We use the usual notation for combinatorics on words. A word of $n$ elements is $w = w[1..n]$, with $w[i]$ being the $i$th element and $w[i..j]$ the factor of elements from position $i$ to position $j$. If $i = 1$ then the factor is a prefix and if $j = n$ then it is a suffix. The letters in $w$ come from some alphabet $A$. The set of all finite words with letters from $A$ is $A^*$. The length of $w$, written $|w|$, is the number of occurrences of letters in $w$ and the number of occurrences of the letter $a$ in $w$ is $|w|_a$. If $w$ is a word whose letters come from an alphabet $a_1, a_2, \ldots, a_t$ then the Parikh vector of $w$, written $\mathcal{P}(w)$, is the vector $[|w|_{a_1}, |w|_{a_2}, \ldots, |w|_{a_t}]$ An abelian square is a factor $uv$ in which $\mathcal{P}(u) = \mathcal{P}(v)$. We say that abelian squares $u_1v_1$ and $u_2v_2$ are equivalent if $\mathcal{P}(u_1) = \mathcal{P}(u_2)$. Thus $abab$ and $abba$ are equivalent abelian squares but $abba$ and $aaaa$ are non-equivalent abelian squares. We are interested in the maximum and minimum number of abelian squares that can occur in a word of length $n$. This depends on the size of the alphabet we use (in what follows the alphabet will usually be binary) and how the squares are counted. We can count the total number of abelian squares, the number of distinct abelian squares or the number of non-equivalent abelian squares. This gives us many problems to consider. We further expand the set of problems by considering the same set of extremal problems but applied to circular words rather than linear words. For circular words we do not consider abelian squares whose length is greater than the word length, that is, we don’t allow abelian squares to overlap themselves. Changing from linear to circular words can simplify problems as we are avoiding messiness at the ends of words. For example the minimum number of non-equivalent abelian squares in a binary circular word was established with a short straight-forward proof in [9], but the same question for linear words has not been settled and appears to be difficult. On the other hand the maximum possible number of distinct non-empty palindromes in a word of length $n$ is
easily shown to be \( n \) but the equivalent problem for circular words is more difficult [19]. To summarise we are looking at the

\[
\begin{aligned}
\{ & \text{maximum} \} \quad \{ & \text{total number} \} \quad \{ & \text{number of distinct} \} \\
\{ & \text{minimum} \} & \quad \{ & \text{number of non-equivalent} \} & \quad \{ & \text{abelian squares in a} \} \\
& \{ & \text{linear} \} & \quad & \{ & \text{circular} \}
\end{aligned}
\]

word of length \( n \) on an alphabet of size \( t \).

This gives 12 problems with possible sub-problems depending on alphabet size. It will be convenient to adopt the following shorthand for these problems. The required bounds for a binary word of length \( n \) are written \( C_1C_2C_3(n) \) where \( C_1 \) is either \( X \) or \( M \) for maximum and minimum respectively, \( C_2 \) is \( T \), \( D \) or \( N \) for total, distinct and non-equivalent and \( C_3 \) is either \( L \) or \( C \) for linear and circular. Bounds for words on larger alphabets will be treated separately. We have

\[
XC_2C_3(n) \geq MC_2C_3(n)
\]

since maximums are at least as large as minimums,

\[
C_1TC_3(n) \geq C_1DC_3(n) \geq C_1NC_3,
\]

which follows immediately from the meanings of \( T \), \( D \) and \( N \), and

\[
C_1C_2C(n) \geq C_1C_2L(n)
\]

since a circular word contains the same abelian squares as the corresponding linear word plus any abelian squares that straddle the ends of the word.

**The maximum number of abelian squares in a linear word**

The maximum total number of abelian squares in a linear word of length \( n \) is \( \lfloor n/2 \rfloor \lceil n/2 \rceil \) and is attained by \( a^n \) so we have

\[
XTL(n) = \lfloor n/2 \rfloor \lceil n/2 \rceil.
\]

For circular words we have

\[
XTC(n) = n\lfloor n/2 \rfloor
\]

which is also attained by \( a^n \).

For the other two ways of counting abelian squares we will only consider the case of binary words. I’m not aware of any work done on larger alphabets and it’s likely that increasing the alphabet size will not increase the maximum number of abelian squares. In fact Fici and Mignosi [6] make the following conjecture.
Conjecture 0.1. If a word of length $n$ contains $k$ many distinct abelian square factors, then there exists a binary word of length $n$ containing at least $k$ many distinct abelian square factors.

Tables 1 and 2 give values of $XDL(n)$ and $XNL(n)$ for small values of $n$.

| $n$ | $XDL(n)$ | Example attaining bound |
|-----|----------|-------------------------|
| 1   | 0        | a                       |
| 2   | 1        | aa                      |
| 3   | 1        | aaa                     |
| 4   | 2        | aaaa                    |
| 5   | 3        | aabba                   |
| 6   | 4        | aababa                  |
| 7   | 5        | aababaa                 |
| 8   | 7        | aabbaabb                |
| 9   | 9        | aaaaabbaaa              |
| 10  | 11       | aaaaabbaaaa             |
| 11  | 13       | aaaaabbaaabb            |
| 12  | 15       | aaaaabbaaaabb           |
| 13  | 17       | aaaaabbaaaabb           |
| 14  | 21       | aaaaaabbaaabb           |
| 15  | 23       | aaaaabbaabbaaa          |
| 16  | 26       | aaaaabaaabaabaa         |
| 17  | 30       | aaaaabaaabaabaa         |
| 18  | 34       | aababaaabaabaaaabb      |
| 19  | 38       | aababaabaabaabaaaabba   |
| 20  | 43       | aababaabaabaaabbaabaa   |

Table 1. Maximum numbers of distinct abelian squares in binary words ($XDL(n)$). The second column is sequence A262249 in the Online Encyclopedia of Integer Sequences.

It is clear that both $XDL(n)$ and $XNL(n)$ are bounded above by $n^2$ since there cannot be more abelian squares in a word than there are factors. In [13] Kociuska et al. show that $XDL(n) = \Theta(n^2)$ by considering the word $a^kba^kbda^{2k}$, which is easily shown to contain a quadratic number of distinct abelian squares. They mention that Gabriel Fici had previously obtained the same result by a different method, and communicated this information to them. The number of abelian squares
Table 2. Maximum numbers of nonequivalent abelian squares in binary words ($XNL(n)$). The second column is sequence A262265 in the Online Encyclopedia of Integer Sequences.

| $n$ | $XNL(n)$ | Example attaining bound |
|-----|----------|-------------------------|
| 1   | 0        | a                       |
| 2   | 1        | aa                      |
| 3   | 1        | aaa                     |
| 4   | 2        | aaaa                    |
| 5   | 3        | aabbaa                  |
| 6   | 4        | aabbaa                  |
| 7   | 4        | aaaaabba                |
| 8   | 6        | aabbaabb                |
| 9   | 6        | aaaabbaaa               |
| 10  | 7        | aaaaabbaaaa             |
| 11  | 8        | aaaaabbaaabb            |
| 12  | 10       | aaaaabbaaaaabb          |
| 13  | 10       | aaaaabbaaaaabb          |
| 14  | 11       | aaaaaabbaaaaabb         |
| 15  | 12       | aaaaaabbaaaaabba        |
| 16  | 15       | aaaaaabbaaababbba       |
| 17  | 16       | aaaaaabbaaababbba       |
| 18  | 17       | aaaaaabbaaababbbaa      |
| 19  | 17       | aaaaaabbaaababbbaa      |
| 20  | 19       | aaaaaabbaaabaaabbbabb   |

In this word gives a lower bound on $XDL(4k + 2)$ of

$$k + (\lfloor \frac{k}{2} \rfloor + 1)^2 + (k + 1)\lceil \frac{k}{2} \rceil.$$ 

Although this is quadratic in $n = 4k + 2$ it is a fair bit smaller than the values in Table 1, for example with $k = 4$ we have $n = 18$ and the formula above gives a lower bound on $XDL(22)$ of 23 compared with the actual value of 34.

In the case of non-equivalent abelian squares the Kucherov et al. \cite{13} show that

$$XDL(n) = \Omega(n^{1.5n}/\log n)$$

by considering words $w_k$ defined as follows:

$$w_1 = ab$$
$$w_k = w_{k-1}a^kb^k \text{ for } k > 1.$$
In this case the proof is distinctly non-trivial\footnote{Professor Rytter has informed me that this lower bound has been improved to $\Omega(n^{1.5})$ using a different set of words.}. The number of abelian squares in $w_k$ is difficult to estimate. As with the distinct abelian squares case the lower bound obtained in this way is not sharp. For example $w_4$ has length 20 and contains 13 non-equivalent abelian squares, compared with the actual value $XNL(20) = 19$. The authors of \cite{13} make the following conjecture:

**Conjecture 0.2.**

$$XNL(n) = O(n^{1.5}).$$

In a slightly different direction Fici and Mignosi \cite{6} consider infinite words in which the number of distinct abelian square factors of length $n$ grows quadratically with $n$ though they replace this condition with the following slightly weaker one. Write $p_w(n)$ for the factor complexity of $w$, $a(w)$ the set of abelian squares in $w$ and $F_w(n)$ the set of length $n$ factors of $w$, so that $p_w(n) = |F_w(n)|$. They say an infinite word $w$ is *uniformly abelian-square rich* if there exists a positive constant $C$ such that

$$\inf_{v \in F_w(n)} |a(v)| \geq Cn^2.$$  

They show that the Thue-Morse word and certain Sturmian words are uniformly abelian-square rich.

The minimum numbers of distinct and non-equivalent abelian squares in words

The situation here parallels that for non-abelian squares.

| Alphabet size | Non-abelian | Abelian |
|---------------|-------------|---------|
| 2             | $\leq 3$ distinct squares | Conjectured $\lfloor n/4 \rfloor$ distinct/non-equivalent |
| 3             | 0           | Conjectured $\leq 3$ distinct/non-equivalent |
| 4             | 0           | 0       |

**Table 3.** Minimum numbers of distinct/non-equivalent non-abelian/abelian squares in a word of length $n$ for different alphabet sizes.
Keränen [12] used an 85-uniform morphism on a four letter alphabet to construct such an infinite word. As far as I know nobody has used a shorter morphism to produce such words. Currie and Fitzpatrick [7] showed that there exist binary non-abelian cube-free circular words of every length.

For a three letter alphabet Mäkelä [15] has made the following conjecture.

**Conjecture 0.3.** There exist infinite ternary words containing no abelian squares of length greater than two (which implies there are at most three distinct abelian squares).

This would be analogous to the non-abelian case where it’s know that there exist infinite binary words containing only the squares $aa$, $bb$ and $abab$. This was first proved by Fraenkel and Simpson [8] in 1995, with much nicer proofs coming later from Rampersad, Shallit and Wang [17] in 2005, Harju and Nowotka [10] in 2006 and Badkobeh and Crochemore [1] in 2011. The longest binary words containing only zero, one and two squares have lengths, respectively, 3 (for example $aba$), 7 ($aaabaaa$) and 18 ($abaabaaabbaabbab$). Mäkelä’s conjecture has been supported by Rampersad, see [7], who has produced a ternary word of length 3160 wherein the only abelian squares are $aa$, $bb$ and $cc$ (each occurring many times). In the appendix there is a word of 2034 letters containing no abelian squares other than 00, 11 and 22. This was kindly provided by Narad Rampersad. The longest ternary words containing 0, 1 and 2 distinct abelian squares have lengths 7, 18 and 63 respectively. Examples of words attaining these bounds are

- $abacaba$
- $abebabccaccccbabceba$
- $abbbccbbacbcceccacccabbbaacbabacccabbcacccbcbabccbccbcbacccbaabbbcccbacccbacacaacccbacccbacacbababccbacccbacacaacacbacacbababccbacccbacacaacca.$

The second and third examples are unique up to permutation of the alphabet.

For a binary alphabet we must distinguish between distinct and non-equivalent abelian squares. Recall that that $MDL(n)$ is the minimum number of distinct abelian squares in a linear binary word of length $n$ and $MNL(n)$ is the minimum number of non-equivalent abelian squares.

For the non-equivalent case Fraenkel, Paterson and Simpson [9] made the following conjecture.

**Conjecture 0.4.** For all positive integers $n$ $MNL(n) = \lfloor n/4 \rfloor$ and, when $n = 4k + 3$ the bound is only obtained by the words $a^{2k+1}ba^{2k+1}$.
and \((ab)^{2k+1}a\) and their complements. If \(n\) is not congruent to 3 modulo 4 just remove 1, 2 or 3 letters from an end of the word.

In [9] Fraenkel, Paterson and Simpson proved that a circular binary word of length \(n = 2k + 2\) contains at least \(k\) non-equivalent abelian squares and this bound is attained only be \((ab)^{k+1}\), its complement and their conjugates. Thus,

\[
MNC(2k + 2) = k.
\]

In [7] Fici and Saarela made the following conjecture.

**Conjecture 0.5.** \(MDL(n) = \lfloor n/4 \rfloor\) and the only such words of length \(4k + 3\) containing only \(k\) distinct abelian squares are \(a^{2k+1}ba^{2k+1}\) and its complement.

Since \(MDL(n) \geq MNL(n)\) a proof of the Fraenkel Paterson Simpson conjecture would imply the truth of the Fici Saarela conjecture.

In the case of distinct abelian squares in circular binary words we have the following conjecture suggested by computer experiments.

**Conjecture 0.6.** The minimum number of distinct abelian squares in a circular word of length \(n\) is:

(a) \((n - 1)/2\) if \(n\) is odd and this bound is attained only by \(a^n, a^{n-1}b\) and their complements and conjugates.

(b) \((n - 2)/2\) if \(n\) is even and this bound is attained only by \(a^k b^{n-k}\) and its complement and their conjugates, where \(k \in \{1, 3, 5, \ldots, n - 1\}\).

The minimum total number of abelian squares in binary words

Now we are counting the abelian squares in a word according to multiplicity, so that \(ababbaaab\) contains a total of 7 abelian square occurrences: three copies of \(aa\) and one each of \(bb, abab, abba\) and \(baab\). This is the minimum number for binary words of length 11, so \(NTL(11) = 7\). I’m not aware of anybody studying the following question.

**Question 0.7.** What is the minimum number of occurrences of abelian squares in a binary word of length \(n\)?

The non-abelian case has been studied by Kucherov, Ochem and Rao [14]. They set \(m(n)\) to be the minimum number of square occurrences in a word of length \(n\). They showed that the sequence \(m(n)/n\) is convergent and called the limit \(\mathcal{M}\). They exhibited an infinite word that had a density of square occurrences of 103/187, thereby showing
that \( M \leq 103/187 = 0.550802 \ldots \) They further showed, using ideas from [20], that \( M > 0.5508 \). It is remarkable that their upper and lower bounds are so close.

The following table shows the minimum number of abelian square occurrences in binary words of length \( n \) for low values of \( n \).

The questions discussed in this essay can be generalised in several ways and such questions have a considerable literature. Most obviously we could consider other powers than 2. For example Dekking [3] showed that you can avoid abelian cubes with a ternary alphabet and abelian fourth powers with a binary alphabet. or abelian fractional powers. Another variation on the theme uses the notion of \( k \)-equivalence introduced by Karhumäki. Let \(|x|_u\) be the number of times the factor \( u \) occurs in the word \( x \). Two words \( x \) and \( y \) are \( k \)-equivalent if \(|x|_u = |y|_u\) for all factors \( u \) of length at most \( k \). We now say that \( xy \) is a \( k \)-abelian square if \( x \) and \( y \) are \( k \)-equivalent. An ordinary abelian square is therefore a \( k \)-abelian square, and a word being a \( k \)-abelian square is, in general, a stronger condition than it being and abelian square and a weaker condition than it being an ordinary square. One defines \( k \)-abalian powers in analagously. One can avoid 3-abelian squares with a ternary alphabet and 2-abelian cubes with a binary alphabet []
Table 4. Minimum numbers of abelian square occurrences in binary words. Note that, up to complementation and reversal the words of length 12 and 17 are unique. The second column is sequence A268084 in the Online Encyclopedia of Integer Sequences.

References

[1] Badkobeh, Golnaz, and Maxime Crochemore. Finite-Repetition threshold for infinite ternary words. In WORDS, vol. 63, pp. 37-43. 2011.
[2] Currie, James D., and D. Sean Fitzpatrick. Circular words avoiding patterns. In Developments in Language Theory, pp. 319-325. Springer Berlin Heidelberg, 2003.
[3] Dekking, Frederik Michel. Strongly non-repetitive sequences and progression-free sets. Journal of Combinatorial Theory, Series A 27, no. 2 (1979): 181-185.
[4] Erdős, Paul. Some unsolved problems. Magyar Tud. Akad. Mat. Kutat Int. Kzl 6 (1961): 221-254.
[5] Evdokimov, A. A. Strongly asymmetric sequences generated by a finite number of symbols Dokl.Akad. Nauk SSSR, 179 (1968): 1268-1271.
[6] Fici, Gabriele, and Filippo Mignosi. Words with the Maximum Number of Abelian Squares. In Combinatorics on Words, pp. 122-134. Springer International Publishing, 2015.

[7] Fici, Gabriel and Aleksi Saarela. On The Minimum Number of Abelian Squares in a Word Combinatorics and Algorithmics of Strings, Dagstuhl Reports, 4(2014), Issue 3, 34-35

[8] Fraenkel, Aviezri S., and R. Jamie Simpson. How many squares must a binary sequence contain. Electronic J. Combinatorics 2, no. 9 (1995).

[9] Fraenkel, Aviezri S., Jamie Simpson, and Mike Paterson. On weak circular squares in binary words. In Combinatorial Pattern Matching, pp. 76-82. Springer Berlin Heidelberg, 1997.

[10] Harju, Tero, and Dirk Nowotka. Binary words with few squares. Bulletin of the EATCS 89 (2006): 164-166.

[11] Karhumaki, Juhani, Aleksi Saarela, and Luca Q. Zamboni. On a generalization of Abelian equivalence and complexity of infinite words. Journal of Combinatorial Theory, Series A 120, no. 8 (2013): 2189-2206.

[12] Keränen, Veikko. Abelian squares are avoidable on 4 letters. In Automata, languages and programming, pp. 41-52. Springer Berlin Heidelberg, 1992.

[13] Kociumaka, Tomasz, Jakub Radoszewski, Wojciech Rytter, and Tomasz Wale. Maximum number of distinct and nonequivalent nonstandard squares in a word. In Developments in Language Theory, pp. 215-226. Springer International Publishing, 2014.

[14] Kucherov, Gregory, Pascal Ochem, and Michal Rao. How many square occurrences must a binary sequence contain?. JOURNAL OF COMBINATORICS 10.1 (2003): R12.

[15] Mäkelä, S. Patterns in words. Master’s thesis, University of Turku, 2002.

[16] Pleasants, Peter A. B. Non-repetitive sequences. Mathematical Proceedings of the Cambridge Philosophical Society. Vol. 68. No. 02. Cambridge University Press, 1970.

[17] Rampersad, Narad, Jeffrey Shallit, and Ming-wei Wang. Avoiding large squares in infinite binary words. Theoretical Computer Science 339, no. 1 (2005): 19-34.

[18] M. Rao, On some generalizations of abelian power avoidability, manuscript, 2014.

[19] Simpson, Jamie. Palindromes in circular words. Theoretical Computer Science 550 (2014): 66-78.

[20] Taramkinov, Yuriy. The minimal density of a letter in an infinite ternary square-free word is 0.2746. J. Integer Sequences 5.2 (2002).
Appendix: a ternary word containing no abelian squares other than 00, 11 and 22

00010002000111000221000111022000111100220001002220001112000222
00221222000222110111200011220001002001110001122001112110200100222112
0001220221011221221000112101112221000222112220221101112221100
012002220001120222110112220221222000100212202221100212200001002
000111000221110220001220222101112100011220001002011110001122101112
11100011200201220122200022220221011121000112200012220221100020001
0022110111211100112200110022201211100200010022100222000100200011
10022200211100112220221011122210200010002122022200222202210111220
222212220001002221222021100122001111222022100200221100211122110111
122010020001122211111220021100012200002220112220212000111002221122011
22022212200021112220010022001212202221112220222002222011022212211000
120222122200010022212200112100112221110021222000100211000120222122
20012220022221100200010002220001112220001000220000111000221011121
02211000200102220022111000112220221011122202100021000211122200222002
001112202221011121000221100011220011220001021122011222000222
1222022222110120221222000100020111000112200012221102221011122111
00011122000100222122011021100022001000222000111000201121110001222
111011211000112200010002000112222102211022211122201022211011221000222
11220122200010022111020102211222022122200100020111200021220011122
2110200011121011222110021220211122211000121002221222022211122201
110222112221102220001222011220002220012220011222001122202122211011221000222
00111200022110212202220221101112221100022001111002221011121020001002220
0011100222111201122200220011222100222001000200011220001000222000111022122
111021122100201112221020001222011011220001110222122200111102221122200111022122
2000222110001121011222022122001022211101121110020001122002122001122011220020001
222110002001112210002222122202211022210222011222021101112220011220000110
0221022200022111222110111022010001122011221012200011222102221101122200100222001
002211000112102220011100221002221122022212200100221122022211100012202221222202
211100011211011220121100022200120001122101112100011220001002200122200010222011111