A parametrized formulation for the maximum number of runs problem

A. Baker, A. Deza, and F. Franek

Advanced Optimization Laboratoty
Dept. of Computing and Software
McMaster University

PRAGUE STRINGOLOGY CONFERENCE
Aug. 29-31, 2011

Outline

Motivation

Based on computational results (*Kolpakov+Kucherov* for binary alphabets up to $n = 60$, *Franek+Smyth* for all alphabets up to $n = 34$), it was hypothesized that

$$\rho(n) = \max\{r(x) \, |x| = n\} \leq n$$

where $r(x)$ denotes the number of runs in a string x.

Motivation, *cont.*

This become known as the *maximum-number-of-runs conjecture*.

Additional conjectures were put forth, for instance that the maximum is atained by a binary string, reflecting the intuitive believe that the binary case is the hardest.

Motivation, *cont.*

To this end *Deza+Franek* introduced *d*-step approach inspired by a similar approach to the Hirsch conjecture. The size of the alphabet is considered an additional parameter to the traditional length of the string.

Hence we investigate $\rho_d(n) = \max\{r(x) : |x| = n \ \&\ x \text{ has exactly } d \text{ distinct symbols}\}$

We could organize the values $\rho_d(n)$ in a 2-dimensional table where d indexes rows and n indexes columns.

For technical reasons, we organize them into a skewed table, where the columns are indexed by $n - d$ rather than n (we refer to is as $(d, n-d)$ table):

Motivation, *cont.*

$(d, n{-}d)$ table

d \ $n-d$	1	2	3	4	5	6	7	8	9	10	11
1	**1**	*1*	1	1	1	1	1	1	1	1	.
2	1	**2**	*2*	3	4	5	5	6	7	8	.
3	1	2	**3**	*3*	4	5	6	6	7	8	.
4	1	2	3	**4**	*4*	5	6	7	7	8	.
5	1	2	3	4	**5**	*5*	6	7	8	8	.
6	1	2	3	4	5	**6**	*6*	7	8	9	$\rho_6(17)$
7	1	2	3	4	5	6	**7**	*7*	8	9	.
8	1	2	3	4	5	6	7	**8**	*8*	9	.
9	1	2	3	4	5	6	7	8	**9**	*9*	.
10	1	2	3	4	5	6	7	8	9	**10**	.
11	.	.	.	.	.	$\rho_{11}(17)$	.	.	.	.	$\rho_{11}(22)$

Motivation, *cont.*

The table indicates remarkable regularities:

- non-decreasing along a row from left-to-right (*proven in D+F*)
- non-decrasing along a column from top-to-down (*proven in D+F*)
- non-decreasing along a diagonal from left-to-right (*proven in D+F*)
- constant below the diagonal (*new result here*)
- below and on the diagonal the values $\geq n - d$ (*proven in D+F*)
- all values $\leq n - d$ (*only for known values,* we conjecture for all values)

Motivation, *cont.*

- the main (red) and the second (green) diagonals are identical (*only for known values, equivalent with the conjecture*)
- the main (red) diagonal increments by 1 (*only for known values, equivalent with the conjecture*)
- the second (green) diagonal increments by 1 (*only for known values, equivalent with the conjecture*)

Motivation, *cont.*

- The value above the main diagonal is strictly greater (*only for known values, equivalent with the conjecture*)
- The structure of all run-maximal strings on the main diagonal is very simple: *aabbcc*... (*only for known values, equivalent with the conjecture*)

Motivation, *cont.*

Proven in $D{+}F$

for any $2 \leq d \leq n$:

- $\rho_d(n) \leq n - d$ iff $\rho_d(2d) = d$
- $\rho_d(2d) = \rho_d(2d+1) \Rightarrow \rho_d(2d) \leq n - d$

i.e. **the diagonals determine all.**

Immediate neigbourhood of the main diagonal

Constant under the main diagonal:

$$\rho_d(n) = \rho_{n-d}(2n - 2d) \text{ for } 2 \leq d \leq n < 2d.$$

Note that this explains the dominance of the main diagonal

Gap at most 1 just above the main diagonal:

$$\rho_d(2d) \leq \rho_{d-1}(2d - 1) + 1 \text{ for } d \geq 3.$$

if exactly 1, the conjectuer holds

Immediate neigbourhood .., *cont.*

The three immediate values above the main diagonal are identical:

$$\rho_{d-1}(2d-1) = \rho_{d-2}(2d-2) = \rho_{d-3}(2d-3)$$

for $d \geq 5$.

Strengthening the previous result:

$$\rho_d(2d) + 1 \geq \rho_d(2d+1) \Rightarrow \rho_d(2d) \leq n - d$$

Structural propertis of run-maximal strings

This section deals with structural properties of run-maximal strings on the main diagonal. The series of results is used to establish the main result of this section:

Theorem

$$\{\rho_d(n) \leq n - d \;\; \textit{for all}\; 2 \leq d \leq n\} \Leftrightarrow \{\rho_d(9d) \leq 8d \;\; \textit{forall}\; d \geq 2\}.$$

Structural propertis of run-maximal strings, *cont.*

Lemma

Let $\rho_{d'}(2d') \leq d'$ for $2 \leq d' < d$. Let x be a run-maximal string in $S_d(2d)$. Either $r(x) = \rho_d(2d) = d$ or x has at least $\lceil \frac{7d}{8} \rceil$ singletons, and no symbol occurs exactly 2, 3, $\ldots$ 8 times in x.

Structural propertis of run-maximal strings, *cont.*

Proof.

Each symbol must be a singleton or occur at least 9 times. Let $x \in S_d(2d)$ be run-maximal. Let m_1 denote the number of singletons and m_2 the number of multiply-occurring symbols of x. Then $m_1 + 9m_2 \leq 2d$ and $m_1 + m_2 = d$. The solution of the two inequalities gives $m2 \leq \frac{d}{8}$.

Let $d = 8d_1 + r$ where $0 \leq r \leq 7$.

(a) $r = 0$: $m_2 \leq d_1$, $m_1 \geq d - d_1 = 8d_1 = \lceil \frac{7d}{8} \rceil$.

(b) $r \geq 1$: then $m_1 \geq d - d_1 = 7d_1 + r$.

$\lceil \frac{7d}{8} \rceil = \lceil \frac{7 \cdot 8d_1 + 7r}{8} \rceil = \lceil 7d_1 + \frac{r}{8} \rceil = 7d_1 + 1 \leq 7d_1 + r \leq m_1$ □

Structural propertis of run-maximal strings, *cont.*

If $2 \leq d \leq 6$, then $\lceil \frac{7d}{8} \rceil = d$ and so for a run-maximal $x \in S_d(2d)$, $r(x) = d$ as otherwise it would have to consist of singletons.

For $7 \leq d \leq 15$, $\lceil \frac{7d}{8} \rceil = d - 1$ and so for a run-maximal $x \in S_d(2d)$, $r(x) = d$ as otherwise it would have to consist of singletons and one repeating letter.

Structural propertis of run-maximal strings, *cont.*

Since the values of $\rho_2(n)$ have been computed for $n \leq 60$, we can determine the values on the main diagonal for $16 \leq d \leq 23$: let $x \in S_d(2d)$, since $\lceil \frac{7d}{8} \rceil = d - 2$, either $r(x) = d$ or $r(x) = \rho_2(d+2) \leq d$.

Structural propertis of run-maximal strings, *cont.*

Lemma

Let $\rho_{d'}(2d') \le d'$ for $2 \le d' < d$. Let $x \in S_d(2d)$ be run-maximal. Either $r(x) = \rho_d(2d) = d$ or x does not contain a pair.

Lemma

Let $\rho_{d'}(2d') \le d'$ for $2 \le d' < d$. Let $x \in S_d(2d)$ be run-maximal. Either $r(x) = \rho_d(2d) = d$ or x does not contain a triple.

Structural propertis of run-maximal strings, *cont.*

Lemma

Let $\rho_{d'}(2d') \leq d'$ for $2 \leq d' < d$. Let $x \in S_d(2d)$ be run-maximal. Either $r(x) = \rho_d(2d) = d$ or x does not contain a k-tuple, $4 \leq k \leq 8$.

The rest of thge results in the full version submitted to JDA.

Conclusion

- The resuluts presented in this paper constrain the behaviour of the entries in the $(d, n-d)$ table below the main diagonal and in an immediate neighbourhood above the main diagonal.
- One of the the main contributions lies in the characterization of structural properties of the run-maximal strings on the main diagonal, giving yet another property equivalent with the maximum number of runs conjecture.

Conclusion, *cont.*

- these results provide a faster way to computationally check the validity of the conjecture for greater lengths
- they also indicate a possible way to prove the conjecture: the first counter-example on the main diagonal could not possibly have a k-tuple for any conceivable k

Thank you