Černý-like problems for finite sets of words *

Arturo Carpi[1] and Flavio D'Alessandro[2]

[1] Dipartimento di Matematica e Informatica, Università degli Studi di Perugia,
Via Vanvitelli 1, 06123 Perugia, Italy.
carpi@dmi.unipg.it

[2] Dipartimento di Matematica, Università di Roma "La Sapienza"
Piazzale Aldo Moro 2, 00185 Roma, Italy.
dalessan@mat.uniroma1.it

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 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.

Keywords: Černý conjecture, synchronizing automaton, incompletable word, synchronizing set, complete set

1 Introduction

The concepts of completeness and synchronization play a central role in Computer Science since they appear in the study of several problems on variable length codes and on finite automata. 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. 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 [18], 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 [5, 14, 15, 18].

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

* This work was partially supported by MIUR project PRIN 2010/2011 "Automi e Linguaggi Formali: Aspetti Matematici e Applicativi".

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 *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 [12] (see also [1–3, 6–12, 16, 17, 20] for some results on the problem). In particular, in [2] (see also [3]), 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. 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).

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 [4, 13] for more details.

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 W of words 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 W of words, 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, for any $u, v \in X$, with $v = uw$, with $w \in A^*$, one has $w = \epsilon$.

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. *A pair $(u, v) \in X^* \times X^*$ is a* synchronizing pair *of X if for every X-completion (r, s) of uv, it holds that*

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

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

The notion of synchronizing pair of a set is strictly related to that of *constant*. A word c of X^* is said to be a *constant* of X if, for every $u_1, u_2, u_3, u_4 \in A^*$ such that $u_1cu_2, u_3cu_4 \in X^*$, one has $u_1cu_4, u_3cu_2 \in X^*$. The following result holds.

Lemma 1. *Let X be a subset of A^*. If (u, v) is a synchronizing pair of X, uv is a constant of X. Conversely, if c is a constant of X, (c, c) is a synchronizing pair of X.*

2.1 Complete and synchronizing codes

The notions of complete and synchronizing sets provide a rich structure in the case that the set is a code. It is worth to shortly describe some fundamental results on such sets. A set X of words over an alphabet A is said to be a *(variable length) code over A* if it fulfills the unique factorization property, that is, for every word $u \in X^*$, there exists a unique sequence $x_1, \ldots, x_k$ of words of X such that $u = x_1 \cdots x_k$. A well-known example of codes is given by prefix sets. The notion of code is strictly related to the one of *monomorphism.* Let A and B be two alphabets. A map $h : A^* \to B^*$ is said to be a monomorphism or (sequential) encoding if h is injective and, for every $u, v \in A^*$, one has $h(uv) = h(u)h(v)$. The following lemma shows a well-known link between the notion of code and that of monomorphism.

Proposition 1. *Let $h : A^* \to B^*$ be a monomorphism. Then the set $X = h(A)$ is a code over B. If X is a code over B and X has the same cardinality of A, then every bijection between A and X can be extended to a (unique) monomorphism $h : A^* \to B^*$.*

In view of Proposition 1, the monomorphism defined by a prefix code will be called *prefix encoding.* A submonoid of A^* is said to be *free* if it is generated by a code. An important result due to Schützenberger provides a characterization of a free submonoid of A^* in terms of a property called *stability.* A submonoid N of A^* is said to be *stable* if, for every word $u \in A^*$, the existence of words n_1, n_2, n_3, n_4 of N such that $n_2 = un_1$ and $n_4 = n_3u$ implies $u \in N$.

Theorem 1. *Let N be a submonoid of A^*. Then N is a free submonoid of A^* if and only if N is stable.*

Let us finally recall some well-known results on codes.

Theorem 2. *(Kraft-McMillan inequality) Let A be a d-letter alphabet with $d \geq 2$ and let $k_1, \ldots, k_n$ be a finite sequence of positive integers such that*

$$\sum_{i=1}^{n} d^{-k_i} \leq 1\,.$$

Then $k_1, \ldots, k_n$ is the sequence of the code-word lengths of a prefix code over A.

Given a code X over an alphabet A, X is said to be *maximal* if it is not properly contained in any other code over A. Let us recall another important result due to Schützenberger. It provides a tight relation between maximal and complete codes.

Theorem 3. *Let A be a d-letter alphabet and let X be a regular code of A^*. The following conditions are equivalent:*

- *X is a maximal code;*
- *X is a complete code;*
- $\sum_{x \in X} d^{-|x|} = 1$.

Given a set X over an alphabet A, X is said to be *maximal prefix* if it is prefix and it is not properly contained in any other prefix code over A. In the case of prefix codes we get:

Theorem 4. *Let X be a prefix code of A^* and suppose that X is a regular language. The following conditions are equivalent:*

- *X is maximal prefix;*
- *X is right complete, that is, for every word $u \in A^*, uA^* \cap XA^* \neq \emptyset$;*
- *X is a maximal code.*

It is worth recalling that Theorems 3 and 4 hold in the more general case of *thin codes* that is, for all the subsets X of A^* such that $X \cap A^*wA^* = \emptyset$, for some word $w \in A^*$. Another relevant result concerning synchronizing and complete codes is the following.

Theorem 5. *Let X be a synchronizing and complete code. Then there exists a constant $c \in X^*$ such that $cA^*c \subseteq X^*$. Conversely, if X is a code such that there exists a word $c \in X^*$ with $cA^*c \subseteq X^*$, then X is synchronizing and complete.*

2.2 Synchronizing automata and the Černý conjecture

A finite non-deterministic automaton is a tuple $\mathcal{A} = \langle Q, A, \delta, I, F\rangle$ where Q is a finite set of elements called *states*, δ is a map

$$\delta : Q \times A \longrightarrow \mathcal{P}(Q),$$

where $\mathcal{P}(Q)$ is the power set of Q, and I, F are subsets of Q. The map δ is called the *transition function* of $\mathcal{A}$ while I is called the *set of the initial states* and F is the *set of the final states*. The canonical extension of the map δ to the set $Q \times A^*$ is still denoted by δ. If P is a subset of Q and u is a word of A^*, we denote by $\delta(P, u)$ and $\delta(P, u^{-1})$ the sets:

$$\delta(P, u) = \{\delta(s, u) \mid s \in P\}, \quad \delta(P, u^{-1}) = \{s \in Q \mid \delta(s, u) \in P\}.$$

If no ambiguity arises, the sets $\delta(P, u)$ and $\delta(P, u^{-1})$ are denoted Pu and Pu^{-1}, respectively. With the automaton $\mathcal{A}$, we can associate a directed labelled multigraph $G = (Q, E)$, where the set E of edges is defined as $E = \{(p, a, q) \in Q \times A \times Q \mid \ q \in \delta(p, a)\}$. We recall that if $p, q \in Q$, then $q \in pu$ with $u \in A^*$, is equivalent to the existence of a path $c = p \xrightarrow{u} q$ in G labelled by u. The label of c is denoted by $||c||$. A word $u \in A^*$ is said to be accepted by $\mathcal{A}$ if $Iu \cap F \neq \emptyset$. The language accepted by $\mathcal{A}$, denoted by $L_{\mathcal{A}}$, is the set of all the words accepted by $\mathcal{A}$. An automaton $\mathcal{A}$ is said to be *unambiguous* if the following condition holds: for every $q_1, q_2, q_3, q_4 \in Q$ and for every $u, v \in A^*$ one has

$$q_2, q_3 \in q_1 u, \quad q_4 \in q_2 v \cap q_3 v \implies q_2 = q_3.$$

An automaton $\mathcal{A}$ is said to be *deterministic* if for every $q \in Q$ and for every $a \in A$, $\mathrm{Card}(qa) \leq 1$.

An automaton $\mathcal{A}$ is said to be *complete* if for every $u \in A^*$, there exists some $q \in Q$ such that $\mathrm{Card}(qu) \geq 1$.

An automaton $\mathcal{A}$ is said to be *transitive* if for every $p, q \in Q$, there exists $u \in A^*$ such that $q \in pu$. In the sequel, we will only consider automata $\mathcal{A}$ such that $I = F = \{1\}$, that is, with a unique initial and final state denoted 1. In particular, the tuple of the automaton $\mathcal{A}$ will be simply denoted $\mathcal{A} = \langle Q, A, \delta, 1\rangle$. Moreover, in the sequel, we will only consider transitive automata.

An unambiguous automaton $\mathcal{A}$ is said to be *synchronizing* if there exist two words $w_1, w_2 \in A^*$ such that $Qw_1 \cap Qw_2^{-1} = \{1\}$.

As is well known, a deterministic automaton $\mathcal{A}$ is synchronizing if and only if there is a word u such that $\mathrm{Card}(Qu) = 1$. Such a word is said to be a *synchronizing word* of $\mathcal{A}$. The following celebrated conjecture has been raised in [12].

Černý Conjecture. *Each synchronizing and complete deterministic automaton with n states has a synchronizing word of length $(n-1)^2$.*

Let us recall an important problem related to the Černý Conjecture. Let G be a finite, directed multigraph with all its vertices of the same outdegree. Then G is said to be *aperiodic* if the greatest common divisor of the lengths of all cycles of the graph is 1. The graph G is called a *Road coloring graph* (*RC-graph* for short) if it is aperiodic and strongly connected. A *synchronizing coloring* of G is a labeling of the edges of G that transforms it into a complete, deterministic and synchronizing automaton. The *Road coloring problem* asks for the existence of a synchronizing coloring for every RC-graph. In 2007, Trahtman proved the following remarkable result [19].

Theorem 6. *Every RC-graph has a synchronizing coloring.*

Let us conclude this section by recalling some well-known properties of the automaton that accepts the submonoid generated by a finite set (see, e.g., [6]). Let X be a finite set of words over an alphabet A. By using a standard construction, one can associate with X a transitive automaton denoted by $\mathcal{A}_X = \langle Q, A, \delta, 1\rangle$ that accepts X^*.

Lemma 2. *Let X be a regular code (resp., prefix set). Then $\mathcal{A}_X$ is an unambiguous (resp., deterministic) automaton. Conversely, let $\mathcal{A}$ be an automaton such that $L_{\mathcal{A}} = X^*$ where $X \cap X^2X^* = \emptyset$. If $\mathcal{A}$ is unambiguous (resp., deterministic), then X is a code (resp. a prefix set).*

Incompletable words of a regular set are characterized by the following.

Lemma 3. *Let X be a regular set and $\mathcal{A} = \langle Q, A, \delta, 1\rangle$ be a transitive automaton accepting X^*. Then $w \in A^*$ is a completable word of X if and only if $Qw \neq \emptyset$.*

In particular, X is complete if and only if $\mathcal{A}$ is complete.

Similarly it is possible to characterize synchronizing pairs of regular codes [6].

Lemma 4. *Let X be a regular code and $\mathcal{A} = \langle Q, A, \delta, 1\rangle$ be a transitive unambiguous automaton accepting X^* and $w_1, w_2 \in A^*$. Then (w_1, w_2) is a synchronizing pair of X if and only if $w_1w_2 \in X^*$ and $Qw_1 \cap Qw_2{}^{-1} = \{1\}$.*

Consequently, X is synchronizing if and only if $\mathcal{A}$ is synchronizing.

3 The main result

The main result of this paper is related to a problem that was formulated in [18] by Restivo. Let $\mathcal{L}$ be a class of finite languages. For all $n > 0$ we set

$$R_{\mathcal{L}}(n) = \sup_{d\geq 1} R_{\mathcal{L}}(n,d)\,, \quad C_{\mathcal{L}}(n) = \sup_{d\geq 1} C_{\mathcal{L}}(n,d)\,.$$

In [18], it was conjectured that if $\mathcal{F}$ is the class of all finite languages, then $R_{\mathcal{F}}(n) \leq 2n^2$. If we restrict ourselves to prefix codes, we get

Proposition 2. *([18]) Let $\mathcal{P}$ be the class of finite prefix codes. Then*

$$R_{\mathcal{P}}(n) \leq 2n^2.$$

However, in the general case, the previous bound was disproved in [14]. A more general and larger counterexample is given in [15]. We can thus state a slightly weaker version of the problem as follows.

Conjecture 1. (Restivo's Conjecture) Let $\mathcal{F}$ be the class of all finite languages. Then $R_{\mathcal{F}}(n) = O(n^2)$.

In this context, the main result of this paper is the following.

Proposition 3. *Let $\mathcal{M}$ be the class of complete finite codes. For all $n, d > 0$,*

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

Before proving Proposition 3, it is convenient to discuss some interesting consequences of this result. First, if Restivo's conjecture is true, we get

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

Moreover, the bound above would be sharp, as we explain below. Consider the prefix code $X_n = aA^{n-1} \cup bA^{n-2}$ on the alphabet $A = \{a, b\}$. The minimal automaton accepting X_n^* has been studied in [1], where it has been proved that the minimal length of its synchronizing words is $n^2 - 3n + 3$. From this, one derives that any synchronizing pair (w_1, w_2) of X_n verifies $|w_1w_2| \geq (n-1)^2$. In particular, a synchronizing pair of X_n of minimal length is $((ab^{n-2})^{n-1}, \epsilon)$. This provides the lower bound

$$\mathcal{C}_{\mathcal{M}}(n,2) \geq \mathcal{C}_{\mathcal{P}}(n,2) \geq (n-1)^2,$$

for the parameter $\mathcal{C}_{\mathcal{M}}(n,2)$.

It is also worth to do a remark on a recent result by Béal and Perrin. In [2] (cf. also [3]), it is proved that a synchronizing complete prefix code X with n code-words has a synchronizing word of length $2(n-2)(n-3)+1$. This result is derived from an upper bound to the length of shortest synchronizing words of synchronizing one-cluster automata. However, in view of Proposition 3 and Restivo's conjecture, this bound seems of no help in obtaining a good evaluation of the parameter $C_{\mathcal{P}}(n,2)$, as one may have $n \simeq 2^{\ell(X)}$. This suggests that a bound in term of the size of X may be more informative than a bound in terms of the cardinality.

3.1 Proof of Proposition 3

Let us now proceed to prove Proposition 3. For this purpose, let X be a finite complete synchronizing code over a d-letter alphabet A and let $n = \ell(X)$. Let $\mathcal{A}_X = \langle Q, A, \delta, 1\rangle$ be the unambiguous automaton that accepts X^*. The proof of Proposition 3 is based upon the following lemma.

Lemma 5. *Let (v_1, v_2) be a synchronizing pair of X. Then, there exist words $w_1, w_2 \in A^*$ such that*

$$|w_1|, |w_2| \leq R_{\mathcal{F}}(n, d+1), \quad Qw_1 \subseteq Qv_1, \quad Q{w_2}^{-1} \subseteq Qv_2^{-1}.$$

Indeed, assume that Lemma 5 holds. As X is complete, the word w_1w_2 has an X-completion (r, s). With no loss of generality, we may suppose that $|r|, |s| \leq n-1$. Since (v_1, v_2) is a synchronizing pair, one has

$$Q(rw_1) \cap Q(w_2s)^{-1} \subseteq Qw_1 \cap Q{w_2}^{-1} \subseteq Qv_1 \cap Q{v_2}^{-1} = \{1\}.$$

Moreover, the word $rw_1w_2s \in X^*$ is accepted by $\mathcal{A}_X$ and therefore there is a state $q \in Q$ such that $q \in 1rw_1$ and $1 \in qw_2s$. Thus, $q \in Q(rw_1) \cap Q(w_2s)^{-1} \subseteq \{1\}$, that is, $q = 1$. This proves that $rw_1, w_2s \in X^*$ and by Lemma 4 (rw_1, w_2s) is a synchronizing pair of X.

Now, our main goal is to prove Lemma 5. For the sake of simplicity, we will prove the existence of the word w_1 that fulfills the conditions of Lemma 5 since the proof of the existence of the word w_2 can be obtained by using a symmetric construction. The main tool of this proof is a new automaton we construct below.

Let (v_1, v_2) be a synchronizing pair of X. If $v_1 = \epsilon$, the statement is trivially verified by $w_1 = v_1$. Thus we assume $v_1 \neq \epsilon$ and set $v_1 = ua$, with $u \in A^*$ and $a \in A$.

Let a' be a symbol not belonging to A and let $A' = A \cup \{a'\}$. We consider a new automaton $\mathcal{A}' = \langle Q, A', \delta', 1\rangle$ where the transition map δ' is defined as follows: for every $q \in Q$ and $a \in A$, $\delta'(q, a) = \delta(q, a)$ and

$$\delta'(q, a') = \begin{cases} \delta(q, a) \cup \{1\} & \text{if } q \notin \delta(Q, u), \\ \delta(q, a) \setminus \{1\} & \text{if } q \in \delta(Q, u). \end{cases} \tag{1}$$

It is useful to remark that, by construction, the automaton $\mathcal{A}'$ is still transitive. Let Y be the minimal generating set of the language accepted by $\mathcal{A}'$. Thus, $L_{\mathcal{A}'} = Y^*$ and $Y \cap Y^2Y^* = \emptyset$.

Lemma 6. *The set Y is incomplete.*

Proof. By (1) one has $\delta'(Q, ua') = \delta(Q, ua)\setminus\{1\} = \delta(Q, v_1)\setminus\{1\}$ and $\delta'(Q, v_2^{-1}) = \delta(Q, v_2)$. Taking into account that (v_1, v_2) is a synchronizing pair of X, one derives

$$\delta'(Q, ua') \cap \delta'(Q, v_2^{-1}) = (\delta(Q, v_1) \cap \delta'(Q, v_2^{-1})) \setminus \{1\} = \emptyset\,.$$

It follows that $\delta'(Q, ua'v_2) = \emptyset$. This equation proves that the automaton $\mathcal{A}$ is not complete. Thus, by Lemma 3, Y is an incomplete set. $\square$

Lemma 7. *It holds that $\ell(Y) \leq \ell(X)$.*

Proof. In order to prove the statement, it is enough to show that, for every $y \in Y$, there exists $x \in X$ with $|y| \leq |x|$.

Let $y = a_1 \cdots a_k \in Y$, with $a_i \in A'$, for $i = 1, \ldots, k$. Since $Y \cap Y^2Y^* = \emptyset$, in the graph of $\mathcal{A}'$ there is a path

$$c' = 1 \xrightarrow{a_1} q_1 \xrightarrow{a_2} q_2 \xrightarrow{a_3} \cdots \xrightarrow{a_{k-1}} q_k \xrightarrow{a_k} 1,$$

where, for every $i = 1, \ldots, k$, $q_i \neq 1$. Let us now construct a path c in the graph of $\mathcal{A}_X$ such that $||c|| = x \in X$, with $|x| \geq |y|$, so completing the proof.

By the definition of $\mathcal{A}'$, any edge $p \xrightarrow{b} q$ of the graph of $\mathcal{A}'$ with $b \neq a'$ is also an edge of the graph of $\mathcal{A}$. Moreover, if $p \xrightarrow{a'} q$ is an edge of the graph of $\mathcal{A}'$ with $q \neq 1$, then $p \xrightarrow{a} q$ is an edge of the graph of $\mathcal{A}$. Thus, by replacing in c', every transition $q_i \xrightarrow{a'} q_{i+1}$, by $q_i \xrightarrow{a} q_{i+1}$ and deleting the last edge $q_k \xrightarrow{a_k} 1$, we find a path

$$d = 1 \xrightarrow{b_1} q_1 \xrightarrow{b_2} q_2 \cdots \xrightarrow{b_{k-1}} q_k$$

of the graph of $\mathcal{A}$. Since $\mathcal{A}$ is transitive, one can catenate d with a simple path from q_k to 1. In such a way, we obtain a path c of the graph of $\mathcal{A}$ starting and ending in 1, with all intermediate states distinct from 1 and length $\geq k+1$. As is well known, as $\mathcal{A}$ is unambiguous, the label x of such a path is a word of the minimal generating set X of X^*. Since $|x| \geq k+1 = |y|$, this completes the proof. □

We now prove the following lemma.

Lemma 8. *Let v be an incompletable word of Y of minimal length. There exists a word $w_1 \in A^*$ such that*

$$|w_1| \leq |v|, \quad Qw_1 \subseteq Qv_1.$$

Proof. Let v be an incompletable word of Y of minimal length, with the number $|v|_{a'}$ as small as possible. Then, by Lemma 3, one has $\delta'(Q, v) = \emptyset$.

The letter a' necessarily occurs in v, since by the completeness of $\mathcal{A}$, one has $\delta'(Q, r) = \delta(Q, r) \neq \emptyset$ for all $r \in A^*$. Thus, we can write $v = u_1 a' u_2$, with $u_1 \in A^*$ and $u_2 \in A'^*$.

Let us verify that $\delta(Q, u_1) \subseteq \delta(Q, u)$. Indeed, suppose the contrary. Then, by (1), one has

$$\delta'(Q, u_1 a') = \delta(Q, u_1 a) \cup \{1\} = \delta'(Q, u_1 a) \cup \{1\}$$

and consequently, $\delta'(Q, u_1 a u_2) \subseteq \delta'(Q, u_1 a' u_2) = \emptyset$. Thus, $u_1 a u_2$ is an incompletable word of Y, but this contradicts the minimality of $|v|_{a'}$.

We conclude that $\delta(Q, u_1) \subseteq \delta(Q, u)$ and therefore taking $w_1 = u_1 a$ and recalling that $v_1 = ua$, one has $\delta(Q, w_1) \subseteq \delta(Q, v_1)$ and $|w_1| \leq |v|$. The statement follows. □

Let us finally remark that Lemma 7 and Lemma 8 yield

$$|w_1| \leq R_{\mathcal{F}}(n, d+1), \quad Qw_1 \subseteq Qv_1.$$

The proof of Lemma 5 is thus complete.

4 Reduction to the binary case

The aim of this section is to study how much the parameters $R_{\mathcal{L}}(n,d)$ and $C_{\mathcal{L}}(n,d)$ vary according to the cardinal number d of the alphabet. We start to analyze the parameter $R_{\mathcal{L}}(n,d)$. In the sequel, B denotes the binary alphabet $B = \{a, b\}$. The following lemmas are needed for the proof of Proposition 4.

Lemma 9. *Let $Y \subseteq A^*$ be a complete finite set. Then any word w of A^* has a Y-completion (y, s) with $y \in Y^*$.*

Lemma 10. *Let $h : A^* \to B^*$ be a prefix encoding and $Y \subseteq A^*$. The set $h(Y)$ is complete if and only if Y and $h(A)$ are complete.*

Encoding a d-letter alphabet on a suitable complete binary prefix code one obtains

Proposition 4. *Let $\mathcal{L}$ be the class of finite languages (resp. codes). Then*

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

Proof. Let A be a d-letter alphabet and let X be a finite incompletable language over A of size n. By Theorems 2 and 3, there exists a maximal prefix code Y over B such that $\mathrm{Card}(Y) = d$ and, for every $y \in Y$, $\lfloor \log_2 d \rfloor \leq |y| \leq \lceil \log_2 d \rceil$. Let $h : A^* \to B^*$ be a monomorphism constructed by Y. In particular, for every $a \in A$, we have

$$\lfloor \log_2 d \rfloor \leq |h(a)| \leq \lceil \log_2 d \rceil. \tag{3}$$

By (3) the size of $h(X)$ is not greater than $n\lceil \log_2 d \rceil$. By Lemma 10, since X is incompletable, $h(X)$ is incompletable as well. Let v be an incompletable word in $h(X)$ of minimal length. Hence we have

$$|v| \leq R_{\mathcal{L}}(\lceil \log_2 d \rceil n, 2). \tag{4}$$

Since $h(A)$ is a complete prefix code, by Theorem 4, there exist $u \in A^*$ and $s \in B^*$ such that $h(u) = vs$ and by (3)

$$|u| \leq \left\lceil \frac{|v|}{\lfloor \log_2 d \rfloor} \right\rceil. \tag{5}$$

Let us check that u is incompletable. By contradiction, deny. Then $r'us' \in X^*$, for some $r', s' \in A^*$. Consequently, $h(r'us') = h(r')vsh(s') \in h(X^*)$, thus implying that v is completable in $h(X)$. Now (2) easily follows from the latter, (4) and (5). □

Let us now analyze the parameter $C_{\mathcal{L}}(n,d)$. The next two lemmas are useful for this purpose. In particular the following lemma is algebraically similar to Lemma 10.

Lemma 11. *Let $h : A^* \to B^*$ be a monomorphism and let $Y \subseteq A^*$ be a complete set. The set $h(Y)$ is synchronizing if and only if Y and $h(A)$ are synchronizing.*

Lemma 12. *Let $k_1, \dots, k_n, d > 0$ be such that*

$$\gcd(k_1, k_2, \dots, k_n) = 1\,, \qquad \sum_{i=1}^{n} d^{-k_i} = 1.$$

Then $k_1, \dots, k_n$ are the code-word lengths of a synchronizing complete prefix code over d letters.

Remark 1. It is worth noticing that a finite synchronizing complete prefix code over d letters satisfies both the conditions of Lemma 12. Indeed, by Theorem 3, if $X = \{x_1, \dots, x_n\}$ is a complete code over d letters, one gets $\sum_{i=1}^{n} d^{-k_i} = 1$. Moreover, by Theorem 5, there exists a constant $x \in X^*$ for X such that $xA^*x \subseteq X^*$. Let $\gamma = \gcd(k_1, k_2, \dots, k_n)$. Since, by the latter result, $x^2, xax \in X^*$, with $a \in A$, γ should divide both $2|x|$ and $2|x|+1$, whence $\gamma = 1$.

As an application of the two lemmas above, encoding a d-letter alphabet on a suitable complete binary synchronizing code, one obtains the following result:

Proposition 5. *Let $\mathcal{L}$ be the class of finite complete languages (resp. codes, prefix codes). Then*

$$C_{\mathcal{L}}(n,d) \le \left\lceil \frac{C_{\mathcal{L}}(\lceil \log_2(d+1) \rceil n, 2)}{\lfloor \log_2(d-1) \rfloor} \right\rceil . \tag{6}$$

A similar bound can be found also in the case where completeness is not required:

Proposition 6. *Let $\mathcal{L}$ be the class of finite languages (resp. codes, prefix codes). Then*

$$C_{\mathcal{L}}(n,d) \le \left\lceil \frac{C_{\mathcal{L}}(\lceil \log_2(d+1) \rceil n, 2)}{\lceil \log_2(d+1) \rceil} \right\rceil . \tag{7}$$

References

1. D. S. Ananichev, V. V. Gusev, M. V. Volkov, Slowly synchronizing automata and digraphs, in: P. Hliněný, A. Kučera eds., *MFCS 2010 Mathematical Foundations of Computer Science, Lecture Notes in Comput. Sci.* Vol. 6281, pp. 55–65, Springer, Berlin, 2010.
2. M.-P. Béal, D. Perrin, A quadratic upper bound on the size of a synchronizing word in one-cluster automata, in: V. Diekert, D. Nowotka eds., *DLT 2009 Developments in Language Theory, Lecture Notes in Computer Science*, Vol. 5583, pp. 81–90, Springer, Berlin, 2009.

3. M.-P. Béal, M. V. Berlinkov, D. Perrin, A quadratic upper bound on the size of a synchronizing word in one-cluster automata, *Int. J. Found. Comput. Sci.,* 22, 277–288, 2011.
4. J. Berstel, D. Perrin, Ch. Reutenauer, Codes and Automata, Encyclopedia of Mathematics and its Applications, 129, Cambridge University Press, 2009.
5. J. M. Boë, A. de Luca, A. Restivo, Minimal complete sets of words, *Theoret. Comput. Sci.,* 12, 325–332, 1980.
6. A. Carpi, On synchronizing unambiguous automata, *Theoret. Comput. Sci.,* 60, 285–296, 1988.
7. A. Carpi, F. D'Alessandro, The Synchronization Problem for Strongly Transitive Automata, in: M. Ito, M. Toyama eds., *DLT 2008 Developments in Language Theory, Lecture Notes in Computer Science*, Vol. 5257, pp. 240–251, Springer, Berlin, 2008.
8. A. Carpi, F. D'Alessandro, Strongly transitive automata and the Černý conjecture *Acta Informatica,* 46, 591–607, 2009.
9. A. Carpi, F. D'Alessandro, The synchronization problem for locally strongly transitive automata, in: R. Královič, D. Niwiński eds., *MFCS 2009 Mathematical Foundations of Computer Science, Lecture Notes in Comput. Sci.,* Vol. 5734, pp. 211–222, Springer, Berlin, 2009.
10. A. Carpi, F. D'Alessandro, On the Hybrid Černý-Road coloring problem and Hamiltonian paths, in: Y. Gao, H. Lu, S. Seki, S. Yu eds., *DLT 2010 Developments in Language Theory, Lecture Notes in Comput. Sci.,* Vol. 6224, pp. 124–135, Springer, Berlin, 2010.
11. A. Carpi, F. D'Alessandro, Independent sets of words and the synchronization problem, *Advances in Applied Mathematics,* 50, 339–355, 2013.
12. J. Černý, Poznámka k. homogénnym experimenton s konečnými automatmi, *Mat. fyz. cas SAV,* 14, 208–215, 1964.
13. A. de Luca, F. D'Alessandro, Teoria degli Automi Finiti, 68, Springer Italia, 2013.
14. G. Fici, E. V. Pribavkina, J. Sakarovitch, On the Minimal Uncompletable Word Problem, CoRR, arXiv: 1002.1928, 2010.
15. V. V. Gusev, E. V. Pribavkina, On Non-complete Sets and Restivo's Conjecture, in: G. Mauri, A. Leporati eds., *DLT 2011 Developments in Language Theory, Lecture Notes in Comput. Sci.,* Vol. 6795, pp. 239–250, Springer, Berlin, 2011.
16. J. E. Pin, *Le problème de la synchronization et la conjecture de Cerny,* Thèse de 3ème cycle, Université de Paris 6, 1978.
17. J. E. Pin, Sur un cas particulier de la conjecture de Cerny, in: G. Ausiello, C. Böhm eds., *5th ICALP Lecture Notes in Computer Science*, Vol. 62, pp. 345–352, Springer, Berlin, 1978.
18. A. Restivo, Some remarks on complete subsets of a free monoid, in: A. de Luca ed., *Non-Commutative Structures in Algebra and Geometric Combinatorics, International Colloquium, Arco Felice, July 1978*, *Quaderni de "La Ricerca Scientifica"*, CNR, 109, 19–25, 1981.
19. A. N. Trahtman, The road coloring problem, *Israel J. Math.,* 172, 51–60, 2009.
20. M. V. Volkov, Synchronizing automata and the Cerny conjecture, in: C. Martín-Vide, F. Otto, H. Fernau eds., *LATA 2008 Language and Automata Theory and Applications, Lecture Notes in Comput. Sci.,* Vol. 5196, pp. 11–27, Springer, Berlin, 2008.