Büchi Automata Recognizing Sets of Reals Definable in First-Order Logic with Addition and Order

Arthur Milchior

Büchi Automata Recognizing Sets of Reals Definable in First-Order Logic with Addition and Order

Arthur Milchior[1]

LACL, Université Paris-Est Créteil arthur.milchior@lacl.fr

Abstract. This work considers encodings of non-negative reals in a fixed base, and their encoding by weak deterministic Büchi automata. A Real Number Automaton is an automaton which recognizes all encodings of elements of a set of reals. We explain in this paper how to decide in linear time whether a set of reals recognized by a given minimal weak deterministic RNA is $\mathrm{FO}[\mathbb{R};+,<,1]$-definable. Furthermore, it is explained how to compute in quasi-quadratic (respectively, quasi-linear) time an existential (respectively, existential-universal) $\mathrm{FO}[\mathbb{R};+,<,1]$-formula which defines the set of reals recognized by the automaton.
As an additional contribution, the techniques used for obtaining our main result lead to a characterization of minimal deterministic Büchi automata accepting $\mathrm{FO}[\mathbb{R};+,<,1]$-definable set.

Introduction

This paper deals with logically defined sets of numbers encoded by weak deterministic Büchi automata. The sets of tuples of integers whose encodings in base b are recognized by a finite automaton are called the b-recognizable sets. By [?], the b-recognizable sets of vectors of integers are exactly the sets which are $\mathsf{FO}\,[\mathbb{N};+,<,V_b]$-definable, where $V_b(n)$ is the greatest power of b dividing n. It was proven in [?,?] that the $\mathsf{FO}\,[\mathbb{N};+,<]$-definable sets are exactly the sets which are b- and b'-recognizable for every $b \geq 2$.

Those results naturally led to the following problem: deciding whether a finite automaton recognizes a $\mathsf{FO}\,[\mathbb{N};+,<]$-definable set of d-tuples of integers for some dimension $d \in \mathbb{N}^{>0}$. In the case of dimension $d = 1$, decidability was proven in [?]. For $d > 1$, decidability was proven in [?]. Another algorithm was given in [?], which solves this problem in polynomial time. For $d = 1$, a quasi linear time algorithm was given in [?].

The above-mentioned results about sets of tuples of natural numbers and finite automata have then been extended to sets of tuples of reals recognized by a Büchi automaton. The notion of Büchi automata is a formalism which describes languages of infinite words, also called ω-words. The Büchi automata are similar to the finite automata. The main difference is that finite automata accept finite words which admit runs ending on accepting state, while Büchi automata accept infinite words which admit runs in which an accepting state appears infinitely often.

One of the main differences between finite automata and Büchi automata is that finite automata can be determinized while deterministic Büchi automata are less expressive than Büchi automata. For example, the language $L_{\text{fin}(a)}$ of words containing a finite number of times the letter a is recognized by a Büchi automaton, but is not recognized by any deterministic Büchi automaton. This statement implies, for example, that no deterministic Büchi automaton recognizes the set of reals of the form nb^p with $n \in \mathbb{N}$ and $p \in \mathbb{Z}$, that is, the set of reals whose encoding ends with 0 or $(b-1)$ repeated infinitely many times.

A main difference between the two classes of deterministic automata is that the class of languages recognized by deterministic finite automata is closed under complement while the class of languages recognized by deterministic Büchi automata is not. For example, $L_{\text{fin}(a)}$ is not recognized by any deterministic Büchi automaton while its complements $L_{\text{inf}(a)}$ is recognized by a deterministic Büchi automaton.

However, the set of weak deterministic Büchi automata is closed under complement. A weak deterministic Büchi automaton is a deterministic Büchi automaton whose set of accepting states is a union of strongly connected components. Handling weak Büchi automata is similar to manipulating finite automata. A set is said to be weakly b-recognizable if it is recognized by a weak automaton in base b. The class of weak deterministic Büchi automata is less expressive than the class of deterministic Büchi automata. For example, as mentionned above, the language $L_{\text{inf}(a)}$ is recognized by a deterministic Büchi automaton, but this language is not recognized by any weak deterministic Büchi automaton. This implies that, for example, no weak deterministic Büchi automaton recognizes the set of reals which are not of the form nb^p with $n \in \mathbb{N}$ and $p \in \mathbb{Z}$, since those reals are the ones whose encoding in base b contains an infinite number of non-0 digits. Furthermore, by [?], weak deterministic Büchi automata can be efficiently minimized.

A Real Vector Automaton (RVA, See e.g. [?]) of dimension d is a Büchi automaton $\mathcal{A}$ over alphabet $\{0,\ldots,b-1\}^d \cup \{\star\}$, which recognizes the set of encodings in base b of the elements of a set of vectors of reals. Equivalently, for w an infinite word encoding a vector of dimension d of real $(r_0,\ldots,r_{d-1})$, if w is accepted by $\mathcal{A}$, then all encodings w' of $(r_0,\ldots,r_{d-1})$ are accepted by $\mathcal{A}$. In the case where the dimension d is 1, those automata are called Real Number Automata (RNA, See e.g. [?]).

The sets of tuples of reals whose encoding in base b is recognized by a RVA are called the b-recognizable sets. By [?], they are exactly the $\mathsf{FO}[\mathbb{R},\mathbb{N};+,<,X_b,1]$-definable sets. The logic $\mathsf{FO}[\mathbb{R},\mathbb{N};+,<,X_b,1]$ is the first-order logic over reals with a unary predicate which holds over integers, addition, order, the constant one, and the function $X_b(x,u,k)$. The function $X_b(x,u,k)$ holds if and only if u is equal to some b^n with $n \in \mathbb{Z}$ and there exists an encoding in base b of x whose digit in position n is k. That is, u and x are of one of the two following forms:

$$\begin{array}{l} u = 0 \dots 0 \star 0 \dots 0\ 1\ 0 \dots \\ x = \quad \dots \quad \star \quad \dots \quad k \quad \dots \end{array} \text{ or } \begin{array}{l} u = 0 \dots 0\ 1\ 0 \dots 0 \star 0 \dots \\ x = \quad \dots \quad k \quad \dots \quad \star \quad \dots \end{array} .$$

By [?], a set is $\mathsf{FO}[\mathbb{R}, \mathbb{N}; +, <]$-definable if and only if its set of encodings is weakly b-recognizable for all $b \geq 2$.

By [?], the logic $\mathsf{FO}[\mathbb{R}; +, <, 1]$ admits quantifier elimination. By [?, Sect. 6], the set of reals which are $\mathsf{FO}[\mathbb{R}; +, <, 1]$-definable are finite unions of intervals with rational bounds. Those sets are called the *simple sets*.

Standard definitions are recalled in Sect. **??**. Sets of states of automata reading reals are studied in Sect. **??**. Furthermore, a method to efficiently solve automaton problem is introduced. In Sect. **??**, given a simple set, an automaton accepting it is constructed. A characterization of minimal deterministic Büchi automata accepting a $\mathsf{FO}[\mathbb{R}; +, <, 1]$-definable set is given in Sect. **??**. This characterization is similar to the insight given in [?] and leads to a linear time algorithm deciding whether a minimal RNA recognizes a $\mathsf{FO}[\mathbb{R}; +, <, 1]$-definable set. This algorithm does not return any false positive on weak deterministic Büchi automata which are not RNA. A false negative is exhibited at the end of Sect. **??**. Given a minimal weak RNA automaton accepting a simple set, it is shown in Sect. **??** that an existential (respectively, existential-universal) $\mathsf{FO}[\mathbb{R}; +, <, 1]$-formula which defines R is computable in quasi-quadratic (respectively quasi-linear) time.

1 Definitions

The definitions used in this paper are given in this section. Some basic lemmas are also given. Most definitions are standard.

Let $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$ and $\mathbb{R}$ denote the set of non-negative integers, integers, rationals and reals, respectively. For $R \subseteq \mathbb{R}$, let $R^{\geq 0}$ and $R^{>0}$ denote the set of non-negative and of positive elements of R, respectively. For $n \in \mathbb{N}$, let $[n]$ represent $\{0, \dots, n\}$. For $a, b \in \mathbb{R}$ with $a \leq b$, let $[a, b]$ denote the closed interval $\{r \in \mathbb{R} \mid a \leq r \leq b\}$, and let (a, b) denote the open interval $\{r \in \mathbb{R} \mid a < r < b\}$. Similarly, let $(a, b]$ (respectively, $[a, b)$) be the half-open interval equals to the union of (a, b) and of $\{b\}$ (respectively, $\{a\}$). For $r \in \mathbb{R}$ let $\lfloor r \rfloor$ be the greatest integer less than or equal to r.

1.1 Finite and Infinite Words

An alphabet is a finite set, its elements are called letters. A finite (respectively ω-) word over alphabet A is a finite (respectively infinite) sequence of letters of A. That is, a function from $[n]$ to A for some $n \in \mathbb{N}$ (respectively from $\mathbb{N}$ to A). A set of finite (respectively ω-) words over alphabet A is called a language (respectively, an ω-language) over alphabet A. The empty word is denoted ϵ.

Let w be a word, its length is denoted $|w|$, it is either a non-negative integer or the cardinality of $\mathbb{N}$. For $n < |w|$, let $w[n]$ denote the n-th letter of w. For v a finite word, let $u = vw$ be the concatenation of v and of w, that is, the word of length $|v| + |w|$ such that $u[i] = v[i]$ for $i < |v|$ and $u[|v| + i] = w[i]$ for $i < |w|$. Let $w\,[< n]$ denote the *prefix* of w of length n, that is, the word u of length n such that $w[i] = u[i]$ for all $i \in [n-1]$. Similarly, let $w\,[\geq n]$ denote the suffix of w without its n-th first letters, that is, the word u of length $|w| - n$ such that $u[i] = w[i + n]$ for all $i \in [|w| - n]$. Note that $w = w\,[< i]\,w\,[\geq i]$ for all $i < |w|$.

Let L be a language of finite words and let L' be either a ω-language or a language of finite words. Let LL' be the set of concatenations of the words of L and of L'. For $i \in \mathbb{N}$, let L^i be the concatenations of i words of L. Let $L^* = \bigcup_{i \in \mathbb{N}} L^i$ and $L^+ = \bigcup_{i \in \mathbb{N}^{>0}} L^i$. If L is a language which does not contain the empty word, let L^ω be the set of infinite sequences of elements of L.

Encoding of Real Numbers. Let us now consider the encoding of numbers in an integer base $b \geq 2$. Let Σ_b be equal to $[b-1]$; it is the set of digits. The base b is fixed for the rest of this paper.

The function sending words to the number they encode are now introduced. Let w be an ω-word with exactly one $\star$. It is of the form $w = w_I \star w_F$, with $w_I \in \Sigma_b^*$ and $w_F \in \Sigma_b^\omega$. The word w_I is called the natural part of w and the ω-word w_F is called its fractional part. Let $[w_I]_b^I = \sum_{i=0}^{|w|-1} b^{|w|-1-i} w_I[i]$, let $[w_F]_b^F = \sum_{i \in \mathbb{N}} b^{-i-1} w_F[i]$ and finally, let $[w_I \star w_F]_b^{\mathsf{R}} = [w_I]_b^I + [w_F]_b^F$. Some properties of concatenation and of encoding of reals are now stated.

Lemma 1. *Let $v \in \Sigma_b^*$, $v' \in \Sigma_b^+$, $w \in \Sigma_b^\omega$ and $a \in \Sigma_b$. Then:*

$$[w]_b^F = [0 \star w]_b^{\mathsf{R}}, \qquad [aw]_b^F = \left(a + [w]_b^F\right)/b,$$
$$[va]_b^I = b\,[v]_b^I + a \quad \textit{and} \quad [v^\omega]_b^F = [v]_b^I \,/\, \left(b^{|v|} - 1\right).$$

Some basic facts about rationals are recalled (see e.g. [?]). The rationals are exactly the numbers which admit encodings in base b of the form $u \star vw^\omega$ with $u, v \in \Sigma_b^*$ and $w \in \Sigma_b^+$. Rationals of the form nb^p, with $n \in \mathbb{N}$ and $p \in \mathbb{Z}$, admit exactly two encodings in base b without leading 0 in the natural part. If $p < 0$, the two encodings are of the form $u \star va(b-1)^\omega$ and $u \star v(a+1)0^\omega$, with $u, v \in \Sigma_b^*$ and $a \in [b-2]$. Otherwise, if $p \geq 0$, the two encodings are of the form $ua(b-1)^q \star (b-1)^\omega$ and $u(a+1)0^q \star 0^\omega$ with $u \in \Sigma_b^*$, $a \in [b-2]$ and $q \in \mathbb{N}$. The rationals which are not of the form nb^p admit exactly one encoding in base b without leading 0 in the natural part.

Encoding of Sets of Reals. Relations between languages and sets of reals are now recalled. Given a language L which is a subset of $\Sigma_b^* \star \Sigma_b^\omega$, let $[L]_b^{\mathsf{R}}$ be the set of reals admitting an encoding in base b in L. The language L is said to be an encoding in base b of the set of reals $[L]_b^{\mathsf{R}}$. Reciprocally, given a set $R \subseteq \mathbb{R}^{\geq 0}$ of reals, $L_b(R)$ is the set of all encodings in base b of elements of R. For L a subset of Σ_b^ω, $[L]_b^F$ is the set of d-tuples of reals, belonging to $[0,1]^d$, which admits an encoding in base b in L.

Following [?], a language L is said to be *saturated* if for any number r which admits an encoding in base b in L, all encodings in base b of r belong to L. The saturated languages are of the form $L_b(R)$ for $R \subseteq \mathbb{R}^{\geq 0}$. Note that $[L_b(R)]_b^{\mathsf{R}} = R$ for all sets $R \subseteq \mathbb{R}^{\geq 0}$. Note also that $L \subseteq L_b([L]_b^{\mathsf{R}})$, and the subset relation is an equality if and only if L is saturated.

All non-empty sets of reals have infinitely many encodings in base b. For example, for $I \subseteq \mathbb{N}$ an arbitrary set, $0^* \star \{0,1\}^\omega \setminus \left\{0^i 1^\omega \mid i \in I\right\}$ is an encoding in base 2 of the simple set $[0,1]$. This language is saturated if and only if $I = \emptyset$.

1.2 Deterministic Büchi Automata

This paper deals with deterministic Büchi automata. This notion is now defined.

A *Deterministic Büchi automaton* $\mathcal{A}$ is a 5-tuple (Q, A, δ, q_0, F), where Q is a finite *set of states*, A is an alphabet, $\delta : Q \times A \to Q$ is the *transition function*, $q_0 \in Q$ is the *initial states* and $F \subseteq Q$ is the set of *accepting states.* A state belonging to $Q \setminus F$ is said to be a *rejecting state.*

From now on in this paper, all automata are assumed to be deterministic. The function δ is implicitly extended on $Q \times A^*$ by $\delta(q, \epsilon) = q$ and $\delta(q, wa) = \delta(\delta(q, w), a)$ for $a \in A$ and $w \in A^*$.

Let $\mathcal{A}$ be an automaton and w be an infinite word. A *run* π of $\mathcal{A}$ on w is a mapping $\pi : \mathbb{N} \mapsto Q$ such that $\pi(0) = q_0$ and $\delta(\pi(i), w[i]) = \pi(i+1)$ for all $i < |w|$. The run is accepting if there exists a state $q \in F$ such that there is an infinite number of $i \in \mathbb{N}$ such that $\pi(i) = q$. Let $\mathcal{A}$ be a finite automaton. Let $L_\omega(\mathcal{A})$ be the set of infinite words w such that a run of $\mathcal{A}$ on w is accepting.

Accessibility and Recurrent States. Some definitions related to the underlying labelled graph of Büchi automata are introduced in this section. A state q is said to be *accessible* from a state q' if there exists a finite non-empty word w such that $\delta(q', w) = q$. Following [?], a state q is said to be *recurrent* if it is accessible from itself and *transient* otherwise. Transient states are called *trivial* in [?]. The *strongly connected component* of a recurrent state q is the set of states q' such that q' is accessible from q and q is accessible from q'. A strongly connected component C is said to be a *leaf* if for all $a \in A$, for all $q \in C$, $\delta(q, a) \in C$. Let C be a strongly connected component. It is said to be a cycle if for each $q \in C$, there exists a unique $s_q \in A$ such that $\delta(q, s_q) \in C$.

The transient states of the automaton pictured in Figure **??** are q_1, q_{10}, q_{11}, $q_{10\star}$ and $q_{11\star}$. All other states are recurrent. The cycles are $\{q_0\}$, $\{q_{0\star}, q_{0\star 0}\}$, $\{q_{10\star 0}\}$, $\{q_{10\star 1}, q_{10\star 10}\}$, $\{q_{11\star 0}\}$ and $\{q_{11\star 1}, q_{11\star 10}\}$. The strongly connected components which are not cycles are $\left\{q_{\emptyset,\mathcal{A}}\right\}$, $\{q_{\infty,\mathcal{A}}\}$ and $\left\{q_{[0,1],\mathcal{A}}\right\}$.

For $q \in Q$, let $\mathcal{A}_q$ be (Q_q, A, δ, q, F_q), where Q_q is the set of states of Q accessible from q, and $F_q = F \cap Q_q$. Note that, if there is no finite word w such that $\delta(q_0, w) = q_0$, then $Q_q \subsetneq Q$ for all $q \neq q_0$.

Quotients, Morphisms and Weak Büchi Automata The Büchi automaton $\mathcal{A} = (Q, A, \delta, q_0, F)$ is said to be minimal if, for each distinct states q and q' of $\mathcal{A}$, $L_\omega(\mathcal{A}_q) \neq L_\omega(\mathcal{A}_{q'})$. Let $\mathcal{A} = (Q, A, \delta, q_0, F)$ be a Büchi automaton and $\mathcal{A}' = (Q', A, \delta', q'_0, F')$ be a minimal Büchi automaton. A surjective function $\mu : Q \to Q'$ is a *morphism* of Büchi automata if and only if $\mu(q_0) = q'_0$ and, for all $q \in Q$, $L_\omega(\mathcal{A}_q) = L_\omega\left(\mathcal{A}'_{\mu(q)}\right)$.

The Büchi automaton $\mathcal{A}$ is said to be *weak* if for each recurrent accepting state q of $\mathcal{A}$, all states of the strongly connected components of q are accepting. An ω-language is said to be (weakly) *recognizable* if it is a set of word accepted by a (weak) Büchi automaton. An example of weak deterministic Büchi automaton is now given. This example is used through this paper to illustrate properties of Büchi automaton reading set of real numbers.

Example 1. Let $R = \left(\frac{1}{3}, 2\right] \cup \left(\frac{8}{3}, 3\right] \cup \left(\frac{11}{3}, \infty\right]$. The set of encodings in base 2 of reals of R is (weakly) recognized by the automaton pictured in Fig. **??**. The run

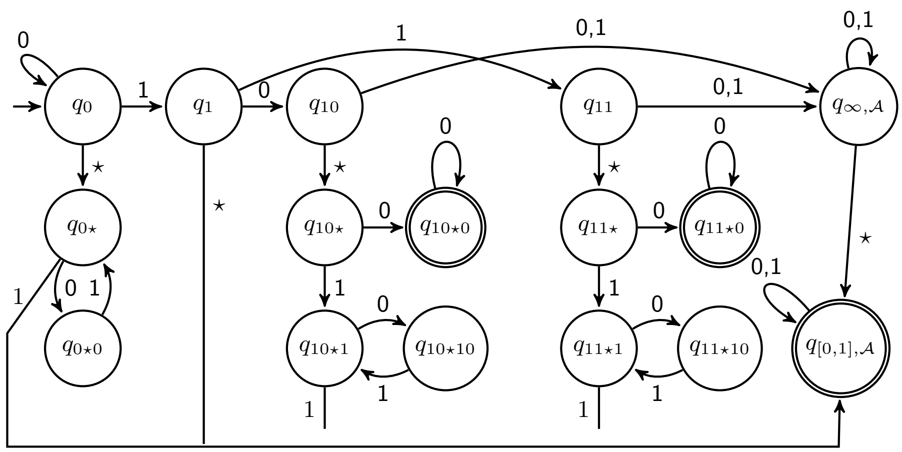

Fig. 1. Automaton $\mathcal{A}_R$ of Ex. **??**

of $\mathcal{A}$ on the ω-word $011 \star (10)^\omega$ is $(q_0, q_0, q_1, q_3, q_{11\star}, q_{11\star1}, q_{11\star10}, \dots)$, with the two last states repeated infinitely often. The Büchi automaton $\mathcal{A}$ does not accept $011\star(10)^\omega$ since this run does not contain any accepting state. The run of $\mathcal{A}$ on ω-word $\star1^\omega$ is $\left(q_0, q_{0\star}, q_{[0,1],\mathcal{A}}, \dots\right)$ with the last state repeated infinitely often. The Büchi automaton $\mathcal{A}$ accepts $\star1^\omega$ since the accepting state $q_{[0,1],\mathcal{A}}$ appears infinitely often in the run.

The main theorem concerning quotient of weak Büchi automata is now recalled.

Theorem 1 ([?]). *Let $\mathcal{A} = (Q, A, \delta, q_0, F)$ be a weak Büchi automaton with n states such that all states of $\mathcal{A}$ are accessible from its initial state. Let c be the* *cardinality of A. There exists a minimal weak Büchi automaton $\mathcal{A}'$ such that there exists a morphism of automaton μ from $\mathcal{A}$ to $\mathcal{A}'$. The automaton $\mathcal{A}'$ and the morphism μ are computable in time $O(n\log(n)c)$ and space $O(nc)$.*

The Büchi automaton $\mathcal{A}_R$ pictured in Figure **??** is weak and is not minimal. Its minimal quotient is pictured in Figure **??**. The following lemma states that

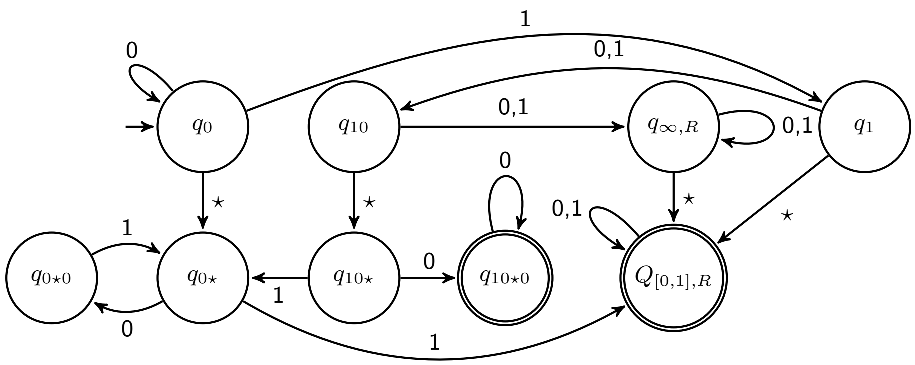

Fig. 2. Minimal quotient of automaton $\mathcal{A}_R$ of Figure **??**

each strongly connected component of a quotient by a morphism μ from an automaton $\mathcal{A}$ is the image of a strongly connected component of $\mathcal{A}$. It allows to prove that some properties, such as being a cycle, is closed under taking quotient.

Lemma 2. *Let $\mathcal{A} = (Q, A, \delta, q_0, F)$ and $\mathcal{A}' = (Q', \Sigma_b, \delta', q'_0, F')$ be two Büchi automata. Let μ be a morphism from $\mathcal{A}$ to $\mathcal{A}'$. Let C' be a strongly connected component of $\mathcal{A}'$. There exists a strongly connected component $C \subseteq Q$ such that $\mu(C) = C'$ and such that, for all $q \in Q \setminus C$ accessible from C, $\mu(q) \notin C'$.*

Real Number Automata. For $\mathcal{A}$ a Büchi automaton over alphabet $\Sigma_b \cup \{\star\}$, let $[\mathcal{A}]_b^{\mathsf{R}}$ denote $[L_\omega(\mathcal{A})]_b^{\mathsf{R}}$. It is said that $\mathcal{A}$ recognizes $[\mathcal{A}]_b^{\mathsf{R}}$. Following [**?**], a Büchi automaton over alphabet $\Sigma_b \cup \{\star\}$ is said to be a Real Number Automaton (RNA) if it recognizes a subset of $\Sigma_b^* \star \Sigma_b^\omega$ and if the language $L_\omega(\mathcal{A})$ is saturated. The Büchi automata pictured in **??** and **??** are RNA. Clearly, the RNAs are the Büchi automata which recognize saturated languages.

1.3 Logic

The logic $\mathsf{FO}[\mathbb{R}; +, <, 1]$ used in this paper is introduced in this section. FO stands for first-order. The first parameter $\mathbb{R}$ means that the (free or quantified) variables are interpreted by non-negative real numbers. The $+$ and $<$ symbols mean that the addition function and the binary order relation over reals can be used in formulas. Finally, the last term, 1, means that the only constant is 1. The logic $\mathsf{FO}\,[\mathbb{R};+,<,1]$ is denoted by $\mathscr{L}$ in [?], where it is proven that this logic admits quantifier elimination. In this paper, all results deal with the quantifier-free, the existential fragment and the existential-universal fragment of $\mathsf{FO}\,[\mathbb{R};+,<,1]$ denoted by $\Sigma_0\,[\mathbb{R};+,<,1]$, $\Sigma_1\,[\mathbb{R};+,<,1]$ and $\Sigma_2\,[\mathbb{R};+,<,1]$ respectively.

In the rest of the paper, rationals are also used in the formulas. Admitting rationals does not change the expressivity since all rational constants are $\Sigma_0\,[\mathbb{R};+,1]$-definable. The length of a formula ϕ is denoted by $|\phi|$. It is such that each symbol takes one bit of space, apart from integers n and rationals n/m which take $\log(1+|n|)$ and $\log(1+|n|+|m|)$ bits of space respectively.

First-Order Definable Sets of Reals. In this section, notations are introduced for the kind of sets studied in this paper: the $\mathsf{FO}\,[\mathbb{R};+,<,1]$-definable sets. Following [?, Sect. 6], the $\mathsf{FO}\,[\mathbb{R};+,<,1]$-definable sets are called the *simple sets*. By [?, Sect. 6], those sets are the finite unions of intervals with rational bounds. It implies that there exists an integer t_R such that for all $x,y\geq t_R$, x belongs to R if and only if y belongs to R. The least such integer t_R is called the *threshold of R*.

Note that every closed and half-closed intervals is the union of an open interval and of singletons, hence it can be assumed that any simple set R is of the form $R=\bigcup_{i=0}^{I-1}(\rho_{i,\mathfrak{L}},\rho_{i,\mathfrak{R}})\cup\bigcup_{i=0}^{J-1}\{\rho_{i,\mathfrak{S}}\}$, with $\rho_{i,\mathfrak{L}},\rho_{i,\mathfrak{S}}\in\mathbb{Q}$ and $\rho_{i,\mathfrak{R}}\in\mathbb{Q}\cup\{\infty\}$. The $\rho_{i,\mathfrak{L}}$'s are the left bounds, the $\rho_{i,\mathfrak{R}}$'s are the right bounds and the $\rho_{i,\mathfrak{S}}$'s are the singletons.

For example, let $R=\left(\frac{1}{3},2\right]\cup\left(\frac{8}{3},3\right]\cup\left(\frac{11}{3},\infty\right]$ as in Ex. **??**. Then t_R is 4, $I=3$, $J=2$, $\rho_{1,\mathfrak{L}}=\frac{1}{3}$, $\rho_{2,\mathfrak{R}}=2$, $\rho_{2,\mathfrak{L}}=\frac{8}{3}$, $\rho_{2,\mathfrak{R}}=3$, $\rho_{3,\mathfrak{L}}=11/3$, $\rho_{3,\mathfrak{R}}=\infty$, $\rho_{1,\mathfrak{S}}=2$ and $\rho_{2,\mathfrak{S}}=3$.

2 Some Sets of states of Automata Reading Reals

We now introduce five sets of states used in the algorithms of this paper.

Definition 1 **($Q_{\emptyset,\mathcal{A}}$, $Q_{[0,1],\mathcal{A}}$, $Q_{\infty,\mathcal{A}}$, $Q_{I,\mathcal{A}}$ and $Q_{F,\mathcal{A}}$).** *Let $\mathcal{A}=(Q,A,\delta,q_0,F)$.*

- *Let $Q_{\emptyset,\mathcal{A}}$ be the set of states q such that $\mathcal{A}_q$ recognizes the empty language.*
- *Let $Q_{[0,1],\mathcal{A}}$ be the set of states q such that $\mathcal{A}_q$ recognizes Σ_b^ω.*
- *Let $Q_{\infty,\mathcal{A}}$ be the set of states q such that $\mathcal{A}_q$ recognizes the language $\Sigma_b^*\star\Sigma_b^\omega$.*
- *Let $Q_{I,\mathcal{A}}$ be the set of states q such that $\mathcal{A}_q$ recognizes a subset of $\Sigma_b^*\star\Sigma_b^\omega$.*
- *Let $Q_{F,\mathcal{A}}$ be the set of states q such that $\mathcal{A}_q$ recognizes a subset of Σ_b^ω.*

In [?], the strongly connected components included in $Q_{\emptyset,\mathcal{A}}$ are called empty *and the ones included in $Q_{[0,1],\mathcal{A}}$ are called* universal.

Intuitively the states belonging to $Q_{I,\mathcal{A}}$ and to $Q_{F,\mathcal{A}}$ are the states which can be visited while the automaton read the natural and the fractional part of the number respectively.

Let $\mathcal{A}$ be the automaton pictured in Figure **??**. Let $q_{\emptyset,\mathcal{A}}$ be the state $\delta(q_{10\star 0}, 1)$, which is not pictured in Figure **??**. Then $Q_{[0,1],\mathcal{A}} = \left\{q_{[0,1],R}\right\}$, $Q_{\infty,\mathcal{A}} = \{q_{\infty,R}\}$ and $Q_{\emptyset,\mathcal{A}} = \left\{q_{\emptyset,\mathcal{A}}\right\}$. Furthermore, the states of $Q_{I,\mathcal{A}}$ are pictured in the top row of Figure **??**, they are q_0, q_1, q_{10}, $q_{\infty,R}$ and $q_{\emptyset,R}$. Finally, the states of $Q_{F,\mathcal{A}}$ are pictured in the second row of Figure **??**, they are $q_{10\star}$, $q_{10\star 0}$, $q_{0\star}$, $q_{0\star 0}$, $Q_{[0,1],R}$ and $q_{\emptyset,R}$.

In a minimal weak Büchi automaton $\mathcal{A}$, let $q_{\emptyset,\mathcal{A}}$, $q_{[0,1],\mathcal{A}}$ and $q_{\infty,\mathcal{A}}$ denote the only state q such that $\mathcal{A}_q$ recognizes the languages $\emptyset$, Σ_b^ω and $\Sigma_b^* \star \Sigma_b^\omega$ respectively. The following lemma states that the five sets introduced in Def. **??** are linear time computable.

Lemma 3. *Let $\mathcal{A}$ be a weak Büchi automaton with n states. Then the sets $Q_{\emptyset,\mathcal{A}}$, $Q_{[0,1],\mathcal{A}}$, $q_{\infty,\mathcal{A}}$, $Q_{I,\mathcal{A}}$ and $Q_{F,\mathcal{A}}$ are computable in time $O(nb)$.*

It is explained how to compute $Q_{\emptyset,\mathcal{A}}$. The other sets are computed similarly.

Proof. Tarjan's algorithm [**?**] can be used to compute the set of strongly connected component in time $O(nb)$, and therefore the set of recurrent states. By definition, $q \in Q_{\emptyset,\mathcal{A}}$ if and only if $\mathcal{A}_q$ accept no ω-word. It is equivalent to the fact that no recurrent state accessible from q are accepting. Equivalently, $Q_{\emptyset,\mathcal{A}}$ is the greatest set of states q such that, q is not a recurrent accepting state, and furthermore, for all $a \in \Sigma_b \cup \{\star\}$, $\delta(q, a) \in Q_{\emptyset,\mathcal{A}}$. This naturally leads to the following greatest fixed-point algorithm.

Two sets `PotentiallyEmpty` and `ToProcess` are used by the algorithm. The algorithm initializes the set `PotentiallyEmpty` to Q and initializes the set `ToProcess` to the empty set. The algorithm runs on each recurrent state q. For each state q, if q is accepting, then q is removed from `PotentiallyEmpty` and added to `ToProcess`. The algorithm then runs on each element q of `ToProcess`. For each state q, the algorithms removes q from `ToProcess` and runs on each predecessors q' of q. For each q', if q' is in `PotentiallyEmpty`, then q' is removed from `PotentiallyEmpty` and added to `ToProcess`. Finally, when `ToProcess` is empty, the algorithm halts and $Q_{\emptyset,\mathcal{A}}$ is the value of `PotentiallyEmpty`.

3 From simple sets to automata

Let us fix a simple non-empty set $R \subsetneq \mathbb{R}^{\geq 0}$. In this section a weak RNA $\mathcal{A}_R$ which recognize $L_\omega(R)$ is constructed. Since R is a simple set, there exists an integer $t_R \in \mathbb{N}^{\geq 0}$ such that $[t_R, \infty)$ is either a subset of R or is disjoint from R. Without loss of generality, it is assumed that $t_R \geq b$. As seen in Section **??**, R can be expressed as $\bigcup_{i=0}^{I-1}(\rho_{i,\mathfrak{L}}, \rho_{i,\mathfrak{R}}) \cup \bigcup_{i=0}^{J-1}\{\rho_{i,\mathfrak{S}}\}$ with $\rho_{i,j} \in \mathbb{Q} \cap [0, t_R]$. Without loss of generality, it can be assumed that all integers n belonging to $[0, t_R]$ are of the form $\rho_{i,j}$ for some i, j. It suffices either to assume that there is some $i \in \mathbb{N}$ such that n is of the form $\rho_{i,\mathfrak{S}}$ if $n \in R$ and of the form $\rho_{i,\mathfrak{L}}$ and $\rho_{i,\mathfrak{R}}$ otherwise.

Since the $\rho_{i,j}$ are rationals, their encodings in base b are of the form $u_{i,j,k}v_{i,j,k}^\omega$ with $u_{i,j,k} \in \Sigma_b^* \star \Sigma_b^*$ such that $u_{i,j,k}[0] \neq 0$ and $v_{i,j,k} \in \Sigma_b^+$. Since there are at most two encodings, a third index, k, is also required. Up to replacing the words $u_{i,j,k}$ by $u_{i,j,k}v_{i,j,k}^{n}$, it can be assumed without loss of generality that, for all i,j,k,i',j',k', the word $u_{i,j,k}$ is not a strict prefix of $u_{i',j',k'}$ and if $u_{i,j,k}=u_{i',j',k'}$ then $v_{i,j,k}=v_{i',j',k'}$. The formal definition of $\mathcal{A}_R$ is now given.

Definition 2 ($\mathcal{A}_R$). *Let $R \subsetneq [0,\infty)$ be a simple non-empty set. Note that $t_R > 0$. Let $\mathcal{A}_R$ be the automaton $(Q, \Sigma_b \cup \{\star\}, \delta, q_0, F)$ where:*

- *Q contains the states $q_{\emptyset,\mathcal{A}}$, $q_{[0,1],\mathcal{A}}$, $q_{\infty,\mathcal{A}}$, and a state q_w for each strict prefix w of a word $u_{i,j,k}v_{i,j,k}$.*
- *F contains $q_{[0,1],\mathcal{A}}$, and the q_w's, for $w \in \Sigma_b^* \star \Sigma_b^\omega$ some non-empty prefix of some $u_{i,\mathfrak{S},k}v_{i,\mathfrak{S},k}$.*
- *and the transition function is such that, for each word w and for each letter a:*
  - *$\delta(q_\epsilon, 0) = q_\epsilon$.*
  - *For wa a strict prefix of some $u_{i,j,k}v_{i,j,k}$, $\delta(q_w, a) = q_{wa}$.*
  - *For wa of the form $u_{i,j,k}v_{i,j,k}$, $\delta(q_w, a) = q_{u_{i,j,k}}$.*

  It is now assumed that wa is not a prefix of or equal to any $u_{i,j,k}v_{i,j,k}$.
  - *If $wa \in \Sigma_b^*$, then $\delta(q_w, a)$ is $q_{\infty,\mathcal{A}}$ if $[t_R,\infty) \subseteq R$ and $q_{\emptyset,\mathcal{A}}$ otherwise.*
  - *For $wa \in \Sigma_b^* \star \Sigma_b^*$, $\delta(q_w, a)$ is $q_{\emptyset,R}$ if $[wa0^\omega]_b^{\mathsf{R}} \notin R$ and $q_{[0,1],\mathcal{A}}$ otherwise.*
  - *For q being $q_{[0,1],R}$, $q_{\infty,R}$ or $q_{\emptyset,R}$, $\delta(q,a) = q$.*
  - *$\delta(q_{\infty,R}, \star) = q_{[0,1],R}$.*
  - *For q being $q_{[0,1],\mathcal{A}}$ or $q_{\emptyset,\mathcal{A}}$ or q_w for $w \in \Sigma_b^* \star \Sigma_b^*$, $\delta(q,\star) = q_{\emptyset,\mathcal{A}}$.*

It can be shown that $\mathcal{A}_R$ recognizes R. Let $R = \left(\frac{1}{3}, 2\right] \cup \left(\frac{8}{3}, 3\right] \cup \left(\frac{11}{3}, \infty\right]$ as in Example **??**. The automaton $\mathcal{A}_R$ is pictured in Figure **??**, without the non accepting state $q_{\emptyset,\mathcal{A}}$. Its minimal quotient is pictured in Figure **??**.

A second example is now given, which shows that the minimal number of intervals of a simple set may be exponential in the number of state of the minimal Büchi automaton accepting this set. For every non-negative integer n, let R_n be $\left\{m2^{-(n-1)} \mid m \in [2^{n-1}]\right\}$. It is the set of reals which admit an encoding w in base 2 whose suffixes $w[\geq n]$ are either equal to 0^ω or to 1^ω. This set can not be described with less than 2^{n-2} intervals and is recognized by the automaton $\mathcal{A}_n$:

$$\mathcal{A}_n = \left(\{q_i \mid i \in [n]\} \cup \left\{q_{n+1,0}, q_{n+1,1}, q_{\emptyset,\mathcal{A}}\right\}, \Sigma_b, \delta, q_0, \{q_{n+1,0}, q_{n+1,1}\}\right),$$

where the transition function is such that, for $a \in \Sigma_2$, and $i \in [n-1] \setminus \{0\}$, $\delta(q_0, \star) = q_1$, $\delta(q_i, a) = q_{i+1}$, $\delta(q_n, a) = q_{n+1,a}$, $\delta(q_{n+1,a}, a) = q_{n+1,a}$. For each state q and letter a such that $\delta(q,a)$ is not defined above, $\delta(q,a) = q_{\emptyset,\mathcal{A}}$.

4 Deciding Whether an Automaton Recognizes a Simple Set

It is explained in this section how to decide whether a minimal weak RNA accepts a simple set. The first main theorem of this paper is now given.

Theorem 2. *It is decidable in time $O(nb)$ and space $O(n)$ whether a minimal weak Büchi RNA with n states recognizes a simple set.*

In order to prove this theorem, a proposition is is now given. This property is a general method used to efficiently decide properties of automata. This method is similar to the method used in [?] and in [?].

Proposition 1. *Let $\mathbb{A}'$ be a class of weak Büchi automata and let $\mathbb{L}'$ be the class of languages $\{L_\omega(\mathcal{A}) \mid \mathcal{A} \in \mathbb{A}'\}$. Let $\mathbb{L}$ be a class of languages over an alphabet such that there exists a class $\mathbb{A}$ of weak Büchi automata such that:*

1. *there exists an algorithm α which decides in time $t(n, b)$ and space $s(n, b)$ whether a Büchi automaton belongs to $\mathbb{A}$, for n the number of states and b the number of letters,*
2. *for each $L \in \mathbb{L} \cap \mathbb{L}'$, there exists an automaton $\mathcal{A} \in \mathbb{A}$ which recognizes L,*
3. *the minimal quotient of any automaton of $\mathbb{A}$ belongs to $\mathbb{A}$ and*
4. *every language recognized by an automaton belonging to $\mathbb{A}$ belongs to $\mathbb{L}$.*

The algorithm α decides in time $t(n, b)$ and space $s(n, b)$ whether a minimal automaton of $\mathbb{A}'$ recognizes a language of $\mathbb{L}$. Furthermore, the algorithm α applied to an automaton belonging to $\mathbb{A}' \setminus \mathbb{A}$ may not return a false positive.

Proof. Let $\mathcal{A}$ be an automaton which recognizes a language L. Let us assume that α accepts $\mathcal{A}$, by Prop. (**??**), $\mathcal{A} \in \mathbb{A}$, hence by Prop. (**??**), $L \in \mathbb{L}$.

Let us now assume that $\mathcal{A} \in \mathbb{A}'$ and that $L \in \mathbb{L}$. By definition of $\mathbb{L}'$, $L \in \mathbb{L}'$, hence $L \in \mathbb{L} \cap \mathbb{L}'$, thus by Prop **??**, there exists $\mathcal{A}' \in \mathbb{A}$ which recognizes L. Since $\mathcal{A}'$ and $\mathcal{A}$ recognize the same language, they have the same minimal quotient, which is $\mathcal{A}$. By Prop. **??**, $\mathcal{A} \in \mathbb{A}$. Thus, by Prop. (**??**), α accepts $\mathcal{A}$.

In this paper, $\mathbb{A}'$ is the set of RNAs, hence $\mathcal{L}'$ is the class of saturated recognizable languages. The class of languages $\mathcal{L}$ is the class of base b encoding of non-empty sets $R \subsetneq \mathbb{R}^{\geq 0}$. The cases of $R = \mathbb{R}^{\geq 0}$ and of $R = \emptyset$ being special cases. The class $\mathbb{A}$ of automata is now introduced.

Definition 3 ($\mathbb{A}$). *Let $\mathbb{A}$ be the set of weak Büchi automata $\mathcal{A}$, of the form $(Q, \Sigma_b \cup \{\star\}, \delta, q_0, F)$, such that, for each strongly connected component $C \subseteq Q_{F,\mathcal{A}} \setminus (Q_{[0,1],\mathcal{A}} \cup Q_{\emptyset,\mathcal{A}})$, there exists $\beta_{<,C}$ and $\beta_{>,C}$, two states of $Q_{[0,1],\mathcal{A}} \cup Q_{\emptyset,\mathcal{A}}$, such that, for all $q \in C$:*

1. *C is a cycle. Recall that s_q is the only letter such that $\delta(q, s_q) \in C$.*
2. *For all $a > s_q$, $\delta(q, a)$ is $\beta_{>,C}$.*
3. *For all $a < s_q$, $\delta(q, a)$ is $\beta_{<,C}$.*
4. *There exists an accepting and a rejecting strongly connected component, accessible from the initial state, belonging to $Q_{F,\mathcal{A}}$.*
5. *The set $Q_{\emptyset,\mathcal{A}}$ contains exactly one recurrent state, denoted $q_{\emptyset,\mathcal{A}}$.*
6. *The set $Q_{\infty,\mathcal{A}}$ contains at most one recurrent state, denoted $q_{\infty,\mathcal{A}}$.*
7. *$\delta(q_0, 0) = q_0$.*
8. *$\delta(q_0, a) \neq q_0$ for all $0 < a < b$.*
9. *If $q_{\infty,\mathcal{A}}$ exists, then $\delta(q, a) \neq q_{\emptyset,\mathcal{A}}$ for all $q \in Q_{I,\mathcal{A}} \setminus \{q_{\emptyset,\mathcal{A}}\}$ and $a \in \Sigma_b$.*
10. *The recurrent states of $Q_{I,\mathcal{A}}$ are $q_{\emptyset,\mathcal{A}}$, q_0 and potentially $q_{\infty,\mathcal{A}}$.*

The automata of $\mathbb{A}$ admits the following property.

Lemma 4. *Let $\mathcal{A} \in \mathbb{A}$ be an automaton with n states recognizing a set R. If $\mathcal{A}$ contains a state $q_{\infty,\mathcal{A}}$, as in Definition **??**, then $(b^{n-1}, \infty) \subseteq R$, otherwise $(b^{n-1}, \infty) \cap R = \emptyset$.*

*Proof (Sketch of proof of Theo. **??**).* Using Lem. **??**, the algorithms checks whether $\mathcal{A}$ accepts a subset L of $\Sigma_b^* \star \Sigma_b^\omega$, if it is not the case, the algorithm rejects. The algorithms also checks whether L is $\emptyset$ or $\Sigma_b^* \star \Sigma_b^\omega$. If it is the case, the algorithm accepts. It is now assumed that $\mathcal{A}$ accepts a non-empty language $L \subsetneq \Sigma_b^* \star \Sigma_b^\omega$. Let $\mathbb{L}'$ be the set of saturated languages and $\mathbb{A}'$ be the set of RNAs. Let $\mathbb{L}$ be the set of encoding of simple non-empty sets $R \subsetneq \mathbb{R}^{\geq 0}$. In order to prove this theorem, it suffices to show that $\mathbb{A}$ admits the four properties of Proposition **??**.

Each property of Def. **??** is testable in time $O(nb)$ and space $O(n)$. Therefore, it is decidable in time $O(nb)$ and space $O(n)$ whether a weak Büchi automaton $\mathcal{A}$ with n states belongs to $\mathbb{A}$. Hence Property (**??**) of Prop. **??** holds.

For $R \subsetneq \mathbb{R}^{\geq 0}$ a non-empty simple set, the automaton $\mathcal{A}_R$ of Def. **??** belongs to $\mathbb{A}$. Therefore Property (**??**) of Prop. **??** holds.

Let $\mathcal{A} \in \mathbb{A}$ be a RNA. Let $\mathcal{A}'$ be its minimal quotient. It can be proven that $\mathcal{A}'$ satisfies the properties of Def. **??**, hence $\mathcal{A}' \in \mathbb{A}$. Therefore Property (**??**) of Prop. **??** holds.

Property (**??**) of Prop. **??** is now considered. Automata satisfying Properties (**??**), (**??**) and (**??**) of Def. **??** are studied in [**?**]. It is shown that automata satisfying those properties accepts a set R such that $R \cap [i, i+1]$ is a finite union of intervals with rationals boundaries for all $i \in \mathbb{N}$. Lemma **??** ensures that furthermore, there is some $t \in \mathbb{N}$ such that $[t, \infty)$ is either a subset of R or is disjoint from R. Thus, an automaton of $\mathbb{A}$ recognize a finite union of interval with rational boundaries, i.e. a simple set. Therefore Property (**??**) of Prop. **??** holds. ☐

The algorithm of Theo. **??** takes as input a minimal weak RNA and runs in time $O(nb)$. It should be noted that it is not known whether it is decidable in time $O(nb)$ whether a minimal Büchi automaton is a RNA. However, if the algorithm of Theo. **??** is applied to a weak Büchi automaton which is not a Real Number Automaton, the algorithm returns no false positive. An example of false negative is now given. The not-saturated language $L = (00)^* (01 + 2) \Sigma_3^* \star \Sigma_3^\omega$ encode the simple set of reals $\mathbb{R}^{>0}$. However, the minimal automaton recognizing L it is not accepted by the algorithm of Theo. **??**.

5 From Automata to Simple Set

It is explained in this section how to compute a first-order formula which defines the simple set accepted by a weak RNA. The exact theorem is now stated.

Theorem 3. *Let $\mathcal{A} = (Q, \Sigma_b \cup \{\star\}, \delta, q_0, F)$ be a be a minimal weak RNA with n states which recognizes a simple set. There exists a $\Sigma_1[\mathbb{R}; +, <, 1]$-formula computable in time $O(n^2 b \log(nb))$ which defines $[\mathcal{A}]_b^{\mathsf{R}}$. There exists a $\Sigma_2[\mathbb{R}; +, <, 1]$-formula computable in time $O(nb \log(nb))$ which defines $[\mathcal{A}]_b^{\mathsf{R}}$.*

The proof of Theo. **??** consists mostly in encoding in a first-order formula $\phi(x)$ the run of $\mathcal{A}$ over an encoding w of x. The following lemma allows to consider two distinct part of the run on the fractional part of w. The first part of the run is of length at most n. The second part on the run begins on a state belonging to a restricted set of states.

Lemma 5. *Let $A \in \mathbb{A}$ be minimal with n states and $q \in Q_{F,\mathcal{A}}$. Let $w_I \in \Sigma_b^*$ and $w_F \in \Sigma_b^\omega$. Let $\mathcal{Q} \subseteq Q_{F,\mathcal{A}}$ be a set containing exactly one state of each strongly connected component. Then, there exists $s \in [n]$ such that $\delta(q, w_I{\star}w_F\,[< s]) \in \mathcal{Q}$.*

The following lemma allows to reduce the size of a formula by adding quantifications.

Lemma 6. *Let $\psi(x, x')$ be a formula of length l and $(x_i)_{i\in[n-1]}$ be n variables. Then $\bigwedge_{i\in[n-2]} \psi(x_i, x_{i+1})$ is equivalent to the following formula of length $O(n + l)$:*

$$\forall y, y'. \left\{ \bigvee_{i\in[n-2]} [y \doteq x_i \wedge y' \doteq x_{i+1}] \right\} \implies \psi(y, y').$$

A sketch of the proof of Theo. **??** can now be given.

*Proof (Proof of Theo. **??**).* Let $R = [\mathcal{A}]_b^{\mathsf{R}}$. As shown in Sect. **??**, it can be assumed that $\mathcal{A}$ belongs to $\mathbb{A}$. By Lem. **??**, in order to construct a formula which defines R it suffices to construct a formula $\phi(x)$ which defines $R \cap [0, b^{n-1})$. The formula $\phi(x)$ is the conjunction of four subformulas of size $O\left(n^2 b \log(nb)\right)$. Let $x \in [0, b^{n-1})$ and let w be an encoding of x without leading 0 in the natural part.

The first formula, $\psi(x, x_I, x_F)$, states that $x = x_I + x_F$ and that $x_I \in \mathbb{N}$. Since $x_I < b^{n-1}$, in order to state that $x_I \in \mathbb{N}$, it suffices to state that x_I is of the form $\sum_{i=0}^{n-2} a_i b^i$ for $a_i \in [b-1]$. More precisely, it suffices to state that x_I is of the form $(c_n + b(c_{n-1} + b(\cdots + b(c_0)\ldots)))$ with the c_i belonging to $[b-1]$. This can be stated by existentially quantifying the $2n$ partial sums and products and taking disjunctions over each c_i. This can be done by a formula $\psi(x_I)$ of size $O(nb\log(b))$.

Let q be the state $\delta(q_0, w_I)$. The second formula, $\phi_I(x_I, q)$, states that the state $\delta(q_0, w_I\star)$ is equal to q. This formula existentially quantifies $2n$ variables. Those variables encode the n first steps of the runs and the values of $w_I\,[< i]$ for $i < n$. Each step of the computation can be encoded by a $\Sigma_0\,[\mathbb{R}; +, <, 1]$-formula of length $O(nb\log(b))$, using the equalities of Lem. **??**. Since $x_I < b^{n-1}$, $|w_I| < n$, the formula $\phi_I(x_I, q)$ have to consider at most n steps of the computation. The formula $\phi_I(x_I, q)$ is a conjunction of n formulas of size $O\left(n^2 b \log(b)\right)$ and thus the size of $\phi_I(x_I, q)$ is $O\left(n^2 b \log(b)\right)$.

Let $\mathcal{Q}$ be a set of states as in Lem. **??** and let q' be the first state of $\mathcal{Q}$ in the run of $\mathcal{A}$ on w. The third formula, $\phi_F(q, x_F, q', x'_F)$ states that there exists $i \in [n]$ such that $\delta(q, w_F\,[< i]) = q'$, that $q' \in \mathcal{Q}$ and that $x'_F = [w_F\,[\geq i]]_b^F$. By Lem. **??**, i is at most n. Hence, similarly to $\phi_I(x_I, q)$, the size of the formula $\phi_F(q, x_F, q', x'_F)$ is $O\left(n^2 b \log(b)\right)$.

Finally, the fourth formula $\phi'_F(q', x'_F)$, states that $\mathcal{A}_{q'}$ accepts $w_F\,[\geq i]$. Let c be the number of strongly connected components in $\mathcal{A}$. For C a strongly connected components, let n_C be its number of state and q_C the only state of $C \cap \mathcal{Q}$. Let us assume that, for each strongly connected component C, there exists a formula $\phi'_C(q', x'_F)$ of length $O\left(n_C b \log(n_C b)\right)$ which states that $\mathcal{A}_{q'}$ accepts $w_F\,[\geq i]$. Then the formula $\phi'_F(q', x'_F)$ is a disjunction of c formulas $q' \doteq q_C \wedge \phi'_C(q', x'_F)$ and its length is $O\left(\sum_C n_C b \log(n_C b)\right) = O\left(nb \log(nb)\right)$.

It is now explained how to construct $\phi'_C(q', x'_F)$. Since $\mathcal{A} \in \mathbb{A}$, by Prop. **??** of Def. **??**, strongly connected components of automata included in $Q_{F,\mathcal{A}}$ are either $\left\{q_{\emptyset,\mathcal{A}}\right\}$, $\left\{q_{[0,1],\mathcal{A}}\right\}$, or a cycle. In the first two cases, $\phi'_C(q', x'_F)$ is the formula False or True respectively. Let us consider the third cases. Let $v_{q'}$ be the word of size n_C such that $\delta(q', v_{q'}) = q'$. Since C is cycle, this word exists and is a unique. Then let $y = \left[v^{\omega}_{q'}\right]^F_b = \left[v_{q'}\right]^I_b / \left(b^{n_C} - 1\right)$. Recall that the notations $\beta_{<,C}$ and $\beta_{>,C}$ are introduced in Def. **??**. Then the formula $\phi'_C(q', x'_F)$ states that $q' \in C'$ and that either ($x'_F < y$ and $\beta_{<,C} \in Q_{[0,1],\mathcal{A}}$), either ($x'_F = y$ and $q' \in F$), or ($x'_F > y$ and $\beta_{>,C} \in Q_{[0,1],\mathcal{A}}$). It is indeed a formula of length $O\left(n_C b \log(n_C b)\right)$.

Finally, in order to reduce the size of the formula to $O\left(nb \log(nb)\right)$, it suffices to replace the conjunctions of $\phi_I(x_I, q)$ and of $\phi_F(q, x_F, q' x'_F)$ by a universal quantifications, as explained in Lem. **??**. □

6 Conclusion

In this paper, we proved that it is decidable in linear time whether a minimal weak Büchi Real Number Automaton $\mathcal{A}$ reading a set of real number R recognizes a finite union of intervals. It is proven that a quasi-linear sized existential-universal formula defining R exists. And that a quasi-quadratic sized existential formula defining R also exists.

The theorems of this paper lead us to consider two natural generalization. We intend to adapt the algorithm of this paper to similar problems for automata reading vectors of reals instead of automata reading reals. We also intend to solve the similar problem of deciding whether an RNA accepts a $\mathsf{FO}\left[\mathbb{R}, \mathbb{N}; +, <\right]$-definable set of reals. Solving this problem requires to solve the problem of deciding whether an automaton reading natural number, beginning by the most-significant digit, recognizes an ultimately-periodic set. Similar problems has already been studied, see e.g [**?**,**?**] and seems to be difficult. Finally, we also intend to consider how to efficiently decide whether an automaton is a Real Number Automaton or a Real Vector Automaton.

The author thanks Bernard Boigelot, for a discussion about the algorithm of Theo. **??**, which led to a decrease of the computation time. He also thanks the anonymous referees of for their remarks and suggestion to improve the paper.