# Abelian Squares in Finite Strings

L.J. Cummings
Faculty of Mathematics
University of Waterloo
Waterloo, Ontario
Canada N2L 3G1

ABSTRACT. A string is strongly square-free if it contains no Abelian squares; that is, adjacent substrings which are permutations of each other. We discuss recent results concerning the construction of strongly square-free finite strings.

## 1 Introduction

A string is one of the most fundamental data structures. Some other names for a string are: sequence, vector, codeword, linear array, and list. The goal is to find what inherent properties can be found in strings independently of their appearance in any algebraic structure. Of course strings may be defined using any ordinal, not just finite ones. One can view the entries of a given string as a coloring of the underlying ordinal, but we do not adopt that language. Much research effort has been directed toward countably infinite strings which do or do not exhibit certain properties, but here we will be mainly concerned with finite strings.

An ordered sequence $\mathbf{x} = x_1x_2\cdots x_m$ of elements chosen from a fixed finite set, $A$, of distinct elements is called a finite *string* of length $|x| = m$ over the *alphabet* $A$. In the interests of notational convenience, and without loss of generality, we often choose $A = \{0, \ldots n-1\}$ for fixed $n > 1$ as the alphabet. Every element of the alphabet is also considered to be a string. The elements of the alphabet will be called entries or letters. For each $a \in A$ we define a function $|x|_a$ to be the number of times that $a$ appears in the string $\mathbf{x}$. We freely concatenate strings and write the concatenation of strings $\mathbf{x}$ and $\mathbf{y}$ as simply $\mathbf{xy}$. If a string $\mathbf{x}$=$\mathbf{uv}$ is the concatenation of two strings $\mathbf{u}$ and $\mathbf{v}$ then $\mathbf{u}$ is said to be a *prefix* of $\mathbf{x}$ and $\mathbf{v}$ is said to be a *suffix*. If $\mathbf{v}$ is not empty then $\mathbf{u}$ is said to be a *proper suffix* of $\mathbf{x}$.

JCMCC 24 (1997), pp. 201-208

If $1 \leq i \leq j \leq m$ then the ordered sequence $x_i x_{i+1} \cdots x_j$ is said to be a *substring* of the string x. Interesting combinatorial problems arise by asking when certain strings can occur as substrings of other strings. One of the first questions to ask is whether there are repetitions in a given string; i.e., a substring consisting of a block of letters immediately followed at least once by the same block of letters in the same order. If a string x contains a substring of the form yy then we say x contains the *square* yy.

For example, over the alphabet $\{0, 1\}$, 0010100101 is a square which contains as substrings the squares

00 0101 1010 010010.

One direction of research has been to determine under what conditions a string can avoid "squares" as well as algorithms for finding if they exist in a given string. A string without any substrings which are squares is said to be *square-free*. The strings 010 and 101 are square-free and, moreover, they cannot be extended by concatenation over the alphabet $\{0, 1\}$ on either the right or the left without creating a substring which is a square. A brief survey of square-free infinite strings and references is given in [1]. Viewing the entries of a given string as a coloring of the underlying ordinal, the "ultimate" generalization was given in [11] where it was shown that the class of all ordinals has a square-free 3-coloring. In a different direction, It was shown by Ross and Winklmann [15] that over any alphabet of at least three elements, the set of strings containing "squares" is not context-free. In what follows we concentrate on the less frequently studied question of "Abelian squares".

An *Abelian square* is a string followed by a permutation of itself. Over the alphabet $\{0, 1\}$, 010100 is an Abelian square which contains the squares 0101, 1010, and 00. Thus 010100 contains 4 Abelian squares.

Apparently Erdös [9] first raised the question of the minimum alphabet size over which there exist countably infinite strings without Abelian squares. This is a variant of the corresponding problem for squares raised and solved by Thue [16] in 1906.

**Definition 1** *An* Abelian square *over the alphabet A is a non-empty string of the form*

$$yy^{\sigma} = y_1 \cdots y_k y_{\sigma(1)} \cdots y_{\sigma(k)}$$

*where $\sigma$ is a permutation of A. A string is said to be* strongly square-free *if it contains no Abelian squares.*

Note that every square is an Abelian square corresponding to the identity permutation. Clearly every strongly square-free string is square-free. Over the alphabet $A = \{0, 1, 2\}$, 012201 is an Abelian square while 0102010 is strongly square-free and cannot be extended on the alphabet $A = \{0, 1, 2\}$ without introducing Abelian squares.

Main [12] has shown that, for every alphabet with at least 16 elements, the set of strings which contain Abelian squares is not context-free. In a different direction, Entringer, Jackson, and Schatz [8] proved that every infinite binary string has arbitrarily long Abelian squares. Dekking [7] has shown that there exist infinite binary strings in which no four adjacent substrings appear which are permutations of one another; i.e., no two Abelian squares are adjacent.

In 1970 Pleasants [14] showed that there existed an infinite strongly square-free string on an alphabet of 5 elements. This result was recently sharpened by Keränen [10] who showed the same was true for an alphabet of 4 elements, with a computer-aided proof.

It is folklore that any strongly square-free string over $\{0, 1, 2\}$ has length at most 7 [1].

This can be established, say, by diligently constructing the tree of possible strongly square-free strings starting with 0 and observing that starting with 1 or 2 would yield the same tree. Knowing this allows one to prove:

**Theorem 1** *There are* 117 *distinct strongly square-free finite strings over the alphabet* $\{0, 1, 2\}$.

This is proved in [5]. Accepting the result by Keränen [10], the case of just three letters is seen to be important because it is the last case for which all strongly square-free strings are finite. We include as an appendix a list of all strongly square-free strings on $\{0, 1, 2\}$.

A great deal of attention has been focused on the case of infinite strings, culminating in the work of Keränen which resolved Erdös's question.

## 2 Central Strongly Square-Free Strings

We are not concerned with just a computer listing of strongly square-free strings, but rather determining their global properties.

**Definition 2** *A string $x$ is said to be* central *if it contains at least one entry $a$ such that $|x|_a = 1$. A set of strings $S$ is said to be* central *with respect to $a \in A$ if $|x|_a = 1$ for all $x \in S$.*

All central strongly square-free strings can be constructed by the following lemma.

**Lemma 1** *If $x$ and $z$ are strongly square-free strings on an alphabet $A$ and $a \in A$ does not appear in $x$ or $z$ then $xaz$ is a strongly square-free central string.*

**Proof.** Suppose $xaz$ contains an Abelian square $yy^\sigma$, where $\sigma$ can be any permutation of $y$. Then, $yy^\sigma$ cannot be a substring of $x$ or of $z$ since they are strongly square-free. Therefore, either $a$ occurs in $y$ or $a$ occurs in $y^\sigma$. But since $y^\sigma$ is a permutation of $y$, $a$ must appear in *both* $y$ and $y^\sigma$, contradicting the assumption that $a$ appears only once in $xaz$. □

Note that if $a \in \{0, 1, \ldots, n-1\}$ does not appear in $x$ or $y$ then both $x$ and $y$ are necessarily strings over alphabets of at most $n-1$ elements.

For each permutation of the underlying alphabet, the following recursive definition constructs a different central string on an alphabet of size $n$ which is strongly square-free by Lemma 1.

**Definition 3** *For each permutation $\pi$ of $\{0, \ldots, n-1\}$ recursively define a string $z_n = z_n(\pi)$ by iterating for $k = 0, \ldots, n-1$*

$$\begin{aligned} z_1 &= \pi(0) \\ z_{k+1} &= z_k \pi(k) z_k. \end{aligned} \tag{1}$$

For example, if $n = 4$ and $\pi$ is the 4-cycle $(0123)$ then

$$z_4(\pi) = 121312101213121.$$

For $|A| = n$, Definition 3 yields $n!$ distinct strings each of length $2^n - 1$. It had been previously shown in [4] that the strings $z_n = z_n(\pi)$ are square-free.

It is the non-central strings which appear to be more difficult to construct. A string $x$ is non-central only if every letter of the alphabet $A$ appears at least twice in $x$ and, consequently, $|x| \geq 2|A|$.

For example, on $\{0, 1, 2\}$ the only non-central strings are of length 6 or 7 [5]. The non-central strings on $\{0, 1, 2\}$ of length 6 are

010212 020121 101202 121020 202101 212010

and those of length 7 are

0121012 0212021 1020102 1202120 2010201 2101210.

For a proof see [5].

## 3 Maximal Strongly Square-Free Strings

We observed earlier that 010 and 101 were binary strings that could not be extended by concatenation right or left without introducing a square. Moreover they are strongly square-free strings which cannot be extended. Another example we have already noted is $z_3(id) = 0102010$ which does not contain Abelian squares and cannot be extended over the alphabet $\{0, 1, 2\}$ without introducing Abelian squares.

**Definition 4** *A finite string $x$ over an alphabet $A$ is a* maximal *strongly square-free string if for every $a \in A$, both $ax$ and $xa$ contain Abelian squares.* Right maximal *and* left maximal *strongly square-free strings are defined in the obvious way.*

For example, 0102010 and the 5 strings obtained from it by permuting the underlying alphabet are all strongly square-free strings. Although strongly square-free implies square-free, a maximal strongly square-free string need not be a maximal square-free string. A simple example is the string 1020102 over $\{0, 1, 2\}$

For each permutation of the underlying alphabet, Definition 3 constructs a maximal strongly square-free string on an alphabet of size $n$. Although the indexing in Definition 3 might seem a bit tricky, in practice any $z_n(\pi)$ is easy to compute. For example, if $n = 4$ and $\pi$ is the identity permutation then

$$z_4(\pi) = 010201030102010.$$

For $|A| = n$, Definition 3 yields $n!$ distinct strings each of length $2^n - 1$. It had been previously shown in [4] that the strings $z_n(\pi)$ are square-free.

**Theorem 2** *The $n!$ strings $z_n = z_n(\pi)$ are maximal strongly square-free strings on $\{0, 1, \ldots n-1\}$.*

**Proof.** Using induction, Lemma 1 implies each string $z_k$ in (1) is strongly square-free because $\pi(k-1)$ cannot appear in $z_k$. The proof that each $z_n$ is maximal is also by induction, but on $n$.

If $n = 1$ then $A = \{0\}$ and $z_1 = 0$ is trivially strongly square-free.

For $n \geq 2$ observe that

$$z_n\pi(n-1) \quad = \quad z_{n-1}\pi(n-1)z_{n-1}\pi(n-1)$$

is a square. Similiarly, $\pi(n-1)z_n$ is a square.

Take $a \in \{0, \ldots, n-1\} \backslash \{\pi(n-1)\}$. Then $a = \pi(i)$ for some $i \neq n-1$ since $\pi$ is a permutation. We have seen by induction that $z_{n-1}$ is a maximal strongly square-free string and so both $az_{n-1}$ and $z_{n-1}a$ have Abelian squares. Therefore, both

$$az_n = az_{n-1}\pi(n-1)z_{n-1}$$

and

$$z_n a = z_{n-1}\pi(n-1)z_{n-1}a$$

have Abelian squares for each choice of $a \in \{0, \ldots, n-1\} \backslash \{\pi(n-1)\}$. □

Gus Simmons noted that the strings $z_n$ are all palindromes so that it is really only necessary to check that they are maximal on the left or on the right. The strings $z_n$ are also central since $\pi(n-1)$ is not repeated in the string. We conjecture that the strings $z_n(\pi)$ also have maximum length in the set of all finite strongly square-free strings over $\{0,\ldots,n-1\}$. If this is true then we are assured that there are only finitely many maximal strongly square-free strings for any alphabet $A$ with $|A|\geq 4$ even though infinite strongly square-free strings are known to exist [10]. Also we needn't look at strings of very short length to find the maximal strongly square-free ones as the following lemma shows.

**Lemma 2** *If $x=x_1\ldots x_m$ is any finite maximal strongly square-free string on an alphabet $A$, then $m\geq 2|A|-1$.*

**Proof.** Suppose that $x=x_1\ldots x_m$ is a maximal strongly square-free string. Then the strings x$a$, $a\in A$, have distinct suffixes which are Abelian squares, necessarily of even length. An easy counting argument shows that the number of possible suffixes that are Abelian squares in any string of length $m$ is given by $\lfloor\frac{m+1}{2}\rfloor$. Hence, $|A|\leq\lfloor\frac{m+1}{2}\rfloor$. That is, $m\geq 2|A|-1$. □

This not a particularly good lower bound. Even for $|A|=3$ it can be checked (see the appendix) that none of the strongly square-free strings of length 5 are maximal. A better lower bound on the lengths of maximal strongly square-free strings would speed up searching algorithms such as the one given in [6].

## 4 Further Research

We have seen that relatively little is known about strongly square-free finite strings. For those only interested in the countably infinite case, it is important to note that infinite strongly square-free strings must avoid maximal finite ones. Although the enumeration problem has been studied for square-free strings [13], we do not have good bounds for the number of strongly square-free strings of length n. While it is relatively easy to write down central strongly square-free strings, we know little about the strings which are non-extendible and non-central. Searching strings for Abelian squares is discussed in [6].

## References

[1] J. Berstel, Some recent results on squarefree words STACS'84, *Lecture Notes in Computer Science* Vol. 166, Springer-Verlag, 14–25.

[2] J. Berstel, Axel Thue's work on repetitions in words, *Publications du LCIM*, Université du Québec à Montréal, 1992, 65–80.

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

[4] L.J. Cummings, On the construction of Thue sequences, *Proc. 9 th S-E Conf. on Combinatorics, Graph Theory, and Computing*, 1978, 235–242.

[5] L.J. Cummings, Strongly Square-Free Strings on Three Letters, *The Australasian Journal of Combinatorics* **14** (1996), 259–266.

[6] L.J. Cummings and W.F. Smythe, Weak repetitions in strings, *J. Combinatorial Math. and Combinatorial Computing* **24** (1997).

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

[8] R.C. Entringer, D.E. Jackson, and J.A. Schatz, On nonrepetitive sequences, *Journal of Combinatorial Theory A*, **16** (1974), 159–164.

[9] P. Erdős, Some unsolved problems, *Hungarian Academy of Sciences Mat. Kutató Intézet Közl.* **6** (1961), 221–254.

[10] V. Keränen, Abelian squares are avoidable on 4 letters, *Lecture Notes in Computer Science*, No. **623** (1992), 41–52.

[11] J.A. Larson, R. Laver, and G.F. McNulty, Square-Free and Cube-Free Colorings of the Ordinals, *Pacific Journal of Mathematics*, **89** (1980), 137–141.

[12] M.G. Main, Permutations are not context-free: an application of the 'Interchange Lemma', *Information Processing Letters*, **15** (1982), 68–71.

[13] H. Prodinger, Non-repetitive sequences and Gray code, *Discrete Mathematics* **43** (1983), 113–116.

[14] P.A. Pleasants, Non-repetitive sequences, *Proc. Cambridge Phil. Soc.* **68** (1970), 267–274.

[15] R. Ross, K. Winklmann, Repetitive strings are not context-free, *RAIRO Informatique Théorique* **16** (1982), 191–199.

[16] A. Thue, Über unendliche Zeichenreihen, *Norske Vid. Selsk. Skr. I, Mat. Nat. Kl., Christiana*, **7** (1906), 1–22

**Strongly Square-Free Finite Strings on $\{0, 1, 2\}$**

| | | | | | |
|---|---|---|---|---|---|
| 0 | 1 | 2 | | | |
| 01 | 10 | 02 | 20 | 12 | 21 |
| 010 | 101 | 020 | 202 | 121 | 212 |
| 012 | 102 | 210 | 021 | 201 | 120 |

| | | |
|---|---|---|
| 1012 | 0102 | 0201 |
| 1021 | 0120 | 0210 |
| 2021 | 2102 | 1201 |
| 2012 | 2120 | 1210 |
| 1202 | 0212 | 0121 |
| 2101 | 2010 | 1020 |

| | | |
|---|---|---|
| 10121 | 01020 | 02010 |
| 10212 | 01202 | 02101 |
| 20121 | 21020 | 12010 |
| 20212 | 21202 | 12101 |
| 12021 | 02120 | 01210 |
| 12012 | 02102 | 01201 |
| 21021 | 20120 | 10201 |
| 21012 | 20102 | 10210 |
| 12102 | 02012 | 01021 |
| 21201 | 20210 | 10120 |

| | | | |
|---|---|---|---|
| 120121 | 120212 | 210121 | 210212 |
| 121021 | 121012 | 212012 | 212021 |
| 021020 | 021202 | 201020 | 201202 |
| 020120 | 020102 | 202102 | 202120 |
| 012010 | 012101 | 102010 | 102101 |
| 010210 | 010201 | 101201 | 101210 |
| 010212 | 020121 | 101202 | 121020 |
| 202101 | 212010 | | |

| | | | |
|---|---|---|---|
| 1210121 | 1210212 | 2120121 | 2120212 |
| 0201020 | 0201202 | 2021020 | 2021202 |
| 0102010 | 0102101 | 1012010 | 1012101 |
| 0121012 | 0212021 | 1020102 | 1202120 |
| 2010201 | 2101210 | | |