Improving the Upper Bound on the Length of the Shortest Reset Words

Marek Szykuła
Institute of Computer Science,
University of Wrocław, Wrocław, Poland
msz@cs.uni.wroc.pl

Abstract

We improve the best known upper bound on the length of the shortest reset words of synchronizing automata. The new bound is slightly better than $114n^3/685+O(n^2)$. The Černý conjecture states that $(n-1)^2$ is an upper bound. So far, the best general upper bound was $(n^3-n)/6-1$ obtained by J.-E. Pin and P. Frankl in 1982. Despite a number of efforts, it remained unchanged for about 35 years.

To obtain the new upper bound we utilize avoiding words. A word is avoiding for a state q if after reading the word the automaton cannot be in q. We obtain upper bounds on the length of the shortest avoiding words, and using the approach of Trahtman from 2011 combined with the well-known Frankl theorem from 1982, we improve the general upper bound on the length of the shortest reset words. For all the bounds, there exist polynomial algorithms finding a word of length not exceeding the bound.

2012 ACM Subject Classification Mathematics of computing → Combinatoric problems, Theory of computation → Formal languages and automata theory

Keywords and phrases Avoiding Word, Černý Conjecture, Reset Length, Reset Threshold, Synchronizing Automaton

Digital Object Identifier 10.4230/LIPIcs.STACS.2018.56

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

Acknowledgements I thank Mikhail Berlinkov, Costanza Catalano, Vladimir Gusev, Jakub Kośmider, Jakub Kowalski, and the anonymous reviewers for proofreading and useful comments.

1 Introduction

We deal with deterministic finite complete (semi)automata $\mathscr{A}(Q,\Sigma,\delta)$, where Q is the set of *states*, Σ is the input *alphabet*, and $\delta\colon Q\times\Sigma\to Q$ is the *transition function*. We extend δ to the function $Q\times\Sigma^*\to Q$ in the usual way. Throughout the paper, by n we denote the number of states $|Q|$.

By $\Sigma^{\le i}$ we denote the set of all words over Σ of length at most i. Given a state $q\in Q$ and a word $w\in\Sigma^*$ we write shortly $q\cdot w=\delta(q,w)$. Given a subset $S\subseteq Q$ we write $S\cdot w$ for the image $\{q\cdot w \mid q\in S\}$. Then, $S\cdot w^{-1}$ is the preimage $\{q\in Q \mid q\cdot w\in S\}$, and when S is a singleton we also write $q\cdot w^{-1}=\{q\}\cdot w^{-1}$.

The *rank* of a word $w\in\Sigma^*$ is the cardinality of the image of Q under the action of this word: $|Q\cdot w|$. A word is *reset* or *synchronizing* if it has rank 1. An automaton is *synchronizing* if it admits a reset word. The *reset threshold* $\mathrm{rt}(\mathscr{A})$ is the length of the shortest reset words.

35th Symposium on Theoretical Aspects of Computer Science (STACS 2018).
Editors: Rolf Niedermeier and Brigitte Vallée; Article No. 56; pp. 56:1–56:13
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

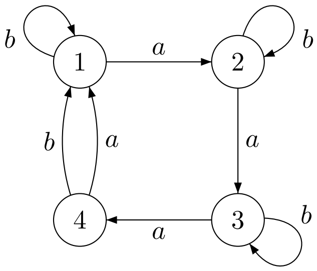

Figure 1 The Černý automaton with 4 states.

We say that a word $w \in \Sigma^*$ *compresses* a subset $S \subseteq Q$ if $|S \cdot w| < |S|$. A word $w \in \Sigma^*$ *avoids* a state $q \in Q$ if $q \notin Q \cdot w$. A state that admits an avoiding word is *avoidable.* We also say that a state q is *avoidable from a subset* S if there exists a word w such that $q \notin S \cdot w$.

The famous Černý conjecture, formally formulated in 1969, is one of the most longstanding open problems in automata theory. It states that every synchronizing n-state automaton has a reset word of length at most $(n-1)^2$. This bound would be tight, since it is reached for every n by the Černý automata [7]. Fig. 1 shows the Černý automaton with $n = 4$ states. Its shortest reset word is ba^3ba^3b.

The first general upper bound for the reset threshold given by Černý in [7] was $2^n - n - 1$. Later, it was improved several times: $\frac{1}{2}n^3 - \frac{3}{2}n^2 + n + 1$ given by Starke [23] in 1966, $\frac{1}{3}n^3 - \frac{3}{2}n^2 + 25/6n - 4$ by Černý, Pirická, and Rosenauerová [8] in 1971, $\frac{7}{27}n^3 - 17/18n^2 + 17/6n - 3$ by Pin [19] in 1978, and $(\frac{1}{2} - \frac{\pi}{36})n^3 + o(n^3)$ by Pin [21] in 1981.

Then, the well known upper bound was established in 1982 by Pin and Frankl through the following combinatorial theorem:

▶ **Theorem 1** ([12, 21]). *Let $\mathscr{A}(Q, \Sigma, \delta)$ be a strongly connected synchronizing automaton, and consider a subset $S \subseteq Q$ of cardinality ≥ 2. Then there exists a word such that $|S \cdot w| < |S|$ of length at most*

$$\frac{(n - |S| + 2) \cdot (n - |S| + 1)}{2}.$$

For integers $1 \leq i, j \leq n$ we define

$$C(j, i) \quad = \quad \sum_{s=i+1}^{j} \frac{(n - s + 2) \cdot (n - s + 1)}{2}.$$

From Theorem 1, $C(j, i)$ is an upper bound on the length of the shortest words compressing a subset of size j to a subset of size at most i: starting from a subset S of size j, we iteratively apply Theorem 1 to bound the length of a shortest word compressing each (in the worst case) of the obtained subsets of sizes $j, j-1, \ldots, i+1$. This yields the well known bound on the length of the shortest reset words:

$$\mathrm{rt}(\mathscr{A}) \leq C(n, 1) = \frac{n^3 - n}{6}.$$

This bound was also discovered independently in [17]. Actually, the best bound was $\frac{n^3-n}{6} - 1$ (for $n \geq 4$), since Pin [21] proved that (for $n \geq 4$) there is a word compressing Q to a subset of size $n - 3$ by a word of length 9 (instead of 10). Theorem 1 also bounds the lengths of a compressing word found by a greedy algorithm (e.g. [1, 11]), which is an algorithm finding a reset word by iterative application of a shortest word compressing the current subset. For about 35 years, there was no progress in improving the bound in the general case.

However, better bounds have been obtained for a lot of special classes of automata, for example for oriented (monotonic) automata [11], circular automata [10], Eulerian automata [15], aperiodic automata [26], generalized and weakly monotonic automata [2, 29], automata with a sink (zero) state [18], one-cluster automata [3, 25], quasi-Eulerian and quasi-one-cluster automata [5], automata respecting intervals of a directed graph [14], decoders of finite prefix codes [4, 6], automata with a letter of small rank [4, 20], and 1-contracting automata [9]. See also [28] for a survey.

In 2011, Trahtman claimed the better upper bound $(7n^3+6n-16)/48$ [27]. Unfortunately, the proof contains an error, and so the result remains unproved. The idea was to utilize avoiding words; [27, Lemma 3] states that for every $q \in Q$ there exists an avoiding word of length at most $n-1$. A counterexample to this was found in [13], where it was also suggested that providing any linear upper bound on the length of avoiding words would also imply an improvement for the upper bound on the reset threshold.

The avoiding word problem is similar to synchronization: instead of bringing the automaton into one state, we ask how long word we require to not being in a particular state. For the automaton from Fig. 1, the shortest avoiding words for states 1, 2, 3, 4 are ba, baa, $baaa$, and b, respectively. So far, only a trivial cubic upper bound $\mathrm{rt}(\mathscr{A})+1$ was known for synchronizing automata. Avoiding words do not necessarily exist in general, but they always do for every state in the case of a synchronizing automaton unless there is a *sink state* ([18]), for which all letters act like identity.

The main contributions in this paper are as follows: We prove upper bounds on the length of the shortest avoiding words, in particular the quadratic bound $(n-1)(n-2)+2$. Also, the length of avoiding words is connected with the length of compressing words. We show that for every state q and a subset of states S, either there is a short avoiding word for q from S or a short compressing word for S. This connection leads to the main idea for the improvement of the general upper bound on the reset threshold: either improve by avoiding words, or use shorter compressing words directly to reduce the bound obtained by Theorem 1. In contrast to the previous approaches, which bounded the length of the compressing words independently for each size $|S|$, the new bound utilizes a conditional approach.

The new upper bound is

$$(85059n^3 + 90024n^2 + 196504n - 10648)/511104,$$

which is slightly better than the much simpler formula $114n^3/685 + O(n^2)$. The latter improves the coefficient of n^3 by $1/4110$. In the last section we discuss open problems and further possibilities for improvements.

2 Avoiding words

For the next lemma, we need to introduce a few definitions from linear algebra for automata (see, e.g., [4, 15, 20]). By $\mathbb{R}^n$ we denote the real n-dimensional linear space of row vectors. Without loss of generality we assume that $Q = \{1, 2, \dots, n\}$. For a vector $v \in \mathbb{R}^n$, we denote the value at an i-th position by $v(i)$. For a subset $S \subseteq Q$, by $[S]$ we denote its characteristic row vector, which has $[S](i) = 1$ if $i \in S$, and $[S](i) = 0$ otherwise. Similarly, for a matrix M, we denote the value at an i-th row and a j-th column by $M(i,j)$. For a word $w \in \Sigma^*$, by $[w]$ we denote the $n \times n$ matrix of the transformation of w: $[w](i,j) = 1$ if $i \cdot w = j$ (state i is mapped to state j by the transformation of w), and $[w](i,j) = 0$ otherwise.

Right matrix multiplication corresponds to concatenation of two words; i.e. for every two words $u, v \in \Sigma^*$ we have $[uv] = [u] \cdot [v]$. For a subset S we have $([S][u])(i)$ equal to the number of states from S mapped by the transformation of u to state i. In particular,

$([S][u])(i) \geq 1$ if and only if $[S \cdot u](i) = 1$. Note that for $w \in \Sigma^*$, the matrix $[w]$ contains exactly one 1 in each row. Therefore, these are *stochastic* matrices, and we have the property that for any $v \in \mathbb{R}^n$, right matrix multiplication by $[w]$ preserves the sum of the entries, i.e. $\sum_{i \in Q}[v](i) = \sum_{i \in Q}([v][w])(i)$.

For example, for the automaton from Fig. 1 we have:

$$[a] = \begin{pmatrix} 0&1&0&0 \\ 0&0&1&0 \\ 0&0&0&1 \\ 1&0&0&0 \end{pmatrix}, \ [b] = \begin{pmatrix} 1&0&0&0 \\ 0&1&0&0 \\ 0&0&1&0 \\ 1&0&0&0 \end{pmatrix}, \ [ba] = \begin{pmatrix} 0&1&0&0 \\ 0&0&1&0 \\ 0&0&0&1 \\ 0&1&0&0 \end{pmatrix}.$$

If $[S] = [1,0,1,1]$, then $[S][ba] = [S][b][a] = [0,2,0,1]$.

The linear subspace *spanned* by a set of vectors V is denoted by $\operatorname{span}(V)$. Given a linear subspace $L \subseteq \mathbb{R}^n$ and an $n \times n$ matrix m, the linear subspace mapped by m is $Lm = \{vm \mid v \in L\}$. The *dimension* of a linear subspace L is denoted by $\dim(L)$.

The following key lemma states that by a short (linear) word we can either avoid a state (or one of the states from some set A) from the current subset or compress the current subset.

▶ **Lemma 2.** *Let $\mathscr{A}(Q, \Sigma, \delta)$ be an n-state automaton. Consider a non-empty subset $S \subseteq Q$ and a non-empty proper subset $A \subsetneq S$. Suppose that there is a word $w \in \Sigma^*$ such that $A \not\subseteq S \cdot w$. Then there exists a word w length at most $n - |A|$ satisfying either*

1. *$A \not\subseteq S \cdot w$, or*
2. *$|S \cdot w| < |S|$.*

Proof. Let $L_i = \operatorname{span}(\{[S][w] \mid w \in \Sigma^{\leq i}\})$. We consider the following sequence of linear subspaces:

$$L_0 \subseteq L_1 \subseteq L_2 \subseteq \ldots,$$

and use the ascending chain condition (see, e.g., [4, 15, 20, 24]):

- If $L_k = L_{k+1}$, then we claim that also $L_{k+1} = L_{k+2} = \ldots$ holds. Observe that for all $i \geq 0$ we have:

  $$L_{i+1} = \operatorname{span}\left(L_i \cup \bigcup_{a \in \Sigma} L_i[a]\right).$$

  Hence, if $L_k = L_{k+1}$, then for $i = k$ we obtain

  $$L_{k+1} = \operatorname{span}\left(L_{k+1} \cup \bigcup_{a \in \Sigma} L_{k+1}[a]\right) = L_{k+2},$$

  and so $L_{k+i} = L_k$ for all $i \geq 0$.
- Let i be the smallest integer such that $L_i = L_{i+1}$. Then $m = \dim(L_i)$ is the maximum among the dimensions of the subspaces from the above sequence.
- $\dim(L_0) = 1$ and the dimensions grow by at least 1 up to m. Hence, we have

  $$\dim(L_{n-|A|}) \geq \min\{m, n - |A| + 1\}.$$

Note that if for a word w the vector $v = [S][w]$ has $v(q) = 0$ for some $q \in A$, then $q \notin S \cdot w$, and we have Case (1). If $v = [S][w]$ has $v(q) \geq 2$ for some $q \in A$, then a pair of states from S is compressed by the action of w (to state q), and we have Case (2).

Now, we show that in the spanning set of $L_{n-|A|}$ there must be a vector that contains either 0 or an integer ≥ 2 at the position corresponding to a state from A, which implies that there exists a word w of length at most $n - |A|$ satisfying either Case (1) or Case (2). Suppose for a contradiction that this is not the case. Every vector $v \in L_k$ is a linear combination of the vectors from the spanning set; let c be the sum of the coefficients of the spanning vectors in such a linear combination. Every vector $[S][w]$ in the spanning set has the sum of elements equal to $|S|$ and has 1 at all the positions corresponding to the states from A. Hence, the sum of the entries in v is equal to $c|S|$, and at every position corresponding to the states from A we have value c. The sum of the entries at the positions corresponding to the states from $Q \setminus A$ equals $c(|S| - |A|)$. Therefore, every $q \in A$ satisfies the following equality:

$$v(q) = \frac{1}{|S| - |A|} \cdot \sum_{p \in Q \setminus A} v(p).$$

It follows that the values at the positions corresponding to the states from A are completely determined by the sum of the values from the other positions, which means that the dimension of $L_{n-|A|}$ is at most $n - |A|$. We assumed in the lemma that there exists a word w avoiding a state from A. Hence, $[S][w]$ has 0 at some position corresponding to a state from A, and therefore breaks the above equality for this state, as the right side is non-zero. Therefore, the subspace $L_{|w|}$ must have a larger dimension than $\dim(L_{n-|A|})$. This means that the dimension of $L_{n-|A|}$ is not maximal, which contradicts $\dim(L_{n-|A|}) \geq \min\{m, n - |A| + 1\}$. ◀

Lemma 2 can be applied iteratively to obtain a word compressing the given subset to the desired size.

▶ **Lemma 3.** *Let $\mathscr{A}(Q, \Sigma, \delta)$ be an n-state automaton. Consider a non-empty subset $S \subseteq Q$ and a non-empty proper subset $A \subsetneq S$. Let $k \geq 1$ be an integer. Suppose that there exists a word $w \in \Sigma^*$ such that $A \not\subseteq S \cdot w$. Then there is a word w of length at most $k(n - |A|)$ satisfying either:*

1. *$A \not\subseteq S \cdot w$, or*
2. *$|S \cdot w| \leq |S| - k$.*

Proof. If Case (1) holds for some $w \in \Sigma^{\leq k(n-|A|)}$ then we are done; suppose this is not the case.

We iteratively apply Lemma 2 k times for subset A starting from subset S: For $i = 1, \ldots, k$ we apply the lemma for the subset $S \cdot w_1 \ldots w_{i-1}$, where $w_j \in \Sigma^{\leq n-|A|}$ is the word obtained from the lemma in the j-th iteration.

In every iteration, we must get Case (2) of Lemma 2 ($|S \cdot w| < |S|$), as otherwise $A \not\subseteq S \cdot w_1 \ldots w_i$, which contradicts our assumption that Case (1) does not hold for every word of length at most $k(n - |A|) \geq i(n - |A|)$. Also, for $i \leq k - 1$, we must have $A \subset S \cdot w_1 \ldots w_i$ (i.e. A is a proper subset); otherwise $A \not\subseteq S \cdot w_1 \ldots w_i a$ for some letter $a \in \Sigma$ as A contains a state that can be avoided from S, and this word has length at most $k(n - |A|)$ which again contradicts our assumption. Therefore, the conditions are met for every iteration so we can apply the lemma k times.

It follows that the obtained word $w_1 \ldots w_k$ is such that $|S \cdot w_1 \ldots w_k| \leq |S| - k$. ◀

If the subset A of states to avoid is large, the following approach can lead to a better bound:

▶ **Lemma 4.** *Let $\mathscr{A}(Q, \Sigma, \delta)$ be an n-state automaton. Consider a non-empty subset $S \subseteq Q$ and a non-empty subset $A \subseteq S$. If there exists a word $w \in \Sigma^*$ such that $A \not\subseteq S \cdot w$, then there exists such a word of length at most $(|S| - |A|)(n - |A|) + 1$.*

Proof. As in the proof of Lemma 3, we iteratively apply Lemma 2 at most $|S|-|A|$ times for subset A starting from subset S, stopping if the conditions are not met. It is possible that we do not do any iteration, which is the case when $A=S$.

In every iteration, we obtain a word w_i of length at most $n-|A|$. If we get $A \not\subseteq S \cdot w_1 \dots w_i$ in some i-th iteration, then we are done as the word $w_1 \dots w_i$ has length at most $(|S|-|A|)(n-|A|)$.

If we get $A = S \cdot w_1 \dots w_i$ for some $i \in \{0,\dots,|S|-|A|\}$, then observe that there must exist a letter $a \in \Sigma$ such that $A \cdot a \neq A$, because A contains an avoidable state from $S \supseteq A$. Note that since $|S \cdot w_1 \dots w_i| < |S \cdot w_1 \dots w_{i-1}|$ for every $i = 1,\dots,k$, after the $(|S|-|A|)$-th iteration we must have $|S \cdot w_1 \dots w_k| \leq |S| - (|S|-|A|) = |A|$, we must get this case after the last iteration. It follows that in any case we obtain the word $w_1 \dots w_i a$ of length at most $(|S|-|A|)(n-|A|)+1$. ◀

We state a quadratic upper bound on the length of the shortest avoiding words:

▶ **Corollary 5.** *For $n \geq 2$, in an n-state automaton $\mathscr{A}(Q,\Sigma,\delta)$, for every non-empty proper subset $A \subset Q$ containing an avoidable state, there exists a word avoiding a state from A of length at most*

$$(n-1-|A|)(n-|A|)+2.$$

Proof. Since there exists an avoidable state in A, there is a letter $a \in \Sigma$ such that $|Q \cdot a| < n$.

If $A \not\subseteq Q \cdot a$ then we are done with a word of length 1. Otherwise $A \subseteq Q \cdot a$, so we use Lemma 4 with subset A and subset $S = Q \cdot a$. Since there exists a word avoiding a state from A, the lemma yields a word w of length at most $(|S|-|A|)(n-|A|)+1 \leq (n-1-|A|)(n-|A|)+1$. Thus, aw avoids a state from A and has length at most $(n-1-|A|)(n-|A|)+2$. ◀

In particular, we obtain the upper bound $(n-2)(n-1)+2$ on the length of the shortest avoiding words for any state ($|A|=1$).

▶ **Theorem 6.** *The words from Lemma 2, Lemma 3, Lemma 4, and Corollary 5 can be found in polynomial time.*

Proof. We use the reduction procedure from [4], which in polynomial time replaces each set $\Sigma^{\leq i}$ in the proof of Lemma 2 with a set W_i containing at most $i+1$ words such that L_i has the same dimension.

The procedure starts for $i=0$ with $\{\varepsilon\}$ (the set with the empty word) and inductively constructs a set W_i assuming we have found W_{i-1}. This is done by considering all words wa for $w \in W_{i-1}$ and $a \in \Sigma$ and setting $W_i = W_{i-1} \cup \{wa\}$ for which the dimension of the corresponding subspace grows. There always exists such a word wa, which is argued by the ascending chain condition.

Then, the set W_m is used to span the first linear subspace with the maximal dimension (L_m), so we can find a word satisfying Case (1) or Case (2) of Lemma 2 in W_m. It is obvious that the corresponding words from the other proofs are constructible in polynomial time. ◀

3 Improved bound on reset threshold

In this section, we consider a synchronizing n-state automaton $\mathscr{A}(Q,\Sigma,\delta)$. Obviously, in such an automaton, every state is avoidable unless there is a sink state (a state q such that $q \cdot a = q$ for all $a \in \Sigma$), which cannot be avoided. For synchronizing automata with a sink state the tight upper bound is $n(n-1)/2$ (see, e.g., [22]). Thus we can assume that $\mathscr{A}$ does not have a sink state, and so Lemma 2 and Lemma 3 can be applied for every non-empty subset A.

▶ **Lemma 7.** *Let $w \in \Sigma^*$ and let $g = \min\{|q \cdot w^{-1}| \mid q \in Q \cdot w\}$. There are at least $(g+1)|Q \cdot w| - n$ states $q \in Q \cdot w$ such that $|q \cdot w^{-1}| = g$.*

Proof. Let d be the number of states $q \in Q \cdot w$ whose preimages under w^{-1} have size equal to g. So $|Q \cdot w| - d$ states have the preimages of size at least $g+1$. Note that $(Q \cdot w) \cdot w^{-1} = Q$, and that the sets $q \cdot w^{-1}$ and $p \cdot w^{-1}$ are disjoint for all pairs of states $q \neq p$. So $Q \cdot w^{-1}$ has cardinality at least $dg + (g+1)(|Q \cdot w| - d) = (g+1)|Q \cdot w| - d$. Since this cannot be larger than $n = |Q|$, we get $d \geq (g+1)|Q \cdot w| - n$. ◀

From Lemma 7, in particular, we get that there are at least $2|Q \cdot w| - n$ states in the image $Q \cdot w$ with a unique state in the preimage.

The following lemma is based on [27, Lemma 4], but with a more general bound:

▶ **Lemma 8.** *Let $w \in \Sigma^*$ be a word of rank $r \geq \lfloor (n+1)/2 \rfloor$. Suppose that for some integer $k \geq 1$, for every $A \subset Q$ of size $1 \leq |A| \leq n-1$, there is a word $v_A \in \Sigma^{\leq k(n-|A|)}$ such that $A \not\subseteq Q \cdot v_A$. Then there is a word of rank at most $n/2$ and length at most*

$$|w| + k\frac{n^2 - (2n-2r-1)^2}{4}.$$

Proof. For $i = r, r-1, \ldots, \lfloor n/2 \rfloor$, we inductively construct words w_i of length $\leq |w| + k(r-i)(2n-r-i-1)$ of rank at most i. First, let $w_r = w$.

Let $i < r$ and suppose that we have already found w_{i+1}. If already $|Q \cdot w_{i+1}| \leq i$ then we just set $w_i = w_{i+1}$. Otherwise, we have $|Q \cdot w_{i+1}| = i+1$.

Because $i+1 \geq (n+1)/2$, there exists a non-empty subset of $Q \cdot w_{i+1}$ of states with a unique state in the unique preimage. By Lemma 7, we let $X \subseteq Q \cdot w_{i+1}$ to be a subset of size $2|Q \cdot w_{i+1}| - n = 2i + 2 - n$ of states $q \in Q \cdot w_{i+1}$ such that $|q \cdot w_{i+1}^{-1}| = 1$. We set $w_i = v_X w_{i+1}$, where v_X is the avoiding word from the assumption of the lemma for set X. We have $p \notin Q \cdot v_X$ for some $p \in X$.

State p is the only state mapped by the transformation of w_{i+1} to some state $q = p \cdot w_{i+1}$, i.e. there is no other state p' such that $p' \cdot w_{i+1} = q$. Hence we know that $q \notin Q \cdot w_i = Q \cdot v_X w_{i+1}$. Since $Q \cdot w_i \subseteq Q \cdot w_{i+1}$, $q \notin Q \cdot w_i$ but $q \in Q \cdot w_{i+1}$, we have $Q \cdot w_i \subsetneq Q \cdot w_{i+1}$. Therefore, we have rank

$$|Q \cdot w_i| \leq |Q \cdot w_{i+1}| - 1 \leq i + 1 - 1 = i,$$

and length

$$\begin{aligned} |w_i| &\leq k(n - |A|) + |w_{i+1}| \\ &\leq 2k(n-i-1) + k(r-(i+1))(2n-r-(i+1)-1) + |w| \\ &= k(r-i)(2n-r-i-1) + |w|. \end{aligned}$$

Finally, for $i = \lfloor n/2 \rfloor$ we obtain:

$$\begin{aligned} &\quad |w| + k(r - \lfloor n/2 \rfloor)(2n - r - \lfloor n/2 \rfloor - 1) \\ &\leq |w| + k(r-(n-1)/2)(2n-r-(n-1)/2-1) \\ &= |w| + k(n^2 - (2n-2r-1)^2)/4. \end{aligned}$$

◀

Note that Lemma 4 also provides an upper bound on the length of the shortest avoiding words, but it is larger than that the corresponding bound from Theorem 1, and so would not yield an improvement when used as in Lemma 8. Therefore, we use there an assumption about the length of the shortest avoiding words.

We observe that it is profitable to use Theorem 1 to find the starting word w, as long as $C(i+1,i)$ is smaller than $k(n-|A|)$. An approximate solution is to find the starting word w of rank at most $n-4k$. The following lemma utilizes this idea.

▶ **Lemma 9.** *Suppose that for some integer k, $1 \leq k \leq n/8$, for every $A \subset Q$ of size $1 \leq |A| \leq n-1$, there is a word $v_A \in \Sigma^{\leq k(n-|A|)}$ such that $A \not\subseteq Q \cdot v_A$. Then there is a word of rank at most $n/2$ and length at most*

$$k\frac{3n^2-64k^2+144k+13}{12}.$$

Proof. From Theorem 1, let w be a word of rank at most $n-4k$ and length at most

$$C(n,n-4k) = 4k(8k^2+6k+1)/3.$$

If w has rank $\geq \lfloor (n+1)/2 \rfloor$, then we apply Lemma 8 and obtain a word of rank at most $n/2$ and length at most

$$\begin{aligned}&\frac{4k(8k^2+6k+1)}{3}+\frac{k(n^2-(2n-2(n-4k)-1)^2)}{4}\\ =&\frac{k(3n^2-64k^2+144k+13)}{12}.\end{aligned}$$

Otherwise, w has rank $< n/2$, and because

$$k(n^2-(2n-2(n-4k)-1)^2)/4 = k(n^2-(8k-1)^2)/4$$

is positive for $1 \leq k \leq n/8$ (and $n \geq 8$), the upper bound is also valid. Thus, w has the desired length. ◀

We prove a parametrized upper bound on the reset threshold, depending on whether the assumption in Lemma 9 holds. When the assumption holds, the lemma provides an upper bound using avoiding words; otherwise, we have a quadratic word of a particular rank that yields an improvement.

▶ **Lemma 10.** *For every integer $1 \leq k \leq n/8$, there exists a reset word of length at most*

$$\max\left\{k\frac{3n^2-64k^2+144k+13}{12},\ k(n-1)+C(n-k,\lfloor n/2\rfloor)\right\}+C(\lfloor n/2\rfloor,1).$$

Proof. We use Lemma 3 with the given k and subset $S=Q$.

Suppose that Case (1) from Lemma 3 holds for every $A \subset Q$ with $1 \leq |A| \leq n-1$. Then by Lemma 9 we obtain a word w of rank $\leq n/2$ and length $\leq k(3n^2-64k^2+144k+13)/12$.

Suppose that Case (2) from Lemma 3 holds for some $A \subset Q$ with $1 \leq |A| \leq n-1$. Then we have a word w of rank $\leq n-k$ and length $\leq k(n-1)$. By Theorem 1, we construct a word compressing $Q \cdot w$ to a subset of size $\leq n/2$. Then $k(n-1)+C(n-k,\lfloor n/2\rfloor)$ is an upper bound for the length of the found word of rank $\leq n/2$.

Finally, we need to take the maximum from both cases, and add $C(\lfloor n/2\rfloor,1)$ to bound the length of a word compressing a subset of size $\lfloor n/2\rfloor$ to a singleton. ◀

Now, by finding a suitable k, we state the new general upper bound on the reset threshold:

▶ **Theorem 11.**

$$\mathrm{rt}(\mathscr{A}) \leq (85059n^3+90024n^2+196504n-10648)/511104.$$

Proof. We use Lemma 10 with a suitable k that minimizes the maximum for large enough n.

First, we bound $C(n-k, \lfloor n/2 \rfloor)$ in the second argument in the maximum. If n is even then

$$\begin{aligned} C(n-k, \lfloor n/2 \rfloor) &= C(n-k, n/2) \\ &= \sum_{s=n/2+1}^{n-k} \frac{(n-s+2)(n-s+1)}{2} \\ &= \frac{n^3+6n^2+8n-8k^3-24k^2-16k}{48}. \end{aligned}$$

If n is odd then

$$\begin{aligned} C(n-k, \lfloor n/2 \rfloor) &= C(n-k, (n-1)/2) \\ &= \sum_{s=(n-1)/2+1}^{n-k} \frac{(n-s+2)(n-s+1)}{2} \\ &= \frac{n^3+9n^2+23n-8k^3-24k^2-16k+15}{48}, \end{aligned}$$

which is larger than the previous one.

Now we discuss our choice of k; any value of k gives a bound but we try to get it minimal. Assume that n is large enough. Note that for the largest possible value $k = n/8$ the first function in the maximum from Lemma 10 yields the coefficient of n^3 equal to $1/48$ (the same as by $C(n, \lfloor n/2 \rfloor)$), hence does not give an improvement. For a similar reason, we reject small values $k \in o(n)$. Within linear values k of n, the first function decreases and the second function increases with k. Since they are continuous, it is enough to consider the values of k such that both functions are equal. The approximate solution is $k \simeq 0.11375462n$. For simplicity of the calculations and the final formula, we use the approximation $k = \lfloor 5/44n \rfloor$. Note that any value of k within the valid range will lead to a correct bound, and we use $5/44$ since it is the best approximation by a rational number using integers with at most two digits.

We assume $n \geq 9$; for the smaller values of n the bound is a valid upper bound since it gives larger values than the bound from Theorem 1.

In the following calculations, we use the fact that $5/44n - 1 < \lfloor 5/44n \rfloor$ and $5/44n - 1$ is non-negative. By substitution, for the first function in the maximum we have

$$\begin{aligned} & k\frac{3n^2-64k^2+144k+13}{12} \\ &< (5/44n)\frac{3n^2-64(5/44n-1)^2+144(5/44n)+13}{12} \\ &= (5n(263n^2+3740n-6171))/63888, \end{aligned} \tag{1}$$

and for the second function we have

$$\begin{aligned} & k(n-1) + \frac{n^3+9n^2+23n-8k^3-24k^2-16k+15}{48} \\ &< (5/44n)(n-1) + \big(n^3+9n^2+23n-8(5/44n-1)^3 \\ &\quad -24(5/44n-1)^2-16(5/44n-1)+15\big)/48 \\ &= (10523n^3+153912n^2+196504n+159720)/511104. \end{aligned} \tag{2}$$

Note that (2) is larger than (1) for all n.

Now we have to bound $C(\lfloor n/2 \rfloor, 1)$. If n is even then

$$C(\lfloor n/2 \rfloor, 1) = C(n/2, 1) = (7n^3 - 6n^2 - 16)/48.$$

If n is odd then

$$C(\lfloor n/2 \rfloor, 1) = C((n-1)/2, 1) = (7n^3 - 9n^2 - 31n - 15)/48,$$

which is smaller than the previous one for $n \geq 2$.

Finally, we obtain

$$\frac{10523n^3 + 152262n^2 + 189244n + 191664}{511104} + \frac{7n^3 - 6n^2 - 16}{48}$$
$$= \frac{85059n^3 + 90024n^2 + 196504n - 10648}{511104}.$$ ◀

The theorem improves the old well known bound $(n^3 - n)/6 - 1$ by the factor 85059/85184, or by the coefficient 125/511104 of n^3. This is slightly better than the simpler formula $114n^3/685 + O(n^2)$.

The bound does not necessarily apply for the words obtained by a greedy compression algorithm for synchronization ([1, 11]), because the words in the proof of Lemma 8 are constructed by appending avoiding words at the beginning. However, we can show that there exists a polynomial algorithm finding words of lengths within the bound.

▶ **Proposition 12.** *A reset word of length within the bound from Theorem 11 can be computed in polynomial time.*

Proof. We use k from the proof of Theorem 11. We follow the construction from the proof of Lemma 8. By Theorem 6, we can compute a word from Lemma 2 for a subset A. If (1) holds every time, then we use the obtained word from Lemma 8. Otherwise, we use the word from Lemma 2 for which (2) holds. Finally, the words of lengths at most $C(j, i)$ are computed using a greedy compression algorithm ([1]). ◀

4 Further remarks

Although the improvement in terms of the cubic coefficient is small, it breaks longstanding persistence of the old bound from [21], and possibly opens the area for further progress.

Tiny improvements of the bound from Theorem 11 are possible with more effort yielding better calculations, for example by tuning the value of k in Theorem 11, better rounding, using better bounds at the beginning (note that one can find a shorter word than the word of rank k when Case (2) holds in Lemma 3 by combining with Theorem 1). These however do not add new ideas.

Avoiding a state

The first natural possibility for improving the bound is to show a better bound on the length of the shortest avoiding words. For strongly connected synchronizing automata, currently the best known lower bound is $2n - 3$ by Vojtěch Vorel[1] (binary series), whereas $2n - 2$ is conjectured to be a tight upper bound based on experiments [16].

▶ **Open Problem 1.** *Is $2n - 2$ the tight upper bound on the length of the shortest avoiding words?*

[1] personal communication, unpublished, 2016

Avoiding a subset

The technique from Lemma 8 can be applied only for compressing Q to a subset of size at most $n/2$, because at this point there can be no states with a unique state in the preimage. To bypass this obstacle, we can generalize the concept of avoiding to subsets, and say that a word w *avoids* a subset $D \subseteq Q$ if $D \cap (Q \cdot w) = \emptyset$. Having a good upper bound on the length of the shortest words avoiding D, we could continue using avoiding words for subsets smaller than $n/2$, since for a word s there are at least $|D| \cdot |Q \cdot s| - n$ states such that $1 \le |q \cdot s^{-1}| \le |D|$ (see Lemma 7).

▶ **Open Problem 2.** *Find a good upper bound (in terms of $|D|$ and n) on the length ℓ such that in every n-state automaton, for every subset $D \subset Q$ there is a word avoiding D of length at most ℓ, unless D is not avoidable.*

In fact, we can prove an upper bound in the spirit of Lemma 2, provided that we have avoiding words for smaller subsets than D.

▶ **Lemma 13.** *For $n \ge 2$, let $\mathscr{A}(Q, \Sigma, \delta)$ be an n-state strongly connected synchronizing automaton. Consider non-empty subsets $S, D \subseteq Q$ such that $|S| \ge 1$ and $|D| \ge 2$. Suppose that there is a state $p \in D$ such that for $D' = D \setminus \{p\}$ there exists a word $w_{D'} \in \Sigma^{\ell}$ that avoids D'. Then there exists a word $w \in \Sigma^{n-1+\ell}$ such that either:*

1. *$(S \cdot w) \cap D = \emptyset$, or*
2. *$|S \cdot w| < |S|$.*

Proof. Let $L_i = \mathrm{span}(\{[S][w] \mid w \in \Sigma^{\le i}\})$. We consider the following sequence of linear subspaces:

$$L_0 \subseteq L_1 \subseteq L_2 \subseteq \dots,$$

and use the ascending chain condition as in the proof of Lemma 2. Since the automaton is synchronizing, there is a reset word u so $[S][u] = n[q]$ for some state q. Since the automaton is strongly connected, for every state p we have a word v such that $q \cdot v = p$, and so $[S][uv] = n[p]$. These vectors generate the whole space $\mathbb{R}^n$, and so the maximal dimension of the linear subspaces from the sequence is n; in particular, $\dim(L_{n-1}) = n$.

Let $P = p \cdot (w_{D'})^{-1}$. Suppose for a contradiction that for every word w of length $\le n-1$, subset S is not compressed by w and $|(S \cdot w) \cap P| = 1$. Then $[S][w]$ contains exactly one 1 and $|P| - 1$ 0s at the positions corresponding to the states from P. Therefore, all vectors v generated by the vectors with this property satisfy:

$$(|S| - 1) \sum_{i \in P} v(i) = \sum_{i \in Q \setminus P} v(i).$$

This means that the dimension of L_{n-1} is at most $n-1$, since in $\mathbb{R}^n$ there are vectors that broke this equality. Hence, we have a contradiction.

Hence, there must be a word w that either compresses S or is such that $|(S \cdot w) \cap P| \neq 1$. In the latter case, if $(S \cdot w) \cap P = \emptyset$ then we obtain $(S \cdot w w_{D'}) \cap D = \emptyset$. If $(S \cdot w) \cap P \ge 2$ then $w_{D'}$ maps at least two states from $(S \cdot w) \cap P$ to p, thus $w w_{D'}$ compresses S. ◀

By an iterative application of the above lemma, we can obtain the upper bound $k(n-1+kn)$ on the length of a word that either avoids two states from the given subset or compresses the subset. This bound is too large to provide a further improvement (at least within the cubic coefficient) for the upper bound on the length of the shortest reset words. However, if the shortest words avoiding a single state are indeed of linear length, then we obtain a quadratic upper bound on the length of the shortest words avoiding two states.

References

1 D. S. Ananichev and V. V. Gusev. Approximation of Reset Thresholds with Greedy Algorithms. *Fundamenta Informaticae*, 145(3):221–227, 2016.

2 D. S. Ananichev and M. V. Volkov. Synchronizing generalized monotonic automata. *Theoretical Computer Science*, 330(1):3–13, 2005.

3 M.-P. Béal, M. V. Berlinkov, and D. Perrin. A quadratic upper bound on the size of a synchronizing word in one-cluster automata. *International Journal of Foundations of Computer Science*, 22(2):277–288, 2011.

4 M. Berlinkov and M. Szykuła. Algebraic synchronization criterion and computing reset words. *Information Sciences*, 369:718–730, 2016.

5 M. V. Berlinkov. Synchronizing Quasi-Eulerian and Quasi-one-cluster Automata. *International Journal of Foundations of Computer Science*, 24(6):729–745, 2013.

6 M. T. Biskup and W. Plandowski. Shortest synchronizing strings for Huffman codes. *Theoretical Computer Science*, 410(38-40):3925–3941, 2009.

7 J. Černý. Poznámka k homogénnym eksperimentom s konečnými automatami. *Matematicko-fyzikálny Časopis Slovenskej Akadémie Vied*, 14(3):208–216, 1964. In Slovak.

8 J. Černý, A. Pirická, and B. Rosenauerová. On directable automata. *Kybernetica*, 7:289–298, 1971.

9 H. Don. The Černý Conjecture and 1-Contracting Automata. *Electronic Journal of Combinatorics*, 23(3):P3.12, 2016.

10 L. Dubuc. Sur les automates circulaires et la conjecture de Černý. *Informatique théorique et applications*, 32:21–34, 1998. In French.

11 D. Eppstein. Reset sequences for monotonic automata. *SIAM Journal on Computing*, 19:500–510, 1990.

12 P. Frankl. An extremal problem for two families of sets. *European Journal of Combinatorics*, 3:125–127, 1982.

13 F. Gonze, R. M Jungers, and A. N. Trahtman. A Note on a Recent Attempt to Improve the Pin-Frankl Bound. *Discrete Mathematics and Theoretical Computer Science*, 17(1):307–308, 2015.

14 M. Grech and A. Kisielewicz. The Černý conjecture for automata respecting intervals of a directed graph. *Discrete Mathematics and Theoretical Computer Science*, 15(3):61–72, 2013.

15 J. Kari. Synchronizing finite automata on Eulerian digraphs. *Theoretical Computer Science*, 295(1-3):223–232, 2003.

16 A. Kisielewicz, J. Kowalski, and M. Szykuła. Experiments with Synchronizing Automata. In *Implementation and Application of Automata*, volume 9705 of *LNCS*, pages 176–188. Springer, 2016.

17 A. A. Klyachko, I. K. Rystsov, and M. A. Spivak. An extremal combinatorial problem associated with the bound on the length of a synchronizing word in an automaton. *Cybernetics*, 23(2):165–171, 1987.

18 P. V. Martugin. A series of slowly synchronizing automata with a zero state over a small alphabet. *Information and Computation*, 206(9-10):1197–1203, 2008.

19 J.-E. Pin. Sur les mots synchronisants dans un automate fini. *Elektron. Informationsverarb. Kybernet.*, 14:293–303, 1978.

20 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.

21 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.

22 I. K. Rystsov. Reset words for commutative and solvable automata. *Theoretical Computer Science*, 172(1-2):273–279, 1997.

23 P. H. Starke. Eine Bemerkung über homogene Experimente. *Elektronishe Informationverarbeitung und Kybernetic*, 2:257–259, 1966. In German.

24 B. Steinberg. The averaging trick and the Černý conjecture. *International Journal of Foundations of Computer Science*, 22(7):1697–1706, 2011.

25 B. Steinberg. The Černý conjecture for one-cluster automata with prime length cycle. *Theoretical Computer Science*, 412(39):5487–5491, 2011.

26 A. N. Trahtman. The Černý conjecture for aperiodic automata. *Discrete Mathematics and Theoretical Computer Science*, 9(2):3–10, 2007.

27 A. N. Trahtman. Modifying the upper bound on the length of minimal synchronizing word. In *Fundamentals of Computation Theory*, volume 6914 of *LNCS*, pages 173–180. Springer, 2011.

28 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.

29 M. V. Volkov. Synchronizing automata preserving a chain of partial orders. *Theoretical Computer Science*, 410(37):3513–3519, 2009.