McMaster University

Advanced Optimization Laboratory

Title:

On singularities of extremal periodic strings

Authors:

Antoine Deza and Frantisek Franek

AdvOL-Report No. 2012/03

November 2012, Hamilton, Ontario, Canada

On singularities of extremal periodic strings *

Antoine Deza and Frantisek Franek

Abstract

Fraenkel and Simpson conjectured in 1998 that the number of distinct squares in a string is at most its length. Similarly, Kolpakov and Kucherov conjectured in 1999 that the number of runs in a string is at most its length. Since then, both conjectures attracted the attention of many researchers and many results have been presented, including asymptotic lower bounds for both, asymptotic upper bounds for runs, and universal upper bounds for distinct squares. We consider the role played by the size of the alphabet of the string in both problems and investigate the functions $\sigma_d(n)$ and $\rho_d(n)$, i.e. the maximum number of distinct primitively rooted squares, respectively runs, over all strings of length n containing exactly d distinct symbols. We revisit earlier results and conjectures and express them in terms of singularities of the two functions where a pair (d,n) is a singularity if $\sigma_d(n) - \sigma_{d-1}(n-2) \geq 2$, or $\rho_d(n) - \rho_{d-1}(n-2) \geq 2$ respectively.

Keywords: *string, square, primitively rooted square, maximum number of distinct primitively rootedsquares, run, maximum number of runs, parameterized approach, $(d, n-d)$ table, singularity*

1 Introduction and motivation

A *square*, or a *tandem repeat* is a fundamental regularity in a string, and a simplest of *repetitions*. We denote this fact as u^2 indicating concatenation of a string u with a copy of itself; u is referred to as the *generator* of the square and the length of u is referred to as the *period* of the square. A *primitively rooted square* is a square whose generator is *primitive*, i.e. not a repetition itself. A *run*, a maximal possibly fractional primitively rooted repetition in a string, was conceptually introduced by Main in 1989 [19]. The term *run* was coined by Iliopoulos, Moore, and Smyth in 1997 [17]. A run in a string x encoded by a four-tuple (s,p,e,t) has a primitive *generator* $x[s \mathrel{..} s{+}p]$ of length p repeating e times ($e \geq 2$), followed by the prefix of the generator of length t ($0 \leq t < p$). More precisely, $x[s{+}i] = x[s{+}i{+}rp]$ for any $1 \leq i < s+p$ and $1 \leq r < e$, and $x[s+i] = x[s+i+rp]$ for any $1 \leq i \leq t$ and $1 \leq r \leq e$. The maximality in this context means that the same is neither true for $s-1$ nor for $s+1$. Thus, the knowledge of all runs succinctly captures the knowledge of all occurrences of all repetitions. It is natural to ask about the maximum number of distinct squares or runs in a string and to expect both to depend on the length of the string.

The problem of the number of distinct squares when the types of the squares in a string are counted rather than their occurrences, was first introduced in 1998 by Fraenkel and

*This work was supported by the *Natural Sciences and Engineering Research Council of Canada* and by the *Canada Research Chair program.*

Simpson [13], showing that the number of distinct squares in a string of length n is at most by $2n$, in particular the number of primitively rooted distinct squares in strings of length $n \geq 5$ is bounded by $2n - 8$, and for binary strings of length $n \geq 22$ is bounded by $2n - 29$. They also gave an infinite sequence of of strictly increasing length with a number of primitively rooted distinct squares asymptomatically approaching the strings length from below, and conjectured that the number of distinct squares is at most the length n of the string. The work relied on an improved Lemma 10 of Crochemore and Rytter [7] stating that *if* u^2, v^2, w^2 *are prefixes of a string* x *and* w *is primitive, and* $|u| > |v| > |w|$*, then* $|u| \geq |v| + |w|$. Ilie [15] provided a simpler proof of the main lemma of [13] and presented an asymptotic upper bound of $2n - \Theta(\log n)$ in [16].

Though there may be as many as $O(n \log n)$ repetitions in a string of length n, see [5], it was hoped that the more succinct notation of runs would eliminate the need to list all repetitions. Kolpakov and Kucherov [18] in 1999 showed that the number of runs in a string is $O(n)$ and conjectured that the maximum number of runs in a string is at most its length n. Let $\rho(n)$ denote the maximum number of runs over all strings of length n. Several authors have presented asymptotic upper and lower bounds for $\rho(n)$, see Crochemore and Ilie [6], respectively Matsubara et al. [20], for upper, respectively lower, bounds, and references therein.

The problems of the number of distinct squares and runs where, in addition to the consideration of the length of a string, the size of its alphabet is considered as an additional parameter, have been studied in [1, 4, 8, 9, 11, 12].

A string x of length n with d distinct symbols is referred to as a (d, n)-string, $\boldsymbol{s}(x)$, respectively $\boldsymbol{r}(x)$, denotes the number of distinct primitively rooted squares, respectively runs, of x. Let $\sigma_d(n)$, respectively $\rho_d(n)$, denote the maximum number of distinct primitively rooted squares, respectively runs, over all (d, n)-strings. A (d, n)-string satisfying $\boldsymbol{s}(x) = \sigma_d(n)$, respectively $\boldsymbol{r}(x) = \rho_d(n)$, is referred to as a *square-maximal*, respectively *run-maximal* string.

Some elementary properties of the function $\sigma_d(n)$ are discussed in [11], where the values of $\sigma_d(n)$ are presented in the form of $(d, n - d)$ table, where the value $\sigma_d(n)$ is the entry at row d and column $n - d$, pointing to ways of applying reductions to the problem of bounding the maximum number of distinct squares.

The computed values with $2 \leq d \leq 15$ and $2 \leq n - d \leq 15$ of the $(d, n - d)$ table for $\sigma_d(n)$ are given in Table 1 with the main diagonal in bold, and the up-to-date table of all computed values is available on-line at [10].

Some elementary properties of the function $\rho_d(n)$ are discussed in [4, 8, 9], where the values are presented in the $(d, n - d)$ *table*, pointing to ways of applying reductions to the problem of bounding the maximum number of runs. The computed values with $2 \leq d \leq 15$ and $2 \leq n - d \leq 15$ of the $(d, n - d)$ table for $\rho_d(n)$ are given in Table 2 with the main diagonal in bold, and the up-to-date table of all computed values is available on-line at [2].

While there are similarities, the investigation of distinct squares is different from the investigation of runs in a string in many ways. For instance the concatenation of two strings may merge some runs from both strings, but would not merge distinct squares, on the other hand all the runs in both strings count, while the same is not true for distinct squares. The computed values of $\sigma_d(n)$ and $\rho_d(n)$ appear to be very close and hint at a simple relationship

d \ $n-d$	2	3	4	5	6	7	8	9	10	11	12	13	14	15
2	**2**	2	3	3	4	5	6	7	7	8	9	10	11	12
3	2	**3**	3	4	4	5	6	7	8	8	9	10	11	12
4	2	3	**4**	4	5	5	6	7	8	9	9	10	11	12
5	2	3	4	**5**	5	6	6	7	8	9	10	10	11	12
6	2	3	4	5	**6**	6	7	7	8	9	10	11	11	12
7	2	3	4	5	6	**7**	7	8	8	9	10	11	12	12
8	2	3	4	5	6	7	**8**	8	9	9	10	11	12	13
9	2	3	4	5	6	7	8	**9**	9	10	10	11	12	13
10	2	3	4	5	6	7	8	9	**10**	10	11	11	12	13
11	2	3	4	5	6	7	8	9	10	**11**	11	12	12	13
12	2	3	4	5	6	7	8	9	10	11	**12**	12	13	13
13	2	3	4	5	6	7	8	9	10	11	12	**13**	13	14
14	2	3	4	5	6	7	8	9	10	11	12	13	**14**	14
15	2	3	4	5	6	7	8	9	10	11	12	13	14	**15**

Table 1: $(d, n-d)$ table for $\sigma_d(n)$ with $2 \le d \le 15$ and $2 \le n-d \le 15$

$\sigma_d(n) \le \rho_d(n)$ as illustrated in Table 3 where the entries are presented in a $(d, n-2d)$ table. The up-to-date table of all computed values $\rho_d(n) - \sigma_d(n)$ is available on-line at [10].

The computed values of $\sigma_d(n)$ and $\rho_d(n)$ lead to the hypothesized universal upper bounds [9]:

$$\sigma_d(n) \le n - d - \lfloor \log_2 \lfloor (n+10-2d)/6 \rfloor \rfloor - \lceil \log_2 \lceil (n+3-2d)/5 \rceil \rceil \text{ for } n \ge 2d+2$$

$$\rho_d(n) \le n - d - \lceil \log_2 \lceil (n+4-2d)/4 \rceil \rceil \text{ for } n \ge 2d$$

The values for $\sigma_d(n)$ and $\rho_d(n)$ computed to date indicate that for $n \ge d \ge 3$, $\sigma_d(n) - \sigma_{d-1}(n-2) = 1$ and $\rho_d(n) - \rho_{d-1}(n-2) = 1$ except for relatively rare pairs (n, d) satisfying $\sigma_d(n) \ge \sigma_{d-1}(n-2) + 2$, respectively $\rho_d(n) = \rho_{d-1}(n-2) + 2$. For $\sigma_d(n)$ function, so far we have found two such pairs, $(3, 35)$ as $\sigma_3(35) = 25$ and $\sigma_2(33) = 23$, and $(3, 36)$ as $\sigma_3(36) = 26$ and $\sigma_2(34) = 24$, see [10]. For $\rho_d(n)$ function, so far we have found three such pairs, $(3, 15)$ as $\rho_3(15) = 10$ and $\rho_2(13) = 8$, see Table 2 and the entries in bold italic, $(3, 43)$ as $\rho_2(41) = 33$ and $\rho_3(43) = 35$, and $(4, 44)$, as $\rho_3(42) = 33$ and $\rho_4(44) \ge \rho_3(43) = 35$, see [2].

Though it is impossible to have 3 consecutive equal values in any row of the $(d, n-d)$ table for $\rho_d(n)$ as $\rho_d(n+2) > \rho_d(n)$, there is no such restriction for $\sigma_d(n)$. Such three times repeating values were found for binary strings at lengths 31, 32, and 33, however it is the only case known to us to date. In general, whenever $\sigma_d(n) = \sigma_d(n+1) = \sigma_d(n+2)$, it follows that either eventually there is an exceptional pair (n', d') so that $n' - 2d' \ge n - 2d$, or $\sigma_{d+k}(n+2k) = \sigma_{d+k}(n+1+2k) = \sigma_{d+k}(n+2+2k)$ for any $k \ge 1$. Moreover, it follows that $\sigma_{d+1}(n+2) > \sigma_d(n) = \sigma_d(n+2)$; that is, the maximum number of squares among all strings of length $n+2$ is not achieved by $(d, n+2)$-strings. In particular, $\sigma_2(31) = \sigma_2(33) = \sigma_2(33) = 23$

		$n-d$													
		2	3	4	5	6	7	8	9	10	11	12	13	14	15
d	2	**2**	2	3	4	5	5	6	7	8	***8***	10	10	11	12
	3	2	**3**	3	4	5	6	6	7	8	9	***10***	11	11	12
	4	2	3	**4**	4	5	6	7	7	8	9	10	11	12	12
	5	2	3	4	**5**	5	6	7	8	8	9	10	11	12	13
	6	2	3	4	5	**6**	6	7	8	9	9	10	11	12	13
	7	2	3	4	5	6	**7**	7	8	9	9	10	11	112	13
	8	2	3	4	5	6	7	**8**	8	9	10	11	11	12	13
	9	2	3	4	5	6	7	8	**9**	9	10	11	12	12	13
	10	2	3	4	5	6	7	8	9	**10**	10	11	12	13	13
	11	2	3	4	5	6	7	8	9	10	**11**	11	12	13	13
	12	2	3	4	5	6	7	8	9	10	11	**12**	12	13	14
	13	2	3	4	5	6	7	8	9	10	11	12	**13**	13	14
	14	2	3	4	5	6	7	8	9	10	11	12	13	**14**	14
	15	2	3	4	5	6	7	8	9	10	11	12	13	14	**15**

Table 2: $(d, n-d)$ table for $\rho_d(n)$ with $2 \leq d \leq 15$ and $2 \leq n-d \leq 15$

		$n-2d$															
		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
d	2	0	0	0	1	1	0	0	0	1	**0**	1	0	0	0	1	1
	3	0	0	0	1	1	0	0	0	1	1	1	0	0	0	1	1
	4	0	0	0	1	1	0	0	0	1	1	1	0	0	0	1	1
	5	0	0	0	1	1	0	0	0	1	1	1	0	0	0	1	1
	6	0	0	0	1	1	0	0	0	1	1	1	0	0	0	1	1
	7	0	0	0	1	1	0	0	0	1	1	1	0	0	0	1	1

Table 3: $(d, n-2d)$ table for $\rho_d(n) - \sigma_d(n)$ with $d \leq 7$ and $n-2d \leq 15$

and so any binary string of length 33 has at most 23 distinct squares while there is a ternary string with 24 distinct squares, see [10].

We hypothesize that though rare, there are infinitely many such pairs for both $\rho_d(n)$ and $\sigma_d(n)$. Our hypothesis implies that the values along a column of Table 3 are constant except for every such pair (d, n) and its corresponding entry in column $n-2d$ and row $d-1$ in the $(d, n-d)$ table. For illustration, for $(3, 15)$ for $\rho_d(n)$, the entry at column 9 and row 2 is depicted in bold in Table 3. This leads us to the following definition: we refer to a pair (d, n) such that $\sigma_d(n) - \sigma_{d-1}(n-2) \geq 2$, or $\rho_d(n) - \rho_{d-1}(n-2) \geq 2$ respectively, as a *singularity*.

2 Basic properties of $\sigma_d(n)$ and $\rho_d(n)$

The following basic properties of $\sigma_d(n)$ were presented in [11] and are summarized in Proposition 1. The values of $\sigma_d(n)$ are increasing when moving right along a row of the $(d, n-d)$ table and the increase is of at most 2, the values are increasing when moving down along a column, the values are strictly increasing when moving along descending diagonals, the values under and on the main diagonal along a column are constant. In addition, the 2 values immediately above the main diagonal are equal and differ from the value on the main diagonal by at most 1 for $d \geq 4$. Note that the main diagonal of the $(d, n-d)$ table corresponds to the values of $\sigma_d(2d)$ for $d \geq 2$.

Proposition 1 ([11]).
(s_1) $0 \leq \sigma_d(n+1) - \sigma_d(n) \leq 2$ *for* $n \geq d \geq 2$,
(s_2) $\sigma_d(n) \leq \sigma_{d+1}(n+1)$ *for* $n \geq d \geq 2$,
(s_3) $\sigma_d(n) < \sigma_{d+1}(n+2)$ *for* $n \geq d \geq 2$,
(s_4) $\sigma_d(n) = \sigma_{d+1}(n+1)$ *for* $2d \geq n \geq d \geq 2$,
(s_5) $\sigma_d(n) \geq n-d$, $\sigma_d(2d+1) \geq d$ *and* $\sigma_d(2d+2) \geq d+1$ *for* $2d \geq n \geq d \geq 2$,
(s_6) $\sigma_{d-1}(2d-1) = \sigma_{d-2}(2d-2)$ *and* $0 \leq \sigma_d(2d) - \sigma_{d-1}(2d-1) \leq 1$ *for* $d \geq 4$,
(s_7) $1 \leq \sigma_{d+1}(2d+2) - \sigma_d(2d) \leq 2$ *for* $d \geq 2$.

Since $\sigma_2(41) = 31$ and $2 \geq \sigma_2(n+1) - \sigma_2(n) \geq 0$, we have the following slight improvement of the upper bound for $\sigma_2(n)$ in Corollary 2. In addition, $\sigma_{15}(30) = 15$ and $\sigma_{d+1}(2d+2) - \sigma_d(2d) \leq 2$ implies $\sigma_d(n) \leq 2n - 15$ for $n \geq 15$, thus, using $\sigma_d(n) \leq \sigma_{n-d}(2n-2d)$ for $n \geq 2d \geq 4$, we can also slightly improve the bound for the maximum number $\sigma(n)$ of distinct squares over all strings of length n, for comparison see [13].

Corollary 2.
(c_1) $\sigma_2(n) \leq 2n - 51$ *for* $n \geq 41$,
(c_2) $\sigma(n) \leq 2n - 19$ *for* $n \geq 30$.

A *singleton* refers to a symbol in a string that occurs exactly once, while a *pair* refers to a symbol that occurs exactly twice. The following structural result for square-maximal strings on the main diagonal was noted in [11] and we presented in here reformulated in terms of singularities.

Proposition 3. *Let* $(d, 2d)$ *be the first singularity on the main diagonal, i.e. the least* d *such that* $\sigma_d(2d) - \sigma_{d-1}(d-2) \geq 2$. *Then any square-maximal* $(d, 2d)$*-string does not contain a pair but must contain at least* $\lceil \frac{2d}{3} \rceil$ *singletons.*

Propositions 1 and 3 yield Theorem 4 underlining the importance of the diagonals of the $(d, n-d)$ table with respect to the conjectured upper bound $n-d$ for $\sigma_d(n)$. In particular, Theorem 4 shows that in order to prove $\sigma_d(n) \leq n-d$ for all n and d it is enough to prove the bound for the special case $n = 2d$ for all d, i.e. for the main diagonal of the $(d, n-d)$ table. In other words, $\sigma_d(2d) \leq d$ for all d implies that the maximum number $\sigma(n)$ of runs over all strings of length $n \geq 3$ satisfies $\sigma(n) \leq n-2$. This equivalence is generalized to the special case $n = 4d$. In addition, the role played by $\sigma_d(2d)$ and $\sigma_d(2d+1)$ is underlined as well as the hypothesis that the run-maximal $(d, 2d)$-strings are, up to relabelling, unique. The original version of Theorem 3 is presented in [11], here we restate it in terms of singularities:

Theorem 4.
(e_1) *no* $(d, 2d)$ *singularity* $\Longleftrightarrow \{\sigma_d(n) \le n-d$ *for* $n \ge d \ge 2\}$,
(e_2) $\{\sigma_d(n) \le n-d$ *for* $n \ge d \ge 2\} \Longleftrightarrow \{\sigma_d(4d) \le 3d$ *for* $d \ge 2\}$,
(e_3) $\{\sigma_d(n) \le n-d$ *for* $n \ge d \ge 2\} \Longleftrightarrow \{\sigma_d(2d+1) - \sigma_d(2d) \le 1$ *for* $d \ge 2\}$,
(e_4) *no* $(d, 2d+1)$ *singularity* $\Longrightarrow$ {*no* $(d, 2d)$ *singularity and* $\sigma_d(n) \le n-d-1$ *for* $n > 2d \ge 4\}$,
(e_5) $\{\sigma_d(2d) = \sigma_d(2d+1)$ *for* $d \ge 2\} \Longrightarrow$ {*no* $(d, 2d)$ *singularity and* $\sigma_d(n) \le n-d-1$ *for* $n > 2d \ge 4\}$,
(e_6) $\{\sigma_d(2d) = \sigma_d(2d+1)$ *for* $d \ge 2\} \Longrightarrow$ {*square-maximal* $(d, 2d)$*-strings are, up to relabelling, unique and equal to* $a_1a_1a_2a_2a_2 \ldots a_da_d\}$.

The following basic properties of $\rho_d(n)$ were presented in [4, 8, 9] and are summarized in Proposition 5. The values of $\rho_d(n)$ are increasing when moving right along a row of the $(d, n-d)$ table, the values are increasing when moving down along a column, the values are strictly increasing when moving along descending diagonals, the values under and on the main diagonal along a column are constant. In addition, the 3 values immediately above the main diagonal are equal and differ from the value on the main diagonal by at most 1 for $d \ge 5$. Note that the main diagonal of the $(d, n-d)$ table corresponds to the values of $\rho_d(2d)$ for $d \ge 2$.

Proposition 5.
(r_1) $\rho_d(n) \le \rho_{d+1}(n+1)$ *for* $n \ge d \ge 2$,
(r_2) $\rho_d(n) \le \rho_d(n+1)$ *for* $n \ge d \ge 2$,
(r_3) $\rho_d(n) < \rho_{d+1}(n+2)$ *for* $n \ge d \ge 2$,
(r_4) $\rho_d(n) = \rho_{d+1}(n+1)$ *for* $2d \ge n \ge d \ge 2$,
(r_5) $\rho_d(n) \ge n-d$, $\rho_d(2d+1) \ge d$ *and* $\rho_d(2d+2) \ge d+1$ *for* $2d \ge n \ge d \ge 2$,
(r_6) $\rho_{d-1}(2d-1) = \rho_{d-2}(2d-2) = \rho_{d-3}(2d-3)$ *and* $0 \le \rho_d(2d) - \rho_{d-1}(2d-1) \le 1$ *for* $d \ge 5$.

The following proposition from [4] is restated in terms of singularities.

Proposition 6. *Let* $(d, 2d)$ *be the first singularity on the main diagonal, i.e. the least* d *such that* $\rho_d(2d) - \rho_{d-1}(2d-2) \ge 2$. *Then any run-maximal* $(d, 2d)$*-string does not contain a symbol occurring exactly* $2, 3, \ldots, 7$ *or* 8 *times, and must contains at least* $\lceil \frac{7d}{8} \rceil$ *singletons.*

Propositions 5 and 6 yield Theorem 7 underlining the importance of the diagonals of the $(d, n-d)$ table with respect to the conjectured upper bound $n-d$ for $\sigma_d(n)$. In particular, Theorem 7 shows that in order to prove $\rho_d(n) \le n-d$ for all n and d it is enough to prove the bound for the special case $n = 2d$ for all d, i.e. for the main diagonal of the $(d, n-d)$ table. In other words, $\rho_d(2d) \le d$ for all d implies that the maximum number $\rho(n)$ of runs over all strings of length $n \ge 3$ satisfies $\rho(n) \le n-2$. This equivalence is generalized to the special case $n = 9d$. In addition, the role played by $\rho_d(2d)$ and $\rho_d(2d+1)$ is underlined as well as the hypothesis that the run-maximal $(d, 2d)$-strings are, up to relabelling, unique.

Theorem 7.
(e_1) *no* $(d, 2d)$ *singularity* $\Longleftrightarrow \{\rho_d(n) \le n-d$ *for* $n \ge d \ge 2\}$,
(e_2) $\{\rho_d(n) \le n-d$ *for* $n \ge d \ge 2\} \Longleftrightarrow \{\rho_d(9d) \le 8d$ *for* $d \ge 2\}$,

(e_3) *$\{\rho_d(n) \le n-d$ for $n \ge d \ge 2\} \Longleftrightarrow \{\rho_d(2d+1)-\rho_d(2d) \le 1$ for $d \ge 2\}$,*
(e_4) *no $(d, 2d+1)$ singularity $\Longrightarrow \{$no $(d, 2d)$ singularity and $\rho_d(n) \le n-d-1$ for $n > 2d \ge 4\}$,*
(e_5) *$\{\rho_d(2d) = \rho_d(2d+1)$ for $d \ge 2\} \Longrightarrow \{$no $(d, 2d)$ singularity and $\rho_d(n) \le n-d-1$ for $n > 2d \ge 4\}$,*
(e_6) *$\{\rho_d(2d) = \rho_d(2d+1)$ for $d \ge 2\} \Longrightarrow \{$square-maximal $(d, 2d)$-strings are, up to relabelling, unique and equal to $a_1a_1a_2a_2a_2 \ldots a_da_d\}$.*

3 Computational substantiation of the hypothesized properties for $\sigma_d(n)$ and $\rho_d(n)$ for tractable instances

The notion of r-cover introduced in [3] was modified for the problem of distinct squares in [12] and used as a basis for a computational framework for determining $\sigma_d(n)$ values. This modification of the r-cover is referred to as the s-cover. A heuristic to obtain an efficient lower bound $\sigma_2^-(n)$ for $\sigma_2(n)$ is given in [12]. Moreover, the value $\sigma_d^-(n) = \max\ \{\sigma_{d-1}(n-1),\ \ \sigma_{d-1}(n-2)+1,\ \ \sigma_d(n-1)\}$ is used there as an efficient lower bound for $\sigma_d(n)$ for $d \ge 3$. In both cases, by *efficient* we mean the fact that for all d and n we have dealt with so far, $\sigma_d^-(n)$ either equals the actual value of $\sigma_d(n)$ or differs by 1. Furthermore, it is shown in [12] that a square-maximal string with more than $\sigma_d^-(n)$ must have an s-cover of specific properties satisfying certain density conditions, and thus a search for a square-maximal string can be limited to such strings only, significantly reducing the search space, allowing the determination $\sigma_d(n)$ for previously intractable values, see [10]. The computations so far support the hypothesis that there are no singularities on the main diagonal for $\sigma_d(n)$.

For the runs, the notion of r-cover as introduced in [3] was generalized in [1] and used as a basis for a computational framework for determining $\rho_d(n)$ values. In a similar fashion as for squares, a heuristic to obtain an efficient lower bound $\rho_2^-(n)$ for $\rho_2(n)$ is given in [3] and the value $\rho_d^-(n) = \max\ \{\rho_{d-1}(n-1), \rho_{d-1}(n-2)+1, \rho_d(n-1)\}$ is used there as an efficient lower bound for $\rho_d(n)$ for $d \ge 3$. Again, by *efficient* we mean the fact that up to now for all d and n we have dealt with, $\rho_d^-(n)$ equals the actual value of $\rho_d(n)$ or differs by 1. Furthermore, [3] shows that a run-maximal string with more than $\rho_d^-(n)$ must have an r-cover of specific properties satisfying certain density conditions, and thus a search for a run-maximal string can be limited to such strings only, significantly reducing the search space, allowing the determination of $\rho_d(n)$ for previously intractable values, see [2]. The computations so far support the hypothesis that there are no singularities on the main diagonal for $\rho_d(n)$.

The subroutine computing the number of distinct squares or runs in a string in the framework uses the C++ implementation of the algorithm introduced in [14].

4 Acknowledgments

The authors would like to thank Andrew Baker and Mei Jiang for their help with computational checking of some hypothesized properties for small instances and for determining all the known values in the respective $(d, n-d)$ tables. This work was supported by grants from the Natural Sciences and Engineering Research Council of Canada, and by the Canada

Research Chairs program, and made possible by the facilities of the Shared Hierarchical Academic Research Computing Network (`http://www.sharcnet.ca/`).

References

[1] A. Baker, A. Deza, and F. Franek. A computational framework for determining run-maximal strings. *Journal of Discrete Algorithms*. to appear.

[2] A. Baker, A. Deza, and F. Franek. Run-maximal strings. `http://optlab.mcmaster.ca/~bakerar2/research/runmax/index.html`.

[3] A. Baker, A. Deza, and F. Franek. On the structure of run-maximal strings. *Journal of Discrete Algorithms*, 14:10–14, 2012.

[4] A. Baker, A. Deza, and F. Franek. A parameterized formulation for the maximum number of runs problem. In J. Holub, B. Watson, and J. Žďárek, editors, *Festschrift for Bořivoj Melichar*, pages 102–117, Prague, Czech Republic, 2012. Czech Technical University.

[5] M. Crochemore. An optimal algorithm for computing the repetitions in a word. *Information Processing Letters*, 12:297–315, 1981.

[6] M. Crochemore, L. Ilie, and L. Tinta. The "runs" conjecture. `http://www.csd.uwo.ca/faculty/ilie/runs.html`.

[7] M. Crochemore and W. Rytter. Squares, cubes, and time-space efficient string searching. *Algorithmica*, 13:405–425, 1995.

[8] A. Deza and F. Franek. A d-step analogue for runs on strings. AdvOL Tech. Report 2010/02, McMaster University.

[9] A. Deza and F. Franek. A parameterized approach to the maximum number of distinct squares and runs in strings. AdvOL Tech. Report 2011/02, McMaster University.

[10] A. Deza, F. Franek, and M. Jiang. Square-maximal strings. `http://optlab.mcmaster.ca/~jiangm5/research/square.html`.

[11] A. Deza, F. Franek, and M. Jiang. A d-step approach for distinct squares in strings. *Lecture Notes in Computer Science*, 5029:77–89, 2011.

[12] A. Deza, F. Franek, and M. Jiang. A computational framework for determining square-maximal strings. In *Proceedings of Prague Stringology Conference 2012*, pages 112–119. Prague, Czech Republic, August 2012.

[13] A. S. Fraenkel and J. Simpson. How many squares can a string contain? *Journal of Combinatorial Theory Series A*, 82:112–120, 1998.

[14] F. Franek, M. Jiang, and C. Weng. An improved version of the runs algorithm based on crochemore's partitioning algorithm. In *Proceedings of Prague Stringology Conference 2011*, pages 98–105. Prague, Czech Republic, 2011.

[15] L. Ilie. A simple proof that a word of length n has at most $2n$ distinct squares. *Journal of Combinatorial Theory Series A*, 112:163–164, 2005.

[16] L. Ilie. A note on the number of squares in a word. *Theoretical Computer Science*, 380:373–376, 2007.

[17] C.S. Iliopoulos, D. Moore, and W.F. Smyth. A characterization of the squares in a Fibonacci string. *Theoretical Computer Science*, 172:281–291, 1997.

[18] R. Kolpakov and G. Kucherov. Finding maximal repetitions in a word in linear time. In *40th Annual Symposium on Foundations of Computer Science*, pages 596–604, 1999.

[19] M.G. Main. Detecting leftmost maximal periodicities. *Discrete Applied Mathematics*, 25:145–153, 1989.

[20] W. Matsubara, K. Kusano, A. Ishino, H. Bannai, and A. Shinohara. Lower bounds for the maximum number of runs in a string. `http://www.shino.ecei.tohoku.ac.jp/runs/`.