Theoretical Computer Science ••• (••••) •••–•••

Contents lists available at ScienceDirect

Theoretical Computer Science

www.elsevier.com/locate/tcs

On incomplete and synchronizing finite sets ☆

Arturo Carpi [a], Flavio D'Alessandro [b]

[a] Dipartimento di Matematica e Informatica, Università degli Studi di Perugia, via Vanvitelli 1, 06123 Perugia, Italy
[b] Dipartimento di Matematica, La Sapienza Università di Roma, Piazzale Aldo Moro 2, 00185 Roma, Italy

ARTICLE INFO

Article history:
Received 26 February 2015
Accepted 31 August 2015
Available online xxxx

Keywords:
Černý conjecture
Synchronizing automaton
Incompletable word
Synchronizing set
Complete set

ABSTRACT

This paper situates itself in the theory of variable length codes and of finite automata where the concepts of completeness and synchronization play a central role. In this theoretical setting, we investigate the problem of finding upper bounds to the minimal length of synchronizing words and incompletable words of a finite language X in terms of the length of the words of X. This problem is related to two well-known conjectures formulated by Černý and Restivo, respectively. In particular, if Restivo's conjecture is true, our main result provides a quadratic bound for the minimal length of a synchronizing pair of any finite synchronizing complete code with respect to the maximal length of its words.

1. Introduction

The concepts of completeness and synchronization play a central role in Formal Language Theory since they appear in the study of several problems on variable length codes and on finite automata [5]. According to a well-known result of Schützenberger, the property of completeness provides an algebraic characterization of finite maximal codes, which are the objects used in Information Theory to construct optimal sequential codings.

Let X be a set of words on an alphabet A and let X^* be its Kleene closure. The set X is *complete* if any word on the alphabet A is a factor of some word belonging to X^*, otherwise it is *incomplete*. In the latter case, any word which is factor of no word of X^* is said to be *incompletable in* X.

In [21], Restivo conjectured that a finite incomplete set X has always an incompletable word whose length is quadratically bounded by the maximal length of the words of X. Results on this problem have been obtained in [6,17,18,21]. The property of synchronization plays a natural role in Information Theory where it is relevant for the construction of decoders that are able to efficiently cope with decoding errors caused by noise during the data transmission. A set X is *synchronizing* if there are two words u, v of X^* such that whenever $ruvs \in X^*$, $r, s \in A^*$, one has also $ru, vs \in X^*$. The pair of words (u, v) is called a *synchronizing pair of* X.

In the study of synchronizing sets, the search for synchronizing words of minimal length in a prefix complete code is tightly related to that of synchronizing words of minimal length for synchronizing complete deterministic automata and the celebrated Černý Conjecture [15] (see also [2–4,7–12,15,19,20,23] for some results on the problem). In particular, in [3] (see also [4]), Béal and Perrin have proved that a complete synchronizing prefix code X on an alphabet of d letters with n code-words has a synchronizing word of length $O(n^2)$.

In this paper we are interested in finding upper bounds to the minimal lengths of incompletable and synchronizing words of a finite set X in terms of the size of X.

☆ This work was partly supported by MIUR project PRIN 2010/2011 "Automi e Linguaggi Formali: Aspetti Matematici e Applicativi".
E-mail addresses: carpi@dmi.unipg.it (A. Carpi), dalessan@mat.uniroma1.it (F. D'Alessandro).

http://dx.doi.org/10.1016/j.tcs.2015.08.042
0304-3975/

We recall that the size of X is the parameter $\ell(X)$ defined as the maximal length of the words of X.

Let $\mathcal{L}$ be a class of finite languages. For all $n, d > 0$, we denote by $R_{\mathcal{L}}(n,d)$ the least positive integer r satisfying the following condition: any incomplete set $X \in \mathcal{L}$ on a d-letter alphabet such that $\ell(X) \leq n$ has an incompletable word of length r. Similarly, we denote by $C_{\mathcal{L}}(n,d)$ the least positive integer c satisfying the following condition: any synchronizing set $X \in \mathcal{L}$ on a d-letter alphabet such that $\ell(X) \leq n$ has a synchronizing pair (u, v) such that $|uv| \leq c$.

In this context, the main result of this paper provides a bridge between the parameters $R_{\mathcal{L}}(n,d)$ and $C_{\mathcal{L}}(n,d)$. More precisely, denoting by $\mathcal{F}$ and by $\mathcal{M}$ the classes of finite languages and of complete finite codes respectively, we show that, for all $n, d > 0$,

$$C_{\mathcal{M}}(n,d) \leq 2R_{\mathcal{F}}(n, d+1) + 2n - 2.$$

In particular, if Restivo's conjecture is true, the latter bound gives

$$C_{\mathcal{M}}(n,d) = O(n^2),$$

thus providing a quadratic bound in the size of the set for the minimal length of a synchronizing pair of a finite synchronizing complete code.

In the second part of the paper, we study the dependence of the parameters $R_{\mathcal{L}}(n,d)$ and $C_{\mathcal{L}}(n,d)$ upon the number of letters d of the considered alphabet, by showing that both the parameters have a low rate of growth. More precisely, we show that, for the class $\mathcal{L}$ of finite languages (resp., codes, prefix codes), we have

$$R_{\mathcal{L}}(n,d) \leq \left\lceil \frac{R_{\mathcal{L}}(\lceil \log_2 d \rceil n, 2)}{\lfloor \log_2 d \rfloor} \right\rceil,$$

and, for the class $\mathcal{L}$ of finite complete languages (resp., codes, prefix codes), we have

$$C_{\mathcal{L}}(n,d) \leq \left\lceil \frac{C_{\mathcal{L}}(\lceil \log_2 (d+1) \rceil n, 2)}{\lfloor \log_2 (d-1) \rfloor} \right\rceil.$$

A similar result is obtained also when $\mathcal{L}$ is the class of finite (not necessarily complete) languages (resp., codes, prefix codes).

All the latter results were presented with a sketch of the proof in [13,14].

The paper is structured as follows. In Section 2, some basic results about complete and synchronizing codes as well as synchronizing automata and the Černý Conjecture are given. In Section 3 we describe our main result. In Section 4, a study of the dependence of the parameters $R_{\mathcal{L}}(n,d)$ and $C_{\mathcal{L}}(n,d)$ from the number d of letters of the alphabet is presented. Finally, in Section 5, some open questions about Restivo Conjecture are formulated.

2. Preliminaries

In this section we shortly recall some basic results of the theory of automata and of the theory of codes which will be useful in the sequel and we fix the corresponding notation used in the paper. The reader can refer to [5,16] for more details.

2.1. Complete and synchronizing sets

Let A be a finite alphabet and let A^* be the free monoid of words over the alphabet A. The identity of A^* is called the *empty word* and is denoted by ϵ. The *length* of a word $w \in A^*$ is the integer $|w|$ inductively defined by $|\epsilon| = 0$, $|wa| = |w| + 1$, $w \in A^*$, $a \in A$. Given $w \in A^*$ and $a \in A$, we denote by $|w|_a$ the number of occurrences of the letter a in w. For any finite set of words W we denote by $\ell(W)$ the maximal length of the words of W. The number $\ell(W)$ will be called the *size* of W. Given words $u, w \in A^*$, u is said to be a *factor* of w if $w = \alpha u \beta$, for some $\alpha, \beta \in A^*$. The set of all factors of w is denoted by $\mathrm{Fact}(w)$. Given a set of words W, the set of the factors of all the words of W is denoted by $\mathrm{Fact}(W)$. Similarly, given a word w, a word u is said to be a *prefix* of w if $w = u\beta$, for some $\beta \in A^*$. A set X is said to be *prefix* if no word of X is a prefix of another word of X.

Definition 1. Let X be a subset of A^*. A pair of words (r, s) is an X-*completion* of a word w if $rws \in X^*$. A word having an X-completion is a *completable* word of X; conversely, a word with no X-completion is an *incompletable* word of X. The set X is *complete* if all words of A^* are completable words of X; X is *incomplete*, otherwise.

Another crucial notion of this paper is that of synchronizing set.

Definition 2. Let X be a subset of A^*. A pair $(u, v) \in X^* \times X^*$ is a *synchronizing pair* of X if for every X-completion (r, s) of uv, one has

$$ru, vs \in X^*.$$

The set X is *synchronizing* if it has a synchronizing pair.