 + \left( \begin{array}{c}
r_0' \\
r_1
\end{array} \right) + 2 \left( \begin{array}{c}
r_1' \\
r_0
\end{array} \right) \right), \\
\left( \begin{array}{c}
r_0 \\
r_1
\end{array} \right) &= 2 \left( \begin{array}{c}
r_1' \\
r_0'
\end{array} \right) + \left( \begin{array}{c}
r_1'' \\
r_0''
\end{array} \right) + 2 \left( \begin{array}{c}
r_0' \\
r_1
\end{array} \right) + 4 \left( \begin{array}{c}
r_0'' \\
r_1'
\end{array} \right) + \left( \begin{array}{c}
r_1' \\
r_0
\end{array} \right) + \left( \begin{array}{c}
r_1' \\
r_0
\end{array} \right), \\
\left( \begin{array}{c}
r_0 \\
r_1
\end{array} \right) - \left( \begin{array}{c}
r_0 \\
r_1
\end{array} \right) &= -2 \left( \begin{array}{c}
r_0' \\
r_1'
\end{array} \right) + 3 \left( \begin{array}{c}
r_1' \\
r_0
\end{array} \right) - 3 \left( \begin{array}{c}
r_0' \\
r_1
\end{array} \right) + \delta_{b,0} \left( 1 + 3 \left( \begin{array}{c}
r_0' \\
r_1
\end{array} \right) + 3 \left( \begin{array}{c}
r_1' \\
r_0
\end{array} \right) \right)
\end{align*}
$$
Since, $p \sim r$, the last two relations evaluated modulo 3 give

$$
|p'|_1 + \delta_{a,1} \equiv |r'|_0 + \delta_{b,0} \pmod{3}. \tag{5}$$

Now, we take into account the fact that $p$, $q$ and $r$ are abelian equivalent to get a contradiction. The following two vectors are equal:

$$
\begin{align*}
\left( \begin{array}{c}
p_l \\
p_r
\end{array} \right) &= \left( \begin{array}{c}
p_l \\
p_r
\end{array} \right), \\
\left( \begin{array}{c}
r_l \\
r_r
\end{array} \right) &= \left( \begin{array}{c}
r_l \\
r_r
\end{array} \right).
\end{align*}
$$
We derive easily that

$$
|p'|_1 - |r'|_1 = 1 + \delta_{a,0} - \delta_{b,0}.
$$
On the one hand, using the latter relation and [8]

$$
|r'|_1 + 1 + \delta_{a,0} - \delta_{b,0} + \delta_{a,1} = |p'|_1 + \delta_{a,1} \equiv |r'|_0 + \delta_{b,0} \pmod{3}
$$
Replacing $|r'|_0$ by $n - |r'|_1$, we get $2|r'|_1 + 2 \equiv n + 2\delta_{b,0} \pmod{3}$, or equivalently

$$
|r'|_1 + 1 \equiv 2n + \delta_{b,0} \pmod{3}.
$$
On the other hand, using [11],

$$
|r'|_1 + 1 + \delta_{a,0} - \delta_{b,0} + \delta_{a,1} = |p'|_1 + \delta_{a,1} \equiv 2n + 2 \pmod{3}
$$
and thus,

$$
|r'|_1 \equiv 2n + \delta_{b,0} \pmod{3}.
$$
We get a contradiction, $2n + \delta_{b,0}$ should congruent to both $|r'|_1$ and $|r'|_1 + 1$ modulo 3. \hfill $\square$

We are ready to prove the main theorem of this section.

**Theorem 4.** Let $h : 0 \mapsto 001, 1 \mapsto 011$. For every 2-binomial-cube-free word $w \in \{0,1\}^*$, $h(w)$ is 2-binomial-cube-free.
Proof. Let \( w \) be a 2-binomial-cube-free binary word. Assume that \( h(w) = z_0 \ldots z_{|w|-1} \) contains a 2-binomial cube \( pqr \) occurring in position \( i \), i.e., \( p \sim q \sim r \) and \( w = w' p q r w'' \), where \( |w'| = i \). We consider three cases depending on the size of \( p \) modulo 3.

As a first case, assume that \( |p| = 3n \). We consider three sub-cases depending on the position \( i \) modulo 3.

1.a) Assume that \( i \equiv 2 \pmod{3} \). Then \( p, q, r \) have 1 as a prefix and the letter following \( r \) in \( h(w) \) is the symbol \( z_{i+9n} = 1 \). Hence, the word \( 1^{-1}pqr1 \) occurs in \( h(w) \) in position \( i + 1 \) and it is again a 2-binomial cube. Indeed, thanks to Proposition \( 3 \) we have \( 1^{-1}p1 \sim 2^{-1}q1 \sim 2^{-1}r1 \). This case is thus reduced to the case where \( i \equiv 0 \pmod{3} \).

1.b) Assume that \( i \equiv 1 \pmod{3} \). Then \( p, q, r \) have 0 as a suffix and the letter preceding \( p \) in \( h(w) \) is the symbol \( z_{i-1} = 0 \). Hence, the word \( 0pqr0^{-1} \) occurs in \( h(w) \) in position \( i - 1 \) and it is also a 2-binomial cube. Thanks to Proposition \( 3 \) we have \( 0p0^{-1} \sim 2q0^{-1} \sim 2r0^{-1} \). Again this case is reduced to the case where \( i \equiv 0 \pmod{3} \).

1.c) Assume that \( i \equiv 0 \pmod{3} \). In this case, we can desubstitute: there exist three words \( p', q', r' \) of length \( n \) such that \( h(p') = p, h(q') = q, h(r') = r \) and \( p'q'r' \) is a factor occurring in \( w \). We have \( \Psi_2(p') = \Psi_2(q') = \Psi_2(r') \). By Proposition \( 2 \) and since \( M_q \) is invertible, we have \( \Psi_2(p') = \Psi_2(q') = \Psi_2(r') \), meaning that \( w \) contains a 2-binomial cube \( p'q'r' \).

As a second case, assume that \( |p| = 3n + 1 \). In this case, one of \( p, q \) and \( r \) occur in position 0 modulo 3, one in position 1 modulo 3, and one in position 2 modulo 3. Suppose w.l.o.g. that \( p \) occur in position 0 modulo 3, and \( q \) in position 1 modulo 3. Then there are three factors \( p', q' \) and \( r' \) in \( w \), and \( a, b \in \{0, 1\} \) such that \( p = h(p')0, q = a1h(q')0b \) and \( r = 1h(r') \). By Lemma \( 4 \) this is impossible.

For the final case, assume that \( |p| = 3n + 2 \). In this case again, one of \( p, q \) and \( r \) occur in position 0 modulo 3, one in position 1 modulo 3, and one in position 2 modulo 3. Suppose w.l.o.g. that \( p \) occur in position 0 modulo 3, and \( q \) in position 1 modulo 3. Then there are three factors \( p', q' \) and \( r' \) in \( w \), and \( a, b \in \{0, 1\} \) such that \( p = h(p')0a, q = 1h(q')0b \) and \( r = 1b1h(r') \). By Lemma \( 5 \) this is impossible. \( \square \)

Corollary 5. The infinite word \( z = 001001011 \cdots \) fixed point of \( h : 0 \mapsto 001, 1 \mapsto 011 \) avoids 2-binomial cubes.

References

[1] F. M. Dekking, Strongly nonrepetitive sequences and progression-free sets, J. Combin. Theory Ser. A 27 (1979), 181-185.
[2] P. Erdős, Some unsolved problems, Magyar Tud. Akad. Mat. Kutató Int. Közl. 6 (1961), 221-254.
[3] M. Huova, Existence of an infinite ternary 64-abelian square-free word, preprint.
[4] M. Huova, J. Karhumäki, Observations and problems on \( k \)-abelian avoidability, In Combinatorial and Algorithmic Aspects of Sequence Processing (Dagstuhl Seminar 11081), (2011) 2215-2219.
[5] M. Huova, J. Karhumäki, On unavoidability of \( k \)-abelian squares in pure morphic words, J. Integer Seq. 16 (2013), no. 2, Article 13.2.9.
[6] J. Karhumäki, A. Saarela, L. Q. Zamboni, On a generalization of Abelian equivalence and complexity of infinite words, J. Combin. Theory Ser. A 120 (2013), 2189-2206.
[7] V. Keränen, Abelian squares are avoidable on 4 letters, Lecture Notes in Comput. Sci. 623 (1992), 41-52.
[8] M. Lothaire, Combinatorics on Words, Cambridge Mathematical Library, Cambridge University Press, (1997).
[9] M. Rao, On some generalizations of abelian power avoidability, preprint (2013).
[10] M. Rigo, P. Salimov, Another generalization of abelian equivalence: binomial complexity of infinite words, WORDS 2013, Lect. Notes in Comput. Sci. 8079, 217-228, Springer Verlag (2013).
[11] A. Thue, Über unendliche Zeichenreihen, Norske vxd. Selsk. Skr. Mat. Nat. Kl. 7 (1906), 1-22.
[12] A. Thue, Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen, Norske vxd. Selsk. Skr. Mat. Nat. Kl. 1 (1912), 1-67.
AVOIDING 2-BINOMIAL SQUARES AND CUBES

(M. Rao) CNRS, LIP, ENS Lyon, 46 alle d'Italie, 69364 Lyon Cedex 07 - France, michael.rao@ens-lyon.fr

(M. Rigo) Dept of Math., University of Liège, Grande traverse 12 (B37), B-4000 Liège, Belgium, M.Rigo@ulg.ac.be

(P. Salimov) Dept of Math., University of Liège, Grande traverse 12 (B37), B-4000 Liège, Belgium and Sobolev Institute of Math., 4 Acad. Koptyug avenue, 630090 Novosibirsk, Russia