Binary 3-compressible automata

Alessandra Cherubini[1⋆] and Andrzej Kisielewicz[2⋆⋆]

[1] Politecnico di Milano, Dipartimento di Matematica
[2] Department of Mathematics and Computer Science, University of Wrocław
alessandra.cherubini@polimi.it, andrzej.kisielewicz@math.uni.wroc.pl

Abstract. A finite deterministic automaton $\mathcal{A} = (Q, \Sigma, \delta)$ is k-compressible if there is a word $w \in \Sigma^+$ such that the image of the state set Q under the natural action of w is reduced by at least k states. In such case w is called a k-compressing word for $\mathcal{A}$. It is known that, for any alphabet Σ and any $k \geq 2$, there exist words that are k-compressing for each k-compressible automaton with the input alphabet Σ. Such words are called k-collapsing. It has been proved that recognizing 2-collapsing words over a 2-element alphabet may be done in polynomial time, while recognizing 2-collapsing words over an alphabet of size ≥ 3 is co-NP-complete. A natural question in this context, whether recognizing 3-collapsing words over a 2-element alphabet is easy or hard, has remained open. In this paper we provide results on 3-compressible binary automata, which allow to prove that that the latter problem is co-NP-complete.

1 Introduction

Let $T(Q)$ be the full transformation monoid on a finite set Q. For $f \in T(Q)$, we define the *deficiency* $df(f)$ of f to be the difference between the cardinalities of Q and the image Imf under f, $df(f) = |Q| - |Imf|$. At the beginning of 1990's Sauer and Stone [12] introduced the property Δ_k defined as follows. For a finite alphabet Σ and a positive integer k, a word $w \in \Sigma^+$ has the property Δ_k for Σ if for all homomorphisms $\phi : \Sigma^+ \to T(Q)$, where Q is any finite set, $df(w\phi) \geq k$ whenever $df(v\phi) \geq k$ for some $v \in \Sigma^+$. They proved the nonobvious fact that such words exist for each positive integer k and for each finite alphabet Σ giving an elegant recursive construction that produces a word whose length is $O(2^{2^k})$. Their construction was improved in [9], where better but yet unrealistic upper bounds for the length of shortest words with the property Δ_k were given.

Words with the property Δ_k have a natural interpretation in the language of finite automata theory. Each complete deterministic automaton $\mathcal{A} = (Q, \Sigma, \delta)$ over the alphabet Σ can be viewed as the transformation monoid generated by transformations on Q induced via δ by the letters of Σ. Namely, for each $\alpha \in \Sigma$ we define the induced transformation by $q\alpha = \delta(q, \alpha)$. This action of the letters

⋆ Supported in part by PRIN: "Automi e linguaggi formali: aspetti matematici e applicativi".
⋆⋆ Supported in part by Polish NCN grant 2012/07/B/ST1/03318.

on Q extends naturally into the action of words $w \in \Sigma^+$ on Q which is denoted briefly by $qw = \delta(q, \alpha)$.

Conversely, to define an automaton it is enough to assign to any letter of Σ a transformation on Q. Thus an automaton $\mathcal{A}$ can be identified with a specific homomorphism $\phi_{\mathcal{A}}$ of Σ^+ in $T(Q)$. If there exists a word $v \in \Sigma^+$ such that $df(v\phi_{\mathcal{A}}) \geq k$—that is, if $|Q| - |Qv| \geq k$—the automaton $\mathcal{A}$ is called *k-compressible* and v is a *k-compressing* word for $\mathcal{A}$ (it *k-compresses* $\mathcal{A}$). A word that is k-compressing for each k-compressible automaton with the input alphabet Σ is called *k-collapsing*. Obviously, a word has the property Δ_k (or witnesses for deficiency k, in the terminology of [9]) if and only if it is k-collapsing. Hence, besides the original motivations coming from combinatorics and algebra, the interest in such words comes from the fact that they can be seen as universal testers whose action on the set of states of an automaton exposes whether or not the automaton is k-compressible. The problem of the length of the shortest k-collapsing word over an alphabet Σ can be considered as a black-box version of the generalized Černý's conjecture stated by Pin [7, 8].

In [10] it is proved that the membership of a given word $w \in \Sigma^+$ to the language of k-collapsing words is decidable. The decision procedure is in the class co-NP and requires linear space, which shows that the language of k-collapsing words is context-sensitive. In [11] it is shown that it is not context-free even in the very simple case of the language of 2-collapsing words over a 2-letter alphabet. Most results so far concern 2-collapsing words. In particular, 2-collapsing words were characterized in [2] and in [5]. From the first characterization, that has a group theoretical flavor, a non-deterministic polynomial algorithm to recognize whether a word $w \in \Sigma^+$ is 2-collapsing was derived [3]. A refinement of this algorithm was used in [4] to give the list of shortest 2-collapsing words over a 3-letter alphabet. The second characterization is in terms of systems of permutation conditions and it is used in [6] to show that the membership problem of the language of 2-collapsing words over an alphabet of size ≥ 3 is a co-NP-complete problem. The algorithms for recognizing whether a word $w \in \Sigma^+$ is 2-collapsing, derived by both the characterizations of 2-collapsing words, become polynomial algorithms when $|\Sigma| = 2$ even if w is represented in the compressed form [5]. In view of these results a question arose whether for $k \geq 3$, k-collapsing words over a binary alphabet can also be recognized in polynomial time. This natural question has remained open so far. In this paper we show that the answer is negative. We prove that the problem of recognizing whether a word $w \in \{\alpha, \beta\}^+$ is 3-collapsing is co-NP-complete.

The difficulty in this case is that we have no characterization of 3-collapsing words similar to that for the case of 2-collapsing words. Yet, we have a certain classification of proper 3-compressible automata on 2 input letters [1], and we can see that the problem looks differently depending on the class. In some classes it is easy to recognize whether a word w 3-compresses all the automata in the class. There is however at least one class where this problem leads to solving a sort of a system of permutation conditions, similarly as in [6], and it seems computationally hard.

Our idea is the following. First, we reduce the problem whether a word w is 3-collapsing to the problem of recognizing those words w that 3-compress all automata in a restricted class $\mathcal{D}$ of 3-compressible automata. Next, we show that in this restricted class the problem is equivalent to the existence of a solution of a certain system of permutation/transformation conditions similar to those considered in [6]. Then, using the tools worked out in [6], we show that the problem of the existence of a solution of such systems is NP-hard. Since our reductions induce suitable polynomial transformations, we obtain a proof that the initial problem is co-NP-complete. In this paper, we present the first part of the plan, showing how to reduce the problem to a problem concerning systems of transformation conditions. The full proof will appear in the extended version of this paper.

2 Preliminaries

We deal with binary automata over the alphabet $\Sigma = \{\alpha, \beta\}$. The letters α and β are identified with transformations they induce on the set Q of the states. The image of $q \in Q$ by a letter (transformation) α is denoted $q\alpha$. The words $w = \alpha_1\alpha_2 \dots \alpha_t$ over Σ are identified with transformations they induce. The inverse image is denoted by $x\alpha^{-1}$.

We use special notation for concrete transformations α (similar to permutation notation), where round brackets $(x_1x_2 \dots x_t)$ denote a cycle: $x_1\alpha = x_2, \dots, x_t\alpha = x_1$, and square brackets $[x_1x_2 \dots x_t]$ denote a path: $x_1\alpha = x_2, \dots, x_{t-1}\alpha = x_t$. This notation is not unique: for example $[123][42](357) = [12][423](357)$. Yet, we always write full cycles, and refer to them as the *cycles* of α, and usually omit all the fixed points. If x is an element of a cycle in α, then we write $Cyc_\alpha(x)$ to denote the cycle containing x, or a set of elements in this cycle, and we write $|Cyc_\alpha(x)|$ to denote the length of this cycle (the number of elements). We will also use a part of the structure of a transformation to speak about transformations of a given form. For example, a transformation (permutation) is of the form $\beta = (12y)(xa)(zb)\dots$ for some elements $y, x, z, a, b \in Q$ if $\beta = (12y)(xa)(zb)\tau$ for some transformation τ. We do not exclude that some of these elements may be equal, and some of these cycles coincide. For example, we may have $x = z$ and (consequently) $a = b$, in which case the cycle (xa) is the same as the cycle (zb). We may have also $x = a$, which means that (xa) is, in fact, a fixpoint $(x) = (a)$. But, assuming $1 \neq 2$, we cannot have $x = y$, because y is in a cycle of length 3, while x is in a cycle of length 2 or 1. Thus saying that a transformation is of the form $\beta = (a_1a_2 \dots a_t)\dots$ we assume that the length of the cycle is t or a divisor of t.

Treating words over $\Sigma = \{\alpha, \beta\}$ as compositions of transformations on the set Q with $1, 2 \in Q$, we shall consider systems of *transformation conditions* of the form

$$1u_1, 1u_2, \dots, 1u_s \in \{1, 2\}$$

stating that the image of 1 by each of words $u_1, u_2, \dots, u_s$ belongs to the set $\{1, 2\}$. If all transformations in Σ fix 1 or all fix the set $\{1, 2\}$, then they form

a solution of the system (1), which is called *trivial*. The problem whether there exists a nontrivial solution for a system of permutation conditions has been proved to be NP-complete in [6]. Similarly one can prove that the problem is hard if we look for a solution in transformations of given types. We will exploit this in our proof.

We recall that a *factor* of a word $w \in \Sigma^+$ is a word $v \in \Sigma^+$ such that $w = uvz$ for some $u, z \in \Sigma^*$. If $\mathcal{A}$ is k-compressible at least one letter of its input alphabet has deficiency greater than 0. It is known that each k-collapsing word over a fixed alphabet Σ is k-*full* [12], that is, contains each word of length k over the alphabet Σ among its factors. A k-*compressible automaton* is called *proper* k-compressible if it cannot be k-compressed by any word of length k. Thus, k-collapsing words are k-full words that k-compress all proper k-compressible automata. In particular, in our consideration we may restrict to proper k-compressible automata.

We will consider types of transformations with regard to which states are sent into the same element, and which states are missing in the image. We say that a transformation α is of type $I\backslash M$, where M is a subset of Q, and I is a family of disjoint subsets of Q, if $M = Q \setminus Q\alpha$ is a set of elements missing in the image $Q\alpha$, while I is the family of those inverse images of elements of Q that have more than one element. In other words, we write that a letter α is of type $[x_{1_1}, \ldots, x_{j_1}][x_{1_2}, \ldots, x_{j_2}] \ldots [x_{1_r}, \ldots, x_{j_r}]\backslash y_1, y_2, ..., y_m$, if $\{y_1, \ldots y_m\} = Q \setminus Im(\alpha)$ and $\{x_{1_1}, \ldots, x_{j_1}\}, \{x_{1_2}, \ldots, x_{j_2}\}, \ldots, \{x_{1_r}, \ldots, x_{j_r}\}$ are the equivalence classes of the kernel of transformation induced by α that have more than one element. We say that $\alpha \in \Sigma$ is a *permutation letter* if it induces a permutation on the set Q of the states, i.e., it has deficiency 0. Otherwise, a letter is a *non-permutation*.

The following fact leading to a classification of proper 3-compressible automata is not difficult to prove (see e.g. [1]).

Proposition 1. *If $\mathcal{A}$ is a proper 3-compressible automaton over the alphabet $\Sigma = \{\alpha, \beta\}$ then each letter in Σ is either a permutation or is one of the following types:*

1. $[x, y, z]\backslash x, y$;
2. $[x, y][z, t]\backslash x, z$;
3. $[x, y]\backslash x$;
4. $[x, y]\backslash z$ *with* $za \in \{x, y\}$.

where x, y, z, t are different states of $\mathcal{A}$.

Since we wish to classify proper 3-compressible automata up to renaming the states in Q, we may assume that $Q = \{1, 2, \ldots, n\}$, and $\{x, y, z, t\} = \{1, 2, 3, 4\}$. Then, following [1], we say that an automaton $\mathcal{A}$ over a two-letter alphabet $\Sigma = \{\alpha, \beta\}$ is an $(\mathbf{i}., \mathbf{j}.)$-*automaton*, where $\mathbf{i}, \mathbf{j} \in \{\mathbf{1}, \mathbf{2}, \mathbf{3}, \mathbf{4}\}$, if the letter α is of type $\mathbf{i}.$, while the letter β is of type $\mathbf{j}.$, above. We say also that $\mathcal{A}$ is a $(\mathbf{i}., \mathbf{p})$ (or $(\mathbf{p}, \mathbf{i}.)$-*automaton*, if it is an automaton in which the letter α (β) is of type $\mathbf{i}.$, while the other letter is a permutation. By Proposition 1, up to renaming the states, each proper 3-compressible automaton over 2 letters is a $(\mathbf{t}, \mathbf{s})$-*automaton* with some $\mathbf{t}, \mathbf{s} \in \{\mathbf{1}., \mathbf{2}., \mathbf{3}., \mathbf{4}., \mathbf{p}\}$.

3 Automata of Type (3., p)

Let $\mathcal{A}$ be a $(\mathbf{3}., \mathbf{p})$-automaton over the alphabet $\Sigma = \{\alpha, \beta\}$ and with the state set $Q = \{1, 2, \ldots, n\}$. Without loss of generality we can also assume that the letter α is of type $[1, 2]\backslash 1$ (and β is a permutation). Given such an automaton, we call an integer $k > 0$ a *good exponent* for the permutation β, or briefly, β-*good*, if $1\beta^k \notin \{1, 2\}$; otherwise, it is called β-*bad* exponent. We will consider transformation conditions of the form $1v \in \{1, 2\}$ stating the the image of 1 by the word v is 1 or 2. For a word $v \in \Sigma^+$ and $Q_1 \subseteq Q$, as in [1], we denote $\mathcal{M}(v) = Q \setminus Qv$, the set of the states missing in the image of Q under the action of v, and $\mathcal{M}(Q_1, v) = Q \setminus (Q \setminus Q_1)v$, the set of states missing in Q after applying the word v, provided that the set Q_1 of states is already missing.

Our first result is the following characterization.

Theorem 1. *Let $\mathcal{A} = \langle Q, \Sigma, \delta\rangle$ be a proper 3-compressible $(\mathbf{3}., \mathbf{p})$-automaton with Q and Σ as above, and w be a word in Σ^+. Then, w* ***does not*** *3-compresses $\mathcal{A}$ if and only if for every factor of w of the form $\alpha u\alpha$ where*

$$u = \beta^{k_1}\alpha^{m_1} \ldots \beta^{k_t}\alpha^{m_t}\beta^{k_{t+1}},$$

$t \geq 1$, and k_1 and k_{t+1} are β-good, while all other k_i are β-bad, the condition $1u \in \{1, 2\}$ holds.

Proof. Let $w = \beta^s\alpha^{r_1}\beta^{s_1}\alpha^{r_2} \ldots \beta^{s_{h-1}}\alpha^{r_h}\beta^{s_h}$, where $s, s_h \geq 0$, $r_h > 0$, and $r_i, s_i > 0$ for $1 \leq i \leq h - 1$. We consider the sequence of missed states by the action of consecutive prefixes of w. Obviously $\mathcal{M}(\beta^s\alpha^{r_1}) = \{1\}$. Moreover if s_1 is a β-bad exponent then $\mathcal{M}(\{1\}, b^{s_1}) = \mathcal{M}(\beta^s\alpha^{r_1}\beta^{s_1}) = \{1\beta^{s_1}\} \subseteq \{1, 2\}$, hence $\mathcal{M}(\beta^s\alpha^{r_1}\beta^{s_1}\alpha^{r_2}) = \{1\} = \mathcal{M}(\beta^s\alpha^{r_1})$. Repeating the same argument it turns out that if there is no β-good exponent among $s_1, s_2, \ldots, s_h$, then w does not 3-compress α.

So we may assume that there is i with $1 \leq i \leq h - 1$ such that s_i is a β-good exponent and s_j is a β-bad exponent for each $j < i$. Then we have $w = p\alpha^{r_i}\beta^{s_i}\alpha^{r_{i+1}}\beta^{s_{i+1}} \ldots$, where $\mathcal{M}(p\alpha^{r_i}) = \{1\}$. So $\mathcal{M}(\{1\}, \beta^{s_i}) = \{1\beta^{s_i}\} \nsubseteq \{1, 2\}$, and $\mathcal{M}(\{1\beta^{s_i}\}, \alpha^{r_{i+1}}) = \{1, x_1\}$ where $x_1 = 1\beta^{s_i}\alpha^{r_{i+1}} \neq 1$ (since α is of type $[1, 2]\backslash 1$). Now if s_{i+1} is β-bad, then $\mathcal{M}(p\alpha^{r_i}\beta^{s_i}\alpha^{r_{i+1}}\beta^{s_{i+1}}) = \mathcal{M}(\{1, x_1\}, \beta^{s_{i+1}}) = \{y, x_1\beta^{s_{i+1}}\}$ with $y \in \{1, 2\}$, whence, whether $x_1\beta^{s_{i+1}} \in \{1, 2\}$ or not, we have $\mathcal{M}(p\alpha^{r_i}\beta^{s_i}\alpha^{r_{i+1}}\beta^{s_{i+1}}\alpha^{r_{i+2}}) = \mathcal{M}(\{y, x_1\beta^{s_{i+1}}\}, \alpha^{r_{i+2}}) = \{1, x_2\}$ where $x_2 = x_1\beta^{s_{i+1}}\alpha^{r_{i+2}} \neq 1$. Then until we encounter β^{s_m} with a β-good exponent s_m the missing state set of each prefix of w ending with α has cardinality 2 and contains 1; in particular, if there no such second β-good exponent, w does not 3-compress α.

So, assume that there is m with $i < m \leq h$ such that s_m is a β-good exponent and each s_t with $i < t < m$ is a β-bad exponent. Then $w = p\alpha^{r_i}\beta^{s_i}\alpha^{r_{i+1}}v\alpha^{r_{m+1}} \ldots$, where $v = \beta^{s_{i+1}}\alpha^{r_{i+2}} \ldots \beta^{s_m}$. Now,

$$\mathcal{M}(\alpha^{r_i}\beta^{s_i}\alpha^{r_{i+1}}v) = \mathcal{M}(\{1, x_1\}, v) = \mathcal{M}(\{1, x_{m-i}\}, beta^{s_m}) = \{1\beta^{s_m}, x_{m-i}\beta^{s_m}\},$$

where $x_{m-i}\beta^{s_m} = 1v$ (since $x_1 = 1\beta^{s_i}\alpha^{r_{i+1}}, x_2 = x_1\beta^{s_{i+1}}\alpha^{r_{i+2}}$, etc.). Hence if $1v \notin \{1,2\}$ then $\mathcal{M}(\{1\beta^{s_m}, 1v\}, \alpha)$ has cardinality 3, and w is a 3-compressing word. If $1v \in \{1,2\}$ then $1v\alpha^{r_{m+1}} = 1\alpha^{r_{m+1}}$, hence $\mathcal{M}(p\alpha^{r_i}\beta^{s_i}\alpha^{r_{i+1}}v\alpha^{r_{m+1}}) = \{1\beta^{s_m}, 1\alpha^{r_{m+1}}\}$. So we are in the analogous situation as with $\mathcal{M}(p\alpha^{r_i}\beta^{s_i}\alpha^{r_{i+1}})$, and we can use the above argument with i replaced by m, the prefix p replaced by $p_1 = p\alpha^{r_i}\beta^{s_i}\ldots\beta^{s_{m-1}}$, and β-good exponent s_i replaced by β-good exponent s_m. Thus, the reasoning may be repeated for consecutive factors $\alpha u\alpha$ with the property stated in the theorem, and the claim easily follows. □

The theorem above means that to check whether there exists a $(\mathbf{3.}, \mathbf{p})$-automaton that fails to be 3-compressed by a given word w, we need to consider all the possibilities for β leading to various sets of good and bad exponents. This means that we consider whether 1 and 2 are in the same orbit of β. If so, we consider various lengths of this orbit and distances between 1 and 2, and if not, the only thing that counts is the length of the orbit of 1. For each such case we establish β-good and β-bad exponents, find all the factors of w of the form described in the theorem above, and form the corresponding system of transformation conditions. If the system has a solution, then w fails to 3-compress the corresponding automaton $\mathcal{A}$ with a non-permutation α and a permutation β. The solution represents an automaton that is not 3-compressed by word w. It is *nontrivial*, if the resulting automaton is 3-compressible (some solutions correspond to automata that fails to be 3-compressible).

In principle, there are infinitely many cases to consider. But given the word w, if k is the largest exponent for β (when w is written in the compact form with exponents), then only $k-1$ elements following 1 in the cycle of 1 in β are what really counts. So, in practice, for every word, we may distinguish and consider only finitely many cases.

In fact, we may view α as (almost) a permutation without 1 (1 is not an image of any state; we need only to keep in mind that 1 has the same image as 2). The system of transformation conditions may be treated very similarly to the system of suitable permutation conditions, and the method of trees with distinguished nodes developed in [6] may be used to solve it. (This is so, in spite of that the two letters involved come in the both cases from very different considerations.)

4 Reduction to a Smaller Class of Automata

To prove that solving systems of transformations in Theorem 1 is hard we need to restrict considerations to one class in which the conditions may be stated in a unique and relatively simple form. Therefore, we define the class $\mathcal{D}$ to contain all proper 3-compressible automata $\mathcal{A}$ over alphabet $\Sigma = \{\alpha, \beta\}$, such that α is a transformation of type $[1,2]\backslash 1$, and β is a permutation of the form $\beta = (12y)\ldots$. For such permutation the β-good exponents are positive integers $k = 2$ modulo 3. We will refer to them as to $(12y)$-good. Other positive integers are $(12y)$-bad.

In order to reduce our problem to automata in $\mathcal{D}$, to each word $w \in \{\alpha, \beta\}^+$ we assign a word $w^* = w'w$ such that w^* is 3-collapsing if and only if w 3-compresses all automata in $\mathcal{D}$. The idea is to find a set of words that 3-compress

all automata not in $\mathcal{D}$, and do not 3-compress automata in $\mathcal{D}$. These words will be used to form w'. We wish that these words have no $(12y)$-good exponents, because then w^* and w have the same factors described in Theorem 1, which guarantees that they 3-compress the same proper automata in $\mathcal{D}$. This cannot be achieved exactly, and the few exceptions we allow will be handled further in a different way.

We start from $(\mathbf{3}., \mathbf{p})$-automata not in $\mathcal{D}$. The following technical lemma established in [1] will be useful here.

Lemma 1. ([1], Lemma 3) *Let $\mathcal{A}$ be a $(\mathbf{3}., \mathbf{p})$-automaton with α of type $[1,2]\backslash 1$. Then $\mathcal{A}$ is not 3-compressible if, and only if (up to renaming the states) one of the following conditions holds:*

(i) β *fixes 1 or the set* $\{1,2\}$,
(ii) $\beta = (13)\pi$ *for some permutation* π *on* $Q \setminus \{1,3\}$ *and* $3\alpha = 3$,
(iii) $\beta = (13)(2)\pi$, $\beta = (132)\pi$ *or* $\beta = (123)\pi$ *for some permutation* π *on* $Q \setminus \{1,2,3\}$ *and* $\{2,3\}\alpha = \{2,3\}$,
(iv) $\beta = (13)(24)\pi$ *or* $\beta = (1324)\pi$ *for some permutation* π *on* $Q \setminus \{1,2,3,4\}$ *and* $\{3,4\}\alpha = \{3,4\}$.

In all other cases $\mathcal{A}$ is a proper 3-compressible automaton, and one of the words $\{ababa, abab^2a, aba^2ba, abab^2aba, ab^2ab^2a, ab^2a^2b^2a, ab^2abab^2a, ab^2aba, ab^3aba, abab^3a, ab^3ab^3a\}$ 3-compresses $\mathcal{A}$.

Using it we get the following.

Lemma 2. *Each proper 3-compressible $(\mathbf{3}., \mathbf{p})$-automaton $\mathcal{A} \notin \mathcal{D}$ can be 3-compressed by a word of the form $\alpha\beta^i\alpha^k\beta^j\alpha$, where $i,j \in \{1,3,4\}, k \in \{1,2,3\}$ or $\alpha\beta^4\alpha\beta^2\alpha\beta^4\alpha$.*

Proof. Let $\mathcal{A}$ be a $(\mathbf{3}., \mathbf{p})$-automaton $\mathcal{A} \notin \mathcal{D}$, then by (i) of Lemma 1, b fixes neither 1 nor the set $\{1,2\}$.

First, assume that $1\beta = x \notin \{1,2\}$. Then, $\mathcal{M}(\alpha\beta\alpha^k) = \{1, x\alpha^k\}$. If $x\alpha^k\beta \notin \{1,2\}$ for some k then the word $\alpha\beta\alpha^k\beta\alpha$ 3-compresses the automaton $\mathcal{A}$. Obviously, if $x\alpha^k\beta \notin \{1,2\}$, for some k we can always assume that $k \leq 3$. Otherwise, either $x\alpha = x$ or $x\alpha^2 = x$ and $x\alpha^k\beta \in \{1,2\}$ for all k.

Then let $x\alpha^k\beta \in \{1,2\}$ for each $k \in \{1,2,3\}$.

First assume $x\alpha = x$. If $x\alpha\beta = x\beta = 1$ then $\beta = (1x)\ldots$ and by (ii), $\mathcal{A}$ is not 3-compressible. So $x\alpha\beta = x\beta = 2$ and $\beta = (1x2\ldots)\ldots$. If $\beta = (1x2)\ldots$ then $2\alpha \neq 2$ otherwise $\mathcal{A}$ is not 3-compressible by (iii). Then $\mathcal{M}(\alpha\beta^{1+3h}\alpha\beta^2) = \{1,2\}$ for each $h \geq 0$. Since $\mathcal{M}(\{1,2\},\alpha) = \{1, 2\alpha\}$ and $2\alpha \notin \{1,2,x\}$, then $\alpha\beta^{1+3h}\alpha\beta^2\alpha\beta^{1+3k}\alpha$ 3-compresses $\mathcal{A}$. So let $\beta = (1x2y\ldots)\ldots$. Then $\mathcal{M}(\alpha\beta^3\alpha) = \{1, y\alpha\}$ and if $y\alpha\beta \notin \{1,2\}$ then $\alpha\beta^3\alpha\beta\alpha$ 3-compresses $\mathcal{A}$. We know that $y\alpha\beta \neq x\alpha\beta = 2$, so let $y\alpha\beta = 1$. If $y\alpha = y$ then $\mathcal{A}$ is not 3-compressible by (iv). If $y\alpha \neq y$ then $\alpha\beta^3\alpha\beta^4\alpha$ 3-compresses $\mathcal{A}$.

Then let $x\alpha \neq x$ and $x\alpha^2 = x$. Put $x\alpha = y$, then $y\alpha = x$ and $y\alpha\beta = x\alpha^2\beta = x\beta \in \{1,2\}$. If $x\beta = 1$ then $\beta = (1x)(y2z\ldots)\ldots$ where y, z, 2 are all distinct by (iii) and (iv). Hence $\alpha\beta\alpha\beta^3\alpha$ 3-compresses $\mathcal{A}$. So let $x\beta \neq 1$, hence $x\beta = 2$,

and again by (iv), $\beta = (1x2z\dots y)\dots$ where $1, x, 2, z, y$ are distinct states, then $\alpha\beta\alpha\beta^4\alpha$ is a 3-compressing $\mathcal{A}$. This completes all the cases when $1\beta = x \neq 1, 2$.

Lastly, let $1\beta = 2$. Then $\beta = (12xy\dots)\dots$ where $1, 2, x, y$ are distinct elements, otherwise $\mathcal{A} \in \mathcal{D}$. Then $\mathcal{M}(\alpha\beta^3\alpha) = \{1, y\alpha\}$. If $y\alpha\beta^3 \notin \{1, 2\}$ then $\alpha\beta^3\alpha\beta^3\alpha$ is 3-compressing, similarly if $y\alpha\beta^4 \notin \{1, 2\}$ then $\alpha\beta^3\alpha\beta^4\alpha$ is 3-compressing. So $y\alpha\beta^3 = 1$ and $y\alpha\beta^4 = 2$ whence $y \neq y\alpha$ and so $y\alpha \neq y\alpha^2$, and so $y\alpha^2\beta^3 \neq 1$. If $y\alpha^3\beta^3 \neq 2$ then $\alpha\beta^3\alpha^2\beta^3\alpha$ is 3-compressing, if $y\alpha^3\beta^3 = 2$, then $y\alpha^3\beta^4 = x$ and $\alpha\beta^3\alpha^2\beta^4\alpha$ is 3-compressing. □

For the automata of types other than $(\mathbf{3}., \mathbf{p})$ we make use of Proposition 1 and other lemmas established in [1]. These lemmas have been established in [1] to compute a bound for the length of the shortest 3-collapsing word. We extract from them information we need to our aim. We have the following.

(i) no 3-compressible $(\mathbf{i}., \mathbf{j}.)$-automaton with $\mathbf{i} \in \{\mathbf{1}, \mathbf{2}\}$, $\mathbf{j} \in \{\mathbf{1}, \mathbf{2}, \mathbf{4}\}$ or with $\mathbf{j} \in \{\mathbf{1}, \mathbf{2}\}$, $\mathbf{i} \in \{\mathbf{1}, \mathbf{2}, \mathbf{4}\}$ is proper; [1, Lemma 5];

(ii) each proper 3-compressible $(\mathbf{1}., \mathbf{p})$-automaton and each proper 3-compressible $(\mathbf{1}., \mathbf{3})$-automaton is 3-compressed by $\underline{\alpha\beta^2\alpha}$; hence (by switching the letters), each proper 3-compressible $(\mathbf{p}., \mathbf{1})$ or $(\mathbf{3}., \mathbf{1})$-automaton is 3 compressed by $\beta\alpha^2\beta$; [1, Lemma 1 and Lemma 6];

(iii) each proper 3-compressible $(\mathbf{2}., \mathbf{p})$ and each proper 3-compressible $(\mathbf{2}., \mathbf{3})$-automaton is 3-compressed by a word in the set $\{\underline{\alpha\beta^2\alpha}, \alpha\beta^3\alpha\}$; each proper 3-compressible $(\mathbf{p}., \mathbf{2})$ and each proper 3-compressible $(\mathbf{3}., \mathbf{2})$-automaton is 3-compressed by a word in the set $\{\beta\alpha^2\beta, \beta\alpha^3\beta\}$; [1, Lemma 2 and Lemma 7];

For the future reference we underline words that have an occurrence of β with a $(12y)$-good exponent *inside* the word. Here, this is limited only to occurrences of β^2. These words will be handled in a different way than those words that have no occurrence β with a $(12y)$-good exponent or they have such an occurrence only at the beginning or at the end of the word. We will see that only the exact form of underlined words is what really counts in our proof.

In order to find a set of words 3-compressing all $(\mathbf{p}, \mathbf{3}.)$-automata we use again Lemma 1, yet switching the letters α and β. It yields the following set: $\{\beta\alpha\beta\alpha\beta, \beta\alpha\beta\alpha^2\beta, \beta\alpha\beta^2\alpha\beta, \beta\alpha\beta\alpha^2\beta\alpha\beta, \beta\alpha^2\beta\alpha^2\beta, \beta\alpha^2\beta^2\alpha^2\beta, \beta\alpha^2\beta\alpha\beta\alpha^2\beta, \beta\alpha^2\beta\alpha\beta, \beta\alpha^3\beta\alpha\beta, \beta\alpha\beta\alpha^3\beta, \beta\alpha^3\beta\alpha^3\beta\}$. Some of these words are factors of others, so we may infer the following:

(iv) each proper 3-compressible $(\mathbf{p}, \mathbf{3}.)$-automaton is 3-compressed by any word with a factor of the form $\beta\alpha\beta\alpha\beta, \underline{\beta\alpha\beta^2\alpha\beta}, \beta\alpha\beta\alpha^2\beta\alpha\beta, \beta\alpha^2\beta\alpha^2\beta, \underline{\beta\alpha^2\beta^2\alpha^2\beta}, \beta\alpha^2\beta\alpha\beta\alpha^2\beta, \beta\alpha^3\beta\alpha\beta, \beta\alpha\beta\alpha^3\beta, \beta\alpha^3\beta\alpha^3\beta$; [1, Lemma 3];

Similarly we get the following

(v) each proper 3-compressible $(\mathbf{4}., \mathbf{p})$ is 3-compressed by any word with a factor of the form $\alpha\beta\alpha\beta\alpha, \underline{\alpha^2\beta^2\alpha^2}, \alpha^2\beta\alpha^2, \alpha^2\beta^3\alpha, \alpha\beta^3\alpha\beta^3\alpha, \underline{\alpha^2\beta\alpha\beta^2\alpha}$; each proper 3-compressible $(\mathbf{p}., \mathbf{4})$ is 3-compressed by any word with a factor of the form $\beta\alpha\beta\alpha\beta, \beta^2\alpha^2\beta^2, \beta^2\alpha\beta^2, \beta^2\alpha^3\beta, \beta\alpha^3\beta\alpha^3\beta, \beta^2\alpha\beta\alpha^2\beta$; [1, Lemma 4];

(vi) each proper 3-compressible $(\mathbf{3}., \mathbf{3})$-automaton is 3-compressed by a word in the set $\{\alpha\beta\alpha\beta, \underline{\alpha\beta^2\alpha\beta}, \alpha\beta\alpha^2\beta, \underline{\alpha\beta^2\alpha^2\beta}, \beta\alpha\beta\alpha, \beta\alpha^2\beta\alpha, \underline{\beta\alpha\beta^2\alpha}, \underline{\beta\alpha^2\beta^2\alpha}\}$; [1, Lemma 9];

(vii) each proper 3-compressible $(\mathbf{3}., \mathbf{4})$-automaton is 3-compressed by a word in the set $\{\beta^2\alpha\beta^2, \beta^2\alpha^2\beta^2, \beta^2\alpha^3\beta^2, \beta^2\alpha\beta\alpha\beta^2\}$; each proper 3-compressible $(\mathbf{4}., \mathbf{3})$-automaton is 3-compressed by a word in the set $\{\alpha^2\beta\alpha^2, \underline{\alpha^2\beta^2\alpha^2}, \alpha^2\beta^3\alpha^2, \alpha^2\beta\alpha\beta\alpha^2\}$; [1, Lemma 10];

(viii) each proper 3-compressible $(\mathbf{4}., \mathbf{4})$ is 3-compressed by a word in the set $\{\alpha^2\beta\alpha^2, \alpha^2\beta^2, \beta^2\alpha\beta^2, \beta^2\alpha^2\}$; [1, Lemma 11].

Using the lemmas above we can see that any word containing as factors the following words

(I) $\underline{\alpha^2\beta\alpha\beta^2\alpha}, \underline{\beta\alpha^2\beta^2\alpha^2\beta}, \underline{\alpha\beta^4\alpha\beta^2\alpha\beta^4\alpha}$,

(II) $\alpha\beta^3\alpha\beta^3\alpha, \beta\alpha^3\beta\alpha^3\beta, \beta^2\alpha\beta\alpha^2\beta, \beta\alpha\beta\alpha^2\beta\alpha\beta, \beta\alpha^2\beta\alpha^2\beta, \beta\alpha^2\beta\alpha\beta\alpha^2\beta$, $\beta\alpha^3\beta\alpha\beta$, $\beta\alpha\beta\alpha^3\beta, \beta\alpha^3\beta\alpha^3\beta, \beta^2\alpha^2\beta^2, \beta^2\alpha^3\beta^2, \beta^2\alpha\beta\alpha\beta^2, \alpha^2\beta^3\alpha^2$.

(III) $\alpha\beta^i\alpha^k\beta^j\alpha$, where $i, j \in \{1, 3, 4\}, k \in \{1, 2, 3\}$.

3-compresses all 3-compressible automata except those in the $\mathcal{D}$.

To form a single word that 3-compresses all automata not in $\mathcal{D}$ it is enough to concatenate all the words listed in (I-III) above. Yet, we wish to have such a word without $(12y)$-good exponents. So, at this moment, we have only a partial solution. Let $w_{\mathcal{D}}$ be the word obtained from concatenation of words in (II) and (III), in arbitrary order, adding the suffix $\alpha^2\beta\alpha$ at the beginning, and replacing all β^2 by β^3 (more precisely replacing all factors $\alpha\beta^2\alpha$ by $\alpha\beta^3\alpha$; note that the factors $\alpha\beta^2\alpha$ in the word obtained from concatenation of words in (II) and (III) can only come from the concatenation of words in (II-III) ending and starting with β). Then obviously, $w_{\mathcal{D}}$ has all words in (II-III) as factors and has no occurrences of $(12y)$-good exponents. The fact that it has the suffix $\alpha^2\beta\alpha$ will be used later in the proof. We observe that $w_{\mathcal{D}}$ is also 3-full, since all words of length 3 appear as factors in words listed in (II). This means we may state the following.

Proposition 2. *Let $w \in \{\alpha, \beta\}^*$ be a word such that $w^* = w_{\mathcal{D}}w$ has as factors all the three words listed in (I) above. Then, w 3-compresses all proper automata in $\mathcal{D}$ if and only if the word $w^* = w_{\mathcal{D}}w$ described above is 3-collapsing.*

5 Reduction to a System of Transformation Conditions

Now, our aim is to express the problem of 3-compressibility of proper automata in $\mathcal{D}$ in terms of solving a system of transformation conditions. The base for this is Theorem 1. Our attention is restricted to transformations α and β satisfying the following conditions.

(C1) α is a transformation of type $[1, 2]\backslash 1$;

(C2) β is a permutation of the form $\beta = (12y)\ldots$ for some $y \notin \{1, 2\}$;

(C3) either 2α or $y\alpha$ is not in $\{2, y\}$.

Given a word $w \in \{\alpha, \beta\}^+$, by $u_1, u_2, \ldots, u_s$ we denote the set of all factors of w such that $\alpha u_1 \alpha$, $\alpha u_2 \alpha$, $\ldots$, $\alpha u_s \alpha$ are all the factors of w defined in Theorem 1 for the permutation $\beta = (12y)\ldots$. Then we have the following.

Proposition 3. *Let $w \in \{\alpha, \beta\}^+$, and let $u_1, u_2, \ldots, u_s$ be the factors of w described above. Then, there exists a proper automaton $A \in \mathcal{D}$ such that w does not 3-compresses A if and only if the system*

$$1u_1, 1u_2, \ldots, 1u_s \in \{1, 2\} \tag{1}$$

has a solution in transformations α, β on a finite set $Q = \{1, 2, \ldots, n\}$ satisfying the conditions (C1-C3).

Proof. First, suppose that a required solution exists, and let $\mathcal{A}$ be an automaton with the state set $Q = \{1, 2, \ldots, n\}$ and two input letters whose transition function is defined by the action of the letters α and β given by the solution. Then conditions (C1) and (C2) mean simply that $A \in \mathcal{D}$, provided it is proper 3-compressible. It is 3-compressible by Lemma 1, item (iii). Finally, it is not difficult to see that any word 3-compressing A has a length exceeding 4. Indeed, we need first a letter α to get 1 missing in the image, then a factor β^2, to get y missing in the image, and the next α, to get two states missing. Thus $A \in \mathcal{D}$ and, by Theorem 1, w does not 3-compress $\mathcal{A}$.

Conversely, if $A \in \mathcal{D}$ and w does not 3-compresses it, then we may assume that its set of the states is $Q = \{1, 2, \ldots, n\}$. Then, the transformations α and β corresponding to the letters of $\mathcal{A}$ satisfy, by definition of $\mathcal{D}$, the condition (C1) and (C2), and as above, by Lemma 1, they satisfy also condition (C3). Thus, by Theorem 1, they form a required solution of the system (1). □

In such a way the problem of 3-compressibility of automata in $\mathcal{D}$ is reduced to solving a system of transformation conditions $1u \in \{1, 2\}$ with all u of the form $u = \beta^k \alpha u' \alpha \beta^\ell$, where k, ℓ are $(12y)$-good exponents, and u' has no occurrence of $(12y)$-bad exponents. We are going to show that solving a certain subclass of such systems is computationally hard.

Given words $v_1, \ldots, v_s \in \{\alpha, \beta\}^+$, we add to them two further words $v_0 = \alpha^2$ and $v_{s+1} = \alpha\beta^4\alpha$, and consider the following system of transformation conditions.

$$1\beta^2 v_0 \beta^2,\ 1\beta^2 v_1 \beta^2,\ \ldots,\ 1\beta^2 v_{s+1} \beta^2 \in \{1, 2\} \tag{2}$$

We define a specific decision problem:

PROBLEM (*)
INSTANCE: words $v_1, \ldots, v_s \in \{\alpha, \beta\}^+$ such that each word v_i starts and ends with α, and has no occurrence of β with a $(12y)$-good exponent;
QUESTION: Is there a solution (α, β) of the system (2) satisfying conditions (C1-C3)?

We have the following.

Theorem 2. *Problem (*) formulated above is NP-complete.*

The proof ot this theorem uses tools worked out in [6], where solving systems of such conditions is expressed in terms of coloring trees with distinguished nodes. It will be given in the extended version of the paper. Having this theorem we can can easily prove our main result.

Theorem 3. *The problem whether a given word $w \in \{\alpha, \beta\}^*$ is 3-collapsing is co-NP-complete.*

Proof. First observe that the problem belongs to co-NP class. Indeed, to see that a word w is not 3-collapsing a nondeterministic algorithm needs only to guess the smallest automaton that is not 3-compressed by w. By [10, Theorem 1] such an automaton has not more than $4|w| + 2$ states, and the facts that it is 3-compressible and that w does not 3-compress it can be checked easily in polynomial time with respect to $|w|$.

We transform problem (*) to our problem. Let $v_1, \ldots, v_s$ be an instance of (*). First we form the word

$$w' = \beta^2 v_0 \beta^2 v_0 \beta^2 v_1 \beta^2 v_2 \beta^2 \ldots \beta^2 v_s \beta^2 v_{s+1} \beta^2 v_{s+1} \beta^2.$$

Note that this word has doubled occurrences of factors v_0 and v_{s+1}. By assumption, the only occurrences of β in w' with $(12y)$-good exponents are β^2 separating factors $v_0, v_1, v_2, \ldots, v_s, v_{s+1}$. Thus, by Proposition 3, system (2) has a solution satisfying conditions (C1-C3) if and only if there exists a proper automaton $A \in \mathcal{D}$ such that w does not 3-compress A.

Now, we observe that $w = w_{\mathcal{D}} w'$, defined as in Proposition 2, has as factors all the three words listed in (I). Indeed, $\alpha^2\beta\alpha\beta^2\alpha$ is a factor of w since $w_{\mathcal{D}}$, by definition, has the suffix $\alpha^2\beta\alpha$, and w' starts from $\beta^2\alpha$. The prefix $\beta^2 v_0 \beta^2 v_0 \beta^2 = \beta^2\alpha^2\beta^2\alpha^2\beta^2$ of w' has the second word in (I) as a factor. Finally, the suffix $v_{s+1}\beta^2 v_{s+1}\beta^2 = \alpha\beta^4\alpha\beta^2\alpha\beta^4\alpha\beta^2$ of w' has the third word in (I) as a factor. Therefore, by Proposition 2, there exists a proper automaton $A \in \mathcal{D}$ such that w does not 3-compresses A if and only if the word $w^* = w_{\mathcal{D}} w$ is not 3-collapsing. Obviously, this transformation may be performed in polynomial time, which completes the proof. □

References

1. A. Frigeri, A. Cherubini and Z. Liu. Composing short 3-compressing words on a 2 letter alphabet. *to appear,* (arxiv.org: 1406.1413v1, 2014).
2. D. S. Ananichev, A. Cherubini, and M. V. Volkov. Image reducing words and subgroups of free groups. *Theor. Comput. Sci.*, 307(1):77–92, 2003.
3. D. S. Ananichev, A. Cherubini, and M. V. Volkov. An inverse automata algorithm for recognizing 2-collapsing words. In *Developments in Language Theory*, volume 2450 of *LNCS*, pages 270–282, 2003.
4. D. S. Ananichev and I. V. Petrov. Quest for short synchronizing words and short collapsing words. In *WORDS. Proc. 4th Int. Conf.*, pages 411–418, 2003.

5. A. Cherubini, P. Gawrychowski, A. Kisielewicz, and B. Piochi. A combinatorial approach to collapsing words. In *MFCS*, pages 256–266, 2006.
6. A. Cherubini and A. Kisielewicz. Collapsing words, permutation conditions and coherent colorings of trees. *Theor. Comput. Sci.*, 410(21-23):2135–2147, 2009.
7. J. E. Pin. Le problème de la synchronization. contribution à l'étudia de la conjecture de Černý. *Thèse 3e cycle, Paris*, 1978.
8. J. E. Pin. Sur le mots synchronisants dans un automata fini. *Elektron. Informationverarbeigtung und Kybernetik*, 14:283–289, 1978.
9. S. W. Margolis, J.-E. Pin, and M. V. Volkov. Words guaranteeing minimum image. *Internat. J. Foundations Comp. Sci.*, 15:259–276, 2004.
10. I. V. Petrov. An algorithm for recognition of n-collapsing words. *Theoret. Comput. Sci.*, 391(1-2):99–108, 2008.
11. E. V. Pribavkina. On some properties of the language of 2-collapsing words. In *Developments in Language Theory*, volume 3572 of *LNCS*, pages 374–384, 2005.
12. N. Sauer and M. G. Stone. Composing functions to reduce image size. *Ars Combinatoria*, 1:171–176, 1991.