

# Synchronizing Strongly Connected Partial DFAs

Mikhail V. Berlinkov ✉

Institute of Natural Sciences and Mathematics, Ural Federal University, Ekaterinburg, Russia

Robert Ferens ✉

Institute of Computer Science, University of Wrocław, Poland

Andrew Ryzhikov ✉

Université Gustave Eiffel, LIGM, Marne-la-Vallée, France

Marek Szykuła ✉

Institute of Computer Science, University of Wrocław, Poland

---

## Abstract

We study synchronizing partial DFAs, which extend the classical concept of synchronizing complete DFAs and are a special case of synchronizing unambiguous NFAs. A partial DFA is called synchronizing if it has a word (called a reset word) whose action brings a non-empty subset of states to a unique state and is undefined for all other states. While in the general case the problem of checking whether a partial DFA is synchronizing is PSPACE-complete, we show that in the strongly connected case this problem can be efficiently reduced to the same problem for a complete DFA. Using combinatorial, algebraic, and formal languages methods, we develop techniques that relate main synchronization problems for strongly connected partial DFAs with the same problems for complete DFAs. In particular, this includes the Černý and the rank conjectures, the problem of finding a reset word, and upper bounds on the length of the shortest reset words of literal automata of finite prefix codes. We conclude that solving fundamental synchronization problems is equally hard in both models, as an essential improvement of the results for one model implies an improvement for the other.

**2012 ACM Subject Classification** Theory of computation → Formal languages and automata theory; Mathematics of computing → Combinatorics

**Keywords and phrases** Černý conjecture, literal automaton, partial automaton, prefix code, rank conjecture, reset threshold, reset word, synchronizing automaton, synchronizing word

**Digital Object Identifier** 10.4230/LIPIcs.STACS.2021.12

**Related Version** *Full Version:* <https://arxiv.org/abs/2101.05057>

**Funding** *Robert Ferens:* Supported in part by the National Science Centre, Poland under project number 2017/25/B/ST6/01920.

*Marek Szykuła:* Supported in part by the National Science Centre, Poland under project number 2017/25/B/ST6/01920.

**Acknowledgements** We are grateful to the anonymous reviewers for useful comments and recent literature references.

## 1 Introduction

Synchronization is a concept in various domains of computer science which consists in regaining control over a system by applying (or observing) a specific set of input instructions. These instructions are usually required to lead the system to a fixed state no matter in which state it was at the beginning. This idea has been studied for automata (deterministic [8, 36], non-deterministic [17], unambiguous [2], weighted and timed [10], partially observable [22], register [1], nested word [9]), parts orienting in manufacturing [13, 23], testing of reactive systems [30], variable length codes [5], and Markov Decision Processes [11, 12].



© Mikhail V. Berlinkov, Robert Ferens, Andrew Ryzhikov, and Marek Szykuła; licensed under Creative Commons License CC-BY 4.0

38th International Symposium on Theoretical Aspects of Computer Science (STACS 2021).

Editors: Markus Bläser and Benjamin Monmege; Article No. 12; pp. 12:1–12:16

Leibniz International Proceedings in Informatics



LIPICs Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany



In this paper, we study the synchronization of partial DFAs, which are a generalization of complete DFAs and a special case of unambiguous NFAs. We are motivated by applications of this model and its connections with others, as well as the need for new techniques applied to partial DFAs. The problems for strongly connected partial DFAs are a motivation for further development and generalization of the methods applied for complete DFAs, since, as we show, these models are closely related. We also hope that our methods will serve as a step toward studying a wider class of strongly connected unambiguous NFAs.

## 1.1 Observing a reactive system

Consider a finite-state reactive system modeled by a partial DFA (by partial we mean that for some states there can be no outgoing transitions corresponding to some letters). The observer knows the structure of the DFA but does not know its current state. At every step, the DFA reads a letter (also known to the observer) and transits to another state. The observer wants to eventually learn the actual state of the DFA. Since the DFA is deterministic, once a state is known, it will be known forever.

In this setting, the actual state is known if and only if the system reads a *reset word* – a word which transits a non-empty set of states to a single state and is undefined for all other states. The presence of undefined transitions indicates that certain actions cannot be performed from certain states, which can be essential for synchronization.

For several identical systems running in parallel and receiving the same input (but possibly starting from different states), the presence of a reset word guarantees that all systems end up in the same state. This idea can be used in robotics, where a sequence of passive obstacles is used for orienting a large number of arbitrarily rotated parts arriving simultaneously on a conveyor belt ([13, 23], see also [36] for an illustrative example).

Reactive systems (such as Web servers, communication protocols, operating systems and processors) are systems developed to run without termination and interact through visible events, so it is natural to assume that the system can return to any state from any other state (NFAs with this property are called *strongly connected*). The probabilistic version of the described problem for strongly connected partial DFAs has been considered in the context of  $\varepsilon$ -machines [34]. In particular, the observer knows the state of an  $\varepsilon$ -machine precisely if and only if a reset word for the underlying partial DFA was applied. Some experimental results on finding shortest reset words for partial DFAs were recently presented in [31].

## 1.2 Synchronizing automata

There exist several definitions that generalize the notion of a synchronizing complete DFA to larger classes of NFAs. In this subsection, we describe the notion which preserves most of the properties of the complete DFAs case, and in 1.5 we briefly describe alternative notions.

An NFA is called *unambiguous* if for every two states  $p, q$  and every word  $w$ , there is at most one path from  $p$  to  $q$  labeled by  $w$  [2]. In the strongly connected case, this is equivalent to a more classical definition of an unambiguous automaton with chosen initial and final states, if there is a unique initial state and a unique final state. An unambiguous NFA is called *synchronizing* if there exist two non-empty subsets  $C, R$  of its states and a word  $w$  (called a *reset word*) such that its action maps every state in  $C$  exactly to the whole set  $R$ , and is undefined for all states outside  $C$  [2]. For partial DFAs, the set  $R$  has size one [5], and, for complete DFAs, the set  $C$  is also the whole set of states [36].

Partial DFAs are a natural intermediate class between unambiguous NFAs and complete DFAs. The bounds on the length of shortest reset words in strongly connected partial DFAs have not been studied before. The famous Černý conjecture, which is one of the most

longstanding open problems in automata theory, states that for an  $n$ -state complete DFA we can always find a reset word of length at most  $(n - 1)^2$ , unless there are no reset words. The best known upper bound is cubic in  $n$  [32, 33], and the problem of deciding whether a complete DFA is synchronizing is solvable in time quadratic in  $n$  [36]. For an  $n$ -state strongly connected unambiguous NFA the best known upper bound on the length of a shortest reset word is  $n^5$ , and the existence of a reset word is verifiable in time  $\mathcal{O}(n^5)$  [28]. The same upper bound holds for the length of the shortest mortal words in strongly connected unambiguous NFAs [21], whereas partial DFAs admit a tight quadratic bound [26].

### 1.3 Synchronizing codes

A *variable length code*  $X$  (which we call a *code*) is a set of finite words over a finite alphabet  $\Sigma$ , such that no word over  $\Sigma$  can be written as a concatenation of codewords of  $X$  in two different ways. Such codes (especially Huffman codes [16]) are widely used for lossless data compression. Since the lengths of codewords can be different, one transmission error can spoil the whole decoding process, causing a major data loss. Also, in general, decoding a part of a message (e.g., a segment of a compressed video stream) is not possible without decoding the whole message.

These issues can be addressed by using synchronizing codes. A code  $X$  is called *synchronizing* if there exists a *synchronizing* word  $w \in X^*$  such that for every  $uwv \in X^*$  we have  $uw, wv \in X^*$ . The occurrence of the word  $w$  thus stops error propagation and allows parallel decoding of the two parts of the message. More generally, each appearance of the word  $w$  in a coded message allows running decoding independently from the position after the first  $w$ .

A code is called *prefix* if none of its codewords is a prefix of another codeword. Such codes allow obtaining the correct partition of a message into codewords one by one by going from left to right. Even if a code is synchronizing, there are no guarantees that a synchronizing word will appear in a message. Codes where every long enough concatenation of codewords is synchronizing are called *uniformly synchronizing* [5, 7]. A prefix code is called *maximal* if it is not a subset of another prefix code. All non-trivial uniformly synchronizing finite prefix codes are non-maximal [5].

### 1.4 Automata for $X^*$

A code recognized by an NFA as a language is called *recognizable*. In particular, every finite code is recognizable. To argue about synchronization properties of a recognizable code  $X$ , special NFAs recognizing  $X^*$  are studied. These NFAs have a unique initial and final state  $r$  such that the set of words labeling paths from  $r$  to itself coincides with  $X^*$ , thus they are also strongly connected. Provided a recognizable code  $X$ , an NFA with the described properties can be chosen to be unambiguous [5]. Moreover, this NFA can be chosen to be a partial (respectively, a complete) DFA if and only if  $X$  is a recognizable prefix (respectively, recognizable maximal prefix) code [5].

For such an unambiguous NFA with the properties as above,  $X$  is synchronizing if and only if the NFA is synchronizing, and the length of a shortest synchronizing word for  $X$  is at most the length of a shortest reset word of the NFA plus twice its number of states [5, Chapter 4].

Finite prefix codes admit a direct construction of partial DFAs with the described properties, called *literal automata*. Let  $X$  be a finite prefix code over an alphabet  $\Sigma$ . The *literal automaton*  $\mathcal{A}_X = (Q, \Sigma, \delta)$  is constructed as follows. The set of states  $Q$  is the set of all proper prefixes of the words in  $X$ , the transition function is defined as follows:  $\delta(q, x) = qx$

if  $qx \notin X$  and  $qx$  is a proper prefix of a word in  $X$ ,  $\delta(q, x) = \varepsilon$  if  $qx \in X$ , and  $\delta(q, x) = \perp$  otherwise. The state corresponding to the empty prefix  $\varepsilon$  is called the *root state*. The *height* of a literal automaton is the length of a longest path of its transitions without repetition of states; equivalently, this is the length of the longest word in  $X$  minus one. Note that the number of states of  $\mathcal{A}_X$  is at most the total length of all codewords of  $X$ , which allows to directly transfer upper bounds from literal automata to finite prefix codes. An example of a literal automaton is shown in Fig. 1(right). The literal automaton of a prefix code can be used as a decoder for this code by adding output labels to the transitions [5].

## 1.5 Carefully synchronizing DFAs

For general NFAs, synchronizability can be generalized to  $D_i$ -directability for  $i = 1, 2, 3$  [17]. As discussed in [38, Section 6.3], for partial DFAs the notions of D1- and D3-directing words both coincide with *carefully synchronizing* words. These are words sending every state of a partial DFA to the same state, not using any undefined transitions at all. A D2-directing word for a partial DFA is either carefully synchronizing or mortal (undefined for every state). The definitions of carefully synchronizing and D2-directing words are different from our definition of synchronizing words.

A carefully synchronizing word can be applied to a partial DFA at any moment without the risk of using an undefined transition. This comes at a high cost: even for strongly connected partial DFAs, the shortest carefully synchronizing words can have exponential length [38, Proposition 9], and the problem of checking the existence of such a word is PSPACE-complete [38, Theorem 12], in contrast with the case of complete DFAs. On the contrary, the notion of a synchronizing partial DFA preserves most of the properties of a synchronizing complete DFA, at least in the strongly connected case. Note that every carefully synchronizing word is synchronizing, but the converse is not true.

While for complete DFAs the property of being strongly connected is not essential for many synchronization properties [36], the situation changes dramatically for partial DFAs. Partial DFAs, which are not strongly connected, can have exponentially long shortest reset words, and the problem of checking the existence of a reset word is PSPACE-complete [3]. Thus, strong connectivity is indeed necessary to obtain good bounds and algorithms. As explained above, for reactive systems and prefix codes this requirement comes naturally.

## 1.6 Our contribution and organization of the paper

We prove a number of results for strongly connected partial DFAs connected with the Černý conjecture and its generalizations. Where possible, we look up for methods that allow relating the partial case with the complete case, instead of directly reproving known results in this more general setting. In this way, we do not have to go into the existing proofs, and future findings concerning the complete case should be often immediately transferable to the partial case.

We start from basic properties and introduce more advanced techniques along with their applications. First, we investigate the rank conjecture, which is a generalization of the Černý conjecture from the case of synchronizing automata to the case of all automata. We show that the rank conjecture for complete DFAs implies it also for partial DFAs (Theorem 10). For this, we introduce our first basic tool called *fixing automaton*, which is a complete automaton obtained from a partial one and sharing some properties. Our result shows a general way for transferring upper bounds from the case of complete DFAs to partial DFAs, e.g., we immediately get that the rank conjecture holds for partial Eulerian automata (Corollary 11).

To connect the Černý conjecture for the cases of complete and partial DFAs, we need more involved techniques, since the construction developed for the rank conjecture does not preserve the property of being synchronizing. We introduce a *collecting automaton*, which extends the concept of the fixing automaton. We use it to show that all upper bounds on the length of the shortest reset words, up to a subquadratic additive component, are equivalent for partial and complete DFAs (Theorem 17). We also use it to prove that the problems of determining synchronizability and finding a reset word of a strongly connected partial DFA can be effectively reduced to the same problem for a complete DFA (Section 3.5). This also means that possible improvements of the complexity of the best-known algorithms for these problems for complete DFAs should directly apply to partial DFAs.

As discussed in 1.3 and 1.4, one of the main motivations for studying synchronization of strongly connected partial DFAs is a direct correspondence with synchronization of recognizable prefix codes. An important special case is when the prefix code is finite. We investigate it by studying literal automata of finite prefix codes and obtain stronger upper bounds than those for the general case of strongly connected partial (or complete) DFAs. We show that the length of the shortest reset words for literal automata of finite prefix codes is at most  $\mathcal{O}(n \log^3 n)$ , where  $n$  is the number of states of the automaton (Corollary 23). This upper bound is the same as the best known for maximal prefix codes (which becomes now a special case), but it is not transferred directly, as key statements do not hold in the same way for non-maximal prefix codes. To prove it, we first show that literal automata of finite prefix codes admit a word of linear length whose action sends all the states to a non-empty subset of small size (Theorem 22). It establishes a natural combinatorial property of finite prefix codes and constitutes the most involved proof in this paper. Once we show the existence of such a word, we use one more construction called the *induced automaton*, which is a generalization of linear algebraic techniques to the case of partial DFAs (Section 3.7). This particular construction extends the existing techniques originally developed for complete DFAs but simultaneously comes with a new simpler and more general proof.

Finally, we show that the lower bounds for strongly connected partial DFAs are asymptotically the same even if we ensure the existence of undefined transitions (Section 4). In other words, undefined transitions do not help in general, as we cannot significantly improve upper bounds for such automata without doing that for the complete case.

## 2 Preliminaries

A *partial deterministic finite automaton*  $\mathcal{A}$  (which we call *partial automaton* throughout the paper) is a triple  $(Q, \Sigma, \delta)$ , where  $Q$  is a set of *states*,  $\Sigma$  is an *input alphabet*, and  $\delta$  is partial function  $Q \times \Sigma \rightarrow Q$  called the *transition function*. Note that the automata we consider do not have any initial or final states. We extend  $\delta$  to a partial function  $Q \times \Sigma^* \rightarrow Q$  as usual: we set  $\delta(q, wa) = \delta(\delta(q, w), a)$  for  $w \in \Sigma^*$  and  $a \in \Sigma$ . For a state  $q \in Q$  and a word  $w \in \Sigma^*$ , if the action  $\delta(q, w)$  is undefined, then we write  $\delta(q, w) = \perp$ . Note that if  $\delta(q, w) = \perp$  for a word  $w \in \Sigma^*$ , then  $\delta(q, wu) = \perp$  for every word  $u \in \Sigma^*$ . An automaton is *complete* if all its transitions are defined, and it is *incomplete* otherwise. An automaton is *strongly connected* if for every two states  $p, q \in Q$  there is a word  $w \in \Sigma^*$  such that  $\delta(p, w) = q$ .

By  $\Sigma^i$  we denote the set of all words over  $\Sigma$  of length exactly  $i$  and by  $\Sigma^{\leq i}$  the set of all words over  $\Sigma$  of length at most  $i$ . For two sets of words  $W_1, W_2 \in \Sigma^*$ , by  $W_1 W_2$  we denote their product  $\{w_1 w_2 \in \Sigma^* \mid w_1 \in W_1, w_2 \in W_2\}$ . The empty word is denoted by  $\varepsilon$ . Throughout the paper, by  $n$  we always denote the number of states  $|Q|$ .

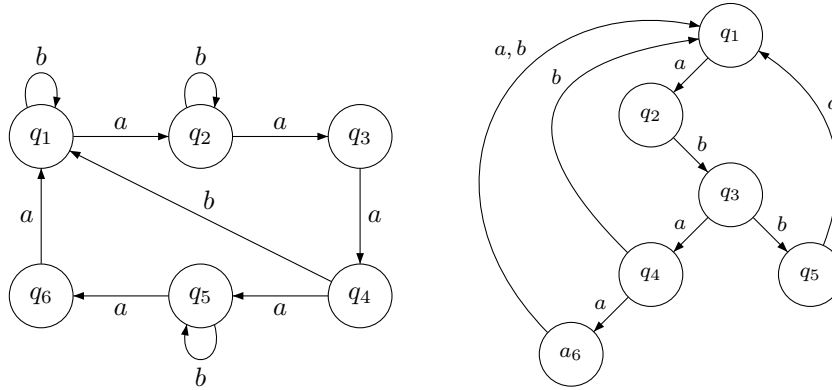
Given  $S \subseteq Q$ , the *image* of  $S$  under the action of  $w$  is  $\delta(S, w) = \{\delta(q, w) \mid q \in S, \delta(q, w) \neq \perp\}$ . The *preimage* of  $S$  under the action of  $w$  is  $\delta^{-1}(S, w) = \{q \in Q \mid \delta(q, w) \in S\}$ . Since

## 12:6 Synchronizing Strongly Connected Partial DFAs

$\mathcal{A}$  is deterministic, for disjoint subsets  $S, T \subseteq Q$ , their preimages under the action of every word  $w \in \Sigma^*$  are also disjoint.

The *rank* of a word  $w$  is the size of the image of  $Q$  under the action of this word, i.e.,  $|\delta(Q, w)|$ . In contrast with complete automata, partial automata may admit words of rank zero; these words are called *mortal*. Words of non-zero rank are called *non-mortal*. A word of rank 1 is called *reset*, and if the automaton admits such a word then it is *synchronizing*. The *reset threshold*  $\text{rt}(\mathcal{A})$  is the length of the shortest reset words of  $\mathcal{A}$ .

We say that a word  $w$  *compresses* a subset  $S \subseteq Q$ , if  $|\delta(S, w)| < |S|$  but  $\delta(S, w) \neq \emptyset$ . A subset that admits a compressing word is called *compressible*. There are two ways to compress a subset  $S \subseteq Q$  with  $|S| \geq 2$ . One possibility is the *pair compression*, which is the same as in the case of a complete automaton, i.e., mapping at least two states  $p, q \in S$  to the same state (but not to  $\perp$ ). The other possibility is to map at least one state from  $S$  to  $\perp$ , but not all states from  $S$  to  $\perp$ . Sometimes, a subset can be compressed in both ways simultaneously.



■ **Figure 1** Left: a strongly connected partial 6-state binary automaton; right: the literal automaton of the prefix code  $\{abaaa, abaab, abab, abba\}$ .

An example of a strongly connected partial automaton is shown in Fig. 1 (left). We have two undefined transitions:  $\delta(q_3, b) = \delta(q_6, b) = \perp$ . The unique shortest reset word is  $bab$ :  $\delta(Q, b) = \{q_1, q_2, q_5\}$ ,  $\delta(Q, ba) = \{q_2, q_3, q_6\}$ , and  $\delta(Q, bab) = \{q_2\}$ . However, in contrast with the case of a complete automaton, the preimage  $\delta^{-1}(\{q_2\}, bab) = \{q_1, q_4\}$  is not  $Q$ .

### 3 Upper Bounds

#### 3.1 Inseparability Equivalence

Let  $\mathcal{A} = (Q, \Sigma, \delta)$  be a partial automaton. We define the inseparability relation  $\equiv$  on  $Q$ . Two states are *separable* if they can be separated by mapping exactly one of them to  $\perp$ , leaving the other one.

► **Definition 1.** The inseparability equivalence  $\equiv$  on  $Q$  is defined as follows:

$$p \equiv q \quad \text{if and only if} \quad \forall u \in \Sigma^* \quad (\delta(p, u) \in Q \Leftrightarrow \delta(q, u) \in Q).$$

The same relation is considered in [5, Section 1.4] if all states of the partial automaton are final. Also, if we replace  $\perp$  with a unique final state, then  $\equiv$  is the well-known Myhill-Nerode congruence on words in a complete automaton. Under a different terminology, it also appears in the context of  $\varepsilon$ -machines, where non-equivalent states are called *topologically distinct* [34].

For a subset  $S \subseteq Q$ , let  $\kappa(S)$  be the number of equivalence classes that have a non-empty intersection with  $S$ . In the automaton from Fig. 1(left) we have three equivalence classes, where  $q_1 \equiv q_4$ ,  $q_2 \equiv q_5$ , and  $q_3 \equiv q_6$ .

Our first auxiliary lemma states that a subset  $S \subseteq Q$  that intersects at least two equivalence classes can be compressed and the number of intersected classes can be decreased with a short word. This is done by mapping to  $\perp$  all the states of  $S$  from at least one equivalence class, but not the whole set  $S$ . A linear upper bound can be inferred from a standard analysis of the corresponding Myhill-Nerode congruence (e.g., [35]), but we will need a more precise bound in terms of  $\kappa(S)$ .

► **Lemma 2.** *Let  $\mathcal{A} = (Q, \Sigma, \delta)$  be a partial automaton, and let  $S \subseteq Q$  be a subset such that  $\kappa(S) \geq 2$ . Then there is a word  $w \in \Sigma^*$  of length at most  $\kappa(Q) - \kappa(S) + 1 \leq n - |S| + 1$  and such that  $1 \leq \kappa(\delta(S, w)) < \kappa(S)$ .*

By an iterative application of Lemma 2, we can easily compress any subset of states to a subset of a single equivalence class.

► **Corollary 3.** *Let  $\mathcal{A} = (Q, \Sigma, \delta)$  be a partial automaton, and let  $S \subseteq Q$  be a non-empty subset. There is a word  $w$  of length at most  $(\kappa(S) - 1)(\kappa(Q) - \kappa(S)/2)$  such that  $\delta(S, w)$  is non-empty and is contained in one inseparability class.*

### 3.2 Fixing Automaton

The other possibility of compressing a subset in a partial automaton is the classical pair compression. This is the only way for compressing a subset with all states in one equivalence class, which is always the case in a complete automaton.

Our next tool to deal with this way of compression is the *fixing automaton*. This is a complete automaton obtained from a partial one, defined as follows.

► **Definition 4 (Fixing automaton).** *For a partial automaton  $\mathcal{A}(Q, \Sigma, \delta)$ , the fixing automaton is the complete automaton  $\mathcal{A}^F = (Q, \Sigma, \delta^F)$  such that the states are fixed instead of having an undefined transition: for every  $q \in Q$  and  $a \in \Sigma$ , we have  $\delta^F(q, a) = q$  if  $\delta(q, a) = \perp$ , and  $\delta^F(q, a) = \delta(q, a)$  otherwise.*

We list some useful properties of the fixing automaton.

► **Lemma 5.** *Let  $\mathcal{A} = (Q, \Sigma, \delta)$  be a partial automaton, let  $S \subseteq Q$ , and let  $w \in \Sigma^*$ . We have  $\delta(S, w) \subseteq \delta^F(S, w)$ . Moreover, if for every state  $q \in S$  we have  $\delta(q, w) \neq \perp$ , then  $\delta(S, w) = \delta^F(S, w)$ .*

► **Lemma 6.** *Let  $\mathcal{A} = (Q, \Sigma, \delta)$  be a partial automaton and let  $S \subseteq Q$  be a non-empty subset. For every word  $w \in \Sigma^*$ , there exists a word  $w' \in \Sigma^*$  of length  $|w'| \leq |w|$  such that  $\emptyset \neq \delta(S, w') \subseteq \delta^F(S, w)$ . In particular, if  $w$  has rank  $r$  in  $\mathcal{A}^F$ , then  $w'$  has rank  $1 \leq r' \leq r$  in  $\mathcal{A}$ .*

► **Corollary 7.** *The minimal non-zero rank of a partial automaton  $\mathcal{A}$  is at most the minimal rank of  $\mathcal{A}^F$ .*

In the general case of a partial automaton, it can happen that we cannot compress some subset  $S$  even if there exists a word of non-zero rank smaller than  $|S|$ . This is the reason why the shortest words of the minimal non-zero rank can be exponentially long and why deciding if there is a word of a given rank is PSPACE-complete [3].

However, in the case of a strongly connected partial automaton, as well as for a complete automaton, every non-mortal word can be extended to a word of the minimal non-zero rank. This is a fundamental difference that allows constructing compressing words iteratively. Note that the fixing automaton of a strongly connected partial one is also strongly connected.

► **Lemma 8.** *Let  $\mathcal{A} = (Q, \Sigma, \delta)$  be a strongly connected partial automaton, and let  $r$  be the minimal non-zero rank over all words. For every non-empty subset  $S \subseteq Q$ , there exists a non-mortal word  $w$  such that  $|\delta(S, w)| \leq r$ .*

### 3.3 Rank Conjecture

The *rank conjecture* (sometimes called *Černý-Pin conjecture*) is a well-known generalization of the Černý conjecture to non-synchronizing automata (e.g., [25]). The rank conjecture is a weaker version of the conjecture originally stated by Pin that was not restricted to the minimal rank and turned out to be false [18]. Some further results on the rank conjecture for strongly connected complete automata are provided in [20].

► **Conjecture 9 (The rank conjecture).** *For an  $n$ -state complete automaton where  $r$  is the minimal rank over all words, there exists a word of rank  $r$  and of length at most  $(n - r)^2$ .*

For partial automata, the rank conjecture is analogous with the exception that  $r$  is the minimal non-zero rank.

► **Theorem 10.** *Let  $\mathcal{A} = (Q, \Sigma, \delta)$  be a strongly connected partial automaton. If the rank conjecture holds for the fixing automaton  $\mathcal{A}^F$ , then it also holds for  $\mathcal{A}$ .*

**Proof.** Let  $r$  be the minimal rank in  $\mathcal{A}^F$  over all words. From the conjecture and by Lemma 6, there exists a word  $w'$  of length at most  $(n - r)^2$  and such that  $\emptyset \neq \delta(Q, w') \subseteq \delta^F(Q, w)$ .

Let  $r' \leq r$  be the minimal rank of  $\mathcal{A}$ . For every  $s = r, r - 1, \dots, r' + 1$ , we inductively construct a word of non-zero rank less than  $s$ , of length at most  $(n - (s - 1))^2$ , and such that  $w'$  is its prefix. Let  $w'v$  be a word of rank at most  $s$  and of length at most  $(n - s)^2$ , and let  $S = \delta(Q, w'v)$ . Suppose that  $\kappa(S) = 1$ . Since  $s$  is not the minimal rank of  $\mathcal{A}$ , by Lemma 8,  $S$  must be compressible. Since its states are inseparable, there must be two distinct states  $p, q \in S$  and a word  $u$  such that  $\delta(q, u) = \delta(p, u) \neq \perp$ . But (from Lemma 5)  $\{p, q\} \subseteq \delta(Q, w'v) \subseteq \delta^F(Q, w'v) \subseteq \delta^F(Q, wv)$ , thus  $\delta^F(Q, wv)$  is compressible in  $\mathcal{A}^F$ , which contradicts that  $w$  has the minimal rank in  $\mathcal{A}^F$ . Hence  $\kappa(S) \geq 2$ , and by Lemma 2,  $\delta(Q, w'v)$  can be compressed with a word  $u$  of length at most  $n - s + 1$ . We have  $|w'vu| \leq (n - s)^2 + n - s + 1 \leq (n - (s - 1))^2$ , which proves the induction step. ◀

The theorem implies that, in the strongly connected case, the rank conjecture is true for complete automata if and only if it is true for partial automata. For instance, we immediately get the result for the class of Eulerian automata. A partial automaton is *Eulerian* if it is strongly connected and the numbers of outgoing and incoming transitions are the same at every state, i.e., for every  $q \in Q$ , we have  $|\{a \in \Sigma \mid \delta(q, a) \in Q\}| = |\{(p, a) \in Q \times \Sigma \mid \delta(p, a) = q\}|$ . The following corollary follows from the facts that the rank conjecture holds for complete Eulerian automata [20] and that the fixing automaton of a partial Eulerian automaton is also Eulerian.

► **Corollary 11.** *The rank conjecture is true for partial Eulerian automata.*

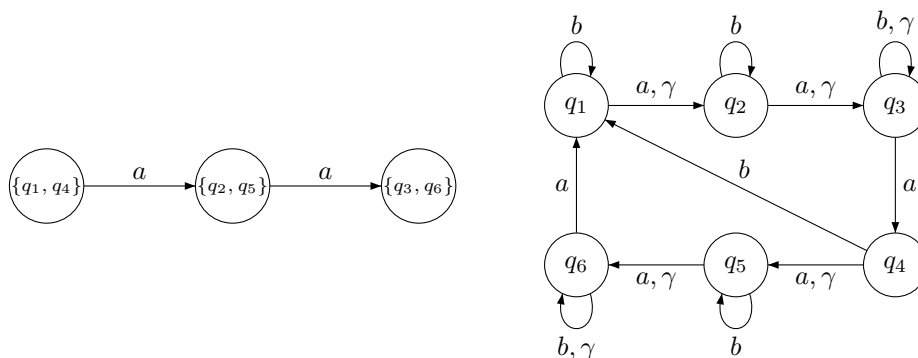


### 3.4 Collecting Automaton

The fixing automaton allows relating the behavior of words in a partial and a complete automaton, but its main disadvantage is that it is not necessarily synchronizing. Therefore, we will need one more tool, called *collecting automaton*. It is an extension of the fixing automaton by an additional letter that allows a quick synchronization into one inseparability class, while it does not affect the length of a shortest synchronizing word for any particular inseparability class.

By  $\mathcal{A}/\equiv = (Q_{\mathcal{A}/\equiv}, \Sigma, \delta_{\mathcal{A}/\equiv})$ , we denote the quotient automaton by the inseparability relation.  $\mathcal{A}/\equiv$  is also a partial automaton, and if  $\mathcal{A}$  is strongly connected, then so is  $\mathcal{A}/\equiv$ . By  $[p] \in Q_{\mathcal{A}/\equiv}$ , we denote the class of a state  $p \in Q$  of the original automaton  $\mathcal{A}$ .

A *collecting tree* of  $\mathcal{A}$  is a tree  $T$  with the set of vertices  $Q_{\mathcal{A}/\equiv}$  and directed edges labeled by letters from  $\Sigma$  in the following way: (a) Edges are labeled by letters from  $\Sigma$  and correspond to transitions in  $\mathcal{A}/\equiv$ : each edge  $([p], [p'], a)$  is such that  $\delta_{\mathcal{A}/\equiv}([p], a) = [p']$ . (b) There is a root  $[r]$  such that the tree is directed toward it. See Fig. 2 in Appendix for an example. Equivalently, it can be seen as a specific partial automaton being a subautomaton of  $\mathcal{A}/\equiv$  whose underlying digraph is a tree directed toward one state. An automaton can have many collecting trees, even for the same  $[r]$ , and every strongly connected automaton has a collecting tree for every class  $[r]$ .



■ **Figure 2** Left: a collecting tree with root  $[q_3] = \{q_3, q_6\}$ ; right: the corresponding collecting automaton of the example from Fig. 1(left).

► **Definition 12** (Collecting automaton). Let  $\mathcal{A} = (Q, \Sigma, \delta)$  be a strongly connected partial automaton and let  $T$  be one of its collecting trees with a root  $[r]$ . The collecting automaton  $\mathcal{A}^{C(T)} = (Q, \Sigma \cup \{\gamma\}, \delta^{C(T)})$  is defined as follows:

- The transition function  $\delta^{C(T)}$  on  $\Sigma$  is defined as in the fixing automaton  $\mathcal{A}^F$ .
- $\gamma \notin \Sigma$  is a fresh letter. Its action is defined according to the edges in  $T$ : Let  $q_1 \in Q \setminus [r]$  be a state. Since  $T$  is a tree directed toward  $[r]$ , there is exactly one edge outgoing from  $[q_1]$ , say  $([q_1], [q_2], a) \in T$  for some  $[q_2] \in Q_{\mathcal{A}/\equiv}$  and  $a \in \Sigma$ . We set  $\delta^{C(T)}(q_1, \gamma) = \delta(q_1, a)$ . Finally, the transition of  $\gamma$  on each state in  $[r]$  is the identity.

A collecting automaton is always strongly connected, as it contains all transitions of the fixing automaton. We prove several properties connecting partial automata and their collecting complete automata. They are preliminary steps toward relating the Černý conjecture for strongly connected partial and complete automata.

► **Lemma 13.** Let  $\mathcal{A} = (Q, \Sigma, \delta)$  be a strongly connected partial automaton, and let  $T$  be one of its collecting trees with a root  $[r]$ . If there is a word over  $\Sigma \cup \{\gamma\}$  synchronizing  $[r]$  in  $\mathcal{A}^{C(T)} = (Q, \Sigma \cup \{\gamma\}, \delta^{C(T)})$ , then there is such a word over  $\Sigma$  that is not longer.

## 12:10 Synchronizing Strongly Connected Partial DFAs

► **Lemma 14.** *Let  $\mathcal{A} = (Q, \Sigma, \delta)$  be a strongly connected partial automaton and let  $T$  be one of its collecting trees. Then  $\mathcal{A}$  is synchronizing if and only if the collecting automaton  $\mathcal{A}^{C(T)} = (Q, \Sigma \cup \{\gamma\}, \delta^{C(T)})$  is synchronizing.*

This also means that the choice of  $T$  does not matter:  $\mathcal{A} = (Q, \Sigma, \delta)$  is synchronizing if and only if all  $\mathcal{A}^{C(T)}$  are synchronizing.

### 3.5 Algorithmic Issues

Checking if a strongly connected partial automaton is synchronizing and finding a minimum-rank word can be done similarly as for a complete automaton, by a suitable generalization of the well-known Eppstein algorithm [13, Algorithm 1]. The same algorithm for checking synchronizability, under different terminology, was described in the context of  $\varepsilon$ -machines [34].

► **Proposition 15.** *Checking if a given strongly connected partial automaton with  $n$  states over an alphabet  $\Sigma$  is synchronizing can be done in  $\mathcal{O}(|\Sigma|n^2)$  time and  $\mathcal{O}(n^2 + |\Sigma|n)$  space. Finding a word of the minimum rank can be done in  $\mathcal{O}(|\Sigma|n^3)$  time and  $\mathcal{O}(n^2 + |\Sigma|n)$  space.*

Furthermore, the problem of checking the synchronizability of a strongly connected partial automaton can be reduced in smaller time to the case of a complete automaton.

► **Theorem 16.** *Given a strongly connected partial automaton, in  $\mathcal{O}(|\Sigma|n \log n)$  time, we can construct a complete automaton that is synchronizing if and only if the given partial automaton is synchronizing.*

**Proof.** We can compute all inseparability classes in  $\mathcal{O}(|\Sigma|n \log n)$  time. This is done by the Hopcroft minimization algorithm [15], if we interpret the partial DFA as a language-accepting DFA with an arbitrary initial state and the sink state  $\perp$  that is its only final state.

Having computed the classes, we can construct a collecting automaton for an arbitrary collecting tree. Note that it can be done in  $\mathcal{O}(|\Sigma|n)$  time by a breadth-first search from a class  $[r]$ . The desired property follows from Lemma 14. ◀

### 3.6 Černý Conjecture

The famous Černý conjecture is the rank conjecture for  $r = 1$ . Let  $\mathfrak{C}(n)$  be the maximum length of the shortest reset words of all  $n$ -state synchronizing complete automata. It is well known that  $\mathfrak{C}(n) \geq (n - 1)^2$  [8]. The Černý conjecture states that  $\mathfrak{C}(n) = (n - 1)^2$ , but the best proved upper bound is cubic [32, 33].

Let  $\mathfrak{C}_P(n)$  be the maximal length of the shortest reset words of all  $n$ -state synchronizing strongly connected partial automata. We show that if the Černý conjecture is true (or another upper bound holds), then a slightly weaker upper bound holds for synchronizing strongly connected partial automata.

To prove the following theorem, we combine several techniques, in particular, the inseparability equivalence, the collecting automaton, and an algebraic upper bound on the reset threshold of a complete automaton with a word of small rank [4].

► **Theorem 17.**

$$\mathfrak{C}_P(n) \leq \mathfrak{C}(n) + \mathcal{O}(n^{4/3}).$$

**Proof.** Let  $\mathcal{A} = (Q, \Sigma, \delta)$  be a synchronizing partial automaton with  $n$  states. Let  $T$  be a collecting tree of  $\mathcal{A}$  with a root class  $[r]$  containing the smallest number of states. We consider the collecting automaton  $\mathcal{A}^{C(T)} = (Q, \Sigma \cup \{\gamma\}, \delta^{C(T)})$ . By Lemma 14,  $\mathcal{A}^{C(T)}$  is synchronizing. We have two cases, depending on the number  $\kappa(Q)$  of inseparability classes of  $\mathcal{A}$ .

First, suppose that  $\kappa(Q) \leq n^{2/3}$ . Then, by Corollary 3 (for  $S = Q$ ), there is a word  $v$  of length at most  $(\kappa(Q) - 1)\kappa(Q)/2 < n^{4/3}$  such that  $\delta(Q, v)$  is non-empty and is contained in one equivalence class, say  $[p]$ . Since  $\mathcal{A}$  is strongly connected, there is a word  $u_{p \rightarrow r}$  of length at most  $n - 1$  whose action maps  $[p]$  into  $[r]$ . Let  $w'$  be a reset word for  $\mathcal{A}^{C(T)}$  of length at most  $\mathfrak{C}(n)$ . In particular,  $w'$  synchronizes  $[r]$ , so by Lemma 13, we get a word  $w$  of length at most  $\mathfrak{C}(n)$  that synchronizes  $[r]$  in  $\mathcal{A}$ . Then,  $vu_{p \rightarrow r}w$  is a reset word for  $\mathcal{A}$  of length at most  $n^{4/3} + n - 1 + \mathfrak{C}(n)$ .

In the second case, we have  $\kappa(Q) > n^{2/3}$ . Then the size of  $[r]$ , which has been chosen to have the smallest size, has at most  $n^{1/3}$  states. Note that  $\gamma^{n-1}$  is a word of rank at most  $n^{1/3}$ . Then we can apply Corollary 20 (cf. [4, Theorem 2]) for  $\mathcal{A}^{C(T)}$  with this word, obtaining that the reset threshold of  $\mathcal{A}^{C(T)}$  is upper bounded by  $(n - 1) + 2(n - 1)\mathfrak{C}(n^{1/3})$ . Using the well-known general cubic upper bound  $n^3/6$  on the reset threshold of a complete automaton (e.g., [25]), we get that there is a reset word  $w'$  for  $\mathcal{A}^{C(T)}$  of length at most  $(n - 1) + 2(n - 1)n/6$ . Now, we return to  $\mathcal{A}$ . By Corollary 3 (for  $S = Q$ ), we get a word  $v$  of length at most  $(n - 1)n/2$  such that  $\delta(Q, v)$  is non-empty and is contained in one equivalence class  $[p]$ . As in the first case, there is a word  $u_{p \rightarrow r}$  of length at most  $n - 1$  whose action maps  $[p]$  into  $[r]$ . By Lemma 13, from  $w'$  we obtain a word  $w$  that synchronizes  $[r]$  and is no longer than  $(n - 1) + 2(n - 1)n/6$ . Finally,  $vu_{p \rightarrow r}w$  is a reset word for  $\mathcal{A}$  of length at most  $(n - 1)n/2 + (n - 1) + (n - 1) + 2(n - 1)n/6 \leq (n - 1)^2$  for  $n \geq 18$ .

From both cases, we conclude that  $\text{rt}(\mathcal{A}) \leq \mathfrak{C}(n) + \mathcal{O}(n^{4/3})$ .  $\blacktriangleleft$

From Theorem 17, it follows that all upper bounds on the reset threshold of a complete automaton transfer to upper bounds for partial automata, up to a subquadratic component. Thus  $\mathfrak{C}_P(n) \leq 0.1654n^3 + \mathcal{O}(n^2)$  [32]. It can be also seen from the proof that if we have a better general upper bound on  $\mathfrak{C}(n)$ , we get a smaller additional term, e.g., if  $\mathfrak{C}(n) \in \mathcal{O}(n^2)$ , then  $\mathfrak{C}_P(n) = \mathfrak{C}(n) + \mathcal{O}(n)$ , so the additional term is at most linear. The additional component is likely not needed, but it is difficult to completely get rid of it in general, as for that we could not lengthen by any means the reset word assumed for a complete automaton. However, it is easy to omit it when reproving particular bounds for complete automata, both combinatorial [25] and based on avoiding words [32, 33]. We conjecture that  $\mathfrak{C}_P(n) = \mathfrak{C}(n)$  for all  $n$ .

### 3.7 Induced Automaton

We develop an algebraic technique applied to partial automata. It will allow us deriving upper bounds on reset thresholds, in particular, in the cases when there exists a short word of a small rank, which is the case of the literal automaton of a prefix code. We base on the results from [4] for complete automata and generalize them to be applied to partial automata. The existing linear algebraic proofs for complete automata do not work for partial ones, because the matrices of transitions may not have a constant sum of the entries in each row.

We need to introduce a few definitions from linear algebra for automata (see, e.g., [4, 19, 24, 33]). Let  $\mathcal{A} = (Q, \Sigma, \delta)$  be a partial DFA. Without loss of generality we assume that  $Q = \{1, \dots, n\}$ . By  $\mathbb{R}^n$ , we denote the real  $n$ -dimensional linear space of row vectors. For a vector  $v \in \mathbb{R}^n$  and an  $i \in Q$ , we denote the vector's value at the  $i$ -th position by  $v(i)$ . Similarly, for a matrix  $M$ , we denote its value in an  $i$ -th row and a  $j$ -th column by  $M(i, j)$ .

## 12:12 Synchronizing Strongly Connected Partial DFAs

A vector  $g$  is *non-negative* if  $g(i) \geq 0$  for all  $i$ , and it is *non-zero* if  $g(i) \neq 0$  for some  $i$ . For a word  $w \in \Sigma^*$ , by  $M(w)$  we denote the  $n \times n$  matrix of the transformation of  $w$  in  $\delta$ :  $M(w)(p, q) = 1$  if  $\delta(p, w) = q$ , and  $M(w)(p, q) = 0$  otherwise. Note that if  $\delta(p, w) = \perp$ , then we have  $M(w)(p, q) = 0$  for all  $q \in Q$ . The usual scalar product of two vectors  $u, v$  is denoted by  $u \odot v$ . The linear subspace *spanned* by a set of vectors  $V$  is denoted by  $\text{span}(V)$ .

Given a transition function  $\delta$  (which defines matrices  $M(w)$ ), call a set of words  $W \subseteq \Sigma^*$  *complete for a subspace*  $V \subseteq \mathbb{R}^n$  *with respect to a vector*  $g \in V$ , if  $V \subseteq \text{span}(\{gM(w) \mid w \in W\})$ . A set of words  $W \subseteq \Sigma^*$  is *complete for a subspace*  $V \subseteq \mathbb{R}^n$  if for every non-negative non-zero vector  $g \in V$ ,  $W$  is complete for  $V$  with respect to  $g$ . Let  $\chi(p)$  denote the characteristic (unitary) vector of  $\{p\}$ . For a subset  $S \subseteq Q$ , we define  $\mathbb{V}(S) = \text{span}(\{\chi(p) \mid p \in S\}) \subseteq \mathbb{R}^n$ .

For example, consider the automaton from Fig. 1(left). Let  $V = \mathbb{V}(\{q_1, q_2, q_5\})$  and let  $W = \{ab, aab\}a^{\leq 5}$ . Let  $g \in V$  be a non-negative non-zero vector, and let  $i$  be such that  $g(i) \neq 0$ . If  $i = 1$  then let  $u = ab$ , and otherwise let  $u = aab$ ; then  $gM(u)$  has exactly one non-zero entry. Then, for each  $j \in \{1, 2, 5\}$ , the vector  $gM(ua^{j'})$  for some  $j'$  has the unique non-zero entry at  $q_j$ . These vectors generate  $V$ , thus  $W$  is complete for  $V$  with respect to  $g$ .

The induced automaton of a partial one is another partial automaton acting on a subset of states  $R \subseteq Q$ . It is built from two sets of words. Let  $W_1$  be a set of words such that  $R = \bigcup_{w \in W_1} \delta(Q, w)$ . For each state in  $R$ , there is some state mapped to it by a word from  $W_1$ . The second set  $W_2$  is any non-empty set of words that enriches its transitions. The induced automaton is  $\mathcal{A}$  restricted to  $R$  with alphabet  $W_2W_1$ . Note that its transition function is well defined, which is ensured by the fact that every word of the form  $w_2w_1 \in W_2W_1$  has the action mapping every state  $q \in Q$  into  $R$  or to  $\perp$ .

► **Definition 18** (Induced automaton). *Let  $W_1, W_2 \subseteq \Sigma^*$  be non-empty and  $R = \{\delta(q, w) \mid q \in Q, w \in W_1, \delta(q, w) \neq \perp\}$ . If  $R$  is non-empty, we define the induced automaton  $\mathcal{A}^{I(W_1, W_2)} = (R, W_2W_1, \delta_{\mathcal{A}^{I(W_1, W_2)}})$ , where the transition function is defined in compliance with the actions of words in  $\mathcal{A}$ , i.e.,  $\delta_{\mathcal{A}^{I(W_1, W_2)}}(q, w) = \delta(q, w)$  for all  $q \in R$  and  $w \in W_2W_1$ .*

We can analyze an induced automaton as a separate one, and synchronize the whole automaton using it, which is particularly profitable when  $R$  is small. Following our previous example, for Fig. 1(left) with  $W_1 = \{b\}$  and  $W_2 = \{ab, aab\}a^{\leq 5}$  we obtain the induced automaton on  $R = \{q_1, q_2, q_5\}$ . Furthermore, it is synchronizing already by a letter from  $W_2W_1$  (e.g.,  $abb$ ), and each its reset word corresponds to a reset word of the original  $\mathcal{A}$ .

The following lemma states that the completeness of a set of words together with the synchronizability and strong connectivity of the whole automaton transfer to the induced automaton. It generalizes [4, Theorem 2] to partial automata, and the proof uses a recursion instead of an augmenting argument.

► **Lemma 19.** *Let  $\mathcal{A} = (Q, \Sigma, \delta)$  be a strongly connected synchronizing partial automaton and let  $W_1$  and  $W_2$  be two non-empty sets of words over  $\Sigma$ . Let  $\mathcal{A}^{I(W_1, W_2)} = (R, W_2W_1, \delta_{\mathcal{A}^{I(W_1, W_2)}})$  be the induced automaton. If  $W_2$  is complete for  $\mathbb{V}(Q) = \mathbb{R}^n$ , then  $W_2W_1$  is complete for  $\mathbb{V}(R)$ , and  $\mathcal{A}^{I(W_1, W_2)}$  is synchronizing and strongly connected.*

► **Corollary 20.** *Let  $\mathcal{A} = (Q, \Sigma, \delta)$  be a strongly connected synchronizing partial automaton with  $n$  states, and let  $w \in \Sigma^*$  be a word such that  $R = \delta(Q, w) \neq \emptyset$ . Let  $W_1 = \{w\}$ ,  $W_2 = \Sigma^{\leq n-1}$ , and  $\mathcal{A}^{I(W_1, W_2)} = (R, \Sigma^{\leq n-1}\{w\}, \delta_{\mathcal{A}^{I(W_1, W_2)}})$  be the induced automaton. Then  $\text{rt}(\mathcal{A}) \leq |w| + (|w| + n - 1) \cdot \text{rt}(\mathcal{A}^{I(W_1, W_2)})$ .*

The corollary directly follows from Lemma 19, since  $\Sigma^{\leq n-1}$  is always complete for  $\mathbb{V}(Q)$  in the case of a strongly connected synchronizing partial automaton. It is useful for deriving upper bounds for automata with a word of small rank. Having such a word  $w$ , we can further synchronize  $R$  through the induced automaton instead of trying to do this directly. Although

every letter of  $\mathcal{A}^{I(W_1, W_2)}$  corresponds to a word of length  $|w| + n - 1$  in the original  $\mathcal{A}$ , if  $R$  is small enough, this yields a better upper bound. We show its application in the next subsection.

### 3.8 The Literal Automaton of a Finite Prefix Code

We use the obtained results about induced automata to get better bounds for partial literal automata of finite prefix codes. To do so, we first need to prove that such automata admit short enough words of small rank. It is known that every complete literal automaton over an alphabet  $\Sigma$  has a word of rank and length at most  $\lceil \log_{|\Sigma|} n \rceil$  ([4, Lemma 16], cf. [6, Lemma 14]). However, this is no longer true for non-mortal words in partial literal automata [29] and no similar statement was known for any wider class than complete literal automata. We prove that there exist  $\mathcal{O}(\log n)$ -rank non-mortal words of length  $\mathcal{O}(n)$  in such automata, excluding the case of a code with only one word. Then we use this result to provide an  $\mathcal{O}(n \log^3 n)$  upper bound on the reset threshold of  $n$ -state synchronizing partial literal automata, matching the known upper bound for complete literal automata [4].

We start with a special case of one-word codes. A non-empty word  $w$  is called *primitive* if it is not a power of another word, i.e.,  $w \neq u^k$  for every word  $u$  and  $k \geq 2$ . The upper bound on  $\text{rt}(\mathcal{A}_X)$  follows from a result of Weinbaum ([14, 39]).

► **Proposition 21.** *Let  $X = \{x\}$  be a one-word prefix code, and suppose that  $x = y^k$ , where  $y$  is a non-empty primitive word and  $k \geq 1$ . Then  $\mathcal{A}_X$  has rank  $k$ . If  $\mathcal{A}_X$  is synchronizing, then  $\text{rt}(\mathcal{A}_X) \leq \frac{|x|}{2}$ , and this bound is tight.*

In the remaining cases, there always exists a word of linear length and logarithmic rank.

► **Theorem 22.** *Let  $X$  be a prefix code with at least two words. Let  $\mathcal{A}_X = (Q, \Sigma, \delta)$  be its partial literal automaton with  $n$  states and height  $h$ . Then there exists a word of length at most  $2h$  and of rank at most  $\lceil \log_2 hn \rceil + \lceil \log_2 h \rceil$  for  $\mathcal{A}_X$ . Moreover, such a word can be found in polynomial time in  $n = |Q|$ .*

**Proof idea.** The general idea is as follows. We construct a word from the theorem in two phases. First, we define an auxiliary *filtering* algorithm computing some function  $\alpha: \Sigma^* \rightarrow \Sigma^*$ . We consider the results of the algorithm for a lot of short (logarithmic) input words  $w$  and show that at least one of them satisfies that  $\alpha(w)$  is non-mortal, has length at most  $h$ , and every state from  $Q$  is either sent to  $\perp$  or goes through the root by its action. Then, we use specific properties of the image  $\delta(Q, \alpha(w))$  to divide it into two disjoint sets: one that has up to  $h$  states, but on a single specific path, and the other one with a small (logarithmic) number of states. In the second phase, we construct a word  $v$  of length also bounded by  $h$ , such that its action map all the states from the mentioned specific path to a subset of at most logarithmic size. The concatenation of both words  $\alpha(w)v$  is a word of length at most  $2h$  satisfying the theorem. ◀

► **Corollary 23.** *Let  $\mathcal{A}_X$  be a partial literal automaton with  $n$  states. If it is synchronizing, its reset threshold is at most  $\mathcal{O}(n \log^3 n)$ . If the Cerny conjecture holds, then it is  $\mathcal{O}(n \log^2 n)$ .*

**Proof.** If  $|X| = 1$  then the bound follows from Proposition 21. If  $|X| \geq 2$ , from Theorem 22, we get a word  $w$  of length  $\mathcal{O}(n)$  and rank  $\mathcal{O}(\log n)$ . Then we use Corollary 20 with  $w$ , which yields the upper bound  $\mathcal{O}(n) + \mathcal{O}(n) \cdot \text{rt}(\mathcal{B})$ , where  $\mathcal{B}$  is an induced automaton with  $\mathcal{O}(\log n)$  states. Then we use upper bounds on the reset threshold of a complete DFA ([32, 33]) transferred to strongly connected partial DFAs by Theorem 17. ◀

#### 4 Lower Bounds for Properly Incomplete Automata

We conclude with observations for transferring lower bounds from the complete case to the partial case. Of course, in general, this is trivial, since a complete automaton is a special case of a partial one. On the other hand, letters with all transitions undefined cannot be used for synchronization. Hence we need to add a restriction to exclude these cases and see the effect of usable incomplete transitions. A partial automaton is *properly incomplete* if there is at least one letter whose transition is defined for some state and is undefined for some other state.

For an automaton  $\mathcal{A}$ , let the length of the shortest words of rank  $r$  be called the *rank threshold*  $\text{rt}(\mathcal{A}, r)$ . We show that bounding the rank/reset threshold of a strongly connected properly incomplete automaton is related to bounding the corresponding threshold of a complete automaton. A general construction for this is the following.

► **Definition 24** (Duplicating automaton). *For a complete automaton  $\mathcal{A} = (Q, \Sigma, \delta_{\mathcal{A}})$ , we construct the duplicating automaton  $\mathcal{A}^{\text{D}} = (Q \cup Q', \Sigma \cup \{\gamma\}, \delta_{\mathcal{A}^{\text{D}}})$  as follows. Assume that  $Q = \{q_1, \dots, q_n\}$ . Then  $Q' = \{q'_1, \dots, q'_n\}$  is a set of fresh states disjoint with  $Q$  and  $\gamma \notin \Sigma$  is a fresh letter. For all  $1 \leq i \leq n$  and  $a \in \Sigma$ , we define:  $\delta_{\mathcal{A}^{\text{D}}}(q_i, a) = q_i$ ,  $\delta_{\mathcal{A}^{\text{D}}}(q_i, \gamma) = q'_i$ ,  $\delta_{\mathcal{A}^{\text{D}}}(q'_i, a) = \delta_{\mathcal{A}}(q_i, a)$ , and  $\delta_{\mathcal{A}^{\text{D}}}(q'_i, \gamma) = \perp$ .*

The duplicating automaton turns out to be a partial DFA counterpart to the recent Volkov's construction of a complete DFA [37]. The duplicating automaton  $\mathcal{A}^{\text{D}}$  has twice the number of states of  $\mathcal{A}$  and is properly incomplete. Also, it is strongly connected if  $\mathcal{A}$  is.

► **Proposition 25.** *Let  $\mathcal{A} = (Q, \Sigma, \delta_{\mathcal{A}})$  be a strongly connected complete automaton. For all  $1 \leq r < n$ , we have  $\text{rt}(\mathcal{A}^{\text{D}}, r) = 2 \text{rt}(\mathcal{A}, r)$ .*

From Proposition 25, it follows that we cannot expect a better upper bound on the reset threshold of a properly incomplete strongly connected automaton than  $0.04135n^3 + \mathcal{O}(n^2)$ , unless we can improve the best general upper bound on the reset threshold of a complete automaton, which currently is roughly  $0.1654n^3 + \mathcal{O}(n^2)$  [32, 33]. We can also show a lower bound on the largest possible reset threshold, using the Rystsov's construction of automata with long shortest mortal words [27].

► **Proposition 26.** *For every  $n$ , there exists a strongly connected properly incomplete  $n$ -state automaton with the reset threshold  $\frac{n^2-n}{2}$ .*

---

#### References

- 1 P. Babari, K. Quaas, and M. Shirmohammadi. Synchronizing data words for register automata. In *MFCS*, pages 15:1–15:15, 2016.
- 2 M.-P. Béal, E. Czeizler, J. Kari, and D. Perrin. Unambiguous automata. *Mathematics in Computer Science*, 1(4):625–638, 2008.
- 3 M. Berlinkov. On Two Algorithmic Problems about Synchronizing Automata. In *Developments in Language Theory*, LNCS, pages 61–67. Springer, 2014.
- 4 M. Berlinkov and M. Szykuła. Algebraic synchronization criterion and computing reset words. *Information Sciences*, 369:718–730, 2016.
- 5 J. Berstel, D. Perrin, and C. Reutenauer. *Codes and Automata*. Encyclopedia of Mathematics and its Applications. Cambridge University Press, 2009.
- 6 M. T. Biskup and W. Plandowski. Shortest synchronizing strings for Huffman codes. *Theoretical Computer Science*, 410(38-40):3925–3941, 2009.
- 7 V. Bruyère. On maximal codes with bounded synchronization delay. *Theoretical Computer Science*, 204(1):11–28, 1998.

- 8 J. Černý. Poznámka k homogénnym experimentom s konečnými automatami. *Matematicko-fyzikálny Časopis Slovenskej Akadémie Vied*, 14(3):208–216, 1964. In Slovak.
- 9 D. Chistikov, P. Martyugin, and M. Shirmohammadi. Synchronizing automata over nested words. *Journal of Automata, Languages and Combinatorics*, 24(2-4):219–251, 2019.
- 10 L. Doyen, L. Juhl, K. G. Larsen, N. Markey, and M. Shirmohammadi. Synchronizing words for weighted and timed automata. In *FSTTCS*, pages 121–132, 2014.
- 11 L. Doyen, T. Massart, and M. Shirmohammadi. Robust Synchronization in Markov Decision Processes. In *CONCUR*, pages 234–248, 2014.
- 12 L. Doyen, T. Massart, and M. Shirmohammadi. The complexity of synchronizing Markov decision processes. *J. Comput. Syst. Sci.*, 100:96–129, 2019.
- 13 D. Eppstein. Reset sequences for monotonic automata. *SIAM Journal on Computing*, 19:500–510, 1990.
- 14 T. Harju and D. Nowotka. On unique factorizations of primitive words. *Theor. Comput. Sci.*, 356(1-2):186–189, 2006.
- 15 J. Hopcroft. An  $n \log n$  algorithm for minimizing states in a finite automaton. In Zvi Kohavi and Azaria Paz, editors, *Theory of Machines and Computations*, pages 189–196. Academic Press, 1971.
- 16 D. A. Huffman. A method for the construction of minimum-redundancy codes. *Proceedings of the IRE*, 40(9):1098–1101, 1952.
- 17 B. Imreh and M. Steinby. Directable nondeterministic automata. *Acta Cybern.*, 14(1):105–115, 1999.
- 18 J. Kari. A counter example to a conjecture concerning synchronizing word in finite. *EATCS Bulletin*, 73:146–147, 2001.
- 19 J. Kari. Synchronizing finite automata on Eulerian digraphs. *Theoretical Computer Science*, 295(1-3):223–232, 2003.
- 20 J. Kari, A. Ryzhikov, and A. Varonka. Words of minimum rank in deterministic finite automata. In *Developments in Language Theory - 23rd International Conference, DLT 2019, Warsaw, Poland, August 5-9, 2019, Proceedings*, pages 74–87, 2019.
- 21 S. Kiefer and C. Mascle. On Finite Monoids over Nonnegative Integer Matrices and Short Killing Words. In *STACS, LIPIcs*, 2019.
- 22 K. G. Larsen, S. Laursen, and J. Srba. Synchronizing strategies under partial observability. In *International Conference on Concurrency Theory*, pages 188–202. Springer, 2014.
- 23 B. K. Natarajan. An algorithmic approach to the automated design of parts orienters. In *Foundations of Computer Science, 27th Annual Symposium on*, pages 132–142, 1986.
- 24 J.-E. Pin. Utilisation de l’algèbre linéaire en théorie des automates. In *Actes du 1er Colloque AFCET-SMF de Mathématiques Appliquées II*, AFCET, pages 85–92, 1978. In French.
- 25 J.-E. Pin. On two combinatorial problems arising from automata theory. In *Proceedings of the International Colloquium on Graph Theory and Combinatorics*, volume 75 of *North-Holland Mathematics Studies*, pages 535–548, 1983.
- 26 I. K. Rystsov. Polynomial complete problems in automata theory. *Information Processing Letters*, 16(3):147–151, 1983.
- 27 I. K. Rystsov. Reset words for commutative and solvable automata. *Theoretical Computer Science*, 172(1-2):273–279, 1997.
- 28 A. Ryzhikov. Mortality and synchronization of unambiguous finite automata. In *Combinatorics on Words - 12th International Conference, WORDS 2019, Loughborough, UK, September 9-13, 2019, Proceedings*, pages 299–311, 2019.
- 29 A. Ryzhikov and M. Szykuła. Finding Short Synchronizing Words for Prefix Codes. In Igor Potapov, Paul Spirakis, and James Worrell, editors, *MFCS 2018*, volume 117 of *LIPIcs*, pages 21:1–21:14. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2018.
- 30 S. Sandberg. Homing and synchronizing sequences. In *Model-Based Testing of Reactive Systems*, volume 3472 of *LNCS*, pages 5–33. Springer, 2005.

## 12:16 Synchronizing Strongly Connected Partial DFAs

- 31 H. Shabana. Exact synchronization in partial deterministic automata. *Journal of Physics: Conference Series*, 1352:012047, 2019.
- 32 Y. Shitov. An Improvement to a Recent Upper Bound for Synchronizing Words of Finite Automata. *Journal of Automata, Languages and Combinatorics*, 24(2–4):367–373, 2019.
- 33 M. Szykuła. Improving the Upper Bound on the Length of the Shortest Reset Word. In *STACS 2018, LIPIcs*, pages 56:1–56:13. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2018.
- 34 N. F. Travers and J. P. Crutchfield. Exact Synchronization for Finite-State Sources. *Journal of Statistical Physics*, 145(5):1181–1201, 2011.
- 35 L. Trevisan. Notes on state minimization. <https://people.eecs.berkeley.edu/~luca/cs172-07/notemindfa.pdf>, 2007.
- 36 M. V. Volkov. Synchronizing automata and the Černý conjecture. In *Language and Automata Theory and Applications*, volume 5196 of *LNCS*, pages 11–27. Springer, 2008.
- 37 M. V. Volkov. Slowly synchronizing automata with idempotent letters of low rank. *Journal of Automata, Languages and Combinatorics*, 24(2–4):375–386, 2019.
- 38 V. Vorel. Subset synchronization and careful synchronization of binary finite automata. *Int. J. Found. Comput. Sci.*, 27(5):557–578, 2016.
- 39 C. M. Weinbaum. Unique subwords in nonperiodic words. *Proceedings of the American Mathematical Society*, 109(3):615–619, 1990.