STRONGLY SQUARE-FREE STRINGS ON THREE LETTERS

L.J. Cummings
Department of Pure 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 construct all 117 strongly square-free finite strings on an alphabet of three letters.

1 Introduction

An ordered sequence $\mathbf{x}= x_1x_2\cdots x_n$ of elements chosen from a fixed finite set, A, of distinct elements is called a *string* of length $|\mathbf{x}| = n$ over the alphabet A. The empty string is denoted by λ. The ordered sequence $x_ix_{i+1}\cdots x_j$ is a *substring* of $\mathbf{x}$ if $1 \leq i \leq j \leq n$. In the interests of notational convenience and without loss of generality, we choose $A = \{0, \ldots n-1\}$ for fixed $n > 1$ as the alphabet. Every element of the alphabet is also a string. For each $a \in A$ we define a function $|\mathbf{x}|_a$ 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 $\mathbf{v}\neq \lambda$ then $\mathbf{u}$ is said to be a *proper prefix* of $\mathbf{x}$. Similiarly, if $\mathbf{u}\neq \lambda$ then $\mathbf{v}$ is said to be a *proper suffix* of $\mathbf{x}$. If S_0, S_1 are sets of strings and $\mathbf{x}$ is a fixed string then we use the notation

$$S_0\mathbf{x}S_1 = \{\mathbf{uxv} \mid \mathbf{u} \in S_0, \mathbf{v} \in S_1\}.$$

A string which consists of the concatenation of two equal substrings is called a *square*. A string without any substrings which are squares is said to be *square-free*.

For example, over the alphabet $\{0, 1\}$, 010010 is a square which contains another square 00 and it is a substring of 10100101. The strings 010 and 101 are square-free and, moreover, they cannot be extended by concatenation on either the right or the left without creating a square. Thue [14] observed that every two-sided infinite square-free string $\mathbf{x}$ over $\{0, 1, 2\}$ is a product of strings from the set:

01 012 0121 02 021 0212.

Australasian Journal of Combinatorics 14 (1996), pp. 259-266

It has been shown by Ross and Winkleman [12] that over any alphabet of at least three elements, the set of strings containing "squares" is not context-free.

An *Abelian square* is the concatenation of a string with a permutation of itself. Over the alphabet $\{0,1\}$, 010100 is an Abelian square which contains the squares 0101, 1010, and 00. Since every square is an Abelian square, 010100 contains 4 Abelian squares. More formally:

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

$$BB^{\sigma} = b_1 \cdots b_k b_{\sigma(1)} \cdots b_{\sigma(k)}$$

where σ is a permutation of $\{1,\ldots,k\}$. A string is said to be strongly square-free *if it contains no Abelian squares.*

Over the alphabet $A = \{0,1,2\}$, the string 012201 is an Abelian square but 0102010 does not contain Abelian squares and cannot be extended over the alphabet $A = \{0,1,2\}$ without introducing Abelian squares. Clearly every strongly square-free string is square-free. Main [10] 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 [7] proved that every infinite binary string has arbitrarily long Abelian squares. Dekking [2] 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.

Apparently Erdős [8] first raised the question of the minimum alphabet size over which there exist infinite strings without Abelian squares. This was a variant of the corresponding problem for squares raised and solved by Thue [13] in 1906. In 1970 Pleasants [11] showed that there existed an infinite strongly square-free string on an alphabet of 5 elements. This result was recently sharpened by Keränen [9] who showed the same was true for an alphabet of 4 elements, with a computer-aided proof. It was shown in [5] that every coordinate sequence of a binary Gray code of order n with one integer in the sequence deleted is a strongly square-free string of length $2^n - 1$. The purpose of this paper is to prove there are 117 distinct strongly square-free strings on $\{0,1,2\}$ and explicitly construct them. Accepting the result by Keränen [9], 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.

It is folklore that any strongly square-free string over $\{0,1,2\}$ has length ≤ 7 [1]. This can be established, say, by diligently constructing the tree of possible strongly square-strings starting with 0 and observing that starting with 1 or 2 would yield the same tree.

Clearly, 0,1, and 2 are strongly square-free and 01 10 02 20 12 21 are the only strings of length 2 that are strongly square-free over $\{0,1,2\}$. It is easy to write down the strongly square-free strings of length 3 as the next step, but we refrain here because they will be listed as elements of larger sets. The strongly square-free strings of each length $1,\ldots,7$ over $\{0,1,2\}$ are listed together as an appendix.

2 Central Strongly Square-Free Strings

In this section we discuss a class of strongly square-free strings which are obtained recursively from strongly square-free strings over smaller alphabets.

Definition 2 *A string is said to be* central *if it contains $a \in A$ such that $|\mathrm{x}|_a = 1$. A set of strings S is said to be* central *if there exists $a \in A$ such that $|\mathrm{x}|_a = 1$ for all* x$\in S$.

It is easy to see that the only strongly square-free strings on the binary alphabet $\{0, 1\}$ are

$$S_{01} = \{0 \;\; 1 \;\; 01 \;\; 10 \;\; 010 \;\; 101\}.$$

Note that each of these strings is central. In the same way, we get complete sets of strongly central square-free strings:

$$S_{02} = \{0 \;\; 2 \;\; 02 \;\; 20 \;\; 020 \;\; 202\}$$

on $\{0, 2\}$ and

$$S_{12} = \{1 \;\; 2 \;\; 12 \;\; 21 \;\; 121 \;\; 212\}$$

on $\{1, 2\}$.

Lemma 1 will show the sets

$$S_{12}0S_{12} \quad S_{02}1S_{02} \quad S_{01}2S_{01}$$

are strongly square-free strings. All the strings in these sets have length at least 3 and each set contains 36 strings. However, there are strings that appear in more than one of the sets. In Lemma 2, we determine the intersections of these sets. Further, all central strings with length ≥ 3 over the alphabet $\{0, 1, 2\}$ are in one of these sets. The following lemma yields a general technique for constructing central strongly square-free strings on arbitrary alphabets.

Lemma 1 *If ω and ω' are strongly square-free strings over an alphabet A and $a \in A$ does not appear in ω or ω' then $\omega a \omega'$ is strongly square-free.*

Proof. Suppose $\omega a \omega'$ contains an Abelian square BB^σ, where σ can be any permutation of B. Then, BB^σ cannot be a substring of ω or ω' since they are strongly square-free. Therefore, either a occurs in B or a occurs in B^σ. But since B^σ is a permutation of B, a must appear in *both* B and B^σ, contradicting the assumption that a appears only once in $\omega a \omega'$. □

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

Lemma 2

$$\begin{aligned}
S_{12}0S_{12} \cap S_{02}1S_{02} &= \{2012 \;\; 2102 \;\; 20212 \;\; 21202\} \\
S_{12}0S_{12} \cap S_{01}2S_{01} &= \{1201 \;\; 1021 \;\; 12101 \;\; 10121\} \\
S_{02}1S_{02} \cap S_{01}2S_{01} &= \{0210 \;\; 0120 \;\; 02010 \;\; 01020\}.
\end{aligned}$$

Proof. The strings in these intersections are all strongly square-free by Lemma 1. Let $\mathbf{x} \in S_{12}0S_{12} \cap S_{02}1S_{02}$. Then

$$\mathbf{x} = \mathbf{u}0\mathbf{v} = \mathbf{y}1\mathbf{z} \tag{1}$$

for some strings $\mathbf{u},\mathbf{v} \in S_{12}$ and $\mathbf{y},\mathbf{z} \in S_{02}$. From (1), either $\mathbf{u}=\mathbf{y}$, or $\mathbf{u}$ is a prefix of $\mathbf{y}$, or $\mathbf{y}$ is a prefix of $\mathbf{u}$. But if $\mathbf{u}=\mathbf{y}$ then $0 = 1$, which is impossible.

If $\mathbf{u}$ is a proper prefix of $\mathbf{y}$ then $\mathbf{u}$ also is in S_{02} because it is easy to see that S_{02} is closed under the taking of prefixes. But $S_{12} \cap S_{02} = \{2\}$. Therefore, $\mathbf{u}= 2$ and so $\mathbf{y}$ can only be one of 20 or 202. If $\mathbf{y}= 20$ then (1) becomes $\mathbf{x}= 20\mathbf{v}= 201\mathbf{z}$ which implies that $\mathbf{v}= 1\mathbf{z}$. In the same way, if $\mathbf{y}= 202$ then (1) becomes $\mathbf{x}= 20\mathbf{v}= 2021\mathbf{z}$ which implies that $\mathbf{v}= 21\mathbf{z}$. In either case, $\mathbf{z}$ is a proper suffix of $\mathbf{v}$. Since S_{12} is closed under the taking of suffixes we conclude that $\mathbf{z} \in S_{12} \cap S_{02} = \{2\}$. Therefore, $\mathbf{z}= 2$ and so $\mathbf{v}$ can only be one of 12 or 212. It follows that

$$S_{12}0S_{12} \cap S_{02}1S_{02} = \{2012 \quad 20212\}.$$

in this case.

If $\mathbf{y}$ is a proper prefix of $\mathbf{u}$ then a similiar argument shows that in this case

$$S_{12}0S_{12} \cap S_{02}1S_{02} = \{2102 \quad 21202\}.$$

We conclude that

$$S_{12}0S_{12} \cap S_{02}1S_{02} = \{2012 \quad 2102 \quad 20212 \quad 21202\}.$$

In the same way we establish that

$$S_{12}0S_{12} \cap S_{01}2S_{01} = \{1201 \quad 1021 \quad 12101 \quad 10121\}$$

and

$$S_{02}1S_{02} \cap S_{01}2S_{01} = \{0210 \quad 0120 \quad 02010 \quad 01020\}.$$

This completes the proof. □

Note that necessarily

$$S_{12}0S_{12} \cap S_{02}1S_{02} \cap S_{01}2S_{01} = \lambda.$$

Hence the union of these three sets contains exactly 96 distinct strings.

3 Non-Central Strongly Square-Free Strings

If $\mathbf{x}$ is any non-central strongly square-free string then $|\mathbf{x}|_a \geq 2$ for all $a \in A$. Further, for any non-central strongly square-free string $\mathbf{x}$, $|\mathbf{x}| \geq 2|A|$. In particular, if $\mathbf{x}$ is a non-central strongly square-free string over $\{0, 1, 2\}$ then $|\mathbf{x}| \geq 6$.

Lemma 3 *There are* 6 *non-central strongly square-free strings of length* 6 *over* $\{0, 1, 2\}$.

Proof. If $|\mathbf{x}| = 6$ and $\mathbf{x}$ is non-central then $\mathbf{x}$ is a permutation of 001122 because every letter of $\{0,1,2\}$ must appear at least twice. Let $\mathbf{x}= x_1 \cdots x_6$. Suppose $x_1 = 0$. Then, $x_2 \neq 0$ and there is exactly one other occurrence of 0 in $\mathbf{x}$. Either $\mathbf{x}$ has the form $0\mathbf{t}0$ or the form $0\mathbf{t}_1 0\mathbf{t}_2$ for nonempty substrings $\mathbf{t}, \mathbf{t}_1, \mathbf{t}_2$. But if $\mathbf{x} = 0\mathbf{t}0$ then $\mathbf{t} \in S_{12}$ has length 4 and necessarily contains Abelian squares and so $\mathbf{x}$ must as well. Therefore, $\mathbf{x}= 0\mathbf{t}_1 0\mathbf{t}_2$ where $\mathbf{t}_1, \mathbf{t}_2 \in S_{12}$. Further, $|\mathbf{t}_1| + |\mathbf{t}_2| = 4$ and since both $\mathbf{t}_1$ and $\mathbf{t}_2$ are square-free, $|\mathbf{t}_1|, |\mathbf{t}_2| \leq 3$.

If $|\mathbf{t}_1| = 1$ let $|\mathbf{t}_2| = a \in \{1,2\}$. Then $\mathbf{x}= 0a0\mathbf{t}_2$ is strongly square-free. It follows that $\mathbf{t}_2 = bab \in S_{1,2}$, where $b \neq a$. We conclude that $\mathbf{x}=$ 010212 and $\mathbf{x}=$ 020121 are the only non-central strongly square-free strings in this case.

If $|\mathbf{t}_1| = 2$ then $|\mathbf{t}_2| = 2$. But if $\mathbf{t}_1 = ab \in S_{1,2}$ then $\mathbf{t}_2 = ab$ or $ba \in S_{1,2}$. In either case $\mathbf{x}$ is an Abelian square.

If $|\mathbf{t}_1| = 3$ then $|\mathbf{t}_2| = 1$ and $\mathbf{t}_1 = aba$ or $bab \in S_{1,2}$. We see that $\mathbf{x}$ is either an Abelian square, or is central contrary to hypothesis.

We conclude that $\mathbf{x}=$ 010212 and $\mathbf{x}=$ 020121 are the only non-central strongly square-free strings beginning with 0.

Making the same arguments for $\mathbf{x}$ beginning with 1 we further obtain

$$101202 \quad 121020.$$

And if $\mathbf{x}$ begins with 2, the only possibilities are

$$202101 \quad 212010.$$

Therefore, there are only 6 strongly square-free strings of length 6 over $\{0,1,2\}$. □

Lemma 4 *There are 6 non-central strongly square-free strings of length 7 over* $\{0,1,2\}$.

Proof. If $\mathbf{x}$ is any strongly square-free non-central string with $|\mathbf{x}| = 7$ then $\mathbf{x}$ has one entry which appears exactly three times, since every entry must appear at least twice. Assume that $x_1 = 0$. Either 0 appears twice or three times in $\mathbf{x}$. First suppose 0 appears three times. Then there are two cases: either $\mathbf{x}$ has the form $\mathbf{x} = 0\mathbf{t}_1 0\mathbf{t}_2 0$ or $\mathbf{x}= 0\mathbf{t}_1 0\mathbf{t}_2 0\mathbf{t}_3$ where the strongly square-free substrings $\mathbf{t}_1, \mathbf{t}_2, \mathbf{t}_3$ are in S_{12}. If $\mathbf{x}= 0\mathbf{t}_1 0\mathbf{t}_2 0$, then $|\mathbf{t}_1| + |\mathbf{t}_2| = 4$ while $|\mathbf{t}_1|, |\mathbf{t}_2| \leq 3$. If $|\mathbf{t}_1| = 1$ so $\mathbf{t}_1 = a \in S_{12}$ then, as in the case of Lemma 3, we argue that $\mathbf{x}= 0a0bab0$ which has an Abelian square suffix. If $|\mathbf{t}_1| = 2$ then $\mathbf{t}_1 = ab \in S_{12}$ and $\mathbf{x}= 0ab0ba0$ which contains Abelian squares. The same is true in the last case: If $|\mathbf{t}_1| = 3$ then $|\mathbf{t}_2| = 1$ and all possibilities contain Abelian squares.

If $\mathbf{x} = 0\mathbf{t}_1 0\mathbf{t}_2 0\mathbf{t}_3$ then $|\mathbf{t}_1| + |\mathbf{t}_2| + |\mathbf{t}_3| = 4$ while $1 \leq |\mathbf{t}_1|, |\mathbf{t}_2|, |\mathbf{t}_3| \leq 3$. The only possibility is that exactly one of the substrings has length 2 and the others length 1. First suppose that $|\mathbf{t}_1| = 2, |\mathbf{t}_2| = 1, |\mathbf{t}_3| = 1$. It follows that $\mathbf{x}= 0ab0a0b$ or $\mathbf{x}= 0ab0b0a$. But both of these contain Abelian squares. The second case is only somewhat different. If $|\mathbf{t}_1| = 1, |\mathbf{t}_2| = 2, |\mathbf{t}_3| = 1$, then $\mathbf{t}_2 = ab$ implies $\mathbf{t}_1 = b$ since otherwise $\mathbf{x}$ would have a square prefix $0a0a$. Thus, either $\mathbf{x}= 0b0ab0a$ or $\mathbf{x}= 0b0ab0b$.

But in the first case, $\mathbf{x}$ contains the square $b0a$ as suffix and in the second case $\mathbf{x}$ would be central, contrary to hypothesis. The last case $|\mathbf{t}_1| = 1, |\mathbf{t}_2| = 1, |\mathbf{t}_3| = 2$ is like the first since all possible strings contain Abelian squares.

Now suppose that 0 appears only twice in $\mathbf{x}$. Then $\mathbf{x}= 0t0$ or $\mathbf{x} = 0t_10t_2$. The first case is impossible since t would belong to S_{12} while $|t| = 5$. If $\mathbf{x} = 0t_10t_2$ then $|\mathbf{t}_1| + |\mathbf{t}_2| = 5$ while $0 \leq |\mathbf{t}_1|, |\mathbf{t}_2| \leq 3$. Consequently, there are only two cases to consider.

If $|\mathbf{t}_1| = 3$ then $\mathbf{t}_1 = aba$ and $\mathbf{t}_1 = ab$ or ba for $a \neq b \in \{1,2\}$. Thus, $\mathbf{x}= 0aba0ab$ or $\mathbf{x}= 0aba0ba$. In the later case $\mathbf{x}$ would have an Abelian square prefix. In the former case, we have only the two strings:

$$0121012 \quad 0212021$$

If $|\mathbf{t}_1| = 2$ then $|\mathbf{t}_2| = 3$ and it is easy to check that all possbilities contain Abelian squares.

We conclude that 0121012 and 0212021 are the only non-central strings starting with 0 of length 7 that are strongly square-free.

In the same way, if $\mathbf{x}$ begins with 1 we obtain strongly square-free strings 1020102 and 1202120. If $\mathbf{x}$ begins with 2 we further obtain strongly square-free strings 2010201 and 2101210. □

Theorem 1 *There are* 117 *distinct strongly square-free strings on any alphabet of three letters.*

Proof. Each of the 3 sets,

$$S_{12}0S_{12} \quad S_{02}1S_{02} \quad S_{01}2S_{01}$$

contain 36 strongly square-free strings of length k, $3 \leq k \leq 7$ and intersect pair-wise in 4 strings. Since there are no strings in the intersection of all 3 sets, it follows that there are 96 distinct strings in the union of these sets. There do not exist non-central strings of length $k \leq 6$ and Lemmas 3 and 4 yield 12 non-central strings of lengths 6 and 7. Finally, since the alphabet provides 3 strongly square-free strings and there are 6 of length 2 we conclude that there are exactly 117 strongly square-free strings on three letters. □

4 Generalizations

The case of just three letters is special because, as already remarked, it is the last case in which all strongly square-free strings are finite. Lemma 1 shows that there will exist finite strongly square-free strings over any finite alphabet. If empty strings are taken into account, then Lemma 1 shows that there are finite strongly square-free strings of every length $n \geq 1$ and gives a recursive way to construct many of them. In particular, it shows that there are at least 13,689 central strongly square-free finite strings over an alphabet of four letters. Obviously, central strings of all orders can be constructed, but the analysis for non-central strongly square-free strings is not yet clear.

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 01020
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

References

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

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

[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, *Gray Codes and Strongly Square-Free Strings*, R. Capocelli, A. de Santis, U. Vacaro, eds., Sequences II, Springer- Verlag, 1993, 439-446.

[6] L.J. Cummings and W.F. Smyth, *Weak repetitions in strings*, J. Combinatorial Math. and Combinatorial Computing, (to appear).

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

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

[9] V. Keränen, *Abelian squares are avoidable on 4 letters*, Lecture Notes in Computer Science, Vol. 623, Springer-Verlag, 1992, 41-52.

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

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

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

[13] A. Thue, *Über unendliche Zeichenreichen*, Kra. Vid. Selsk. Skr. I, Mat. Nat. Kl., Christiana, 1906, 1-22.

[14] A. Thue, *Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen*, Kra. Vid. Selsk. Skr. I, Mat. Nat. Kl., Christiana, 1912, 1-67.

(Received 29/1/96)