More Results on Overlapping Squares[*]

Frantisek Franek[1], Robert C. G. Fuller[1], Jamie Simpson[2,3], and W. F. Smyth[1,3]

[1] Algorithms Research Group, Department of Computing & Software
McMaster University, Hamilton, Ontario, Canada L8S 4K1
`{franek,fullerrc,smyth}@mcmaster.ca`

[2] Department of Mathematics & Statistics
Curtin University, GPO Box U1987, Perth WA 6845, Australia
`simpson@maths.curtin.edu.au`

[3] Centre for Stringology & Applications
Digital Ecosystems & Business Intelligence Institute
Curtin University, GPO Box U1987, Perth WA 6845, Australia

Abstract. Three recent papers [5, 11, 8] have considered in complementary ways the combinatorial consequences of assuming that three squares overlap in a string. In this paper we provide a unifying framework for these results: we show that in 12 of 14 subcases that arise the postulated occurrence of three neighbouring squares forces a breakdown into highly periodic behaviour, thus essentially trivial and easily recognizable. In particular, we provide a proof of Subcase 4 for the first time, and we simplify and refine the previously established results for Subcases 11–14.

1 Introduction

In this paper we build on the results of three recent papers [5, 11, 8] to extend the combinatorial analysis of overlapping squares occurring at the same or neighbouring positions in a given string. In order to reduce proliferation of notation, we adopt throughout the convention that a string denoted $\boldsymbol{x}$ (in **mathbold**) has length x (regular *mathmode*). For a complete description of the background and motivation of this research, see [8]. The combinatorial problem considered in this paper was first discussed in [4], though in a much simplified form ($k = 0$).

In [5, 8] the following problem was considered:

> **(P)** Suppose that a string $\boldsymbol{x}$ has prefixes $\boldsymbol{u}^2$ and $\boldsymbol{v}^2$, $3u/2 < v < 2u$, and suppose further that a third square $\boldsymbol{w}^2$ occurs at position $k+1$ of $\boldsymbol{x}$, where $v-u < w < v$, $w \neq u$, and $0 \leq k < v-u$. What can be said about the periodicity of $\boldsymbol{x}$?

An easily-proved result specifies a basic structure imposed on $\boldsymbol{x}$ by the conditions on u and v:

Lemma 1 ([11, 8]) *Suppose $x = v^2$ has prefix u^2. Then*

$$x = u_1 u_2 u_1 u_1 u_2 u_1 u_2 u_1 u_1 u_2 \tag{1}$$

with $u_1 = 2u - v > 0$, $u_2 = 2v - 3u > 0$ if and only if $3u/2 < v < 2u$.

We observe that (P) is only one of several cases that may arise. For example, [8, 11] also consider the case in which $u < v \leq 3u/2$; it turns out that this condition, even without the introduction of w^2, is sufficient to guarantee the occurrence of repetitions of small period and indeed to enable the number, size and location of the repetitions that occur to be exactly specified [8]. More generally, it will also be of interest to consider the various cases that arise when u^2 and v^2 do not occur at the same position — this is an open area of research.

Specifically, for the case defined by (P), it turns out to be convenient to consider 14 subcases, defined as shown in Table 1 in terms of the four parameters (u, v, k, w) (thus also u_1, u_2). For each subcase we specify the location in (1) of the start, the end and the center of the square w^2 in terms of the parameters u, v, k. A diagram of one of these subcases is shown in Figure 1.
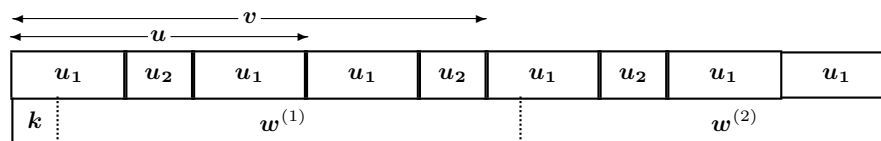
Fig. 1. Subcase 6

Based on a computation of all strings determined by $1 \leq u_1 \leq 30$, $1 \leq u_2 \leq 30$, together with all corresponding allowable values of w and k, [8] formulates conjectures as shown in Table 2. In this table, σ is the largest alphabet size compatible with the particular parameters u, v, k, w, while

$$d = \gcd(u_1, u_2, w) = \gcd(u, v, w) \tag{2}$$

and

$$s = \gcd(u - w, v - u); \quad \alpha = \lfloor u/s \rfloor; \quad \gamma = \lfloor v/s \rfloor; \quad \varepsilon = (v - u)/s. \tag{3}$$

For subcases 11–14 with $\sigma > d$, no conjecture was provided in [8], an oversight that we address in Section 3. In general, we summarize the status of the Table 2 conjectures as follows:

(1) The conjectures for subcases 5, 6 and 10 were first proved in [11]; then in [8] the conjectures for all the subcases in the first group, 1, 2, 5, 6, and 8–10, were proved.

(2) In Section 2 we prove the correctness of the conjecture for Subcase 4.

Table 1. The 14 subcases identified in [5], slightly modified, for three neighbouring squares $\boldsymbol{u}$, $\boldsymbol{v}$, $\boldsymbol{w}$ (with $v-u < w < v$, $w \neq u$, $0 \leq k < v-u$).

Subcase S	k	$k+w$	$k+2w$	Special Conditions
1	$0 \leq k \leq u_1$	$k+w \leq u$	$k+2w \leq u+u_1$	$k \geq u_2$
2	$0 \leq k \leq u_1$	$k+w \leq u$	$k+2w \leq u+u_1$	$k < u_2$
3	$0 \leq k \leq u_1$	$k+w \leq u$	$k+2w > u+u_1$	—
4	$0 \leq k \leq u_1$	$u < k+w \leq u+u_1$	—	—
5	$0 \leq k \leq u_1$	$u+u_1 < k+w \leq v$	—	—
6	$0 \leq k \leq u_1$	$v < k+w < 2u$	—	—
7	$u_1 < k < u_1+u_2$	$k+w \leq u+u_1$	$k+2w \leq 2u$	—
8	$u_1 < k < u_1+u_2$	$k+w \leq u+u_1$	$k+2w > 2u$	—
9	$u_1 < k < u_1+u_2$	$u+u_1 < k+w \leq v$	—	$w < u$
10	$u_1 < k < u_1+u_2$	$k+w \leq v$	$k+2w \leq u+v$	$w > u$
11	$u_1 < k < u_1+u_2$	$k+w \leq v$	$u+v < k+2w \leq 2v-u_2$	—
12	$u_1 < k < u_1+u_2$	$k+w \leq v$	$2v-u_2 < k+2w$	—
13	$u_1 < k < u_1+u_2$	$v < k+w \leq 2u$	—	—
14	$u_1 < k < u_1+u_2$	$2u < k+w < 2u+u_2-1$	—	—

Table 2. Overview of Conjectures

Subcases S	Conditions	Breakdown of $\boldsymbol{x}/\boldsymbol{v}^2$
$1,2,5,6,8\text{–}10$	$(\forall \boldsymbol{x}, \sigma = d)$	$\boldsymbol{x} = \boldsymbol{d}^{(x/d)}$
$3,4,7$	$\sigma = d$	$\boldsymbol{x} = \boldsymbol{d}^{(x/d)}$
	$\sigma > d$	$\boldsymbol{x} = \boldsymbol{s}^\alpha \boldsymbol{s}[1..u_1 \bmod s]\boldsymbol{s}^\gamma \boldsymbol{s}[1..u_1 \bmod s]\boldsymbol{s}^\varepsilon$
$11\text{–}14$	$\sigma = d$	$\boldsymbol{x} = \boldsymbol{d}^{(x/d)}$
	$\sigma > d$	(???)

(3) It was shown in [11] that the conjectures shown for Subcases 11–14 with $\sigma = d$ hold except in two Exceptional Cases, that we discuss further in Section 3. We establish necessary and sufficient conditions that $\sigma = d$, both in general and for Subcases 11-14; the latter result leads to a much simplified treatment of these subcases.

(4) Also in Section 3 we state new and more precise results for Subcases 11–14 when $\sigma > d$.

Thus overall we are able to confirm that in 12 of the 14 subcases arising from problem (P), the prefix $\boldsymbol{u}$ is itself a repetition of small periodicity, while $\boldsymbol{x}$ is either a repetition of small periodicity or breaks down into repetitive substrings of small periodicity. This means that for these 12 subcases the occurrence of the three squares $\boldsymbol{u}^2, \boldsymbol{v}^2, \boldsymbol{w}^2$ is impossible except in easily-recognized circumstances.

2 Subcase 4

We begin with a brief introduction of necessary terminology; in general, usage follows [12]. We write a string x of length x as $x[1..x]$. If $x = uvw$ for possibly empty strings u, v, w, then u is a **prefix**, v a **substring**, and w a **suffix** of x. If x has prefix u and also suffix u, $0 \leq u < x$, then u is a **border** of x and $x-u$ a **period** of x.

Given a nonempty string x on a finite alphabet Σ, a **repetition** in x is a substring u^e, u nonempty, integer $e \geq 2$, where $x = vu^e w$ for some (possibly empty) strings v, w. We call e the **exponent** of the repetition, u its **generator**, and u its **period**. For $e = 2, 3$, we say that u^e is a **square** or **cube**, respectively. There are well-known algorithms [3, 1, 10] that compute all the repetitions in x in $O(x \log x)$ time, asymptotically optimal since Fibonacci strings f contain $\Omega(f \log f)$ repetitions [3]. A repetition in x can be represented in constant space by a triple (i, p, e), where u^e is said to **occur** at position i in x and $p = u$.

A **run** in x (originally introduced in [9] as a **maximal periodicity**) is a substring w of x of minimum period $p \leq w/2$ occurring at some position i, where neither $x[i-1..i+w-1]$ nor $x[i..i+w]$ (whenever these are well-defined) has period p. Note that a run always has a prefix u^e, $p = u$, $e = \lfloor w/p \rfloor \geq 2$, that is a repetition. A run can be specified by a four-tuple (i, p, e, t), where i, p, e are defined as for a repetition, and the **tail** $t = w \bmod p$. In general, each run (i, p, e, t) determines $t+1$ repetitions of exponent e, and computing all the runs in x implicitly computes all the repetitions. Using heavy preprocessing, the runs can be computed in $O(x)$ time [7, 2].

To establish Subcase 4 we will need the following results:

Lemma 2 ("Periodicity Lemma", see [6]) *Let p and q be two periods of x, and let $d = \gcd(p, q)$. If $p+q \leq x+d$, then d is also a period of x.*

Lemma 3 ([7]) *If $x = uvw$, where both uv and vw have period $p \leq v$, then x has period p.*

Lemma 4 ([11]) *Let $x = arb$. If $x = r''r^t r'$, where t is a nonnegative integer and r', r'' are a prefix and a suffix, respectively, of r, then x has period $\gcd(r, r''-a)$.*

Theorem 5 (Subcase 4) *Suppose that string x has prefixes u^2 and v^2, $3u/2 < v < 2u$, and suppose further that a square w^2 occurs at position $k+1$ of x, where $u < w < v$, $0 \leq k \leq u_1$ and $u < k+w \leq u+u_1$. Then $x[1..u]$, $x[u+1..u+v]$ and $x[u+v+1..2v]$ each has period s, as defined in (3). Furthermore, if $k+2w \leq u+v$ and $s \leq k+w-u$, then x has period d, as defined in (2).*

Proof. By Lemma 1

$$x = u_1 u_2 u_1 u_1 u_2 u_1 u_2 u_1 u_1 u_2.$$

The first half of the w square equals $u_1' u_2 u_1 u_1''$ where u_1' is a suffix and u_1'' is a prefix of u_1. The second half of the square is a prefix of $u_1''' (u_2 u_1)^2 u_1 u_2$

where u_1''' is a suffix of u_1. Let p be the prefix of w of length

$$\min\{w, u + v - k - w\}.$$

The copy of this in the second half of the w-square has period $u_2 + u_1$. We claim the copy of p in the first half of this square extends to at least $x[2u_1+u_2]$. This is clearly so if $p = w$. Suppose instead that $p = u + v - k - w$ and, for the sake of contradiction, that the claim does not hold. Then

$$k+u+v-k-w < 2u_1 + u_2$$
$$\Rightarrow 3u_1+2u_2 < w$$
$$\Rightarrow v < w$$

which is a contradiction. Thus the claim holds. This allows us to apply Lemma 4 with p in the role of x, $u_2 u_1$ in the role of r, u_1''' in the role of r'' and u_1' in the role of a. We see that p has period

$$\gcd(|u_1'''|-|u_1'|, u_2+u_1) = \gcd((3u_1+u_2-k-w)-(u_1-k), u_1+u_2)$$
$$= \gcd(2u_1+u_2-w, u_1+u_2)$$
$$= \gcd(u-w, v-u)$$
$$= s, \text{ see } (3).$$

Thus $u_2 u_1 = s^\varepsilon$ where $\varepsilon = (v-u)/s$ is an integer. Hence $u_1 u_2 u_1$, $u_1 u_2 u_1 u_2 u_1$, and $u_1 u_2$ all have period s. The concatenation of these three factors makes up v^2 so that $x[1\ldots u]$, $x[u+1\ldots u+v]$ and $x[u+v+1\ldots 2v]$ all have period s. This is the first part of the theorem.

If $k+2w \le u + v$ then $p = w$. Hence the second half of the w-square is contained in $x[u+1\ldots u+v]$ and so w has period s. Now suppose that we also have $k+w-u \ge s$. The intersection of $x[k+1\ldots k+w]$ (the first half of the square) with $x[u+1\ldots u+v]$ has length $k+w-u$, so by Lemma 3, the whole of $x[k+1\ldots u+v]$ has period s. Another application of Lemma 3 using $x[1\ldots u]$ and $x[k+1\ldots u+w]$ shows that all of $x[1\ldots u+w]$ has this period. In particular $x[1..2u]$ has period s as well as period u. By the Periodicity Lemma it therefore has period $\gcd(u, \gcd(u-w, v-u)) = \gcd(u, v, w) = d$. This periodicity clearly extends to the whole of x, completing the proof. □

3   Subcases 11–14

We paraphrase here a version of the main result of [11] (compare with the problem statement (P) in Section 1):

Theorem 6 *Suppose that a string x has prefixes u^2 and v^2, $3u/2 < v < 2u$, and suppose further that a square w^2 occurs at position $k+1$ of x, where $u < w < v$ and $0 \le k \le v-u$. Then x has period d, as defined in (2), except possibly in two Exceptional Cases that, setting $d' = \gcd(u, v-w)$, are defined as follows:*

EC1 *if* $k > 2u - v$, $v > k + w$, $u + v < k + 2w$, *and*

$$d' > 2u - (k+w) + d, \tag{4}$$

then $\boldsymbol{x}[1..2u]$ *and* $\boldsymbol{x}[k+w+1..2v]$ *both have period* d';
EC2 *if* $k > 2u - v$, $v \leq k + w$, *and*

$$d' > 2u - v, \tag{5}$$

then $\boldsymbol{x}[1..2u]$ *and* $\boldsymbol{x}[v+1..2v]$ *both have period* d'.

We make several remarks about this result:

(R1) Theorem 6 requires $w > u$, while in problem (P) of Section 1 the condition $w > v - u$ is sufficient, a weaker one since by Lemma 1

$$u - (v - u) = (2u_1 + u_2) - (u_1 + u_2) = u_1 > 0.$$

It is tedious but straightforward (see the Appendix) to verify that for Subcases 5, 6 and 10–14 of Table 1, $w > u$, and so Theorem 6 applies to these cases. In fact, for these cases it is slightly more general in its application, since it allows $k = v - u$.

(R2) The Exceptional Cases identified in Theorem 6 both require $k > 2u - v = u_1 + u_2$ and so by Table 1 cannot apply to Subcases 5 and 6. Thus the main result of Theorem 6, that $\boldsymbol{x}$ has period d, agrees exactly with the conjectures for Subcases 5 and 6 shown in Table 2 and proved in [8].

(R3) In addition EC1 requires $v > k + w$ and $u + v < k + 2w$, conditions that we see from Table 1 restrict its application to Subcases 11 and 12. At the same time EC2 requires $v \leq k + w$, conditions compatible with any of Subcases 11–14. Thus the main result of Theorem 6 again agrees exactly with the conjecture for Subcase 10 as shown in Table 2 and proved in [8].

(R4) As given in [11], Theorem 6 includes also the case $u < v \leq 3u/2$, showing that in this case also $\boldsymbol{x}$ has period $\gcd(u, v, w)$. But [8] gives the stronger result that, without introducing $\boldsymbol{w}$, $\boldsymbol{x}$ contains exactly $m + 5$ specified runs (repetitions), where $m = \lfloor u/(v-u) \rfloor$, and otherwise contains only runs (repetitions) of period strictly less than $v - u \leq u/2$.

(R5) We can relate the new quantity d' to d, as follows. Observing that $d = \gcd(u_1, u, w) = \gcd(u_1, u, w + u_1)$, while

$$d' = \gcd(2u_1 + u_2, 3u_1 + 2u_2 - w) = \gcd(w + u_1, u),$$

we see

$$d = \gcd(d', u_1). \tag{6}$$

Based on these remarks, we see that Theorem 6 establishes the breakdown of $\boldsymbol{u}$ (as well as of $\boldsymbol{v}$ and $\boldsymbol{w}$) into a repetition of small period d for Subcases 11-14, except when EC1 holds (for Subcases 11 and 12) and when EC2 holds (Subcases 11–14).

Note that Theorem 6 requires $u < w < v$, in other words

$$0 < v-w < v-u = u_1+u_2 < u.$$

It follows that $d' = \gcd(u, v-w)$ is a positive integer strictly less than u that divides u. Suppose first that EC1 holds. Since therefore $\boldsymbol{x}[1..2u]$ has period d', it follows that $\boldsymbol{u}$ is a repetition of period d'. Thus, since $\boldsymbol{v} = \boldsymbol{u}\boldsymbol{u_1}\boldsymbol{u_2}$, it is immediate that $\boldsymbol{v}$ also has period d'. Similarly, for EC2, $\boldsymbol{u}$ is again a repetition of period d' and $\boldsymbol{v}$ has period d'.

If now we suppose that the conditions of Theorem 6 are satisfied, but neither EC1 nor EC2 holds, then of course $\boldsymbol{u}$ and $\boldsymbol{v}$ are repetitions of period d. But since by (6) $d \mid d'$ and $d' \mid u$, therefore again $\boldsymbol{u}$ is a repetition of period d' and so $\boldsymbol{v}$ has period d'. Thus we have established a somewhat stronger version of Theorem 6:

Lemma 7 *Suppose the conditions of Theorem 6 hold, with $d = \gcd(u, v, w)$, $d' = \gcd(u, v-w)$. Then $\boldsymbol{u}$ is a repetition of period d', and $\boldsymbol{v}$ also has period d'. Moreover, $\boldsymbol{u}$ and $\boldsymbol{v}$ are both repetitions of period d if and only if neither EC1 nor EC2 of Theorem 6 holds.*

This result makes it clear that for Subcases 11–14, if $\boldsymbol{x} \neq \boldsymbol{d}^{(x/d)}$ as specified in Table 2, then an appropriate replacement for (???) would be

$$\boldsymbol{u} = \boldsymbol{d'}^{(u/d')}.$$

The puzzle that remains is to show the equivalence of the conditions stated for EC1 and EC2 and the conjectured (and simpler) condition $\sigma > d$ given in Table 2. It was shown informally in [8] that $\sigma \not< d$ and that, over all the subcases 1–14, $\sigma = d$ is a condition necessary for periodicity d, but the following lemma goes further. It establishes for all subcases that $\sigma = d$ if and only if $\boldsymbol{x}$ has period d, hence that $\sigma > d$ if and only if $\boldsymbol{x}$ does not have periodicity d — in particular, for Subcases 11–14, if and only if one of EC1/EC2 holds.

Lemma 8 *Suppose $\boldsymbol{x} = \boldsymbol{v}^2$ has prefix $\boldsymbol{u}^2$, and that $\boldsymbol{x}[k+1, k+2w]$ is a square with $0 \leq k \leq u_1$. Let σ be the maximum alphabet size that $\boldsymbol{x}$ can have consistent with these periodicities. Then $\boldsymbol{x}$ has period $d = \gcd(u, v, w)$ if and only if $\sigma = d$.*

Proof. The periodicities in $\boldsymbol{x}$ engender a set of equations among the letters of $\boldsymbol{x}$ ($\boldsymbol{x}[i] = \boldsymbol{x}[i+v]$ for $i \in 1..v$, etc.). This produces a set of equivalence classes on $\{1, \ldots, |\boldsymbol{x}|\}$, with $i \sim j$ if $\boldsymbol{x}[i] = \boldsymbol{x}[j]$. Clearly σ is the number of resulting equivalence classes. If $i \not\equiv j \mod d$ then i and j belong to different equivalence classes, since the equations can only equate letters with indices differing by a multiple of d. Thus $\sigma \geq d$.

If $\boldsymbol{x}$ has period d the letters used in $\boldsymbol{x}[1..d]$ will be the same as those used in the whole of $\boldsymbol{x}$. This establishes the lemma in one direction.

For the other direction suppose $\sigma = d$. Then the set of equivalence classes is $\{S_n : n = 1, \ldots, d\}$ where $S_n = \{i : i \equiv n \mod d\}$. We can then label the letters of $\boldsymbol{x}$ with the index of their equivalence class. Clearly $\boldsymbol{x}$ has period d. □

Consider, for example, the string

$$\boldsymbol{x} = \boldsymbol{v}^2 = 12112112112\ 12112112112 \tag{7}$$

corresponding to $u_1 = 1$, $u_2 = 4$, $u = 6$, $v = 11$, $k = 3$, $w = 8$. Since $\sigma = 2 > 1 = d$, we conclude from Lemma 8 that $\boldsymbol{u}$ does not have period d; moreover, since $\boldsymbol{x}$ is an example of Subcase 12, it follows from Lemma 7 that $\boldsymbol{u}$ is a repetition of period $d' = 3$. Note that since $v = 11 = k+w$, therefore $\boldsymbol{x}$ satisfies EC2 of Theorem 6.

For Subcases 11–14 we can establish another necessary and sufficient condition that $\sigma = d$:

Lemma 9 *For Subcases 11–14, let $d = \gcd(u, v, w)$, $d' = \gcd(u, v-w)$. Then $\sigma > d$ if and only if $d' > u_1$.*

Proof. Suppose $\sigma > d$. Then by Lemma 8 $\boldsymbol{x}$ does not have period d, and so by Lemma 7 either EC1 or EC2 of Theorem 6 holds. If EC1 holds, then

$$\begin{aligned}
d' &> 2u-(k+w)+d \\
&> 2u-v+d \\
&= u_1+d > u_1.
\end{aligned}$$

However, if EC2 holds, then $d' > 2u-v = u_1$. This proves necessity.

Suppose then that $d' \leq u_1$. We show that therefore $\sigma = d$. If $d' = u_1$, then by (6) $d = d'$, so that by Lemma 7 $\boldsymbol{u}$ is a repetition of period d, and so by Lemma 8 $\sigma = d$. On the other hand, if $d' < u_1$, two cases arise depending on the value of $t = v \bmod d'$:

$(t = 0)$ In this case $\boldsymbol{u}$ and $\boldsymbol{v}$, hence $\boldsymbol{u_1}$ and $\boldsymbol{u_2}$, are repetitions of period d'. Therefore in particular $d' \mid u_1$ (d divides u_1), so that by (6) $d = d'$; hence $\boldsymbol{u}$ is a repetition of period d, implying as above that $\sigma = d$.

$(t > 0)$ Here, since by Lemma 7 $\boldsymbol{u}$ is a repetition of period d', it follows that the prefix $\boldsymbol{u}^2$ of $\boldsymbol{x}$ has suffix $\boldsymbol{tu_1} = \boldsymbol{d'}^r$ for some $r \geq 2$, where $\boldsymbol{tu_1}$ has prefix $\boldsymbol{d'}$. Thus $\boldsymbol{d'} = R_t(\boldsymbol{d'})$, the tth rotation of $\boldsymbol{d'}$, and so by [12, Lemma 1.4.2] $\boldsymbol{d'}$ is a repetition of period $g = \gcd(t, d'-t)$. Since $\boldsymbol{v} = \boldsymbol{d'}^{\lfloor v/d' \rfloor}\boldsymbol{t}$, it follows that $\boldsymbol{u}$ and $\boldsymbol{v}$ are repetitions of period g, hence so also is $\boldsymbol{u_1}$. We find then that $g \mid d'$ and $g \mid u_1$, while $d = \gcd(d', u_1)$, telling us that $g \mid d$. Thus $\boldsymbol{u}$ and $\boldsymbol{v}$ are also repetitions of period d, and so by Lemma 9 $\sigma = d$, as required. $\square$

Thus Lemma 9 replaces conditions EC1 and EC2 of Theorem 6 by the condition $d' > u_1$. Note that for the string (7), an example of Subcase 12, $d' = 3 > 1 = u_1$ and $\sigma = 2 > 1 = d$.

We now prove a result that provides a more precise periodic breakdown for Subcases 11–14 whenever $d' > u_1$:

Lemma 10 *For Subcases 11–14, suppose $d' = \gcd(u, v-w) > u_1$. Then $\boldsymbol{d'}$ has period $t = v \bmod d' > 0$ ($\boldsymbol{t}$ is a nonempty border of $\boldsymbol{v}$ and a generator of $\boldsymbol{d'}$).*

Proof. If d' divides v, then

$$d' = \gcd(u, v-w, v) = \gcd(u, v, w) = d.$$

Since by Lemma 7 $\boldsymbol{u}$ and $\boldsymbol{v}$ have period d', it follows therefore from (2) that $\boldsymbol{x}$ is a repetition of period $d = d'$, hence by Lemma 8 that $\sigma = d$, and so by Lemma 9 that $d' \leq u_1$, contradicting the hypothesis of the lemma. Therefore $t = v \bmod d' > 0$, and so

$$d' \lfloor v/d' \rfloor < v < 2u.$$

Setting $\boldsymbol{d'} = \boldsymbol{x}[1..d']$, we see that $\boldsymbol{x}\big[d'\lfloor v/d'\rfloor+1..d'\lfloor v/d'\rfloor+d'\big]$ is a copy of $\boldsymbol{d'}$, and so, by the v-periodicity, is $\boldsymbol{x}[v+1..v+d']$. These two copies intersect, so $\boldsymbol{d'}$ has a border of length $d'\lfloor v/d'\rfloor+d'-v$. Hence $\boldsymbol{d'}$ has period

$$\begin{aligned} d' - (d'\lfloor v/d'\rfloor+d'-v) &= v - d'\lfloor v/d'\rfloor \\ &= v \bmod d', \end{aligned}$$

as required. $\qquad\qquad\square$

Putting together the results of Lemmas 7, 9 and 10, we have the following succinct formulation:

> Suppose that the parameters (u, v, k, w) correspond to any one of Subcases 11–14. Then $\boldsymbol{u}$ is a repetition of period $d' = \gcd(u, v-w)$. If $d' > 2u-v$, then $\boldsymbol{d'}$ has period $t = v \bmod d' > 0$, so that in fact
>
> $$\boldsymbol{v} = (\boldsymbol{t}^{d'/t})^{v/d'};$$
>
> otherwise, $\boldsymbol{u}$ and $\boldsymbol{v}$ are repetitions of period $d = \gcd(u, v, w)$.

Thus, computing d, d', $2u-v$, and t determines $\boldsymbol{x}$ up to an isomorphism on the alphabet. Note that d' may or may not be a run (possibly $t > d/2$). For example, in the string (7), since $d' > u_1 = 2u-v$, therefore $\boldsymbol{d'} = 121$ of length 3 has period $t = 11 \bmod 3 = 2 > 3/2$.

4   Subcase Reduction to Canonical Form

In this section we introduce ideas that may in future assist in the simplification of both results related to neighbouring squares and their proofs. Observe that whenever a subcase is ***regular*** — that is, $\boldsymbol{x}$ breaks down into a repetition of period $d = \gcd(u_1, u_2, w)$ —, the value of k can be replaced by zero. That is, since each copy of $\boldsymbol{w}$ must be a repetition of period d, the given subcase can be replaced by an equivalent subcase in which $\boldsymbol{w}^2$ is shifted left by k positions. Consider, for example, an instance of Subcase 12 determined by

$$u = 16, \ v = 26, \ k = 7, \ w = 20,$$

hence $u_1 = 6$, $u_2 = 4$, $d = d' = 2$. Since $d' < u_1$, we know by Lemma 9 that $\sigma = d$, hence that the resulting string (on maximum alphabet size 2) is

just $\boldsymbol{x} = (12)^{26}$ with $\boldsymbol{w} = (21)^{10}$. Clearly we can also regard this string as an instance of Subcase 4 with parameters

$$u = 16, \ v = 26, \ k = 0, \ w = 20,$$

where $u = 16 < 20 = k+w < u+u_1 = 22$. (Of course, $(12)^{26}$ is an instance of many strings in many other subcases.) We call this left shift by k of a regular string a **reduction** and the corresponding repetition $\boldsymbol{x} = (1 \ 2 \ \cdots \ d)^{x/d}$ its **canonical form**.

Observation 11 *An instance (u, v, k, w) of problem (P) is regular if and only if the maximum alphabet size $\sigma = \gcd(u, v, w)$.*

Observation 12 *Every regular instance (u, v, k, w) of Subcase S, $7 \le S \le 14$, reduces to a regular instance $(u, v, 0, w)$ of some Subcase $S' \le 6$.*

In other words, regular instances of all 14 subcases (assuming the conjectures for Subcases 3 & 7 hold) reduce to the Three Squares Lemma [4]!

5 Conclusions & Future Work

In this paper we have considered the problem (P) on overlapping squares parameterized by (u, v, k, w), which breaks down into 14 subcases depending on the parameter values. We have provided a proof for Subcase 4, previously unresolved, and we have clarified, simplified and strengthened the results for Subcases 11–14. We have also provided a context that may be useful for future study of these and related problems. The immediate tasks that remain are proofs for the conjectures given in Table 2 relating to Subcases 3 and 7.

Appendix

We show that for Subcases 5, 6 and 10–14 of Table 1, $w > u$:

(5) Since $k \le u_1$ and $k+w > u+u_1$, $w > u$.
(6) Since $k+w > v > u+u_1$, the result follows from (5).
(10) The "special condition" is $w > u$.
(11) Since $k+2w > u+v$ and $k < u_1+u_2$, we have

$$
\begin{aligned}
2w &> u+v-k \\
&\ge u+v-(u_1+u_2) \\
&= u+v-(2u-v)-(2v-3u) \\
&= 2u,
\end{aligned}
$$

and so $w > u$.
(12) Here $k+2w > 2v-u_2 > u+v$ and so the result follows from (11).

(13) Since $k+w > v$ and $k < u_1+u_2$, we have

$$
\begin{aligned}
w &> v-k \\
&\geq v-(u_1+u_2) \\
&= v-(2u-v)-(2v-3u) \\
&= u.
\end{aligned}
$$

(14) Here $k+w > 2u > v$ and so the result follows from (13).

References

1. Alberto Apostolico & Franco P. Preparata, **Optimal off-line detection of repetitions in a string**, *Theoret. Comput. Sci. 22* (1983) 297–315.
2. Gang Chen, Simon J. Puglisi & W. F. Smyth, **Fast & practical algorithms for computing all the runs in a string**, *Proc. 18th Annual Symp. Combinatorial Pattern Matching*, B. Ma & K. Zhang (eds.), LNCS 4580, Springer-Verlag (2007) 307–315.
3. Maxime Crochemore, **An optimal algorithm for computing all the repetitions in a word**, *Inform. Process. Lett. 12–5* (1981) 244–248.
4. Maxime Crochemore and Wojciech Rytter, **Squares, cubes, and time-space efficient strings searching**, *Algorithmica 13* (1995), pp. 405–425.
5. Kangmin Fan, Simon J. Puglisi, W. F. Smyth & Andrew Turpin, **A new periodicity lemma**, *SIAM J. Discrete Math. 20–3* (2006) 656–668.
6. N. J. Fine and H. S. Wilf, **Uniqueness theorems for periodic functions**, *Proc. Amer. Math. Soc. 16* (1965) 109–114.
7. Roman Kolpakov & Gregory Kucherov, **On maximal repetitions in words**, *J. Discrete Algorithms 1* (2000) 159–186.
8. Evguenia Kopylov & W. F. Smyth, **The three squares lemma revisited**, *J. Discrete Algorithms* (2011) to appear.
9. Michael G. Main, **Detecting leftmost maximal periodicities**, *Discrete Applied Maths. 25* (1989) 145–153.
10. Michael G. Main & Richard J. Lorentz, **An $O(n\log n)$ algorithm for finding all repetitions in a string**, *J. Algorithms 5* (1984) 422–432.
11. R. J. Simpson, **Intersecting periodic words**, *Theoret. Comput. Sci. 374* (2007) 58–65.
12. Bill Smyth, *Computing Patterns in Strings*, Pearson Addison-Wesley, 2003, 423pp.