Completeness and synchronization for finite sets of words

Arturo Carpi*
and
Flavio D'Alessandro†

June 22, 2014

Abstract

We investigate the problem of finding upper bounds to the minimal lengths of incompletable and synchronizing 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.

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

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.

We recall that a set of words X is *incomplete* if there exists a word which is not factor of any word of X^*. Such a word is called *incompletable*. The set X is *synchronizing* if there is a pair of words $u, v \in X^*$ such that whenever $ruvs \in X^*$, one has $ru, vs \in X^*$. Such a pair is called *synchronizing*.

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. A problem that naturally arises in this context concerns the minimal length of an incompletable word in a finite incomplete set of words. This problem was first addressed and studied by Restivo in [20]. In particular, it was conjectured that a finite incomplete set X has always an incompletable word whose length is quadratic in the *size* of X, that is, the maximal length of the words of X. Results on this problem have been obtained in [5, 15, 16, 20].

The property of synchronization plays a natural role for the formal description of fault-tolerant systems. In particular, in the context of Information Theory, this property is relevant for the construction of decoders that are able to efficiently cope with decoding errors caused by noise during the data transmission. 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 minimal length synchronizing words of synchronizing deterministic automata and the celebrated Černý Conjecture [13] (see also [1, 2, 3, 4, 6, 7, 9, 10, 11, 12, 13, 14, 17, 18, 19, 21] 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 $O(n^2)$.

*Dipartimento di Matematica e Informatica, Università degli Studi di Perugia, via Vanvitelli 1, 06123 Perugia, Italy,

†Dipartimento di Matematica, Università di Roma "La Sapienza", Piazzale Aldo Moro 2, 00185 Roma, Italy.

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.

Let $\mathcal{L}$ be a class of finite languages and denote by $\ell(X)$ the size of a finite set X of words. 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, our main result provides a bridge between the parameters $R_{\mathcal{L}}(n,d)$ and $C_{\mathcal{L}}(n,d)$.

Proposition 1 *Let $\mathcal{F}$ and $\mathcal{M}$ be, respectively, the classes of finite languages and of complete finite codes. For all $n, d > 0$ one has*

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

In particular, if Restivo's conjecture is true, Proposition 1 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.

We notice that, in rather general cases, quadratic lower bounds for $R_{\mathcal{L}}$ and $C_{\mathcal{L}}$ are known. Indeed, if $\mathcal{P}$ is the class of prefix codes, then

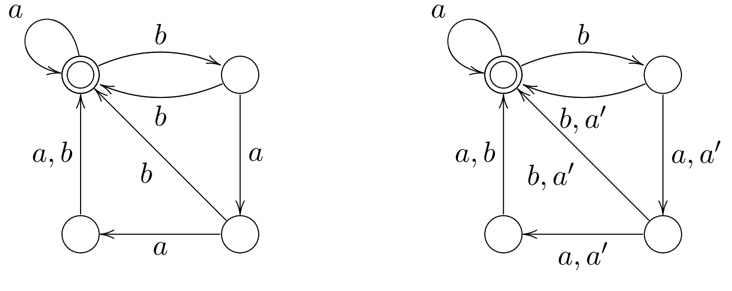
$$R_{\mathcal{P}}(n,2) \geq n^2 + n - 1, \qquad C_{\mathcal{P}}(n,2) \geq (n-1)^2.$$

The first inequality is proved in [19], the second follows from a result of [1].

We outline the proof of Proposition 1 in the particular case that X is a prefix code. Let $\mathcal{A}$ be the minimal deterministic finite automaton accepting X^*, q_0 be its initial state, and a be the terminal letter of some word of X. We construct a non deterministic finite automaton $\mathcal{A}'$ adding to $\mathcal{A}$ a new letter a' with transitions defined by $qa' = \{qa, q_0\}$ if $qa \neq q_0$, $qa' = \emptyset$, otherwise. It turns out that the language accepted by $\mathcal{A}'$ is equal to Y^* for a suitable incomplete set Y of the same size of X. Moreover, replacing by a all the occurrences of a' in a minimal incompletable word of Y, one obtains a synchronizing word of X. This allows to establish, for the class $\mathcal{P}$ of complete prefix codes, the inequality

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

For instance, if $X = \{a, baab, bab, bb\}$, then $\mathcal{A}$ and $\mathcal{A}'$ are respectively the automata below, a minimal incompletable word of Y is aaa', and aaa is a synchronizing word of the prefix code X.



The proof of the general case [8] makes use of unambiguous finite state automata and of their combinatorial properties.

Afterwards, 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.

Proposition 2 *Let $\mathcal{L}$ be the class of finite languages (resp. codes, prefix codes). Then, for all $d > 2$ one has*

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

Proposition 3 *Let $\mathcal{L}$ be the class of complete finite languages (resp. codes, prefix codes). Then, for all $d > 2$ one has*

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

The proofs of the propositions above are based on suitable binary encodings preserving incompleteness or synchronization. Among the combinatorial tools used in this work, we mention in particular the following one.

Lemma 1 *Let $d, n \geq 1$. The positive integers $k_1, \ldots, k_n$ are the code-word lengths of a synchronizing complete prefix code over d letters if and only if*

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

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, Černý-like problems for finite sets of words, manuscript, 2014.

[9] A. Carpi, F. D'Alessandro, Strongly transitive automata and the Černý conjecture *Acta Informatica,* 46, 591–607, 2009.

[10] 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.

[11] 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.

[12] A. Carpi, F. D'Alessandro, Independent sets of words and the synchronization problem, *Advances in Applied Mathematics,* 50, 339–355, 2013.

[13] J. Černý, Poznámka k. homogénnym experimenton s konečnými automatmi, *Mat. fyz. cas SAV,* 14, 208–215, 1964.

[14] A. de Luca, F. D'Alessandro, Teoria degli Automi Finiti, 68, Springer Italia, 2013.

[15] G. Fici, J. Sakarovitch, E. V. Pribavkina, On the Minimal Uncompletable Word Problem, CoRR, arXiv: 1002.1928, 2010.

[16] 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.

[17] J. E. Pin, *Le problème de la synchronization et la conjecture de Cerny,* Thèse de 3ème cycle, Université de Paris 6, 1978.

[18] 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.

[19] E. V. Pribavkina, Slowly synchronizing automata with zero and incomplete sets, CoRR, arXiv:0907.4576, 2009.

[20] 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.

[21] 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.