# On Repetition Languages


## Orna Kupferman
School of Engineering and Computer Science, Hebrew University, Jerusalem, Israel
orna@cs.huji.ac.il

## Ofer Leshkowitz
School of Engineering and Computer Science, Hebrew University, Jerusalem, Israel
ofer.leshkowitz@mail.huji.ac.il



## Abstract

A regular language $R$ of finite words induces three *repetition languages* of infinite words: the language $\lim(R)$, which contains words with infinitely many prefixes in $R$, the language $\infty R$, which contains words with infinitely many disjoint subwords in $R$, and the language $R^\omega$, which contains infinite concatenations of words in $R$. Specifying behaviors, the three repetition languages provide three different ways of turning a specification of a finite behavior into an infinite one. We study the expressive power required for recognizing repetition languages, in particular whether they can always be recognized by a deterministic Büchi word automaton (DBW), the blow up in going from an automaton for $R$ to automata for the repetition languages, and the complexity of related decision problems. For $\lim R$ and $\infty R$, most of these problems have already been studied or are easy. We focus on $R^\omega$. Its study involves some new and interesting results about additional repetition languages, in particular $R^\#$, which contains exactly all words with unboundedly many concatenations of words in $R$. We show that $R^\omega$ is DBW-recognizable iff $R^\#$ is $\omega$-regular iff $R^\# = R^\omega$, and there are languages for which these criteria do not hold. Thus, $R^\omega$ need not be DBW-recognizable. In addition, when exists, the construction of a DBW for $R^\omega$ may involve a $2^{O(n \log n)}$ blow-up, and deciding whether $R^\omega$ is DBW-recognizable, for $R$ given by a nondeterministic automaton, is PSPACE-complete. Finally, we lift the difference between $R^\#$ and $R^\omega$ to automata on finite words and study a variant of Büchi automata where a word is accepted if (possibly different) runs on it visit accepting states unboundedly many times.



**2012 ACM Subject Classification** Automata on Infinite Words

**Keywords and phrases** Büchi automata, Expressive power, Succinctness

**Digital Object Identifier** 10.4230/LIPIcs.MFCS.2020.


## 1 Introduction

Finite *automata on infinite objects* were first introduced in the 60's, and were the key to the solution of several fundamental decision problems in mathematics and logic [6, 14, 17]. Today, automata on infinite objects are used for specification, verification, and synthesis of nonterminating systems. The automata-theoretic approach reduces questions about systems and their specifications to questions about automata [11, 22], and is at the heart of many algorithms and tools. Industrial-strength property-specification languages such as the IEEE 1850 Standard for Property Specification Language (PSL) [7] include regular expressions and/or automata, making specification and verification tools that are based on automata even more essential and popular.

One way to classify an automaton is by the type of its branching mode, namely whether it is *deterministic*, in which case it has a single run on each input word, or *nondeterministic*, in which case it may have several runs, and the input word is accepted if at least one of them is accepting. A run of an automaton on finite words is accepting if it ends in an accepting state. A run of an automaton on infinite words does not have a final state, and acceptance is determined with respect to the set of states visited infinitely often during the run. Another way to classify an automaton on infinite words is the class of its acceptance condition. For example, in *Büchi* automata, some of the states are designated as accepting states, and a run is accepting iff it visits states from the accepting set infinitely often [6].

The different classes of automata have different *expressive power*. For example, unlike automata on finite words, where deterministic and nondeterministic automata have the same expressive power, deterministic Büchi automata (DBWs) are strictly less expressive than nondeterministic Büchi automata (NBWs). That is, there exists a language $L$ over infinite words such that $L$ can be recognized by an NBW but cannot be recognized by a DBW. The different classes also differ in their *succinctness*. For example, while translating a nondeterministic automaton on finite words (NFW) into a deterministic one (DFW) is always possible, the translation may involve an exponential blow-up [19].

There has been extensive research on expressiveness and succinctness of automata on infinite words [21, 9]. Beyond the theoretical interest, the research has received further motivation with the realization that many algorithms, like synthesis and probabilistic model checking, need to operate on deterministic automata [5, 4], as well as the discovery that many natural specifications correspond to DBWs. In particular, it is shown in [10] that given a *linear temporal logic* (LTL) formula $\psi$, there is an *alternation-free $\mu$-calculus* (AFMC) formula equivalent to $\forall\psi$ iff $\psi$ can be recognized by a DBW. Since AFMC is as expressive as weak alternating automata and the weak monadic second-order theory of trees [18, 16, 3], this relates DBWs also with them.

Proving that NBWs are more expressive than DBWs, Landweber characterized languages that are DBW-recognizable as these that are the *limit* of some regular language on finite words. Formally, for an alphabet $\Sigma$ and a language $R \subseteq \Sigma^*$, we define $\lim(R)$ as the set of infinite words in $\Sigma^\omega$ that have infinitely many prefixes in $R$. For example, if $R = (0+1)^* \cdot 0$, namely the set of finite words over $\{0,1\}$ that end with a 0, then $\lim(R) = ((0+1)^* \cdot 0)^\omega$, namely the set of words with infinitely many 0's. On the other hand, we cannot point to a language $R$ such that $\lim(R)$ is the set of all words with only finitely many 0's. Landweber proved that a language $L \subseteq \Sigma^\omega$ is DBW-recognizable iff there is a regular language $R$ such that $L = \lim(R)$ [12].

Beyond the limit operator, another natural way to obtain a language of infinite words from a language $R$ of finite words is to require the words in $R$ to repeat infinitely often. This actually induces two "repetition languages". The first is $\infty R$, where $w \in \infty R$ iff $w$ contains infinitely many disjoint subwords in $R$. Formally, $\infty R = \{\Sigma^* \cdot w_1 \cdot \Sigma^* \cdot w_2 \cdot \Sigma^* \cdot w_3 \cdots : \ w_i \in R \text{ for all } i \geq 1 \}$. The second is $R^\omega$, where $w \in R^\omega$ iff $w$ is an infinite concatenation of words in $R$. Formally, $R^\omega = \{w_1 \cdot w_2 \cdot w_3 \cdots : \ w_i \in R \text{ for all } i \geq 1\}$. For example, for the language $R = (0+1)^* \cdot 0$ above, we have $\lim(R) = \infty R = R^\omega = ((0+1)^* \cdot 0)^\omega$. In order to see that the three repetition languages may be different, consider the language $R = 0 \cdot (0+1)^* \cdot 0$, namely of all words that start and end with 0. Now, $\lim(R) = 0 \cdot ((0+1)^* \cdot 0)^\omega$, $\infty R = ((0+1)^* \cdot 0)^\omega$, and $R^\omega = 0 \cdot ((0+1)^* \cdot 0 \cdot 0)^\omega$. When specifying on-going behaviors, the three repetition languages induce three different ways for turning a finite behavior into an infinite one. For example, if $R = \mathit{call} \cdot \mathit{true}^* \cdot \mathit{return}$ describes a sequence of events that starts with a call and ends with a return, then $\lim R$ describes behaviors that start with a call followed by infinitely many returns, $\infty R$ behaviors with infinitely many calls and returns, and $R^\omega$ behaviors with infinitely many successive calls and returns.

In this paper we study expressiveness, succinctness, and complexity of repetition languages. We start with expressiveness, where we examine which of the repetition languages are $\omega$-regular, and for those that are $\omega$-regular, whether they are also DBW-recognizable. By [12], for $\lim(R)$ the answer is positive – it is DBW-recognizable for all regular languages $R$. For a *finite* regular language $R$, we show that $R^\omega = \lim(R^*)$, implying a positive answer too. Our main result is a negative answer in the general $R^\omega$ case: we point to a regular language $R$ such that $R^\omega$ is not DBW-recognizable. In order to find such a language, we study repetition languages in general, and introduce the language $R^\# = \{w \in \Sigma^\omega :$ for all $i \geq 1$ there exists a prefix of $w$ in $R^i\}$, namely the language of exactly all words with unboundedly many concatenations of words in $R$. As detailed below, $R^\#$ is strongly related to $R^\omega$ and turns out to be also strongly related to our question. We show that when $R^\#$ is $\omega$-regular, then $R^\# = R^\omega$, in which case, by Landweber's characterization of DBW-recognizable languages as countable intersections of open sets in the product topology over $\Sigma^\omega$, both are DBW-recognizable. In other words, we show (Theorem 5) that $R^\#$ is $\omega$-regular iff $R^\# = R^\omega$ iff $R^\omega$ is DBW-recognizable.

The above characterization enables us to point to a language $R$ that does not satisfy the three criteria (Theorem 9). In short, $R = \$ + (0 \cdot \{0, 1, \$\}^* \cdot 1)$. It is easy to see that for every word $w \in R^\omega$, if $w$ contains infinitely many 1's, then $w$ contains infinitely many 0's. Hence, the word $w = 011\$1\$\$1\$\$\$1\$\$\$\$ \cdots = 0 \cdot \prod_{k=0}^{\infty} 1\$^k$ is not in $R^\omega$, yet for all $i \geq 1$, its prefix $0 \cdot \prod_{k=0}^{i} 1\$^k = (0 \cdot (\prod_{k=0}^{i-1} 1\$^k) \cdot 1) \cdot \$^i$ is in $R^i$, and so $w \in R^\#$. It follows that $w \in R^\# \setminus R^\omega$, which by our characterization implies that $R^\omega$ is not DBW-recognizable. We also study the problem of deciding, given an NFW $\mathcal{A}$, whether $L(\mathcal{A})^\omega$ is DBW-recognizable, and show that it is PSPACE-complete. We lift the difference between $R^\#$ and $R^\omega$ to automata on finite words and define the *#-language* of a Büchi automaton $\mathcal{A}$ as the set of words $w$ such that for all $i \geq 1$, there is a run of $\mathcal{A}$ on $w$ that visits the set of accepting states at least $i$ times. We show that the #-language of $\mathcal{A}$ is $\omega$-regular iff the #-language of $\mathcal{A}$ coincides with its $\omega$-regular language, iff $L(\mathcal{A})$ is DBW-recognizable.

We continue and study the size of automata for the repetition languages. We consider the cases $R$ is given by a DFW or an NFW, and the automaton for the repetition language is a DBW or an NBW. By [12], going from a DFW for $R$ to a DBW for $\lim(R)$ involves no blow up – we only have to view the DFW as a Büchi automaton. We show that the cases of $\infty R$ and $R^\omega$ are more complicated, and involve a $2^{O(n)}$ and a $2^{O(n \log n)}$ blow-up, respectively. Beyond the relevancy to our study, the family of languages we use is a new witness to the known lower bound for NBW determinization [13]. The succinctness analysis for the cases the automata for the repetition languages are nondeterministic are much easier, as we show that, except for the case of $\lim(R)$, simple constructions with no blow-ups are possible, even when we start with an NFW for $R$. For the case of $\lim(R)$, going from an NFW for $R$ to an NBW for $\lim(R)$ is not trivial and the best known upper bound is $O(n^3)$ [2]. Our results are summarized in Section 7.

# 2 Preliminaries

## 2.1 Automata

An *alphabet* $\Sigma$ is a finite set of letters. A *word* over $\Sigma$ is a finite or infinite sequence $w = \sigma_1, \sigma_2, \sigma_3, \cdots$ of letters from $\Sigma$. We use $|w|$ to denote the length of $w$, with $|w| = \infty$ for an infinite word $w$. For $1 \leq i \leq |w|$, we use $w[i]$ to denote $\sigma_i$, that is, the $i$-th letter in $w$, and for $1 \leq i \leq j \leq |w|$, we use $w[i, j]$ to denote the *infix* $\sigma_i, \sigma_{i+1}, \cdots, \sigma_j$ of $w$. We use $\Sigma^*$ and $\Sigma^\omega$ to denote the set of all finite and infinite words over $\Sigma$, respectively. For two words $x \in \Sigma^*$ and $y \in \Sigma^* \cup \Sigma^\omega$, we use $x \cdot y$ to denote the *concatenation* of $x$ and $y$. We say that $x$ is a *prefix* of a $w$, denoted $x \prec w$, if there is $1 \leq i \leq |w|$ such that $x = w[1, i]$. Equivalently, if $x \neq \varepsilon$, and there is $y \in \Sigma^* \cup \Sigma^\omega$, such that $x \cdot y = w$. Thus, $y = [i + 1, |w|]$, and we call it a *suffix* of $w$. Note that we do not consider the empty word $\varepsilon$ as a prefix of a word.

A *nondeterministic automaton* is $\mathcal{A} = \langle \Sigma, Q, \delta, Q_0, \alpha \rangle$, where $\Sigma$ is a finite input alphabet, $Q$ is a finite set of states, $\delta : Q \times \Sigma \to 2^Q$ is a transition function, $Q_0 \subseteq Q$ is a set of initial states, and $\alpha \subseteq Q$ is an acceptance condition. Intuitively, $\delta(q, \sigma)$ is the set of states $\mathcal{A}$ may move to when reading the letter $\sigma$ from state $q$. Formally, a *run* of $\mathcal{A}$ on a word $w$ is the function $r : \{i \in \mathbb{N}_0 : 0 \leq i \leq |w|\} \to Q$, such that $r(0) \in Q_0$, i.e., the run starts from an initial state, and for all $i \geq 0$, we have that $r(i+1) \in \delta(r(i), \sigma_{i+1})$, i.e., the run obeys the transition function. Note that as $\mathcal{A}$ may have several initial states and the transition function may specify several possible successor states, the automaton $\mathcal{A}$ may have several runs on $w$. If $|Q_0| = 1$ and for all $q \in Q$ and $\sigma \in \Sigma$, it holds that $|\delta(q, \sigma)| = 1$, then $\mathcal{A}$ has a single run on $w$, and we say that $\mathcal{A}$ is *deterministic*. We sometimes refer to a run also as a sequence of states; that is, $r = r(0), r(1), \ldots \in Q^{|w|+1}$.

When $\mathcal{A}$ runs on finite words, the run $r$ is finite, and it is *accepting* iff it ends in an accepting state, thus $r(|w|) \in \alpha$. When $\mathcal{A}$ runs on infinite words, acceptance depends on the set $\inf(r)$, of the states that $r$ visits infinitely often. Formally $\inf(r) = \{q \in Q :$ for infinitely many $i \in \mathbb{N}$, we have that $r(i) = q\}$. As $Q$ is finite, the set $\inf(r)$ is guaranteed not to be empty. In *Büchi* automata, the run $r$ is *accepting* iff $\inf(r) \cap \alpha \neq \emptyset$. Otherwise, $r$ is *rejecting*. The automaton $\mathcal{A}$ accepts a word $w$ if there exists an accepting run $r$ of $\mathcal{A}$ on $w$. The *language* of $\mathcal{A}$, denoted $L(\mathcal{A})$, is the set of words that $\mathcal{A}$ accepts. We also say that $\mathcal{A}$ *recognizes* $L(\mathcal{A})$.

We use three letter acronyms in $\{\text{D,N}\} \times \{\text{F,B}\} \times \{\text{W}\}$ to denote classes of word automata. The first letter indicates whether the automaton is deterministic or nondeterministic, and the second letter indicates whether it is an automaton on finite words or a Büchi automaton on infinite words. For example, DBW is a deterministic Büchi automaton.

Throughout the paper, we use $R$ and $L$ to represent languages of finite and infinite words, respectively. A language $R \subseteq \Sigma^*$ is *finite* if $|R| < \omega$, where $|R|$ is the cardinality of $R$ as a set. A language $R \subseteq \Sigma^*$ is *regular* if there is an NFW that recognizes $R$. Likewise, a language $L \subseteq \Sigma^\omega$ is $\omega$-*regular* if there is an NBW that recognizes $L$. We sometimes refer to the three-letter acronyms as describing sets of languages, thus NBW is also the set of $\omega$-regular languages, and DBW is its subset of languages recognizable by DBW.

## 2.2 Repetition languages

Consider a language $R \subseteq \Sigma^*$, and assume $\varepsilon \notin R$. We refer to $R$ as the *base language* and define the following *repetition languages* of words induced by $R$. We start with languages of finite words:

1. For $i \geq 0$, we define $R^i = \{w_1 \cdot w_2 \cdots w_i : w_j \in R \text{ for all } 1 \leq j \leq i\}$.
2. $R^* = \bigcup_{i \geq 0} R^i$.
3. $R^+ = \bigcup_{i \geq 1} R^i$.

We continue with languages of infinite words:

3. $\lim(R) = \{w \in \Sigma^\omega : \; w[1,i] \in R \text{ for infinitely many } i\text{'s}\}$.
4. $\infty R = \{\Sigma^* \cdot w_1 \cdot \Sigma^* \cdot w_2 \cdot \Sigma^* \cdot w_3 \cdots : \; w_i \in R \text{ for all } i \geq 1 \}$.
5. $R^\omega = \{w_1 \cdot w_2 \cdot w_3 \cdots : \; w_i \in R \text{ for all } i \geq 1\}$.
6. $R^\# = \{w \in \Sigma^\omega : \text{for all } i \geq 1 \text{ there exists } j \geq 1 \text{ such that } w[1,j] \in R^i\}$.

Thus, $R^i, R^*, R^+$, and $R^\omega$ are the standard bounded, finite, finite and positive, and infinite concatenation operators. Then, $\lim(R)$ contains exactly all infinite words with infinitely many prefixes in $R$, and $\infty R$ contains exactly all infinite words with infinitely many disjoint infixes in $R$. Finally, $R^\#$ contains exactly all words with prefixes with unboundedly many concatenations of words in $R$. The language $R^\#$ may seem equivalent to $R^\omega$, and the difference between $R^\omega$ and $R^\#$ is in fact one of our main results.

▶ **Example 1.** Let $R = (0+1)^* \cdot 0$. Then, $\lim(R) = \infty R = R^\omega = R^\# = \infty 0$.

Now, let $R = \{0^n \cdot 1^m, 1^n \cdot 0^m : 0 \leq m \leq n\}$. While $R$ is not regular, we have that $\lim(R) = \{0^\omega, 1^\omega\}$ and $\infty(R) = R^\omega = R^\# = \{0,1\}^\omega$ are in DBW.

Finally, for all $R \subseteq \Sigma^*$, we have $R^\omega \subseteq \lim(R^*)$ and $R^\omega \subseteq \infty R$. Thus, $R^\omega \subseteq \lim(R^*) \cap \infty R$. One may suspect that $R^\omega = \lim(R^*) \cap \infty R$. As a counterexample, consider $R = 0 \cdot (0+1)^* \cdot 0$.

Then, $R^\omega = 0 \cdot ((0+1)^* \cdot 0 \cdot 0)^\omega$, $\lim(R^*) = 0 \cdot ((0+1)^* \cdot 0)^\omega$, and $\infty R = ((0+1)^* \cdot 0)^\omega = \infty 0$. Thus, the word $0 \cdot (1 \cdot 0)^\omega$ is in $\lim(R^*) \cap (\infty R)$ but is not in $R^\omega$. ◀

As another warm up, we state the following lemma, which would be helpful in the sequel.

▶ **Lemma 2.** *Consider languages $R \subseteq \Sigma^*$ and $P \subset \Sigma^\omega$ such that $\varepsilon \notin R$. If $P \subseteq R \cdot P$, then $P \subseteq R^\omega$.*

**Proof.** Consider a word $w_0 \in P$. Since $P \subseteq R \cdot P$, then $w_0 = x_1 \cdot w_1$, for some $x_1 \in R$ and $w_1 \in P$. Have defined $x_1, \ldots, x_i \in R$ and $w_i \in P$, such that $w_0 = x_1 \cdots x_i \cdot w_i$, we can continue and define $x_{i+1} \in R$ and $w_{i+1} \in P$ such that $w_i = x_{i+1} \cdot w_{i+1}$. Overall, we have defined $\{x_i\}_{i=1}^\infty \subseteq R$ such that $w_0 = x_1 \cdot x_2 \cdot x_3 \cdots$. Hence, $w_0 \in R^\omega$, and we are done. ◀

Note that if $\varepsilon \in R$, then $P \subseteq R \cdot P$ trivially holds for all $P \subseteq \Sigma^\omega$, whereas possibly $P \nsubseteq R^\omega$. Also, if $\varepsilon \in R$, then $\infty R$, $R^\omega$, and $R^\#$ as defined above include also finite words, in particular $\varepsilon$ is a member of all of those languages. In order to circumvent the technical issues that the above entails, for $R \subseteq \Sigma^*$ such that $\varepsilon \in R$, we define $\infty R = \infty(R \setminus \{\varepsilon\})$, $R^\omega = (R \setminus \{\varepsilon\})^\omega$, and $R^\# = (R \setminus \{\varepsilon\})^\#$, and accordingly assume, throughout the paper, that $\varepsilon \notin R$.

We conclude the preliminaries with the case the base language $R$ is finite. As we shall see, then $R^\omega = R^\# = \lim(R^*)$, implying that they are all in DBW.

▶ **Theorem 3.** *For every finite language $R \subseteq \Sigma^*$, we have that $R^\omega = R^\# = \lim(R^*)$.*

**Proof.** Consider a finite language $R \subseteq \Sigma^*$. We prove that $R^\omega \subseteq \lim(R^*) \subseteq R^\# \subseteq R^\omega$. First, it is easy to see, regardless of $R$ being finite, that $R^\omega \subseteq \lim(R^*)$.

We prove that $\lim(R^*) \subseteq R^\#$. Clearly $R^{n+1} \cdot \Sigma^\omega \subseteq R^n \cdot \Sigma^\omega$, and thus we only need to show that for all $w \in \lim(R^*)$, we have that $w \in R^n \cdot \Sigma^\omega$ for infinitely many $n$'s. Since $R$ is finite, there exists some $k \geq 1$ such that for all $x \in R$, we have that $|x| \leq k$. It follows that for all $x \in R^*$, if $|x| \geq m \cdot k$, then $x \in R^n$ for some $n \geq m$. Consider some word $w \in \lim(R^*)$. By definition, $w$ has infinitely many prefixes in $R^*$, thus for all $m \geq 1$, there exists a prefix $x \in R^*$ of $w$ such that $|x| \geq m \cdot k$. Hence, $x \in R^n$ for some $n \geq m$, implying that $w \in R^n \cdot \Sigma^\omega$ for infinitely many $n$'s, and we are done.

It is left to prove that $R^\# \subseteq R^\omega$. Consider a word $w \in R^\#$. Intending to use König's Lemma, we build a tree with set of nodes $V = \{(x, i) : x \prec w \text{ and } x \in R^i\}$. Since $w \in R^\#$, the set $V$ is infinite. As the parent of a node $(x, i+1) \in V$, we set some $(y, i) \in V$ that satisfies $x = y \cdot z$ for some $z \in R$. Since $x \in R^{i+1}$, such a prefix $y$ exists. Note that there might be several $y$'s and only a single $(y, i)$ is chosen to be the parent of $(x, i+1)$. Observe that all nodes $(x, i)$ are connected to $(\varepsilon, 0)$ by a single path of length $i$, and thus we have defined an infinite tree above $V$. The out degree of each node is bounded by $|R| < \infty$. Hence, by König's Lemma, the tree has an infinite path $\pi = \langle(\varepsilon, 0), (x_1, 1), (x_2, 2), \ldots\rangle$. By construction, for all $i \geq 0$ there exists some $y_i$ such that $x_{i+1} = x_i \cdot y_i$ and $y_i \in R$. It follows that $w = y_1 \cdot y_2 \cdots$, and hence $w \in R^\omega$, and we are done. ◀

For every regular language $R \subseteq \Sigma^*$, the language $R^*$ is regular. Hence, by [12], the language $\lim(R^*)$ is in DBW, and so Theorem 3 implies the following.

▶ **Corollary 4.** *For every finite language $R \subseteq \Sigma^*$, we have that $R^\omega$ and $R^\#$ are in DBW.*

As we shall see in Section 3, the case of an infinite base language $R$ is much more difficult.

## 3 Expressiveness

In this section we examine which of the repetition languages are $\omega$-regular, and for these that are $\omega$-regular, whether they are also DBW-recognizable. Note that going in the other direction need not be possible. For example, the language $L = 0 \cdot 1^\omega$ is DBW-recognizable, but there is no regular language $R$ such that $L = \infty R$, $L = R^\#$, or $L = R^\omega$. By [12], a language $L \subseteq \Sigma^\omega$ is in DBW iff there exists a regular language $R \subseteq \Sigma^*$ such that $L = \lim(R)$. In particular, this means that for every $R \subseteq \Sigma^*$ regular, we have that $\lim(R) \in$ DBW. We study this question for $\infty R$, $R^\omega$, and $R^\#$.

It is well known that for every regular language $R$, the language $R^\omega$ is $\omega$-regular. This follows, for example, from the translation of $\omega$-regular expressions to NBWs. Studying whether $R^\omega$ is always DBW-recognizable is much harder, and is out main result:

▶ **Theorem 5.** *For all regular languages $R \subseteq \Sigma^*$, the following are equivalent.*
*(1)* $R^\omega = R^\#$*.*
*(2)* $R^\omega$ *is in DBW.*
*(3)* $R^\#$ *is $\omega$-regular.*

The proof of Theorem 5 is partitioned into Lemmas 6, 7, and 8.

▶ **Lemma 6.** [**(1)** $\to$ **(2)** **and** **(3)**] *If $R^\omega = R^\#$, then $R^\omega$ is in DBW and $R^\#$ is $\omega$-regular.*

**Proof.** By Landweber's Theorem [12], an $\omega$-regular language $L$ is in DBW iff $L$ is a countable intersection of open sets in the product topology over $\Sigma^\omega$, induced by the discrete topology over $\Sigma$. Specifically, the topology that is induced by the basis $\mathcal{B} = \{N_x : x \in \Sigma^*\}$, where $N_x = x \cdot \Sigma^\omega$. That is, $A \subseteq \Sigma^\omega$ is an open set in the product topology if there is a $B \subseteq \Sigma^*$ such that $A = \cup_{x \in B} N_x = B \cdot \Sigma^\omega$. Equivalently, the topology induced by the metric $d : \Sigma^\omega \times \Sigma^\omega \to \mathbb{R}_{\geq 0}$, defined $d(x, y) = \frac{1}{2^n}$, where $n$ is the first position that $x$ and $y$ differ, and $d(x, y) = 0$, if $x = y$. That is, $A \subseteq \Sigma^\omega$ is an open set if for all $x \in A$ there exists $\gamma > 0$ such that $\{y : d(x, y) < \gamma\} \subseteq A$.

As discussed above, an open set is a set of the form $K \cdot \Sigma^\omega$ for some $K \subseteq \Sigma^*$. Thus, Landweber's Theorem states that an $\omega$-regular language $L$ is in DBW iff there exists $\{K_i\}_{i \in \mathbb{N}}$, $K_i \subseteq \Sigma^*$, such that $L = \bigcap_i K_i \cdot \Sigma^\omega$. By definition, the language $R^\#$ fulfills the topological condition in Landweber's Theorem. Hence, if $R^\#$ is $\omega$-regular, then $R^\#$ is in DBW.

Since $R$ is regular, the language $R^\omega$ is $\omega$-regular. Thus, $R^\# = R^\omega$ is $\omega$-regular, and by the above, both are also in DBW. ◀

▶ **Lemma 7.** [**(2)** $\to$ **(1)**] *If $R^\omega$ is in DBW, then $R^\omega = R^\#$.*

**Proof.** Since, by definition, $R^\omega \subseteq R^\#$, we only have to prove that $R^\# \subseteq R^\omega$. Assume that $\mathcal{A} = \langle \Sigma, Q, q_0, \delta, \alpha \rangle$ is a DBW for $R^\omega$. Let $n = |Q|$. Consider a word $w \in R^\#$, and let $r : \mathbb{N} \to Q$ be the run of $\mathcal{A}$ on $w$. Let $t$ be a position from which $r$ is contained in $\inf(r)$, i.e., for all $t' \geq t$, we have that $r(t') \in \inf(r)$. Let $w = w_1 \cdot w_2 \cdots w_{t+n} \cdot y$ be a partition of $w$ to words such that for all $1 \leq j \leq t+n$, we have that $w_j \in R$. Since $w \in R^\#$, such a partition exists. Let $q_j = r(|w_1 \cdots w_j|)$, i.e., $q_j$ is the state $\mathcal{A}$ reaches when reading the prefix $w_1 \cdots w_j$. Observe that since there are only $n$ states, there must exist indices $j_1$ and $j_2$ such that $t \leq j_1 < j_2 \leq t+n$ and $q_{j_1} = q_{j_2}$.

Consider the word $w' = w_1 \cdot w_2 \cdots w_{j_1} \cdot (w_{j_1+1} \cdots w_{j_2})^\omega$, and let $r'$ be the run of $\mathcal{A}$ on $w'$. Since all the words $w_j$ are in $R$, then $w' \in R^\omega$, and so $\inf(r') \cap \alpha \neq \emptyset$. Moreover, since $|w_1 \cdots w_{j_1}| \geq t$ and $w_{j_1+1} \cdots w_{j_2}$ closes a cycle from $q_{j_1}$, then $\inf(r') \subseteq \inf(r)$. Hence, $\inf(r) \cap \alpha \neq \emptyset$. Thus, the run of $\mathcal{A}$ on $w$ is accepting, and so $w \in R^\omega$. ◀

▶ **Lemma 8.** [**(3)** $\to$ **(1)**] *If $R^\#$ is $\omega$-regular, then $R^\omega = R^\#$.*

**Proof.** Since, by definition, $R^\omega \subseteq R^\#$, we only have to prove that $R^\# \subseteq R^\omega$. Let $R \subseteq \Sigma^*$ be such that $R^\#$ is $\omega$-regular. Then, as $R^\omega$ is $\omega$-regular, so is $K = R^\# \setminus R^\omega$. Let $\mathcal{A}$ be an NBW with $n$ states for $K$. Assume by way of contradiction that $L(\mathcal{A}) \neq \varnothing$. There exist some accepting state $q$ that is reachable from both an initial state by a path labeled with some $u \in \Sigma^*$, and from itself by a cycle labeled with some $v \in \Sigma^*$. Thus, the word $w = u.v^\omega$ is a lasso-shaped word in $L(\mathcal{A})$. Let $x$ be a prefix of $w$ with $x \in R^{|u|+|v|}$, and let $x = y_0.y_1...y_{|v|}$ be a partition of $x$ such that $y_0 \in R^{|u|}$ and $y_i \in R$ for all $i > 0$. Note that $|y_0| \geq |u|$, thus for $i > 0$, the $y_i$'s are nonempty subwords of $\{v\}^+$. For $0 \leq i \leq |v|$, let $k_i$ be the position in $v$ that is reached after reading $y_0.y_1...y_i$. I.e., $k_i = j$, for $0 \leq j \leq |v| - 1$, such that $y_0...y_i = u.v^t.v[1,j]$ for some $t \geq 0$. For example, if $y_0 = u.v$, then $k_0 = 0$, and if $y_0.y_1 = u.v.v.v[1,2]$, then $k_1 = 2$. Since $0 \leq k_i \leq |v| - 1$ for all $0 \leq i \leq |v|$, there are indices $i$ and $j$ such that $i < j$, and $k_i = k_j$. Therefore, there exist $t_1, t_2 \geq 0$ such that the following hold:

1. $z_1 = y_0...y_i = u.v^{t_1}.v[1,k_i] \in R^+$, and
2. $z_2 = y_{i+1}...y_j = v[k_i+1,|v|].v^{t_2}.v[1,k_j] = v[k_i+1,|v|].v^{t_2}.v[1,k_i] \in R^+$.

Clearly, $z_1.(z_2)^\omega \in R^\omega$. Also, $(z_2)^\omega = v[k_i+1,n].v^\omega$, thus $z_1.(z_2)^\omega = u.v^\omega = w$. Thus, $w \in R^\omega$, contradicting the assumption that $L(\mathcal{A}) = R^\# \setminus R^\omega$. Hence, $L(\mathcal{A}) = \varnothing$; thus $R^\# \subseteq R^\omega$. ◀

This completes the proof of Theorem 5. We now show that the theorem is not trivial, thus there is a language $R$ that does not satisfy the three criteria in the theorem, in particular the criteria about DBW, which is our main interest.

▶ **Theorem 9.** *There exists a regular language $R \subseteq \Sigma^*$, such that $R^\omega$ is not in DBW.*

**Proof.** We define the regular language $R \subseteq \{0,1,\$\}^*$ by the regular expression $R = (\$ + 0 \cdot \{0,1,\$\}^* \cdot 1)$. It is easy to see that for every word $w \in R^\omega$, if $w$ contains infinitely many 1's, then $w$ contains infinitely many 0's. Indeed, the only way to have only finitely many 1's in a word in $R^\omega$ is to have an infinite tail of \$'s. Hence, the word

$$w = 011\$1\$\$1\$\$\$1\$\$\$\$1\$\$\$\$\$\ldots = 0 \cdot \prod_{i=0}^{\infty} 1\$^i$$

is not in $R^\omega$. We prove that $w \in R^\#$. For $n \in \mathbb{N}$, consider the word $w_n = 0 \cdot \prod_{i=0}^{n} 1\$^i = (0 \cdot (\prod_{i=0}^{n-1} 1\$^i) \cdot 1) \cdot \$^n$. It is easy to see that $w_n \in R^{n+1}$. Since all of the $w_n$'s are prefixes of $w$, it follows that $w \in R^\#$.

Thus, $w \in R^\# \setminus R^\omega$, implying that $R^\# \neq R^\omega$. Then, by Theorem 5, we have that $R^\omega$ is not in DBW, and we are done. ◀

▶ **Corollary 10.** *For every regular language $R \subseteq \Sigma^*$, we have that $R^\omega$ is $\omega$-regular. Yet, $R^\omega$ need not be in DBW, and $R^\#$ need not be $\omega$-regular.*

We continue to $\infty R$, showing it is an easy special case of $R^\omega$. Given a regular language $R \subseteq \Sigma^*$, let $P = \Sigma^* \cdot R$. It is easy to see that $\infty R = P^\omega$. As we argue below, the special form of $P$ implies it satisfies all the three criteria in Theorem 5:

▶ **Theorem 11.** *For every regular language $R \subseteq \Sigma^*$, we have that $(\Sigma^* \cdot R)^\# = (\Sigma^* \cdot R)^\omega$.*

**Proof.** Let $P = \Sigma^* \cdot R$. We prove that $P^\# \subseteq P \cdot P^\#$. By Lemma 2, the latter implies that $P^\# = P^\omega$. Consider a word $w \in P^\#$, and let $x_0 \prec w$ be a word of minimal length such that $x_0 \in P$. Let $w' \in \Sigma^\omega$ be such that $w = x_0 \cdot w'$. We prove that $w' \in P^\#$, implying that $w \in P \cdot P^\#$. For all $i \geq 1$, let $x_i \cdot y_i \prec w$, with $x_i \in P$ and $y_i \in P^i$. Note that by the minimality of $x_0$, it holds that $x_0 \prec x_i$ for all $i \geq 1$. Now, for all $i \geq 1$, let $z_i \in \Sigma^*$ be the suffix of $x_i$, with $x_i = x_0 \cdot z_i$, and consider $u_i = z_i \cdot y_i \prec w'$. Observe that for all $i \geq 1$ we have $u_i \in \Sigma^* \cdot P^i = (\Sigma^* \cdot R) \cdot P^{i-1} = P^i$. Hence, $w' \in P^\#$. ◀

▶ **Corollary 12.** *For every regular language $R \subseteq \Sigma^*$, the language $\infty R$ is in DBW.*

## 4 Complexity

In this section we study the complexity of deciding, given an NFW $\mathcal{A}$, whether $L(\mathcal{A})^\omega$ is DBW-recognizable. We first describe a simple linear translation of an NFW for $R$ to an NBW for $R^\omega$.

▶ **Theorem 13.** *For every NFA $\mathcal{A}$ with $n$ states, there exists an NBW $\mathcal{A}'$ with $O(n)$ states such that $L(\mathcal{A}') = L(\mathcal{A})^\omega$.*

**Proof.** Consider an NFW $\mathcal{A} = \langle Q, \Sigma, q_0, \delta, \alpha \rangle$, and let $R = L(\mathcal{A})$. For simplicity, we assume that $\mathcal{A}$ has a single initial state. We construct an NBW $\mathcal{A}'$ for $R^\omega$. We define $\mathcal{A}' = \langle Q', \Sigma, Q'_0, \delta', \alpha' \rangle$, where $Q' = Q \cup \{q'_0\}$, for some state $q'_0 \notin Q$, and $Q'_0 = \alpha' = \{q'_0\}$. Intuitively, we simulate a run of $\mathcal{A}$ and allow non-deterministic "jumps" from states in $\alpha$ to $q_0$. We accept words for which the simulation makes infinitely many jumps. The positions of the jumps partition the input word into a concatenation of infinitely many words in $R$. The NBW $\mathcal{A}'$ implements a "jump" by introducing a new state $q'_0$, which replaces $q_0$ and inherits its outgoing transitions. In addition, whenever $\delta$ may move to a state in $\alpha$, the NBW $\mathcal{A}'$ may move to $q'_0$ instead. Consequently, closing a loop from $q'_0$ corresponds to reading a word in $R$. Thus, visiting $q'_0$ infinitely many times corresponds to reading a word in $R^\omega$. Formally, the transition function $\delta'$ is defined, for every $q \in Q'$ and $\sigma \in \Sigma$, as follows.

$$\delta'(q,\sigma) = \begin{cases} \delta(q,\sigma) & \text{if } q \neq q'_0 \text{ and } \delta(q,\sigma) \cap \alpha = \varnothing, \\ \delta(q,\sigma) \cup \{q'_0\} & \text{if } q \neq q'_0 \text{ and } \delta(q,\sigma) \cap \alpha \neq \varnothing, \\ \delta(q_0,\sigma) & \text{if } q = q'_0 \text{ and } \delta(q_0,\sigma) \cap \alpha = \varnothing, \\ \delta(q_0,\sigma) \cup \{q'_0\} & \text{if } q = q'_0 \text{ and } \delta(q_0,\sigma) \cap \alpha \neq \varnothing. \end{cases}$$

In Appendix A.1 we prove that indeed $L(\mathcal{A}') = R^\omega$. ◀

▶ **Theorem 14.** *Deciding whether $L(\mathcal{A})^\omega$ is DBW-recognizable, for an NFW $\mathcal{A}$, is PSPACE-complete.*

**Proof.** We start with the upper bound. As described in the proof of Theorem 13, given an NFW $\mathcal{A}$ with $n$ states, we can construct an NBW for $L(\mathcal{A})^\omega$ with $n+1$ states. By [10], deciding whether the language of a given NBW is DBW-recognizable can be done in PSPACE. Hence, membership in PSPACE for our result follows.

For the lower bound, we describe a logspace reduction from the universality problem for NFWs, proved to be PSPACE-hard in [15]. For two alphabets $\Sigma_1$ and $\Sigma_2$, and two words $w_1 \in \Sigma_1^\omega$ and $w_2 \in \Sigma_2^\omega$, let $w_1 \oplus w_2 \in (\Sigma_1 \times \Sigma_2)^\omega$ be the word obtained by merging $w_1$ and $w_2$. Formally, if $w_1 = \sigma_1^1 \cdot \sigma_2^1 \cdot \sigma_3^1 \cdots$ and $w_2 = \sigma_1^2 \cdot \sigma_2^2 \cdot \sigma_3^2 \cdots$, then $w_1 \oplus w_2 = \langle \sigma_1^1, \sigma_1^2 \rangle \cdot \langle \sigma_2^1, \sigma_2^2 \rangle \cdot \langle \sigma_3^1, \sigma_3^2 \rangle \cdots$. We use the operator $\oplus$ also for merging two finite words $w_1 \in \Sigma_1^*$ and $w_2 \in \Sigma_1^*$ of the same length. Note that then, $|w_1 \oplus w_2| = |w_1| = |w_2|$.

Consider an NFW $\mathcal{A}$ over some alphabet $\Sigma$, and assume $\perp \notin \Sigma$. Consider the language $R = \$^* + 0 \cdot \{0,1,\$\}^* \cdot 1$. Note that $R$ is similar to the language used in the proof of Theorem 9 – here we include in $R$ words in $\$^*$. This does not change $R^\#$ or $R^\omega$, and the word $0 \cdot \prod_{i=0}^{\infty} 1 \cdot \$^i$ is in $R^\# \setminus R^\omega$, witnessing that $R^\omega$ is not DBW-recognizable.

We define the language $R_\mathcal{A}$ over the alphabet $(\Sigma \cup \{\perp\}) \times \{0,1,\$\}$ as follows.

$$R_\mathcal{A} = \{(w_1 \cdot \perp) \oplus w_2 : w_1 \in L(\mathcal{A}) \text{ or } w_2 \in R\}.$$

Note that since NFWs for $R$ and for $(\Sigma \cup \{\perp\})^* \cdot \perp$ are of a fixed size, the size of an NFW for $R_\mathcal{A}$ is linear in the size of $\mathcal{A}$ and it can be constructed from $\mathcal{A}$ in logspace. We prove that

$L(\mathcal{A}) = \Sigma^*$ iff $R^\omega_\mathcal{A} \in$ DBW. First, observe that if $L(\mathcal{A}) = \Sigma^*$, then $R^\omega_\mathcal{A} = (\infty\perp) \oplus \{0, 1, \$\}^\omega$, and so $R^\omega_\mathcal{A} \in$ DBW. For the other direction, assume that $L(\mathcal{A}) \neq \Sigma^*$, and consider a word $x \in \Sigma^* \setminus L(\mathcal{A})$. Let $w_x = (x \cdot \perp)^\omega$. Observe that for every partition $y_1 \cdot y_2 \cdot y_3 \cdots$ of $w_x$ into subwords with $y_i \in (\Sigma \cup \{\perp\})^* \cdot \perp$, for all $i \geq 1$, it must be that $y_i \notin L(\mathcal{A}) \cdot \perp$ for all $i \geq 1$. It follows that for every $w \in \{0, 1, \$\}^\omega$, if $w_x \oplus w \in R^\omega_\mathcal{A}$, then $w \in R^\omega$.

Let $m = |x \cdot \perp|$, and consider the word $w = 0^m \cdot \prod_{i=0}^\infty 1^m \cdot \$^{im}$, obtained from $0 \cdot \prod_{i=0}^\infty 1 \cdot \$^i$ by replacing each letter $\sigma \in \{0, 1, \$\}$ by the word $\sigma^m$. Using the same arguments used in the proof of Theorem 9, we have that $w \notin R^\omega$. Hence, $w_x \oplus w \notin R^\omega_\mathcal{A}$.

We prove that $w_x \oplus w \in R^\#_\mathcal{A}$. Note that $w_x \oplus w = (x \cdot \perp)^\omega \oplus 0^m \cdot \prod_{i=0}^\infty 1^m \cdot \$^{im} = ((x \cdot \perp) \oplus 0^m) \cdot \prod_{i=0}^\infty ((x \cdot \perp) \oplus 1^m) \cdot ((x \cdot \perp) \oplus \$^m)^i$. For all $j \geq 1$, we have $((x \cdot \perp) \oplus \$^m)^j) \in R^j_\mathcal{A}$, and $((x \cdot \perp) \oplus 0^m) \cdot (\prod_{i=0}^{j-1} ((x \cdot \perp) \oplus 1^m) \cdot ((x \cdot \perp) \oplus \$^m)^i) \cdot ((x \cdot \perp) \oplus 1^m) \in R_\mathcal{A}$. Hence, $y^j = ((x \cdot \perp) \oplus 0^m) \cdot \prod_{i=0}^{j} ((x \cdot \perp) \oplus 1^m) \cdot ((x \cdot \perp) \oplus \$^m)^i \in R^{j+1}_\mathcal{A}$. Since $y^j \prec w_x \oplus w$, for all $j \geq 1$, we conclude that $w_x \oplus w \in R^\#_\mathcal{A}$.

Thus, $w_x \oplus w \in R^\#_\mathcal{A} \setminus R^\omega_\mathcal{A}$, and so, by Theorem 5, we have that $R^\omega_\mathcal{A} \notin$ DBW. ◀

## 5 Succinctness

In this section we study the blow-up in going from an automaton for $R$ to automata for $\lim(R)$, $\infty R$, and $R^\omega$. Note that, by Theorem 5, a DBW for $R^\omega$ is also a DBW for $R^\#$, and thus we do not consider $R^\#$ explicitly.

Studying succinctness, we also refer to the *Rabin* acceptance condition. There, $\alpha = \{\langle G_1, B_1 \rangle, \dots, \langle G_k, B_k \rangle\} \subseteq 2^Q \times 2^Q$, and a run $r$ is accepting iff there is a pair $\langle G, B \rangle \in \alpha$ such that $\inf(r) \cap G \neq \emptyset$ and $\inf(r) \cap B = \emptyset$. We use DRW to denote deterministic Rabin word automata. By [8], DRWs are *Büchi type*: if a DRW $\mathcal{A}$ recognizes a DBW-recognizable language, then a DBW for $L(\mathcal{A})$ can be defined on top of $\mathcal{A}$. In other words, if $L(\mathcal{A})$ is in DBW, then we can obtain a DBW for $L(\mathcal{A})$ by redefining the acceptance condition of $\mathcal{A}$.

Our study of succinctness considers the cases $R$ is given by a DFW or an NFW, and the automaton for the repetition language is DBW, DRW, or NBW. We start with the case both automata are deterministic. Then, the case of $\lim(R)$ is easy and well known: Given a DFW $\mathcal{A}$ for $R$, viewing $\mathcal{A}$ as a DBW results in an automaton for $\lim(R)$ [12]. Hence, there is no blow-up in going from a DFW for $R$ to a DBW for $\lim(R)$. We continue to the case of $\infty R$. We first consider the case we are given an NFW or DFW for $\Sigma^* \cdot R$.

▶ **Theorem 15.** *For every regular language $R \subseteq \Sigma^*$, there is no blow-up in going from an NFW (DFW) for $\Sigma^* \cdot R$ to an NBW (resp. DBW) for $\infty R$.*

**Proof.** Let $\mathcal{A} = \langle Q, \Sigma, \delta, q_0, \alpha \rangle$ be an NFW with a single initial state that recognizes $\Sigma^* \cdot R$. We define an NBW $\mathcal{A}'$ for $(\Sigma^* \cdot R)^\omega = \infty R$ as follows. Intuitively, $\mathcal{A}'$ simulates a run of $\mathcal{A}$, each time the simulation reaches a state in $\alpha$ it "restarts" the simulation, and it accepts an infinite word iff simulation has been restarted infinitely often. The partition to successful simulations also partitions accepted words to infixes in $L(\mathcal{A})^\omega$, thus accepted words are in $\infty R$. In addition, if a word is in $\infty R$, then a word in $\Sigma^* \cdot R$ start in all positions, implying that a successful simulation is always eventually completed. Formally, $\mathcal{A}' = \langle Q, \Sigma, \delta', q_0, \alpha \rangle$, where $\delta'$ is defined for all $q \in Q$ and $\sigma \in \Sigma$ as follows:

$$\delta'(q, \sigma) = \begin{cases} \delta(q, \sigma) : & q \notin \alpha, \\ \delta(q_0, \sigma) : & q \in \alpha. \end{cases}$$

In Appendix A.2, we prove that $L(\mathcal{A}') = (\Sigma^* \cdot R)^\omega = \infty R$. Note that since $\varepsilon \notin R$, then $q_0 \notin \alpha$. Also, note that when $\mathcal{A}$ is deterministic, so is $\mathcal{A}'$. ◀

Going from a DFW for $R$ to a DFW for $\Sigma^* \cdot R$ may involve an exponential blow-up. To see this, consider for example the language $R = 0 \cdot (0+1)^n$. While it can be recognized by a DFW with $n+2$ states, a DFA for $(0+1)^* \cdot 0 \cdot (0+1)^n$ needs at least $2^n$ states. Theorem 16 shows that this blow-up is inherited to the construction of a DBW for $\infty R$.

▶ **Theorem 16.** *The blow-up in going from a DFW for $R$ to a DBW for $\infty R$ is $2^{O(n)}$.*

**Proof.** For the upper bound, starting with a DFW with $n$ states for $R$, one can construct an NFW with $n+1$ states for $\Sigma^* \cdot R$. Its determinization results in a DFW with $2^{n+1}$ states for $\Sigma^* \cdot R$. Then, by Theorem 15, we end up with a DBW with $2^{n+1}$ states for $\infty R$.

For the lower bound, we describe a family of languages $R_1, R_2, \ldots$ of finite words, such that for all $n \geq 1$, the language $R_n$ can be recognized by a DFW with $O(n)$ states, yet a DBW for $\infty R_n$ needs at least $\frac{2^{n-1}}{n}$ states.

Let $\Sigma = \{0,1\}$. For $n \geq 1$, we define $R_n \subseteq \Sigma^*$ as the set of words of length $n+1$ that start and end with the same letter. That is, $R_n = \{\sigma \cdot w \cdot \sigma : \text{for } \sigma \in \Sigma \text{ and } w \in \Sigma^{n-1}\}$. Equivalently, $R_n = 0 \cdot (0+1)^{n-1} \cdot 0 + 1 \cdot (0+1)^{n-1} \cdot 1$. It is easy to see that $R_n$ can be recognized by a DFW with $2n+3$ states. In Appendix A.3 we prove that a DBW for $\infty R_n$ needs at least $\frac{2^{n-1}}{n}$ states. ◀

We continue to $R^\omega$. While it is easy, given a DFW for $R$, to construct an NBW for $R^\omega$ (see Theorem 13), staying in the deterministic model is complicated, and not only in terms of expressive power. Formally, we have the following.

▶ **Theorem 17.** *The blow-up in going from a DFW for $R$ to a DBW for $R^\omega$, when exists, is $2^{O(n \log n)}$.*

**Proof.** For the upper bound, one can determinize the NBW for $R^\omega$. Thus, starting with a DFW with $n$ states for $R$, we construct an NBW with $n+1$ states for $R^\omega$, and determinize it to a DRW with $2^{O(n \log n)}$ states [20]. Since DRWs are Büchi type, the result follows.

For the lower bound, we describe a family of languages $R_1, R_2, \ldots$ of finite words, such that for all $n \geq 1$, the language $R_n$ can be recognized by a DFW with $\mathrm{O}(n)$ states, $R_n^\omega$ is in DBW, yet a DBW for $R_n^\omega$ needs at least $n!$ states.

Given $n \geq 1$, let $\Sigma_n = [n] \cup \{\#\}$, where $[n] = \{1, \ldots, n\}$. We define the language $R_n \subseteq \Sigma_n^*$ as the set of all finite words that start and end with the same letter from $[n]$. That is, $R_n = \{\sigma \cdot x \cdot \sigma : \text{for } x \in \Sigma_n^* \text{ and } \sigma \in [n]\}$. It is easy to see that $R_n$ is regular and a DFW for $R_n$ needs $2n+1$ states. In Appendix A.4, we prove that $R_n^\omega$ is in DBW, and that a DBW for $R_n^\omega$ needs at least $n!$ states. ◀

Since DRWs are Büchi type, Lemma 23 and Theorems 16 and 17 imply the following.

▶ **Theorem 18.** *The blow-ups in going from a DFW with $n$ states for $R$ to DRWs for $\infty R$ and $R^\omega$ are $2^{O(n)}$ and $2^{O(n \log n)}$, respectively.*

The succinctness analysis for case the automaton for the repetition languages is non-deterministic is much easier, as the construction described above involve no blow-up, and except for the case of $\lim(R)$, they are valid also when $R$ is given by an NFW. The case of $\lim(R)$ is more complicated and is studied in [2]. It is easy to see that just viewing an NFW for $R$ as a Büchi automaton does not result in an NBW for $\lim(R)$. For example, an NFW for $(0+1) \cdot 0$ that guesses whether each 0 is the last letter, in which case it moves to an accepting state with no successors, is empty when viewed as an NBW. The best known construction of an NBW for $\lim(R)$ from an NFW $\mathcal{A}$ for $R$ is based on a characterization of the limit of $L(\mathcal{A})$ as the union of languages, each associated with a state $q$ of $\mathcal{A}$ and containing words that have infinitely many prefixes whose accepting run reaches $q$. Following this characterization, it is possible to construct, starting with $\mathcal{A}$ with $n$ states, an NBW with $O(n^3)$ states for $\lim(R)$ [2].

# 6 On Unboundedly Many vs. Infinitely Many

Essentially, the definition of $R^{\#}$ replaces the "infinite" nature of $R^{\omega}$ by an "unbound" one. In this section we examine an analogous change in the definition of acceptance in Büchi automata. Consider a nondeterministic automaton $\mathcal{A} = \langle \Sigma, Q, \delta, Q_0, \alpha \rangle$. When we view $\mathcal{A}$ as a $\#$-*automaton*, it accepts a word $w \in \Sigma^{\omega}$ if for all $i \geq 0$, there is a run of $\mathcal{A}$ on $w$ that visits $\alpha$ at least $i$ times. Formally, for for all $i \geq 0$, there is a run $r_i = q_0^i, q_1^i, q_2^i, \ldots$ of $\mathcal{A}$ on $w$ such that $r_j^i \in \alpha$ for at least $i$ indices of $j \geq 0$. The $\#$-*language* of $\mathcal{A}$, denoted $L_{\#}(\mathcal{A})$, is the set of words that $\mathcal{A}$ accepts as above. We use the notations $L_F(\mathcal{A})$ and $L_B(\mathcal{A})$ to refer the languages of $\mathcal{A}$ when viewed as an automaton on finite words and a Büchi automaton, respectively. It is not hard to see that when $\mathcal{A}$ is deterministic, then $L_B(\mathcal{A}) = L_{\#}(\mathcal{A})$. Indeed, in both cases, $\mathcal{A}$ accepts a word $w$ if its single run on $w$ visits $\alpha$ infinitely often. When, however, $\mathcal{A}$ is nondeterministic, its $\#$-language may contain words accepted via infinitely many different runs, none of which visits $\alpha$ infinitely often.

▶ **Example 19.** Consider the automaton $\mathcal{A}_1$ in Figure 1. Note that $L_F(\mathcal{A}_1) = R$, for $R = (\$+0\cdot\{0,1,\$\}^* \cdot 1)$, namely the language used in Theorem 9 for demonstrating a language with $R^{\#} \neq R^{\omega}$. Here, we have that $L_{\#}(\mathcal{A}_1) \neq L_B(\mathcal{A}_1)$. For example, $w = 011\$1\$\$1\$\$\$1\$\$\$\$1\$\$\$\$\$ \cdots = 0 \cdot \prod_{i=0}^{\infty} 1\$^i \in L_{\#}(\mathcal{A}_1) \setminus L_B(\mathcal{A}_1)$.

Consider now the automaton $\mathcal{A}_2$. Here, $L_B(\mathcal{A}_2) = (0+1)^* \cdot 1^{\omega}$. On the other hand, for every $i \geq 1$, there is a run of $\mathcal{A}_2$ on $w = 01011011101111 \cdots = \prod_{i=0}^{\infty} 01^i$ that visits $\alpha$ at least $i$ times. Thus, $w \in L_{\#}(\mathcal{A}_2)$ even though it has infinitely many 0's and is not in $L_B(\mathcal{A}_2)$. Note that the word $w$ is also used to differentiate the Büchi and *prompt-Büchi* acceptance conditions. A prompt-Buch automaton $\mathcal{A}$ accepts a word $w$ iff there is $i \geq 1$ and a run $r$ of $\mathcal{A}$ on $w$, such that $r$ visits $\alpha$ at least once in every $i$ successive states [1]. It is not hard to see that $w$ is not accepted by all DBWs for $(1^* \cdot 0)^{\omega}$. ◀

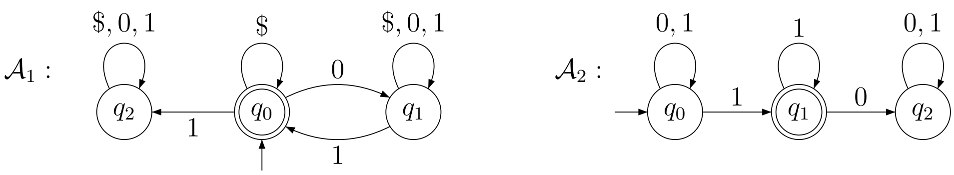


**Figure 1** Automata with a non-regular $\#$-language.

▶ Remark 20. Defining $L_{\#}(\mathcal{A})$, we require the transition function $\delta$ of $\mathcal{A}$ to be defined for all states and letters. Indeed, a rejecting sink in a $\#$-automaton may support acceptance. To see this, consider $\mathcal{A}_2$ from Example 19, and assume that rather than going with the letter 0 to the rejecting sink $q_2$, the state $q_1$ would have no outgoing transitions labeled 0. Then, no run of $\mathcal{A}_2$ on the word $w$ from the example can visit $q_1$ even once without getting stuck. Note that rather than requiring $\delta$ to be total, we could also define $L_{\#}(\mathcal{A})$ as these words for which, for all $i \geq 0$, there is a run of $\mathcal{A}$ on a prefix of $w$ that visits $\alpha$ at least $i$ times. ◀

Interestingly, the relation between $L_{\#}(\mathcal{A})$ and $L_B(\mathcal{A})$ is similar to the one obtained for $R^{\#}$ and $R^{\omega}$. Formally, we have the following.

▶ **Theorem 21.** *For all finite automata $\mathcal{A}$, the following are equivalent.*
***(1)*** *$L_{\#}(\mathcal{A})$ is $\omega$-regular.*
***(2)*** *$L_B(\mathcal{A}) = L_{\#}(\mathcal{A})$.*
***(3)*** *$L_{\#}(\mathcal{A})$ is in DBW.*

**Proof.** Clearly, both (2) $\rightarrow$ (1) and (3) $\rightarrow$ (1). We prove that (1) $\rightarrow$ (2) and (1) $\rightarrow$ (3).

We start with (1) $\rightarrow$ (2). First, clearly, for all automata $\mathcal{A}$, we have that $L_B(\mathcal{A}) \subseteq L_\#(\mathcal{A})$. We prove that $L_\#(\mathcal{A}) \subseteq L_B(\mathcal{A})$. Since $L_\#(\mathcal{A})$ is $\omega$-regular, then, as $\omega$-regular languages are closed under complementation, there is an NBW $\mathcal{B}$ for $L_\#(\mathcal{A}) \setminus L_B(\mathcal{A})$. If $L_\#(\mathcal{A}) \not\subseteq L_B(\mathcal{A})$, then $L_B(\mathcal{B})$ is not empty, which implies $\mathcal{B}$ accepts a lasso-shaped word, namely a word of the form $u \cdot v^\omega$ for $u, v \in \Sigma^* \setminus \{\varepsilon\}$. But $L_\#(\mathcal{A})$ and $L_B(\mathcal{A})$ agree on all lasso-shaped words. Indeed, $u \cdot v^\omega \in L_\#(\mathcal{A})$ iff $\mathcal{A}$ has a cycle that visits $\alpha$ and is traversed when the $v^\omega$ suffix is read, iff $u \cdot v^\omega \in L_B(\mathcal{A})$. Hence, $\mathcal{B}$ is empty, $L_\#(\mathcal{A}) \subseteq L_B(\mathcal{A})$, and we are done.

We continue to (1) $\rightarrow$ (3). For all $i \geq 0$, let $L_i$ be the set of words $w \in \Sigma^*$ such that there exists a run of $\mathcal{A}$ on $w$ that visits $\alpha$ exactly $i$ times. Observe that $L_\#(\mathcal{A}) = \bigcap_{i \geq 0} L_i \cdot \Sigma^\omega$. Thus, $L_\#(\mathcal{A})$ is a countable intersection of open sets. Hence, by Landweber, $L_\#(\mathcal{A})$ being $\omega$-regular implies that $L_\#(\mathcal{A})$ is in DBW. ◀

## 7 Discussion

The expressiveness and succinctness of different classes of automata on infinite words have been studied extensively in the early days of the automata-theoretic approach to formal verification [21]. Specification formalisms that combine regular expressions or automata with temporal-logic modalities have been the subject of extensive research too [23, 22]. Quite surprisingly, the expressiveness and succinctness of repetition languages, which are at the heart of this study, have been left open. The research described in this paper started following a question asked by Michael Kaminski about $R^\omega$ being DBW-recognizable for every regular language $R$. We had two conjectures about this question. First, that the answer is positive, and second, that this must have been studied already. We were not able to prove either conjecture, and in fact refuted the first. In the process, we developed the full theory of repetition languages, their expressiveness, and succinctness, as well the notion of #-languages which goes beyond $\omega$-regular languages. Our results are summarized in Table 2 below. The $\surd$ and $\times$ symbols indicate whether a translation always exists. All blow-ups except for the one from [2] are tight. The blow-ups in translations to DBWs apply also to DRWs (Th. 18). Finally, for $R^\#$, translations exist whenever $R^\omega$ is DBW-recognizable (Th. 5), in which case the blow-ups agree with the one described for $R^\omega$.

| | $\lim(R)$ | $\infty R$ | $R^\omega$ |
|---|---|---|---|
| DFW to DBW | $\surd$<br>$O(n)$<br>[12] | $\surd$<br>$2^{O(n)}$<br>Ths. 15 and 16 | $\times$<br>$2^{O(n \log n)}$<br>Ths. 9 and 17 |
| DFW to NBW | $\surd$<br>$O(n)$<br>[12] | $\surd$<br>$O(n)$<br>Th. 15 | $\surd$<br>$O(n)$<br>Th. 13 |
| NFW to NBW | $\surd$<br>$O(n^3)$<br>[2] | $\surd$<br>$O(n)$<br>Th. 15 | $\surd$<br>$O(n)$<br>Th. 13 |

**Figure 2** Translations from an automaton for $R$ to automata for its repetition languages.

**Acknowledgment** We thank Michael Kaminski for asking us whether $R^\omega$ is DBW-recognizable for every regular language $R$.

## References


1 S. Almagor, Y. Hirshfeld, and O. Kupferman. Promptness in omega-regular automata. In *8th Int. Symp. on Automated Technology for Verification and Analysis*, volume 6252, pages 22–36, 2010.

2 B. Aminof and O. Kupferman. On the succinctness of nondeterminizm. In *4th Int. Symp. on Automated Technology for Verification and Analysis*, volume 4218 of *Lecture Notes in Computer Science*, pages 125–140. Springer, 2006.

3 A. Arnold and D. Niwiński. Fixed point characterization of weak monadic logic definable sets of trees. In M. Nivat and A. Podelski, editors, *Tree Automata and Languages*, pages 159–188. Elsevier, 1992.

4 C. Baier, L. de Alfaro, V. Forejt, and M. Kwiatkowska. Model checking probabilistic systems. In *Handbook of Model Checking.*, pages 963–999. Springer, 2018.

5 R. Bloem, K. Chatterjee, and B. Jobstmann. Graph games and reactive synthesis. In *Handbook of Model Checking.*, pages 921–962. Springer, 2018.

6 J.R. Büchi. On a decision method in restricted second order arithmetic. In *Proc. Int. Congress on Logic, Method, and Philosophy of Science. 1960*, pages 1–12. Stanford University Press, 1962.

7 C. Eisner and D. Fisman. *A Practical Introduction to PSL*. Springer, 2006.

8 S.C. Krishnan, A. Puri, and R.K. Brayton. Deterministic $\omega$-automata vis-a-vis deterministic Büchi automata. In *Algorithms and Computations*, volume 834 of *Lecture Notes in Computer Science*, pages 378–386. Springer, 1994.

9 O. Kupferman. Automata theory and model checking. In *Handbook of Model Checking*, pages 107–151. Springer, 2018.

10 O. Kupferman and M.Y. Vardi. From linear time to branching time. *ACM Transactions on Computational Logic*, 6(2):273–294, 2005.

11 R.P. Kurshan. *Computer Aided Verification of Coordinating Processes*. Princeton Univ. Press, 1994.

12 L.H. Landweber. Decision problems for $\omega$–automata. *Mathematical Systems Theory*, 3:376–384, 1969.

13 C. Löding. Optimal bounds for the transformation of $\omega$-automata. In *Proc. 19th Conf. on Foundations of Software Technology and Theoretical Computer Science*, volume 1738 of *Lecture Notes in Computer Science*, pages 97–109, 1999.

14 R. McNaughton. Testing and generating infinite sequences by a finite automaton. *Information and Control*, 9:521–530, 1966.

15 A.R. Meyer and L.J. Stockmeyer. The equivalence problem for regular expressions with squaring requires exponential time. In *Proc. 13th IEEE Symp. on Switching and Automata Theory*, pages 125–129, 1972.

16 D.E. Muller, A. Saoudi, and P.E. Schupp. Alternating automata, the weak monadic theory of the tree and its complexity. In *Proc. 13th Int. Colloq. on Automata, Languages, and Programming*, volume 226 of *Lecture Notes in Computer Science*, pages 275 – 283. Springer, 1986.

17 M.O. Rabin. Decidability of second order theories and automata on infinite trees. *Transaction of the AMS*, 141:1–35, 1969.

18 M.O. Rabin. Weakly definable relations and special automata. In *Proc. Symp. Math. Logic and Foundations of Set Theory*, pages 1–23. North Holland, 1970.

19 M.O. Rabin and D. Scott. Finite automata and their decision problems. *IBM Journal of Research and Development*, 3:115–125, 1959.

20 S. Safra. On the complexity of $\omega$-automata. In *Proc. 29th IEEE Symp. on Foundations of Computer Science*, pages 319–327, 1988.

21 W. Thomas. Automata on infinite objects. *Handbook of Theoretical Computer Science*, pages 133–191, 1990.

22 M.Y. Vardi and P. Wolper. Reasoning about infinite computations. *Information and Computation*, 115(1):1–37, 1994.

23 P. Wolper. Temporal logic can be more expressive. In *Proc. 22nd IEEE Symp. on Foundations of Computer Science*, pages 340–348, 1981.

# A Proofs

## A.1 Correctness of the Construction in the proof of Theorem 13

We prove that a finite word $x \in \Sigma^*$ closes a cycle from $q_0'$ iff $x \in R^*$. Thus, $\mathcal{A}'$ has a run on an infinite word $w \in \Sigma^\omega$ that visits $q_0'$ infinitely often iff $w \in R^\omega$. It is sufficient to prove that $\mathcal{A}'$ has a run on $x$ that visits $q_0'$ exactly twice, at the first and last states, iff $x \in R$.

Consider a word $x \in R$. We first prove that there is a cycle from $q_0'$ labeled by $x$. Let $l = |x|$, and let $r : \{0, \ldots, l\} \to Q$ be an accepting run of $\mathcal{A}$ over $x$. Clearly $q_0', r(1), r(2), \ldots, r(l-1)$ is a legal run of $\mathcal{A}'$ on $x[1, l-1]$. Since $r$ is accepting, we have that $r(l) \in \alpha$, and hence $\delta(r(l-1), x[l]) \cap \alpha \neq \varnothing$. If $l = 1$, then $q_0' \in \delta'(q_0', x[l])$, otherwise $l > 1$, and $q_0' \in \delta'(r(l-1), x[l])$. In either case, we have that $q_0', r(1), r(2), \ldots, r(l-1), q_0'$ is a finite run of $\mathcal{A}'$ on $x$. That is, $x$ closes a cycle from $q_0'$, and we are done.

For the other direction, we show that any cycle from $q_0'$, that visits $q_0'$ exactly twice, is labeled by some word in $R$. Consider a cycle $\pi = q_0', s_1, \ldots, s_{l-1}, q_0'$ in $\mathcal{A}'$, and assume that $s_i \neq q_0'$ for all $1 \leq i \leq l-1$. Let $x \in \Sigma^*$ be a word that traverses $\pi$, note that $|x| = l$. We prove that $x \in R$. If $l = 1$, that is $\pi = q_0', q_0'$, it follows that $\delta(q_0, x[1]) \cap \alpha \neq \varnothing$. Hence, $x \in L(\mathcal{A}) = R$, and we are done. If $l \geq 2$, then as $s_i \neq q_0'$ for $1 \leq i \leq l-1$, we have that $q_0, s_1, \ldots, s_{l-1}$ is a run of $\mathcal{A}$ over $x[1, l-1]$. In addition, since $q_0' \in \delta'(s_{l-1}, x[l])$ and $s_{l-1} \neq q_0'$, there exist $s_l \in \delta(s_{l-1}, u[l]) \cap \alpha$. Thus, $q_0, s_1, \ldots, s_{l-1}, s_l$ is an accepting run of $\mathcal{A}$ on $x$. Hence, $x \in L(\mathcal{A}) = R$, and we are done.

## A.2 Correctness of the Construction in the Proof of Theorem 15

We first prove that $L(\mathcal{A}') \subseteq (\Sigma^* \cdot R)^\omega$. Consider a word $w \in L(\mathcal{A}')$, and let $r$ be an accepting run of $\mathcal{A}'$ on $w$. Since $r$ is accepting, there is an infinite sequence of positions $0 = i_0 < i_1 < i_2 < \ldots$ such that for all $j \geq 1$, we have that $r(i_j) \in \alpha$, and for all $i_{j-1} < t < i_j$, the state $r(t)$ is not in $\alpha$. Thus, $i_1, i_2, \ldots$ is the sequence of positions in which $r$ visits $\alpha$. We show that for all $j \geq 0$, it holds that $w[i_j + 1, i_{j+1}] \in \Sigma^* \cdot R$. Consider the segment $r(i_j), r(i_j + 1), \ldots, r(i_{j+1})$ of the run $r$ on the infix $w[i_j + 1, i_{j+1}]$ of $w$. Recall that $r(i_j) \in \alpha$ and $r(t) \notin \alpha$ for $i_j < t < i_{j+1}$. It follows that the sequence $q_0, r(i_j + 1), \ldots, r(i_{j+1})$ is a legal run of $\mathcal{A}$, the NFW for $\Sigma^* \cdot R$, on $w[i_j + 1, i_{j+1}]$. This run is accepting, and so $w[i_j + 1, i_{j+1}]$ is in $\Sigma^* \cdot R$. Hence, $w = w[i_0 + 1, i_1] \cdot w[i_1 + 1, i_2] \cdots \in (\Sigma^* \cdot R)^\omega$.

It is left to prove that $(\Sigma^* \cdot R)^\omega \subseteq L(\mathcal{A}')$. Consider a sequence of words $w_1, w_2, \ldots \in \Sigma^* \cdot R$, and assume that for all $n \geq 1$, the word $w_n$ is minimal with respect to '$\prec$'. I.e. if $x \prec w_n$ and $x \in \Sigma^* \cdot R$, then $x = w_n$. For $n \geq 1$, let $l_n$ be the length of $w_n$, and $r_n = \langle q_0, q_1^n, \ldots, q_{l_n}^n \rangle$ be an accepting run of $\mathcal{A}$ on $w_n$. We prove that for all $i, j \geq 0$, the concatenation $r_i \cdot r_j[1, l_j] = \langle q_0, \ldots, q_{l_i}^i, q_1^j, \ldots, q_{l_j}^j \rangle$ is a legal run of $\mathcal{A}'$ on $w_i \cdot w_j$. Observe that since the $w_n$'s are minimal, it holds that $q_t^n \in \alpha$ iff $t = l_n$. Thus, by definition of $\delta'$, we only need to show that $q_1^j \in \delta'(q_{l_i}^i, w_j[1])$. Since $q_{l_i}^i \in \alpha$, we have that $\delta'(q_{l_i}^i, w_j[1]) = \delta(q_0, w_j[1])$. By definition of $r_j$, it holds that $q_1^j \in \delta(q_0, w_j[1])$, and thus $q_1^j \in \delta'(q_{l_i}^i, w_j[1])$. It follows that the infinite concatenation $r = r_1 \cdot \prod_{i=2}^{\infty} r[1, l_i] = \langle q_0, q_1^1, \ldots, q_{l_1}^1, q_1^2, \ldots, q_{l_2}^2, q_1^3, \ldots \rangle$, is a legal run of $\mathcal{A}'$ on $w_1 \cdot w_2 \cdot w_3 \cdots$. Clearly, $r$ visits $\alpha$ infinitely many times, and hence $w_1 \cdot w_2 \cdot w_3 \cdots \in L(\mathcal{A}')$. Thus, we are only left showing that if $w \in (\Sigma^* \cdot R)^\omega$, then there exists a sequence of minimal words $w_1, w_2, \ldots \in \Sigma^* \cdot R$, such that $w = w_1 \cdot w_2 \cdots$. Equivalently, let $K$ be the set of minimal words in $\Sigma^* \cdot R$, and prove that $(\Sigma^* \cdot R)^\omega \subseteq K^\omega$. We prove that $(\Sigma^* \cdot R)^\omega \subseteq K \cdot (\Sigma^* \cdot R)^\omega$, and conclude by Lemma 2 that $(\Sigma^* \cdot R)^\omega \subseteq K^\omega$. Consider a word $w \in (\Sigma^* \cdot R)^\omega$, and let $x \prec w$, be the minimal prefix of $w$ with which $x \in \Sigma^* \cdot R$, and let $y \in \Sigma^\omega$ be such that $w = x \cdot y$. As a suffix of $w$, we have that $y \in (\Sigma^* \cdot R)^\omega$, and thus, $w \in K \cdot (\Sigma^* \cdot R)^\omega$, and we are done.

## A.3 Lower Bound on the DBWs from the Proof of Theorem 16

For a word $x \in \Sigma^*$, denote by $\overline{x} \in \Sigma^*$ the *negative* of $x$; that is, the word obtained from $x$ by flipping 0's and 1's. Formally, for all $1 \leq i \leq |x|$, we have that $\overline{x}[i] = 0$ iff $x[i] = 1$. For example, if $x = 010$, then $\overline{x} = 101$, and if $x = 1110$, then $\overline{x} = 0001$. Proving a lower bound on the number of states of a DBW for $\infty R_n$, it is more convenient to characterize $\infty R_n$ through its complement. Observe that a word $w \in \Sigma^\omega$ is not in $\infty R_n$ iff there exists a position $t_0 \geq 1$, such that for all $t \geq t_0$, we have that $w[t] \neq w[t+n]$. Equivalently, there exists a finite word $u \in \Sigma^n$, such that $(u \cdot \overline{u})^\omega$ is a suffix of $w$.

For two words $u_1, u_2 \in \Sigma^n$, we say that $u_1 \sim u_2$ iff $u_1 \cdot \overline{u_1}$ is a *cyclic shift* of $u_2 \cdot \overline{u_2}$. That is, $u_1 \sim u_2$ iff exists $1 \leq i \leq 2n$ such that $u_1 \cdot \overline{u_1} = (u_2 \cdot \overline{u_2})[i, 2n] \cdot (u_2 \cdot \overline{u_2})[1, i-1]$. For example, $011 \sim 000$, as 011100 is a cyclic shift of 000111. It is easy to see that $\sim$ is an equivalence relation. Moreover, since there are $2n$ cyclic shifts of a word of length $2n$, it follows that each equivalence class has at most $2n$ members. Thus, there are at least $\frac{2^n}{2n} = \frac{2^{n-1}}{n}$ equivalence classes.

We are now ready to prove that a DBW for $\infty R_n$ needs at least $\frac{2^{n-1}}{n}$ states. Consider a DBW $\mathcal{A}_n = \langle \Sigma, Q, q_0, \delta, \alpha \rangle$ for $\infty R_n$, and consider a finite word $u \in \Sigma^n$. Let $w_u = (u \cdot \overline{u})^\omega$, and recall that $w_u \notin \infty R_n$. Let $r_u$ be the rejecting run of $\mathcal{A}_n$ on $w_u$, and let $S_u$ be the set of states visited by $r_u$ infinitely often. We prove that for every two finite words $u_1$ and $u_2$ of length $n$ such that $u_1 \not\sim u_2$, it must be that $S_{u_1} \cap S_{u_2} = \varnothing$. Since there are at least $\frac{2^{n-1}}{n}$ different equivalence classes, this implies that $\mathcal{A}_n$ must have at least $\frac{2^{n-1}}{n}$ states. Assume by way of contradiction that $u_1$ and $u_2$ are such that $S_{u_1} \cap S_{u_2} \neq \varnothing$. For brevity, for $i \in \{1, 2\}$, let $r_i = r_{u_i}$ and $S_i = S_{u_i}$. Let $q \in Q$ be a state in $S_1 \cap S_2$. For $i \in \{1, 2\}$, let $h_i \cdot y_i^\omega$ be a word induced by $r_i$, such that the following hold.

- $h_i$ is a prefix of $(u_i.\overline{u_i})^\omega$ with which $r_i$ moves from $q_0$ to $q$ and stays in $S_i$. That is, $r_i(|h_i|) = q$ and $r_i(t) \in S_i$ for all $t \geq |h_i|$, and
- $y_i$ labels a cycle from $q$ to itself that includes $(u_i.\overline{u_i})$ as a subword.

Consider the word $w = h_1 \cdot (y_1 \cdot y_2)^\omega$, and let $r_w$ be the run of $\mathcal{A}_n$ over $w$. Observe that $\inf(r_w) \subseteq S_1 \cup S_2$. Thus, since $S_1$ and $S_2$ are both rejecting, so is $\inf(r_w)$. Hence $w \notin \infty R_n$, implying that there is $x \in \Sigma^n$ such that $(x \cdot \overline{x})^\omega$ is a suffix of $(y_1.y_2)^\omega$. Since $u_1 \cdot \overline{u_1}$ and $u_2 \cdot \overline{u_2}$, are subwords of $y_1$ and $y_2$, respectively, it follows that both are also subwords of $(x \cdot \overline{x})^\omega$. This is possible only if $u_1 \cdot \overline{u_1}$ and $u_2 \cdot \overline{u_2}$ are cyclic shifts of $x \cdot \overline{x}$. That is, $u_1 \sim x \sim u_2$, which contradicts the assumption that $u_1 \not\sim u_2$, and we are done.

## A.4 On $R_n^\omega$ from the Proof of Theorem 17

In Lemma 22, we prove that $R_n^\omega$ is in DBW. Then, in Lemma 23, we prove that a DBW for $R_n^\omega$ needs at least $n!$ states.

▶ **Lemma 22.** *For all $n \geq 1$, we have that $R_n^\omega$ is in DBW.*

**Proof.** We prove that for all $n \geq 1$, we have that $R_n^\omega = R_n^\#$. By Theorem 5, we then have that $R_n^\omega$ is in DBW. We need to show that $R_n^\# \subseteq R_n^\omega$. Consider a word $w \in R_n^\#$. For $i \geq 1$, let $x_i, y_i, u_i, \in \Sigma_n^*$ be such that $x_i$ and $y_i$ are in $R_n$, $u_i$ is in $R_n^i$, the word $x_i \cdot y_i \cdot u_i$ is a prefix of $w$, and all the words $y_i$ start and end with the same letter $\sigma \in [n]$. Since $[n]$ is finite and $w \in R_n^\#$, we know that such $x_i, y_i$, and $u_i$ exist.

For $i \geq 1$, let $l_i = |x_i|$. Let $t \geq 1$ be such that $l_t$ is minimal, and let $w' \in \Sigma_n^\omega$ be such that $w = x_t \cdot w'$. Observe that, by the minimality of $t$, we have that $x_t \prec x_j$ for all $j \geq 1$. Specifically, $x_j = x_t \cdot (x_j[l_t + 1, l_j])$. Note that $x_j[l_t + 1, l_j] = \varepsilon$ when $l_j = l_t$. For $j \geq 1$, let $z_j = x_j[l_t + 1, l_j] \cdot y_j$. Note that the first letter of $z_j$ is the $(l_t + 1)$-th letter of $w$, which is also the first letter of $y_t$. Therefore, the word $z_j$ starts and ends with the letter $\sigma$. It follows that for all $j \geq 1$, we have that $z_j \in R_n$ and $x_j \cdot y_j = x_t \cdot z_j$. Hence, for all $j \geq 1$, the word $z_j \cdot u_j$ is a prefix of $w'$, implying that $w' \in R_n^{\#}$. Recall that $w = x_t \cdot w'$. Hence, $w \in R_n \cdot R_n^{\#}$, and so, by Lemma 2, we have that $w \in R_n^{\omega}$, and we are done. ◀

▶ **Lemma 23.** *A DBW for $R_n^{\omega}$ needs at least $n!$ states.*

**Proof.** Consider a DBW $\mathcal{A}_n = \langle \Sigma_n, Q, q_0, \delta, \alpha \rangle$ for $R_n^{\omega}$, and consider a permutation $\pi = \langle \sigma_1, \ldots, \sigma_n \rangle$ of $\{1, \ldots, n\}$. Note that the word $w_\pi = (\sigma_1 \cdots \sigma_n \cdot \#)^{\omega}$ is not in $R_n^{\omega}$. Thus, $w_\pi$ is not accepted by $\mathcal{A}_n$. Let $r_\pi$ be the rejecting run of $\mathcal{A}_n$ on $w_\pi$, and let $S_\pi \subseteq Q$ be the set of states that are visited infinitely often in $r_\pi$. We prove that for every two different permutations $\pi_1$ and $\pi_2$ of $\{1, ..., n\}$, it must be that $S_{\pi_1} \cap S_{\pi_2} = \varnothing$. Since there are $n!$ different permutations, this implies that $\mathcal{A}_n$ must have at least $n!$ states. Assume by way of contradiction that $\pi_1$ and $\pi_2$ are such that $S_{\pi_1} \cap S_{\pi_2} \neq \varnothing$. For brevity, for $i \in \{1, 2\}$, let $w_i = w_{\pi_i}$, $r_i = r_{\pi_i}$, and $S_i = S_{\pi_i}$. Let $q \in Q$ be a state in $S_1 \cap S_2$. We define four finite words in $\Sigma_n^*$, for $i = 1, 2$, as follows.

- Let $h_i$ be a prefix of $w_i$ with which $r_i$ moves from $q_0$ to $q$ and stays in $S_i$. That is, $r_i(|h_i|) = q$ and $r_i(t) \in S_i$ for all $t \geq |h_i|$.
- Let $z_i$ be the suffix of $w_i$ such that $w_i = h_i \cdot z_i$. Let $u_i$ be a prefix of $z_i$ such that $u_i$ includes the permutation $\pi_i$ and with which $r_i$ moves from $q$ back to $q$ by visiting exactly all the states in $S_{\pi_i}$.

Consider the word $w = h_1 \cdot (u_1 \cdot u_2)^{\omega}$, and let $r_w$ be the run of $\mathcal{A}_n$ on $w$. We prove that $w \in R_n^{\omega}$ and $w \notin L(\mathcal{A}_n)$, which is a contradiction for $\mathcal{A}_n$ being a DBW for $R_n^{\omega}$. First, observe that $\inf(r_w) = S_1 \cup S_2$. Since both $w_1$ and $w_2$ are not in $R_n^{\omega}$, then $(S_1 \cup S_2) \cap \alpha = \varnothing$. Thus, $\inf(r_w) \cap \alpha = \varnothing$. Hence, $w \notin L(\mathcal{A}_n)$. It is left to prove that $w \in R_n^{\omega}$. Let $\pi_1 = \langle \sigma_1^1, \ldots, \sigma_n^1 \rangle$ and $\pi_2 = \langle \sigma_1^2, \ldots, \sigma_n^2 \rangle$. We prove the following two claims.

▷ Claim 24. There exists $v = v_0, v_1, \ldots, v_{k-1} \in [n]^*$ such that for all $0 \leq t \leq k-1$, the pairs $v_t \cdot v_{t+1 (\text{mod } k)}$ appear in $w$ infinitely often.

**Proof.** Let $j \geq 1$ be the minimal index for which $\sigma_j^1 \neq \sigma_j^2$. Note that $j < n$. There must exist $m$ and $l$ such that $j < m, l \leq n$, $\sigma_j^1 = \sigma_m^2$, and $\sigma_j^2 = \sigma_l^1$. We define $v = v_0, v_1, \ldots, v_{k-1} = \sigma_j^1, \ldots, \sigma_{l-1}^1, \sigma_j^2, \ldots, \sigma_{m-1}^2$. Indeed, the pairs $v_t \cdot v_{t+1 (\text{mod } k)}$, for $0 \leq t \leq k-1$, repeat in $(u_1 \cdot u_2)^{\omega}$ infinitely many times as they all are subwords of $\pi_1$ or $\pi_2$, and $\pi_i$ is a subword of $u_i$ for $i \in \{1, 2\}$. ◀

▷ Claim 25. For all $1 \leq j \leq n$, there exists $x_j \in R_n^*$ such that $x_j \cdot \sigma_j^1 \prec w$.

**Proof.** For $j = 1$, observe that since $h_1 \cdot u_1 \prec w_1$, we have that $\sigma_1^1 \prec h_1 \cdot (u_1 \cdot u_2)^{\omega} = w$. Thus, we can take $x_1 = \varepsilon$. Assume that $x_j$ has been defined for some $j < n$. The pair $\sigma_j^1 \cdot \sigma_{j+1}^1$ repeats infinitely many times in $w$, as it is a subword of $\pi_1$. Hence, there exists $y \in \Sigma_n^*$ such that $x_j \cdot \sigma_j^1 \cdot y \cdot \sigma_j^1 \cdot \sigma_{j+1}^1 \prec w$. Thus, by taking $x_{j+1} = x_j \cdot (\sigma_j^1 \cdot y \cdot \sigma_j^1) \in R_n^*$, we have that $x_{j+1} \cdot \sigma_{j+1}^1 \prec w$. ◀

We conclude that $w = x \cdot w'$, for $w' \in v_0 \cdot \Sigma_n^{\omega}$, and for all $0 \leq t \leq k-1$, the pair $v_t \cdot v_{t+1 (\text{mod } k)}$ appears in $w'$ infinitely often. Thus, we may iteratively define $y_j \in \Sigma_n^*$, for $j \geq 0$, such that $(\prod_{t=0}^{j} (v_{t (\text{mod } k)} \cdot y_t \cdot v_{t (\text{mod } k)})) \cdot v_{j+1 (\text{mod } k)} \prec w'$. It follows that $w' = \prod_{t=0}^{\infty} (v_{t (\text{mod } k)} \cdot y_t \cdot v_{t (\text{mod } k)}) \in R_n^{\omega}$. Hence, $w = x \cdot w' \in R_n^* \cdot R_n^{\omega} = R_n^{\omega}$, and we are done. ◀