Discrete Applied Mathematics 217 (2017) 488–494

Contents lists available at ScienceDirect

Discrete Applied Mathematics

journal homepage: www.elsevier.com/locate/dam

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

Antoine Deza [*], Frantisek Franek

Advanced Optimization Laboratory, Department of Computing and Software, McMaster University, Hamilton, Ontario, Canada

ARTICLE INFO

Article history:
Received 30 April 2015
Received in revised form 19 September 2016
Accepted 27 September 2016
Available online 20 October 2016

Keywords:
Strings
Runs
Lyndon root
d-step approach

ABSTRACT

Inspired by the d-step approach used for investigating the diameter of polytopes, Deza and Franek introduced the d-step conjecture for runs stating that the number of runs in a string of length n with exactly d distinct symbols is at most $n-d$. Bannai et al. showed that the number of runs in a string is at most $n-3$ for $n \geq 5$ 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. The number of runs in a string of length n is shown to be at most $n-4$ for $n \geq 9$.

1. Introduction

A *run* in a string $x[1..n]$ is a succinct notion of a maximal repetition. A run is usually encoded by in a triple (s, e, p) such that the substring $x[s..e]$ has a minimal period of p, $x[s..s+p-1]$ is primitive, $s+2p-1 \leq e$ and so $x[s..s+p-1]$ repeats at least twice, and either $s = 1$ or $x[s-1] \neq x[s+p-2]$ and either $e = n$ or $x[e-p] \neq x[e+1]$, i.e. the periodicity can be extended neither to the left nor to the right. Thus, s encodes the start of the run, e the end of the run, and p its period. The substring $x[s..s+p-1]$ is the *root* of the run. For example, in the string $a\underline{abababa}a$, the underlined run is encoded by $(2, 8, 2)$, and its root ab is repeated 4 times, with the last repeat being incomplete. Runs, equal up to translation, may occur more than once in a string. For example, in the string $a\underline{abababa}aaaa\underline{abababa}a$, the underlined runs encoded by $(2, 8, 2)$ and $(13, 19, 2)$ are both counted.

Crochemore [4] showed in 1981 that the order of the number of maximal repetitions in a string of length n is $\mathcal{O}(n \log n)$. In 1999, Kolpakov and Kucherov [18] showed that the order of the largest number $\rho(n)$ of runs over all strings of length n is $\mathcal{O}(n)$, without exhibiting an explicit constant, and conjectured that $\rho(n) \leq n$. Rytter [23,24] determined such a constant in 2006, and the following years witnessed a tightening of the lower and upper bounds for the limit of $\rho(n)/n$, see [5,6,14–16,19,21,20,22]. In 2015, the conjecture was proven by Bannai et al. [3] who showed that $\rho(n) \leq n-1$, and $\rho(n) \leq n-3$ for $n \geq 5$, by using starts of specific Lyndon roots of each run; that is by mapping all runs to mutually disjoint subsets of the string indices.

Deza and Franek investigated the largest number $\rho_d(n)$ of runs over all strings of length n with exactly d distinct symbols. Similarities between $\rho_d(n)$ and the largest diameter $\Delta(d, n)$ over all polytopes of dimension d having n facets triggered the formulation of the d-step conjecture for strings stating that $\rho_d(n) \leq n-d$, see [8]. The proposed d-step approach proved that the following statements are equivalent $\{\rho_d(n) \leq n-d$ for all d and $n\}$, $\{\rho_d(2d) \leq d$ for all $d\}$, and $\{\rho_d(2d)$ is achieved for all d by a, up to relabelling, unique string $\}$. Considering binary strings, Fischer et al. [12] showed that $\rho_2(n) \leq \lceil 22n/23 \rceil$. While it is widely believed that $\rho_{d+1}(n) \leq \rho_d(n)$, and thus that $\rho(n) = \rho_2(n)$, no such results are known.

* Corresponding author.
E-mail addresses: deza@mcmaster.ca (A. Deza), franek@mcmaster.ca (F. Franek).

http://dx.doi.org/10.1016/j.dam.2016.09.036

Some properties concerning maximal strings are rather counterintuitive. For example, consider the largest number $\sigma_d(n)$ of distinct primitively rooted squares over all strings of length n with exactly d distinct symbols. It was similarly believed that the binary case is the key one; i.e. that $\sigma_{d+1}(n) \le \sigma_d(n)$, and thus that $\sigma(n) = \sigma_2(n)$, till a counterexample was provided for $n = 33$ with $\sigma_3(33) > \sigma_2(33)$, see [9].

This paper aims at combining the Bannai et al. and d-step approaches in order to highlight the structural properties of run-maximal strings. Besides strengthening by one the upper bound to $\rho(n) \le n - 4$ for $n \ge 9$, these structural properties may provide preliminary substantiation for the hypothesis that $\rho(n) \le n - \lceil \log_2 n \rceil$. For more details and additional results concerning runs in strings we refer to [3] and references therein. Before presenting the main results in Section 2, we briefly recall the Bannai et al. and d-step approaches in the remainder of this section.

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 $x[1]x[2]\dots x[n]$. The *alphabet* of a string x is the set of all symbols occurring in x. A (d, n)-string refers to a string of length n with exactly d distinct symbols. 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. The lexicographic order of strings is induced in the usual manner by the order of the alphabet. Note that $\rho_1(1) = 0$ and $\rho_1(n) = 1$ for $n \ge 2$. Thus, we can assume that both d and n are at least 2 in the remainder of the paper.

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

We briefly recall the d-step approach used to investigate the Hirsch bound for the diameter of polytopes, and its continuous analogue, and provide some basic references.

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 with n facets. The diameter $\delta(P)$ of a polytope P is the smallest integer such that 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 largest diameter over all (d, n)-polytopes. The Hirsch conjecture, posed in 1957, states that $\Delta(d, n) \le n - d$. 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 checked: $\Delta(d, n) \le \Delta(d, n + 1)$, $\Delta(d, n) < \Delta(d + 1, n + 2)$, $\Delta(d, n) \le \Delta(d + 1, n + 1)$ for $n \ge d$; and $\Delta(d, n) = \Delta(d + 1, n + 1)$ for $2d \ge n \ge d$. 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 [17] who showed the equivalency between the Hirsch conjecture and the d-step conjecture stating that $\Delta(d, 2d) \le d$ for all d. The Hirsch conjecture was disproved by Santos [25] 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. Note that the d-cube is a $(d, 2d)$-polytope having diameter d and therefore $\Delta(d, 2d) \ge d$ for all d. The string $a_1a_1a_2a_2\dots a_da_d$ is a $(d, 2d)$-string with d runs and therefore $\rho(d, 2d) \ge d$ for all d. While there is no obvious way to map the n facets of a (d, n)-polytope and the n characters of a (d, n)-string in general, one may map the d squares a_ia_i of the string $a_1a_1a_2a_2\dots a_da_d$ and the d pairs of disjoint facets of the d-cube.

A d-step approach for curvature-maximal polytopes

Considering links between the currently most computationally successful algorithms for linear optimization; i.e., the simplex and central-path following primal–dual interior point methods, Deza et al. [11] proposed a continuous analogue of the Hirsch conjecture. 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 total curvature of the associated central path, can be regarded as the continuous analogue of its diameter. Considering the largest curvature $\Lambda(d, n)$ over all (d, n)-polytopes, Deza et al. [11] 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 exhibiting a $(d, 3d/2)$-polytope with 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 [8] where it was shown that $\rho_d(n)$ and $\Delta(d, n)$ exhibit similarities and, in particular, that $\rho_d(n) \le \rho_d(n + 1)$, $\rho_d(n) < \rho_{d+1}(n + 2)$, $\rho_d(n) \le \rho_{d+1}(n + 1)$ for $n \ge d$; and $\rho_d(n) = \rho_{d+1}(n + 1)$ for $2d \ge n \ge d$. Consequently, the value of $\rho_d(n)$ is presented 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 $\rho_d(n)$. These properties noted in [8] show that the maximum of $\rho_d(n)$ within a column is achieved on the main diagonal and all values below a value on