BINARY COMPLETELY REACHABLE AUTOMATA

DAVID CASAS AND MIKHAIL V. VOLKOV

ABSTRACT. A deterministic finite automaton in which every non-empty set of states occurs as the image of the whole state set under the action of a suitable input word is called completely reachable. We study completely reachable automata with two input letters.

1. INTRODUCTION

Completely reachable automata are complete deterministic finite automata in which every non-empty subset of the state set occurs as the image of the whole state set under the action of a suitable input word. Such automata appeared in the study of descriptional complexity of formal languages [2, 10] and in relation to the Černý conjecture [4]. A systematic study of completely reachable automata was initiated in [2, 3] and continued in [1]. In [1, 3] completely reachable automata were characterized in terms of a certain finite sequence of directed graphs (digraphs): the automaton is completely reachable if and only if the final digraph in this sequence is strongly connected. In [1, Theorem 11] it was shown that given an automaton $\mathscr{A}$ with n states and m input letters, the k-th digraph in the sequence assigned to $\mathscr{A}$ can be constructed in $O(mn^{2k}\log n)$ time. However, this does not yet ensure a polynomial-time algorithm for recognizing complete reachability: a series of examples in [1] demonstrates that the length of the digraph sequence for an automaton with n states may reach $n-1$.

Here we study completely reachable automata with two input letters; for brevity, we call automata with two input letters *binary*. Our main results provide a new characterization of binary completely reachable automata, and the characterization leads to a quasilinear time algorithm for recognizing complete reachability for binary automata.

Our prerequisites are minimal: we only assume the reader's acquaintance with basic properties of strongly connected digraphs, subgroups, and cosets.

2. PRELIMINARIES

A *complete deterministic finite automaton* (DFA) is a triple $\mathscr{A} = \langle Q, \Sigma, \delta \rangle$ where Q and Σ are finite sets called the *state set* and, resp., the *input alphabet* of $\mathscr{A}$, and $\delta \colon Q \times \Sigma \to Q$ is a totally defined map called the *transition function* of $\mathscr{A}$.

The elements of Σ are called *input letters* and finite sequences of letters are called *words over* Σ. The empty sequence is also treated as a word, called the *empty word* and denoted ε. The collection of all words over Σ is denoted Σ^*.

INSTITUTE OF NATURAL SCIENCES AND MATHEMATICS, URAL FEDERAL UNIVERSITY, 620000 EKATERINBURG, RUSSIA, DAFECATO4@GMAIL.COM, M.V.VOLKOV@URFU.RU

Key words and phrases. Deterministic finite automaton; Complete reachability; Strongly connected graph; Tree.

The authors were supported by the Ministry of Science and Higher Education of the Russian Federation, project FEUZ-2020-0016.

The transition function δ extends to a function $Q \times \Sigma^* \to Q$ (still denoted by δ) via the following recursion: for every $q \in Q$, we set $\delta(q,\varepsilon) = q$ and $\delta(q,wa) = \delta(\delta(q,w),a)$ for all $w \in \Sigma^*$ and $a \in \Sigma$. Thus, every word $w \in \Sigma^*$ induces the transformation $q \mapsto \delta(q,w)$ of the set Q. The set $T(\mathscr{A})$ of all transformations induced this way is called the *transition monoid* of $\mathscr{A}$; this is the submonoid generated by the transformations $q \mapsto \delta(q,a)$, $a \in \Sigma$, in the monoid of all transformations of Q. A DFA $\mathscr{B} = \langle Q, \Theta, \zeta \rangle$ with the same state set as $\mathscr{A}$ is said to be *syntactically equivalent* to $\mathscr{A}$ if $T(\mathscr{B}) = T(\mathscr{A})$.

The function δ can be further extended to non-empty subsets of the set Q. Namely, for every non-empty subset $P \subseteq Q$ and every word $w \in \Sigma^*$, we let $\delta(P,w) = \{\delta(q,w) \mid q \in P\}$.

Whenever there is no risk of confusion, we tend to simplify our notation by suppressing the sign of the transition function; this means that we write $q\,.\,w$ for $\delta(q,w)$ and $P\,.\,w$ for $\delta(P,w)$ and specify a DFA as a pair $\langle Q, \Sigma \rangle$.

We say that a non-empty subset $P \subseteq Q$ is *reachable* in $\mathscr{A} = \langle Q, \Sigma \rangle$ if $P = Q\,.\,w$ for some word $w \in \Sigma^*$. A DFA is called *completely reachable* if every non-empty subset of its state set is reachable. Observe that complete reachability is actually a property of the transition monoid of $\mathscr{A}$; hence, if a DFA $\mathscr{A}$ is completely reachable, so is any DFA that is syntactically equivalent to $\mathscr{A}$.

Given a DFA $\mathscr{A} = \langle Q, \Sigma \rangle$ and a word $w \in \Sigma^*$, the *image* of w is the set $Q\,.\,w$ and the *excluded set* $\mathrm{excl}(w)$ of w is the complement $Q \backslash Q\,.\,w$ of the image. The number $|\mathrm{excl}(w)|$ is called the *defect* of w. If a word w has defect 1, its excluded set consists of a unique state called the *excluded state* for w. Further, for any $w \in \Sigma^*$, the set $\{p \in Q \mid p = q_1\,.\,w = q_2\,.\,w \text{ for some } q_1 \neq q_2\}$ is called the *duplicate set* of w and is denoted by $\mathrm{dupl}(w)$. If w has defect 1, its duplicate set consists of a unique state called the *duplicate state* for w. We identify singleton sets with their elements, and therefore, for a word w of defect 1, $\mathrm{excl}(w)$ and $\mathrm{dupl}(w)$ stand for its excluded and, resp., duplicate states.

For any $v \in \Sigma^*$, $q \in Q$, let $qv^{-1} = \{p \in Q \mid p\,.\,v = q\}$. Then for all $u, v \in \Sigma^*$,

$$\mathrm{excl}(uv) = \{q \in Q \mid qv^{-1} \subseteq \mathrm{excl}(u)\}, \tag{1}$$

$$\mathrm{dupl}(uv) = \{q \in Q \mid qv^{-1} \cap \mathrm{dupl}(u) \neq \varnothing \text{ or } |qv^{-1} \backslash \mathrm{excl}(u)| \geq 2\}. \tag{2}$$

The equalities (1) and (2) become clear as soon as the definitions of excl( ) and dupl( ) are deciphered. . Fig. 1 provides a supporting illustration.

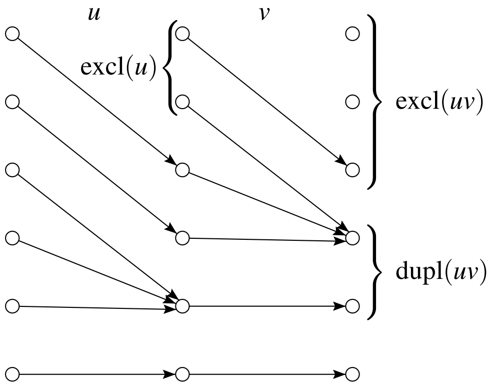

FIGURE 1. An illustration for the equalities (1) and (2)

Recall that DFAs with two input letters are called *binary*. The question of our study is: under which conditions is a binary DFA completely reachable? The rest of the section presents a series of reductions showing that to answer this question, it suffices to analyze DFAs of a specific form.

Let $\mathscr{A} = \langle Q, \{a,b\}\rangle$ be a binary DFA with $n > 1$ states. If neither a nor b has defect 1, no subset of size $n-1$ is reachable in $\mathscr{A}$. Therefore, when looking for binary completely reachable automata, we must focus on DFAs possessing a letter of defect 1. We will always assume that a has defect 1.

The image of every non-empty word over $\{a,b\}$ is contained in either $Q\,.\,a$ or $Q\,.\,b$. If the defect of b is greater than or equal to 1, then at most two subsets of size $n-1$ are reachable (namely, $Q\,.\,a$ and $Q\,.\,b$), whence $\mathscr{A}$ can only be completely reachable provided that $n = 2$. The automaton $\mathscr{A}$ is then nothing but the classical flip-flop, see Fig. 2.

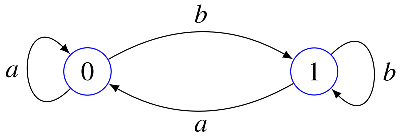

FIGURE 2. The flip-flop. Here and below a DFA $\langle Q,\Sigma\rangle$ is depicted as a digraph with the vertex set Q and a labeled edge $q \xrightarrow{a} q'$ for each triple $(q,a,q') \in Q \times \Sigma \times Q$ such that $q\,.\,a = q'$.

Having isolated this exception, we assume from now on that $n \geq 2$ and the letter b has defect 0, which means that b acts as a permutation of Q. The following fact was first stated in [2]; for a proof, see, e.g., [1, Sect. 6].

Lemma 1. *If $\mathscr{A} = \langle Q, \{a,b\}\rangle$ is a completely reachable automaton in which the letter b acts as a permutation of Q, then b acts as a cyclic permutation.*

Taking Lemma 1 into account, we restrict our further considerations to DFAs with $n \geq 2$ states and two input letters a and b such that a has defect 1 and b acts a cyclic permutation. Without any loss, we will additionally assume that these DFAs have the set $\mathbb{Z}_n = \{0,1,\dots,n-1\}$ of all residues modulo n as their state set and the action of b at any state merely adds 1 modulo n. Let us also agree that whenever we deal with elements of $\mathbb{Z}_n$, the signs $+$ and $-$ mean addition and subtraction modulo n, unless the contrary is explicitly specified.

Further, we will assume that $0 = \mathrm{excl}(a)$ as it does not matter from which origin the cyclic count of the states start.

Since b is a permutation, for each $k \in \mathbb{Z}_n$, the transformations $q \mapsto q\,.\,b^k a$ and $q \mapsto q\,.\,b$ generate the same submonoid in the monoid of all transformations of $\mathbb{Z}_n$ as do the transformations $q \mapsto q\,.\,a$ and $q \mapsto q\,.\,b$. This means that if one treats the word $b^k a$ as a new letter a_k, say, one gets the DFA $\mathscr{A}_k = \langle \mathbb{Z}_n, \{a_k, b\}\rangle$ that is syntactically equivalent to $\mathscr{A}$. Therefore, $\mathscr{A}$ is completely reachable if and only if so is $\mathscr{A}_k$ for some (and hence for all) k. Hence we may choose k as we wish and study the DFA $\mathscr{A}_k$ for the specified value of k instead of $\mathscr{A}$.

What can we achieve using this? From (1) we have $\mathrm{excl}(b^k a) = \mathrm{excl}(a) = 0$. Further, let $q_1 \neq q_2$ be such that $q_1\,.\,a = q_2\,.\,a = \mathrm{dupl}(a)$. Choosing $k = q_1$ (or $k = q_2$), we get $0\,.\,b^k a = \mathrm{dupl}(a)$. Thus, we will assume that $0\,.\,a = \mathrm{dupl}(a)$.

Summarizing, we will consider DFAs $\langle \mathbb{Z}_n, \{a,b\}\rangle$ such that:

- the letter a has defect 1, $\operatorname{excl}(a)=0$, and $0\,.\,a=\operatorname{dupl}(a)$;
- $q\,.\,b=q+1$ for each $q\in\mathbb{Z}_n$.

We call such DFAs *standardized*. For the purpose of complexity considerations at the end of Sect. 5, observe that given a binary DFA $\mathscr{A}$ in which one letter acts as a cyclic permutation while the other has defect 1, one can 'standardize' the automaton, that is, construct a standardized DFA syntactically equivalent to $\mathscr{A}$, in linear time with respect to the size of $\mathscr{A}$.

3. A NECESSARY CONDITION

Let $\langle\mathbb{Z}_n,\{a,b\}\rangle$ be a standardized DFA and $w\in\{a,b\}^*$. A subset $S\subseteq\mathbb{Z}_n$ is said to be *w-invariant* if $S\,.\,w\subseteq S$.

Proposition 1. *If $\langle\mathbb{Z}_n,\{a,b\}\rangle$ is a completely reachable standardized DFA, then no proper subgroup of $(\mathbb{Z}_n,+)$ is a-invariant.*

Proof. Arguing by contradiction, assume that $H\subsetneqq\mathbb{Z}_n$ is a subgroup such that $H\,.\,a\subseteq H$. Let d stand for the index of the subgroup H in the group $(\mathbb{Z}_n,+)$. The set $\mathbb{Z}_n$ is then partitioned into the d cosets

$$H_0=H,\ H_1=H\,.\,b=H+1,\ \dots,\ H_{d-1}=H\,.\,b^{d-1}=H+d-1.$$

For $i=0,1,\dots,d-1$, let T_i be the complement of the coset H_i in $\mathbb{Z}_n$. Then we have $T_i=\cup_{j\neq i}H_j$ and $T_i\,.\,b=T_{i+1\ (\mathrm{mod}\ d)}$ for each $i=0,1,\dots,d-1$.

Since $\mathscr{A}$ is completely reachable, each subset T_i is reachable. Take a word w of minimum length among words with the image equal to one of the subsets $T_0,T_1,\dots,T_{d-1}$. Write w as $w=w'c$ for some letter $c\in\{a,b\}$.

If $c=b$, then for some $i\in\{0,1,\dots,d-1\}$, we have

$$\mathbb{Z}_n\,.\,w'b=T_i=T_{i-1\ (\mathrm{mod}\ d)}\,.\,b.$$

Since b^n acts as the identity mapping, applying the word b^{n-1} to this equality yields $\mathbb{Z}_n\,.\,w'=T_{i-1\ (\mathrm{mod}\ d)}$ whence the image of w' is also equal to one of the subsets $T_0,T_1,\dots,T_{d-1}$. This contradicts the choice of w.

Thus, $c=a$, whence the set $\mathbb{Z}_n\,.\,w$ is contained in $\mathbb{Z}_n\,.\,a$. The only T_i that is contained in $\mathbb{Z}_n\,.\,a$ is T_0 because each T_i with $i\neq0$ contains H_0, and $H_0=H$ contains 0, the excluded state of a. Hence, $\mathbb{Z}_n\,.\,w=T_0$, that is, $\mathbb{Z}_n\,.\,w'a=T_0$. For each state $q\in\mathbb{Z}_n\,.\,w'$, we have $q\,.\,a\in T_0$, and this implies $q\in T_0$ since H_0, the complement of T_0, is a-invariant. We see that $\mathbb{Z}_n\,.\,w'\subseteq T_0$ and the inclusion cannot be strict because T_0 cannot be the image of its proper subset. However, the equality $\mathbb{Z}_n\,.\,w'=T_0$ again contradicts the choice of w. □

We will show that the condition of Proposition 1 is not only necessary but also sufficient for complete reachability of a standardized DFA. The proof of sufficiency requires a construction that we present in full in Sect. 5, after studying its simplest case in Sect. 4.

4. RYSTSOV'S GRAPH OF A BINARY DFA

Recall a sufficient condition for complete reachability from [2]. Given a (not necessarily binary) DFA $\mathscr{A}=\langle Q,\Sigma\rangle$, let $W_1(\mathscr{A})$ stand for the set of all words in Σ^* that have defect 1 in $\mathscr{A}$. Consider a digraph with the vertex set Q and the edge set

$$E=\{(\operatorname{excl}(w),\operatorname{dupl}(w))\mid w\in W_1(\mathscr{A})\}.$$

We denote this digraph by $\Gamma_1(\mathscr{A})$. The notation comes from [2], but much earlier, though in a less explicit form, the construction was used by Rystsov [11] for some special species of DFAs. Taking this into account, we refer to $\Gamma_1(\mathscr{A})$ as the *Rystsov graph* of $\mathscr{A}$.

Theorem 1 ([2, Theorem 1])**.** *If a DFA $\mathscr{A} = \langle Q, \Sigma\rangle$ is such that the graph $\Gamma_1(\mathscr{A})$ is strongly connected, then $\mathscr{A}$ is completely reachable.*

It was shown in [2] that the condition of Theorem 1 is not necessary for complete reachability, but it was conjectured that the condition might characterize binary completely reachable automata. However, this conjecture has been refuted in [1, Example 2] by exhibiting a binary completely reachable automaton with 12 states whose Rystsov graph is not strongly connected. Here we include a similar example which we will use to illustrate some of our results.

Consider the standardized DFA $\mathscr{E}'_{12} = \langle \mathbb{Z}_{12}, \{a, b\}\rangle$ where the action of the letter a is specified as follows:

q	0	1	2	3	4	5	6	7	8	9	10	11
$q\,.\,a$	10	1	2	8	4	5	10	9	3	7	6	11

.

(The DFA $\mathscr{E}'_{12}$ only slightly differs from the DFA $\mathscr{E}_{12}$ used in [1, Example 2], hence the notation.) The DFA $\mathscr{E}'_{12}$ is shown in Fig. 3, in which we have replaced edges that should have been labeled a and b with solid and, resp., dashed edges.

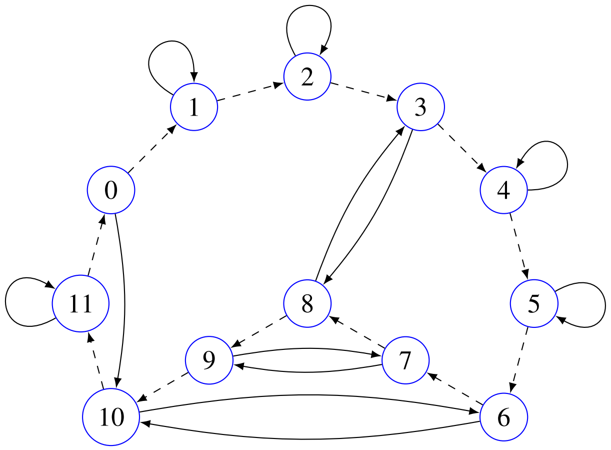

FIGURE 3. The DFA $\mathscr{E}'_{12}$; solid and dashed edges show the action of a and, resp., b

We postpone the description of the digraph $\Gamma_1(\mathscr{E}'_{12})$ and the proof that the DFA $\mathscr{E}'_{12}$ is completely reachable until we develop suitable tools that make the description and the proof easy.

We start with a characterization of Rystsov's graphs of standardized DFAs. Let $\mathscr{A} = \langle \mathbb{Z}_n, \{a, b\}\rangle$ be such a DFA. It readily follows from (1) and (2) that $\mathrm{excl}(w)\,.\,b = \mathrm{excl}(wb)$ and $\mathrm{dupl}(w)\,.\,b = \mathrm{dupl}(wb)$ for every word $w \in W_1(\mathscr{A})$. Therefore, the edge set E of the digraph $\Gamma_1(\mathscr{A})$ is closed under the *translation* $(q, p) \mapsto (q\,.\,b, p\,.\,b) = (q+1, p+1)$. As a consequence, for any edge $(q, p) \in E$ and any k, the pair $(q+k, p+k)$ also constitutes an edge in E.

Denote by $D_1(\mathscr{A})$ the set of ends of edges of $\Gamma_1(\mathscr{A})$ that start at 0, that is, $D_1(\mathscr{A}) = \{p \in \mathbb{Z}_n \mid (0,p) \in E\}$. We call $D_1(\mathscr{A})$ the *difference set* of $\mathscr{A}$. Our first observation shows how to recover all edges of $\Gamma_1(\mathscr{A})$, knowing $D_1(\mathscr{A})$.

Lemma 2. *Let $\mathscr{A} = \langle \mathbb{Z}_n, \{a,b\}\rangle$ be a standardized DFA. A pair $(q,p) \in \mathbb{Z}_n \times \mathbb{Z}_n$ forms an edge in the digraph $\Gamma_1(\mathscr{A})$ if and only if $p-q \in D_1(\mathscr{A})$.*

Proof. If $p-q \in D_1(\mathscr{A})$, the pair $(0,p-q)$ is an edge in E, and therefore, so is the pair $(0+q,(p-q)+q) = (q,p)$. Conversely, if (q,p) is an edge in E, then so is $(q+(n-q),p+(n-q)) = (0,p-q)$, whence $p-q \in D_1(\mathscr{A})$. □

By Lemma 2, the presence or absence of an edge in $\Gamma_1(\mathscr{A})$ depends only on the difference modulo n of two vertex numbers. This means that $\Gamma_1(\mathscr{A})$ is a *circulant* digraph, that is, the Cayley digraph of the cyclic group $(\mathbb{Z}_n,+)$ with respect to some subset of $\mathbb{Z}_n$. Recall that if D is a subset in a group G, the *Cayley digraph of G with respect to D*, denoted $\mathrm{Cay}(G,D)$, has G as its vertex set and $\{(g,gd) \mid g \in G,\ d \in D\}$ as its edge set. The following property of Cayley digraphs of finite groups is folklore[1].

Lemma 3. *Let G be a finite group, D a subset of G, and H the subgroup of G generated by D. The strongly connected components of the Cayley digraph $\mathrm{Cay}(G,D)$ have the right cosets Hg, $g \in G$, as their vertex sets, and each strongly connected component is isomorphic to $\mathrm{Cay}(H,D)$. In particular, the digraph $\mathrm{Cay}(G,D)$ is strongly connected if and only if G is generated by D.*

Let $H_1(\mathscr{A})$ stand for the subgroup of the group $(\mathbb{Z}_n,+)$ generated by the difference set $D_1(\mathscr{A})$. Specializing Lemma 3, we get the following description for Rystsov's graphs of standardized DFAs.

Proposition 2. *Let $\mathscr{A} = \langle \mathbb{Z}_n, \{a,b\}\rangle$ be a standardized DFA. The digraph $\Gamma_1(\mathscr{A})$ is isomorphic to the Cayley digraph $\mathrm{Cay}(\mathbb{Z}_n, D_1(\mathscr{A}))$. The strongly connected components of $\Gamma_1(\mathscr{A})$ have the cosets of the subgroup $H_1(\mathscr{A})$ as their vertex sets, and each strongly connected component is isomorphic to the Cayley digraph $\mathrm{Cay}(H_1(\mathscr{A}), D_1(\mathscr{A}))$. In particular, the digraph $\Gamma_1(\mathscr{A})$ is strongly connected if and only if the set $D_1(\mathscr{A})$ generates $(\mathbb{Z}_n,+)$ or, equivalently, if and only if the greatest common divisor of $D_1(\mathscr{A})$ is coprime to n.*

Proposition 2 shows that structure of the Rystsov graph of a standardized DFA $\mathscr{A}$ crucially depends on its difference set $D_1(\mathscr{A})$. The definition of the edge set of $\Gamma_1(\mathscr{A})$ describes $D_1(\mathscr{A})$ as the set of duplicate states for all words w of defect 1 whose excluded state is 0, that is, $D_1(\mathscr{A}) = \{\mathrm{dupl}(w) \mid \mathrm{excl}(w) = 0\}$. Thus, understanding of difference sets amounts to a classification of transformations caused by words of defect 1. It is such a classification that is behind the following handy description of difference sets.

Proposition 3. *Let $\mathscr{A} = \langle \mathbb{Z}_n, \{a,b\}\rangle$ be a standardized DFA. Let $r \neq 0$ be such that $r\,.\,a = \mathrm{dupl}(a)$. Then*

$$D_1(\mathscr{A}) = \{\mathrm{dupl}(a)\,.\,v \mid v \in \{a, b^r a\}^*\}. \tag{3}$$

Proof. Denote by N the image of the letter a, that is, $N = \mathbb{Z}_n \backslash \{0\}$. If $q\,.\,a = p$ for some $q \in \mathbb{Z}_n$ and $p \in N$, then, clearly, $(q-r)\,.\,b^r a = p$. Hence the only state in N that has a preimage of size 2 under the actions of both a and $b^r a$ is

[1] In fact, our definition is the semigroup version of the notion of a Cayley digraph, but this makes no difference since in a finite group, every subsemigroup is a subgroup.

$$\operatorname{dupl}(a) = \begin{cases} 0\,.\,a = r\,.\,a, \\ (n-r)\,.\,b^r a = 0\,.\,b^r a, \end{cases}$$

and in both cases 0 belongs to the preimage. Thus, the preimage of every $p \in N$ under both a and $b^r a$ contains a unique state in N, which means that both a and $b^r a$ act on the set N as permutations. Hence every word $v \in \{a, b^r a\}^*$ acts on N as a permutation. Then the word av has defect 1 and $\operatorname{excl}(av) = 0$. Applying the equality (2) with a in the role of u, we derive that $\operatorname{dupl}(av) = \operatorname{dupl}(a)\,.\,v$. Thus, denoting the right-hand side of (3) by D, we see that every state in D is the duplicate state of some word whose only excluded state is 0. This means that $D_1(\mathscr{A}) \supseteq D$.

To verify the converse inclusion, take an arbitrary state $p \in D_1(\mathscr{A})$ and let w be a word of defect 1 such that $\operatorname{excl}(w) = 0$ and $\operatorname{dupl}(w) = p$. Since $\operatorname{excl}(w) = 0$, the word w ends with the letter a. We prove that p lies in D by induction on the number of occurrences of a in w. If a occurs in w once, then $w = b^k a$ for some $k \in \mathbb{Z}_n$. We have $p = \operatorname{dupl}(w) = \operatorname{dupl}(b^k a) = \operatorname{dupl}(a) \in D$.

If a occurs in w at least twice, write $w = w' b^k a$ where w' ends with a. Then the word w' has defect 1 and $\operatorname{excl}(w') = 0$. As w' has fewer occurrences of a, the inductive assumption applies and yields $\operatorname{dupl}(w') \in D$. Denoting $\operatorname{dupl}(w')$ by p', we have $p = p'\,.\,b^k a$. If we prove that $k \in \{0, r\}$, we are done since the set D is both a-invariant and $b^r a$-invariant by its definition. Arguing by contradiction, assume $k \notin \{0, r\}$. Let $\ell = k\,.\,a$; then k is the only state in ℓa^{-1}. Hence $\ell a^{-1} = \operatorname{excl}(w' b^k)$, and the equality (1) (with $u = w' b^k$ and $v = a$) shows that $\ell \in \operatorname{excl}(w' b^k a) = \operatorname{excl}(w)$. Clearly, $\ell \neq 0$ as ℓ lies in the image of a. Therefore the conclusion $\ell \in \operatorname{excl}(w)$ contradicts the assumption $\operatorname{excl}(w) = 0$. □

For an illustration, we apply (3) to compute the difference set for the DFA $\mathscr{E}'_{12}$ shown in Fig. 3. In $\mathscr{E}'_{12}$, we have $r = 6$ and $\operatorname{dupl}(a) = 10$. Acting by a and $b^6 a$ gives $10\,.\,a = 6$ and $10\,.\,b^6 a = (10+6)\,.\,a = 4\,.\,a = 4$. Thus, $4, 6 \in D_1(\mathscr{E}'_{12})$. Acting by a or $b^6 a$ at 4 and 6 does not produce anything new: $4\,.\,a = 4$ and $4\,.\,b^6 a = (4+6)\,.\,a = 10\,.\,a = 6$ while $6\,.\,a = 10$ and $6\,.\,b^6 a = (6+6)\,.\,a = 0\,.\,a = 10$. We conclude that $D_1(\mathscr{E}'_{12}) = \{4, 6, 10\}$. Since 2, the greatest common divisor of $\{4, 6, 10\}$, divides 12, we see that the digraph $\Gamma_1(\mathscr{E}'_{12})$ is not strongly connected. The subgroup $H_1(\mathscr{E}'_{12})$ consists of even residues modulo 12 and has index 2. Hence the digraph $\Gamma_1(\mathscr{E}'_{12})$ has two strongly connected components whose vertex sets are $\{0, 2, 4, 6, 8, 10\}$ and $\{1, 3, 5, 7, 9, 11\}$, and for each $q \in \mathbb{Z}_{12}$, it has the edges $(q, q+4)$, $(q, q+6)$, and $(q, q+10)$.

In fact, formula (3) leads to a straightforward algorithm that computes the difference set of any standardized DFA $\mathscr{A}$ in time linear in n. This, together with Proposition 2, gives an efficient way to compute the Rystsov graph of $\mathscr{A}$.

Let $D_1^0(\mathscr{A}) = D_1(\mathscr{A}) \cup \{0\}$. It turns out that $D_1^0(\mathscr{A})$ is always a union of cosets of a nontrivial subgroup.

Proposition 4. *Let $\mathscr{A} = \langle \mathbb{Z}_n, \{a, b\} \rangle$ be a standardized DFA. Let $r \neq 0$ be such that $r\,.\,a =$* $\operatorname{dupl}(a)$. *Then the set $D_1^0(\mathscr{A})$ is a union of cosets of the subgroup generated by r in the group $H_1(\mathscr{A})$.*

Proof. It is easy to see that the claim is equivalent to the following implication: if $d \in D_1^0(\mathscr{A})$, then $d + r \in D_1^0(\mathscr{A})$. This clearly holds if $d + r = 0$. Thus, assume that $d \in D_1^0(\mathscr{A})$ is such that $d + r \neq 0$. Then $(d+r)\,.\,a \in D_1(\mathscr{A})$. Indeed, if $d = 0$, then $(d+r)\,.\,a = r\,.\,a = \operatorname{dupl}(a) \in D_1(\mathscr{A})$. If $d \neq 0$, then $d \in D_1(\mathscr{A})$, whence $(d+r)\,.\,a = d\,.\,b^r a \in D_1(\mathscr{A})$ as formula (3) ensures that the set $D_1(\mathscr{A})$ is closed under the action of the word $b^r a$.

We have observed in the first paragraph of the proof of Proposition 3 that a acts on the set $N = \mathbb{Z}_n \backslash \{0\}$ as a permutation. Hence for some k, the word a^k acts on N as the identity map. Then $d+r = (d+r) \,.\, a^k = ((d+r) \,.\, a) \,.\, a^{k-1} \in D_1(\mathscr{A})$ since we have already shown that $(d+r) \,.\, a \in D_1(\mathscr{A})$ and formula (3) ensures that the set $D_1(\mathscr{A})$ is a-invariant. □

In our running example $\mathscr{E}'_{12}$, $r = 6$ and the set $D^0_1(\mathscr{E}'_{12}) = \{0,4,6,10\}$ is the union of the subgroup $\{0,6\}$ with its coset $\{4,10\}$ in the group $H_1(\mathscr{E}'_{12})$.

Let $\mathscr{A} = \langle \mathbb{Z}_n, \{a,b\} \rangle$ be a standardized DFA. Proposition 4 shows that then the set $D^0_1(\mathscr{A})$ is situated between the subgroup $H_1(\mathscr{A})$ and the subgroup R generated by $r \neq 0$ such that $r \,.\, a = \mathrm{dupl}(a)$:

$$R \subseteq D^0_1(\mathscr{A}) \subseteq H_1(\mathscr{A}). \tag{4}$$

Formula (3) implies that the difference set $D_1(\mathscr{A})$ is a-invariant, and so is the set $D^0_1(\mathscr{A})$ since $0 \,.\, a = \mathrm{dupl}(a) \in D_1(\mathscr{A})$. By Proposition 1, if the automaton $\mathscr{A}$ is completely reachable, then either $H_1(\mathscr{A}) = \mathbb{Z}_n$ or $H_1(\mathscr{A})$ is a proper subgroup and both inclusions in (4) are strict. Recall that by Proposition 2 $H_1(\mathscr{A}) = \mathbb{Z}_n$ if and only if the digraph $\Gamma_1(\mathscr{A})$ is strongly connected. In the other case, n must be a product of at least three (not necessarily distinct) prime numbers. Indeed, the subgroups of $(\mathbb{Z}_n, +)$ ordered by inclusion are in a 1-1 correspondence to the divisors of n ordered by division, and no product of only two primes can have two different proper divisors d_1 and d_2 such that d_1 divides d_2. We thus arrive at the following conclusion.

Corollary 1. *A binary DFA $\mathscr{A}$ with n states where n is a product of two prime numbers is completely reachable if and only if one of its letters acts as a cyclic permutation of the state set, the other letter has defect* 1, *and the digraph $\Gamma_1(\mathscr{A})$ is strongly connected.*

Corollary 1 allows one to show that the number of states in a binary completely reachable automata whose Rystsov graph is not strongly connected is at least 12. (Thus, our examples of such automata ($\mathscr{E}_{12}$ from [1, Example 2] and $\mathscr{E}'_{12}$ from the present paper) are of minimum possible size.) Indeed, Corollary 1 excludes all sizes less than 12 except 8. If a standardized DFA $\mathscr{A}$ has 8 states and the digraph $\Gamma_1(\mathscr{A})$ is not strongly connected, then the group $H_1(\mathscr{A})$ has size at most 4 and its subgroup R generated by the non-zero state in $\mathrm{dupl}(a)a^{-1}$ has size at least 2. By Proposition 4 the set $D^0_1(\mathscr{A})$ is a union of cosets of the subgroup R in the group $H_1(\mathscr{A})$, whence either $D_0(\mathscr{A}) = R$ or $D_0(\mathscr{A}) = H_1(\mathscr{A})$. In either case, we get a proper a-invariant subgroup, and Proposition 1 implies that the DFA $\mathscr{A}$ is not completely reachable.

5. Subgroup sequences for standardized DFAs

In [1,3] Theorem 1 is generalized in the following way. A sequence of digraphs $\Gamma_1(\mathscr{A})$, $\Gamma_2(\mathscr{A}), \dots, \Gamma_k(\mathscr{A}), \dots$ is assigned to an arbitrary (not necessarily binary) DFA $\mathscr{A}$, where $\Gamma_1(\mathscr{A})$ is the Rystsov graph of $\mathscr{A}$ while the 'higher level' digraphs $\Gamma_2(\mathscr{A}), \dots, \Gamma_k(\mathscr{A}), \dots$ are defined via words that have defect $2, \dots, k, \dots$ in $\mathscr{A}$. (We refer the interested reader to [1,3] for the precise definitions; here we do not need them.) The length of the sequence is less than the number of states of $\mathscr{A}$, and $\mathscr{A}$ is completely reachable if and only if the final digraph in the sequence is strongly connected.

For the case when $\mathscr{A}$ is a standardized DFA, Proposition 2 shows that the Rystsov graph $\Gamma_1(\mathscr{A})$ is completely determined by the difference set $D_1(\mathscr{A})$ and the subgroup $H_1(\mathscr{A})$ that $D_1(\mathscr{A})$ generates. This suggests that for binary automata, one may substitute the 'higher level' digraphs of [1,3] by suitably chosen 'higher level' difference sets and their generated subgroups.

Take a standardized DFA $\mathscr{A} = \langle \mathbb{Z}_n, \{a,b\}\rangle$ and for each $k > 1$, inductively define the set $D_k(\mathscr{A})$ and the subgroup $H_k(\mathscr{A})$:

$$\begin{aligned} D_k(\mathscr{A}) = \{p \in \mathbb{Z}_n \mid p \in \operatorname{dupl}(w) \text{ for some } w \in \{a,b\}^*\\ \text{such that } 0 \in \operatorname{excl}(w) \subseteq H_{k-1}(\mathscr{A}),\ |\operatorname{excl}(w)| \le k\}, \qquad (5)\\ H_k(\mathscr{A}) \text{ is the subgroup of } (\mathbb{Z}_n,+) \text{ generated by } D_k(\mathscr{A}). \end{aligned}$$

Observe that if we let $H_0(\mathscr{A}) = \{0\}$, the definition (5) makes sense also for $k = 1$ and leads to exactly the same $D_1(\mathscr{A})$ and $H_1(\mathscr{A})$ as defined in Sect. 4.

Using the definition (5), it is easy to prove by induction that $D_k(\mathscr{A}) \subseteq D_{k+1}(\mathscr{A})$ and $H_k(\mathscr{A}) \subseteq H_{k+1}(\mathscr{A})$ for all k.

Proposition 5. *If $\mathscr{A} = \langle \mathbb{Z}_n, \{a,b\}\rangle$ is a standardized DFA and $H_\ell(\mathscr{A}) = \mathbb{Z}_n$ for some ℓ, then $\mathscr{A}$ is a completely reachable automaton.*

Proof. As $\mathscr{A}$ is fixed, we write D_k and H_k instead of $D_k(\mathscr{A})$ and, resp., $H_k(\mathscr{A})$.

Take any non-empty subset $S \subseteq \mathbb{Z}_n$. We prove that S is reachable in $\mathscr{A}$ by induction on $n - |S|$. If $n - |S| = 0$, there is nothing to prove as $S = \mathbb{Z}_n$ is reachable via the empty word. Now let S be a proper subset of $\mathbb{Z}_n$. We aim to find a subset $T \subseteq \mathbb{Z}_n$ such that $S = T\,.\,v$ for some word $v \in \{a,b\}^*$ and $|T| > |S|$. Since $n - |T| < n - |S|$, the induction assumption applies to the subset T whence $T = \mathbb{Z}_n\,.\,u$ for some word $u \in \{a,b\}^*$. Then $S = \mathbb{Z}_n\,.\,uv$ is reachable as required.

Thus, fix a non-empty subset $S \subsetneqq \mathbb{Z}_n$. Since cosets of the trivial subgroup H_0 are singletons, S is a union of cosets of H_0. On the other hand, since $H_\ell = \mathbb{Z}_n$, the only coset of H_ℓ strictly contains S, and so S is not a union of cosets of H_ℓ. Now choose $k \ge 1$ to be the maximal number for which S is a union of cosets of the subgroup H_{k-1}. The subgroup H_k already has a coset, say, $H_k + t$ being neither contained in S nor disjoint with S; in other words, $\varnothing \ne S \cap (H_k + t) \subsetneqq H_k + t$.

By Lemma 3, the coset $H_k + t$ serves as the vertex set of a strongly connected component of the Cayley digraph $\operatorname{Cay}(\mathbb{Z}_n, D_k)$. Therefore, some edge of $\operatorname{Cay}(\mathbb{Z}_n, D_k)$ connects $(H_k + t) \setminus S$ with $S \cap (H_k + t)$ in this strongly connected component, that is, the head q of this edge lies in $(H_k + t) \setminus S$ while its tail p belongs to $S \cap (H_k + t)$. Let $p' = p - q$; then $p' \in D_k$ by the definition of the Cayley digraph. By (5) there exists a word $w \in \{a,b\}^*$ such that $p' \in \operatorname{dupl}(w)$ and $\operatorname{excl}(w) \subseteq H_{k-1}$. Then $p = p' + q = p'\,.\,b^q \in \operatorname{dupl}(w)\,.\,b^q = \operatorname{dupl}(wb^q)$ and $\operatorname{excl}(wb^q) = \operatorname{excl}(w)\,.\,b^q = \operatorname{excl}(w) + q \subseteq H_{k-1} + q$. From $p \in \operatorname{dupl}(wb^q)$ we conclude that there exist $p_1, p_2 \in \mathbb{Z}_n$ such that $p = p_1\,.\,wb^q = p_2\,.\,wb^q$. Since S is a union of cosets of the subgroup H_{k-1}, the fact that $q \notin S$ implies that the whole coset $H_{k-1} + q$ is disjoint with S, and the inclusion $\operatorname{excl}(wb^q) \subseteq H_{k-1} + q$ ensures that S is disjoint with $\operatorname{excl}(wb^q)$. Therefore, for every $s \in S \setminus \{p\}$, there exists a state $s' \in \mathbb{Z}_n$ such that $s'\,.\,wb^q = s$. Now letting $T = \{p_1, p_2\} \cup \{s' \mid s \in S \setminus \{p\}\}$, we conclude that $S = T\,.\,wb^q$ and $|T| = |S| + 1$. □

For an illustration, return one last time to the DFA $\mathscr{E}'_{12}$ shown in Fig. 3. We have seen that the subgroup $H_1(\mathscr{E}'_{12})$ consists of even residues modulo 12. Inspecting the word ab^3a gives $\operatorname{excl}(ab^3a) = \{0, 8\} \subseteq H_1(\mathscr{E}'_{12})$ and $1 \in \operatorname{dupl}(ab^3a)$, whence $1 \in D_2(\mathscr{E}'_{12})$. Therefore the subgroup $H_2(\mathscr{E}'_{12})$ generated by $D_2(\mathscr{E}'_{12})$ is equal to $\mathbb{Z}_{12}$, and $\mathscr{E}'_{12}$ is a completely reachable automaton by Proposition 5.

To illustrate the next level of the construction (5), consider the standardized DFA $\mathscr{E}_{48} = \langle \mathbb{Z}_{48}, \{a,b\}\rangle$ shown in Fig. 4. We have replaced edges that should have been labeled a and b with solid and, resp., dashed edges and omitted all loops to lighten the picture. The action of a in $\mathscr{E}_{48}$ is defined by $0\,.\,a = 24\,.\,a = 18$, $13\,.\,a = 14$, $14\,.\,a = 13$, $18\,.\,a = 24$, $30\,.\,a = 32$, $32\,.\,a = 30$, and $k\,.\,a = k$ for all other $k \in \mathbb{Z}_{48}$.

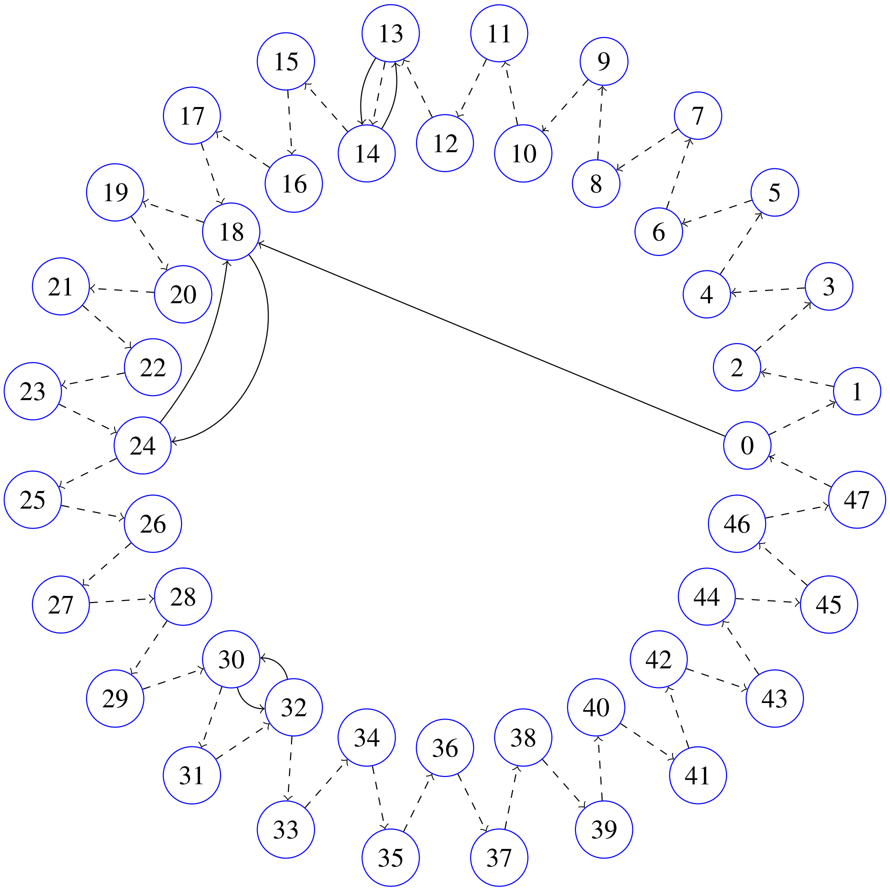

FIGURE 4. The DFA $\mathscr{E}_{48} = \langle \mathbb{Z}_{48}, \{a,b\} \rangle$ with $H_2(\mathscr{E}_{48}) \neq \mathbb{Z}_{48}$. Solid and dashed edges show the action of a and, resp., b; loops are not shown

One can calculate that $D_1(\mathscr{E}_{48}) = \{18, 24, 42\}$ whence the subgroup $H_1(\mathscr{E}_{48})$ consists of all residues divisible by 6. Computing $D_2(\mathscr{E}_{48})$, one sees that this set consists of even residues and contains 2 (due to the word $ab^{32}a$ that has $\mathrm{excl}(ab^{32}a) = \{0, 30\} \subseteq H_1(\mathscr{E}_{48})$ and $\mathrm{dupl}(ab^{32}a) = \{2, 18\}$). Hence the subgroup $H_2(\mathscr{E}_{48})$ consists of all even residues. Finally, the word $ab^{24}ab^{12}ab^{8}$ has $\{0, 8, 20\} \subseteq H_1(\mathscr{E}_{48})$ as its excluded set while its duplicate set contains 13. Hence $13 \in D_3(\mathscr{E}_{48})$ and the subgroup $H_3(\mathscr{E}_{48})$ coincides with $\mathbb{Z}_{48}$. We conclude that the DFA $\mathscr{E}_{48}$ is completely reachable by Proposition 5.

As mentioned, the subgroups of $(\mathbb{Z}_n, +)$ ordered by inclusion correspond to the divisors of n ordered by division whence for any standardized DFA $\mathscr{A}$ with n states, the number of different subgroups of the form $H_k(\mathscr{A})$ is $O(\log n)$. Therefore, if the subgroup sequence $H_0(\mathscr{A}) \subseteq H_1(\mathscr{A}) \subseteq \cdots \subseteq H_k(\mathscr{A}) \subseteq \dots$ strictly grows at each step, then it reaches $\mathbb{Z}_n$ after at most $O(\log n)$ steps, and by Proposition 5 $\mathscr{A}$ is a completely reachable automaton. What happens if the sequence stabilizes earlier? Our next result answers this question.

Proposition 6. *If for a standardized DFA $\mathscr{A} = \langle \mathbb{Z}_n, \{a,b\} \rangle$, there exists ℓ such that $H_\ell(\mathscr{A}) = H_{\ell+1}(\mathscr{A}) \subsetneqq \mathbb{Z}_n$, then $\mathscr{A}$ is not completely reachable.*

Proof. As in the proof of Proposition 5, we use D_k and H_k instead of $D_k(\mathscr{A})$ and, resp., $H_k(\mathscr{A})$ in our arguments.

It suffices to prove the following claim:

Claim: *the equality $H_\ell = H_{\ell+1}$ implies that the subgroup H_ℓ is a-invariant.*

Indeed, since $H_\ell \subsetneqq \mathbb{Z}_n$, we get a proper a-invariant subgroup, and Proposition 1 then shows that $\mathscr{A}$ is not completely reachable.

Technically, it is more convenient to show that if $H_\ell = H_{\ell+1}$, then $H_k\,.\,a \subseteq H_\ell$ for every $k = 0,1,\dots,\ell$. We induct on k. The base $k = 0$ is clear since $H_0 = \{0\}$ and $0\,.\,a = \operatorname{dupl}(a) \in D_1 \subseteq H_1 \subseteq H_\ell$.

Let $k < \ell$ and assume $H_k\,.\,a \subseteq H_\ell$; we aim to verify that $p\,.\,a \in H_\ell$ for every $p \in H_{k+1}$. Since the subgroup H_{k+1} is generated by D_{k+1} and contains H_k, we may choose a representation of p as the sum

$$p = q + d_1 + \dots + d_m, \quad q \in H_k,\ d_1,\dots,d_m \in D_{k+1} \setminus H_k,$$

with the least number m of summands from $D_{k+1} \setminus H_k$. We show that $p\,.\,a \in H_\ell$ by induction on m. If $m = 0$, we have $p = q \in H_k$ and $p\,.\,a \in H_\ell$ since $H_k\,.\,a \subseteq H_\ell$.

If $m > 0$, we write p as $p = d_1 + s$ where $s = q + d_2 + \dots + d_m$. By (5), there exists a word $w \in \{a,b\}^*$ such that $d_1 \in \operatorname{dupl}(w)$, $0 \in \operatorname{excl}(w) \subseteq H_k$ and $|\operatorname{excl}(w)| \le k+1$. Consider the word wb^sa. We have $p\,.\,a = (d_1+s)\,.\,a = d_1\,.\,b^sa$, and the equality (2) gives $p\,.\,a \in \operatorname{dupl}(wb^sa)$. From the equality (1), we get $\operatorname{excl}(wb^sa) = (\operatorname{excl}(w)+s)\,.\,a \cup \{0\}$ if $\operatorname{dupl}(a)a^{-1}$ is either contained in or disjoint with $\operatorname{excl}(w)+s$, and $\operatorname{excl}(wb^sa) = \big((\operatorname{excl}(w)+s)\setminus \operatorname{dupl}(a)a^{-1}\big)\,.\,a \cup \{0\}$ if $|\operatorname{dupl}(a)a^{-1} \cap (\operatorname{excl}(w)+s)| = 1$. In any case, we have the inclusion

$$\operatorname{excl}(wb^sa) \subseteq (\operatorname{excl}(w)+s)\,.\,a \cup \{0\} \tag{6}$$

and the inequality

$$|\operatorname{excl}(wb^sa)| \le |(\operatorname{excl}(w)+s))\,.\,a| + 1 \le |\operatorname{excl}(w))| + 1 \le (k+1)+1 \le \ell+1. \tag{7}$$

For any $t \in \operatorname{excl}(w) \subseteq H_k$, the number of summands from $D_{k+1} \setminus H_k$ in the sum $t+s = t+q+d_2+\dots+d_m$ is less than m. By the induction assumption, we have $(t+s)\,.\,a \in H_\ell$. Hence, $(\operatorname{excl}(w)+s)\,.\,a \subseteq H_\ell$, and since 0 also lies in the subgroup H_ℓ, we conclude from (6) that $\operatorname{excl}(wb^sa) \subseteq H_\ell$. From this and the inequality (7), we see that the word wb^sa satisfies the conditions of the definition of $D_{\ell+1}$ (cf. (5)) whence every state in $\operatorname{dupl}(wb^sa)$ belongs to $D_{\ell+1}$. We have observed that $p\,.\,a \in \operatorname{dupl}(wb^sa)$. Hence $p\,.\,a \in D_{\ell+1} \subseteq H_{\ell+1}$. Since $H_\ell = H_{\ell+1}$, we have $p\,.\,a \in H_\ell$, as required. □

Now we deduce a criterion for complete reachability of binary automata.

Theorem 2. *A binary DFA $\mathscr{A}$ with n states is completely reachable if and only if either $n = 2$ and $\mathscr{A}$ is the flip-flop or one of the letters of $\mathscr{A}$ acts as a cyclic permutation of the state set, the other letter has defect 1, and in the standardized DFA $\langle \mathbb{Z}_n, \{a,b\}\rangle$ syntactically equivalent to $\mathscr{A}$, no proper subgroup of $(\mathbb{Z}_n,+)$ is a-invariant.*

Proof. Necessity follows from the reductions in Sect. 2 and Proposition 1.

For sufficiency, we can assume that $\mathscr{A} = \langle \mathbb{Z}_n, \{a,b\}\rangle$ is standardized. If no proper subgroup of $(\mathbb{Z}_n,+)$ is a-invariant, then the claim from the proof of Proposition 6 implies that the sequence $H_0(\mathscr{A}) \subseteq H_1(\mathscr{A}) \subseteq \dots \subseteq H_k(\mathscr{A}) \subseteq \dots$ strictly grows as long as the subgroup $H_k(\mathscr{A})$ remains proper. Hence, $H_\ell(\mathscr{A}) = \mathbb{Z}_n$ for some ℓ and $\mathscr{A}$ is a completely reachable automaton by Proposition 5. □

Remark 1. The proof of Theorem 2 shows that only subgroups that contain $H_1(\mathscr{A})$ matter. Therefore, one can combine Theorem 1, Proposition 2 and Theorem 2 as follows: *a standardized DFA* $\mathscr{A} = \langle\mathbb{Z}_n, \{a,b\}\rangle$ *is completely reachable if and only if either* $H_1(\mathscr{A}) = \mathbb{Z}_n$ *or no proper subgroup of* $(\mathbb{Z}_n, +)$ *containing the subgroup* $H_1(\mathscr{A})$ *is a-invariant.*

The condition of Theorem 2 can be verified in low polynomial time. We sketch the corresponding algorithm.

Given a binary DFA $\mathscr{A}$ with n states, we first check if $n = 2$ and $\mathscr{A}$ is the flip-flop. If **yes**, $\mathscr{A}$ is completely reachable. If **not**, we check whether one of the letters of $\mathscr{A}$ acts as a cyclic permutation of the state set while the other letter has defect 1. If **not**, $\mathscr{A}$ is not completely reachable. If **yes**, we pass to the standardized DFA $\langle\mathbb{Z}_n, \{a,b\}\rangle$ syntactically equivalent to $\mathscr{A}$. As a preprocessing, we compute and store the set $\{(k, k\,.\,a) \mid k \in \mathbb{Z}_n\}$.

The rest of the algorithm can be stated in purely arithmetical terms. Call a positive integer d a *nontrivial divisor* of n if d divides n and $d \neq 1, n$. We compute all nontrivial divisors of n by checking through all integers $d = 2, \dots, \lfloor\sqrt{n}\rfloor$: if such d divides n, we store d and $\frac{n}{d}$. If for some nontrivial divisor d of n, all numbers $(td)\,.\,a$ with $t = 0, 1, \dots, \frac{n}{d} - 1$ are divisible by d, then d generates a proper a-invariant subgroup in $(\mathbb{Z}_n, +)$ and $\mathscr{A}$ is not completely reachable. If for every nontrivial divisor d of n, there exists $t \in \{0, 1, \dots, \frac{n}{d} - 1\}$ such that $(td)\,.\,a$ is not divisible by d, then no proper subgroup of $(\mathbb{Z}_n, +)$ is a-invariant and $\mathscr{A}$ is completely reachable.

To estimate the time complexity of the described procedure, observe that one has to check at most $\frac{n}{d}$ numbers for each nontrivial divisor d of n. Clearly,

$$\sum_{\substack{1<d<n\\ d\mid n}} \frac{n}{d} = \sum_{\substack{1<d<n\\ d\mid n}} d = \sigma(n) - n - 1,$$

where $\sigma(n)$ stands for the sum of all divisors of n, a well-studied function in the theory of numbers; see, e.g., [8, Chapters XVI–XVIII]. It is known that $\limsup \frac{\sigma(n)}{n \log\log n} = e^{\gamma}$ where γ is the Euler–Mascheroni constant [8, Theorem 323]; this implies that the number of checks in our procedure is $O(n \log\log n)$. The total complexity depends on the time spent for verifying the divisibility condition. If one uses the transdichotomous model [7] (as suggested by one of the referees), assuming constant time for division, the whole procedure can be implemented in $O(n \log\log n)$ time.

One can speed up the above algorithm, using Remark 1, which implies that only the divisors $d > 1$ of the g.c.d. of n and $0\,.\,a$ have to be checked. However, the improvement only reduces the constant behind the $O(\ )$ notation.

6. Conclusion

We have characterized binary completely reachable automata; our characterization leads to an algorithm that given a binary DFA $\mathscr{A}$, decides whether or not $\mathscr{A}$ is completely reachable in quasilinear time with respect to the size of $\mathscr{A}$. Very recently, after the original version of the present paper was submitted, Ferens and Szykuła [6] have devised a polynomial-time algorithm for recognizing complete reachability of arbitrary DFAs, but the complexity of their algorithm is higher.

Our results heavily depend on the fact that apart from a single exception, binary completely reachable automata are *circular*, that is, have a letter acting as a cyclic permutation of the state set. In the literature, one can find several situations when a problem that remains open in general, admits quite a nontrivial solution when restricted to circular automata. Here we mention only Dubuc's result [5] on the Černý conjecture and the recent paper by Yong He *et al* [9] on Trahtman's conjecture. It appears that circular automata may behave in a similar way with respect to complete reachability, and our follow-up work aims at extending the results of the present paper to arbitrary (not necessarily binary) circular automata. We also plan to study an 'orthogonal' extension, aiming to characterize completely reachable automata in which one letter has defect 1 while the other letters act as permutations and generate a group that transitively acts on the state set.

Acknowledgement. We thank the anonymous reviewers of the conference version of our paper for their careful reading and their many useful comments and suggestions that are incorporated in the present version.

References

[1] Bondar, E.A., Casas, D., Volkov, M.V.: Completely reachable automata: an interplay between automata, graphs, and trees. CoRR **abs/2201.05075** (2022), https://arxiv.org/abs/2201.05075

[2] Bondar, E.A., Volkov, M.V.: Completely reachable automata. In: Câmpeanu, C., Manea, F., Shallit, J. (eds.) DCFS 2016. Lect. Notes Comput. Sci., vol. 9777, pp. 1–17. Springer (2016)

[3] Bondar, E.A., Volkov, M.V.: A characterization of completely reachable automata. In: Hoshi, M., Seki, S. (eds.) DLT 2018. Lect. Notes Comput. Sci., vol. 11088, pp. 145–155. Springer (2018)

[4] Don, H.: The Černý conjecture and 1-contracting automata. Electr. J. Combinatorics **23**(3), 3–12 (2016)

[5] Dubuc, L.: Sur les automates circulaires et la conjecture de Černý. RAIRO Informatique Théorique et Applications **32**, 21–34 (1998), in French

[6] Ferens, R., Szykula, M.: Completely reachable automata: A polynomial solution and quadratic bounds for the subset reachability problem. CoRR **abs/2208.05956** (2022), https://arxiv.org/abs/2208.05956

[7] Fredman, M.L., Willard, D.E.: Surpassing the information theoretic bound with fusion trees. J. Comput. Syst. Sci. **47**(3), 424–436 (1993)

[8] Hardy, G.H., Wright, E.M.: An Introduction to the Theory of Numbers. Oxford University Press, 6th edn. (2008)

[9] He, Y., Chen, X., Li, G., Sun, S.: Extremal synchronizing circular automata. Information and Computation **281**, article no. 104817 (2021)

[10] Maslennikova, M.I.: Reset complexity of ideal languages. In: Bieliková, M., Friedrich, G., Gottlob, G., Katzenbeisser, S., Špánek, R., Turán, G. (eds.) SOFSEM 2012. vol. II, pp. 33–44. Institute of Computer Science Academy of Sciences of the Czech Republic (2012), see also http://arxiv.org/abs/1404.2816

[11] Rystsov, I.K.: Estimation of the length of reset words for automata with simple idempotents. Cybernetics and Systems Analysis **36**(3), 339–344 (2000)