Local Squares, Periodicity and Finite Automata*

Mari Huova, Juhani Karhumäki, Aleksi Saarela, and Kalle Saari

Department of Mathematics and Turku Centre for Computer Science TUCS
University of Turku, FI-20014 Turku, Finland
{mahuov,karhumak,amsaar,kasaar}@utu.fi

Abstract. We consider the general problem when local regularity implies the global one in the setting where local regularity means the existence of a square of certain length in every position of an infinite word. The square can occur as centered or to the left or to the right from each position. In each case there are three variants of the problem depending on whether the square is that of words, that of abelian words or, as an in between case, that of so called k-abelian words. The above nine variants of the problem are completely solved, and some open problems are addressed in the k-abelian case. Finally, an amazing unavoidability result for 2-abelian squares is obtained.

1 Introduction

Questions when local properties imply global ones are of fundamental importance in many area of mathematics. Among the most well known problems of this type is Burnside Problem [3]. It asks whether a finitely generated group, where each of its subgroups generated by a single element is finite, is necessarily finite as well. A remarkable paper of Adian and Novikov [1] shows that the answer to this question is "no" – that is we do not have the above desired implication. In the case of free semigroups the situation is the same, only the proof is much simpler application of avoidability properties of words, see [14] or [11].

Positive results of the above nature are recently discovered and studied in connection with infinite words. Here the local regularity is described, e.g., as a property that the word contains a certain repetition everywhere, and the global one as the requirement that the word is (ultimately) periodic. A remarkable result here is the characterization of [13] stating that a one-way infinite word is ultimately periodic if and only if each of its long enough prefixes ends up with a repetition of a word of order at least $\varphi + 1$, where φ is the golden ratio. Consequently, the local regularity "having always a square" does not imply the global one while "having always a cube" does so. The analysis of [13] was refined in [8], see also [10], by restricting the length of the repetition in the local regularity condition. A general treatment of these questions can be found in Chapter 8 of [12].

* Supported by the Academy of Finland under grants 121419 and 134190 and by the Väisäla Foundation.

C.S. Calude, G. Rozenberg, A. Salomaa (Eds.): Maurer Festschrift, LNCS 6570, pp. 90–101, 2011.

In this note we define the local regularity as the existence of a square at each position of the word. We obtain three variants of the problem depending on whether the square is centered, to the left or to the right of each position, respectively. In addition, in each case we obtain three subproblems corresponding to the cases where squares are those of ordinary words, those of abelian words or, as an in between case, those of so-called k-abelian words described in a moment. We parameterize all of these nine problems by adding a restriction on the length of the square. With this setting we completely characterize when the local regularity, i.e. having a square of certain length everywhere, implies the global periodicity, i.e. being ultimately periodic.

The method used here is that from [8]. Namely, for a fixed length n we can construct a finite automaton accepting, as infinite runs, exactly those words which obey a given local regularity constrain. Then the existence of intersecting loops characterizes whether a nonperiodic word of the required form can exist. It follows from the method that whenever a non-ultimately periodic word is found, actually nondenumerably many of those are constructed. Despite of the simplicity of the approach, computer verifications are crucial for our results. The above is easily modified for two-way infinite words.

Our results have another interesting interpretation. Let us consider a one-way infinite word as a dynamical process where it is defined step by step. Assume further that in this process we preserve a given local regularity property. Then our results have identified the exact borderlines, with respect to the length of squares, of *predictable* vs *chaotic behavior* in all of our nine problems. In the predictable case we obtain only periodic words while in the chaotic case we obtain nondenumerably many (nonperiodic) words in our process.

We conclude this introduction by defining informally the notion of *k-abelian equivalence*. For a natural number k two words are k-abelian equivalent if they possess common prefixes and suffixes of length $k-1$, respectively, and each factor of length k occurs equally many times in these words. The value $k = 1$ corresponds to the *abelian equivalence*. The condition for prefixes and suffixes is introduced in order to locate the k-abelian equivalence properly in between the ordinary equivalence (equality) and the abelian equivalence. Finally, a k-abelian square is a word of the form uv where u and v are k-abelian equivalent.

Our last section analyzes some basic properties of the k-abelian equivalence, as well as points out some intriguing open questions. More precisely, we characterize in the binary alphabet the 2-abelian and 3-abelian equivalence classes, as well as analyze the numbers of these equivalence classes of words of length n. In the case $k = 2$ we obtain an exact quadratic formula, while in the case $k = 3$ we obtain an estimate $\Omega(n^4)$. In general, we obtain a polynomial upper bound, although the degree of the polynomial is exponential in k. We conclude by considering avoidability questions for k-abelian powers. More precisely, we ask what is the smallest alphabet size when 2-abelian cubes (resp. 2-abelian squares) can be avoided in infinite words. Obviously, these values are in between the corresponding values for word- and abelian repetitions, that is either 2 or 3, and 3 or 4, respectively, see [11] and [5]. For cubes we check by a computer that there exist binary words of length 100000, avoiding 2-abelian cubes. This makes us to conjecture that 2-abelian cubes can be avoided in the binary alphabet, exactly as ordinary cubes. On the other hand 2-abelian squares – quite surprisingly – can not be avoided in a ternary alphabet. The longest word avoiding such squares is of length 537. Hence the smallest alphabet they can be avoided in is quartic.

2 Preliminaries

In this section we fix our terminology, and in particular define our crucial notions. For words we refer to [12] and for automata to [7].

We denote by Σ a finite alphabet and Σ^+ and Σ^ω the sets of nonempty finite words and one-way infinite words over Σ, respectively. An element $w \in \Sigma^\omega$ is *ultimately periodic* if it can be written in the form $w = uvv\cdots = uv^\omega$ for some finite words u and v. This notion extends in a natural way to two-way infinite words: two-way infinite word w is *ultimately periodic to the right* if it can be written in the form $w = uv$ where u is one-way infinite word to the left and v is ultimately periodic word in Σ^ω. Similarly, we define *two-way ultimately periodic words to both directions*. In our considerations, words of these forms are viewed as *globally regular* words. *Local regularity* is defined in our considerations as a repetition such as a square, a cube etc.

We consider three types of equivalence relations on Σ^+: the ordinary equality of words, the abelian (or commutative) equivalence of words, and, as an in between case, so-called k-abelian equivalence of words. The last one is defined as follows. Let k be a natural number. Two finite words u and v are *k-abelian equivalent* if and only if

– $\mathrm{pref}_{k-1}(u) = \mathrm{pref}_{k-1}(v)$ and $\mathrm{suf}_{k-1}(u) = \mathrm{suf}_{k-1}(v)$
– for all $x \in \Sigma^k$, $\#(x,u) = \#(x,v)$,

where pref_{k-1} (resp. suf_{k-1}) refers to the maximal prefix (resp. suffix) of length at most $k-1$ and $\#$ counts the number of (possibly overlapping) occurrences of x in the word. Let us denote this equivalence relation by $\equiv_{a,k}$. Then, clearly, $\equiv_{a,1}$ coincides with $\equiv_a$, the usual abelian equivalence relation on words.

The first condition is introduced in order to have the following relations, which are straightforward to conclude:

$$u = v \Rightarrow u \equiv_{a,k} v \Rightarrow u \equiv_a v, \quad \text{for all k.}$$

Indeed, for words $u = 1101$ and $v = 0110$ we have $u \not\equiv_a v$ while u and v satisfy the second condition above for $k = 2$. The notion of abelian and k-abelian repetitions are now defined in a natural way. For example, word w is *k-abelian square* if it can be written as $w = uv$ where $u \equiv_{a,k} v$. It follows that we can talk about words *avoiding*, e.g., 2-abelian squares.

3 Local Squares vs. Periodicity

We examine the following problem: for a given number n, if a binary right-infinite word contains at every position a square of a word of length at most n, is the word necessarily ultimately periodic?

To make this question precise, we must define what having a square at every position means. We give three different definitions, which lead to three variations of the problem.

A word w contains everywhere a

- *left square* of length at most n, if every factor of w of length $2n$ has a nonempty square as a suffix,
- *right square* of length at most n, if every factor of w of length $2n$ has a nonempty square as a prefix,
- *centered square* of length at most n, if every factor of w of length $2n$ has a nonempty square exactly in the middle, i.e. is of the form $uxxv$, where $|u| = |v|$ and $x \neq 1$.

In addition to ordinary squares, we can give similar definitions for k-abelian squares. We will study the cases $k = 1$ and $k = 2$, i.e. abelian and 2-abelian equivalences. This gives a total of nine variations of the problem.

We begin by giving a method to solve the following problem: for a number n, if a binary right-infinite word contains everywhere a left square of length at most n, is the word necessarily ultimately periodic? This method can then be modified for the other eight variations of the problem.

We define an automaton as follows: the set of states is Σ^{2n-1}, where $\Sigma = \{a, b\}$. If $c, d \in \Sigma$, $u \in \Sigma^{2n-2}$ and cud has a nonempty square as a suffix, then the value of the transition function at (cu, d) is defined as $\delta(cu, d) = ud$; otherwise δ is undefined. All states are initial and final. The automaton is deterministic, except that there are many initial states.

The following theorems give a way to solve our problem.

Theorem 1. *Let $w = uv$, where $u \in \Sigma^{2n-1}$ and $v \in \Sigma^\omega$. The word w contains a left square of length at most n everywhere if and only if the above automaton accepts every prefix of v starting from the state u.*

Proof. Follows from the definition of the automaton. □

Theorem 2. *There exists a binary aperiodic right-infinite word that contains a left square of length at most n everywhere if and only if the above automaton has two intersecting cycles.*

Proof. If the automaton has two intersecting cycles, then there is a state u and words s, t such that $\delta(u, as) = u = \delta(u, bt)$. Now every word in $\{as, bt\}^\omega$ contains a left square of length at most n everywhere, and there are aperiodic words in this set.

If there are no intersecting cycles and $uv \in \Sigma^\omega$ is such that $\delta(u, t)$ is defined for every prefix t of v, then there must be a cycle such that $\delta(u, t)$ is in that cycle for every long enough prefix t. Then v must be ultimately periodic. □

If the automaton does not have intersecting cycles, then there are only ultimately periodic words that contain a left square of length at most n everywhere. The number of these words can be countably infinite. For example, for every m the word $(ab)^m a^\omega$ has a left square of length at most two everywhere. However, there are only finitely many possible periodic parts, because the automaton has only finitely many cycles. Thus the situation is very structured.

On the other hand, if the automaton has two intersecting cycles, then there are aperiodic words that contain a left square of length at most n everywhere. Moreover, the number of such words is uncountable, and the situation can be viewed as chaotic.

The existence of intersecting cycles is fairly easy to check algorithmically: we determine the strongly connected components (for example with Tarjan's algorithm) and check whether in some component there is a state from which there are two transitions into the same component.

Using this method we can see that the smallest value of n for which there is an aperiodic word containing a left square of length at most n everywhere is $n = 5$. This was already proved in [8]. The automaton has two strongly connected components that contain intersecting cycles. A representation of one of them is in Fig. 1; the other component can be obtained by exchanging the letters. Here we have omitted those states, which have only one transition coming to them and only one transition leaving from them. In addition to these two components, there are also components consisting of a single cycle, and components consisting of a single state. Every cycle generates ultimately periodic words and single states may generate prefixes of infinite words, but only intersecting cycles can generate aperiodic infinite words.

We can use the same method for k-abelian squares: in the definition of the transition function we simply require cud to have a suffix that is a k-abelian square instead of an ordinary square. For 2-abelian left squares the smallest possible value is also $n = 5$; the automaton still has two strongly connected components that contain intersecting cycles, but they are slightly more complicated, see Fig. 2. For abelian left squares the smallest possible value is $n = 3$. The automaton has two strongly connected components that contain intersecting cycles, which have sizes 18 and 12, and two one-state cycles. There are so many transitions that it would be difficult to draw a clear picture. In fact, the automaton is almost the whole De Bruijn graph; there are only eight states from which there is only one transition.

The method can also be modified for right squares: in the definition of the transition function we require cud to have a prefix that is a square. If δ_L, δ_R are the transition functions for left and right squares, then $\delta_L(cu, d) = ud$ if and only if $\delta_R(du^R, c) = u^R c$, where u^R is the reverse of u. Thus the automaton

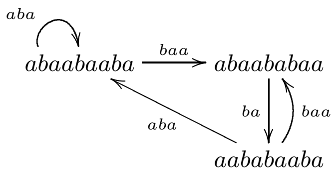

Fig. 1. A component of the automaton for left squares of length $n = 5$

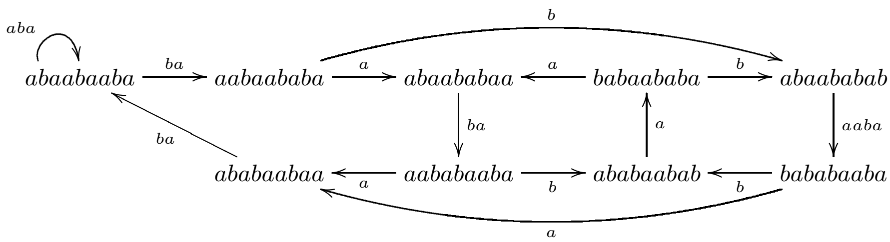

Fig. 2. A component of the automaton for left 2-abelian squares of length $n = 5$

for right squares is obtained from the automaton for left squares by reversing transitions and relabeling states and transitions. It follows that the automaton for left squares has two intersecting cycles if and only if the automaton for right squares has.

Finally, we can define a similar automaton for centered squares. It turns out that in this case the ultimate periodicity is harder to avoid: in the abelian case the smallest value of n for which there is an aperiodic word containing a centered abelian square of length at most n everywhere is $n = 8$. The automaton has one strongly connected component that contains intersecting cycles. This component has 148 states. In the 2-abelian case the smallest possible value is $n = 12$. The automaton has two strongly connected components that contain intersecting cycles. Both components have 222 states. Small parts of these two automata are represented in Fig. 3. These parts are sufficient to generate some examples of aperiodic words that satisfy the conditions. The automata contain also many other intersecting cycles, which are not represented here. Abelian squares (left, right and centered) were studied in [2]. In particular there it was proved that there are some connections with the Thue-Morse morphism and its generalizations.

For ordinary centered squares the situation changes completely: no finite value of n is large enough. This can be proved using the Critical Factorization Theorem (see [4]). The first part of the next theorem is Theorem 8.3.5 in [12].

Theorem 3. *If there is an n such that $w \in \Sigma^\omega$ has a centered square of length at most n everywhere, then w is ultimately periodic. On the other hand, there are aperiodic two-way infinite words that have centered squares everywhere.*

Proof. From the existence of such a number n it follows that the local period of w is at most n at every position. By the Critical Factorization Theorem, also the

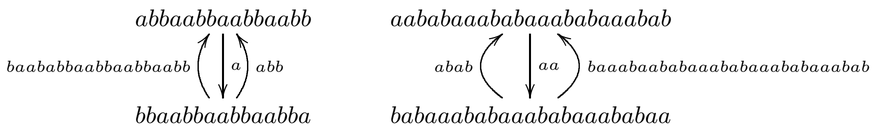

Fig. 3. Small parts of the automata for centered abelian and 2-abelian squares of lengths $n = 8$ and $n = 12$

global period of every prefix of w is bounded by n. Now w must be ultimately periodic by the theorem of Fine and Wilf.

Suppose that a finite word w does not have a centered square at some position, say $w = uv$, where no suffix of u is a prefix of v. Now w can be extended to both directions so that it has a centered square at this position: $vtuvtu$ is a proper extension for every word t. We can repeat this for other positions, and as a limit we get a two-way infinite word that has a centered square everywhere. The t-words can be chosen easily so that the infinite word is aperiodic. □

Above we have studied one-way infinite words. The same questions can be asked also for two-way infinite words. In this case the answers are the same as for one-way infinite words.

Theorem 4. *There is a binary aperiodic right-infinite word containing an ordinary (2-abelian, abelian) left (right, centered) square of length at most n everywhere if and only if there is such a two-way infinite word.*

Proof. If there is an aperiodic right-infinite word, then the automaton has two intersecting cycles, say $\delta(u, as) = u = \delta(u, bt)$ for some state u and words s, t. Now every two-way infinite word formed of as and bt satisfies the conditions, and there are aperiodic such words.

On the other hand, if there is an aperiodic two-way infinite word, then it is aperiodic to the left or to the right. If it is aperiodic to the right, then the automaton has two intersecting cycles. If it is aperiodic to the left, then the automaton, where all transitions are reversed, has two intersecting cycles. But then also the original automaton has two intersecting cycles. □

We summarize the results obtained in this section by presenting in Table 1 the minimal values of n for which there are aperiodic right-infinite words containing an ordinary (or 2-abelian or abelian) left (or right or centered) square of length at most n everywhere.

We conclude this section with two remarks on general k-abelian case. First, for left and right squares the values of Table 1 would remain as 5; the k-abelian case is in between of those of words and 2-abelian cases. For the centered variant of the problem the exact borderline for k-abelian repetitions when $k \geq 3$ is unknown. It looks that our straightforward computations might easily be infeasible.

Table 1. Optimal values for local regularity which does not imply global regularity in our problems

	words	2-abelian	abelian
left	5	5	3
right	5	5	3
centered	∞	12	8

4 *k*-Abelian Equivalence: Observations and Open Problems

In this section we give characterizations of the equivalence classes of 2-abelian and 3-abelian words over a binary alphabet. We count the number of the equivalence classes of 2-abelian words over a binary alphabet and the size of each such an equivalence class. We examine the number of the equivalence classes also in general and give an open problem of the subject. In addition, we formulate an open problem concerning avoidability of 2-abelian repetitions. We then end this section by a discussion that indicates that 2-abelian words sometimes behave like ordinary words and sometimes like abelian words.

First we give characterizations for the equivalence classes for 2- and 3-abelian words over binary alphabet by which we mean that we define a representative for each equivalence class.

Example 1. In a binary alphabet $\Sigma = \{a, b\}$ the characterization of the equivalence classes of 2-abelian words can be given in the form:

$$aa^k b^l (ab)^m a^n \quad \text{or} \quad bb^k a^l (ba)^m b^n,$$

where $k, l, m \geq 0$ and $n \in \{0, 1\}$. So, we have in the beginning of the word all the factors of form aa and bb.

We remark that the above characterization is not unambiguous in a few cases, for example if there does not exist the factor bb but aa exists. Then we may express the same class in two different ways, $aa^k b^1 (ab)^m a^n$ or $aa^{k-1} b^0 (ab)^{m+1} a^n$. The following characterization of the equivalence classes of 3-abelian words is not unambiguous either in some cases. The equivalence class of a word depends in fact on the number of factors of the form aaa, bbb and 3-letter factors containing aa or bb. Clearly, the length of the word and the first and the last two letters are significant, too.

Example 2. In the alphabet $\Sigma = \{a, b\}$ the characterization of the equivalence classes of 3-abelian words can be given in the following form containing eight possible combinations:

$$\left.\begin{array}{l} aaa^k b^l (aabb)^m \\ bbb^k a^l (aabb)^m \\ abb^k a^l (aabb)^m \\ baa^k b^l (aabb)^m \end{array}\right\} \text{or} \; * \text{ connected with } * \left\{\begin{array}{l} (aab)^g (ab)^h b^i a^j \;\text{ or} \\ (abb)^g (ab)^h b^i a^j , \end{array}\right.$$

where $k, l, m, g, h \geq 0$, $i \in \{0, 1\}$ and $j \in \{0, \ldots, 2 - i\}$. If we restrict to classes with $k > 1$ and $l > 2$, the given representation is unambiguous.

Now we can count the number of the equivalence classes of 2-abelian words in the binary alphabet $\Sigma = \{a, b\}$.

Example 3. If the length of the word is one there exist two equivalence classes, namely a and b. If the length is two there exist four equivalence classes aa, ab, ba and bb.

We consider next the words of length $n > 2$ and containing k a-letters and hence $n-k$ b-letters. We have a correspondence between the number of different letters and the number of the equivalence classes:

number of a-letters	number of classes
$k = 0 \vee k = n$	$\Rightarrow 1$
$k = 1 \vee k = n-1$	$\Rightarrow 3$
$1 < k < n-1$	$\Rightarrow 2\min(k,n-k) + \min(k-1,n-k) + \min(k,n-k-1)$

From these we obtain the number of the equivalence classes of words with length $n > 2$:

$$\begin{cases} 8 + \sum_{k=2}^{n-2}(4\min(k,n-k)-1), & \text{if } 2 \nmid n \\ 6 + 2n + \sum_{k=2}^{\frac{n}{2}-1}(4\min(k,n-k)-1) + \sum_{k=\frac{n}{2}+1}^{n-2}(4\min(k,n-k)-1), & \text{if } 2 | n \end{cases}$$

If we count the given sums we get as a conclusion the following theorem.

Theorem 5. *The number of 2-abelian equivalence classes consisting of words of length n over a binary alphabet is $n^2 - n + 2$ and thus the number is $\Theta(n^2)$, where $n > 0$ is the length of words and alphabet is binary.*

We can also examine the sizes of the equivalence classes in the case of binary alphabet and 2-abelian equivalence. Consider the words beginning with a, ending with b and containing factors of the form aa, ab, ba and bb k-, $(k-1)$-, l- and m-times, respectively. Then the equivalence class contains $\binom{l+k-1}{k-1}\binom{m+k-1}{k-1}$ such words. Similarly, there exist $\binom{l+k}{k}\binom{m+k-1}{k-1}$ words in the equivalence class containing words beginning and ending with a and having factors of the form aa, ab, ba and bb k-, k-, l- and m-times, respectively. Results for words beginning with b are similar and these four cases cover all the possible 2-abelian words over binary alphabet.

From the characterization of the equivalence classes of 3-abelian words in example 2 we see that the number of the equivalence classes in this case is $\Omega(n^4)$. We have five independent variables, k, l, m, g and h, and if we restrict $k > 1$ and $l > 2$, each combination of these five values gives a different equivalence class. If we now fix the length of the words to be n we obtain a relation

$$k + l + 4m + 3g + 2h + \alpha = n,$$

where $\alpha \in \{2, 3, 4\}$ depending on i and j. We may restrict to analyze words long enough and to subsets of the equivalence classes and hence the equation can be modified to the form:

$$12k' + 12l' + 4(3m') + 3(4g') + 2(6h') = 12n'.$$

Now we may count the number of solutions of equation $\sum_{i=1}^{5} x_i = N$, where $x_i > 0$ for all $i \in \{1, \ldots, 5\}$ and N is fixed. The number of solutions is $\Theta(N^4)$ which implies that the number of 3-abelian equivalence classes of words of length n is $\Omega(n^4)$.

Contrary to the 2-abelian case the exact formula for the number of the 3-abelian equivalence classes is not a polynomial. This can be concluded from the list of the equivalence classes for small values of n, see [6].

Theorem 6. *The number of k-abelian equivalence classes over Σ^n is $O(n^{|\Sigma|^k-1})$.*

Proof. An equivalence class over Σ^n with a representative $u \in \Sigma^n$ is uniquely determined by the $|\Sigma|^k$ nonnegative integers $|u|_z$, where $z \in \Sigma^k$, and the prefix and suffix of u of length $k-1$. Since the number of length-k factors in a word of length n is $n-k+1$, counting multiplicities, the number of distinct equivalence classes is at most

$$\begin{aligned}|\Sigma|^{2(k-1)} \cdot \Big|\big\{(i_1, i_2, \ldots, i_{|\Sigma|^k}) \;\big|\; i_j \geq 0 \text{ and } i_1 + \cdots + i_{|\Sigma|^k} = n-k+1\big\}\Big| \\ = |\Sigma|^{2(k-1)} \cdot \binom{n-k+|\Sigma|^k}{|\Sigma|^k - 1} \\ = O(n^{|\Sigma|^k-1}).\end{aligned}$$

□

It is likely that the real upper bound is smaller than the given one and hence we can state an open problem:

Open problem 1: Give a better estimate to the number of the equivalence classes of k-abelian words.

Finally, we give an open problem concerning k-abelian repetitions. It is known that cubes are avoidable over the binary alphabet, for example the infinite word of Thue-Morse accomplishes this property, see [11]. On the other hand, it is easy to see that abelian cubes are not avoidable. Though, for abelian words over binary alphabet the repetitions of fourth order are avoidable (see [5]) and so in k-abelian case the order of repetition that can be avoided in a binary alphabet is either three or four. We formulate this in the case of 2-abelian words as an open problem:

Open problem 2: Does there exist an infinite binary word that avoids 2-abelian cubes?

The analysis we made with computers reveals that there exist cube-free 2-abelian words longer than 100 000 letters. This does not prove the existence of such an infinite word but supports the assumption that cubes would be avoidable for 2-abelian words over binary alphabet, exactly as in the case of words.

The problem could also be expressed in terms of the size of an alphabet in which cubes can be avoided. The size of the alphabet is now two or three because abelian cubes are avoidable over ternary alphabet, see [5].

A similar question can be asked for 2-abelian squares. Already squares are avoidable for words over 3-letter alphabet (see [11]) but for abelian words over ternary alphabet the maximal length of a word avoiding squares is seven, as can be easily checked. It is known, although this is not easy to prove, that there exists an infinite abelian word avoiding squares over 4-letter alphabet, see [9]. This indicates that for k-abelian words avoiding squares the size of a smallest alphabet is at least three and at most four. In the case $k = 2$ we discovered the following result:

Theorem 7. *The smallest alphabet in which the 2-abelian squares can be avoided is a 4-letter alphabet.*

We executed with a computer a similar construction of square-free 2-abelian words over ternary alphabet than in the case of cube-free 2-abelian words over binary alphabet. The maximal length of square-free 2-abelian words that could be constructed is 537 letters and every longer word over ternary alphabet contains a 2-abelian square. The constructed word of this maximal length 537 is given in example 4. This word is unique up to the permutations of the alphabet. With earlier results mentioned above this shows that the alphabet to avoid squares have to contain at least four letters, and this is, indeed, enough. The behavior of 2-abelian words is in this situation similar with abelian words.

Example 4. The word of length 537 over ternary alphabet $\Sigma = \{a, b, c\}$ that avoids 2-abelian squares:

abcbabcacbacabacbabcbacabcbabcabacabcacbacabacbabcbacbcacbabcacbcabcba
bcabacbabcbacbcacbacabacbabcbacabcbabcabacabcacbacabacbabcbacbcacbacab
acbcabacabcacbcabcbacbcacbacabacbabcbacbcacbabcacbcabcbabcabacbabcbacb
cacbacabacbabcbacabcbabcabacabcacbacabacbabcbacbcacbacabacbcabacabcacb
cabcbabcabacabcacbacabacbabcbacabcbabcabacabcacbcabcbabcabacbabcbacbca
cbabcacbcabcbabcabacabcacbcabcbacbcacbacabacbcabacabcacbcabcbabcabacab
cacbacabacbabcbacabcbabcabacabcacbcabcbabcabacbabcbacbcacbabcacbcabcba
bcabacabcacbacabacbabcbacabcbabcabacabcacbabcba.

References

1. Adian, S.I., Novikov, P.S.: Infinite periodic groups I, II, III. Izv. Akad. Nauk SSSR. Ser. Mat. 32(1,2,3), 212–244, 251–524, 709–731 (1968)
2. Avgustinovich, S., Karhumäki, J., Puzynina, S.: On abelian versions of Critical Factorization Theorem. In: Proceedings of the 13th Mons. Theoretical Computer Science Days (2010)
3. Burnside, W.: On an unsettled question in the theory of discontinuous groups. Quart. J. Pure and Appl. Math. 33, 230–238 (1902)
4. Choffrut, C., Karhumäki, J.: Combinatorics of words. In: Rozenberg, G., Salomaa, A. (eds.) Handbook of Formal Languages, vol. 1, pp. 329–438. Springer, Heidelberg (1997)
5. Dekking, F.M.: Strongly nonrepetitive sequences and progression-free sets. J. Combin. Theory Ser. A 27(2), 181–185 (1979)
6. Harmaala, E.: Private communication
7. Hopcroft, J.E., Ullman, J.D.: Introduction to Automata Theory, Languages, and Computation. Addison-Wesley, Reading (1979)

8. Karhumäki, J., Lepistö, A., Plandowski, W.: Locally periodic versus globally periodic infinite words. J. Combin. Theory Ser. A 100(2), 250–264 (2002)
9. Keränen, V.: Abelian squares are avoidable on 4 letters. In: Kuich, W. (ed.) ICALP 1992. LNCS, vol. 623, pp. 41–52. Springer, Heidelberg (1992)
10. Lepistö, A.: On Relations between Local and Global Periodicity. PhD thesis, University of Turku (2002)
11. Lothaire, M.: Combinatorics on Words. Addison-Wesley, Reading (1983)
12. Lothaire, M.: Algebraic Combinatorics on Words. Cambridge University Press, Cambridge (2002)
13. Mignosi, F., Restivo, A., Salemi, S.: Periodicity and the golden ratio. Theoret. Comput. Sci. 204(1-2), 153–167 (1998)
14. Morse, M., Hedlund, G.A.: Unending chess, symbolic dynamics and a problem in semigroups. Duke Math. J. 11, 1–7 (1944)