Discrete Mathematics 341 (2018) 1662–1674

Contents lists available at ScienceDirect

Discrete Mathematics

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

On cyclic strings avoiding a pattern

Petros Hadjicostas [a], Lingyun Zhang [b]

[a] Department of Mathematical Sciences, University of Nevada, Las Vegas, Box 454020, 4505 S. Maryland Pkwy., Las Vegas, NV 89154, USA
[b] Unit 35, 25 Tacy Street, Kilbirnie, Wellington 6022, New Zealand

ARTICLE INFO

Article history:
Received 23 June 2017
Received in revised form 3 March 2018
Accepted 5 March 2018

Keywords:
Autocorrelation of a pattern
Avoiding a pattern
Cyclic string
Euler's totient function
Generating function
Necklace

ABSTRACT

We count the number of cyclic strings over an alphabet that avoid a single pattern under two different assumptions. In the first case, we assume that the symbols of the alphabet are on numbered positions on a circle, while in the second case we assume that the symbols can be freely rotated on the circle (i.e., we deal with necklaces). In each case, we provide a generating function, and we explain how these two cases are related. For the situation of avoiding more than one pattern, we formulate a general conjecture for the first case, and a conditional result for the second case. We also explain the differences between our theory and the theory of Edlin and Zeilberger (2000) by emphasizing how we modified the definition of the enumeration of cyclic strings that avoid one or more patterns when their lengths are less than the length of the longest pattern.

1. Introduction

When we talk about strings of symbols in an alphabet, we usually mean *L-type strings*, where 'L' stands for 'Linear'. An L-type string of length n (where $n \in \mathbb{Z}_{>0}$) is just an array (list) that contains n symbols. We can 'bend' an L-type string of length n onto a circle with n numbered positions to create a *C-type string*, where 'C' stands for 'Cyclic'. Throughout this paper,

- Ω denotes the *alphabet*, i.e. the set of all symbols that can be used to form strings;
- $(x_1, \ldots, x_n)_{\mathrm{L}}$ and $(x_1, \ldots, x_n)_{\mathrm{C}}$, where $x_i \in \Omega$ for $1 \le i \le n$, denote an L-type and a C-type string, respectively;
- *pattern* A is an arbitrary but fixed L-type string with length $|A| = d$ (where $d \in \mathbb{Z}_{>0}$).

C-type and L-type strings have different properties; e.g., the L-type string $(1, 0, 1)_{\mathrm{L}}$ does not contain two consecutive ones, but the C-type string $(1, 0, 1)_{\mathrm{C}}$ does.

Many authors have considered this problem:

(P0) Let $f_L(n)$ be the number of L-type strings of length n that avoid pattern A. What is $f_L(n)$, and what is the generating function for $f_L(n)$?

For example, Rényi [10], Siegel [11] and Zhang [12] examined the special case where $\Omega = \{0, 1\}$ and $A = (1, 1)_{\mathrm{L}}$. Gani and Irle [4] studied problems that are closely related to **(P0)**. We must mention Guibas and Odlyzko [7], because not only did they derive the generating function for $f_L(n)$, but also their method is powerful enough to deal with the case of avoiding more than one pattern. The same problem was also solved in full generality by Goulden and Jackson [5]; see also Bassino et al. [1] and Noonan and Zeilberger [9]. Motivated and inspired by Guibas and Odlyzko [7] and Zhang and Hadjicostas [13], in this paper, we solve the following two problems, (**P1**) and (**P2**).

E-mail addresses: peterhadji1@gmail.com (P. Hadjicostas), lyzhang10@gmail.com (L. Zhang).

https://doi.org/10.1016/j.disc.2018.03.007
0012-365X/

(P1) Let $f_C(n)$ be the number of C-type strings of length n that avoid pattern A. What is $f_C(n)$, and what is the generating function for $f_C(n)$?

A *necklace*, called a *CR-type* cyclic string in Zhang and Hadjicostas [13], is an equivalence class of all C-type strings that can be obtained from each other by a cyclic shift. Note that the C-type strings $(1, 0, 1)_C$, $(0, 1, 1)_C$ and $(1, 1, 0)_C$ are equivalent because each one can be obtained from the other by rotation.

(P2) Let $f_{CR}(n)$ be the number of necklaces of length n that avoid pattern A. What is $f_{CR}(n)$, and what is the generating function for $f_{CR}(n)$?

Problem **(P1)** was solved by Edlin and Zeilberger [3] for the general case of avoiding more than one pattern. For solving the problem, the authors provide a Maple program, which is available at their personal websites. In this paper, however, we take a different approach regarding the definition of $f_C(n)$ for $1 \le n \le d-1 = |A|-1$ (when $d \ge 2$). By doing that, we derive an elegant generating function for $f_C(n)$, which allows us to solve problem **(P2)** as well.

Section 2 contains several results necessary for solving problems **(P1)** and **(P2)**. In Section 3, using the theory in Guibas and Odlyzko [7], we formulate a general conjecture on the generating function for $f_C(n)$ when we avoid more than one pattern. (This is achieved by modifying the definition of $f_C(n)$ when the positive integer n is less than the length of the longest pattern.) This conjecture leads to a (conditional) solution of a generalized version of problem **(P2)** when we avoid more than one pattern. In Section 4, we illustrate the differences between our theory and the theory in Edlin and Zeilberger [3]. In Section 5, we show that a modification of a formula by Burstein and Wilf [2] about the number of C-type cyclic strings without constant blocks $> m$ is a special case of our conjecture, and we use their result to derive the corresponding result for necklaces (see Corollary 5.1 in this paper). Proofs (or references to proofs) of all the results in the paper are given in Section 6.

2. Main results

Consider the alphabet $\Omega = \{w_1, w_2, \ldots, w_q\}$ with q distinct symbols (where $q \in \mathbb{Z}_{>0}$). The pattern to be avoided is $A = (a_1, \ldots, a_d)_L$, where $d \in \mathbb{Z}_{>0}$ and $a_i \in \Omega$ for $i = 1, \ldots, d$. Under the usual definition of $f_C(n)$ for $1 \le n \le d-1$ (when $d \ge 2$), used by Burstein and Wilf [2] and Edlin and Zeilberger [3], we have $f_C(n) = q^n$ because no C-type string of length $n < d$ can contain pattern A.

Zhang and Hadjicostas [13] allowed C-type strings of length n with $1 \le n \le d-1$ to wrap around themselves when considering whether they avoid pattern A. This is formulated more precisely in the following definition.

Definition 2.1. If $n \in \{1, \ldots, d-1\}$ and $d \ge 2$, where $d = |A|$, we say that the C-type string $(x_1, \ldots, x_n)_C$ contains pattern A if (and only if) the L-type string

$$\Big((x_1, \ldots, x_n), \overbrace{\underbrace{(x_1, \ldots, x_n), \ldots, (x_1, \ldots, x_n)}_{kn}, x_1, \ldots, x_s}^{d-1}\Big)_L \tag{1}$$

contains A, where $k = \lfloor (d-1)/n \rfloor$ and $s = d-1-kn$.

Remark 2.2. The above definition is inspired by the fact that, for $n \ge d$, a C-type string $(x_1, \ldots, x_n)_C$ contains pattern A if and only if the L-type string

$$(x_1, \ldots, x_n, x_1, \ldots, x_{d-1})_L$$

contains A.

Let $f_A(n)$ be the number of all L-type strings of length n that end with A, but otherwise avoid A. (This notation is used in Guibas and Odlyzko [7].) Hence, $f_A(n) = 0$ for $0 \le n \le d-1$ and $f_A(d) = 1$.

For $n \in \mathbb{Z}_{>0}$, define

$$f_{A,A}(n) = qf_A(n-1) - f_A(n). \tag{2}$$

Hence, $f_{A,A}(n) = 0$ for $1 \le n \le d-1$, and $f_{A,A}(d) = -1$. For $n \ge d+1$, we can easily show that $f_{A,A}(n)$ is the number of L-type strings of length n that start with A, end with A, but otherwise avoid A.

Using terminology from Guibas and Odlyzko [6,7], suppose the *autocorrelation* of A is given by

$$AA = (c_d, \ldots, c_1).$$

In other words, if $A = (a_1, \ldots, a_d)_L$, we have

$$c_i = \begin{cases} 1, & \text{if } (a_{d-i+1}, \ldots, a_d) = (a_1, \ldots, a_i); \\ 0, & \text{otherwise} \end{cases}$$

for $i = 1, \ldots, d$. Clearly, $c_d = 1$.

When $d \ge 2$ and $1 \le n \le d-1$, define $f_C(n)$ to be the number of C-type strings of length n that do not contain A according to Definition 2.1. We then have the following result.