Families of DFAs as Acceptors of $\omega$-Regular Languages

Dana Angluin$^1$, Udi Boker$^2$, and Dana Fisman$^3$

1 Yale University, New Haven, CT, USA
2 The Interdisciplinary Center (IDC), Herzliya, Israel
3 University of Pennsylvania, Philadelphia, PA, USA

Abstract

Families of DFAs (FDFAs) provide an alternative formalism for recognizing $\omega$-regular languages. The motivation for introducing them was a desired correlation between the automaton states and right congruence relations, in a manner similar to the Myhill-Nerode theorem for regular languages. This correlation is beneficial for learning algorithms, and indeed it was recently shown that $\omega$-regular languages can be learned from membership and equivalence queries, using FDFAs as the acceptors.

In this paper, we look into the question of how suitable FDFAs are for defining $\omega$-regular languages. Specifically, we look into the complexity of performing Boolean operations, such as complementation and intersection, on FDFAs, the complexity of solving decision problems, such as emptiness and language containment, and the succinctness of FDFAs compared to standard deterministic and nondeterministic $\omega$-automata.

We show that FDFAs enjoy the benefits of deterministic automata with respect to Boolean operations and decision problems. Namely, they can all be performed in nondeterministic logarithmic space. We provide polynomial translations of deterministic Büchi and co-Büchi automata to FDFAs and of FDFAs to nondeterministic Büchi automata (NBAs). We show that translation of an NBA to an FDFA may involve an exponential blowup. Last, we show that FDFAs are more succinct than deterministic parity automata (DPAs) in the sense that translating a DPA to an FDFA can always be done with only a polynomial increase, yet the other direction involves an inevitable exponential blowup in the worst case.

1998 ACM Subject Classification F.1.1 Automata, D.2.4 Formal methods

Keywords and phrases Finite automata, Omega Regular Languages

Digital Object Identifier 10.4230/LIPIcs.MFCS.2016.11

1 Introduction

The theory of finite-state automata processing infinite words was developed in the early sixties, starting with Büchi [3] and Muller [13], and motivated by problems in logic and switching theory. Today, automata for infinite words are extensively used in verification and synthesis of reactive systems, such as operating systems and communication protocols.

An automaton processing finite words makes its decision according to the last visited state. On infinite words, Büchi defined that a run is accepting if it visits a designated set of states infinitely often. Since then several other accepting conditions were defined, giving rise to various $\omega$-automata, among which are Muller, Rabin, Streett and parity automata.

* Research of the third author was supported by US NSF grant CCF-1138996
Families of DFAs as Acceptors of \(\omega\)-Regular Languages

The theory of \(\omega\)-regular languages is more involved than that of finite words. This was first evidenced by Büchi’s observation that nondeterministic Büchi automata are more expressive than their deterministic counterpart. While for some types of \(\omega\)-automata the nondeterministic and deterministic variants have the same expressive power, none of them possesses all the nice qualities of acceptors for finite words. In particular, none has a corresponding Myhill-Nerode theorem [16], i.e., a direct correlation between the states of the automaton and the equivalence classes corresponding to the canonical right congruence of the recognized language.

The absence of a Myhill-Nerode like property in \(\omega\)-automata has been a major drawback in obtaining learning algorithms for \(\omega\)-regular languages, a question that has received much attention lately due to applications in verification and synthesis, such as black-box checking [17], assume-guarantee reasoning [14], error localization [5], regular model checking [15] and more. The reason is that learning algorithms typically build on this correspondence between the automaton and the right congruence.

Recently, two algorithms for learning an unknown \(\omega\)-regular language were proposed, both using non-conventional acceptors. One uses a reduction due to [4] named \(L_S\)-automata of \(\omega\)-regular languages to regular languages [6], and the other uses a representation termed families of DFAs [1]. Both representations are founded on the following well known property of \(\omega\)-regular languages: two \(\omega\)-regular languages are equivalent if they agree on the set of ultimately periodic words. An ultimately periodic word \(uv\omega\), where \(u \in \Sigma^*\) and \(v \in \Sigma^+\), can be represented as a pair of finite words \((u, v)\). Both \(L_S\)-automata and families of DFAs process such pairs and interpret them as the corresponding ultimately periodic words. Families of DFAs have been shown to be up to exponentially more succinct than \(L_S\)-automata [1].

A family of DFAs (F DFA) is composed of a leading automaton \(\mathbb{Q}\) with no accepting states and for each state \(q\) of \(\mathbb{Q}\), a progress DFA \(\mathcal{P}_q\). Intuitively, the leading automaton is responsible for processing the non-periodic part \(u\), and depending on the state \(q\) reached when \(\mathbb{Q}\) terminated processing \(u\), the respective progress DFA \(\mathcal{P}_q\) processes the periodic part \(v\), and determines whether the pair \((u, v)\), which corresponds to \(uv\omega\), is accepted. (The exact definition is more subtle and is provided in Section 3.) If the leading automaton has \(n\) states and the size of the maximal progress DFA is \(k\), we say that the F DFA is of size \((n, k)\). An earlier definition of FDFAs, given in [9], provided a machine model for the families of right congruences of [10]. They were redefined in [1], where their acceptance criterion was adjusted, and their size was reduced by up to a quadratic factor. We follow the definition of [1].

In order for an F DFA to properly characterize an \(\omega\)-regular language, it must satisfy the saturation property: considering two pairs \((u, v)\) and \((u', v')\), if \(uv\omega = u'v'\omega\) then either both \((u, v)\) and \((u', v')\) are accepted or both are rejected (cf. [4, 20]). Saturated FDFAs are shown to exactly characterize the set of \(\omega\)-regular languages. Saturation is a semantic property, and the check of whether a given F DFA is saturated is shown to be in PSPACE. Luckily, the FDFAs that result from the learning algorithm of [1] are guaranteed to be saturated.

Saturated FDFAs bring an interesting potential – they have a Myhill-Nerode like property, and while they are “mostly” deterministic, a nondeterministic aspect is hidden in the separation of the prefix and period parts of an ultimately periodic infinite word. This gives rise to the natural questions of how “dominant” are the determinism and nondeterminism in FDFAs, and how “good” are they for representing \(\omega\)-regular languages. These abstract questions translate to concrete questions that concern the succinctness of FDFAs and the complexity of solving their decision problems, as these measures play a key role in the usefulness of applications built on top of them.

Our purpose in this paper is to analyze the F DFA formalism and answer these questions.
Specifically, we ask: What is the complexity of performing the Boolean operations of complementation, union, and intersection on saturated \( \text{fdfa} \)? What is the complexity of solving the decision problems of membership, emptiness, universality, and language containment for saturated \( \text{fdfa} \)? How succinct are saturated \( \text{fdfa} \), compared to deterministic and nondeterministic \( \omega \)-automata?

We show that saturated \( \text{fdfa} \) enjoy the benefits of deterministic automata with respect to Boolean operations and decision functions. Namely, the Boolean operations can be performed in logarithmic space, and the decision problems can be solved in nondeterministic logarithmic space. The constructions and algorithms that we use extend their counterparts on standard \( \text{dfa} \). In particular, complementation of saturated \( \text{fdfa} \) can be obtained on the same structure, and union and intersection is done on a product of the two given structures. The correctness proof of the latter is a bit subtle.

As for the succinctness, which turns out to be more involved, we show that saturated \( \text{fdfa} \) properly lie in between deterministic and nondeterministic \( \omega \)-automata. We provide polynomial translations from deterministic \( \omega \)-automata to \( \text{fdfa} \) and from \( \text{fdfa} \) to nondeterministic \( \omega \)-automata, and show that an exponential state blowup in the opposite directions is inevitable in the worst case.

Specifically, a saturated \( \text{dfa} \) of size \((n,k)\) can always be transformed into an equivalent nondeterministic Büchi automaton (NBA) with \(O(n^2k^3)\) states. As for the other direction, transforming an NBA with \(n\) states to an equivalent \( \text{fda} \) is shown to be in \(2^{\Theta(n \log n)}\). This is not surprising since, as shown by Michel [12], complementing an NBA involves a \(2^{\Theta(n \log n)}\) state blowup, while \( \text{fda} \) complementation requires no state blowup.

Considering deterministic \( \omega \)-automata, a Büchi or co-Büchi automaton (DBA or DCA) with \(n\) states can be transformed into an equivalent \( \text{fda} \) of size \((n,2n)\), and a deterministic parity automaton (DPA) with \(n\) states and \(k\) colors can be transformed into an equivalent \( \text{fda} \) of size \((n, kn)\). As for the other direction, since DBA and DCA do not recognize all the \( \omega \)-regular languages, while saturated \( \text{fdfa} \) do, a transformation from an \( \text{fda} \) to a DBA or DCA need not exist. Comparing \( \text{fdfa} \) to DPA, which do recognize all \( \omega \)-regular languages, we get that \( \text{fdfa} \) can be exponentially more succinct: We show a family of languages \( \{L_n\}_{n \geq 1} \), such that for every \( n \), there exists an \( \text{fda} \) of size \((n+1, n^2)\) for \( L_n \), but any DPA recognizing \( L_n \) must have at least \(2^{n-1}\) states. (A deterministic Rabin or Streett automaton for \( L_n \) is also shown to be exponential in \( n \), requiring at least \(2^2\) states.)

Due to lack of space, some proofs are omitted and can be found in the full version, on the authors’ home pages.

## 2 Preliminaries

An alphabet \( \Sigma \) is a finite set of symbols. The set of finite words over \( \Sigma \) is denoted by \( \Sigma^* \), and the set of infinite words, termed \( \omega \)-words, over \( \Sigma \) is denoted by \( \Sigma^\omega \). As usual, we use \( x^* \), \( x^+ \), and \( x^\omega \) to denote finite, non-empty finite, and infinite concatenations of \( x \), respectively, where \( x \) can be a symbol, a finite word, or a language. We use \( \epsilon \) for the empty word and \( \Sigma^+ \) for \( \Sigma^* \setminus \{\epsilon\} \). An infinite word \( w \) is \emph{ultimately periodic} if there are two finite words \( u \in \Sigma^* \) and \( v \in \Sigma^+ \), such that \( w = uv^\omega \). A language is a set of finite words, that is, a subset of \( \Sigma^* \), while an \( \omega \)-language is a set of \( \omega \)-words, that is, a subset of \( \Sigma^\omega \). For natural numbers \( i \) and \( j \) and a word \( w \), we use \([i..j]\) for the set \( \{i, i+1, \ldots, j\} \), \( w[i] \) for the \( i \)-th letter of \( w \), and \( w[i..j] \) for the subword of \( w \) starting at the \( i \)-th letter and ending at the \( j \)-th letter, inclusive.

An automaton is a tuple \( A = \langle \Sigma, Q, \iota, \delta \rangle \) consisting of an alphabet \( \Sigma \), a finite set \( Q \) of states, an initial state \( \iota \in Q \), and a transition function \( \delta : Q \times \Sigma \to 2^Q \). A run of an
Families of DFAs as Acceptors of $\omega$-Regular Languages

An automaton on a finite word $v = a_1a_2 \ldots a_n$ is a sequence of states $q_0, q_1, \ldots, q_n$ such that $q_0 = i$, and for each $i \geq 0$, $q_{i+1} \in \delta(q_i, a_i)$. A run on an infinite word is defined similarly and results in an infinite sequence of states. The transition function is naturally extended to a function $\delta : Q \times \Sigma^* \to 2^Q$, by defining $\delta(q, \varepsilon) = \{q\}$, and $\delta(q, av) = \cup_{p \in \delta(q, a)} \delta(p, v)$ for $q \in Q$, $a \in \Sigma$, and $v \in \Sigma^*$. We often use $A(v)$ as a shorthand for $\delta(i, v)$ and $|A|$ for the number of states in $Q$. We use $A^v$ to denote the automaton $\langle Q, q, \delta \rangle$ obtained from $A$ by replacing the initial state with $q$. We say that $A$ is deterministic if $|\delta(q, a)| \leq 1$ and complete if $|\delta(q, a)| \geq 1$, for every $q \in Q$ and $a \in \Sigma$. We use the term automaton to mean complete. (As is known, every automaton can be linearly translated to an equivalent complete automaton.)

By augmenting an automaton with an acceptance condition $\alpha$, thereby obtaining a tuple $\langle \Sigma, Q, i, \delta, \alpha \rangle$, we get an acceptor, a machine that accepts some words and rejects others. An acceptor accepts a word if at least one of the runs on that word is accepting. For finite words the acceptance condition is a set $P$ of states, which is a set of accepting states. For infinite words, there are various acceptance conditions in the literature; here we mention three: Büchi, co-Büchi, and parity. The Büchi and co-Büchi acceptance conditions are also a set $F \subseteq Q$. A run of a Büchi automaton is accepting if it visits $F$ infinitely often. A run of a co-Büchi automaton is accepting if it visits $F$ only finitely many times. A parity acceptance condition is a function $\kappa : Q \to \{0, 1\}$ assigning each state a color (or rank). A run is accepting if the minimal color visited infinitely often is odd. We use $[A]$ to denote the set of words accepted by a given acceptor $A$, and say that $A$ accepts or recognizes $[A]$. Two acceptors $A$ and $B$ are equivalent if $[A] = [B]$.

We use three letter acronyms to describe acceptors, where the first letter is either D or N depending on whether the automaton is deterministic or nondeterministic, respectively. The second letter is one of F, B, C, or P if it is an acceptor over finite words, B, C, or P if it is an acceptor over infinite words with Büchi, co-Büchi, or parity acceptance condition, respectively. The third letter is always A for acceptor.

For finite words, NFA and DFA have the same expressive power. A language is said to be regular if it is accepted by an NFA. For infinite words, the theory is more involved. While NPAs, DPAs, and NBFAs have the same expressive power, DBAs, NCCAs, and DCCAs are strictly weaker than NBAs. An $\omega$-language is said to be $\omega$-regular if it is accepted by an NBA.

3 Families of DFAs (FDFAs)

It is well known that two $\omega$-regular languages are equivalent if they agree on the set of ultimately periodic words (this is a consequence of McNaughton’s theorem [11]). An ultimately periodic word $uv^\omega$, where $u \in \Sigma^*$ and $v \in \Sigma^+$, is usually represented by the pair $(u, v)$. A canonical representation of an $\omega$-regular language can thus consider only ultimately periodic words, namely define a language of pairs $(u, v) \in \Sigma^* \times \Sigma^+$ such a representation $F$ should satisfy the saturation property: considering two pairs $(u, v)$ and $(u', v')$, if $uv^\omega = u'v'^\omega$ then either both $(u, v)$ and $(u', v')$ are accepted by $F$ or both are rejected by $F$.

A family of DFAs (F DFA) accepts such pairs $(u, v)$ of finite words. Intuitively, it consists of a leading automaton $Q$ with no acceptance condition that runs on the prefix word $u$, and for each state $q$ of $Q$, a progress automaton $P_q$, which is a DFA that runs on the period word $v$.

A straightforward definition of acceptance for a pair $(u, v)$, could have been that the run of the leading automaton $Q$ on $u$ ends at some state $q$, and the run of the progress automaton $P_q$ on $v$ is accepting. This goes along the lines of $L_2$-automata [4]. However,
We provide below algorithms for performing the Boolean operations of complementation, while an unsaturated language with respect to a very large automaton \cite{1}. The intuitive reason is that every progress automaton might need to handle the period-words of all prefix-words.

To better fit the saturation requirement, the acceptance condition of an fdfa is defined with respect to a normalization of the input pair \((u, v)\). The normalization is a new pair \((x, y)\), such that \(xy^\omega = uv^\omega\), and in addition, the run of the leading automaton \(Q\) on \(xy^i\) ends at the same state for every natural number \(i\). Over the normalized pair \((x, y)\), the acceptance condition follows the straightforward approach discussed above. This normalization resembles the implicit flexibility in the acceptance conditions of \(\omega\)-automata, such as the Büchi condition, and allows saturated fdfs to be up to exponentially more succinct than \(L_\omega\)-automata \cite{1}.

Below, we formally define an fdfa, the normalization of an input pair \((u, v)\), and the acceptance condition. We shall use \(\Sigma^*\) as a shorthand for \(\Sigma^* \times \Sigma^*\), whereby the input to an fdfa is a pair \((u, v) \in \Sigma^*\).

\begin{definition}[A family of dfas (fdfa)]
A family of dfas (fdfa) is a pair \((Q, P)\), where \(Q = (\Sigma, Q, \iota, \delta)\) is a deterministic leading automaton, and \(P\) is a set of \(|Q|\) dfas, including for each state \(q \in Q\), a progress dfa \(P_q = (\Sigma, P_q, \iota_q, \delta_q, F_q)\).

Given a pair \((u, v) \in \Sigma^*\) and an automaton \(A\), the normalization of \((u, v)\) w.r.t. \(A\) is the pair \((x, y) \in \Sigma^*\), such that \(x = uv^i\), \(y = v^i\), and \(i \geq 0\), \(j \geq 1\) are the smallest numbers for which \(A(uv^i) = A(uv^{i+j})\). (Since we consider complete automata, such a unique pair \((x, y)\) is guaranteed.)

Let \(F = (Q, P)\) be an fdfa, \((u, v) \in \Sigma^*\), and \((x, y) \in \Sigma^*\) the normalization of \((u, v)\) w.r.t. \(Q\). We say that \((u, v)\) is accepted by \(F\) if \(Q(x) = q\) for some state \(q\) of \(Q\) and \(P_q(y)\) is an accepting state of \(P_q\).

We use \([F]\) to denote the set of pairs accepted by \(F\).

We define the size of \(F\), denoted \(|F|\), as the pair \((|Q|, \max\{|P_q| | q \in Q\})\).

An fdfa \(F\) is saturated if for every two pairs \((u, v)\) and \((u', v')\) such that \(uv^\omega = u'v'^\omega\), either both \((u, v)\) and \((u', v')\) are in \([F]\) or both are not in \([F]\).

A saturated fdfa can be used to characterize an \(\omega\)-regular language (see Theorem 10), while an unsaturated fdfa cannot.

An unsaturated fdfa is depicted in Figure 1 on the left. Consider the pairs \((b, a)\) and \((ba, aa)\). Though \(b(a)^\omega = ba(aa)^\omega\), \((b, a)\) is normalized to \((b, aa)\) and \(P_{ba}^f\) accepts aa but \((ba, aa)\) is normalized to itself and \(P_{ba}^f\) rejects aa. A saturated fdfa is depicted in Figure 1 on the right. It accepts pairs of the forms \((\Sigma^*, a^+)\) and \((\Sigma^*, b^+)\), and characterizes the \(\omega\)-regular language \((a + b)^*(a^\omega + b^\omega)\).

\section{4 Boolean Operations and Decision Procedures}

We provide below algorithms for performing the Boolean operations of complementation, union, and intersection on saturated fdfs, and deciding the basic questions on them, such as emptiness, universality, and language containment. All of these algorithms can be done in

\footnote{The fdfs defined here follow the definition in \cite{1}, which is a little different from the definition of fdfs in \cite{9}; the latter provide a machine model for the families of right congruences introduced in \cite{10}. The main differences between the two definitions are: i) In \cite{9}, a pair \((u, v)\) is accepted by an fdfa \(F = (Q, P)\) if there is some factorization \((x, y)\) of \((u, v)\), such that \(Q(x) = q\) and \(P_q(x)\) accepts \(v\); and ii) in \cite{9}, the fdfa \(F\) should also satisfy the constraint that for all words \(u \in \Sigma^*\) and \(v, v' \in \Sigma^*\), if \(P_{Q(u)}(v) = P_{Q(u)}(v')\) then \(Q(uv) = Q(uv')\).}
Families of DFAs as Acceptors of $\omega$-Regular Languages

nondeterministic logarithmic space, taking advantage of the partial deterministic nature of DFAs. We conclude the section with the decision problem of whether an arbitrary FDFA is saturated, showing that it can be resolved in polynomial space.

Boolean operations

Saturated DFAs are closed under Boolean operations as a consequence of Theorem 10, which shows that they characterize exactly the set of $\omega$-regular languages. We provide below explicit algorithms for these operations, showing that they can be done effectively.

- **Theorem 2.** Let $F$ be an FDFA. There is a constant-space algorithm to obtain an FDFA $F^c$, such that $\{F^c\} = \Sigma^{*+} \setminus \{F\}$, $|F^c| = |F|$, and $F^c$ is saturated iff $F$ is.

Union and intersection of saturated DFAs also resemble the case of DFAs, and are done by taking the product of the leading automata and each pair of progress automata. Yet, the correctness proof is a bit subtle, and relies on the following lemma, which shows that for a normalized pair $(x, y)$, the period-word $y$ can be manipulated in a certain way, while retaining normalization.

- **Lemma 3.** Let $Q$ be an automaton, and let $(x, y)$ be the normalization of some $(u, v) \in \Sigma^{*+}$ w.r.t. $Q$. Then for every $i \geq 0$, $j \geq 1$ and finite words $y', y''$ such that $y = y'y''$, we have that $(xy'y', (y'y''))^j$ is the normalization of itself w.r.t. $Q$.

- **Theorem 4.** Let $F_1$ and $F_2$ be saturated DFAs of size $(n_1, k_1)$ and $(n_2, k_2)$, respectively. There exist logarithmic-space algorithms to obtain saturated DFAs $H$ and $H'$ of size $(n_1n_2, k_1k_2)$, such that $\{H\} = \{F_1\} \cap \{F_2\}$ and $\{H'\} = \{F_1\} \cup \{F_2\}$.

Decision procedures

All of the basic decision problems can be resolved in nondeterministic logarithmic space, using the Boolean operations above and corresponding decision algorithms for DFAs.

The first decision question to consider is that of membership: given a pair $(u, v)$ and an FDFA $F = (Q, P)$, does $F$ accept $(u, v)$? The question is answered by normalizing $(u, v)$

\[2\] Another model that lies in between deterministic and nondeterministic automata are “semi-deterministic Büchi automata” [25], which are Büchi automata that are deterministic in the limit: from every accepting state onward, their behaviour is deterministic. Yet, as opposed to DFAs, complementation of semi-deterministic Büchi automata might involve an exponential state blowup [2].
into a pair \((x, y)\) and evaluating the runs of \(Q\) over \(x\) and of \(P_{Q(x)}\) over \(y\). A normalized pair is determined by traversing along \(Q\), making up to \(|Q|\) repetitions of \(v\). Notice that memory wise, \(x\) and \(y\) only require a logarithmic amount of space, as they are of the form \(x = uv^n\) and \(y = v^i\), where the representation of \(i\) and \(j\) is bounded by \(\log |Q|\). The overall logarithmic-space solution follows from the complexity of algorithms for deterministically traversing along an automaton.

- **Proposition 5.** Given a pair \((u, v)\) \(\in \Sigma^+\) and an FDF A \(\mathcal{F}\) of size \((n, k)\), the membership question, of whether \((u, v)\ \in [\mathcal{F}]\), can be resolved in deterministic space of \(O(\log n + \log k)\).

The next questions to consider are those of emptiness and universality, namely given an FDF A \(\mathcal{F} = (Q, P)\), whether \([\mathcal{F}] = \emptyset\), and whether \([\mathcal{F}] = \Sigma^+\), respectively. Notice that the universality problem is equivalent to the emptiness problem over the complement of \(\mathcal{F}\). For nondeterministic automata, the complement automaton might be exponentially larger than the original one, making the universality problem much harder than the emptiness problem. Luckily, FDF complementation is done in constant space, as is the case with deterministic automata, making the emptiness and universality problems equally easy.

The emptiness problem for an FDF A \((Q, P)\) cannot be resolved by only checking whether there is a nonempty progress automaton in \(P\), since it might be that the accepted period \(v\) is not part of any normalized pair. Yet, the existence of a prefix-word \(x\) and a period-word \(y\), such that \(Q(x) = Q(xy)\) and \(P_{Q(x)}\) accepts \(y\) is a sufficient and necessary criterion for the nonemptiness of \(\mathcal{F}\). This can be tested in NLOGSPACE. Hardness in NLOGSPACE follows by a reduction from graph reachability [8].

- **Theorem 6.** Emptiness and universality for FDFAs are NLOGSPACE-complete.

The complexity for equality and containment is easily derived from that of emptiness, intersection and complementation.

- **Proposition 7.** Equality and containment for saturated FDFAs are NLOGSPACE-complete.

**Saturation check**

All of the operations and decision problems above assumed that the given FDFAs are saturated. This is indeed the case when learning FDFAs via the algorithm of [1], and when translating \(\omega\)-automata to FDFAs (see Section 5). We show below that the decision problem of whether an arbitrary FDFA is saturated is in PSPACE. We leave the question of whether it is PSPACE-complete open.

- **Theorem 8.** The problem of deciding whether a given FDFA is saturated is in PSPACE.

**Proof Sketch.** We first show that if an FDF A \(\mathcal{F}\) of size \((n, k)\) is unsaturated then there exist words \(u, v', v''\) such that \(|u| \leq n\) and \(|v'\), \(|v''| \leq n^a k^{2k}\), and non-negative integers \(l, r \leq k\) such that \((u, (v'v''))' \in \mathcal{F}\) while \((w, (v'v''))' \notin \mathcal{F}\).

Let \(Q, P,\) and \(P'\) be the leading automaton and two relevant progress automata, with state spaces \(Q, P\) and \(P'\), respectively. We achieve the bound on the length of \(v'\) and \(v''\), by considering for every word \(v \in \Sigma^*\), the function \(\chi_v\) from \((Q, P, P')\) to \((Q, P, P')\) defined as \(\chi_v(q, p, p') = (\delta_{Q}(q, v), \delta_{P}(p, v), \delta_{P'}(p', v))\). Note that there are up to \(n^a k^{2k}\) different such functions. Hence, if \(v'\) and \(v''\) are longer than \(n^a k^{2k}\), we can replace them with shorter words that are completely equivalent w.r.t. \(Q, P,\) and \(P'\).

A coNPSpace algorithm guesses integers \(l, r \leq k\) and, letter by letter, some words \(u, v', v''\) such that \(|u| \leq n\) and \(|v'|, \, |v''| \leq n^a k^{2k}\). Along the way, it constructs \(\chi_{v'v''}\) and \(\chi_{v'v''}\).
It then verifies whether one of \((\chi\omega v')^l\) and \((\chi\omega v')^r\), applied to the relevant initial states, is accepting and the other is not. By Savitch’s and Immerman–Szelepcsényi’s theorems, the problem is in \(PSPACE\).

\section{Translating To and From \(\omega\)-Automata}

As two \(\omega\)-regular languages are equivalent iff they agree on the set of ultimately periodic words \([11]\), an \(\omega\)-regular language can be characterized by a language of pairs of finite words, and in particular by a saturated \(\text{fdFA}\). We shall write \(L \equiv L'\) to denote that a language \(L \subseteq \Sigma^\ast\) characterizes an \(\omega\)-regular language \(L'\). Formally:

\begin{definition}
A language \(L \subseteq \Sigma^\ast\) characterizes an \(\omega\)-regular language \(L' \subseteq \Sigma^\omega\), denoted by \(L \equiv L'\), if for every pair \((u, v) \in L\), we have \(uv^\omega \in L'\), and for every ultimately periodic word \(uv^\omega \in L'\), we have \((u, v) \in L\).
\end{definition}

The families of DFAs defined in \([9]\), as well as the analogous families of right congruences of \([10]\), are known to characterize exactly the set of \(\omega\)-regular languages \([9, 10]\). This is also the case with our definition of saturated \(\text{fdFA}\).

\begin{theorem}
Every saturated \(\text{fdFA}\) characterizes an \(\omega\)-regular language, and for every \(\omega\)-regular language, there is a saturated \(\text{fdFA}\) characterizing it.
\end{theorem}

\begin{proof}
The two directions are proved in Theorems 12 and 16, below.
\end{proof}

\subsection{From \(\omega\)-Automata to FDFAs}

We show that DBA, DCA, and DPA have polynomial translations to saturated FDFAs, whereas translation of NBAs to FDFAs may involve an inevitable exponential blowup.

\textbf{From deterministic \(\omega\)-automata.} The constructions of a saturated \(\text{fdFA}\) that characterize a given DBA, DCA, or DPA \(D\) share the same idea: The leading automaton is equivalent to \(D\), except for ignoring the acceptance condition, and each progress automaton consists of several copies of \(D\), memorizing the acceptance level of the period-word. For a DBA or a DCA, two such copies are enough, memorizing whether or not a Büchi (co-Büchi) accepting state was visited. For a DPA with \(k\) colors, \(k\) such copies are required.

\begin{theorem}
Let \(D\) be a DBA or a DCA with \(n\) states. There exists a saturated \(\text{fdFA}\) \(F\) of size \((n, 2n)\), such that \([F] \equiv [D]\).
\end{theorem}

\begin{theorem}
Let \(D\) be a DPA with \(n\) states and \(k\) colors. There exists a saturated \(\text{fdFA}\) \(F\) of size \((n, kn)\), such that \([F] \equiv [D]\).
\end{theorem}

\textbf{From nondeterministic \(\omega\)-automata.} An NBA \(A\) can be translated into a saturated \(\text{fdFA}\) \(F\), by first determinizing \(A\) into an equivalent DPA \(A'\) \([18, 7]\) (which might involve a \(2^{O(n \log n)}\) state blowup and \(O(n)\) colors \([23]\)), and then polynomially translating \(A'\) into an equivalent \(\text{fdFA}\) (Theorem 12).

\begin{proposition}
Let \(B\) be an NBA with \(n\) states. There is a saturated \(\text{fdFA}\) that characterizes \([B]\) with a leading automaton and progress automata of at most \(2^{O(n \log n)}\) states each.
\end{proposition}
A \(2^{O(n \log n)}\) state blowup in this case is inevitable, based on the lower bound for complementing NBAs [12, 26, 22], the constant complementation of FDFAs, and the polynomial translation of a saturated FDDA to an NBA:

\[ \text{Lemma 15.} \quad \text{Michel [12] has shown that there exists a family of languages \(\{L_n\}_{n \geq 1}\), such that for every } n \in \mathbb{N}, \text{ a saturated FDDA that characterizes } [B_n] \text{ must be of size } (m, k), \text{ such that } \max(m, k) \geq 2^{O(n \log n)}. \]

\[ \text{Proof.} \quad \text{Theorem 16.} \quad \text{Michel [12] has shown that there exists a family of languages \(\{L_n\}_{n \geq 1}\), such that for every } n \in \mathbb{N}, \text{ a saturated FDDA that characterizes } [B_n] \text{ must be of size } (m, k), \text{ such that } \max(m, k) \geq 2^{O(n \log n)}. \]

\[ \text{Proof.} \quad \text{Michel [12] has shown that there exists a family of languages \(\{L_n\}_{n \geq 1}\), such that for every } n \in \mathbb{N}, \text{ a saturated FDDA that characterizes } [B_n] \text{ must be of size } (m, k), \text{ such that } \max(m, k) \geq 2^{O(n \log n)}. \]

5.2 From FDFAs to \(\omega\)-Automata

We show that saturated FDFAs can be polynomially translated into NBAs, yet translations of saturated FDFAs to DPAs may involve an inevitable exponential blowup.

\[ \text{To nondeterministic } \omega \text{-automata.} \quad \text{We show below that every saturated FDDA can be polynomially translated to an equivalent NBA. Since an NBA can be viewed as a special case of an NPA, a translation of saturated FDFAs to NPAs follows. Translating saturated FDFAs to NCAs is not always possible, as the latter are not expressive enough.} \]

The translation goes along the lines of the construction given in [4] for translating an \(L_\omega\)-automaton into an equivalent NBA. We prove below that the construction is correct for saturated FDFAs, despite the fact that saturated FDFAs can be exponentially smaller than \(L_\omega\)-automata.

We start with a lemma from [4], which will serve us for one direction of the proof.

\[ \text{Lemma 15 ([4]).} \quad \text{Let } M, N \subseteq \Sigma^* \text{ such that } M \cdot N^* = M \text{ and } N^+ = N. \text{ Then for every ultimately periodic word } w \in \Sigma^* \text{ we have that } w \in M \cdot N^* \text{ iff there exist words } u \in M \text{ and } v \in N \text{ such that } uv^\omega = w. \]

We continue with the translation and its correctness.

\[ \text{Theorem 16.} \quad \text{For every saturated FDDA } \mathcal{F} \text{ of size } (n, k), \text{ there exists an NBA } \mathcal{B} \text{ with } O(n^2k^3) \text{ states, such that } [\mathcal{F}] \equiv [\mathcal{B}]. \]

\[ \text{Proof.} \quad \text{Construction:} \quad \text{Consider a saturated FDDA } \mathcal{F} = (\mathcal{Q}, \mathcal{P}), \text{ where } \mathcal{Q} = (\Sigma, Q, \iota, \delta), \text{ and for each state } q \in Q, \mathcal{P} \text{ has the progress DFA } \mathcal{P}_q = (\Sigma, P_q, \iota_q, \delta_q, F_q). \]

For every \(q \in Q\), let \(M_q\) be the language of finite words on which \(Q\) reaches \(q\), namely \(M_q = \{ u \in \Sigma^* \mid Q(u) = q \}\). For every \(q \in Q\) and for every accepting state \(f \in F_q\), let \(N_{q,f}\) be the language of finite words on which \(Q\) makes a self-loop on \(q\), \(P_q\) reaches \(f\), and \(P_q\) makes a self-loop on \(f\), namely \(N_{q,f} = \{ v \in \Sigma^* \mid (\delta(q, v) = q) \land (P_q(v) = f) \land (\delta_q(f, v) = f) \}\). We define the \(\omega\)-regular language

\[ L = \bigcup_{\{(q,f) \mid (q \in Q) \land (f \in F_q)\}} M_q \cdot N_{q,f}^\omega \quad (1) \]

One can construct an NBA \(\mathcal{B}\) that recognizes \(L\) and has up to \(O(n^2k^3)\) states.
Families of DFAs as Acceptors of $\omega$-Regular Languages

**Correctness:** Consider an ultimately periodic word $uw^\omega \in \mathcal{B}$. By the construction of $\mathcal{B}$, $uw^\omega \in L$, where $L$ is defined by Equation (1). Hence, $uw^\omega \in M_q \cdot N_{q,f}$, for some $q \in Q$ and $f \in F_q$. By the definitions of $M_q$ and $N_{q,f}$, we get that $M_q$ and $N_{q,f}$ satisfy the hypothesis of Lemma 15, namely $N_{q,f}^+ = N_{q,f}$ and $M_q \cdot N_{q,f}^+ = M_q$. Therefore, by Lemma 15, there exist finite words $u' \in M_q$ and $v' \in N_{q,f}$ such that $u'v^\omega = uw^\omega$. From the definitions of $M_q$ and $N_{q,f}$, it follows that the run of $Q$ on $u'$ ends in the state $q$, and $P_q$ accepts $v'$. Furthermore, by the definition of $N_{q,f}$, we have $\delta(q,v') = q$, implying that $(u',v')$ is the normalization of itself. Hence, $(u',v') \in \mathcal{F}$. Since $\mathcal{F}$ is saturated and $u'v^\omega = uw^\omega$, it follows that $(u,v) \in \mathcal{F}$, as required.

As for the other direction, consider a pair $(u,v) \in \mathcal{F}$, and let $(x,y)$ be the normalization of $(u,v)$ w.r.t. $Q$. We will show that $xy^\omega \in L$, where $L$ is defined by Equation (1), implying that $uw^\omega \in \mathcal{B}$. Let $q = Q(x)$, so we have that $P_q(y)$ reaches some accepting state $f$ of $P_q$. Note, however, that it still does not guarantee that $y \in N_{q,f}$, since it might be that $\delta_q(f,y) \neq f$.

To prove that $xy^\omega \in L$, we will show that there is a pair $(x,y') \in \Sigma^+$ and an accepting state $f' \in P_q$, such that $y' = y^t$ for some positive integer $t$, and $y' \in N_{q,f'}$; namely $\delta(q,y') = q$, $P_q(y') = f'$, and $\delta_q(f',y') = f'$. Note first that since $\mathcal{F}$ is saturated, it follows that for every positive integer $i$, $(x,y^i) \in \mathcal{F}$, as $x(y^i)^\omega = xy^\omega$.

Now, for every positive integer $i$, $P_q$ reaches some accepting state $f_i$ when running on $y^i$. Since $P_q$ has finitely many states, for a large enough $i$, $P_q$ must reach the same accepting state $f$ twice when running on $y^i$. Let $h$ be the smallest positive integer such that $P_q(y^h) = f$, and $r$ the smallest positive integer such that $\delta_q(f',y^r) = f$. Now, one can verify that choosing $t$ to be an integer that is bigger than or equal to $h$ and is divisible by $r$ guarantees that $\delta(q,y^t) = q$ and $\delta_q(f',y^r) = f'$, where $f' = P_q(y^r)$. \hfill \blacktriangleleft

**To deterministic $\omega$-automata.** Deterministic Büchi and co-Büchi automata are not expressive enough for recognizing every $\omega$-regular language. We thus consider the translation of saturated DFAs to deterministic parity automata. A translation is possible by first polynomially translating the DFA into an NBA (Theorem 16) and then determinizing the latter into a DPA (which might involve a $2^{O(n \log n)}$ state blowup [12]).

**Proposition 17.** Let $\mathcal{F}$ be a saturated D DFA of size $(n,k)$. There exists a DPA $\mathcal{D}$ of size $2^{O(n^{2k^2} \log n^{2k^2})}$, such that $\mathcal{F} \equiv \mathcal{D}$.

We show below that an exponential state blowup is inevitable. The family of languages $\{L_n\}_{n \geq 1}$ below demonstrates the inherent gap between DFAs and D PAs; an DFA for $L_n$ may only “remember” the smallest and biggest read numbers among $\{1,2,\ldots,n\}$, using $n^2$ states, while a DPA for it must have at least $2^{n-1}$ states.

We define the family of languages $\{L_n\}_{n \geq 1}$ as follows. The alphabet of $L_n$ is $\{1,2,\ldots,n\}$, and a word belongs to it if and only if the following two conditions are met:

- A letter $i$ is always followed by a letter $j$, such that $j < i + 1$. For example, 53245 $\ldots$ is a bad prefix, since 2 was followed by 4, while 55234122 $\ldots$ is a good prefix.
- The number of letters that appear infinitely often is odd. For example, 2331(22343233)$^\omega$ is in $L_n$, while 1(233)$^\omega$ is not.

We show below how to construct, for every $n \geq 1$, a saturated D DFA of size polynomial in $n$ that characterizes $L_n$.

**Lemma 18.** Let $n \geq 1$. There is a saturated DFA of size $(n+1,n^2)$ characterizing $L_n$. 


Proof Sketch. The leading automaton handles the safety condition of $L_n$, having $n + 1$ states, and ensuring that a letter $i$ is always followed by a letter $j$, such that $j \leq i + 1$. The progress automata, which are identical, maintain the smallest and biggest number-letters that appeared, denoted by $s$ and $b$, respectively. Since a number-letter $i$ cannot be followed by a number-letter $j$, such that $j > i + 1$, it follows that the total number of letters that appeared is equal to $b - s + 1$. Then, a state is accepting iff $b - s + 1$ is odd. □

A DPA for $L_n$ cannot just remember the smallest and largest letters that were read, as these letters might not appear infinitely often. Furthermore, we prove below that the DPA must be of size exponential in $n$, by showing that its state space must be doubled when moving from $L_n$ to $L_{n+1}$.

Lemma 19. Every DPA that recognizes $L_n$ must have at least $2^{n-1}$ states.

Proof. The basic idea behind the proof is that the DPA cannot mix between 2 cycles of $n$ different letters each. This is because a mixed cycle in a parity automaton is accepting/rejecting if its two sub-cycles are, while according to the definition of $L_n$, the mixed cycle should reject if both its sub-cycles accept, and vice versa. Hence, whenever adding a letter, the state space must be doubled.

In the formal proof below, we dub a reachable state from which the automaton can accept some word a live state. Consider a DPA $D_n$ that recognizes $L_n$, and let $q$ be some live state of $D_n$. Observe that $[D_n^q]$, namely the language of the automaton that we get from $D_n$ by changing the initial state to $q$, is the same as $L_n$ except for having some restriction on the word prefixes. More formally, if a word $w \in [D_n^q]$ then $w \in L_n$, and if $w \in L_n$ then there is a finite word $u$, such that $uw \in [D_n^q]$. For every $n \in \mathbb{N}$, and every $u \in \Sigma^*$, let $L_{n,u} = \{w \mid uw \in L_n\}$ and let $\mathcal{L}_n$ denote the set of languages $\{L_{n,u} \mid u \in \Sigma^*\}$. Note that there is actually only a finite number of prefixes $u$ to consider (this follows e.g. from [10, Thm. 22]). Moreover, for every state $q$ of $D_n$ there is a corresponding word $u_q$ such that $[D_n^q] = L_{n,u_q}$.

We prove by induction on $n$ the following claim, from which the statement of the lemma immediately follows: Let $D_n$ be a DPA over $\Sigma = \{1, 2, \ldots, n\}$ that recognizes some language in $\mathcal{L}_n$. Then there are finite words $u, v \in \Sigma^*$, such that:

(i) $v$ contains all the letters in $\Sigma$;
(ii) the run of $D_n$ on $u$ reaches some live state $p$; and
(iii) the run of $D_n$ on $v$ from $p$ returns to $p$, while visiting at least $2^{n-1}$ different states.

The base cases, for $n \in \{1, 2\}$, are trivial, as they mean a cycle of size at least 1 over $v$, for $n = 1$, and a cycle of size 2 for $n = 2$.

In the induction step, for $n \geq 2$, we consider a DPA $D_{n+1}$ that recognizes some language $L \in \mathcal{L}_{n+1}$. We shall define $D'$ and $D''$ to be the DPAs that result from $D_{n+1}$ by removing all the transitions over the letter $n + 1$ and by removing all the transitions over the letter 1, respectively.

Observe that for every state $q$ that is live w.r.t. $D_{n+1}$, we have that $[D'^q] \in \mathcal{L}_n$, namely the language of the DPA that results from $D_{n+1}$ by removing all the transitions over the letter $n + 1$ and setting the initial state to $q$ is in $\mathcal{L}_n$. (Note that $q$ might only be reachable via the letter $n + 1$, yet it must have outgoing transitions over letters in $[2, n]$.) Analogously, $[D''q]$ is isomorphic to a language in $\mathcal{L}_n$ via the alphabet mapping of $i \rightarrow (i - 1)$. Hence, for every state $q$ that is live w.r.t. $D_{n+1}$, the induction hypothesis holds for $D'^q$ and $D''q$.

We shall prove the induction claim by describing words $u, v \in \Sigma^*$, and showing that they satisfy the requirements above w.r.t. $D_{n+1}$. We construct $u$ by iteratively concatenating the words $u'_1, v'_1, u''_1, v''_1$, which we define below, until the starting and ending states in
Families of DFAs as Acceptors of $\omega$-Regular Languages

some iteration $k$ are the same. We then define the word $v$ to be the last iteration, namely $u'_k \cdot v'_k \cdot u''_k \cdot v''_k$. Let $q_1$ be the initial state of $D_{n+1}$. We define for every $i \in [1..k]$:

- $u'_i$ and $v'_i$ are the words that follow from the induction hypothesis on $D^{\omega i}$, where $q_i$ is the state that $D_{n+1}$ reaches when reading $u'_1 \cdot v'_1 \cdot u''_1 \cdot v''_1 \ldots u'_{i-1} \cdot v'_{i-1} \cdot u''_{i-1} \cdot v''_{i-1}$.
- $u''_i$ and $v''_i$ are the words that follow from the induction hypothesis on $D^{\omega i}$, where $q'_i$ is the state that $D_{n+1}$ reaches when reading $u'_1 \cdot v'_1 \cdot u''_1 \cdot v''_1 \ldots u'_{i-1} \cdot v'_{i-1} \cdot u''_{i-1} \cdot v''_{i-1} \cdot u'_i \cdot v'_i$.

The word $v$ obviously contains all the letters in $\Sigma$, as it is composed of subwords that contain all the letters in $\Sigma \setminus \{1\}$ and in $\Sigma \setminus \{n+1\}$. By the definition of $u$ and $v$, we also have that the run of $D_{n+1}$ on $u$ reaches some live state $p$, and the run of $D_{n+1}$ on $v$ from $p$ returns to $p$. Now, we need to prove that the run of $D_{n+1}$ on $v$ from $p$ visits at least $2^n$ states.

We claim that when $D_{n+1}$ runs on $v$ from $p$, it visits disjoint set of states when reading $v'_k$ and $v''_k$. This will provide the required result, as $D_{n+1}$ visits at least $2^{n-1}$ states when reading each of $v'_k$ and $v''_k$.

Assume, by way of contradiction, that $D_{n+1}$ visits some state $s$ both when reading $v'_k$ and when reading $v''_k$. Let $l'$ and $r'$ be the parts of $s_k$ that $D_{n+1}$ reads before and after reaching $s$, respectively, and $l''$ and $r''$ the analogous parts of $v''_k$. Now, define the infinite words $m' = u \cdot u'_k \cdot (l' \cdot r')^\omega$, $m'' = u \cdot u''_k \cdot (l'' \cdot r'')^\omega$, and $m = u \cdot u''_k \cdot (l'' \cdot r'')^\omega$.

Observe that $m'$ and $m''$ both belong or both do not belong to $L$, since there is the same number of letters (of length) that appear infinitely often in each of them. The word $m$, on the other hand, belongs to $L$ if $m'$ and $m''$ do not belong to $L$, and vice versa, since $n+1$ letters appear infinitely often in it. However, the set of states that are visited infinitely often in the run of $D_{n+1}$ on $m$ is the union of the sets of states that appear infinitely often in the runs of $D_{n+1}$ on $m'$ and $m''$. Thus, if $D_{n+1}$ accepts both $m'$ and $m''$ it also accepts $m$, and if it rejects both $m'$ and $m''$ it rejects $m$. (This follows from the fact that the minimal number in a union of two sets is even/odd if the minimum within both sets is even/odd.) Contradiction. ▲

Theorem 20. ³ There is a family of languages $\{L_n\}_{n \geq 1}$ over the alphabet $\{1, 2, \ldots, n\}$, such that for every $n \geq 1$, there is a saturated $\text{FDFA}$ of size $(n + 1, n^2)$ that characterizes $L_n$, while a $\text{DPA}$ for $L_n$ must be of size at least $2^{n-1}$.

Proof. By Lemmas 18 and 19. ▲

6 Discussion

The interest in $\text{FDFA}$ as a representation for $\omega$-regular languages stems from the fact that they possess a correlation between the automaton states and the language right congruences, a property that traditional $\omega$-automata lack. This property is beneficial in the context of

³ A small adaptation to the proof of Lemma 19 shows an inevitable exponential blowup also when translating a saturated $\text{FDFA}$ to a deterministic $\omega$-automaton with a stronger acceptance condition of Rabin [19] or Streett [24]: A mixed cycle in a Rabin automaton is rejecting if its two sub-cycles are, and a mixed cycle in a Streett automaton is accepting if its two sub-cycles are. Hence, the proof of Lemma 19 holds for both Rabin and Streett automata if proceeding in the induction step from an alphabet of size $n$ to an alphabet of size $n + 2$, yielding a Rabin/Streett automaton of size at least $2^7$. As for translating a saturated $\text{FDFA}$ to a deterministic Muller automaton [13], it is known that translating a $\text{DFA}$ of size $n$ into a deterministic Muller automaton might require the latter to have an accepting set of size exponential in $n$ [21]. (The family of languages in [21] uses an alphabet of size exponential in the number of states of the $\text{DFA}$, however it can easily be changed to use an alphabet of linear size.) Hence, by Theorem 12, which shows a polynomial translation of $\text{DPAs}$ to $\text{FDFA}$s, we get that translating an $\text{FDFA}$ to a deterministic Muller automaton entails an accepting set of exponential size, in the worst case.
learning, and indeed an algorithm for learning \( \omega \)-regular languages by means of saturated DFAs was recently provided [1]. Analyzing the succinctness of saturated DFAs and the complexity of their Boolean operations and decision problems, we believe that they provide an interesting formalism for representing \( \omega \)-regular languages. Indeed, Boolean operations and decision problems can be performed in nondeterministic logarithmic space and their succinctness lies between deterministic and nondeterministic \( \omega \)-automata.

References

Families of DFAs as Acceptors of $\omega$-Regular Languages


