Lower Bounds on Avoiding Thresholds

Robert Ferens
Institute of Computer Science, University of Wrocław, Wrocław, Poland

Marek Szykuła
Institute of Computer Science, University of Wrocław, Wrocław, Poland

Vojtěch Vorel
Faculty of Informatics and Management, University of Hradec Králové, Czech Republic

Abstract

For a DFA, a word *avoids* a subset of states, if after reading that word the automaton cannot be in any state from the subset regardless of its initial state. A subset that admits an avoiding word is *avoidable*. The *k-avoiding threshold* of a DFA is the smallest number such that every avoidable subset of size k can be avoided with a word no longer than that number. We study the problem of determining the maximum possible k-avoiding thresholds. For every fixed $k \geq 1$, we show a general construction of strongly connected DFAs with n states and the k-avoiding threshold in $\Theta(n^k)$. This meets the known upper bound for $k \geq 3$. For $k = 1$ and $k = 2$, the known upper bounds are respectively in $\mathcal{O}(n^2)$ and in $\mathcal{O}(n^3)$. For $k = 1$, we show that $2n - 3$ is attainable for every number of states n in the class of strongly connected synchronizing binary DFAs, which is supposed to be the best possible in the class of all DFAs for $n \geq 8$. For $k = 2$, we show that the conjectured solution for $k = 1$ (an upper bound in $\mathcal{O}(n)$) also implies a tight upper bound in $\mathcal{O}(n^2)$ on 2-avoiding threshold. Finally, we discuss the possibility of using k-avoiding thresholds of synchronizing automata to improve upper bounds on the length of the shortest reset words.

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

Keywords and phrases avoiding word, Černý conjecture, rank conjecture, reset threshold, reset word, synchronizing automaton, synchronizing word

Digital Object Identifier 10.4230/LIPIcs.MFCS.2021.46

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.

1 Introduction

A *deterministic finite complete semi-automaton* (called simply *automaton*) is a 3-tuple (Q, Σ, δ), where Q is a finite set of *states*, Σ is an *input alphabet*, and $\delta\colon Q \times \Sigma \to Q$ is the *transition function*. The transition function is naturally extended to a function $Q \times \Sigma^* \to Q$. Throughout the paper, by n we always denote the number of states in Q.

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

The *rank* of a word w is the cardinality of the image $\delta(Q, w)$. A word w is *reset* if it has rank 1, i.e., its action maps all the states to one state. If an automaton admits a reset word, then it is called *synchronizing*, and its *reset threshold* is the length of the shortest reset word.

The central problem in the theory of synchronizing automata is the famous Černý conjecture, which states that every synchronizing n-state automaton has reset threshold at most $(n-1)^2$. Fig. 1 shows the 4-state automaton from the well-known Černý series [2], which

46th International Symposium on Mathematical Foundations of Computer Science (MFCS 2021).
Editors: Filippo Bonchi and Simon J. Puglisi; Article No. 46; pp. 46:1–46:14
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

meets the conjectured upper bound for every n. The Černý conjecture is one of the most longstanding open problems in automata theory, with constantly growing literature around the topic; see an old survey [15] and a recent special issue dedicated to the problem [17]. The best known general upper bound on reset threshold is cubic and equals $\sim 0.1654n^3 + o(n^3)$ [11]. Better bounds are known for many subclasses of automata.

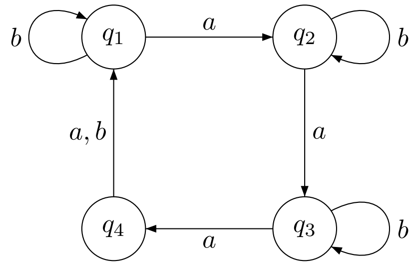

Figure 1 The Černý automaton with $n = 4$ states.

1.1 Avoiding words

Avoiding words are defined similarly to reset words. For a state $q \in Q$, a word w is *avoiding* if $q \notin \delta(Q, w)$, i.e., no state is mapped by the action of w to q. More generally, a word w *avoids* a subset $S \subset Q$ if $\delta(Q, w) \cap S = \emptyset$. Note that a word of rank $n - k$ is also a word avoiding some subset S of size k. In this way, a reset word is a specific case of an avoiding word.

A subset S is *avoidable* if there exists an avoiding word for S. Then, the *S-avoiding threshold* is the length of the shortest words avoiding S. The *k-avoiding threshold* is the maximum S-avoiding threshold over all subsets $S \subset Q$ of size k. In other words, every avoidable subset of size k can be avoided with a word of length not exceeding the k-avoiding threshold.

Obviously, a k-avoiding threshold is never larger than the $(k + 1)$-avoiding threshold. In a synchronizing automaton, every subset of size $\leq n - 1$ is avoidable.

For example, for the automaton from Fig. 1, the k-avoiding thresholds for $k = 1, 2, 3$ are respectively equal to 4, 8, 12. For instance, the shortest word avoiding subset $\{q_2, q_3\}$ is ba^3ba^3, and no other subset of size two requires a longer word.

Avoiding words are closely related to reset words and can be interesting as such for similar reasons. Yet, so far the focus was put on their application to bounding reset thresholds.

Originally, the concept was first used by Trahtman as a tool for improving the cubic upper bound on the reset threshold [14]. This turned out to be wrong as is based on the claim that 1-avoiding threshold is at most n, whereas it can be larger [5]. Nevertheless, the idea of applying avoiding words has been shown to be useful. A non-trivial quadratic upper bound on 1-avoiding threshold already led to the first improvement [12] of the old and well-known upper bound on the reset threshold [8]. This was later refined to $\sim 0.1654n^3 + o(n^3)$ [11], using the same method but improving the counting argument.

Better upper bounds on avoiding thresholds should lead to better upper bounds on reset thresholds. Yet, the problem of avoiding may be of similar difficulty.

Avoiding a subset is closely related to subset reachability. The latter is the question for a given subset $T \subseteq Q$, are there and how long are words w such that $\delta(Q, w) = T$. It is known that the decision problem is PSPACE-complete even if the automaton is strongly connected [18] and the shortest such words can have a length larger than $2^n/n$ [4]. It holds similarly in the weaker *included* version where w is such that $\delta(Q, w) \subseteq T$ [4]. This is equivalent to avoiding the complement of T: $\delta(Q, w) \subseteq T$ is equivalent to $\delta(Q, w) \cap (Q \setminus T) = \emptyset$. It is also equivalent to that w is *totally extending*, i.e., $\delta^{-1}(T, w) = Q$, which was shown to be PSPACE-complete even if the automaton is strongly connected and binary [1]. It follows that avoiding a subset, in general, should require exponentially long words, but precise bounds were not shown so far. In particular, these results do not say what happens in the case of a synchronizing automaton nor if the size of the subset S to avoid is small. Yet, as we note, these cases are particularly important.

1.2 Known upper bounds

For the 1-avoiding threshold, the best known upper bound is quadratic in n. It is derived through linear algebraic methods applied to automata.

▶ **Theorem 1** (rephrased [12, Corollary 5]). *For an n-state automaton, the 1-avoiding threshold is at most $(n-2)(n-1)+2$.*

In the general case, for k-avoiding threshold, we have the following asymptotic bound:

▶ **Theorem 2** (rephrased [1, Theorem 12]). *Let $\mathscr{A} = (Q, \Sigma, \delta)$ be an n-state automaton, let r be the minimal rank in $\mathscr{A}$ over all words, and let m be the length of the shortest words of the minimal rank. Then the k-avoiding threshold is at most $\mathcal{O}(n^{\min(r,k)} + m)$.*

Since for m we have only a cubic upper bound $\mathcal{O}(n^3)$, this component is dominating for $k = 1$ and $k = 2$. In these cases, it is unlikely to be tight. There is also the well-known rank (Pin-Černý) conjecture [8] stating that $m \le (n-r)^2$.

No better bounds on avoiding thresholds are known in the case of a synchronizing and/or strongly connected automaton. Except for the obvious fact that for a synchronizing automaton $\mathcal{O}(n^3)$ is an upper bound on every k-avoiding threshold, and $\mathcal{O}(n^2)$ would be an upper bound if the (weak version of) Černý conjecture holds.

We note that it is also an easy exercise to prove that avoiding thresholds are small (at most linear) for many subclasses of automata. For example, for Eulerian automata, the k-avoiding threshold is at most $k(n-1)$ [6]. For aperiodic automata [13, 16], or more generally, for automata with letters whose transitions contain only trivial cycles (self-loops), it is at most $n-1$.

1.3 Known lower bounds

Concerning automata with the largest possible avoiding thresholds, only a few particular examples of automata were described so far. Moreover, they were limited to 1-avoiding threshold.

The first such example is a 4-state automaton with the 1-avoiding threshold equal to 6, which was found as a counterexample to the conjecture that the 1-avoiding threshold is bounded above by the number of states [5]. Later experiments [7] revealed several other examples with $n \le 11$ states, suggesting that $2n-2$ may be an upper bound.

It is also known that deciding whether 1-avoiding threshold or $\{q\}$-avoiding threshold (for a given q) is smaller than a given integer is NP-complete [1].

1.4 Contribution

We tackle the problem of constructing automata with large avoiding thresholds. In this paper, we summarize our efforts.

First, we show that 1-avoiding threshold can be equal to $2n-3$, for every number of states $n \geq 2$, and this is met by a series of strongly connected synchronizing automata. We conjecture that it is best possible, except for finitely many examples meeting $2n-2$.

For the general case, we show a series of automata with k-avoiding thresholds in $\Theta(n^k)$. These automata are strongly connected but not synchronizing. This matches the asymptotic upper bound for $k \geq 3$, leaving open the cases $k = 2$ and $k = 1$. Yet, we show that the case of $k = 2$ can be reduced to a possible solution of the case of $k = 1$.

Finally, we note the potential of avoiding words for improving upper bounds on the reset threshold. So far, only an application of 1-avoiding threshold was considered; being in $\mathcal{O}(n)$, it would imply an upper bound on the reset threshold equal to $7/48n^3 + \mathcal{O}(n^2)$ [14]. However, with the generalized concept of avoiding subsets of size k, we can achieve more. If the k-avoiding threshold is subquadratic for all k bounded by any growing function in n (e.g., if for all $k \leq \log n$, the k-avoiding threshold is $o(n^2)$), then the reset threshold is subcubic.

2 1-avoiding threshold

We show a series of binary strongly connected synchronizing automata with 1-avoiding threshold equal to $2n-3$. The existence of the series has been mentioned several times ([7, 12]), but it has not been described yet.

For each $n \geq 5$, we define $\mathscr{A}_n = (Q = \{q_1, \ldots, q_n\}, \Sigma = \{a, b\}, \delta)$, which is shown in Fig. 2. The transition function is defined as follows: $\delta(q_1, a) = q_2$; $\delta(q_1, b) = q_1$; $\delta(q_2, a) = q_1$; $\delta(q_2, b) = q_3$; $\delta(q_3, a) = q_1$; $\delta(q_3, b) = q_4$; $\delta(q_i, a) = \delta(q_i, b) = q_{i+1}$ for all $i \in \{4, \ldots, n-1\}$; $\delta(q_n, a) = q_4$; $\delta(q_n, b) = q_2$.

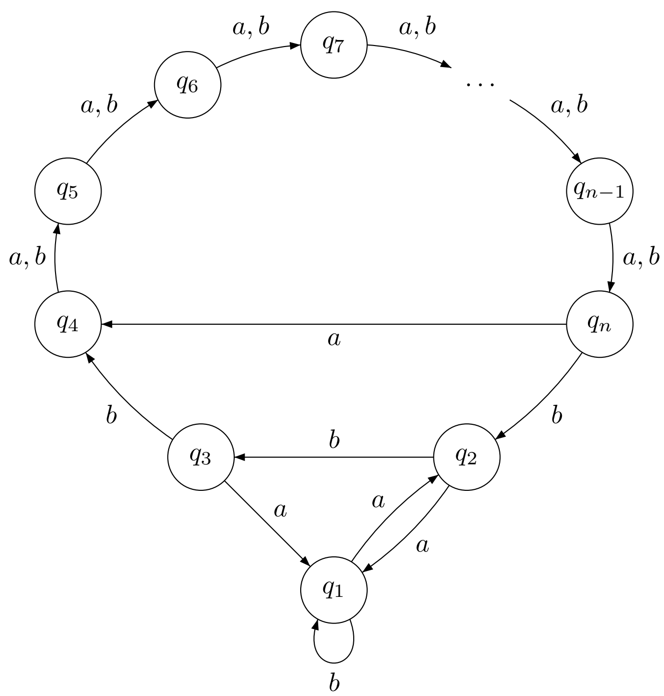

Figure 2 The binary automaton $\mathscr{A}_n$ with the 1-avoiding threshold equal to $2n-3$.

▶ **Proposition 3.** *For $n \geq 5$, the automaton $\mathscr{A}_n$ is strongly connected and synchronizing, and its reset threshold is at most $3n-4$.*

Proof. We can avoid all even-indexed states on the a-cycle by word $(ab)^k$ for some k:

$$Q \xrightarrow{ab} Q \setminus \{q_4\} \xrightarrow{ab} Q \setminus \{q_4, q_6\} \xrightarrow{ab} \cdots \xrightarrow{ab} \{q_1, q_2, q_3\} \cup \{q_5, q_7, \ldots\}.$$

Moreover, if n is even, then we can continue applying word ab to avoid also all odd-indexed states on the a-cycle:

$$\{q_1, q_2, q_3\} \cup \{q_5, q_7, \ldots\} \xrightarrow{ab} \{q_1, q_2, q_3\} \cup \{q_7, q_9, \ldots\} \xrightarrow{ab} \cdots \xrightarrow{ab} \{q_1, q_2, q_3\} \xrightarrow{ab} \{q_1, q_3\}.$$

If n is odd, then we can apply one additional a letter to avoid all odd-number states and repeat the procedure of avoiding even-number states, that is:

$$\{q_1, q_2, q_3\} \cup \{q_5, q_7, \ldots\} \xrightarrow{a} \{q_1, q_2, q_3\} \cup \{q_4, q_6, \ldots\} \xrightarrow{ab} \{q_1, q_2, q_3\} \cup \{q_6, q_8, \ldots\}$$
$$\xrightarrow{ab} \cdots \xrightarrow{ab} \{q_1, q_2, q_3, q_{n-1}\} \xrightarrow{ab} \{q_1, q_2, q_3\}.$$

Overall, we can compress the automaton to the set $\{q_1, q_2, q_3\}$ using a word of length at most $2(n-3)$ for an even and $2(n-3)+1$ for an odd n, since each ab application decreases the size of the image by 1 until there are only three states left. The set $\{q_1, q_2, q_3\}$ can be easily synchronized to q_1 by the word $ab^{n-3}aba$. ◀

Informally, the key property of the construction is the following. To avoid q_1, we must avoid both q_2 and q_3. But to do so, we must first avoid two consecutive states in the cycle $q_4, \ldots, q_n$ on a. This requires one full round on this cycle. In particular, the avoided states on the cycle will be always q_4 and q_5 at some point. Then, we need a second round to map these two gaps to q_2 and q_3, respectively. Hence, the shortest avoiding words for q_1 have length $\sim 2n$.

▶ **Theorem 4.** *For $n \geq 7$, the 1-avoiding threshold of $\mathscr{A}_n$ equals $2n-3$.*

Proof. We show that the length of the shortest avoiding words for state q_1 is $2n-3$.

If n is even, then $(ab)^{n-3}bba$ is an avoiding word for q_1 of length $2n-3$. We have:

$$Q \xrightarrow{(ab)^{n-3}} \{q_1, q_2, q_3\} \xrightarrow{bba} \{q_2, q_4, q_5\}.$$

If $n \geq 9$ is odd, then $aba^{n-5}bab^{n-3}a$ is the desired word. We have:

$$Q \xrightarrow{ab} Q \setminus \{q_4\} \xrightarrow{a^{n-5}} Q \setminus \{q_3, q_{n-1}\} \xrightarrow{ba} Q \setminus \{q_3, q_4, q_5\} \xrightarrow{b^{n-3}} Q \setminus \{q_n, q_2, q_3\} \xrightarrow{a} Q \setminus \{q_1, q_4, q_3\}$$

For $n = 7$, a shortest avoiding word is $abaabaaabba$.

To prove the lower bound, we start with an auxiliary claim.

Claim 1: If a word v avoids state q_n, then $|v| \geq n-2$. It follows because the only state that can be avoided with one letter (a) is q_3, and the shortest words with an action mapping q_3 to q_n have length $n-3$.

Let w be a shortest word avoiding state q_1. Since $\delta(q_1, b) = q_1$, $\delta(q_1, aa) = q_1$, and $\delta(q_1, aba) = q_1$, it follows that w must end with bba, as otherwise it would not be a shortest such word. Let write $w = w'bba$. We have $\delta(q_n, bba) = \delta(q_{n-1}, bba) = q_1$, which implies that w' must avoid both q_n and q_{n-1}. From Claim 1, $|w'| \geq n-2$, so we can write $w = w''xyubba$, where $|u| = n-5$, $|x| = |y| = 1$, and $|w''| \geq 1$. Since $w' = w''xyu$ avoids both q_n and q_{n-1}, and these states are mapped respectively to q_n and q_{n-1} by the action of every word of length $n-5$, we know that $w''xy$ avoids both q_5 and q_4.

We consider three cases:

1. If $y = a$, then $w''x$ must avoid q_n (as $\delta(q_n, a) = q_4$). From Claim 1, we get that $|w''x| \geq n - 2$, hence $|w| \geq (n-2) + 1 + (n-5) + 3 = 2n - 3$.
2. If $y = b$ and $x = a$, then w'' must avoid q_n (as $\delta(q_n, ab) = q_5$). Similarly as in the previous case, we get that $|w| \geq 2n - 2$.
3. If $y = b$ and $x = b$, then w'' must avoid both q_2 and q_3. Then, however, $w''a$ would avoid q_1, which contradicts that w is a shortest such word. ◀

The proof covers the cases $n \geq 7$, whereas the lower bound for the cases $n \leq 6$ was confirmed by experiments [7].

Automata $\mathscr{A}_n$ have another extremal property: The quadratic upper bound on the 1-avoiding threshold is derived by an iterative application of [12, Lemma 3], which for a given subset S gives a word w of length at most $n - |S| + 1$ that either avoids a fixed state $q \in Q$ or compresses S (i.e., $|\delta(S, w)| < |S|$). In the worst case, we must apply this lemma a linear number of times, hence a quadratic upper bound follows. The automata $\mathscr{A}_n$ demonstrate that this may be the case: we may be forced to apply the lemma $\Theta(n)$ times obtaining each time the word $w = ba$, which compresses the subset. On the other hand, ba is very short compared to the linear upper bound $n - |S| + 1$. Yet, it would be possible that also this bound can be met. There also exists another series of automata (similar to the automaton from Fig. 4) showing that the upper bound $n - |S| + 1$ on the length of shortest avoiding or compressing word is tight for each $n \geq |S| \geq 1$. However, it is an open question whether both these bounds can be met simultaneously.

2.1 Exceptional examples

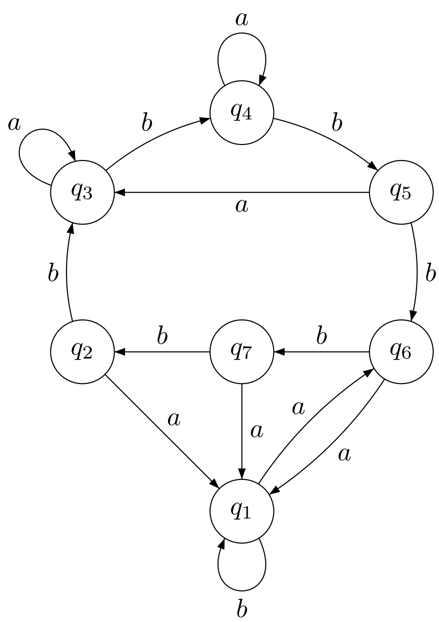

Figure 3 An automaton with the 1-avoiding threshold equal to $2n - 2 = 12$ (for state q_1).

There are several particular examples of synchronizing automata with 1-avoiding threshold equal to $2n - 2$. For instance, for the automaton $\mathscr{A}_5$ it is 8. Another example is shown Fig. 3, which is a largest known automaton meeting this bound.

No such example seems to belong to a series keeping that 1-avoiding threshold. It is known that there are no binary synchronizing automata exceeding 1-avoiding threshold $2n-3$ in range $n \in \{8, 9, 10, 11\}$ [7] (the case of $n = 11$ was verified later).

3 k-avoiding thresholds

3.1 General lower bound

For every fixed $k \geq 1$, we show an infinite series of (non-synchronizing) automata such that their k-avoiding threshold is in $\Theta(n^k)$. The construction is built from gadgets of two types.

3.1.1 One-track counting gadget

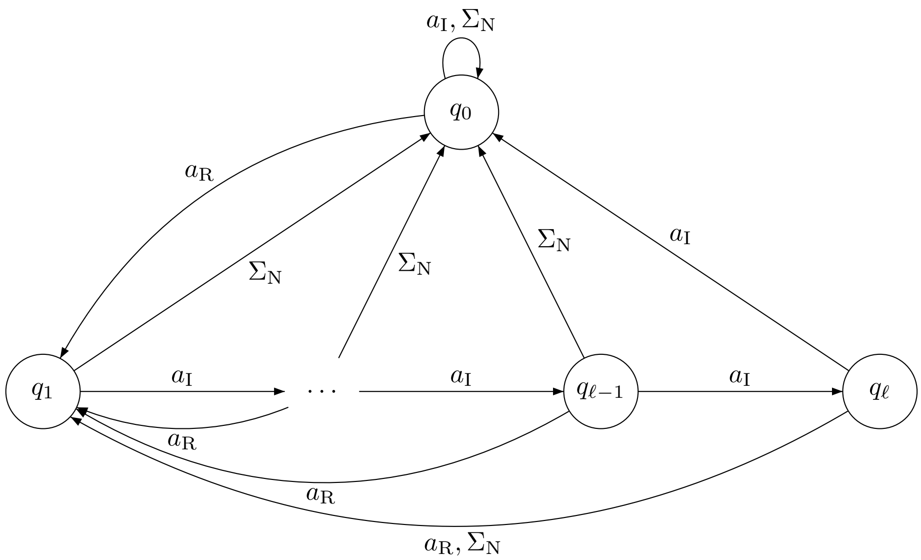

Figure 4 The one-track counting gadget. The identity action of Σ_P is not drawn.

Let $\ell \geq 2$ be an integer, $\Sigma_\mathrm{P}, \Sigma_\mathrm{N}$ be disjoint sets of letters, and $a_\mathrm{R}, a_\mathrm{I} \notin \Sigma_\mathrm{P} \cup \Sigma_\mathrm{N}$ be two other distinct letters. We define the *one-track counting gadget* $\mathscr{T}(\ell, a_\mathrm{R}, \Sigma_\mathrm{P}, a_\mathrm{I}, \Sigma_\mathrm{N})$ (shown in Fig. 4), which is the automaton $(Q_\mathscr{T}, \Sigma_\mathscr{T}, \delta_\mathscr{T})$, where $P = \{q_0, q_1, \ldots, q_\ell\}$, $\Sigma_\mathscr{T} = \{a_\mathrm{R}, a_\mathrm{I}\} \cup \Sigma_\mathrm{P} \cup \Sigma_\mathrm{N}$ and $\delta_\mathscr{T}$ is defined as follows. Letter a_R is the *reset letter* with the action mapping all the states to q_1:

$$\delta_\mathscr{T}(q_i, a_\mathrm{R}) = q_1 \text{ for } i \in \{0, 1, \ldots, \ell\}.$$

Letter a_I is the *incrementing letter* with the action shifting the states $q_1, \ldots, q_k$:

$$\delta_\mathscr{T}(q_i, a_\mathrm{I}) = q_{i+1} \text{ for } i \in \{1, \ldots, \ell - 1\}; \quad \delta_\mathscr{T}(q_\ell, a_\mathrm{I}) = q_0; \quad \delta_\mathscr{T}(q_0, a_\mathrm{I}) = q_0.$$

The letters from Σ_P are called *previous letters* and they all act as the identity. The letters from Σ_N are called *next letters*; they have the same action mapping all the states to q_0 except q_ℓ, which is mapped to q_1:

$$\delta_\mathscr{T}(q_i, a) = q_0 \text{ for } i \in \{0, 1, \ldots, \ell - 1\}, a \in \Sigma_\mathrm{N}; \quad \delta_\mathscr{T}(q_\ell, a) = q_1 \text{ for } a \in \Sigma_\mathrm{N}.$$

The mechanism of the gadget is that, in order to avoid q_0, applications of next letters are restricted. We must start with a_{R} and then keep applying $a_{\mathrm{I}}^{\ell-1}$ alternatingly with one next letter. Additionally, previous letters can be applied interleaving at any time. Formally, we have the following:

▶ **Lemma 5.** *Consider $\mathscr{T}$ and a word $w \in (\{a_{\mathrm{I}}\} \cup \Sigma_{\mathrm{P}} \cup \Sigma_{\mathrm{N}})^*$. If the word $a_{\mathrm{R}}w$ avoids q_0, then in w, the number of occurrences of a_{I} is at least $\ell - 1$ times larger than the total number of occurrences of letters from Σ_{N}.*

Proof. In the analysis, we can ignore the occurrences of Σ_{P} and assume equivalently that $w \in (\{a_{\mathrm{I}}\} \cup \Sigma_{\mathrm{N}})^*$.

Since w does not contain a_{R} and this is the only letter with the action mapping q_0 to another state, for every prefix w' of w, $a_{\mathrm{R}}w'$ must also avoid q_0. In the beginning, $\delta_{\mathscr{T}}(Q_{\mathscr{T}}, a_{\mathrm{R}}) = \{q_1\}$. The only possibility to keep q_0 avoided is to apply $a_{\mathrm{I}}^{\ell-1}$, which must be followed by a letter from Σ_{N}. We end with the singleton $\{q_1\}$ and the argument repeats, keeping the proportion between the numbers of occurrences of a_{I} and of letters from Σ_{N}. ◀

3.1.2 Two-track counting gadget

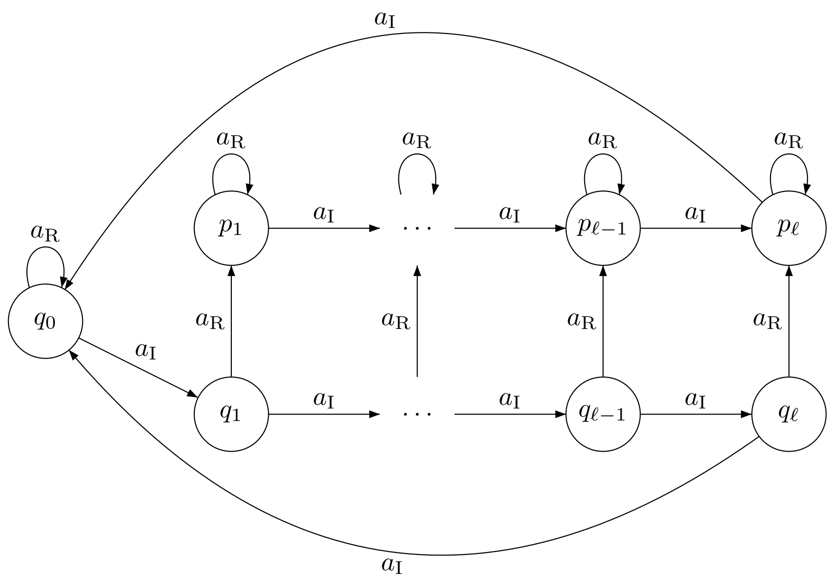

Figure 5 The two-track counting gadget. The identity action of Σ_{P} is not drawn.

Let $\ell \geq 1$ be an integer, Σ_{P} be a set of letters, and $a_{\mathrm{R}}, a_{\mathrm{I}}$ be two other distinct letters. We define the *two-track counting gadget* $\mathscr{D}(\ell, a_{\mathrm{R}}, \Sigma_{\mathrm{P}}, a_{\mathrm{I}})$ (shown in Fig. 5), which is the automaton $(Q_{\mathscr{D}}, \Sigma_{\mathrm{P}} \cup \{a_{\mathrm{R}}, a_{\mathrm{I}}\}, \delta_{\mathscr{D}})$, where $Q_{\mathscr{D}} = \{q_0, q_1, \ldots, q_\ell, p_1, \ldots, p_\ell\}$, and $\delta_{\mathscr{D}}$ is defined as follows. Letter a_{R} is the *reset letter*, whose action maps the corresponding states q_i to p_i:

$$\delta_{\mathscr{D}}(q_i, a_{\mathrm{R}}) = p_i \text{ for } i \in \{1, \ldots, \ell\}; \quad \delta_{\mathscr{D}}(p_i, a_{\mathrm{R}}) = p_i \text{ for } i \in \{1, \ldots, \ell\}; \quad \delta_{\mathscr{D}}(q_0, a_{\mathrm{R}}) = q_0.$$

Letter a_{I} is the *incrementing letter* acting as follows:

$$\delta_{\mathscr{D}}(q_i, a_{\mathrm{I}}) = q_{i+1},\ \delta_{\mathscr{D}}(p_i, a_{\mathrm{I}}) = p_{i+1} \text{ for } i \in \{1, \ldots, \ell - 1\}; \quad \delta_{\mathscr{D}}(p_\ell, a_{\mathrm{I}}) = \delta_{\mathscr{D}}(q_\ell, a_{\mathrm{I}}) = q_0.$$

Finally, the letters from Σ_{P} act as identity.

The point of the gadget is that, in order to avoid p_ℓ, we have to apply $\ell - 1$ times letter a_{I} without applying letter a_{R} in between.

▶ **Lemma 6.** *For $\mathscr{D}$, if a word $w \in \Sigma_{\mathscr{D}}^*$ avoids p_ℓ, then w contains at least ℓ occurrences of a_{I} without any occurrence of a_{R} in between.*

Proof. Observe that for every word $u \in \Sigma_{\mathscr{D}}^*$, we have $\delta_{\mathscr{D}}(Q_{\mathscr{D}}, ua_{\mathrm{R}}) = \{q_0, p_1, \ldots, p_\ell\}$. Thus, if p_ℓ is avoided by w, then w must contain a subword (factor) $w' \in \Sigma_{\mathscr{D}} \setminus \{a_{\mathrm{R}}\}$ such that $p_\ell \notin \delta_{\mathscr{D}}(\{q_0, p_1, \ldots, p_\ell\}, v)$. Ignoring the letters from Σ_{P}, the only such words are a_{I}^i for $i \geq \ell$. ◀

3.1.3 The construction

We build the construction as a union of gadgets. For each $k \geq 1$ and $\ell \geq 2$, we build the automaton $\mathscr{K}(k,\ell) = (Q_{\mathscr{K}(k,\ell)}, \Sigma_{\mathscr{K}(k,\ell)}, \delta_{\mathscr{K}(k,\ell)})$. Let $\Sigma_{\mathscr{K}(k,\ell)} = \{a_{\mathrm{R}}, a_1, \ldots, a_k\}$ be the input alphabet. The automaton is the disjoint union of the following gadgets:

$$\mathscr{D} = \mathscr{D}(\ell, a_{\mathrm{R}}, \{a_k, \ldots, a_2\}, a_1);$$
$$\mathscr{T}_i = \mathscr{T}(\ell, a_{\mathrm{R}}, \{a_k, \ldots, a_{i+1}\}, a_i, \{a_{i-1}, \ldots, a_1\}) \text{ for all } i \in \{2, \ldots, k\}.$$

As it contains states from several gadgets, when denoting a state, we specify the owning gadget in the superscript. Finally, we define the subset to be avoided:

$$S_k = \{p_\ell^{\mathscr{D}}\} \cup \bigcup_{i \in \{1,\ldots,k-1\}} \{q_0^{\mathscr{T}_i}\}.$$

Observe that the number of states n of $\mathscr{K}(k,\ell)$ equals $(k-1)(\ell+1) + (1+2\ell) = k\ell + k + \ell = \ell(k+1) + k$.

Note that for every k, we have $Q_{\mathscr{K}(k,\ell)} \subsetneq Q_{\mathscr{K}(k+1,\ell)}$, $\Sigma_{\mathscr{K}(k,\ell)} \subsetneq \Sigma_{\mathscr{K}(k+1,\ell)}$, $\delta_{\mathscr{K}(k,\ell)} \subsetneq \delta_{\mathscr{K}(k+1,\ell)}$, and also $S_k \subsetneq S_{k+1}$. Hence, every letter of $\mathscr{K}(k,\ell)$ acts the same on the same common states in every $\mathscr{K}(k+i)$ for $i \geq 0$.

For every word $u \in \Sigma^*$, further applying a_{R} yields the same fixed image, that is:

$$\delta_{\mathscr{K}(k,\ell)}(Q_{\mathscr{K}(k,\ell)}, ua_{\mathrm{R}}) = \delta_{\mathscr{K}(k,\ell)}(Q_{\mathscr{K}(k,\ell)}, a_{\mathrm{R}}) = \{q_0^{\mathscr{D}}\} \cup \bigcup_{j \in \{1,\ldots,\ell\}} \{p_j^{\mathscr{D}}\} \cup \bigcup_{i \in \{1,\ldots,k\}} \{q_1^{\mathscr{T}_i}\}. \quad (1)$$

▶ **Lemma 7.** *For $\mathscr{K}(k,\ell)$, the subset S_k is avoidable and the length of the shortest avoiding words for S_k equals $1 + \ell^k$ (and ℓ if $k = 1$).*

Proof. For $k = 1$, the shortest avoiding word for S_1 is a_1^ℓ. For the remaining part, assume that $k \geq 2$.

Let w_k be a shortest avoiding word for S_k in $\mathscr{K}(k,\ell)$. First, we observe that w_k must contain a_{R}, since otherwise the states $q_0^{\mathscr{C}_i}$ could not be avoided. From (1), we know that w_k may contain only one occurrence of a_{R} and it must appear at the beginning; otherwise, there would exist a shorter avoiding word.

To show that S_k is avoidable with a word of length $1+\ell^k$, we use induction on k. We show that there is an avoiding word w_k from $a_{\mathrm{R}}\{a_1, \ldots, a_k\}^*$ of the required length. For $k = 1$, there is only the gadget $\mathscr{D}$, and the word $w_1 = a_{\mathrm{R}} a_1^\ell$ does the job. Assuming the statement for k, we show that it holds for $k+1$. Then w_k acts the same in $\mathscr{K}(k+1,\ell)$ as in $\mathscr{K}(k,\ell)$ on the common states. Let w_{k+1} be the word obtained from w_k by inserting $a_{k+1}^{\ell-1}$ before each occurrence of every letter from $\{a_1, \ldots, a_k\}$. Since a_{k+1} works as the identity in the gadgets $\mathscr{D}, \mathscr{T}_1, \ldots, \mathscr{T}_k$, S_k is avoided. Also, we can see that $\delta_{\mathscr{K}(k+1,\ell)}(\{q_0^{\mathscr{T}_{k+1}}, \ldots, q_\ell^{\mathscr{T}_{k+1}}\}, w_{k+1}) = \{q_1^{\mathscr{T}_{k+1}}\}$. Thus w_{k+1} avoids S_{k+1} and has length $1 + \ell^k \cdot \ell = 1 + \ell^{k+1}$.

To show that there are no shorter avoiding words, we also use induction on k. We show that every avoiding word for S_k contains at least ℓ^k occurrences of letters from $\{a_1, \ldots, a_k\}$. For $k = 1$, by Lemma 6, w has at least ℓ occurrences of a_1. Assuming the statement for k, we show that it holds for $k+1$. Let $w_{k+1} = a_{\mathrm{R}} w'_{k+1}$ be a shortest avoiding word for S_{k+1}. Since $S_k \subset S_{k+1}$, w_{k+1} also avoids S_k. Let w'_k be the word obtained from w'_{k+1} by removing every occurrence of a_{k+1}. Then, $a_{\mathrm{R}} w'_k$ is over the alphabet of $\mathscr{K}(k,\ell)$ and is an avoiding word for S_k in $\mathscr{K}(k,\ell)$. Thus, by the inductive assumption, it has at least ℓ^k occurrences of letters from $\{a_1,\dots,a_k\}$. Then w'_{k+1} contains them as well. By Lemma 4 for the last gadget $\mathscr{T}_{k+1}$, where the set of the next letters is $\{a_1,\dots,a_k\}$, we get that w'_{k+1} also must contain at least $(\ell-1)\cdot\ell^k$ occurrences of a_{k+1}. Altogether, w'_{k+1} contains at least ℓ^{k+1} occurrences of letters from $\{a_1,\dots,a_{k+1}\}$. ◀

3.1.4 Strong connectivity

Each particular gadget is already strongly connected. We can make the whole construction strongly connected by redefining the special action of a_{R} so that its transitions work cyclically on the gadgets. Let

$$\delta_{\mathscr{K}(k,\ell)}(q_0^{\mathscr{D}}, a_{\mathrm{R}}) = q_1^{\mathscr{T}_1};$$
$$\delta_{\mathscr{K}(k,\ell)}(q_j^{\mathscr{T}_i}, a_{\mathrm{R}}) = q_1^{\mathscr{T}_{i+1}} \text{ for } i \in \{1,\dots,k-1\},\ j \in \{0,\dots,\ell\};$$
$$\delta_{\mathscr{K}(k,\ell)}(q_j^{\mathscr{T}_k}, a_{\mathrm{R}}) = q_0^{\mathscr{D}} \text{ for } j \in \{0,\dots,\ell\};$$

and the action is left unchanged for the other states of $\mathscr{D}$: $q_1^{\mathscr{D}},\dots,q_\ell^{\mathscr{D}},p_1^{\mathscr{D}},\dots,p_\ell^{\mathscr{D}}$.

Since (1) still holds for the modified construction, Lemma 7 works as well. We conclude the construction with the following:

▶ **Theorem 8.** *For every $k \ge 2$, there exists an infinite series of strongly connected automata such that its k-avoiding threshold is at least $1 + \left(\frac{n-k}{k+1}\right)^k$. For a fixed k, its k-avoiding threshold is in $\Theta(n^k)$.*

Proof. For an integer $k \ge 2$, the build the automata $\mathscr{K}(k,\ell)$ for all $\ell \ge 2$. Each $\mathscr{K}(k,\ell)$ has $n = \ell(k+1)+k$ states and its k-avoiding threshold is at least $1+\ell^k$, thus we get the lower bound.

Note that a_{R} is the shortest word of the minimal rank, so for a fixed k, the lower bound asymptotically coincides with the upper bound from Theorem 2 for $k \ge 2$, thus the k-avoiding threshold is in $\Theta(n^k)$. ◀

3.2 2-avoiding threshold reduction

The bound $\mathcal{O}(n^k)$ is asymptotically tight for $k \ge 3$, but the cases of $k=1$ and $k=2$ remain open. This is due to the cubic upper bound on the length of the shortest minimal-rank words. However, we can reduce the problem for $k=2$ to the case of $k=1$. If 1-avoiding threshold is bounded by $f(n)$, then 2-avoiding threshold is at most $\mathcal{O}(n^2 + n\cdot f(n))$. Thus, if 1-avoiding threshold is at most linear, then 2-avoiding threshold is at most quadratic, which would be a tight bound.

A similar result, but restricted to strongly connected synchronizing automata, was shown in [12, Lemma 13]. Here, we show it in the general case, which in essence combines both techniques behind Theorem 8 and Theorem 4.

▶ **Theorem 9.** *For every n-state automaton, if the 1-avoiding threshold is at most $f(n)$ for some function f, then the 2-avoiding threshold is at most $\mathcal{O}(n^2 + n\cdot f(n)))$.*

Proof. A *strongly connected bottom component* is a minimal non-empty set of states $X \subseteq Q$ such that for every word w we have $\delta(X, w) \subseteq X$. A synchronizing automaton has exactly one strongly connected bottom component, whereas a non-synchronizing automaton can have many of them. Let z be a shortest word such that all states in $\delta(Q, z)$ are in the strongly connected bottom components of the automaton. It is well-known that the length of z is in $\mathcal{O}(n^2)$ [9].

Let $\{q_1, q_2\}$ be a subset to avoid. States q_1 and q_2 are either in the same strongly connected component or in separate ones. In any case, we can ignore every other strongly connected component, since their states cannot be mapped to q_1 or q_2. Let C_1 and C_2 be these components, respectively of q_1 and q_2. From now, consider the automaton containing only C_1 and C_2.

Let u_1 and u_2 be avoiding words respectively for q_1 and q_2, both of length at most $f(n)$.

We build iteratively some words $w_1, w_2, \ldots, w_{n-1}$. Each w_i will be of length $\mathcal{O}(i(n+f(n)))$. Let $w_0 = \varepsilon$.

Assume that we have built w_i, and let $X = \delta(Q, w_i)$. Now, we use a linear algebraic argument to infer that one of three possibilities hold:

(1) there exists a word v of length at most $n-1$ such that $q_1 \notin \delta(X, vu_2)$, or

(2) there exists a word v of length at most $n-1$ such that there are at least two distinct states $r_1, r_2 \in X$ such that $\delta(r_1, vu_2) = \delta(r_2, vu_2) = q_1$, or

(3) there does not exist any word v (of any length) satisfying (1) or (2).

We omit to repeat the argument here since it follows in the same way as in the proof of [12, Lemma 13] and requires introducing many linear algebraic definitions. In [12, Lemma 13], however, (3) cannot happen due to the assumption that the automaton is synchronizing and strongly connected, so such a word v always exist.

If (1) holds, then $w_i v u_2$ is an avoiding word for $\{q_1, q_2\}$ and it has a desired length. If (2) holds, then let $w_{i+1} = w_i v u_2$, which is longer than w_i by at most $n - 1 + f(n)$. If (3) holds, then it means that w_i in the automaton restricted to C_1 has the minimal rank. Otherwise, we could map two distinct states from $X \cap C_1$ to the same state, and then map it to q_1. In this case, we also stop with w_i.

If we have stopped with (3), then we repeat the construction symmetrically for q_2. If we also do not find an avoiding word, then we obtain a word $w'_{i'}$ that has the minimal rank in the automaton restricted to C_2. Altogether, $w_i w_{i'}$ has the minimal rank the automaton with both C_1 and C_2.

Then, it remains to use the upper bound from Theorem 2. Since the length of $w_i w_{i'}$ is an upper bound on m, we get the upper bound $\mathcal{O}(n^2 + n \cdot f(n))$.

Finally, we have to add z at the beginning to construct an avoiding word in the original automaton, and the length of z is also at most $\mathcal{O}(n^2)$. ◀

4 Bounding reset threshold with avoiding words

If 1-avoiding threshold is subquadratic, it would yield an upper bound on reset threshold equal to $7/48n^3 + o(n^3)$. However, the application of 1-avoiding threshold to bound reset threshold can be generalized to the usage of k-avoiding thresholds, if they are small enough. It turns out that, in a synchronizing automaton, a subquadratic value of k-avoiding thresholds already for small values of k is enough to imply a subcubic upper bound on the reset threshold. In the following calculations, we disregard particular constants and focus only on asymptotic bounds.

▶ **Lemma 10.** *Let (Q, Σ, δ) be an n-state synchronizing automaton and w be a word of rank $r \geq 2$. There exist at least one state $q \in \delta(Q, w)$ such that $|\delta^{-1}(\{q\}, w)| \leq \lfloor n/r \rfloor$.*

Proof. For each q, the states in $\delta^{-1}(\{q\}, w)$ are pairwise disjoint and they cover the whole Q. As we have r states in R, there exist a state $q \in R$ such that the cardinality of $\delta^{-1}(\{q\}, w)$ is at most n/r. ◀

▶ **Lemma 11.** *Let (Q, Σ, δ) be an n-state synchronizing automaton and w be a word of rank $r \geq 2$. There is a word of rank at most $r - 1$ and of length at most $|w| + d$, where d is the $\lfloor n/r \rfloor$-avoiding threshold.*

Proof. We take the state q from Lemma 10. Since $r \geq 2$, so $\lfloor n/r \rfloor < n$, thus every subset of that size is avoidable. We first avoid the subset $\delta^{-1}(\{q\}, w)$ by a word u of length at most d, and then apply w. We have $\delta(Q, uw) \subseteq \delta(Q, w)$ and $q \in \delta(Q, w)$, but also $q \notin \delta(Q, uw)$, thus we obtain that $\delta(Q, uw) \subsetneq \delta(Q, w)$. ◀

Taking the usual notation, $\omega(1)$ is the set of $\mathbb{R} \to \mathbb{R}$ functions growing faster than a constant.

▶ **Theorem 12.** *If there exists a function $f \in \omega(1)$ such that for every n-state synchronizing automaton and every $k \leq f(n)$, the k-avoiding threshold is at most $\mathcal{O}(n^2)/f(n)$, then the reset threshold is in $o(n^3)$.*

Proof. We build a reset word in two phases. First, we start with the empty word and iteratively apply Lemma 11 until the built word reaches a rank of at most $n/f(n)$. Thus, there are at most $n - n/f(n)$ applications of the lemma. The rank is at least $n/f(n)$ every time, so we need the k-avoiding threshold for $k \leq n/(n/f(n)) = f(n)$. Hence, d from the lemma is bounded above by $\mathcal{O}(n^2)/f(n)$ from the assumption of the theorem. It follows that the resulted word of rank $n/f(n)$ has length at most $(n-n/f(n)) \cdot \mathcal{O}(n^2)/f(n) = \mathcal{O}(n^3)/f(n)$.

In the second phase, we use the usual pair compression [8]. We need to compress a pair at most $n/f(n) - 1$ times, each time appending a word of length smaller than n^2. Thus, the word from this phase also has length at most $\mathcal{O}(n^3)/f(n)$.

Altogether, our reset word has length at most $n^3/f(n)$, which is in $o(n^3)$. ◀

5 Conclusions, discussion, and open problems

In the general case of an automaton and its k-avoiding threshold, we have a complete asymptotic solution for $k \geq 3$. The cases of $k = 2$ and $k = 1$ are open, yet a conjectured solution to the latter would solve also the former.

As for now, determining the maximum 1-avoiding threshold is the core problem, in particular in the class of synchronizing automata, but the general case of an automaton does not seem different for $k = 1$. Here, we considered a precise bound and have shown that $2n - 3$ is attainable for every $n \geq 2$. Yet, a few isolated examples reach $2n - 2$. We have the following conjecture:

A *trivial extension* of an automaton (Q, Σ, δ) is any automaton $(Q, \Sigma \cup \Sigma', \delta \cup \delta')$ such that each letter from Σ' act either as an identity or as a letter from Σ.

▶ **Conjecture 13.** *For an n-state automaton, the 1-avoiding threshold is at most $2n - 3$, except for a finite number of cases and their trivial extensions, where it is equal to $2n - 2$.*

A weak version (with any linear bound) of Conjecture 13 implies the following:

▶ **Conjecture 14.** *For an n-state automaton, a tight upper bound on 2-avoiding threshold is $\mathcal{O}(n^2)$.*

Conjecture 13 was verified for small cases [7], in particular up to 11 states for binary synchronizing automata. We have also experimented with the case of $k = 2$. The maximum 2-avoiding threshold of an n-state binary synchronizing automaton for $n = 3, \ldots, 10$ equals respectively 6, 8, 12, 17, 19, 23, 25, and 28. However, for this problem, the range is much too small to reveal the tendency.

The case of synchronizing automata is much harder, and the problem remains open for all k. We have the trivial upper bound $\mathcal{O}(n^3)$, thus the situation is surely different than in the general case already for $k \geq 4$ and most likely already for $k \geq 2$.

We have made an effort to find a lower bound on the maximum possible k-avoiding threshold also for $k \geq 2$. We have not found any series of automata that would exceed the upper bound $\mathcal{O}(kn)$ (when k is a variable). This bound is easily met, for example by the well-known Rystsov series [10] with a sink (zero) state and reset threshold $n(n-1)/2$. We should consider the following:

▶ **Conjecture 15.** *For an n-state synchronizing automaton, the k-avoiding threshold is at most $\mathcal{O}(kn)$ (when k is a variable).*

It turns out that this is a weaker (by several means) version of the disproved Don's conjecture [3, Conjecture 18], which states that if a subset T is reachable, then it is reachable with a word of length at most $n(n - |T|)$. We know that Conjecture 15 does not hold for non-synchronizing automata, and even for synchronizing automata with the original non-asymptotic bound (by e.g., Theorem 4 for $k = 1$). On the other hand, if the reset threshold is at most quadratic, then the conjecture trivially holds for $k \in \Theta(n)$.

Finally, we have noted that a subquadratic upper bound on k-avoiding threshold of a synchronizing automaton for small, but non-constant, values of k, is enough to imply a subcubic upper bound on reset threshold. This should motivate further efforts on bounding avoiding thresholds. Since the current upper bound on 1-avoiding threshold is quadratic, this requires both decreasing it and generalizing to k-avoiding thresholds for small k. We know that a subquadratic upper bound for $k = 2$ is not possible in the class of non-synchronizing automata. Yet, our conjectures are much stronger than the required upper bound to give further improvements.

References

1. M. V. Berlinkov, R. Ferens, and M. Szykuła. Preimage problems for deterministic finite automata. *Journal of Computer and System Sciences*, 115:214–234, 2021.
2. J. Černý. Poznámka k homogénnym eksperimentom s konečnými automatami. *Matematicko-fyzikálny Časopis Slovenskej Akadémie Vied*, 14(3):208–216, 1964. In Slovak.
3. H. Don. The Černý Conjecture and 1-Contracting Automata. *Electronic Journal of Combinatorics*, 23(3):P3.12, 2016.
4. F. Gonze and R. M. Jungers. Hardly Reachable Subsets and Completely Reachable Automata with 1-Deficient Words. *Journal of Automata, Languages and Combinatorics*, 24(2–4):321–342, 2019.
5. 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.
6. J. Kari. Synchronizing finite automata on Eulerian digraphs. *Theoretical Computer Science*, 295(1-3):223–232, 2003.

7 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.

8 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.

9 I. K. Rystsov. Polynomial complete problems in automata theory. *Information Processing Letters*, 16(3):147–151, 1983.

10 I. K. Rystsov. Quasioptimal Bound for the Length of Reset Words for Regular Automata. *Acta Cybernetica*, 12(2):145–152, 1995.

11 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.

12 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.

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

14 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.

15 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.

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

17 M. V. Volkov, editor. *Special Issue: Essays on the Černý Conjecture*, volume 24 (2–4) of *Journal of Automata, Languages and Combinatorics*, 2019.

18 V. Vorel. Subset synchronization and careful synchronization of binary finite automata. *Int. J. Found. Comput. Sci.*, 27(5):557–578, 2016.