McMaster University

Advanced Optimization Laboratory

Title:

Bannai et al. method proves the d-step conjecture for strings

Authors:

Antoine Deza and Frantisek Franek

AdvOL-Report No. 2015/1

January 2015, Hamilton, Ontario, Canada

Bannai et al. method proves the d-step conjecture for strings

Antoine Deza and Frantisek Franek

Abstract

Inspired by the d-step approach used for investigating the diameter of polytopes, the following d-step conjecture was introduced by Deza and Franek : the number of runs in a string of length n with d distinct symbols is at most $n - d$. Bannai et al. showed that the number of runs in a string is less than its length by mapping each run to a set of starting positions of Lyndon roots. We show that Bannai et al. method proves that the d-step conjecture for runs holds, and stress the structural properties of run-maximal strings. In particular, we show that, up to relabelling, there is a unique run-maximal string of length $2d$ with d distinct symbols. As corollary, we obtain a slight improvement of Bannai et al. bound : the number of runs in a string of length n is at most $n - 4$ for $n \geq 9$.

1 Introduction

This paper aims at combining the Bannai et al. method introduced in [3] and the d-step approach for strings introduced in [5] to highlight the structural properties of run-maximal strings. Besides yielding a slight improvement of the current best upper bound for the number of run in a string, these structural properties may provide preliminary substantiation that the number of runs in a string of length n is at most $n - \lceil \log_2 n \rceil$. The main results are presented after briefly recalling the Bannai et al. method and the d-step approach. We refer to [3] and references therein for the extensive literature on the maximum number of runs problem.

1.1 Preliminaries

Strings are indexed starting with 1, i.e. a string x of length n can be written either as $x[1\,..\,n]$ or $x[1]x[2]\ldots x[n]$. Given a string $x[1\,..\,n]$, $x[i\,..\,j]$ is a *run* with period p if $j - i \geq 2p \geq 2$, $x[i\,..\,i{+}p]$ is *primitive* – that is not a concatenation of two or more identical strings, $x[i{+}l{+}p] = x[i{+}l{+}kp]$ for $l = 0, \ldots, p-1$ and $k \geq 0$ such that $i{+}l{+}kp \leq j$, while either $i = 1$ or $x[i{-}1] \neq x[i{+}p{-}1]$ and either $j = n$ or $x[j{-}p{-}1] \neq x[j{+}1]$. Consequently, the run $x[i\,..\,j]$ is encoded by the triple (i, j, p). The *alphabet* of a string x is the set of all symbols occurring in x. A string x is a *rotation* of a string y if there are u and v such that $x = uv$ and $y = vu$, and the rotation is *trivial* when either u or v is the empty string. Let $\prec$ be a total order over the alphabet of a string x. The string x is *Lyndon with respect to* $\prec$ if x is

lexicographically strictly smaller than any of its non-trivial rotations or, equivalently, if x is lexicographically strictly smaller than any of its suffixes.

1.2 A d-step approach for polytopes and its continuous analogue

This work being inspired by the d-step approach used for investigating the Hirsch bound for the diameter of polytopes, we first recall this method and its continuous analogue.

1.2.1 A d-step approach for diameter-maximal polytopes

A polyhedron is the intersection of finitely many closed half-spaces, and a polytope is a bounded polyhedron. A (d, n)-polytope is a polytope of dimension d having n facets. The diameter $\delta(P)$ of a polytope P is the smallest integer such any pair of vertices of P can be connected by an edge-path of length at most $\delta(P)$. Let $\Delta(d, n)$ denote the maximum possible diameter over all (d, n)-polytopes. The Hirsch conjecture, posed in 1957, states that $\Delta(d, n) \leq n - d$. The values for $\Delta(d, n)$ are usually listed in a $(d, n - d)$ table where d is the index for the rows and $n - d$ the index for the columns. The following properties can be easily checked: $\Delta(d, n) \leq \Delta(d, n+1)$, $\Delta(d, n) < \Delta(d+1, n+2)$, and $\Delta(d, n) \leq \Delta(d+1, n+1)$ for $n \geq d \geq 2$; and that $\Delta(d, n) = \Delta(d + 1, n + 1)$ for $2d \geq n \geq d \geq 2$. In other words, the maximum of $\Delta(d, n)$ within a column is achieved on the main diagonal and all values below a value on the main diagonal are equal to that value. The role played by the main diagonal of the $(d, n - d)$ table was underlined by Klee and Walkup [10] who showed the equivalency between the Hirsch conjecture and the d-step conjecture stating that $\Delta(d, 2d) \leq d$ for all $d \geq 2$. Note that the d-cube is a $(d, 2d)$-polytope having diameter d and therefore $\Delta(d, 2d) \geq d$ for any d. In other words, the string $a_1a_1a_2a_2 \ldots a_da_d$ can be viewed as an analogue of the d-cube. The Hirsch conjecture was disproved by Santos [11] by exhibiting a violation on the main diagonal with $(d, n) = (43, 86)$; that is, Santos constructed a polytope in dimension 43 with 86 facets and a diameter of at least 44.

1.2.2 A d-step approach for curvature-maximal polytopes

A continuous analogue of the Hirsch conjecture was proposed Deza et al. [7] by considering the currently most computationally successful algorithms for linear optimization; i.e., the simplex and central-path following primal-dual interior point methods. The value of $\Delta(d, n)$ provides a lower bound for the number of iterations of simplex methods for the worst case behaviour. The curvature of a polytope, defined as the largest possible total curvature of the associated central path, can be regarded as the continuous analogue of its diameter. Considering the largest curvature $\Lambda(d, n)$, Deza et al. [7] proved the following continuous analogue of the equivalence between the Hirsch conjecture and the d-step conjecture: if $\Lambda(d, 2d) = \mathcal{O}(d)$ for all d, then $\Lambda(d, n) = \mathcal{O}(n)$. Using a tropical linear optimization setting, Allamigeon et al. [1] constructed an exponential counterexample to the continuous analogue of the polynomial Hirsch conjecture by exhibition a polytope in dimension d with $3d/2$ facets and a curvature of at least $2^{d/2}$.

1.3 A d-step approach for strings

A d-step formulation for strings was proposed in [5] by considering $\rho_d(n)$; that is, the maximum number of runs over all strings of length n with d distinct symbols. It was shown in [5] that $\rho_d(n)$ and $\Delta(d,n)$ exhibit similarities and, in particular, that $\rho_d(n) \leq \rho_d(n+1)$, $\rho_d(n) < \rho_{d+1}(n+2)$, $\rho_d(n) \leq \rho_{d+1}(n+1)$ for $n \geq d \geq 2$; and $\rho_d(n) = \rho_{d+1}(n+1)$ for $2d \geq n \geq d \geq 2$. Consequently, the authors proposed to present the value of $\rho_d(n)$ in a $(d, n-d)$ table where d is the index for the rows and $n-d$ the index for the columns, see Table 1 for a 20×20 portion of the $(d, n-d)$ table for $\rho_d(n)$. In other words, the properties remarked in [5] show that the maximum of $\rho_d(n)$ within a column is achieved on the main diagonal and all values below a value on the main diagonal are equal to that value. The main results and conjectures yielded by the d-step approach for strings are given in proposition 1 and Conjecture 2.

Proposition 1 ([5])**.** *Let $\rho_d(n)$ be the maximum of runs over all strings of length n with d distinct symbols, then*

(*i*) $\rho_d(n) \leq n-d \iff \rho_d(2d) \leq d$,

(*ii*) $\rho_d(2d) = \rho_d(2d+1) \implies$ *the string $a_1a_1a_2a_2\cdots a_da_d$ is, up to a permutation of the alphabet, the unique run-maximal string of length $2d$ with d distinct symbols,*

(*iii*) $\rho_d(2d+1) = \rho_d(2d+2) = \rho_d(2d+3)$.

Conjecture 2 ([5])**.** *A string of length n with d distinct symbols has at most $n-d$ runs; that is, $\rho_d(n) \leq n-d$.*

Note that the d-step formulation was used in [2] to determine $\rho_d(n)$ for previously intractable values of d and n. In particular, the maximum number of runs has been determined for binary strings of length up to 74.

1.4 Bannai et al. method for strings

The main idea of Bannai et al. method is to map the runs of a string $x = x[1\,..\,n]$ to mutually disjoint subsets of indices of x. Given a total order $\prec$ of the alphabet of a string x, let $\prec^{-1}$ denote the reverse order. Consider a run $t = (i,j,p)$ in x. For $i \leq k \leq j-p$, all the substrings $x[k\,..\,k{+}p]$ are primitive, and at least one of them is Lyndon with respect to $\prec$, and at least one of them is Lyndon with respect to $\prec^{-1}$. This observation motivated the notion of $\mathrm{L}^{\prec}$-roots for a run $t = (i,j,p)$:

Case 1: $j = n$ or $x[j{-}p{+}1] \succ x[j{+}1]$; then every factor $x[k\,..\,k{+}p], i < k \leq j{-}p$, that is Lyndon with respect to $\prec$, is necessarily a maximal Lyndon factor and is referred to as an L-root of t.

Case 2: $j < n$ and $x[j{-}p{+}1] \prec x[j{+}1]$; then every factor $x[k\,..\,k{+}p], i < k \leq j{-}p$, that is Lyndon with respect to $\prec^{-1}$, is necessarily a maximal Lyndon factor and is referred to as an L-root of t.

Note that $x[j-p+1] \neq x[j+1]$ as otherwise t could be extended by one position to the right, contradicting the maximality condition from the definition of run. Thus, exactly one of the two cases holds. If the considered order is clear from the context, we simply use the term L-root. Note a slight modification of Bannai et al. terminology: our definition of L-roots excludes Lyndon subwords, of the length of the period of t, starting at the beginning of t. A run t is mapped to the set $\mathrm{Beg}(t)$ of the starting positions of all its L-roots. Bannai et al. [3] showed that $\mathrm{Beg}(t_1) \cap \mathrm{Beg}(t_2) = \emptyset$ for distinct runs t_1 and t_2; that is, the L-roots of two distinct runs never start at the same position – recall that L-roots of a run never start at the run's beginning. This mapping implies that the number of runs of a string is at most its length. In addition, since no L-root starts at position 1, the number of runs of a string is strictly less than its length.

		$n-d$																		
		2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
d	2	**2**	*2*	*3*	*4*	*5*	***5***	***6***	***7***	***8***	***8***	***10***	10	11	12	13	14	15	15	16
	3	2	**3**	*3*	*4*	*5*	*6*	***6***	***7***	***8***	***9***	***10***	***11***	11	12	13	14	15	16	16
	4	2	3	**4**	*4*	*5*	*6*	*7*	***7***	***8***	***9***	***10***	***11***	***12***	12	13	14	15	16	17
	5	2	3	4	**5**	*5*	*6*	*7*	*8*	***8***	***9***	***10***	***11***	***12***	***13***	13	14	15	16	17
	6	2	3	4	5	**6**	*6*	*7*	*8*	*9*	***9***	***10***	***11***	***12***	***13***	***14***	14	15	16	17
	7	2	3	4	5	6	**7**	*7*	*8*	*9*	*10*	***10***	***11***	***12***	***13***	***14***	***15***	15	16	17
	8	2	3	4	5	6	7	**8**	*8*	*9*	*10*	*11*	***11***	***12***	***13***	***14***	***15***	***16***	16	17
	9	2	3	4	5	6	7	8	**9**	*9*	*10*	*11*	*12*	***12***	***13***	***14***	***15***	***16***	***17***	17
	10	2	3	4	5	6	7	8	9	**10**	*10*	*11*	*12*	*13*	***13***	***14***	***15***	***16***	***17***	***18***
	11	2	3	4	5	6	7	8	9	10	**11**	*11*	*12*	*13*	*14*	***14***	***15***	***16***	***17***	***18***
	12	2	3	4	5	6	7	8	9	10	11	**12**	*12*	*13*	*14*	*15*	***15***	***16***	***17***	***18***
	13	2	3	4	5	6	7	8	9	10	11	12	**13**	*13*	*14*	*15*	*16*	***16***	***17***	***18***
	14	2	3	4	5	6	7	8	9	10	11	12	13	**14**	*14*	*15*	*16*	*17*	***17***	***18***
	15	2	3	4	5	6	7	8	9	10	11	12	13	14	**15**	*15*	*16*	*17*	*18*	***18***
	16	2	3	4	5	6	7	8	9	10	11	12	13	14	15	**16**	*16*	*17*	*18*	*19*
	17	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	**17**	*17*	*18*	*19*
	18	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	**18**	*18*	*19*
	19	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	**19**	*19*
	20	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	**20**

Table 1: $(d, n-d)$ table for $\rho_d(n)$ with $2 \leq d \leq 20$ and $2 \leq n-d \leq 20$

2 Bannai et al. method and the d-step conjecture for strings

The authors contacted Hideo Bannai in the summer of 2014 to point out the d-step conjecture for runs and a proof of $\rho_d(n) \leq n-d$ was subsequently added to [3] in January 2015 along with $\rho_d(n) \leq n-d-1$ for $n \geq 2d+1$ and the directly implied bound : the number of runs in a string of length n is at most $n-3$ for $n \geq 5$. Besides providing alternate proofs for Lemmas 9

and 10, see [3], we show additional properties and a slight strengthening of the current best bound for the number of runs. We wish to point out related results of Crochemore and Mercaş [4] and Fischer et al. [8] that both build up on Bannai et al. method to bound the number of runs. Note that the notions of *multiplicities of Lyndon roots for cubic runs* in [4] and of *overloaded* in [8] and what we call *redundant* overlap. Focusing on binary strings, Fischer et al. showed that $\rho_2(n) \leq \lceil 22n/23 \rceil$.

2.1 Main results

The following propositions are obtained by combining the d-step approach and the Bannai et al. method. Proposition 3 is a slight improvement of the current best upper bound for the number of run in a string of length n, Proposition 4 illustrates that, in contrast with polytopes, the d-step conjecture holds for strings and the uniqueness of run-maximal stings whose length is twice its number of symbols, and Proposition 5 deals with strings whose length is at most twice its number of symbols plus 10.

Proposition 3. *Let $\rho(n)$ be the maximum of runs over all strings of length n, then $\rho(n) \leq n-4$ for $n \geq 9$.*

Proof. proposition 3 is a direct corollary of Lemma 9, respectively Lemma 10 and Lemma 11, for $d \geq 4$, respectively for $d = 3$ and $d = 2$. □

Proposition 4. *The string $a_1a_1a_2a_2 \cdots a_da_d$ is, up to a permutation of the alphabet $\{a_1, a_2, \ldots, a_d\}$, the unique run-maximal string of length $2d$ with d distinct symbols.*

Proof. proposition 4 is a direct corollary of item (ii) of Proposition 1 and item (ii) of Proposition 5. □

Proposition 5. *Let $\rho_d(n)$ be the maximum of runs over all strings of length n with d distinct symbols, then*

(i) $\rho_d(n) = n - d$ *for* $2d \geq n \geq 2$,

(ii) $\rho_d(n) = n - d - 1$ *for* $2d + 4 \geq n \geq 2d + 1 \geq 5$,

(iii) $\rho_d(n) = n - d - 2$ *for* $2d + 10 \geq n \geq 2d + 5 \geq 9$ *except for* $(d, n) = (2, 13)$ *as* $\rho_2(13) = 8$.

Proof. The fact that $\rho_{d+1}(d + 2) > \rho_d(n)$ and $\rho_2(4) = 2$ implies that $\rho_d(2d) \geq d$. Thus, Lemma 9 implies that $\rho_d(2d) = d$; that is, item (i) holds as $\rho_d(n) = n - d$ for $2d \geq n \geq 2$ since $\rho_d(n) = \rho_{d+1}(n + 1)$ for $2d \geq n \geq d \geq 2$. Similarly, the fact that $\rho_{d+1}(d + 2) > \rho_d(n)$ and $\rho_2(8) = \rho_2(7) + 1 = \rho_2(6) + 2 = \rho_2(5) + 3 = 5$ implies that $\rho_d(n) \geq n - d - 1$ for $2d + 4 \geq n \geq 2d + 1 \geq 5$. Thus, Lemma 10 implies that $\rho_d(n) = n - d - 1$ for $2d + 4 \geq n \geq 2d + 1 \geq 5$; that is, item (ii) holds. The proof for item (iii) is almost the same as for item (ii) except that Lemma 11 is used instead of Lemma 10 and the base values are $\rho_2(14) = \rho_3(14) + 1 = \rho_2(12) + 2 = \rho_2(11) + 3 = \rho_2(10) + 4 = \rho_2(9) + 5 = 10$. □

See Table 1 for an illustration of Proposition 5 where the main diagonal corresponding to $n = 2d$ is in bold, the diagonals corresponding to $2d < n \leq 2d + 4$ are in italic, and the diagonals corresponding to $2d + 4 < n \leq 2d + 10$ are in bold italic. Note that these values are computationally intractable for non-trivial (d, n); i.e. to compute the maximum numbers of runs over all strings of length 35 and having 15 symbols is beyond current computational means while Proposition 5 shows that this number is 18.

Remark 6. *A generalization of the proof of Lemma 11 to higher values of $n - 2d$ may substantiate the hypothesis that $\rho_d(2d+k)-d$ is a step function independent of d. Proposition 5 might be considered as a preliminary substantiation that the number of runs in a string of length n with d symbols is at most $n - d - \lceil \log_2 \lceil (n+4-2d)/4 \rceil \rceil$ for $n \geq 2d \geq 4$ as hypothesized in [5]. Assuming that run-maximal strings include binary ones and thus setting $d = 2$ in the previous inequality, the number of runs in a string of length n was hypothesized in [5] to be at most $n - \lceil \log_2 n \rceil$. A d-step approach was introduced for distinct primitively rooted squares in strings as well as hypothesized upper bounds [5]. We recall the bound of Fraenkel and Simpson [9] was strengthened in [6] to : the number of distinct primitively rooted squares in a string of length n is at most $\lfloor 11n/6 \rfloor$.*

2.2 Observations and auxiliary lemmas

Observation 7 points out a few straightforward properties of L-roots.

Observation 7. *Given a string x over the alphabet $\{a_1, \ldots, a_d\}$,*

- (*i*) *no L-root starts at position 1;*
- (*ii*) *if an L-root of a run t starts at the position of a symbol $\boldsymbol{a_i}$:*

  case (1)*: there is a farther occurrence of $\boldsymbol{a_i}$ and $x = \ldots (u\underline{\boldsymbol{a_i} v})(\underline{u\widehat{a_i} v}) \cdots (u a_i v)\mu \ldots$*

  $$\text{and for any } \nu \in u, v, \begin{cases} a_i \preceq \nu & if\ \mu \prec u[1]\ \ (i.e.\ \prec\ is\ used) \\ \nu \preceq a_i & if\ \mu \succ u[1]\ \ (i.e.\ \prec^{-1}\ is\ used) \end{cases}$$

  where $\widehat{a_i}$ is the farther copy of $\boldsymbol{a_i}$, the L-root is underlined, and $(\ )(\ )\cdots(\ )$ indicates the rightmost repetition of the run t containing $\boldsymbol{a_i}$ in its generator,

  case (2)*: there is a previous occurrence of $\boldsymbol{a_i}$ and $x = \ldots (\widehat{a_i} w)(\underline{\boldsymbol{a_i} w}) \cdots (a_i w)\mu \ldots$*

  $$\text{and for any } \nu \in w, \begin{cases} a_i \preceq \nu & if\ \mu \prec \boldsymbol{a_i}\ \ (i.e.\ \prec\ is\ used) \\ \nu \preceq a_i & if\ \mu \succ \boldsymbol{a_i}\ \ (i.e.\ \prec^{-1}\ is\ used) \end{cases}$$

  where $\widehat{a_i}$ is the previous copy of $\boldsymbol{a_i}$, the L-root is underlined, and $(\ )(\ )\cdots(\ )$ indicates the rightmost repetition of the run t starting with $\widehat{a_i}$.

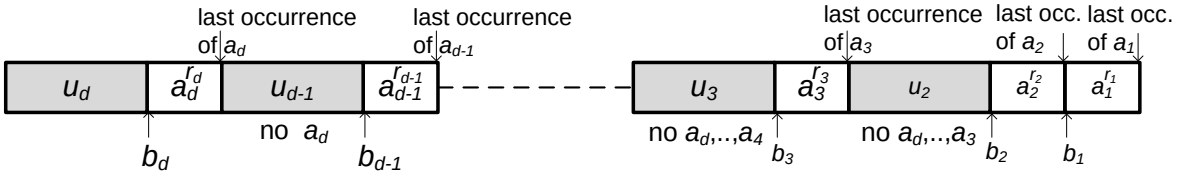
- (*iii*) *if a symbol a_i occurs only once, no L-root starts at the position of a_i;*
- (*iv*) *if a symbol a_i occurs exactly twice, at most one L-root starts at the positions of the occurrence of a_i since they belong to at most one run;*
- (*v*) *if a symbol a_i occurs exactly three times, at most two L-root starts at the positions of the occurrences of a_i since they belong to at most two runs.*

Observation 7 leads to the following notion of redundancy : if a run has $k \geq 2$ L-roots, we consider that $k-1$ of them are redundant. For example, a_i^{k+1} has k L-roots and $k-1$ of them are redundant. The number of runs in a string is clearly at most the number of its non-redundant L-roots. In the rest of the paper, a string of length n with d distinct symbols is referred to as a (d,n)-string, and the number of runs in a string x is denoted by $r(x)$.

Definition 8. *Given a string x containing the symbols $\{a_1,\ldots,a_d\}$ ordered by $\prec$, the string is $\prec$-*labelled *if:*

(*i*) $a_1 \prec a_2 \prec \cdots \prec a_{d-1} \prec a_d$,
(*ii*) $x = u_d a_d^{r_d} u_{d-1} a_{d-1}^{r_{d-1}} \ldots u_2 a_2^{r_2} a_1^{r_1}$ *where u_k does not contain any symbol from $\{a_{k+1},\ldots,a_d\}$ and $r_k \geq 1$ for $k \in 1\,..\,d$,*
(*iii*) *Set $b_k = 1 + (|u_k| + \cdots + |u_d|) + (r_{k+1} + \cdots + r_d)$ for $k \in 1\,..\,d$. Then $x[b_k{-}1] \prec a_k$ for $k \in 1\,..\,d{-}1$, and $x[b_d{-}1] \prec a_d$ if $d > 1$.*

see the following illustration:



Note that b_k is the position of the beginning of the *block of last occurrence of* a_k, i.e. $a_k^{r_k}$. Note also that $|u_k|$ may be possibly zero and that r_k is maximal as $x[b_k{-}1] \prec a_k$ implies that the preceding symbol, if any, differs from a_k. Clearly, any string x can be $\prec$-labelled by a simple act of relabelling the alphabet symbols. The structure of the $\prec$-labelled strings indicates places where an L-root cannot start and where redundant L-roots might occur. The first places to look for positions with no L-roots are, of course, the beginnings of the blocks, i.e. the b_k's.

Lemma 9. *Let $\rho_d(n)$ be the maximum of runs over all (d,n)-strings, then $\rho_d(n) \leq n-d$ for $n \geq 2d \geq 2$.*

Proof. Consider a $\prec$-labelled run-maximal (d,n)-string x, that is x is a $\prec$-labelled string with $\rho_d(n)$ runs. We show that the number of L-roots of $x = u_d a_d^{r_d} u_{d-1} a_{d-1}^{r_{d-1}} \ldots u_2 a_2^{r_2} a_1^{r_1}$ is at most $n-d$, and thus $\rho_d(n) \leq n-d$, by remarking that no L-root starts at b_k for any $k \in 1\,..\,d$. Assume by contradiction that an L-root starts at b_{k_0} for some k_0. The L-root cannot be the case (ii) (1) of Observation 7 as there is no farther occurrence of a_{k_0} past the block $a_{k_0}^{r_{k_0}}$. Thus, the L-root must be the case (ii) (2) must true. Since $x[b_{k_0}{-}1] \prec x[b_{k_0}]$, the $\prec^{-1}$ order must be used, but a_{k_0} does not precede any symbol past $a_{k_0}^{r_{k_0}}$ – hence a contradiction. □

Another natural place to look for no L-root is the beginning of the string – all we need to guarantee is that b_d, the beginning of the first block $a_{a_d}^{r_{a_d}}$, does not coincide with the beginning of the string. The sufficient length of the string is enough to guarantee it.

Lemma 10. *Let $\rho_d(n)$ be the maximum of runs over all (d,n)-strings, then $\rho_d(n) \leq n-d-1$ for $n \geq 2d+1 \geq 5$.*

Proof. Consider a $\prec$-labelled run-maximal (d,n)-string x, that is x is a $\prec$-labelled string with $\rho_d(n)$ runs. We need to show that, besides the d positions corresponding to the b_k's, there is at least one position with no or a redundant L-root.

The case when $|u_k| = 0$ for $k = 1, \ldots, d$. Then $x = a_d^{r_d} a_{d-1}^{r_{d-1}} \ldots a_2^{r_2} a_1^{r_1}$ and since $n \geq 2d+1$, there is a k_0 such that $r_{k_0} \geq 3$ and so at least one L-root is redundant.

The case when $|u_{k_0}| \geq 1$ for some k_0. We show that without loss of generality we can assume that $k_0 = d$. Assume otherwise that $|u_d| = 0$; that is, $x = a_d^{r_d} u_{d-1} a_{d-1}^{r_{d-1}} \ldots u_2 a_2^{r_2} a_1^{r_1}$. Since $r(x) = r(x[r_d{+}1\,..\,n]x[1\,..\,r_d])$, we can move $a_d^{r_d}$ to the end of the string and relabel the symbols and repeat this process until we run into the first k_0 such that $|u_{k_0}| \geq 1$.

Thus, we have a run-maximal string with non-empty u_d and so $b_d > 1$ and so the number of positions with no L-roots is at least $d+1$. □

Lemma 11. *Let $\rho_d(n)$ be the maximum of runs over all (d,n)-strings, then $\rho_d(n) \leq n-d-2$ for $n \geq 2d+5 \geq 9$.*

Proof. Consider a $\prec$-labelled run-maximal (d,n)-string x, that is x is a $\prec$-labelled string with $\rho_d(n)$ runs. Besides the d positions b_k's, we need to exhibit at least two additional positions with no or redundant L-roots.

The case when $|u_k| = 0$ for $k = 1, \ldots, d$. Then is $x = a_d^{r_d} a_{d-1}^{r_{d-1}} \ldots a_2^{r_2} a_1^{r_1}$ and since $n \geq 2d+5$, there are k_0, k_1, k_2, k_3 and k_4 such that $r_{k_0} + r_{k_1} + r_{k_2} + r_{k_3} + r_{k_4} \geq 15$ and so at least 5 L-roots are redundant.

The case when $|u_{k_0}| \geq 1$ for some k_0. As in the proof of Lemma 10, we can assume that $k_0 = d$; that is $|u_d| \geq 1$, and thus $1 < b_d < b_{d-1} < \cdots < b_2 < b_1$ are positions with no L-roots. Therefore, to complete the proof, we need to exhibit one additional position with no or a redundant L-root.

We can assume that $r_k \leq 2$ for all $k \in 1\,..\,d$ as otherwise one additional L-root would be redundant and the proof would be completed. Define m as the smallest k such that $|u_k| \geq 1$, i.e. $x = u_d a_d^{r_d} \ldots u_m a_m^{r_m} a_{m-1}^{r_{m-1}} \ldots a_2^{r_2} a_1^{r_1}$ – note that such $m \leq d$ must exists as $|u_d| \geq 1$. Let a_ℓ the last symbol of u_m and note that $\ell \leq m-1$: if $\ell > m$, then a_ℓ would be occurring past the block of its last occurrence $a_\ell^{r_\ell}$ which is to the left of u_m, which is not possible; if $\ell = m$, then it should be a part of the block $a_m^{r_m}$. Thus, $x = \ldots \mu \boldsymbol{a_\ell^i} a_m^{r_m} a_{m-1}^{r_{m-1}} \ldots a_{\ell+1}^{r_{\ell+1}} a_\ell^{r_\ell} a_{\ell-1}^{r_{\ell-1}} \ldots a_2^{r_2} a_1^{r_1}$ for some $\mu \neq a_\ell$ and some $i \geq 1$. Since $i \geq 3$ would give at least one redundant L-root in one of the positions of $\boldsymbol{a_\ell^i}$, we can assume that i is at most 2.

We show that either there is no L-root at the beginning of $\boldsymbol{a_\ell^i}$ or there is a position before the beginning of $\boldsymbol{a_\ell^i}$ with no L-root. To do so, we assume that there is an L-root at the beginning of $\boldsymbol{a_\ell^i}$ and find a prior position with no L-root.

Consider that the L-root at the beginning is of $\boldsymbol{a_\ell^i}$ is of type (ii) (2) of Observation 7. Since the L-root starts with $\boldsymbol{a_\ell^i}$ followed by a_m and $\ell < m$, and so $a_\ell \prec a_m$, it follows that it is Lyndon with respect to $\prec$, and so the trailing square of the run must be followed by a symbol smaller than a_ℓ and so it must reach past the block $a_\ell^{r_\ell}$. If it reached past $a_{\ell-1}^{r_{\ell-1}}$, the suffix starting with $a_{\ell-1}$ would be lexicographically smaller than the L-root that starts with $\boldsymbol{a_\ell}$, a contradiction. So it must actually end inside the block $a_{\ell-1}^{r_{\ell-1}}$. But then again the suffix starting with $a_{\ell-1}$ would be lexicographically smaller than the L-root, giving again a contradiction.

Therefore, the L-root must be of type (ii) (1) of Observation 7. There are only 4 possibilities for an L-root to occur at the beginning of $\boldsymbol{a_k^i}$ (the root is underlined):

case $(i, r_\ell) = (1, 1)$: $x = \underbrace{\ldots.(a_m^{r_m} a_{m-1}^{r_{m-1}} .. \boldsymbol{a_{\ell+1}^{r_{\ell+1}}}\, \underline{\boldsymbol{a_\ell}})(a_m^{r_m} a_{m-1}^{r_{m-1}} .. \underline{a_{\ell+1}^{r_{\ell+1}} a_\ell})}_{u_m} a_{\ell-1}^{r_{\ell-1}} .. a_2^{r_2} a_1^{r_1}$

case $(i, r_\ell) = (1, 2)$: $x = \underbrace{\ldots.(a_m^{r_m} a_{m-1}^{r_{m-1}} .. \boldsymbol{a_{\ell+1}^{r_{\ell+1}}}\, \underline{\boldsymbol{a_\ell}})(a_m^{r_m} a_{m-1}^{r_{m-1}} .. \underline{a_{\ell+1}^{r_{\ell+1}} a_\ell})}_{u_m} a_\ell a_{\ell-1}^{r_{\ell-1}} .. a_2^{r_2} a_1^{r_1}$

case $(i, r_\ell) = (2, 1)$: $x = \underbrace{\ldots.(a_\ell a_m^{r_m} a_{m-1}^{r_{m-1}} .. \boldsymbol{a_{\ell+1}^{r_{\ell+1}}}\, \underline{\boldsymbol{a_\ell}})(\boldsymbol{a_\ell} a_m^{r_m} a_{m-1}^{r_{m-1}} .. \underline{a_{\ell+1}^{r_{\ell+1}} a_\ell})}_{u_m} a_{\ell-1}^{r_{\ell-1}} .. a_2^{r_2} a_1^{r_1}$

case $(i, r_\ell) = (2, 2)$: $x = \underbrace{\ldots.(a_m^{r_m} a_{m-1}^{r_{m-1}} .. \boldsymbol{a_{\ell+1}^{r_{\ell+1}}}\, \underline{\boldsymbol{a_\ell a_\ell}})(a_m^{r_m} a_{m-1}^{r_{m-1}} .. \underline{a_{\ell+1}^{r_{\ell+1}} a_\ell a_\ell})}_{u_m} a_{\ell-1}^{r_{\ell-1}} .. a_2^{r_2} a_1^{r_1}$

We show that in all four cases, there is no L-root at the beginning of the first $\boldsymbol{a_{\ell+1}^{r_{\ell+1}}}$ (denoted in bold). First note that if there were an L-root, it would have to contain the $\boldsymbol{a_\ell}$ that follows the block $\boldsymbol{a_{\ell+1}^{r_{\ell+1}}}$, so the L-root would have to be determined by $\prec^{-1}$. This excludes case (ii)(1) of Observation 7 as the next available $a_{\ell+1}$ is not followed by a bigger symbol. Thus, if there were an L-root, it would have to be the case (ii) (2) of Observation 7 and then the L-root would have the same length as the L-root starting at the beginning of $\boldsymbol{a_\ell^i}$ as it would span from the first $\boldsymbol{a_{\ell+1}^{r_{\ell+1}}}$ to the next $a_{\ell+1}^{r_{\ell+1}}$, hence they would both belong to the same run as they overlap - which is impossible.

We established that there cannot be an L-root starting at the beginning of $\boldsymbol{a_{\ell+1}^{r_{\ell+1}}}$, and so the last step we need to consider is to show that the beginning of $\boldsymbol{a_{\ell+1}^{r_{\ell+1}}}$ does not coincide with the beginning of the string. The only cases for which the beginning of $\boldsymbol{a_{\ell+1}^{r_{\ell+1}}}$ is the beginning of the string correspond to $(i, r_\ell) = (1, 1)$, $(1, 2)$, or $(2, 2)$ and when $\ell + 1 = m = d$, but these cases cannot occur if $n \geq 2d + 5$ as detailed below:

case $(i, r_\ell) = (1, 1)$ and $\ell + 1 = m = d$: $x = \underbrace{(a_d^{r_d} \underline{\boldsymbol{a_{d-1}}})(a_d^{r_d} a_{d-1})}_{u_d} a_{d-2}^{r_{d-2}} \ldots a_2^{r_2} a_1^{r_1}$

implying $n = 1 + 2r_d + r_{d-1} + \cdots + r_1 \leq 3 + 2d$,

case $(i, r_\ell) = (1, 2)$ and $\ell + 1 = m = d$: $x = \underbrace{(a_d^{r_d}\boldsymbol{a_{d-1}})}_{u_d}(a_d^{r_d}a_{d-1})a_{d-1}a_{d-2}^{r_{d-2}}\dots a_2^{r_2}a_1^{r_1}$

implying $n = 1 + 2r_d + r_{d-1} + \dots + r_1 \leq 3 + 2d$,

case $(i, r_\ell) = (2, 2)$ and $\ell + 1 = m = d$: $x = \underbrace{(a_d^{r_d}\boldsymbol{a_{d-1}a_{d-1}})}_{u_d}(a_d^{r_d}a_{d-1}a_{d-1})a_{d-2}^{r_{d-2}}\dots a_2^{r_2}a_1^{r_1}$

implying $n = 2 + 2r_d + r_{d-1} + \dots + r_1 \leq 4 + 2d$. □

Acknowledgments. The authors wish to thank Maxime Crochemore and Robert Mercaş for informing us of the preprints [4, 8]. This work was supported by the Natural Sciences and Engineering Research Council of Canada, the Digiteo Chair C&O program, and by the Canada Research Chairs program.

References

[1] X. Allamigeon, P. Benchimol, S. Gaubert, and M. Joswig. Long and winding central paths. *arXiv:1405.4161*, 2014.

[2] A. Baker, A. Deza, and F. Franek. A computational framework for determining run-maximal strings. *Journal of Discrete Algorithms*, 20:43 – 50, 2013.

[3] H. Bannai, T. I, S. Inenaga, Y. Nakashima, M. Takeda, and K. Tsuruta. The "Runs" Theorem. *arXiv:1406.0263v6*, 2015.

[4] M. Crochemore and R. Mercaş. Fewer runs than word length. *arXiv:412.4646v1*, 2014.

[5] A. Deza and F. Franek. A d-step approach to the maximum number of distinct squares and runs in strings. *Discrete Applied Mathematics*, 163:268–274, 2014.

[6] A. Deza, F. Franek, and A. Thierry. How many double squares can a string contain? *Discrete Applied Mathematics*, pages 52–69, 2015.

[7] A. Deza, T. Terlaky, and Y. Zinchenko. A continuous d-step conjecture for polytopes. *Discrete and Computational Geometry*, 41:318–327, 2009.

[8] J. Fischer, Š. Holub, T. I, and M. Lewenstein. Beyond the Runs theorem. *arXiv:1502.04644v2*, 2015.

[9] A.S. Fraenkel and J. Simpson. How many squares can a string contain? *Journal of Combinatorial Theory, Series A*, 82(1):112–120, 1998.

[10] Victor Klee and David W. Walkup. The d-step conjecture for polyhedra of dimension $d < 6$. *Acta Mathematica*, 117:53–78, 1967.

[11] F. Santos. A counterexample to the Hirsch conjecture. *Annals of Mathematics*, 176:383–412, 2012.