

Deciding Reachability in Automata on Words Indexed by the Reals and Rationals^{*}

Bernard Boigelot^a, Pascal Fontaine^a, Baptiste Vergain^{a,*}

^a*Montefiore Institute, B28, Université de Liège, Liège, Belgium*

Abstract

Automata on linear orderings generalize the notions of finite-word and infinite-word automata. They recognize words indexed by linear orderings, defined as mappings from the elements of such orderings to a finite alphabet. The reachability problem has already been solved for automata on scattered linear orderings, i.e., those that do not involve density, as well as for automata over all linear orderings. The main contribution of this work is to provide a polynomial algorithm that decides reachability over the specific linear orderings $\mathbb{R}$ and $\mathbb{Q}$. Given an automaton on linear orderings and a pair (p, q) of its states, this procedure decides in polynomial time whether this automaton admits a path from p to q labeled by a word indexed by $\mathbb{R}$ or $\mathbb{Q}$. This also solves the non-emptiness problem for the language accepted by these automata restricted to the orderings of the reals and rationals. This algorithm constitutes an essential step towards designing an automata-based decision procedure for the first-order monadic theory of order interpreted over $\mathbb{R}$ or $\mathbb{Q}$.

Keywords: Automata, Linear Orderings, Real Domain, Rational Domain, Reachability, Non-Emptiness Test

1. Introduction

In [1], Bruyère and Carton introduce automata on linear orderings, that generalize the notions of finite-word, infinite-word and ordinal-word automata. They recognize words indexed by linear orderings, defined as mappings from the elements of such orderings to a finite alphabet. Although initially restricted to scattered linear orderings (i.e., orderings that do not contain a dense sub-ordering), these automata have later been extended to deal with all linear orderings [2]. Notably, a Kleene-like theorem asserting the equivalence between languages accepted by automata on linear orderings and languages described by

^{*}This work is partially supported by the FNRS-DFG PDR Weave (SMT-ART) grant 40019202.

^{*}Corresponding author

an extended form of rational expressions is proved in [1, 2]. The strong connections between rational languages and the monadic second-order theory of linear orderings $\text{MSO}(<)$ is investigated in [3], namely, every language accepted by an automaton on linear orderings is definable in $\text{MSO}(<)$. The converse property however only holds when the orderings are countable and scattered.

We aim at designing effective decision procedures for monadic first-order theories of order $\text{MFO}(D, <)$ interpreted over fixed dense domains D , in particular $\mathbb{R}$ and $\mathbb{Q}$. The decidability of $\text{MFO}(\mathbb{Q}, <)$ derives directly from the decidability of $\text{MSO}(\mathbb{Q}, <)$ [4]. For $\mathbb{R}$ however, Shelah [5] showed that $\text{MSO}(\mathbb{R}, <)$ is undecidable. The decidability of $\text{MFO}(\mathbb{R}, <)$ has first been established in [6]. Another proof of this result is presented in [7], together with a symbolic notation system for models of satisfiable formulas. Some theoretical bounds for the complexity of decision procedures for $\text{MFO}(\mathbb{Q}, <)$ and $\text{MFO}(\mathbb{R}, <)$ have been obtained in [8, 9].

In [10] Büchi introduced a decision procedure for the monadic theory of one successor S1S , which roughly corresponds to $\text{MSO}(\mathbb{N}, <)$, and later extended it to the domain of countable ordinals [11]. We believe that the same approach can be used to tackle $\text{MFO}(\mathbb{R}, <)$ and $\text{MFO}(\mathbb{Q}, <)$, relying on automata on linear orderings. The first step would consist in building an automaton recognizing the models of a given formula without any restriction on the domain. Then, a dedicated procedure would decide the non-emptiness of the language over a specific linear ordering, that is, $\mathbb{R}$ or $\mathbb{Q}$ in our case. This work focuses on the latter operation.

Testing for the non-emptiness of the language accepted by an automaton on linear orderings reduces to deciding whether an accepting state is reachable from an initial state. Reachability for automata on scattered linear orderings is discussed in [12]. A generalization over arbitrary linear orderings is sketched in [13]. In this paper, we design a procedure that decides reachability over the specific domains $\mathbb{R}$ and $\mathbb{Q}$. A preliminary version of this result appeared in [14].

The paper is structured as follows. We first recall in Section 2 the useful definitions related to linear orderings, rational expressions, and automata on linear orderings. In Section 3, we discuss the state of the art regarding the reachability problem. In Section 4, we provide an algorithm for deciding reachability over $\mathbb{R}$ in polynomial time. Section 5 adapts the previous algorithm to deciding reachability over $\mathbb{Q}$. A correctness proof is provided in Section 6.

2. Preliminaries

This section first provides basic definitions and results about linear orderings; we refer to [15] for further details. It then introduces the notions of words indexed by linear orderings, rational expressions on linear orderings, and automata on linear orderings [1, 2].

2.1. Linear Orderings

A *linear ordering* J is a totally ordered set, i.e., a set equipped with a binary relation $<_J$ that is total, irreflexive, asymmetric, and transitive. Two

linear orderings J and K — respectively associated with the order relations $<_J$ and $<_K$ — are *order-isomorphic* if there exists an order-preserving bijection between J and K . Formally, let b be such a bijection (from J to K), then $b(j_1) <_K b(j_2)$ iff $j_1 <_J j_2$ for all $j_1, j_2 \in J$. We denote by $-J$ the *backwards* linear ordering that corresponds to J with its ordering reversed. The class of all linear orderings is denoted by $\mathcal{L}$.

The *order type* of a linear ordering J is the class of all linear orderings that are order-isomorphic to J . The order types of a singleton, the set $\{1, 2, \dots, n\}$, $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, and $\mathbb{R}$ are respectively denoted by 1 , n , ω , ζ , η , and λ . Notice that the order type of any non-empty open interval of $\mathbb{R}$ is λ , and that the order type of any non-empty open interval of $\mathbb{Q}$ is η .

The concatenation of two linear orderings J and K (with respective order relations $<_J$ and $<_K$) is denoted by $J + K$. It corresponds to the linear ordering composed of the set of pairs $\{(j, 1) \mid j \in J\} \cup \{(k, 2) \mid k \in K\}$, equipped with the order relation $<$ defined by $(j_1, 1) < (j_2, 1)$ if $j_1 <_J j_2$, $(k_1, 2) < (k_2, 2)$ if $k_1 <_K k_2$, and $(j, 1) < (k, 2)$ for every $j \in J$ and $k \in K$.

More generally, let K and J_k for all $k \in K$ be linear orderings. The linear ordering $\Sigma_{k \in K} J_k$ is obtained by concatenating the orderings J_k w.r.t. the ordering K . Formally, $\Sigma_{k \in K} J_k$ is the linear ordering defined over the set of pairs (j, k) such that $j \in J_k$ and $k \in K$, with the order relation $(j_1, k_1) < (j_2, k_2)$ iff either $k_1 <_K k_2$, or $k_1 = k_2$ and $j_1 <_{J_{k_1}} j_2$. When $J_k = J$ for every $k \in K$, we may write $J \cdot K$ in place of $\Sigma_{k \in K} J_k$. These operators naturally extend to order types. For instance, the order type $\omega \cdot \omega$ is the class of all linear orderings that are order-isomorphic to the linear ordering $\mathbb{N} \cdot \mathbb{N}$ (whose corresponding set is $\mathbb{N} \times \mathbb{N}$).

A *cut* of a linear ordering J is a partition of J into a pair (K, L) of sets such that for every $(k, \ell) \in K \times L$, one has $k < \ell$. The set of all cuts of an ordering J is denoted by $\hat{J}$. Notice that $\hat{J}$ is a linear ordering as well, with $(K_1, L_1) < (K_2, L_2)$ iff $K_1 \subsetneq K_2$, and that $\hat{J}$ always has a first element $(\emptyset, J)$ and a last element $(J, \emptyset)$. For every $j \in J$, the *consecutive* cuts $(\{k \mid k < j\}, \{ \ell \mid j \leq \ell \})$ and $(\{k \mid k \leq j\}, \{ \ell \mid j < \ell \})$ are (respectively) denoted by j^- and j^+ . Let $c = (K, L)$ be a cut that is neither the first nor the last one of J , i.e., $K \neq \emptyset$ and $L \neq \emptyset$. If K admits a greatest element and L admits a smallest element, then c is called a *jump*. If K does not admit a greatest element and L does not admit a smallest element, then c is called a *gap*.

A linear ordering J is (*Dedekind*) *complete* if the set $\hat{J}$ does not contain any gap. For instance, $\mathbb{N}$ and $\mathbb{R}$ are complete, while $\mathbb{Q}$ is incomplete. A linear ordering J is *dense* if between any two distinct elements lies another one, i.e., for every $j_1, j_2 \in J$ such that $j_1 <_J j_2$, there exists j_3 such that $j_1 <_J j_3 <_J j_2$. Notice that a linear ordering is dense iff it does not admit two consecutive elements. A linear ordering is *scattered* if it does not contain any dense subordering. Given a dense linear ordering J , a subordering $J' \subseteq J$ is said to be *dense in* J if between any two distinct elements of J lies an element of J' . For instance, $\mathbb{Q}$ and $\mathbb{R} \setminus \mathbb{Q}$ are both dense in $\mathbb{R}$.

Example 1. Consider the linear ordering $\mathbb{N} + \mathbb{Z}$ consisting in the set of natural

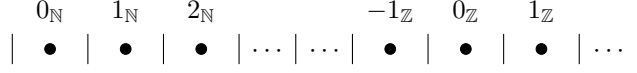


Figure 1: The linear ordering $\mathbb{N} + \mathbb{Z}$.

numbers followed by the set of integers. A depiction of this set and its cuts is given in Figure 1. The elements of this ordering are represented as dots, and its cuts as vertical lines. Notice that this ordering admits a single gap, hence it is incomplete. It is also scattered.

2.2. Words and Rational Expressions

A *word indexed by a linear ordering* is a totally ordered sequence of letters. Formally, given an alphabet Σ and a linear ordering J , a word $w = (\alpha_j)_{j \in J}$ indexed by J is a mapping $w : J \rightarrow \Sigma$. The *length* $|w|$ of w is J itself. The *empty word* ε denotes the word indexed by the empty set. Two words $w_1 = (\alpha_j)_{j \in J}$ and $w_2 = (\beta_k)_{k \in K}$ are considered to be equal if there exists an order-preserving bijection b from J to K such that for all $j \in J$, $\alpha_j = \beta_{b(j)}$. Informally, for a given word w , only the order type of its underlying linear ordering $|w|$ is relevant, and not the linear ordering $|w|$ itself.

We now define a product operator on words. Let K and J_k for all $k \in K$ be linear orderings. Let $w_k = (\alpha_{j,k})_{j \in J_k}$ be a word of length J_k , for every $k \in K$. The *product* $\prod_{k \in K} w_k$ is the word $w = (\alpha_{j,k})_{(j,k) \in |w|}$ of length $|w| = \sum_{k \in K} J_k$. The product of two words w_1 and w_2 is denoted by $w_1 \cdot w_2$.

A generalized notion of rational expressions has been introduced in [1, 2] to describe languages of words indexed by linear orderings. Recall that in this paper, we consider a fixed domain that is either $\mathbb{R}$ or $\mathbb{Q}$. For this reason, we only introduce the rational operators that are relevant to describe words indexed by $\mathbb{R}$ or $\mathbb{Q}$. They include the usual rational operators: α (for any $\alpha \in \Sigma$), union, concatenation, and finite repetition. Let X be a class of words on linear orderings. We define the following additional operators:

- *Infinite repetition*: $X^\omega = \{\prod_{i \in \mathbb{N}} x_i \mid \forall i \in \mathbb{N}. x_i \in X\}$,
- *Reverse infinite repetition*: $X^{-\omega} = \{\prod_{i \in (-\mathbb{N})} x_i \mid \forall i \in -\mathbb{N}. x_i \in X\}$.

We also define a *shuffle* operator that generates density. Let $\Sigma^\diamond$ denote the class of all the words indexed by any linear ordering¹ over the alphabet Σ .

Let $X_1, \dots, X_n \subseteq \Sigma^\diamond$, for some $n \geq 1$. We denote by $\text{sh}(X_1, \dots, X_n)$ the class of words of $\Sigma^\diamond$ that can be written as $\prod_{r \in R} x_r$ such that (1) R is a non-empty, dense, and complete linear ordering without first or last element, and (2) there exists a partition $\{R_1, \dots, R_n\}$ of R such that every R_i is dense in R , and for every $r \in R$ and $i \in [1, n]$, if $r \in R_i$ then $x_r \in X_i$.

¹This definition slightly differs from the one in [1]: it implies $\varepsilon \in \Sigma^\diamond$, while the original definition excludes the empty word from the class $\Sigma^\diamond$.

Notice that the order type of the dense and complete ordering R mentioned in the shuffle definition can differ from λ . This will be further discussed in Section 4.1. Furthermore, as shown in [2], although the shuffle operator relies on a complete ordering R , it is possible for a word $w \in \text{sh}(X_1, \dots, X_n)$ to be such that $|w|$ is incomplete. This occurs when there exists $i \in [1, n]$ such that $\varepsilon \in X_i$, and there exists $j \in [1, n]$ such that $X_j \neq \{\varepsilon\}$. Intuitively, every element of R_i corresponding to ε does not belong to $|w|$, but matches a gap in $\widehat{|w|}$.

2.3. Automata

Definition 1. An *automaton on linear orderings* is a tuple $\mathcal{A} = (Q, \Sigma, I, F, \Delta^s, \Delta^\ell)$, where Q is a finite set of states, Σ is a finite alphabet, $I \subseteq Q$ is a set of initial states, $F \subseteq Q$ is a set of accepting states, $\Delta^s \subseteq Q \times \Sigma \times Q$ is a set of *successor transitions*, and $\Delta^\ell \subseteq ((2^Q \setminus \{\emptyset\}) \times Q) \cup (Q \times (2^Q \setminus \{\emptyset\}))$ is a set of (*left and right*) *limit transitions*.

The set Δ^s contains *successor transitions* of the form (q, α, q') , with $q, q' \in Q$ and $\alpha \in \Sigma$, alternatively written $q \xrightarrow{\alpha} q'$, which are similar to the transitions of finite-word and infinite-word automata. Notice however that automata on linear orderings do not admit ε -transitions. The set Δ^ℓ comprises *left-limit transitions* of the form (P, q) , with $P \subseteq Q$ and $q \in Q$, written $P \rightarrow q$, and *right-limit transitions* of the form (q, P) , with $q \in Q$ and $P \subseteq Q$, written $q \rightarrow P$. We sometimes write sets of limit transitions in a more compact form, e.g., $p, q \rightarrow \{p, q\} \rightarrow p$ in place of the set $\{p \rightarrow \{p, q\}, q \rightarrow \{p, q\}, \{p, q\} \rightarrow p\}$. The *size* of a limit transition $q \rightarrow P$ or $P \rightarrow q$ is equal to the cardinality $|P|$ of the set P .

We now define the notion of *path* in such an automaton. Let J be a linear ordering, Q a finite set of states, and $\pi = (q_c)_{c \in \widehat{J}}$ a word over the alphabet Q (i.e., each cut $c \in \widehat{J}$ is associated with an element of Q). The *left-limit set* and *right-limit set* of π at c are the two subsets $\lim_c^- \pi$ and $\lim_c^+ \pi$ of Q defined as follows:

- $\lim_c^- \pi = \{q \in Q \mid \forall i < c. \exists k. i < k < c \wedge q = q_k\},$
- $\lim_c^+ \pi = \{q \in Q \mid \forall c < i. \exists k. c < k < i \wedge q = q_k\}.$

The left limit of π at c is therefore the set of states that are visited by π before c and infinitely close to c . If c has a predecessor in the ordering $\widehat{J}$, then this set is empty (intuitively it means that a *successor transition*, and not a *limit transition*, must be taken to reach the state mapped to c). The case of a right limit is handled symmetrically. For the first cut $c_{\min}$ and the last cut $c_{\max}$ of J , we deviate from the previous definition by setting $\lim_{c_{\min}}^- \pi = \emptyset$, and $\lim_{c_{\max}}^+ \pi = \emptyset$.

Definition 2. Let $\mathcal{A} = (Q, \Sigma, I, F, \Delta^s, \Delta^\ell)$ be an automaton and $w = (\alpha_j)_{j \in J}$ a word of length J . A *path π labeled by w* is a sequence of states $\pi = (q_c)_{c \in \widehat{J}}$ of length $\widehat{J}$ such that:

- For every $j \in J$, $(q_{j-}, \alpha_j, q_{j+})$ is a successor transition in Δ^s .
- For every cut $c \in \widehat{J}$ that is not the first cut and does not have a predecessor in $\widehat{J}$, $\lim_c^- \pi \rightarrow q_c$ is a left-limit transition in Δ^ℓ .
- For every cut $c \in \widehat{J}$ that is not the last cut and does not have a successor in $\widehat{J}$, $q_c \rightarrow \lim_c^+ \pi$ is a right-limit transition in Δ^ℓ .

A path is *accepting* if it starts in a state $p \in I$ and ends in a state $q \in F$. A word is accepted by $\mathcal{A}$ if it labels an accepting path. A path is *k-limited* if each limit transition $P \rightarrow q$ and $q \rightarrow P$ in the path is such that $|P| \leq k$.

Note that a *finite-word automaton* can be seen as an automaton on linear orderings that has an empty set of limit transitions, i.e., such that $\Delta^\ell = \emptyset$. Infinite-word automata [10, 16] and automata on ordinal words [11, 17] also correspond to particular cases of automata on linear orderings. Both correspond to automata with only left-limit transitions, with the additional constraint for the former that every left-limit transition leads to a unique accepting state that admits no outgoing transition.

Example 2. The automaton depicted in Figure 2 accepts $\{w\}$, where w is the word $w = a^\omega \cdot b^{-\omega} \cdot b^\omega$ defined over the alphabet $\{a, b\}$. Notice that w is indexed by the linear ordering $\mathbb{N} + \mathbb{Z}$ considered in Example 1. An accepting path can be described as follows: it maps the cuts (including the first one) of the first subordering $\mathbb{N}$ to the state q_1 , the gap to the state q_2 , the cuts of the subordering $\mathbb{Z}$ to the state q_3 , and the last cut to the state q_4 .

Example 3. Consider the automata depicted in Figures 3a and 3b. Both accept a non-empty language. The automaton in Figure 3a accepts in particular words indexed by $\mathbb{R}$, and more generally by any non-empty dense and complete ordering J with no first and no last element, as long as the set of indices labeled by the letter a and those labeled by b partition J , and are both dense in J . Consider for instance the word $w : \mathbb{R} \rightarrow \{a, b\}$, such that $w(x) = a$ if $x \in \mathbb{Q}$, and $w(x) = b$ if $x \in \mathbb{R} \setminus \mathbb{Q}$. Let us describe an accepting path π labeled by w . For the sake of clarity, we abbreviate the first cut $(\emptyset, \mathbb{R})$ of $\mathbb{R}$ by $-\infty$ and its last cut $(\mathbb{R}, \emptyset)$ by $+\infty$. We set $\pi(-\infty) = q_1$ and $\pi(+\infty) = q_6$. Since $\mathbb{R}$ is complete, every other cut of $\mathbb{R}$ is of the form r^- or r^+ for some $r \in \mathbb{R}$. If $r \in \mathbb{Q}$,

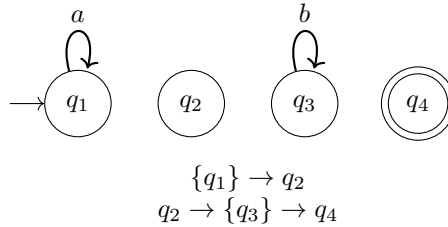


Figure 2: An automaton accepting $a^\omega \cdot b^{-\omega} \cdot b^\omega$.

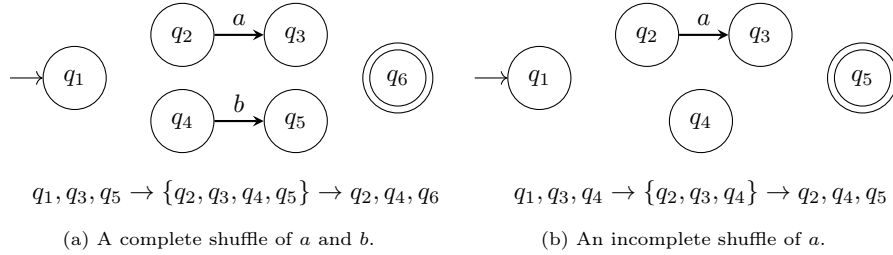


Figure 3: Automata accepting words indexed by dense orderings.

then we set $\pi(r^-) = q_2$ and $\pi(r^+) = q_3$. If $r \in \mathbb{R} \setminus \mathbb{Q}$, then we set $\pi(r^-) = q_4$ and $\pi(r^+) = q_5$. The automaton depicted in Figure 3a does not however accept words indexed by $\mathbb{Q}$, or any incomplete ordering, since this automaton does not contain any state that can be mapped onto a gap. The language accepted by this automaton is described by the rational expression $\text{sh}(a, b)$.

On the other hand, the automaton depicted in Figure 3b accepts words indexed by $\mathbb{Q}$, but does not accept any word indexed by $\mathbb{R}$. Indeed, any accepting path necessarily visits the state q_4 , that can only correspond to a gap, since it does not admit any incoming or outgoing successor transition. The language accepted by this automaton is described by the rational expression $\text{sh}(a, \varepsilon)$.

3. State of the Art

Given an automaton on linear orderings $\mathcal{A}$ and two of its states p and q , the reachability problem consists in deciding whether $\mathcal{A}$ admits a path starting at p and ending in q . In [12], Carton describes a polynomial procedure for deciding reachability on scattered linear orderings only, i.e., whether there exists a path labeled by a word indexed by such an ordering. Given the input automaton $\mathcal{A}$, Carton's algorithm generates a finite-word automaton $\mathcal{A}_F$ sharing the same set of states as $\mathcal{A}$, that accounts for every possible path in $\mathcal{A}$ while storing in the labels read along the paths the set of states that are visited. More precisely, the existence of a transition (p, P, q) in $\mathcal{A}_F$ implies that there exists in $\mathcal{A}$ a path $p \rightarrow q$ labeled by a word indexed by a scattered linear ordering that visits exactly the set of states P . Reciprocally, the existence in $\mathcal{A}$ of a path $p \rightarrow q$ labeled by a word indexed by a scattered linear ordering implies the existence of a matching path $p \rightarrow q$ in $\mathcal{A}_F$, although it may not necessarily be composed of a single transition. By not explicitly building a transition for every possible path, Carton's solution remains polynomial in the size of the automaton. Reachability in $\mathcal{A}$ then reduces to reachability in $\mathcal{A}_F$. This algorithm can be used to solve also the non-emptiness problem over scattered linear orderings, by checking whether an accepting state is reachable from an initial state.

A generalization of this algorithm over arbitrary linear orderings is sketched in [13], where the author introduces a specific additional rule for dealing with shuffles. To the best of our knowledge, a proof of this generalization has not

been published, and it seems that the additional rule is actually only able to deal with complete shuffles. Actually, a small modification of this rule (somewhat similar to what will be done in Section 5) is enough to deal with both complete and incomplete shuffles altogether, so as to accurately cover all linear orderings.

4. Reachability over the Reals

4.1. Characterization

Many properties characterizing $\mathbb{R}$, such as completeness, density, and not having a first or last element, can easily be expressed by automata on linear orderings. One could therefore hope that a procedure to check the non-emptiness (or equivalently decide the reachability problem) of an automaton over $\mathbb{R}$ may be designed by intersecting [13] this automaton with automata describing sufficiently many of the properties of $\mathbb{R}$, and then check the non-emptiness of the resulting automata.

Unfortunately, as shown by the following result, if accepting words indexed by complete and dense orderings without first or last element is a necessary condition for accepting words indexed by $\mathbb{R}$, this criterion is not sufficient.

Theorem 1. *There exists an automaton on linear orderings $\mathcal{A}_{>\mathbb{R}}$ that accepts words indexed by non-empty, dense, and complete linear orderings that do not have a first or last element, and that does not accept any word indexed by $\mathbb{R}$.*

The automaton $\mathcal{A}_{>\mathbb{R}}$ depicted in Figure 4 matches the description given in Theorem 1. The proof of Theorem 1 ensues from the two following lemmas.

Lemma 1. *The automaton $\mathcal{A}_{>\mathbb{R}}$ does not accept any word indexed by $\mathbb{R}$.*

PROOF. Let w be a word accepted by $\mathcal{A}_{>\mathbb{R}}$. We first prove that the ordering $|w|$ must be of the form $\Sigma_{k \in K} J_k$, where for all $k \in K$, J_k is an infinite, dense, and complete ordering with first and last elements, and K is an infinite, dense, and complete ordering without a first or last element. Intuitively, each J_k corresponds to a subpath that starts in q_2 , continues in q_3 , then visits a dense combination of q_4 and q_5 before reaching q_6 , and finally ends in q_7 .

The absence of first and last elements in K (and hence in $|w|$) results from the fact that the unique initial state q_1 does not admit an outgoing successor transition, and the unique accepting state q_8 does not admit an incoming successor

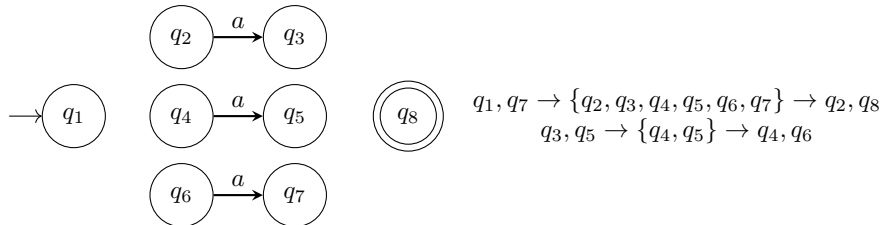


Figure 4: Automaton $\mathcal{A}_{>\mathbb{R}}$ that does not accept any word indexed by $\mathbb{R}$.

transition. On the other hand, each J_k admits a first element that corresponds to the successor transition (q_2, a, q_3) , and a last element that corresponds to the successor transition (q_6, a, q_7) .

Furthermore, since there does not exist a state q in $\mathcal{A}_{>\mathbb{R}}$ that admits both limit transitions $P \rightarrow q$ and $q \rightarrow P'$, it is impossible to map a gap of an incomplete ordering to a state of $\mathcal{A}_{>\mathbb{R}}$, hence $|w|$, and therefore K , as well as every J_k , are necessarily complete.

Recall that a linear ordering is dense if and only if it does not admit two consecutive elements. The density of K and each J_k naturally ensues from the fact that there does not exist a state q in $\mathcal{A}_{>\mathbb{R}}$ that admits both an incoming successor transition $p \rightarrow q$ and an outgoing successor transition $q \rightarrow p'$.

We now show that K and every J_k are infinite. By definition, if a path π in some automaton on linear orderings follows a limit transition, then an infinite set of states is necessarily visited in π . It ensues that π follows an infinite set of successor transitions, and that the word labeling π must be indexed by an infinite linear ordering. In our case, every subpath from q_2 to q_7 necessarily follows the limit transitions $q_3, q_5 \rightarrow \{q_4, q_5\} \rightarrow q_4, q_6$. Therefore, every J_k is infinite. Moreover, the only possibility of leaving (resp. reaching) the state q_1 (resp. q_8) is to follow the limit transition $q_1 \rightarrow \{q_2, q_3, q_4, q_5, q_6, q_7\}$ (resp. $\{q_2, q_3, q_4, q_5, q_6, q_7\} \rightarrow q_8$). It follows that K is also infinite.

The ordering $\mathbb{R}$ cannot however be expressed as an ordering $\Sigma_{k \in K} J_k$ such that K is an infinite, dense, and complete linear ordering without a first or a last element, and such that each J_k is an infinite, dense, and complete linear ordering with first and last elements. By contradiction, assume that $\mathbb{R}$ can be partitioned into an infinite, dense, and complete set I of pairwise disjoint closed intervals $[a, b]$ with $a < b$. The density of I implies that between any two intervals $[a, b], [e, f] \in I$ with $b < e$, there exists $[c, d] \in I$ such that $b < c < d < e$.

We first show that I must be uncountable. Otherwise, it would mean that I is both countable and complete. This contradicts Cantor's isomorphism theorem [18] which states that any infinite, dense, and countable ordering without first and last elements is order-isomorphic to $\mathbb{Q}$, and therefore incomplete.

Then, we show that I must be countable. Any set of disjoint intervals of $\mathbb{R}$ with non-zero length (such as I) is necessarily countable [19], since each of them contains a rational value that does not belong to the others.

Since I cannot be both countable and uncountable, we have a contradiction. $\square$

Lemma 2. *The automaton $\mathcal{A}_{>\mathbb{R}}$ accepts a word indexed by a non-empty, dense, and complete linear ordering that does not have a first or a last element.*

PROOF. Consider the linear ordering $S = [0, 1] \cdot \mathbb{R}$, where $[0, 1]$ denotes the closed interval of $\mathbb{R}$ between 0 and 1. Intuitively, S corresponds to the real line in which every real number has been replaced by a copy of the closed interval $[0, 1]$. The ordering S is dense, complete, and does not admit a first or a last element. We can represent elements of S as pairs $(x, y) \in [0, 1] \times \mathbb{R}$, ordered by the relation $<_S$ defined by $(x_1, y_1) <_S (x_2, y_2)$ if either $y_1 < y_2$, or $y_1 = y_2$

and $x_1 < x_2$, where $<$ denotes the usual order relation on both $\mathbb{R}$ and $[0, 1]$. Now, consider the word w indexed by S (i.e., $|w| = S$) obtained by labeling each element of S with the symbol a . We show that there exists an accepting run of $\mathcal{A}_{>\mathbb{R}}$ reading w . We denote by $\hat{s}_{min}$ (resp. $\hat{s}_{max}$) the first (resp. last) cut of S . Consider the mapping $\pi : \hat{S} \rightarrow Q$ defined as follows:

- $\pi(\hat{s}_{min}) = q_1$ and $\pi(\hat{s}_{max}) = q_8$,
- for all $y \in \mathbb{R}$, $\pi((0, y)^-) = q_2$ and $\pi((0, y)^+) = q_3$,
- for all $y \in \mathbb{R}$, $\pi((1, y)^-) = q_6$ and $\pi((1, y)^+) = q_7$,
- for all $x \in (0, 1)$ and for all $y \in \mathbb{R}$, $\pi((x, y)^-) = q_4$,
- for all $x \in (0, 1)$ and for all $y \in \mathbb{R}$, $\pi((x, y)^+) = q_5$.

It is immediate to establish that π describes an accepting run. $\square$

This concludes the proof of Theorem 1.

4.2. Algorithm

We now introduce an algorithm that decides reachability over $\mathbb{R}$. Let $\mathcal{A} = (Q, \Sigma, I, F, \Delta^s, \Delta^\ell)$ be an automaton on linear orderings, $q_I, q_F \in Q$ be two of its states, and M be the maximum size of limit transitions in $\mathcal{A}$, i.e., $M = \max\{|P| \mid \exists q \in Q. q \rightarrow P \in \Delta^\ell \vee P \rightarrow q \in \Delta^\ell\}$.

We consider the problem of deciding reachability between q_I and q_F over $\mathbb{R}$, i.e., deciding whether there exists a path in $\mathcal{A}$ that starts in q_I , ends in q_F , and is labeled by a word indexed by $\mathbb{R}$. Following an analogous approach as in [12], this algorithm first generates a sequence of finite-word automata $\mathcal{A}_0, \dots, \mathcal{A}_M$ sharing the same set of states as $\mathcal{A}$. Each automaton $\mathcal{A}_j$ is such that the existence of a path in $\mathcal{A}_j$ (obeying specific restrictions that we will characterize) from a state q to a state q' is equivalent to the existence of a j -limited path from q to q' in $\mathcal{A}$ that is labeled by a word indexed by $\mathbb{R}$. Recall that each limit transition $P \rightarrow q$ and $q \rightarrow P$ in a j -limited path is such that $|P| \leq j$. Therefore all paths are systematically M -limited. Once the last automaton of the sequence has been constructed, it only remains to test whether there exists a path (of the appropriate form) from q_I to q_F in $\mathcal{A}_M$.

For every $j \in [0, M]$, the automaton $\mathcal{A}_j$ (conveniently represented here as an automaton on linear orderings without any limit transition) is of the form $(Q, 2^Q, I, F, \Delta_j, \emptyset)$, i.e., the successor transitions in Δ_j (which we define later) are labeled by subsets of Q . We introduce the following useful definitions. An *open $\mathbb{R}$ -path* $\bar{r}$ from q_1 to q'_n in $\mathcal{A}_j$ labeled by $P = P_1 \cup \dots \cup P_n$ is a finite sequence of transitions $(q_1, P_1, q'_1), \dots, (q_n, P_n, q'_n) \in \Delta_j$ such that, for every $i \in [1, n-1]$, $(q'_i, \alpha_i, q_{i+1}) \in \Delta^s$ for some $\alpha_i \in \Sigma$. Intuitively, such a path corresponds to a finite concatenation of open intervals of $\mathbb{R}$ (corresponding to transitions in $\mathcal{A}_j$), connected by single elements (corresponding to successor transitions in $\mathcal{A}$), resulting in an open interval of $\mathbb{R}$. Analogously, a *closed $\mathbb{R}$ -path* $\bar{r}$ from p to p' in $\mathcal{A}_j$ labeled by $P \cup \{p, p'\}$ is a triple consisting of an open

$\mathbb{R}$ -path from a state q_1 to a state q'_n (labeled by P) together with two transitions $(p, \alpha, q_1), (q'_n, \alpha', p') \in \Delta^s$, for some $\alpha, \alpha' \in \Sigma$.

The transition relation Δ_j is defined by induction over j , with $\Delta_0 = \emptyset$, and, for $0 < j \leq M$, $\Delta_j = \Delta_{j-1} \cup S_j \cup L_j \cup R_j$, where S_j , L_j , and R_j are sets of transitions constructed from $\mathcal{A}_{j-1}$ by the following procedure. The sets S_j , L_j , and R_j are initially empty and filled according to the shuffle, infinite repetition, and reverse infinite repetition rules developed in the next sections. The final step of the algorithm consists in searching whether there exists an open $\mathbb{R}$ -path in $\mathcal{A}_M$ that starts in q_I and ends in q_F . In the positive case, $\mathcal{A}$ admits a path labeled by a word indexed by $\mathbb{R}$ from q_I to q_F , otherwise it does not.

Shuffle rule. For every pair of transitions $p_I \rightarrow P \in \Delta^\ell$ and $P \rightarrow p_F \in \Delta^\ell$ sharing a common set P such that $|P| = j$, the transition $(p_I, P \cup \{p_I, p_F\}, p_F)$ is added to S_j if and only if there exist

- $m > 0$ successor transitions $(s_1, \gamma_1, s'_1), \dots, (s_m, \gamma_m, s'_m)$ in Δ^s , such that, for each $i \in [1, m]$, $s_i, s'_i \in P$, $P \rightarrow s_i \in \Delta^\ell$, and $s'_i \rightarrow P \in \Delta^\ell$, and
- $n \geq 0$ closed $\mathbb{R}$ -paths $\bar{r}_1, \dots, \bar{r}_n$ in $\mathcal{A}_{j-1}$, where each $\bar{r}_i$ is labeled by a set of states Q_i , starts in a state q_i such that $P \rightarrow q_i \in \Delta^\ell$, and ends in a state q'_i such that $q'_i \rightarrow P \in \Delta^\ell$

satisfying the following condition:

$$\bigcup_{i \in [1, m]} \{s_i, s'_i\} \cup \bigcup_{i \in [1, n]} Q_i = P.$$

Intuitively, there are two kinds of shuffles. A *simple shuffle* involves no closed $\mathbb{R}$ -path (corresponding to the case $n = 0$ in the shuffle rule) and is a dense combination of single elements only. The simple shuffle rule is depicted in Figure 5. The solid transitions correspond to successor transitions in $\mathcal{A}$, and the dotted transition shows the new transition introduced in $\mathcal{A}_j$. A *Cantor shuffle* is a dense combination of intervals of order type $1 + \lambda + 1$ (corresponding to closed $\mathbb{R}$ -paths in $\mathcal{A}_{j-1}$) and single elements (corresponding to successor transitions

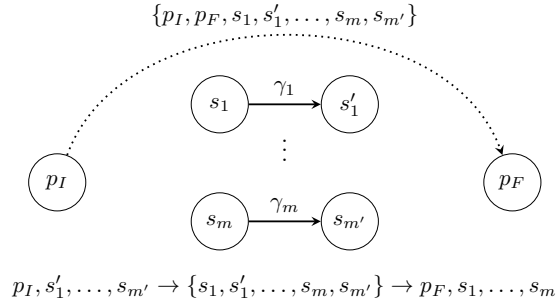


Figure 5: Simple shuffle rule.

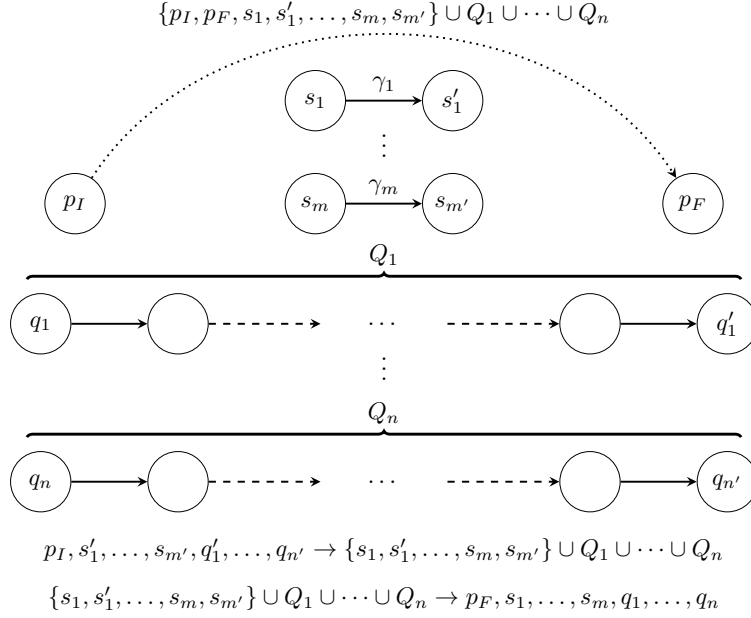


Figure 6: Cantor shuffle rule.

in $\mathcal{A}$). The Cantor shuffle rule is depicted in Figure 6. The dashed transitions correspond to preexisting transition in the automaton $\mathcal{A}_{j-1}$.

As a side note, consider again the automaton $\mathcal{A}_{>\mathbb{R}}$ mentioned in Lemma 1. The requirement $m > 0$ prevents the shuffle rule to be applicable with $p_I = q_1$ and $p_F = q_8$, in accordance with the fact that the automaton does not admit a path from q_1 to q_8 labeled by a word indexed by $\mathbb{R}$.

Infinite repetition rule. For every transition $P \rightarrow q \in \Delta^\ell$ such that $|P| = j$, and for every state $s \in P$, the transition $(s, P \cup \{q\}, q)$ is added to L_j if and only if there exists $s' \in P$ satisfying the two following conditions:

- $\mathcal{A}_{j-1}$ admits an open $\mathbb{R}$ -path from s to s' labeled by P of the form $(s_1, P_1, s'_1), \dots, (s_n, P_n, s'_n)$ with $s = s_1$ and $s' = s'_n$.
- $\mathcal{A}$ admits a successor transition of the form $(s', \alpha, s) \in \Delta^s$.

Intuitively, this rule looks for a cycle that alternates between open intervals of order type λ (corresponding to open $\mathbb{R}$ -paths in $\mathcal{A}_{j-1}$) and single elements (corresponding to successor transitions in $\mathcal{A}$), such that the set of states visited by the whole cycle is exactly P . If such a cycle exists, then it can be repeated infinitely many times, therefore allowing to follow the limit transition $P \rightarrow q$. Notice that the starting point $s = s_1$ of such a cycle can be chosen arbitrarily, hence the rule can also be applied for every state s_i visited by the cycle, for $i \in [1, n]$, as illustrated in Figure 7.

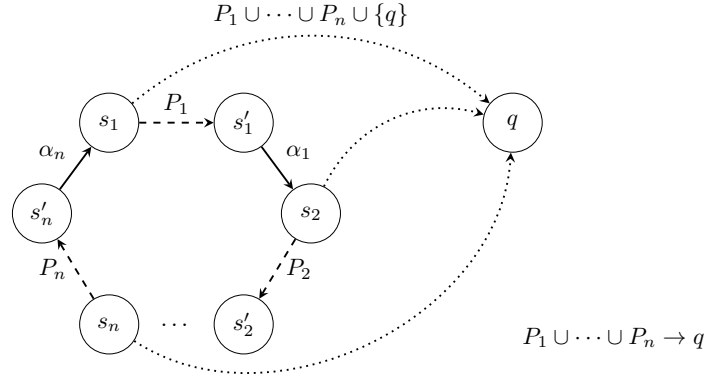


Figure 7: Infinite repetition rule.

Reverse infinite repetition rule. This rule follows a similar principle as the infinite repetition rule, but for a reversed ordering. For every transition $q \rightarrow P \in \Delta^\ell$ such that $|P| = j$, and for every state $s' \in P$, the transition $(q, P \cup \{q\}, s')$ is added to R_j if and only if there exists $s \in P$ satisfying the two following conditions:

- $\mathcal{A}_{j-1}$ admits an open $\mathbb{R}$ -path from s to s' labeled by P of the form $(s_1, P_1, s'_1), \dots, (s_n, P_n, s'_n)$ with $s = s_1$ and $s' = s'_n$.
- $\mathcal{A}$ admits a successor transition of the form $(s', \alpha, s) \in \Delta^s$.

4.3. Correctness

We state in the two following theorems that the algorithm described in Section 4.2 is sound and complete.

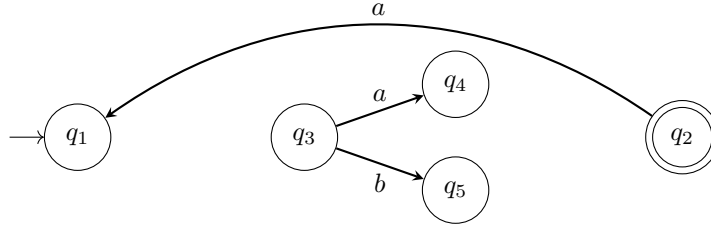
Theorem 2 (Soundness). *If there exists an open $\mathbb{R}$ -path in $\mathcal{A}_M$ from q_I to q_F labeled by P , then there exists a path in $\mathcal{A}$ from q_I to q_F , labeled by a word indexed by $\mathbb{R}$, that visits exactly the states in P .*

Theorem 3 (Completeness). *If the automaton $\mathcal{A}$ admits a path from q_I to q_F , labeled by a word indexed by $\mathbb{R}$, that visits a set of states P , then there exists an open $\mathbb{R}$ -path in $\mathcal{A}_M$ from q_I to q_F labeled by P .*

The proofs of these theorems are given in Section 6.

4.4. Example

We illustrate our algorithm, and more specifically the construction of the automata $\mathcal{A}_0, \dots, \mathcal{A}_M$ with a simple example. Consider the automaton $\mathcal{A}$ depicted in Figure 8. The automaton $\mathcal{A}_0$ comprises the states of automaton $\mathcal{A}$ without any transition. Since $\mathcal{A}$ does not admit a limit transition of size smaller than 2, automata $\mathcal{A}_2$ and $\mathcal{A}_1$ are identical to $\mathcal{A}_0$. The automaton $\mathcal{A}_3$ shown in Figure 9 has two transitions $(q_4, \{q_3, q_4, q_5\}, q_3)$ and $(q_5, \{q_3, q_4, q_5\}, q_3)$ resulting from the application of the simple shuffle rule on the set $\{q_3, q_4, q_5\}$ with



$$q_4, q_5 \rightarrow \{q_3, q_4, q_5\} \rightarrow q_3$$

$$q_1, q_4, q_5 \rightarrow \{q_1, q_2, q_3, q_4, q_5\} \rightarrow q_2, q_3$$

Figure 8: The input automaton $\mathcal{A}$.

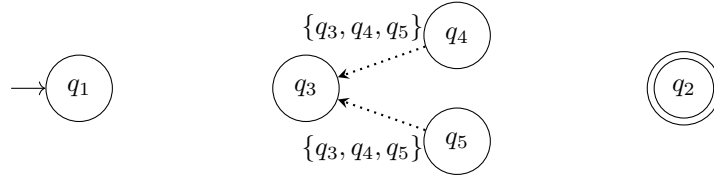


Figure 9: The finite-word automata $\mathcal{A}_3$ and $\mathcal{A}_4$.

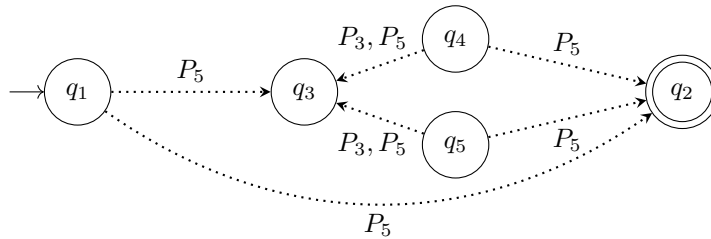
$p_F = q_3$, and respectively $p_I = q_4$ and $p_I = q_5$. Since $\mathcal{A}$ does not admit a limit transition of size 4, $\mathcal{A}_4$ is identical to $\mathcal{A}_3$. The final automaton $\mathcal{A}_5$ is illustrated in Figure 10. The new transitions result from the application of the shuffle rule on the set $\{q_1, q_2, q_3, q_4, q_5\}$, for every $p_I \in \{q_1, q_4, q_5\}$ and $p_F \in \{q_2, q_3\}$.

4.5. Complexity

The algorithm described in Section 4.2 runs in polynomial time w.r.t. the size of $\mathcal{A}$, i.e., $|Q| + |\Delta^s| + |\Delta^\ell|$.

Theorem 4. *The automaton $\mathcal{A}_M$ can be computed in polynomial time in the size of $\mathcal{A}$, by systematic application of the shuffle, infinite repetition and reverse infinite repetition rules.*

PROOF. First, for each $j \in [0, M]$, each rule is called only polynomially many times during the construction of the automaton $\mathcal{A}_j$. Indeed, the shuffle rule



with $P_5 = \{q_1, q_2, q_3, q_4, q_5\}$, and $P_3 = \{q_3, q_4, q_5\}$.

Figure 10: The finite-word automaton $\mathcal{A}_5$.

is only considered for every combination of two limit transitions in $\mathcal{A}$, and the (reverse) infinite repetition rule is only considered for every pair consisting of a limit transition in $\mathcal{A}$ together with a state of $\mathcal{A}$.

It remains to show that each rule can be applied in polynomial time. The case of the simple shuffle rule is immediate. The other rules require to be able to check either the existence of an open $\mathbb{R}$ -path labeled by a given set of states P (to deal with the infinite and reverse infinite repetition rules), or the existence of a set of closed $\mathbb{R}$ -paths and successor transitions such that their combined labels and visited states are equal to P (to deal with the Cantor shuffle rule). These two operations can be carried out by a polynomial procedure that, given a set of states P , two states $q, q' \in P$, and the automaton $\mathcal{A}_j$, computes the set $S_{\text{open}}(\mathcal{A}_j, P, q, q') = \{r \in R \mid R \text{ labels an } \mathbb{R}\text{-path in } \bar{R} \wedge R \subseteq P\}$ where $\bar{R}$ is the set of open $\mathbb{R}$ -paths from q to q' in $\mathcal{A}_j$, and similarly for $S_{\text{closed}}(\mathcal{A}_j, P, q, q')$. This procedure works as follows. Let $\delta_{P,j,q}^{\rightarrow} \subseteq \Delta_j$ (resp. $\delta_{P,j,q'}^{\leftarrow}$) be the set of transitions occurring in (open or closed) $\mathbb{R}$ -paths of $\mathcal{A}_j$ starting in q (resp. ending in q'), and labeled by a subset of P . The sets $\delta_{P,j,q}^{\rightarrow}$ and $\delta_{P,j,q'}^{\leftarrow}$ can be computed in polynomial time by performing a (respectively forward and backward) traversal of the automaton $\mathcal{A}_j$ that respects the structure of $\mathbb{R}$ -paths, i.e., alternating between transitions in Δ_j and transitions in Δ^s . The set $S_{\text{open}}(\mathcal{A}_j, P, q, q')$ (resp. $S_{\text{closed}}(\mathcal{A}_j, P, q, q')$) is then obtained by taking the union of the labels of the transitions in $\delta_{P,j,q}^{\rightarrow} \cap \delta_{P,j,q'}^{\leftarrow}$.

For the (reverse) infinite repetition rule, consider a set of states P , and a transition $(p', \alpha, p) \in \Delta^s$ with $p, p' \in P$. If $S_{\text{open}}(\mathcal{A}_j, P, p, p') = P$, there necessarily exists an open $\mathbb{R}$ -path from p to p' in $\mathcal{A}_j$ labeled by P . Such a path can be constructed by first choosing, for each $q \in P$, an open $\mathbb{R}$ -path $\bar{r}_q$ labeled by $R \subseteq P$ such that $q \in R$, and then concatenating these paths with the help of the transition $(p', \alpha, p) \in \Delta^s$.

Finally, for the shuffle rule, consider a set of states P involved in a limit transition. We define $L_P = \{p \in P \mid P \rightarrow p \in \Delta^\ell\}$, and $R_P = \{p \in P \mid p \rightarrow P \in \Delta^\ell\}$. Let the set of states $S \subseteq P$ correspond to the states $s_1, s'_1, \dots, s_m, s'_m$ involved in the successor transitions satisfying the shuffle rule constraints. Then, using a similar argument as before, one shows that checking the existence of a set of closed $\mathbb{R}$ -paths $\bar{r}_1, \dots, \bar{r}_n$, respectively labeled by $Q_1, \dots, Q_n$, starting from a state in L_P , and ending in a state in R_P , such that $P = S \cup \bigcup_{i \in [1, n]} Q_i$, reduces to checking that $P = S \cup \bigcup_{(q, q') \in L_P \times R_P} S_{\text{closed}}(\mathcal{A}_j, P, q, q')$. $\square$

4.6. Pseudo Code

We now describe in pseudo-code the algorithm computing the set of transitions Δ_j of the automaton $\mathcal{A}_j$. Recall that the transitions in Δ_j must account for every j -limited path labeled by a word indexed by $\mathbb{R}$ in the input automaton $\mathcal{A} = (Q, \Sigma, I, F, \Delta^s, \Delta^\ell)$. The set of transitions Δ_j is computed recursively, with $\Delta_0 = \emptyset$, and Δ_{j-1} being used in the computation of Δ_j for $j > 0$. The procedure **Compute- Δ_j** described in Algorithm 1 computes the set of transitions Δ_j given $\mathcal{A}$ and Δ_{j-1} . This procedure relies on the subroutine **Labels- $\mathbb{R}$ -path** described in Algorithm 2, that works as follows.

Algorithm 1: Compute- Δ_j

Input: $\Delta_{j-1}, \Delta^s, \Delta^\ell$
Output: The set of transitions Δ_j corresponding to the automaton $\mathcal{A}_j$.

```

1  $L_j, R_j, S_j \leftarrow \emptyset$ 
2 foreach  $P \subseteq Q$  such that  $|P| = j \wedge \exists q \in Q. (q \rightarrow P) \in \Delta^\ell \vee (P \rightarrow q) \in \Delta^\ell$  do
3   foreach  $(p', \alpha, p) \in \Delta^s$  such that  $p, p' \in P$  do
4     if  $\text{Labels-}\mathbb{R}\text{-path}(\Delta_{j-1}, \Delta^s, P, p, p', \text{open}) = P$  then
5       foreach  $q \in Q$  such that  $(P \rightarrow q) \in \Delta^\ell$  do
6          $L_j \leftarrow L_j \cup \{(p, P \cup \{q\}, q)\}$  /* Infinite rep. */
7       foreach  $q \in Q$  such that  $(q \rightarrow P) \in \Delta^\ell$  do
8          $R_j \leftarrow R_j \cup \{(q, P \cup \{q\}, p')\}$  /* Rev. infinite rep. */
9    $\text{Left}_P \leftarrow \{q \in Q \mid (P \rightarrow q) \in \Delta^\ell\}$  /* Shuffles */
10   $\text{Right}_P \leftarrow \{q \in Q \mid (q \rightarrow P) \in \Delta^\ell\}$ 
11   $\text{atLeastOneIndividualPoint} \leftarrow \text{false}$ 
12   $\text{StatesVisited} \leftarrow \emptyset$ 
13  foreach  $(\ell, r) \in (\text{Left}_P \cap P) \times (\text{Right}_P \cap P)$  do
14    foreach  $(\ell, \alpha, r) \in \Delta^s$  do
15       $\text{atLeastOneIndividualPoint} \leftarrow \text{true}$ 
16       $\text{StatesVisited} \leftarrow \text{StatesVisited} \cup \{\ell, r\}$ 
17      break
18   $\text{StatesVisited} \leftarrow \text{StatesVisited} \cup \text{Labels-}\mathbb{R}\text{-path}(\Delta_{j-1}, \Delta^s, P, \ell, r, \text{closed})$ 
19  if  $\text{StatesVisited} = P \wedge \text{atLeastOneIndividualPoint}$  then
20    foreach  $(p_I, p_F) \in \text{Right}_P \times \text{Left}_P$  do
21       $S_j \leftarrow S_j \cup \{(p_I, P \cup \{p_I, p_F\}, p_F)\}$ 
22  $\Delta_j \leftarrow \Delta_{j-1} \cup L_j \cup R_j \cup S_j$ 

```

For a given automaton $\mathcal{A}_j$, a set of states $P \subseteq Q$, and two states $q, q' \in P$, the subroutine **Labels- $\mathbb{R}$ -path** computes either the set $S_{\text{open}}(\mathcal{A}_j, P, q, q')$, or the set $S_{\text{closed}}(\mathcal{A}_j, P, q, q')$ mentioned in Section 4.5, depending on the value of the **type** parameter. It first performs a search (or traversal) in the automaton $\mathcal{A}_j$ starting from the state q , and collects every transition in Δ_j that allows to progress in the search by following an $\mathbb{R}$ -path labeled by subsets of P . Once this search is over, it performs a backward search in $\mathcal{A}_j$ starting from q' , that collects every transitions allowing to progress in this traversal by following an $\mathbb{R}$ -path labeled by a subset of P . In this backward search, we actually only consider the transitions in $\mathcal{A}_j$ that have been collected during the first forward search. Once this second search is done, we return the union of the labels of every transition collected during the backward search, which by construction was first collected during the forward search, hence belonging to an $\mathbb{R}$ -path from q to q' in $\mathcal{A}_j$. The procedure **Labels- $\mathbb{R}$ -path** runs in polynomial time.

5. Reachability over the Rationals

The reachability algorithm over the reals can be adapted to accommodate reachability over the rational numbers $\mathbb{Q}$. Because of the incompleteness of $\mathbb{Q}$, the rules must be adapted in order to account for the presence of gaps.

Algorithm 2: Labels- $\mathbb{R}$ -path

Input: $\Delta_{j-1}, \Delta^s, P \subseteq Q, q \in P, q' \in P, \text{type} \in \{\text{open}, \text{closed}\}$.
Output: The set of states (subset of P) occurring in labels of *open* or *closed* (depending on the **type** parameter) $\mathbb{R}$ -paths from q to q' , in $\mathcal{A}_{j-1}$.

```
1  $E_I \leftarrow \emptyset$  /* Transitions from  $\Delta_{j-1}$  reachable from  $q$  */
2  $E_P \leftarrow \emptyset$  /* Transitions from  $\Delta^s$  reachable from  $q$  */
3  $\text{VisitedI} \leftarrow \emptyset$  /* States already visited when looking for an interval */
4  $\text{VisitedP} \leftarrow \emptyset$  /* States already visited when looking for a point */
5 if  $\text{type} = \text{open}$  then
6    $(\text{ToCheckI}, \text{ToCheckP}) \leftarrow (\{q\}, \emptyset)$ 
7 else
8    $(\text{ToCheckI}, \text{ToCheckP}) \leftarrow (\emptyset, \{q\})$ 
9 while  $(\text{ToCheckI}, \text{ToCheckP}) \neq (\emptyset, \emptyset)$  do
10   foreach  $p \in \text{ToCheckI}$  do
11      $(\text{VisitedI}, \text{ToCheckI}) \leftarrow (\text{VisitedI} \cup \{p\}, \text{ToCheckI} \setminus \{p\})$ 
12     foreach  $(p, R, r) \in \Delta_{j-1}$  such that  $R \subseteq P$  do
13        $E_I \leftarrow E_I \cup \{(p, R, r)\}$ 
14       if  $r \notin \text{VisitedP}$  then  $\text{ToCheckP} \leftarrow \text{ToCheckP} \cup \{r\}$ 
15   foreach  $p \in \text{ToCheckP}$  do
16      $(\text{VisitedP}, \text{ToCheckP}) \leftarrow (\text{VisitedP} \cup \{p\}, \text{ToCheckP} \setminus \{p\})$ 
17     foreach  $(p, \alpha, r) \in \Delta^s$  such that  $\{p, r\} \subseteq P$  do
18        $E_P \leftarrow E_P \cup \{(p, \alpha, r)\}$ 
19       if  $r \notin \text{VisitedI}$  then  $\text{ToCheckI} \leftarrow \text{ToCheckI} \cup \{r\}$ 
20   /* We only keep the transitions in  $E_I \cup E_P$  that allow to reach  $q'$ , */
21   /* by performing the same search but backwards, starting from  $q'$ . */
22    $S \leftarrow \emptyset$ 
23    $(\text{VisitedIBack}, \text{VisitedPBack}) \leftarrow (\emptyset, \emptyset)$ 
24   if  $\text{type} = \text{open}$  then
25      $(\text{ToCheckIBack}, \text{ToCheckPBack}) \leftarrow (\{q'\}, \emptyset)$ 
26   else
27      $(\text{ToCheckIBack}, \text{ToCheckPBack}) \leftarrow (\emptyset, \{q'\})$ 
28   while  $(\text{ToCheckIBack}, \text{ToCheckPBack}) \neq (\emptyset, \emptyset)$  do
29     foreach  $p \in \text{ToCheckIBack}$  do
30        $(\text{VisitedIBack}, \text{ToCheckIBack}) \leftarrow (\text{VisitedIBack} \cup \{p\}, \text{ToCheckIBack} \setminus \{p\})$ 
31       foreach  $(r, R, p) \in E_I$  do /*  $p$  is the endpoint */
32          $S \leftarrow S \cup R$ 
33         if  $r \notin \text{VisitedPBack}$  then  $\text{ToCheckPBack} \leftarrow \text{ToCheckPBack} \cup \{r\}$ 
34     foreach  $p \in \text{ToCheckPBack}$  do
35        $(\text{VisitedPBack}, \text{ToCheckPBack}) \leftarrow (\text{VisitedPBack} \cup \{p\}, \text{ToCheckPBack} \setminus \{p\})$ 
36       foreach  $(r, \alpha, p) \in E_P$  do /*  $p$  is the endpoint */
37          $S \leftarrow S \cup \{r, p\}$ 
38         if  $r \notin \text{VisitedIBack}$  then  $\text{ToCheckIBack} \leftarrow \text{ToCheckIBack} \cup \{r\}$ 
39 return  $S$ 
```

We first define a notion of valid path over $\mathbb{Q}$, analogous to the notion of $\mathbb{R}$ -path. An open $\mathbb{Q}$ -path from q_1 to q'_n in $\mathcal{A}_j$ labeled by $P = P_1 \cup \dots \cup P_n$ is a finite sequence of transitions $(q_1, P_1, q'_1), \dots, (q_n, P_n, q'_n) \in \Delta_j$ such that for each $i \in [1, n-1]$, either $q'_i = q_{i+1}$, or there exists a successor transition $(q'_i, \alpha_i, q_{i+1}) \in \Delta^s$. Intuitively, such a path corresponds to a finite concatenation of open intervals of $\mathbb{Q}$ (corresponding to transitions in $\mathcal{A}_j$), in which two consecutive intervals are either connected by a gap (if the shared bound between the two intervals is an irrational number in $\mathbb{R} \setminus \mathbb{Q}$, e.g., $\sqrt{2}$), or by a single element (corresponding to a successor transition in $\mathcal{A}$, if the shared bound belongs to $\mathbb{Q}$). Like in the case of $\mathbb{R}$, we also define *closed* $\mathbb{Q}$ -paths, i.e., open $\mathbb{Q}$ -paths preceded and followed by a successor transition so as to form a closed interval, as well as *half-open* $\mathbb{Q}$ -paths, i.e., open $\mathbb{Q}$ -paths preceded or (exclusively) followed by one successor transition.

For a given set of states P involved in a limit transition, both the simple and Cantor shuffle rules additionally require a non-empty set of states $(g_1, \dots, g_k)$ such that $g_i \rightarrow P \rightarrow g_i$, for all $i \in [1, k]$, corresponding to gaps. Also, the left and right infinite repetition rules must be adapted to allow for the required $\mathbb{Q}$ -path to be cyclic, i.e., $q'_n = q_1$. We now describe these rules formally.

Shuffle rule. For every pair of transitions $p_I \rightarrow P \in \Delta^\ell$ and $P \rightarrow p_F \in \Delta^\ell$ sharing a common set P such that $|P| = j$, the transition $(p_I, P \cup \{p_I, p_F\}, p_F)$ is added to S_j if and only if there exist

- $m \geq 0$ successor transitions $(s_1, \gamma_1, s'_1), \dots, (s_m, \gamma_m, s'_m)$ in Δ^s , such that, for each $i \in [1, m]$, $s_i, s'_i \in P$, $P \rightarrow s_i \in \Delta^\ell$, and $s'_i \rightarrow P \in \Delta^\ell$,
- $n \geq 0$ open, closed, or half-open $\mathbb{Q}$ -paths $\bar{q}_1, \dots, \bar{q}_n$ in $\mathcal{A}_{j-1}$, where each $\bar{q}_i$ is labeled by a set of states Q_i , starts in a state q_i such that $P \rightarrow q_i \in \Delta^\ell$, and ends in a state q'_i such that $q'_i \rightarrow P \in \Delta^\ell$, and
- $k > 0$ states $g_1, \dots, g_k$ in P such that, for every $i \in [1, k]$, both transitions $g_i \rightarrow P$ and $P \rightarrow g_i$ belong to Δ^ℓ

satisfying the two following conditions:

- $\bigcup_{i \in [1, m]} \{s_i, s'_i\} \cup \bigcup_{i \in [1, k]} \{g_i\} \cup \bigcup_{i \in [1, n]} Q_i = P$.
- The inequality $n + m > 0$ holds (otherwise, there would only be gaps).

Infinite repetition rule. For every transition $P \rightarrow q \in \Delta^\ell$ such that $|P| = j$, and for every state $s \in P$, the transition $(s, P \cup \{q\}, q)$ is added to L_j if and only if there exists $s' \in P$ satisfying the two following conditions:

- $\mathcal{A}_{j-1}$ admits an open $\mathbb{Q}$ -path from s to s' labeled by P of the form $(q_1, P_1, q'_1), \dots, (q_n, P_n, q'_n)$ with $s = q_1$ and $s' = q'_n$.
- Either $s = s'$, or $\mathcal{A}$ admits a successor transition of the form $(s', \alpha, s) \in \Delta^s$.

Reverse infinite repetition rule. For every transition $q \rightarrow P \in \Delta^\ell$ such that $|P| = j$, and for every state $s' \in P$, the transition $(q, P \cup \{q\}, s')$ is added to R_j if and only if there exists $s \in P$ satisfying the two following conditions:

- $\mathcal{A}_{j-1}$ admits an open $\mathbb{Q}$ -path from s to s' labeled by P of the form $(q_1, P_1, q'_1), \dots, (q_n, P_n, q'_n)$ with $s = q_1$ and $s' = q'_n$.
- Either $s = s'$, or $\mathcal{A}$ admits a successor transition of the form $(s', \alpha, s) \in \Delta^s$.

5.1. Example

First, observe that running the algorithm for $\mathbb{Q}$ over the automaton $\mathcal{A}$ (depicted in Figure 8 and studied in Section 4.4) does not generate a single new transition. This is because the automaton $\mathcal{A}$ lacks states that can be matched with gaps, which implies that no shuffle rule over $\mathbb{Q}$ can be applied.

Now, let us illustrate our algorithm for $\mathbb{Q}$ over the automaton $\mathcal{B}$ depicted in Figure 11. The automaton $\mathcal{B}_2$ shown in Figure 12a has four transitions. The two transitions $(q_2, \{q_1, q_2\}, q_1)$ and $(q_2, \{q_1, q_2\}, q_2)$ result from the application of the shuffle rule on the set $\{q_1, q_2\}$ with $p_I = q_2$, and respectively $p_F = q_1$ and $p_F = q_2$. Notice that in this application of the shuffle rule, the state q_2 behaves both as the endpoint of a successor transition, and as a state that can be matched to a gap. The two transitions $(q_1, \{q_1, q_3\}, q_1)$ and $(q_1, \{q_1, q_3\}, q_3)$ result from a similar application of the shuffle rule on the set $\{q_1, q_3\}$ with $p_I = q_1$, and respectively $p_F = q_1$ and $p_F = q_3$. This time, it is q_1 that acts both as the endpoint of a successor transition, and as a state that can be matched to a gap.

The automaton $\mathcal{B}_3$ shown in Figure 12b has two additional transitions. The transition $(q_1, \{q_1, q_2, q_3, q_4\}, q_4)$ results from the application of the infinite repetition rule on the set $\{q_1, q_2, q_3\}$, with $s = s' = q_1$, which is valid thanks to the existence of the open $\mathbb{Q}$ -path from q_1 to itself $q_1 \xrightarrow{\{q_1, q_3\}} q_3 \xrightarrow{a} q_2 \xrightarrow{\{q_1, q_2\}} q_1$. The transition $(q_2, \{q_1, q_2, q_3, q_4\}, q_4)$ results from the application of the infinite repetition rule on the set $\{q_1, q_2, q_3\}$, with $s = q_2$ and $s' = q_3$, using the open $\mathbb{Q}$ -path $q_2 \xrightarrow{\{q_1, q_2\}} q_1 \xrightarrow{\{q_1, q_3\}} q_3$, and the successor transition (q_3, a, q_2) .

Notice that the application of the infinite repetition rule does not generate the transition $(q_3, \{q_1, q_2, q_3, q_4\}, q_4)$. This is because it is not possible to find an open $\mathbb{Q}$ -path that starts in q_3 , i.e., there does not exist a transition in $\mathcal{B}_2$ that originates from q_3 .

6. Proof of Correctness

We now prove the correctness of the algorithm introduced in Section 4.2.

6.1. Soundness

In order to prove the soundness of the algorithm (Theorem 2), we first show (Lemma 3) that an open $\mathbb{R}$ -path $\bar{r}$ in $\mathcal{A}_M$ corresponds to a path in $\mathcal{A}$ indexed by $\mathbb{R}$, provided that each transition of $\mathcal{A}_M$ involved in $\bar{r}$ corresponds to a path in $\mathcal{A}$ indexed by $\mathbb{R}$. Then, it suffices to prove (Lemma 5) that each transition in $\mathcal{A}_M$ corresponds to a path in $\mathcal{A}$ indexed by $\mathbb{R}$.

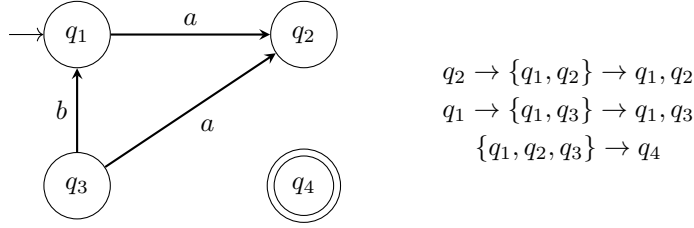


Figure 11: The input automaton $\mathcal{B}$.

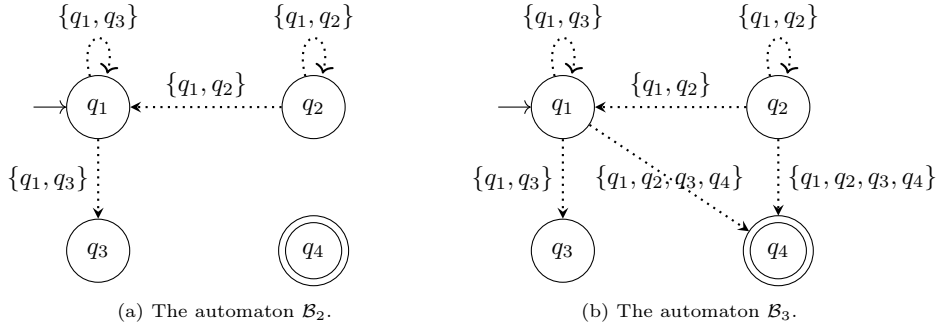


Figure 12: The automata obtained at the main steps of the algorithm for $\mathbb{Q}$.

Lemma 3. *Let $(q_1, P_1, q'_1), \dots, (q_n, P_n, q'_n)$ be an open $\mathbb{R}$ -path in $\mathcal{A}_M$, such that for every $i \in [1, n]$, there exists a path $q_i \rightarrow q'_i$ in $\mathcal{A}$ labeled by a word indexed by $\mathbb{R}$, visiting the set of states P_i . Then, there exists a path $q_1 \rightarrow q'_n$ in $\mathcal{A}$ labeled by a word indexed by $\mathbb{R}$, that visits exactly the set of states $P_1 \cup \dots \cup P_n$.*

PROOF. The proof relies on the fact that a path labeled by a word indexed by $\mathbb{R}$ can also be labeled by a word indexed by any non-empty open interval $(x, y) \subseteq \mathbb{R}$, since both share the order type λ .

We consider a sequence of words $w_1, w_2, \dots, w_n$ such that:

- w_1 is indexed by $(-\infty, 1)$ and w_n is indexed by $(n-1, +\infty)$;
- for all $i \in [2, n-1]$, w_i is indexed by $(i-1, i)$;
- for all $i \in [1, n]$, w_i labels a path π_i in $\mathcal{A}$ from q_i to q'_i .

Recall that for $(q_1, P_1, q'_1), \dots, (q_n, P_n, q'_n)$ to be an open $\mathbb{R}$ -path, there must exist a successor transition of the form $(q'_i, \alpha_i, q_{i+1}) \in \Delta^s$, for every $i \in [1, n-1]$. In order to construct a word indexed by $\mathbb{R}$ that labels a path from q_1 to q'_n , we simply concatenate together the words w_i , using these successor transitions to fill the gaps between consecutive open intervals. The paths π_i can be combined into a path π from q_1 to q'_n , labeled by the word $w = w_1 \cdot \alpha_1 \cdot w_2 \cdot \dots \cdot w_{n-1} \cdot \alpha_{n-1} \cdot w_n$. By construction, the path π is indexed by $\mathbb{R}$. $\square$

Before advancing, we first introduce the following result which will be needed in the remainder of the proof of Theorem 2.

Lemma 4. *Let $m > 0$. There exists a partition of $\mathbb{R}$ into m non-empty pairwise-disjoint subsets $R_1, \dots, R_m$ such that for every $i \in [1, m]$, R_i is dense in $\mathbb{R}$.*

PROOF. Let us construct such a partition of $\mathbb{R}$. We define, for every $k \in \mathbb{N}_{>1}$, the set S_k denoting the set of (non-integer) rational numbers that can be represented by a fraction with a strictly positive power of k as its denominator. Formally, $S_k = \{p/k^j \mid p \in \mathbb{Z} \text{ and } j \in \mathbb{N}_{>0}\} \setminus \mathbb{Z}$.

Let us show that, for any $k > 1$, the set S_k is dense in $\mathbb{R}$. We simply need to show that within any non-empty open interval $(x, y) \subseteq \mathbb{R}$, there exists $z \in S_k$ such that $x < z < y$. Since $k > 1$, we know that there exists $j \in \mathbb{N}$ such that $1/k^j < (y - x)/2$. Let us pick such a value j . There also exists an integer $p \in \mathbb{Z}$ such that $p/k^j \leq (x + y)/2 \leq (p + 1)/k^j$, one could pick $p = \lfloor (x + y) \cdot (k^j)/2 \rfloor$ for instance. Notice that $(x + y)/2$ is the middle of the interval (x, y) . Since the difference between p/k^j and $(p + 1)/k^j$ is equal to $1/k^j$ and since $1/k^j < (y - x)/2$, we have that either $p/k^j \in (x, y)$ or $(p + 1)/k^j \in (x, y)$ (it could be both).

Let us now show that for two distinct prime numbers p_1 and p_2 , the sets S_{p_1} and the set S_{p_2} are disjoint. Assume that there exist $n_1/(p_1)^i \in S_{p_1}$ and $n_2/(p_2)^j \in S_{p_2}$ for some $i, j > 0$, such that $n_1/(p_1)^i = n_2/(p_2)^j$. This yields the equality $n_1 p_2^j = n_2 p_1^i$. Since p_1 and p_2 are prime numbers, this implies $n_1 = k(p_1^i)$ and $n_2 = k(p_2^j)$ for some $k \in \mathbb{Z}$. This is impossible because $k(p_1^i)/p_1^i = k \in \mathbb{Z}$, and we require $S_{p_1} \cap \mathbb{Z} = \emptyset$ (the same development applies to S_{p_2}).

Let $P = (p_i)_{i \in [1, m]}$ be the sequence of the m first prime numbers (i.e., $2, 3, 5, 7, \dots$). We then simply set $R_1 = \mathbb{R} \setminus \mathbb{Q}$, $R_i = S_{p_i}$ for $i \in [2, m - 1]$, and $R_m = \mathbb{Q} \setminus (\bigcup_{i \in [2, m - 1]} S_{p_i})$. By construction, it is clear that the sets $(R_i)_{i \in [1, m]}$ form a partition of $\mathbb{R}$. The density of each R_i for $i \in [2, m - 1]$ has already been proven. The set R_1 is also dense in $\mathbb{R}$ since it corresponds to the set of irrational numbers. Finally, the density of R_m is obtained by noticing that $S_{p_m} \subsetneq R_m$, and remembering that we already proved that S_{p_m} is dense in $\mathbb{R}$. $\square$

Lemma 5. *For each transition (p, P, q) of $\mathcal{A}_j$ for $j \in [0, M]$, there exists a path from p to q in $\mathcal{A}$, labeled by a word indexed by $\mathbb{R}$, that visits exactly the states in P .*

PROOF. The proof is constructive, explicitly building a word indexed by $\mathbb{R}$ that yields a path in $\mathcal{A}$. The form of the word depends on the rule that generated the transition of $\mathcal{A}_j$.

The lemma is proved by induction over j . The base case is trivial, since $\mathcal{A}_0$ does not have transitions. Assuming that every transition of $\mathcal{A}_{j-1}$ verifies the lemma, we now prove that the lemma also holds for any transition (p, P, q) of $\mathcal{A}_j$ that does not belong to $\mathcal{A}_{j-1}$.

First, consider that (p, P, q) results from the application of the *simple shuffle* rule, i.e., the shuffle rule involving only successor transitions without any closed $\mathbb{R}$ -path. Formally, we have $p \rightarrow P \in \Delta^\ell$ and $P \rightarrow q \in \Delta^\ell$, and a non-empty set of successor transitions $\{(s_1, \gamma_1, s'_1), \dots, (s_m, \gamma_m, s'_m)\} \subseteq \Delta^s$ such that $s_i, s'_i \in P$, $P \rightarrow s_i \in \Delta^\ell$, and $s'_i \rightarrow P \in \Delta^\ell$, for all $i \in [1, m]$. Let $\{R_1, \dots, R_m\}$ be a partition of $\mathbb{R}$ into m disjoint infinite dense sets, i.e., for any two real numbers x and y such that $x < y$, there exist $z_1, \dots, z_m$ such that $x < z_i < y$ and $z_i \in R_i$, for all $i \in [1, m]$. By Lemma 4, such a partition of $\mathbb{R}$ can always be found. Now, consider the word $w = (\alpha_r)_{r \in \mathbb{R}}$ indexed by $\mathbb{R}$ such that $\alpha_r = \gamma_i$ if $r \in R_i$. To build a suitable path π in $\mathcal{A}$ from p to q reading w , it suffices to impose $\pi(-\infty) = p$ and $\pi(+\infty) = q$, where $-\infty$ and $+\infty$ denote respectively the first and last cut of $\mathbb{R}$, and furthermore require $\pi(r^-) = s_i$ and $\pi(r^+) = s'_i$ if $r \in R_i$, for each $r \in \mathbb{R}$. By construction, it is clear that π is a valid path in $\mathcal{A}$ that visits exactly the states in P .

For the three remaining rules, notice that for any $\mathbb{R}$ -path $(p_1, P_1, p'_1), \dots, (p_n, P_n, p'_n)$ in $\mathcal{A}_{j-1}$, the inductive hypothesis and Lemma 3 ensure that there exists a path $p_1 \rightarrow p'_n$ in $\mathcal{A}$, labeled by a word indexed by $\mathbb{R}$ (or any non-empty open interval of $\mathbb{R}$), that visits exactly the states in $P_1 \cup \dots \cup P_n$.

If (p, P, q) is generated by the *infinite repetition* rule, i.e., it belongs to L_j but not to Δ_{j-1} , the rule's conditions imply that there exists an open $\mathbb{R}$ -path $(p_1, P_1, p'_1), \dots, (p_n, P_n, p'_n)$ in $\mathcal{A}_{j-1}$, such that:

- $p_1 = p$,
- there exists a transition $(p'_n, \alpha, p) \in \Delta^s$,
- $P = P_1 \cup \dots \cup P_n$, and
- $(P \rightarrow q) \in \Delta^\ell$.

Consider an infinite sequence of words $(w_k)_{k \in \mathbb{N}}$, each of them labeling a valid path from p to p'_n , such that w_0 is indexed by $(-\infty, 0)$, and for every $k > 0$, w_k is indexed by $(k-1, k)$. Now consider the word $w = \prod_{k \in \mathbb{N}} (w_k \cdot \alpha) = w_0 \cdot \alpha \cdot w_1 \cdot \alpha \cdot \dots$ indexed by $\mathbb{R}$. By simply concatenating the paths labeled by the words $(w_k)_{k \in \mathbb{N}}$ (each starting in p and ending in p'_n) with the transition (p'_n, α, p) , and assigning the state q to the last cut $(+\infty)$ of $\mathbb{R}$, we obtain a valid path reading w that starts in p , ends in q , and visits exactly P .

The case of *reverse infinite repetitions* is handled symmetrically. The accepted word indexed by $\mathbb{R}$ can also be expressed as an infinite concatenation of words but in reverse, i.e., w_0 is indexed by $(0, +\infty)$, and each w_k is indexed by $-(k+1, -k)$ for $k \in \mathbb{N}_{>0}$.

Finally, the case of the Cantor shuffle can be dealt with by considering the set of intervals and the set of individual points involved in the Cantor ternary set construction [20, 21], that we will describe explicitly. For simplicity, we will actually build a word indexed by the interval $(0, 1)$ instead of $\mathbb{R}$. A corresponding word indexed by $\mathbb{R}$ can be directly derived by applying an increasing bijective function from $(0, 1)$ to $\mathbb{R}$, e.g., $f : (0, 1) \rightarrow \mathbb{R}; x \mapsto \frac{1}{1-x} - \frac{1}{x}$.

Consider the pair of states p, q , some closed $\mathbb{R}$ -paths $\bar{r}_1, \dots, \bar{r}_n$ in $\mathcal{A}_j$ (respectively starting in $q_1, \dots, q_n$, ending in $q'_1, \dots, q'_n$, and labeled by $Q_1, \dots, Q_n$), as well as some successor transitions $(s_1, \gamma_1, s'_1), \dots, (s_m, \gamma_m, s'_m)$ in Δ^s meeting the requirements for the application of the Cantor shuffle rule generating the transition (p, P, q) , for some set of states P .

In the spirit of the Cantor ternary set construction, we first define an infinite sequence of sets of intervals $(\bar{I}_k)_{k \in \mathbb{N}}$ by induction. We start by setting $\bar{I}_0 = \{[\frac{1}{3}, \frac{2}{3}]\}$. We then define inductively $\bar{I}_{k+1}$ as the set of closed intervals that correspond to the middle third of the maximal intervals that remain from $(0, 1)$ after removing every interval in $\bar{I}_j$, for every $j \in [1, k]$. For instance, $\bar{I}_2$ is equal to $\{[\frac{1}{9}, \frac{2}{9}], [\frac{7}{9}, \frac{8}{9}]\}$, and $\bar{I}_3$ is equal to $\{[\frac{1}{27}, \frac{2}{27}], [\frac{7}{27}, \frac{8}{27}], [\frac{19}{27}, \frac{20}{27}], [\frac{25}{27}, \frac{26}{27}]\}$. We then define n sets of closed intervals $\bar{H}_1, \dots, \bar{H}_n$ by setting

$$\bar{H}_i = \bigcup_{\substack{k \in \mathbb{N} \\ (k=i) \bmod n}} \bar{I}_k,$$

for every $i \in [1, n]$. This definition ensures density among all these sets, i.e., for any $i \in [1, n]$, $\bar{H}_i$ is dense in $\bigcup_{j \in [1, n]} \bar{H}_j$ (where the implicit order relation between disjoint intervals is $[a, b] < [c, d]$ iff $b < c$).

For every $i \in [1, n]$ and interval $[a, b]$ in $\bar{H}_i$, we consider a word $w^{[a, b]} = (\sigma_r^{[a, b]})_{r \in [a, b]}$ indexed by $[a, b]$ that labels a path $q_i \rightarrow q'_i$ visiting the set of states Q_i . The existence of such words results from our assumption that every open $\mathbb{R}$ -path admits a corresponding word indexed by any non-empty open interval of $\mathbb{R}$, and that a closed $\mathbb{R}$ -path is simply an open $\mathbb{R}$ -path with additional successor transitions at the beginning and at the end, corresponding to the bounds of the interval.

In order to obtain a complete path over $(0, 1)$, it remains to assign a state to both the first and last cuts 0^+ and 1^- , as well as to all the individual points left aside by the Cantor ternary set construction. The first task is easy to achieve, by simply assigning the state p to 0^+ and q to 1^- . For the second task, let C be the set of points left out by the construction of the sets $\bar{H}_i$, i.e., $C = (0, 1) \setminus \bigcup_{[a, b] \in (\bar{H}_1 \cup \dots \cup \bar{H}_n)} [a, b]$. We define a partition $C_1, \dots, C_m$ of C , such that each C_j is dense in C , i.e., $\forall k \in [1, m]. \forall x, y \in C. (x < y) \Rightarrow (\exists z. x < z < y \wedge z \in C_k)$. This can be achieved, for instance, by considering the representation of each element in C in base 3. Indeed, a notable property of the Cantor ternary set construction (using closed intervals for the definition of the sequence $(\bar{I}_k)_{k \in \mathbb{N}}$) is that the set C corresponds exactly to the real numbers in $(0, 1)$ that admit a (unique) representation in base 3 of the form $0.t_1 t_2 t_3 \dots$ where $t_i \in \{0, 2\}$ for every $i \in \mathbb{N}$, that does not end with an infinite sequence of 0 or 2 digits. For $k \in [1, m-1]$, we define the set C_k as the subset of C consisting of the elements that admit a notation ending with $(0^k 2^k)^\omega$, and we define $C_m = C \setminus (\bigcup_{k \in [1, m-1]} C_k)$. It is easy to show that for every $k \in [1, m]$, C_k is dense in C . Then, we simply associate to each real number $c \in C_k$ the transition (s_k, γ_k, s'_k) , for every $k \in [1, m]$.

Let $u = (\beta_r)_{r \in (0, 1)}$ be a word indexed by $(0, 1)$ defined as follows. For all $r \in C$ and $k \in [1, m]$, if $r \in C_k$ then $\beta_r = \gamma_k$. For all $r \in (0, 1) \setminus C$ and

$i \in [1, n]$, if there exists $[a, b] \in \overline{H}_i$ such that $r \in [a, b]$, then $\beta_r = \sigma_r^{[a, b]}$. This defines u entirely since by construction, the sets $C_1, \dots, C_m$ together with the intervals in $\overline{H}_1, \dots, \overline{H}_n$ form a partition of $(0, 1)$. A path π reading u can then be constructed by properly assembling the paths reading each word $w^{[a, b]}$ for every $[a, b] \in \overline{H}_1 \cup \dots \cup \overline{H}_n$, and every successor transition (s_k, γ_k, s'_k) , for every $k \in [1, m]$. It follows that π is of the required form, i.e., π is a path from p to q , that reads a word indexed by $(0, 1)$, and visits exactly the states in P . $\square$

6.2. Completeness

The completeness of the reachability algorithm described in Section 4.2 (Theorem 3) is a direct corollary of the following, more general, result.

Theorem 5. *Assume that the automaton $\mathcal{A}$ admits a path from a state q_I to state q_F visiting the set of states P , labeled by a word w indexed by $\mathbb{R}$. If this path is j -limited, then there exists an open $\mathbb{R}$ -path in $\mathcal{A}_j$ from q_I to q_F labeled by P .*

The remainder of this section is dedicated to proving Theorem 5. The proof is essentially based on mechanisms introduced in [22] for establishing the decidability of the first-order theory of order. The same mechanisms have been adapted to develop a self-contained proof of decidability for $\text{MFO}(\mathbb{R}, <)$ in [7].

The proof is by induction over j . The base case $j = 0$ is trivial. Indeed, every path labeled by $\mathbb{R}$ must involve a limit transition, and limit transitions cannot involve the empty set of states. For the inductive case, let $j \in [1, M]$ be such that Theorem 5 holds for all $(j-1)$ -limited paths. Since every k -limited path with $k < j$ is also a $(j-1)$ -limited path (every transition in A_k is also a transition in A_{j-1}), it suffices to consider the case of j -limited paths that are not $(j-1)$ -limited paths, i.e., that involve at least one limit transition of size j .

Throughout the remainder of this section, we consider a j -limited path π in $\mathcal{A}$ labeled by a word $w = (\alpha_r)_{r \in \mathbb{R}}$ indexed by $\mathbb{R}$.

Let $(x, y) \subseteq \mathbb{R}$ be a non-empty open interval. We denote by $S_\pi(x, y)$ the set of states visited by π between the cuts x^+ and y^- . In the particular cases $x = -\infty$ and $y = +\infty$, x^+ and y^- respectively correspond to the first and the last cuts of $\mathbb{R}$. Formally, we have

$$S_\pi(x, y) = \{q \in Q \mid \exists z. x < z < y \wedge (\pi(z^-) = q \vee \pi(z^+) = q)\}.$$

We say that an interval (x, y) is π -stable if for every interval (a, b) with $x \leq a < b \leq y$, one has $S_\pi(a, b) = S_\pi(x, y)$. In other words, an interval is π -stable if within any subinterval, the set of states visited in the corresponding subpath of π remains the same.

We now prove that any π -stable interval corresponds to a structure satisfying the simple shuffle rule. We first show that every limit transition followed within a π -stable interval involves the same set of states, namely, the set of all the states visited by π within that interval.

Lemma 6. *Let $(x, y) \subseteq \mathbb{R}$ be a π -stable interval. The set of limit transitions followed by π over (x, y) is exactly $\{S_\pi(x, y) \rightarrow q \mid \exists x < z < y. \pi(z^-) = q\} \cup \{q \rightarrow S_\pi(x, y) \mid \exists x < z < y. \pi(z^+) = q\}$. Furthermore, π also follows the right-limit transition $\pi(x^+) \rightarrow S_\pi(x, y)$ to leave $\pi(x^+)$ at the cut x^+ , and follows the left-limit transition $S_\pi(x, y) \rightarrow \pi(y^-)$ to reach $\pi(y^-)$ at the cut y^- .*

PROOF. Let us prove the left-limit transition case. The right-limit transition case is handled in a symmetrical way.

Let C_L be the set of left cuts z^- associated with the elements $z \in (x, y)$, and let Q_L be the set of states associated by π to the cuts in C_L . Notice that since $\mathbb{R}$ is complete, any subinterval of $\mathbb{R}$ is complete as well. Therefore, C_L also coincides with the set of cuts in $\widehat{(x, y)} \setminus \{y^-\}$ that do not admit a predecessor. Formally, Q_L is defined as follows:

$$Q_L = \{q \in S_\pi(x, y) \mid \exists z. x < z < y \text{ and } \pi(z^-) = q\}.$$

Consider a state $q \in Q_L \cup \{\pi(y^-)\}$. By contradiction, assume that either no left-limit transition of the form $P \rightarrow q$ is followed in π over the interval (x, y) , or that a left-limit transition of the form $P \rightarrow q$ with $P \neq S_\pi(x, y)$ is followed by π over (x, y) .

In the former case, if π does not follow any left-limit transition of the form $P \rightarrow q$ over (x, y) , then it can never reach q . Therefore, there exists a cut $z^- \in \widehat{(x, y)}$ that does not have a predecessor, such that $\lim_{z^-} \pi \rightarrow \pi(z^-)$ is not a transition in $\mathcal{A}$. This contradicts the fact that π is a valid path.

In the latter case, by definition of S_π , one has $P \subsetneq S_\pi(x, y)$. Let $s \in S_\pi(x, y)$ be such that $s \notin P$. Let $z \in (x, y]$ be such that $\pi(z^-) = q$ and $\lim_{z^-} \pi = P$. The definition of Q_L , and its non-emptiness, ensure that such a z exists. Let $z' \in (x, z)$ be such that for all z'' with $z' < z'' < z$, one has $\pi(z''^-) \neq s$ and $\pi(z''^+) \neq s$. The existence of such a z' is justified by the fact that the left-limit transition $P \rightarrow q$ must be followed in π to reach $\pi(z^-)$, otherwise s would belong to P . Now consider the interval (z', z) . This interval is non-empty, contained entirely within (x, y) , and such that $S_\pi(z', z) = P \subsetneq S$. This contradicts the original hypothesis about the π -stability of the interval (x, y) . $\square$

We are now ready to show that for any π -stable interval $(x, y) \subseteq \mathbb{R}$, the set of states $S_\pi(x, y)$ can always be rearranged into a structure satisfying the simple shuffle rule.

Lemma 7. *For any π -stable interval $(x, y) \subseteq \mathbb{R}$, the automaton $\mathcal{A}_j$ admits a corresponding transition $(\pi(x^+), S_\pi(x, y) \cup \{\pi(x^+), \pi(y^-)\}, \pi(y^-))$.*

PROOF. Since the interval (x, y) is π -stable, by Lemma 6 we have:

- A starting and an ending state $p_I = \pi(x^+)$ and $p_F = \pi(y^-)$, such that
 - $\pi(x^+) \rightarrow S_\pi(x, y) \in \Delta^\ell$, and
 - $S_\pi(x, y) \rightarrow \pi(y^-) \in \Delta^\ell$.

- A set of states Q_L and a set of states Q_R , such that

- $Q_L \cup Q_R = S_\pi(x, y)$,
- for each $q \in Q_L$, there exists $q' \in Q_R$, $\alpha \in \Sigma$ such that $(q, \alpha, q') \in \Delta^s$,
- for each $q \in Q_R$, there exists $q' \in Q_L$, $\alpha \in \Sigma$ such that $(q', \alpha, q) \in \Delta^s$,
- for each $q \in Q_L$, the transition $S_\pi(x, y) \rightarrow q$ belongs to Δ^ℓ , and
- for each $q \in Q_R$, the transition $q \rightarrow S_\pi(x, y)$ belongs to Δ^ℓ .

These are exactly the requirements for applying the simple shuffle rule, generating the transition $(p_I, S_\pi(x, y) \cup \{p_I, p_F\}, p_F)$. $\square$

Notice that unless $(-\infty, +\infty)$ is π -stable, no set of pairwise disjoint π -stable intervals can form a partition of $\mathbb{R}$, since such a partition would only be composed of open intervals. We denote by π -unstable point any real number that is not contained within a π -stable interval. We will show that any π -unstable point is necessarily an individual point, i.e., there does not exist a non-zero length interval of π -unstable points. Formally, let $\mathcal{S} = \bigcup_{(x,y) \text{ is } \pi\text{-stable}} (x, y)$, we have the following result.

Lemma 8. *For all $z \in \mathbb{R} \setminus \mathcal{S}$ and $\varepsilon > 0$, there exist $(x_1, y_1), (x_2, y_2) \subseteq \mathbb{R}$ such that:*

- $x_1 < y_1 < z < x_2 < y_2$,
- (x_1, y_1) and (x_2, y_2) are π -stable intervals,
- both $z - x_1 < \varepsilon$ and $y_2 - z < \varepsilon$.

PROOF. Let $z \in \mathbb{R} \setminus \mathcal{S}$ and $\varepsilon > 0$. The intervals (x_1, y_1) and (x_2, y_2) can be constructed by initially setting $x_1 = z - \varepsilon/2$, $y_1 = z - \varepsilon/4$, $x_2 = z + \varepsilon/4$, and $y_2 = z + \varepsilon/2$. If (x_1, y_1) is not directly π -stable, then there must exist a subinterval (x'_1, y'_1) of (x_1, y_1) such that $S_\pi(x'_1, y'_1) \subsetneq S_\pi(x_1, y_1)$. If (x'_1, y'_1) is π -stable, then the construction stops. Otherwise the same operation is repeated after replacing (x_1, y_1) by (x'_1, y'_1) until reaching a π -stable subinterval. This will necessarily happen after a finite number of steps, since the set of states Q of $\mathcal{A}$ is finite. The same construction can be applied to find a suitable interval (x_2, y_2) . $\square$

The remainder of the proof of Theorem 5 boils down to showing that the π -stable intervals and the π -unstable points can be combined into a complete partition of $\mathbb{R}$, and that the different ways to combine them correspond to the rules implemented in our algorithm.

We now introduce the notion of π -pseudo-stable interval, corresponding to a non-empty open interval $(x, y) \subseteq \mathbb{R}$ such that $\mathcal{A}_j$ admits an open $\mathbb{R}$ -path from $\pi(x^+)$ to $\pi(y^-)$ labeled by $S_\pi(x, y) \cup \{\pi(x^+), \pi(y^-)\}$. We denote by $\bar{R}_{(x,y)}$ the set of all the open $\mathbb{R}$ -paths in $\mathcal{A}_j$ that start in $\pi(x^+)$ and end in $\pi(y^-)$, of length $n \leq |\Delta_j|^2$, labeled by $S_\pi(x, y) \cup \{\pi(x^+), \pi(y^-)\}$, where Δ_j is the set

of transitions of the automaton $\mathcal{A}_j$. Bounding the length of the $\mathbb{R}$ -paths by $|\Delta_j|^2$ ensures that the set $\bar{R}_{(x,y)}$ is finite, while also guaranteeing that if $\mathcal{A}_j$ admits an open $\mathbb{R}$ -path from the state $\pi(x^+)$ to the state $\pi(y^-)$ labeled by $S_\pi(x,y) \cup \{\pi(x^+), \pi(y^-)\}$, then an analogous one exists in $\bar{R}_{(x,y)}$. The $|\Delta_j|^2$ bound provides an upper bound for the minimal length of an $\mathbb{R}$ -path visiting every transition in Δ_j .

Notice that any π -stable interval is also π -pseudo-stable, since by Lemma 7, an open $\mathbb{R}$ -path of length one (i.e., a single transition of $\mathcal{A}_j$) is associated with any π -stable interval. Furthermore, we call a π -pseudo-stable interval π -maximal if it is not contained within a greater π -pseudo-stable interval. Formally, an interval $(x,y) \subseteq \mathbb{R}$ is π -maximal if and only if

- $x = \min\{z \in \mathbb{R} \mid \text{there exists an open } \mathbb{R}\text{-path from } \pi(z^+) \text{ to } \pi(y^-) \text{ labeled by } S_\pi(z,y) \cup \{\pi(z^+), \pi(y^-)\}\}$, and
- $y = \max\{z \in \mathbb{R} \mid \text{there exists an open } \mathbb{R}\text{-path from } \pi(x^+) \text{ to } \pi(z^-) \text{ labeled by } S_\pi(x,z) \cup \{\pi(x^+), \pi(z^-)\}\}$.

Before moving to the case of the (reverse) infinite repetition rule, let us first introduce the following useful result.

Lemma 9. *Let $\ell \in [1, j-1]$, and let $\bar{r}$ be an open $\mathbb{R}$ -path in $\mathcal{A}_\ell$ from a state q to a state q' , labeled by a set of states P . For every $p \in P$, $\mathcal{A}_\ell$ also admits an open $\mathbb{R}$ -path $\bar{r}'$ from q to q' labeled by P of the form $(r_1, R_1, r'_1), \dots, (r_n, R_n, r'_n)$, such that:*

- $r_1 = q$ and $r'_n = q'$, and
- there exists $i \in [1, n]$ such that either $r_i = p$ or $r'_i = p$.

PROOF. The existence of $\bar{r}$ implies that $\mathcal{A}$ admits a path μ from q to q' visiting exactly the states in P and labeled by a word w indexed by $\mathbb{R}$, such that $\mu(-\infty) = q$ and $\mu(+\infty) = q'$. For any $p \in P$, there exists by definition of μ a point $x \in \mathbb{R}$ such that either $\mu(x^-) = p$ or $\mu(x^+) = p$.

Consider the two subpaths of μ over the intervals $(-\infty, x)$ and $(x, +\infty)$. These paths are labeled by a word indexed by a non-empty open interval of $\mathbb{R}$, i.e., of order type λ . Therefore, there also exists a path μ' (resp. μ'') from q to $\mu(x^-)$ (resp. from $\mu(x^+)$ to q') labeled by a word w' (resp. w'') indexed by $\mathbb{R}$.

Given the inductive hypothesis on the completeness of $\mathcal{A}_\ell$, there must exist two open $\mathbb{R}$ -paths $\bar{r}_1$ and $\bar{r}_2$ in $\mathcal{A}_\ell$, such that $\bar{r}_1$ goes from q to $\mu(x^-)$ and $\bar{r}_2$ from $\mu(x^+)$ to q' . Furthermore, either $\bar{r}_1$ ends with a transition of the form (r_i, R_i, p) (if $\mu(x^-) = p$) or $\bar{r}_2$ starts with a transition of the form (p, R_i, r'_i) (if $\mu(x^+) = p$). Additionally, since μ is a valid path, there must also exist a transition $(\mu(x^-), \alpha, \mu(x^+)) \in \Delta^s$ for some $\alpha \in \Sigma$. Therefore, one can concatenate $\bar{r}_1$ with $\bar{r}_2$ to obtain the open $\mathbb{R}$ -path of the desired form. The fact that the union of the labels of $\bar{r}_1$ and $\bar{r}_2$ is equal to P is immediate. $\square$

We are now ready to deal with the case where a set of π -pseudo-stable intervals can be combined thanks to the (reverse) infinite repetition rule.

Lemma 10. *Let (x, y) be a non-empty open interval in $\mathbb{R}$. If every interval (z, z') such that $x < z < z' < y$ is π -pseudo-stable, then (x, y) is π -pseudo-stable as well.*

PROOF. Consider the subpath of π over the interval (x, y) . By definition, this subpath starts in $\pi(x^+)$ and ends in $\pi(y^-)$. It therefore ends with the left-limit transition $P \rightarrow \pi(y^-)$, where $P = \lim_{y^-} \pi$ and $|P| = k \leq j$. Let $z \in (x, y)$ be such that $S_\pi(z, y) = P$. Let $(p', \alpha, p) \in \Delta^s$ be such that, for every $y' < y$, π follows the transition (p', α, p) within the interval (y', y) , where $p, p' \in P$. Intuitively, this means that π follows this transition in every left neighborhood of y^- . Such a transition always exists since the set of successor transitions Δ^s is finite.

Let $(z_i)_{i \in \mathbb{N}}$ be a sequence of real numbers such that $\lim_{i \rightarrow +\infty} z_i = y$, and for every $i \in \mathbb{N}$, one has $z < z_i < z_{i+1} < y$, $\pi(z_i^-) = p'$, and $\pi(z_i^+) = p$. The existence of such a sequence is a direct consequence of the fact that π follows the transition (p', α, p) within every interval (y', y) with $y' < y$. By the infinite version of Ramsey's theorem [23], we can assume that this sequence satisfies $S_\pi(z_i, z_{i+1}) = P$, for all $i \in \mathbb{N}$.

We now show that there necessarily exists an open $\mathbb{R}$ -path from $\pi(z_0^+)$ to $\pi(y^-)$ labeled by $P \cup \{\pi(y^-)\}$ in $\mathcal{A}_k$, and therefore also in $\mathcal{A}_j$ since $k \leq j$. Notice that since $z_0 > z$, one necessarily has $P = S_\pi(z_0, y)$, with additionally $\pi(z_0^+) \in P$. Every interval (z_i, z_{i+1}) is π -pseudo-stable by hypothesis.

In the case where $\mathcal{A}_{k-1}$ admits an open $\mathbb{R}$ -path from p to p' labeled by P , the existence of such an open $\mathbb{R}$ -path in $\mathcal{A}_{k-1}$ together with the existence of a successor transition $(p', \alpha, p) \in \Delta^s$ implies that L_k necessarily contains the transition $(\pi(z_0^+), P \cup \{\pi(y^-)\}, \pi(y^-))$. Therefore, there exists an open $\mathbb{R}$ -path from $\pi(z_0^+)$ to $\pi(y^-)$ labeled by $P \cup \{\pi(y^-)\}$ in $\mathcal{A}_k$, i.e., (z_0, y) is π -pseudo-stable.

Consider now the case where $\mathcal{A}_k$ admits an open $\mathbb{R}$ -path $\bar{r}$ from p to p' labeled by P , but $\mathcal{A}_{k-1}$ does not. In this situation, a transition allowing to reach $\pi(y^-)$ from $p = \pi(z_0^+)$ may not be generated, since such a transition can only be created at Step k (since $|P| = k$), but relies on an open $\mathbb{R}$ -path that itself only appears at the same Step k . We now show that even if such transitions are not generated, there will nonetheless exist an open $\mathbb{R}$ -path from p to $\pi(y^-)$.

Let $(r, R, r') \in \Delta_k$ be a transition generated at Step k that is involved in $\bar{r}$, i.e., $\bar{r}$ is of the form $(r_1, R_1, r'_1), \dots, (r_n, R_n, r'_n)$ such that $r_1 = p$, $r'_n = p'$, and there exists $i \in [1, n]$ such that $(r_i, R_i, r'_i) = (r, R, r')$. In the eventuality where multiple transitions generated at Step k are involved in the open $\mathbb{R}$ -path $\bar{r}$ from p to p' , we assume that (r, R, r') is the leftmost one in $\bar{r}$.

We now show that for every rule that may have produced the transition (r, R, r') , some other transitions must also have been generated, resulting in the existence of an open $\mathbb{R}$ -path in $\mathcal{A}_k$ from $\pi(z_0^+)$ to $\pi(y^-)$ labeled by $P \cup \{\pi(y^-)\}$. A first observation is that, independently from the rule that has produced the transition, R must be equal to P . Indeed, since (r, R, r') has been generated at step k , one has $|R| \geq k$. Furthermore, since this transition is involved in an open $\mathbb{R}$ -path labeled by P , one also has $R \subseteq P$. Since $|P| = k$, it follows that $P = R$.

Consider first the case where the transition (r, P, r') has been generated by an application of the (simple or Cantor) shuffle rule. Since the conditions of the shuffle rule were satisfied to generate the transition (r, P, r') , and since $\mathcal{A}$ admits the limit transition $P \rightarrow \pi(y^-)$, the construction of S_k also necessarily yielded the transition $(r, P \cup \{\pi(y^-)\}, \pi(y^-))$. By hypothesis, there exists an open $\mathbb{R}$ -path from p to p' labeled by P that involves the transition (r, P, r') , therefore an open $\mathbb{R}$ -path from p to $\pi(y^-)$ labeled by $P \cup \{\pi(y^-)\}$ can be easily constructed by taking the first part of $\bar{r}$ up to the transition (r, P, r') (excluded) and simply appending the transition $(r, P \cup \{\pi(y^-)\}, \pi(y^-))$.

Similarly, to deal with the case where the transition (r, P, r') has been generated by an application of the infinite repetition rule, one simply considers the conditions used to apply the infinite repetition rule to generate the transition (r, P, r') , this time replacing the endpoint state r' by $\pi(y^-)$, which is valid since $\mathcal{A}$ admits the limit transition $P \rightarrow \pi(y^-)$. It follows that L_k necessarily contains the transition $(r, P \cup \{\pi(y^-)\}, \pi(y^-))$, hence there exists an open $\mathbb{R}$ -path in $\mathcal{A}_k$ from $p = \pi(z_0^+)$ to $\pi(y^-)$ labeled by $P \cup \{\pi(y^-)\}$.

It remains to consider the case where the transition (r, P, r') has been generated by an application of the reverse infinite repetition rule. This case is more tricky as the application of the reverse infinite repetition rule does not straightforwardly implies that an analogous rule has been applied to reach $\pi(y^-)$.

Since the transition (r, P, r') has been generated at Step k by the reverse infinite repetition rule, the following conditions are necessarily satisfied:

- The automaton $\mathcal{A}_{k-1}$ admits an open $\mathbb{R}$ -path $\bar{r}'$ from a state r'' to the state r' labeled by P , such that $(r', \alpha, r'') \in \Delta^s$ for some $\alpha \in \Sigma$.
- The states r, r' and r'' belong to P , by definition of the label of $\mathbb{R}$ -paths for r' and r'' . If r did not belong to P , then the generation of the transition (r, P, r') would have relied on a limit transition involving the set $P \setminus \{r\}$ of cardinality $k - 1$, which could therefore not have been generated at Step k .

Since r belongs to P , by Lemma 9, $\mathcal{A}_{k-1}$ admits an open $\mathbb{R}$ -path from r'' to r' of the form $(q_1, Q_1, q'_1), \dots, (q_m, Q_m, q'_m)$ such that $q_1 = r''$, $q'_m = r'$, and such that either $q_\ell = r$ or $q'_\ell = r$, for some $\ell \in [1, m]$.

Consider the first case, i.e., assume that there exists $\ell \in [1, m]$ such that $q_\ell = r$. Given the existence of

- the open $\mathbb{R}$ -path $\bar{r}'$ in $\mathcal{A}_{k-1}$ labeled by P that starts in r'' and ends in r' ,
- a successor transition $(r', \alpha, r'') \in \Delta^s$, and
- the limit transition $P \rightarrow \pi(y^-)$,

it follows that at Step k , the construction of the set L_k must produce the transition $(q_j, P \cup \{\pi(y^-)\}, \pi(y^-))$ for every $j \in [1, m]$, by application of the infinite repetition rule, in particular for $j = \ell$. Notice that the way we defined the infinite repetition rule implies that its application only generates a transition

for the state associated with the starting point of the first transition of the open $\mathbb{R}$ -path. But since this $\mathbb{R}$ -path actually forms a cycle, there exists a similar open $\mathbb{R}$ -path sharing the same label rooted at every state q_j visited by the cycle, on which the infinite repetition rule can also be applied.

Recall that the transition (r, P, r') appears in the $\mathbb{R}$ -path $\bar{r}$. By concatenating the first transitions composing $\bar{r}$ up to the $(i-1)$ -th one (i.e., stopping just before the occurrence of (r, P, r') , ensuring we end up in a state s such that there exists a transition $(s, \alpha, r) \in \Delta^s$ for some $\alpha \in \Sigma$) with the transition $(r, P, \pi(y^-))$, we obtain an open $\mathbb{R}$ -path of the desired form.

Now consider the second case, i.e. assume that there exists $\ell \in [1, m]$ such that $q'_\ell = r$. Considering the existence of

1. the open $\mathbb{R}$ -path $\bar{r}'$ in $\mathcal{A}_{k-1}$ labeled by P that starts in r'' and ends in r' ,
2. a successor transition $(r', \alpha, r'') \in \Delta^s$,
3. the limit transition $r \rightarrow P \in \Delta^\ell$ mandatory to apply the reverse infinite repetition rule and generate the transition (r, P, r') , and
4. the limit transition $P \rightarrow \pi(y^-)$,

it follows that:

- [(1),(2),(3)] At Step k , the construction of the set R_k must yield the transition (r, P, q'_j) for every $j \in [1, m]$, by application of the reverse infinite repetition rule, and in particular for $j = \ell$.
- [(1),(2),(4)] At Step k , the construction of the set L_k must yield the transition $(q_j, P \cup \{\pi(y^-)\}, \pi(y^-))$ for every $j \in [1, m]$, by application of the infinite repetition rule, and in particular for $j = 1$.

One can therefore build an open $\mathbb{R}$ -path that coincides with $\bar{r}$ (and therefore starts in p) up to the transition (r, P, r') , then follows the transition (r, P, r) before following the subpath of $\bar{r}'$ that starts right after the transition $(q_\ell, Q_\ell, q'_\ell)$ (which is possible because $q'_\ell = r$) and ends in r' , and finally follows the transition $(r'', P \cup \{\pi(y^-)\}, \pi(y^-))$.

A similar approach can be used to show that there exists a sequence $(z'_i)_{i \in \mathbb{N}}$ converging to x , such that there exists an open $\mathbb{R}$ -path from $\pi(x^+)$ to $\pi(z'_0^-)$ labeled by $S_\pi(x, z'_0) \cup \{\pi(x^+), \pi(z'_0^-)\}$ in $\mathcal{A}_j$. We therefore have two π -pseudo-stable intervals (x, z'_0) and (z_0, y) such that $x < z'_0 < z_0 < y$.

By hypothesis, the interval (z'_0, z_0) is also π -pseudo-stable. Therefore, by concatenating some open $\mathbb{R}$ -paths corresponding to the intervals (x, z'_0) , (z'_0, z_0) , and (z_0, y) in that order, we obtain an open $\mathbb{R}$ -path for the whole interval (x, y) , which entails that (x, y) is also π -pseudo-stable. $\square$

It remains to deal with π -unstable points that are neither a junction point between two consecutive π -pseudo-stable intervals, nor the limit of an infinite sequence of π -pseudo-stable intervals. A first step in that direction is to characterize the distribution of the π -maximal intervals.

Lemma 11. *The linear ordering between the π -maximal intervals of $\mathbb{R}$ is dense, i.e., for all π -maximal intervals $(x_1, y_1), (x_2, y_2)$ such that $y_1 < x_2$, there exists a π -maximal interval (x_3, y_3) such that $y_1 < x_3 < y_3 < x_2$.*

PROOF. By contradiction, we prove that two such π -maximal intervals cannot be consecutive. Assume that it is the case, i.e., there exist two π -maximal intervals $(x_1, y_1), (x_2, y_2)$ such that $y_1 < x_2$, and such that there does not exist another maximal interval between y_1 and x_2 .

Since $y_1 < x_2$, the interval (y_1, x_2) is not empty, and any non-empty open interval of $\mathbb{R}$ contains a π -stable interval. Otherwise, it would mean that (y_1, x_2) only contains π -unstable points, which contradicts Lemma 8.

To any π -stable interval corresponds a greater π -maximal interval (it can be build by applying the rules generating π -pseudo-stable intervals, and then extending it as much as possible). However, the corresponding maximal interval cannot be contained strictly within (y_1, x_2) by hypothesis, and it cannot span before y_1 or beyond x_2 since $(x_1, y_1), (x_2, y_2)$ are both π -maximal by hypothesis, hence the contradiction. $\square$

Finally, the last step in proving Theorem 5 consists in showing that any non-trivial dense distribution of π -maximal intervals — i.e., containing at least two distinct π -maximal intervals, and therefore an infinity of them thanks to the density property — would be captured by the Cantor shuffle rule. This leads to the following result that states that there only exists a single π -maximal interval.

Lemma 12. *The only π -maximal interval of $\mathbb{R}$ is $\mathbb{R}$ itself.*

PROOF. By contradiction, assume that there exist two distinct π -maximal intervals of $\mathbb{R}$. Let us pick two distinct π -maximal intervals (x_1, y_1) and (x_2, y_2) with $x_1 < y_1 < x_2 < y_2$, such that for any two π -maximal intervals $(a_1, b_1), (a_2, b_2)$ with $y_1 < a_1 < b_1 < a_2 < b_2 < x_2$, one has $S_\pi(y_1, x_2) = S_\pi(a_1, b_2)$. In other words, the set of states visited between the intervals (x_1, y_1) and (x_2, y_2) is minimal w.r.t. any subinterval of $\mathbb{R}$ containing at least two distinct π -maximal intervals. It is always possible to chose two such π -maximal intervals. Indeed, one can start from any two π -maximal intervals, and replace them if needed by two other intervals that appear between them, such that the set of states that π visits in the interval between them is strictly smaller. This operation can be repeated until reaching a fixed point, which will necessarily occur after a finite number of steps, since the set of states Q of $\mathcal{A}$ is finite.

Remember that any π -maximal interval is also π -pseudo-stable by definition. Therefore, for each π -maximal interval (a, b) , there exists an open $\mathbb{R}$ -path in $\mathcal{A}_j$, and more precisely in $\bar{R}_{(a,b)}$, from $\pi(a^+)$ to $\pi(b^-)$ labeled by $S_\pi(a, b) \cup \{\pi(a^+), \pi(b^-)\}$. By combining them with the successor transitions $(\pi(a^-), \alpha_a, \pi(a^+))$ and $(\pi(b^-), \alpha_b, \pi(b^+))$ of Δ^s , these open $\mathbb{R}$ -paths can be turned into closed $\mathbb{R}$ -paths from $\pi(a^-)$ to $\pi(b^+)$, labeled by $S_\pi(a, b) \cup \{\pi(a^-), \pi(a^+), \pi(b^-), \pi(b^+)\}$. For each π -maximal interval (a, b) in (y_1, x_2) , let $\bar{r}^{[a,b]}$ be one such closed $\mathbb{R}$ -path. Using a similar argument as in the proof of Lemma 6, one can also show that both limit transitions $S_\pi(y_1, x_2) \rightarrow \pi(a^-)$ and $\pi(b^+) \rightarrow S_\pi(y_1, x_2)$ belong to Δ^ℓ . Furthermore, to each π -unstable point $z \in (y_1, x_2)$ we associate the successor transition $(\pi(z^-), \alpha_z, \pi(z^+)) \in \Delta^s$. Following the same

approach as before, one can show that both limit transitions $S_\pi(y_1, x_2) \rightarrow \pi(z^-)$ and $\pi(z^+) \rightarrow S_\pi(y_1, x_2)$ belong to Δ^ℓ .

Notice that since Δ^s and each $\bar{R}_{(a,b)}$ are finite, both the set of successor transitions and the set of closed $\mathbb{R}$ -paths, respectively associated with the π -unstable points and π -maximal intervals in (y_1, x_2) , are finite as well.

Let P denote the set of states corresponding to the union of the labels of all closed $\mathbb{R}$ -paths $\bar{r}^{[a,b]}$ for every π -maximal interval (a, b) within (y_1, x_2) , together with the states involved in the successor transitions associated with all π -unstable points that lie between y_1 and x_2 . Formally, we have

$$P = \{q \mid z \text{ is a } \pi\text{-unstable point in } (y_1, x_2), \text{ and either } \pi(z^-) = q \text{ or } \pi(z^+) = q\} \\ \cup \{q \mid (a, b) \text{ is a } \pi\text{-maximal interval in } (y_1, x_2), R \text{ labels } \bar{r}^{[a,b]}, \text{ and } q \in R\}.$$

In order to show that the Cantor shuffle rule can be applied to the pair of transitions $\pi(y_1^+) \rightarrow S_\pi(y_1, x_2)$ and $S_\pi(y_1, x_2) \rightarrow \pi(x_2^-)$, it remains to show that $P = S_\pi(y_1, x_2)$ holds. It is clear that $P \subseteq S_\pi(y_1, x_2)$ holds. Indeed, for every π -unstable point z , we considered the states $\pi(z^-)$ and $\pi(z^+)$, which belong to $S_\pi(y_1, x_2)$ by definition. Similarly, the definition of the closed $\mathbb{R}$ -path $\bar{r}^{[a,b]}$ for each π -maximal interval (a, b) implies that its label is a subset of $S_\pi(y_1, x_2)$. In order to show $S_\pi(y_1, x_2) \subseteq P$, let us assume by contradiction that there exists $q \in S_\pi(y_1, x_2) \setminus P$. Then, there must exist $z \in (y_1, x_2)$ such that either $q = \pi(z^-)$ or $q = \pi(z^+)$, and z is neither a π -unstable point, nor contained within a π -maximal interval. Such a real number z cannot exist, as the set of π -unstable points combined with the set of π -maximal intervals (extended with their upper and lower bound to form closed intervals) form a partition of the interval (y_1, x_2) .

All the requirements of the Cantor shuffle rule are therefore met. It follows that the construction of the automaton $\mathcal{A}_j$ yields the transition $(\pi(y_1^+), S_\pi(y_1, x_2) \cup \{\pi(y_1^+), \pi(x_2^-)\}, \pi(x_2^-))$. Since (x_1, y_1) (resp. (x_2, y_2)) is π -pseudo-stable, there exists an open $\mathbb{R}$ -path $\bar{r}_1$ (resp. $\bar{r}_2$) from $\pi(x_1^+)$ to $\pi(y_1^-)$ (resp. from $\pi(x_2^+)$ to $\pi(y_2^-)$). By concatenating $\bar{r}_1$, the transition $(\pi(y_1^+), S_\pi(y_1, x_2) \cup \{\pi(y_1^+), \pi(x_2^-)\}, \pi(x_2^-))$ of $\mathcal{A}_j$, and $\bar{r}_2$ (in that order), we obtain an open $\mathbb{R}$ -path from $\pi(x_1^+)$ to $\pi(y_2^-)$ labeled by $S_\pi(x_1, y_2) \cup \{\pi(x_1^+), \pi(y_2^-)\}$. This implies that (x_1, y_2) is π -pseudo-stable, which contradicts the assumption that (x_1, y_1) and (x_2, y_2) are π -maximal intervals. $\square$

This concludes the proof of Theorem 5.

The proof of correctness developed in this section can be straightforwardly adapted to the algorithm of Section 5, dealing with the reachability problem for words indexed by $\mathbb{Q}$.

7. Conclusions and Perspectives

In this work, we addressed the reachability problem for automata on linear orderings, where the words recognized by such automata are restricted to be indexed by $\mathbb{R}$ or $\mathbb{Q}$. In order to restrict the ordering to the reals, it might be

tempting to try to characterize the ordering on the reals with an automaton, for instance expressing that the only acceptable orderings are dense, complete, and without first and last elements. As a first result (Section 4.1), we however exhibit an automaton on linear orderings that accepts words indexed by dense and complete orderings without first and last elements, but that does not accept any word indexed by $\mathbb{R}$. The existence of such an automaton highlights the fact that a full characterization of $\mathbb{R}$ by an automaton is unlikely to exist, and motivates the search of a dedicated reachability algorithm over $\mathbb{R}$.

Our second result (Section 4.2) is a polynomial algorithm to decide the reachability (and equivalently the emptiness) problem for automata on linear orderings, over $\mathbb{R}$ specifically. We provide a pseudo-code implementation (Section 4.6), further emphasizing its polynomial complexity. This reachability algorithm is also adapted (Section 5) for the case of $\mathbb{Q}$. An efficient implementation of those algorithms has recently been proposed in [24].

We believe this work is a stepping stone towards a complete decision procedure for the monadic first-order theory of order over the reals and the rationals. The idea would be to build automata on linear orderings recognizing the set of models of a formula being decided. Usual operations on automata, adapted to automata on linear orderings, can be used to compute union and intersection of sets of models. Existential quantification reduces to a simple projection operation on the automaton. The last required operation for a full decision procedure is the universal first-order quantification. This is an unavoidable step since these automata do not admit complementation [3]. We presented in [25] such an operation for the case of natural numbers and Büchi automata. The next step in our quest towards an automata-based decision procedure for $\text{MFO}(\mathbb{R}, <)$ and $\text{MFO}(\mathbb{Q}, <)$ will be to extend this result to automata on linear orderings.

Acknowledgments. The authors wish to thank Thomas Braipson and Tom Clara, as well as the anonymous reviewers, for their valuable comments on earlier versions of this work.

References

- [1] V. Bruyère, O. Carton, Automata on linear orderings, *Journal of Computer and System Sciences* 73 (1) (2007) 1–24.
- [2] A. Bès, O. Carton, A Kleene theorem for languages of words indexed by linear orderings, *International Journal of Foundations of Computer Science* 17 (03) (2006) 519–541.
- [3] N. Bedon, A. Bès, O. Carton, C. Rispal, Logic and rational languages of words indexed by linear orderings, *Theory of Computing Systems* 46 (2010) 737–760.
- [4] M. O. Rabin, Decidability of second-order theories and automata on infinite trees., *Transactions of the American Mathematical Society* 141 (1969) 1–35.

- [5] S. Shelah, The monadic theory of order, *Annals of Mathematics* 102 (3) (1975) 379–419.
- [6] J. P. Burgess, Y. Gurevich, The decision problem for linear temporal logic, *Notre Dame Journal of Formal Logic* 26 (2) (1985) 115–128.
- [7] B. Boigelot, P. Fontaine, B. Vergain, On the theory of order over real numbers, *Mathematics in Computer Science* (To be published) (2025).
- [8] M. Reynolds, The complexity of temporal logic over the reals, *Annals of Pure and Applied Logic* 161 (8) (2010) 1063–1096.
- [9] A. Rabinovich, Temporal logics over linear time domains are in PSPACE, *Information and Computation* 210 (2012) 40–67.
- [10] J. R. Büchi, On a decision method in restricted second order arithmetic, in: *Proc. Intl. Congress on Logic, Methodology and Philosophy of Science*, 1960.
- [11] J. R. Büchi, Transfinite automata recursions and weak second order theory of ordinals, in: *Proc. Intl. Congress on Logic, Methodology, and Philosophy of Science*, 1965.
- [12] O. Carton, Accessibility in automata on scattered linear orderings, in: *Proc. Intl. Symp. on Mathematical Foundations of Computer Science*, Springer, 2002, pp. 155–164.
- [13] J. Cristau, Automata and temporal logic over arbitrary linear time, in: *Proc. Intl. Conf. on Foundations of Software Technology and Theoretical Computer Science*, Vol. 4 of LIPIcs, 2009, pp. 133–144.
- [14] B. Boigelot, P. Fontaine, B. Vergain, Non-emptiness test for automata over words indexed by the reals and rationals, in: S. Z. Fazekas (Ed.), *Implementation and Application of Automata*, Vol. 15015 of *Lecture Notes in Computer Science*, Springer Nature Switzerland, 2024, pp. 94–108.
- [15] J. G. Rosenstein, *Linear orderings*, Academic press, 1982.
- [16] D. E. Muller, Infinite sequences and finite machines, in: *Proc. Annual Symp. on Switching Circuit Theory and Logical Design*, IEEE Computer Society, 1963, pp. 3–16.
- [17] Y. Choueka, Finite automata, definable sets, and regular expressions over ω^n -tapes, *Journal of Computer and System Sciences* 17 (1) (1978) 81–97.
- [18] G. Cantor, Beiträge zur Begründung der transfiniten Mengenlehre, *Mathematische Annalen* 46 (1895) 481–512.
- [19] W. Sierpiński, *Cardinal and ordinal numbers* (2nd edition), PWN, Warszawa, 1965.

- [20] G. Cantor, Über unendliche, lineare Punktmannichfaltigkeiten, *Mathematische Annalen* 20 (1) (1882) 113–121.
- [21] E. DiBenedetto, *Real analysis*, Springer, 2002.
- [22] H. Läuchli, J. Leonard, On the elementary theory of linear order, *Fundamenta Mathematicae* 59 (1) (1966) 109–116.
- [23] F. P. Ramsey, On a problem of formal logic, *Proceedings of the London Mathematical Society* 30 (1930) 264–286.
- [24] T. Braipson, Non-emptiness test for automata on linear orderings: an efficient implementation, Master’s thesis, Université de Liège, Belgium, <http://hdl.handle.net/2268.2/23218> (2025).
- [25] B. Boigelot, P. Fontaine, B. Vergain, Universal first-order quantification over automata, *International Journal of Foundations of Computer Science* (2025).