Automata on linear orderings*

Véronique Bruyère — Institut d'Informatique† — Université de Mons-Hainaut

Olivier Carton — LIAFA‡ — Université Paris 7

December 2, 2004

Abstract

We consider words indexed by linear orderings. These extend finite, (bi-)infinite words and words on ordinals. We introduce finite automata and rational expressions for these words. We prove that for countable scattered linear orderings, these two notions are equivalent. This result extends Kleene's theorem.

1 Introduction

The theory of automata finds its origin in a paper of S. C. Kleene of 1956 where the basic theorem, known as Kleene's theorem, is proved for finite words [17]. Since then, the study of automata has become a branch of theoretical computer science in its own right and has developed in many directions, including study of automata working on infinite words, trees, and traces ([19, 24]).

In this paper we focus on automata working on linearly ordered objects. Examples of such objects are finite, infinite, bi-infinite and ordinal words, where the underlying linear ordering is respectively a finite ordering, the ordering of the positive integers, the ordering of the integers and the ordering of an ordinal. Each such class of words has its corresponding family of automata and in all cases, a Kleene-like theorem exists. Historically, Büchi introduced the so-called Büchi automata working on infinite words, to show the decidability of the monadic second order theory of $\langle \mathbb{N}, < \rangle$ [7]. He later extended the method to countable ordinals by using appropriate automata [8]. Büchi introduced ω-rational operations for infinite words in [8]. For words indexed by an ordinal less than ω^ω, Choueka defined rational operations and proved an analogue of Kleene's theorem in [12]. This result is extended to any countable ordinal by Wojciechowski in [25]. The case of bi-infinite words is treated in [18, 14].

*partially supported by the CGRI-CNRS cooperation project *Combinatoire et automates*

†Université de Mons-Hainaut, Le Pentagone, 6 avenue du Champ de Mars, B-7000 Mons, Belgium, Email: `Veronique.Bruyere@umh.ac.be`, Url: `http://sun1.umh.ac.be/~vero/`

‡Université Paris 7 - case 7014, 2 place Jussieu, F-75251 Paris cedex 05, France, Email: `Olivier.Carton@liafa.jussieu.fr`, Url: `http://www.liafa.jussieu.fr/~carton/`

The goal of this paper is to provide a unified approach of the study of words indexed by any countable linear ordering. We introduce a new notion of automaton which is simple, natural, and includes previously defined automata. We also define rational expressions and we prove the corresponding extension of Kleene's theorem (Theorem 20).

Words indexed by arbitrary countable linear ordering were first considered in [13] as solutions of systems of equations. They can be viewed as the frontiers of labeled binary trees in which the labels of the leaves are read from left to right. For these words, some kind of rational expressions have been studied in [13, 16, 23]. They have the nice property that the rational operations are total operators. For instance, an ω-power can be concatenated with a finite word and the resulting word can be iterated by a reversed ω-power. These operations lead to a characterization of the words which are the frontier of regular trees.

In our model, there are three groups of rational operations. First, there are the usual union, concatenation, finite iteration and omega iteration, as well as the ordinal iteration introduced in [25] for ordinal words. Next, there are reverse omega iteration and reverse ordinal iteration. They capture the left-infinite ordinal words and the bi-infinite words. Finally, a last operation is necessary. It is the iteration for all linear orderings. This binary operation (denoted by $\diamond$ below) is subtle since it takes into account the cuts of a linear ordering as defined in [21].

We also define automata that work on words indexed by linear orderings. In our formulation, the notion of a path in an automaton depends heavily on cuts. In the case of a finite word w of length n, the underlying ordering is $1 < 2 < \cdots < n$. The $n + 1$ states of a path for w are inserted between the letters of w, i.e. at the cuts of the ordering. In general, for a word w indexed by a linear ordering, the states of a path labeled by w are indexed by the cuts of the ordering. Our automata have three types of transitions: the usual successor transitions, left limit transitions, and right limit transitions. For two consecutive states in a path, there is always a successor transition labeled by the letter in between. For a state q which has no predecessor in a path, there is always a left limit transition between the left limit set P and q. Right limit transitions are used when a state has no successor in a path. For a Muller automaton ([24]), the left limit set P is nothing else than the set of states appearing infinitely often along the path. In our model, there is an additional left limit transition to a state q at the end of the path. The path is then successful if this last state q is final.

We think that our approach may have applications to the area of timed automata which are used for the specification and verification of real-time systems. Recently, ordinal words (called Zeno words) were considered as models of infinite sequences of actions which occur in a finite interval of time [15, 3]. While the intervals of time are finite, infinite sequences of actions can be concatenated. A Kleene-like theorem already exists for standard timed automata (where infinite sequences of actions are supposed to generate divergent sequences of times) [1]. In [3], automata of Choueka and Wojciechowski are adapted to Zeno words. A kind of Kleene's theorem is proved, that is, the class of Zeno languages is the closure under an operation called refinement of the class of languages accepted by standard timed automata.

The paper is organized as follows. Basic notions on linear orderings are given in Section 2. Words indexed by linear orderings are introduced in Section 3. Rational expressions denoting sets of such words are defined in Section 4. Automata accepting words on linear orderings are introduced in Section 5. Finally, Section 6 is devoted to the proof of the equivalence between rational expressions and automata.

Part of the results of the present paper has been presented at the conference MFCS'2001 [4].

2 Linear orderings

In this paper, we consider words on linear orderings, that is, totally ordered sequences of letters. We are only interested in countable linear orderings. Furthermore, we focus on scattered linear orderings. In this section, we provide the needed material on orderings. We recall the definitions and we fix the terminology. We refer the reader to [21] for a complete introduction to linear orderings.

A *linear ordering* J is a set equipped with an ordering $<$ which is total, that is, for all $j \neq k$ in J, either $j < k$ or $k < j$ holds. The *order type* of an ordering J is the class of all orderings isomorphic to J. The order types of $\mathbb{N}$, $\mathbb{Z}$ and $\mathbb{Q}$ under the usual ordering are respectively denoted by ω, ζ and η. For every non-negative integer n, the order type of an ordering with n elements is denoted by $\mathbf{n}$.

Recall that an *ordinal* is a linear ordering which is well-ordered. This means that any nonempty subset has a least element. In this paper, we only consider countable ordinals, that is, ordinals less than the ordinal ω_1.

Given a linear ordering J, we denote by $-J$ the backwards linear ordering obtained by reversing the ordering relation. Suppose that the ordering relation of J is denoted by the symbol $<$. The ordering $-J$ has the same underlying set but its ordering relation $<^*$ is defined by $j <^* k$ if and only if $k < j$. For instance, the ordering $-\mathbb{N}$ is the ordering $\ldots, 3, 2, 1, 0$. Its order type is denoted by $-\omega$.

Example 1 Let $\mathbb{Z}^2$ be the set of all pairs (k, j) of integers. Define the relation $(k_1, j_1) < (k_2, j_2)$ by $j_1 < j_2$ or $j_1 = j_2$ and $k_1 < k_2$. The relation $<$ linearly orders $\mathbb{Z}^2$.

Example 2 Let $\mathbb{Z}^\omega$ be the set of all sequences $(j_n)_{n\geq 0}$ of integers in which there are finitely many n such that $j_n \neq 0$. Define the relation $(j_n)_{n\geq 0} < (k_n)_{n\geq 0}$ if and only if $j_m < k_m$ where m is the greatest integer such that $j_m \neq k_m$. The relation $<$ endows $\mathbb{Z}^\omega$ with a linear ordering. It can be proved that if m is taken as the least integer such that $j_m \neq k_m$, one obtains an ordering isomorphic to $\mathbb{Q}$.

Two elements j and k of a linear ordering J are called *consecutive* if $j < k$ and if there is no element $i \in J$ such that $j < i < k$. The element j is then called the *predecessor* of k and k is called the *successor* of j. Note that an element k has a predecessor if and only if the set $\{j \mid j < k\}$ has a greatest element.

Let K be a subset of a linear ordering J. If the least or the greatest element of K exists, it is denoted by $\min(K)$ or $\max(K)$, and K is then said to be *left closed* or *right closed*. Otherwise, it is said to be *left open* or *right open*. If the the greatest lower bound or the least upper bound of K exists, it is denoted by $\inf(K)$ or $\sup(K)$. Note that if K is left closed, its least element is also the greatest lower bound of K, that is, $\inf(K) = \min(K)$. An ordering J is *complete* if any subset K which is lower bounded has a greatest lower bound and if any subset K which is upper bounded has a least upper bound.

An ordering K is a *subordering* of an ordering J if K is a subset of J and if the order of K is the restriction to K of the order of J. An *interval* of a linear ordering J is a subset K of J such that for all $j, k \in K$ and $i \in J$, the relation $j < i < k$ implies that i belongs to K.

Let J and K be two linear orderings. The linear ordering $J + K$ is the ordering obtained by juxtaposition of J and K. More formally, it is the linear ordering on the disjoint union $J \cup K$ extended with $j < k$ for all $j \in J$ and $k \in K$. For instance, if J and K are the orderings $-\mathbb{N}$ and $\mathbb{N}$, the ordering $J + K$ is isomorphic to $\mathbb{Z}$.

More generally, let J be a linear ordering and let K_j be a linear ordering for all $j \in J$. The linear ordering $\sum_{j \in J} K_j$ is the set of all pairs (k, j) such that $k \in K_j$. The relation $(k_1, j_1) < (k_2, j_2)$ holds if and only if $j_1 < j_2$ or $j_1 = j_2$ and $k_1 < k_2$ in K_{j_1}. For instance, if $J = \mathbb{Z}$ and $K_j = \mathbb{Z}$ for $j \in \mathbb{Z}$, the sum $\sum_{j \in J} K_j$ is isomorphic to the ordering $\mathbb{Z}^2$ already considered in Example 1.

2.1 Cuts

A *Dedekind cut* or simply a *cut* of a linear ordering J is a pair (K, L) of intervals such that $J = K \cup L$ and such that for all $k \in K$ and $l \in L$, $k < l$. Note that K and L are disjoint — in fact they form a partition of J. The set of all cuts of the ordering J is denoted by $\hat{J}$.

Example 3 Let J be the ordering $\{1, 2, 3\}$. The set $\hat{J}$ contains the four cuts $(\varnothing, \{1, 2, 3\})$, $(\{1\}, \{2, 3\})$, $(\{1, 2\}, \{3\})$ and $(\{1, 2, 3\}, \varnothing)$. More generally, if J contains n elements, $\hat{J}$ contains $n + 1$ cuts.

The two trivial cuts $(\varnothing, J)$ and $(J, \varnothing)$ are usually not considered as cuts in the literature [21] but they are essential for our purpose and they are included in the set $\hat{J}$. However, it is sometimes convenient to ignore these two cuts and we denote by $\hat{J}^*$ the set $\hat{J} - \{(\varnothing, J), (J, \varnothing)\}$ of non-trivial cuts.

The set $\hat{J}$ can be linearly ordered as follows. For all cuts $c_1 = (K_1, L_1)$ and $c_2 = (K_2, L_2)$, define the relation $c_1 < c_2$ if and only if $K_1 \subsetneq K_2$. This inclusion implies $L_1 \supsetneq L_2$ and the definition is therefore symmetric. The cuts $(\varnothing, J)$ and $(J, \varnothing)$ are the least and the greatest element of $\hat{J}$. They are respectively called the *first* and the *last cut* of J. We denote them by $c_{\min}$ and $c_{\max}$.

Example 4 Let J be the ordering $\mathbb{N}$. The set $\hat{J}$ contains the cut $(\{0,\ldots,n-1\},\{n,n+1,\ldots\})$ for all integer n and the last cut $(\mathbb{N},\varnothing)$. The ordering $\hat{J}$ is thus the ordering $\mathbb{N}+1$ of order type $\omega+1$.

More generally, if J is an ordinal α, the ordering $\hat{J}$ is the ordinal $\alpha+1$. Indeed, for every cut $c=(K,L)$ different from the last one, the interval L has a least element j_c and the function which maps the cut c to j_c is one-to-one.

From the previous examples, it may seem that the ordering $\hat{J}$ is close to the ordering J. The following example shows that this is not always true.

Example 5 Let J be the ordering $\mathbb{Q}$. The set $\hat{J}$ is not countable since each irrational number defines a unique cut of $\mathbb{Q}$.

More generally, if J contains a subordering of order type η, then $\hat{J}$ is not countable. Therefore, we will only consider countable scattered linear orderings as defined in the next section.

For all J, the ordering $\hat{J}$ is complete and we often use this fact in the sequel. Indeed, the least upper bound of a subset M of $\hat{J}$ is the cut (K_0,L_0) where the intervals K_0 and L_0 are defined by

$$K_0=\bigcup_{(K,L)\in M} K \quad\text{and}\quad L_0=\bigcap_{(K,L)\in M} L.$$

The greatest lower bound of M is given analogously.

2.2 Scattered linear orderings

A linear ordering J is said to be *dense* if it contains at least two elements and if for all $i<k$ in J, there is $j\in J$ such that $i<j<k$. A linear ordering is *scattered* if it contains no dense subordering. Thus scattered ordering are those that do not contain a subordering of order type η. The following characterization of countable scattered linear orderings is due to Hausdorff.

Theorem 6 (Hausdorff) *A countable linear ordering J is scattered if and only if J belongs to $\bigcup_{\alpha<\omega_1} V_\alpha$ where the classes V_α are inductively defined by*

1. $V_0=\{J \mid J$ *of order type* $\mathbf{0}$ *or* $\mathbf{1}\}$*;*
2. $V_\alpha=\{\sum_{j\in J} K_j \mid J$ *of order type* $\mathbf{n}$, ω, $-\omega$, *or* ζ *and* $K_j\in\bigcup_{\beta<\alpha} V_\beta\}$.

The orderings $\mathbb{Z}^2$ and $\mathbb{Z}^\omega$ considered in Examples 1 and 2 are scattered. They respectively belong to the classes V_2 and V_ω defined above. In the sequel, we denote by $\mathcal{S}$ the class of all countable scattered linear orderings. It follows from the previous theorem that if J is a countable scattered linear ordering, then $\hat{J}$ is also a countable scattered linear ordering. Conversely, if $\hat{J}$ is countable, then J is countable and scattered.

2.3 The ordering $J \cup \hat{J}$

The orderings of J and $\hat{J}$ can be extended to an ordering on the disjoint union $J \cup \hat{J}$ as follows. This means that $J \cup \hat{J}$ can be endowed with a linear ordering such that J and $\hat{J}$ are then two of its suborderings. For $j \in J$ and a cut $c = (K, L)$, define the relations $j < c$ and $c < j$ by respectively $j \in K$ and $j \in L$. Note that exactly one of these two relations holds since (K, L) is a partition of J. These relations together with the orderings of J and $\hat{J}$ endow $J \cup \hat{J}$ with a linear ordering. This ordering will be used in Section 4. Notice that for two consecutive elements $j_1 < j_2$ of J, there is exactly one cut c such that $j_1 < c < j_2$. Analogously, for two cuts $c_1 < c_2$ that are consecutive in $\hat{J}$, there is exactly one element $j \in J$ such that $c_1 < j < c_2$. The elements of J and the cuts of J interleave.

$$\Big| \;\cdots\; \Big| \;\cdots\; |\bullet|\bullet|\bullet\cdots \Big| \;\cdots\; |\bullet\underset{c_j^-}{|}\overset{j}{\bullet}\underset{c_j^+}{|}\bullet\cdots \Big| \;\cdots\; |\bullet|\bullet|\bullet \;\;\cdots \Big| \;\cdots\; \Big|$$

Figure 1: Ordering $J \cup \hat{J}$ for $J = \mathbb{Z}^2$

Example 7 The ordering $J = \mathbb{Z}^2$ is equal to the sum $\sum_{j \in \mathbb{Z}} \mathbb{Z}$ introduced in Example 1. There is a cut between each pair of consecutive elements in each copy of $\mathbb{Z}$, but there is also a cut between consecutive copies of $\mathbb{Z}$. There are also the first and the last cuts. The ordering $J \cup \hat{J}$ is pictured in Figure 1 where each element of J is represented by a bullet and each cut by a vertical bar.

For every element j, there are two consecutive cuts c_j^- and c_j^+ such that $c_j^- < j < c_j^+$. They are defined as $c_j^- = (K, \{j\} \cup L)$ and $c_j^+ = (K \cup \{j\}, L)$ with $K = \{k \mid k < j\}$ and $L = \{k \mid j < k\}$. Thus there is then a one-to-one correspondence between elements of J and pairs of consecutive cuts in $\hat{J}$. Notice that a cut $c = (K, L)$ has a predecessor if and only if K has a greatest element j. In this case the cut c is equal to c_j^+ and its predecessor (in $\hat{J}$) is c_j^-. The cut $c = (K, L)$ has a successor if and only if L has a least element j. It is then equal to c_j^- and its successor (in $\hat{J}$) is c_j^+.

The first and last cut of J are respectively the least and the greatest element of the ordering $J \cup \hat{J}$. Thus the ordering $J \cup \hat{J}^*$ is the one obtained by removing the first and the last cut from $J \cup \hat{J}$. By Theorem 6, it can be shown by a straightforward induction on the rank that if J is a countable scattered ordering, then both orderings $J \cup \hat{J}$ and $J \cup \hat{J}^*$ are also countable and scattered.

Both orderings $J \cup \hat{J}$ and $J \cup \hat{J}^*$ are complete. The following lemma gives a characterization of the ordering $J \cup \hat{J}$. It is needed in Section 6.2.

Lemma 8 *Let K be a complete scattered linear ordering with a least and a greatest element. Let J and J' be two suborderings such that $J \cap J' = \varnothing$ and $K = J \cup J'$. Suppose that for consecutive elements k and k' of K, either $k \in J$*

and $k' \in J'$ or $k \in J'$ and $k' \in J$. Suppose also that any element of J has a predecessor and a successor in J'. Then J' is isomorphic to $\hat{J}$ and K is isomorphic to $J \cup \hat{J}$.

Proof We define a function f from K into $J \cup \hat{J}$ as follows. For any $k \in K$, define

$$f(k) = \begin{cases} k & \text{if } k \in J \\ \big(\{j \in J \mid j < k\}, \{j \in J \mid k < j\}\big) & \text{if } k \in J'. \end{cases}$$

Since $J \cap J' = \varnothing$ and $K = J \cup J'$, the function f is well defined. The restriction of f to J is the identity. The image of an element of J' is a cut of J. Therefore f is a function from K into $J \cup \hat{J}$.

We claim that the function f is one-to-one. Thus we must show that $k \neq k'$ implies $f(k) \neq f(k')$. If $k \in J$ or $k' \in J$, the result is trivial. Suppose then that $k, k' \in J'$ and that $k < k'$. Since K is scattered, there are two consecutive elements $j < j'$ of K such that $k \leq j < j' \leq k'$. Since j and j' are consecutive, either j belongs to J or j' belongs to J. Therefore the two cuts $f(k)$ and $f(k')$ are different.

We now prove that the function f is onto. It is clear that $J \subseteq f(K)$. Let (L, M) be a cut of J. We claim that there is $k \in J'$ such that $(L, M) = f(k)$. Since K is complete and has a least and a greatest element, any subset of K has a greatest lower bound and a least upper bound. Define the two elements l and m of K by $l = \sup(L)$ and $m = \inf(M)$. If l belongs to L, it has a successor k in J' and one has $(L, M) = f(k)$. If m belongs to M, it has a predecessor k in J' and one has $(L, M) = f(k)$. If l and m do not belong to L and M, they belong to J' and their image by f is the cut (L, M). $\square$

2.4 Condensation

The following notion is needed in Section 6.2. Let J be a linear ordering. A *condensation* of J is an equivalence relation $\sim$ on J such that each of its classes is an interval. The ordering of J induces a linear ordering of the quotient $J/\sim$.

The quotient $J/\sim$ inherits some properties of J. If J has a least or a greatest element, then $J/\sim$ also has a least or a greatest element which is the class of the least or greatest element of J. If J is complete, then $J/\sim$ is also complete and if J is scattered, then $J/\sim$ is also scattered.

We mention here some useful properties of consecutive classes k and k' of $J/\sim$ when the ordering J is complete. Define j and j' by $j = \sup(k)$ and $j' = \inf(k')$. By definition, one has $j \leq j'$. If $j < j'$, then j and j' respectively belong to k and k' since the classes k and k' are consecutive. In that case, the interval k is right closed and k' is left closed. If $j = j'$, then j belongs to either k or k'. In the former case, k is right closed and k' is left open and in the latter case, k is right open and k' is left closed. Note that it is impossible that k is right open and k' is left open.

In the sequel, we consider only linear orderings which are *countable* and *scattered.* This restriction is needed in the proof of the main result (Theorem 20) which states that rational expressions and automata are equivalent. The definition of the rational operations given in Section 4 also depends on this restriction. The ordinal iterations in particular are restricted to countable ordinals. However, the definition of automata that we introduce in Section 5 makes sense for all linear orderings.

3 Words on linear orderings

Let A be a finite alphabet whose elements are called letters. For a linear ordering J, a *word* of *length* J over A is a function which maps each element of J to a letter of A. A word on a linear ordering J can be viewed as a labeled ordering where each point of J has been decorated by a letter. A word is denoted in a sequence-like notation by $(a_j)_{j\in J}$ where $a_j \in A$ is the image of the element j. The word whose length is the empty set is called the *empty word* and it is denoted by ε.

The notion of word just introduced generalizes notions common in the literature. If the ordering J is finite with n elements, a word of length J is a finite sequence $a_1 \dots a_n$ [19]. A word of length $\mathbb{N}$ is a sequence $a_0a_1a_2\dots$ which is usually called an ω-*word* or an *infinite* word [24]. A word of length $\mathbb{Z}$ is a sequence $\dots a_{-2}a_{-1}a_0a_1a_2\dots$ which is usually called a *bi-infinite* word. An ordinal word is a word indexed by a countable ordinal.

Example 9 Recall that the ordering $\mathbb{N} + (-\mathbb{N})$ is the ordering of $\{(n,0) \mid n \geq 0\} \cup \{(-n,1) \mid n \geq 0\}$ with $(k_1,j_1) < (k_2,j_2)$ if and only if $j_1 < j_2$ or $j_1 = j_2$ and $k_1 < k_2$. The word $x = b^\omega(ab)^{-\omega}$ is the word of length $\mathbb{N}+(-\mathbb{N})$ defined by

$$x_{k,j} = \begin{cases} a & \text{if } j = 1 \text{ and } k \text{ odd} \\ b & \text{otherwise.} \end{cases}$$

Let $x = (a_j)_{j\in J}$ and $y = (b_k)_{k\in K}$ be two words of length J and K. The *product* (or the *concatenation*) of x and y is the word $z = (c_l)_{l\in J+K}$ of length $J + K$ such that

$$c_l = \begin{cases} a_l & \text{if } l \in J \\ b_l & \text{if } l \in K \end{cases}$$

The product of x and y is denoted xy. More generally, let J be a linear ordering and for each $j \in J$, let x_j be a word of length K_j. The *product* $\prod_{j\in J} x_j$ is the word z of length $K = \sum_{j\in J} K_j$ defined as follows. Suppose that each word x_j is equal to $(a_{k,j})_{k\in K_j}$ and recall that K is the set of all pairs (k,j) such that $k \in K_j$. The product z is then equal to $(a_{k,j})_{(k,j)\in K}$. This definition can be extended to sets. Given sets $(X_j)_{j\in J}$, the *product* $\Pi_{j\in J}X_j$ is the set $\{\Pi_{j\in J}x_j \mid x_j \in X_j\}$. When J is finite or isomorphic to ω, the product is called a finite product or an ω-product.

Two words $x = (a_j)_{j\in J}$ and $y = (b_k)_{k\in K}$ of length J and K are *isomorphic* if there is an order-preserving isomorphism f from J to K such that $a_j = b_{f(j)}$

for all j in J. This obviously defines an equivalence relation on words. A class of words up to isomorphism is called a *type*. In this paper, we identify isomorphic words. This makes sense since the automata and rational expressions we introduce do not distinguish isomorphic words: an automaton $\mathcal{A}$ accepting a word x, for example, also accepts any word isomorphic to x. Since isomorphic words are identified, the length of word is more an order type than an ordering. For instance, a word of length $\mathbb{N}$ (or $\mathbb{Z}$), is rather called a word of length ω (or ζ).

Note that some orderings, such as $\mathbb{Z}$, have non trivial internal isomorphisms. This induces some unexpected isomorphisms between words. For instance, let x and y be the two words $(a_j)_{j\in\mathbb{Z}}$ and $(b_j)_{j\in\mathbb{Z}}$ defined by

$$a_j = \begin{cases} a & \text{if } j = 0 \\ b & \text{otherwise} \end{cases} \qquad \text{and} \qquad b_j = \begin{cases} a & \text{if } j = 1 \\ b & \text{otherwise.} \end{cases}$$

The two words x and y are isomorphic since the function f given by $f(x) = x+1$ is an isomorphism from $\mathbb{Z}$ to $\mathbb{Z}$.

Example 10 For $j \in \mathbb{Z}$, define the word x_j by $x_j = b^{-\omega}$ if j is even and by $x_j = ab^{\omega}$ if j is odd. The product $\prod_{j\in J} x_j$ is the word $(b^{-\omega}ab^{\omega})^{\zeta}$ of length $\mathbb{Z}^2$.

4 Rational expressions

In this section, the rational operations used to define rational sets of words on scattered linear orderings are introduced. These *rational operations* include the usual Kleene operations for finite words—union, concatenation, and finite iteration (though these must be appropriately extended to our notion of word). They also include omega iteration (usually used to construct ω-words) and the ordinal iteration introduced by Wojciechowski [25] for ordinal words. Three new operations are also needed: backwards omega iteration, backwards ordinal iteration, and an additional binary operation, a kind of general iteration for scattered linear orderings.

Union, concatenation and finite iteration are denoted as usual by the symbols $+$, $\cdot$ and $*$. Omega iteration and ordinal iteration are denoted by the symbols ω and $\sharp$ whereas backwards omega iteration and backwards ordinal iteration are denoted by the symbols $-\omega$ and $-\sharp$. Iteration for all scattered linear orderings is denoted by the symbol $\diamond$.

We first define the various sorts of iteration in a unified framework. Given a set X of words and a class $\mathcal{J}$ of linear orderings, define iteration $X^{\mathcal{J}}$ of X with respect to $\mathcal{J}$ by

$$X^{\mathcal{J}} = \{\prod_{j\in J} x_j \mid J \in \mathcal{J} \text{ and } x_j \in X\}.$$

The sets X^*, X^{ω}, $X^{-\omega}$, $X^{\sharp}$ and $X^{-\sharp}$ are then respectively equal to $X^{\mathcal{J}}$ for $\mathcal{J}$ equal to the class of all finite linear orderings, the class $\{\omega\}$ which only contains the ordering ω, the class $\{-\omega\}$ which only contains the ordering $-\omega$, the class of all countable ordinals and the class $\{-\alpha \mid \alpha < \omega_1\}$ of all countable backwards ordinals.

We now define the binary operations. Let X and Y be two sets of words. The sets $X+Y$, $X \cdot Y$ and $X \diamond Y$ are defined by

$$X + Y = X \cup Y, \qquad X \cdot Y = \{xy \mid x \in X \text{ and } y \in Y\} \qquad \text{and}$$
$$X \diamond Y = \{ \prod_{j \in J \cup \hat{J}^*} z_j \mid J \in \mathcal{S} \setminus \{0\} \text{ and } z_j \in X \text{ if } j \in J \text{ and } z_j \in Y \text{ if } j \in \hat{J}^* \}.$$

A word x belongs to $X \diamond Y$ if and only if there is a non empty countable scattered linear ordering J such that x is the product of a sequence of length $J \cup \hat{J}^*$ of words where each word indexed by an element of J belongs to X and each word indexed by a cut in $\hat{J}^*$ belongs to Y. Note that if the empty word ε does not belong to X, it does not belong to $X \diamond Y$.

An abstract *rational expression over* A is a well-formed term of the free algebra over $\{\varnothing\} \cup A$ with the symbols denoting the rational operations as function symbols. A set of words is *rational* if it can be denoted by a rational expression. As usual, the dot denoting concatenation omitted in rational expressions. Thus, the expression $a \cdot (a + a \cdot b^\omega)^\sharp$ is written $a(a+ab^\omega)^\sharp$. We also use the following abbreviations. If E is a rational expression, the expressions ε, E^ζ and $E^\diamond$ respectively abbreviate $\varnothing^*$, $E^{-\omega}E^\omega$ and $E \diamond \varepsilon$. If the alphabet A is the set $\{a_1, \dots, a_n\}$, the symbol A is also used as an abbreviation for the expression $a_1 + \cdots + a_n$.

Example 11 The expressions $A^\diamond + \varepsilon$ and $(A^\diamond + \varepsilon)a(A^\diamond + \varepsilon)$ denote respectively the set of all words and the set of words having an occurrence of the letter a.

Example 12 The expression A^* denotes the set of finite words and the expression $(A^\diamond)^\omega(A^\diamond + \varepsilon) + (A^\diamond + \varepsilon)(A^\diamond)^{-\omega}$ denotes its complement. Indeed, a linear ordering J is not finite if and only if it has some cut (K, L) such that either K does not have a greatest element or L does not have a least element. The rational expression $(A^\diamond)^\omega$ denotes the set of words whose length does not have a last element. Therefore, the expression $(A^\diamond)^\omega(A^\diamond + \varepsilon)$ denotes the set whose length has a cut (K, L) such that K does not have a greatest element. Symmetrically, the expression $(A^\diamond + \varepsilon)(A^\diamond)^{-\omega}$ denotes the set whose length has a cut (K, L) such that L does not have a least element.

In the sequel, we will often use the following lemma.

Lemma 13 *The class of rational sets over an alphabet A is closed under the rational operations.*

Proof It is clear that if the sets X and Y are respectively denoted by the expressions E and F, then the sets $X+Y$, $X \cdot Y$, X^*, X^ω, $X^{-\omega}$, $X^\sharp$, $X^{-\sharp}$ and $X \diamond Y$ are respectively denoted by the expressions $E+F$, $E \cdot F$, E^*, E^ω, $E^{-\omega}$, $E^\sharp$, $E^{-\sharp}$ and $E \diamond F$. $\square$

The previous lemma has the following corollary. Let $X_1, \ldots, X_n$ be sets of words over an alphabet A. Any word w over the alphabet $\Sigma = \{X_1, \ldots, X_n\}$ (where $X_1, \ldots, X_n$ are now seen as letters) gives a set X_w of words over A by interpretation of each letter X_i by the corresponding set X_i. Furthermore, each set W of words over the alphabet Σ gives the set $X_W = \bigcup_{w \in W} X_w$ of words over A by interpretation of each X_i.

Corollary 14 *If each set X_i is rational and if W is rational, then the set X_W is also rational.*

Then, we say that the set X_W *has a rational expression over the sets X_i*. The proof of the corollary is an easy induction on the rational expression denoting the set W. A rational expression denoting the set X_W can be obtained by substituting each letter X_i by an expression denoting the set X_i in the expression denoting W.

5 Automata

In this section, automata on words on linear orderings are defined. As in the rest of the paper, linear orderings considered in this section are countable and scattered. However, it is worth pointing out that our definition of automata is suitable for all linear orderings. Automata that we define are a natural generalization of Büchi automata [8] on ordinal words. Automata introduced by Büchi are usual (Kleene) automata with left limit transitions of the form $P \to p$ used for limit ordinals. The automata that we introduce have limit transitions of the form $p \to P$ as well.

Definition 15 *Let A be a finite alphabet. An* automaton $\mathcal{A}$ *over A is a 4-tuple (Q, E, I, F) where Q is a finite set of states, $E \subseteq (Q \times A \times Q) \cup (\mathcal{P}(Q) \times Q) \cup (Q \times \mathcal{P}(Q))$ is the set of transitions, $I \subseteq Q$ is the set of initial states and $F \subseteq Q$ is the set of final states.*

Since the alphabet and the set of states are finite, the set of transitions is also finite. Transitions are either of the form (p, a, q) or of the form (P, q) or of the form (q, P) where P is a subset of Q. A transition of the former case is called a *successor transition* and it is denoted by $p \xrightarrow{a} q$. A transition of the two latter cases are respectively called a *left limit* or a *right limit transition* and they are denoted by $P \to q$ and $q \to P$. Before explaining what a path is and how an automaton accepts words, we illustrate this definition with the following example.

Example 16 The automaton pictured in Figure 2 has 3 successor transitions which are pictured as labeled edges of a graph. It also has a left limit transition $\{2\} \to 0$ and a right limit transition $0 \to \{1\}$. The state 0 is the only initial state and the only final state.

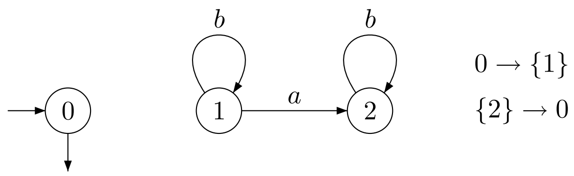


Figure 2: Automaton of Example 16

We say that a transition *leaves* a state q if it is either a successor transition $q \xrightarrow{a} p$ for some state p or a right limit transition $q \to P$ for some subset P of states. We say that it *enters* a state q if it is either a successor transition $p \xrightarrow{a} q$ or a left limit transition $P \to q$. The sets of transitions leaving and entering a state q are respectively denoted by $\mathrm{Out}(q)$ and $\mathrm{In}(q)$. More generally, for a set P of states, we define $\mathrm{Out}(P) = \bigcup_{q \in P} \mathrm{Out}(q)$ and $\mathrm{In}(P) = \bigcup_{q \in P} \mathrm{In}(q)$.

In order to define the notion of path in such an automaton, the following notion of limits is needed. We define it for an arbitrary linear ordering J but we use it when the considered ordering is actually the ordering $\hat{J}$ of cuts of a given ordering J. Let Q be a finite set, let J be a linear ordering and let $\gamma = (q_j)_{j \in J}$ be a word over Q. Let j be a fixed element of J. The *left limit set* and *right limit set* of γ at j are the two subsets $\lim_{j^-} \gamma$ and $\lim_{j^+} \gamma$ of Q defined as follows.

$$\lim_{j^-} \gamma = \{q \in Q \mid \forall k < j \ \exists i \quad k < i < j \text{ and } q = q_i\},$$
$$\lim_{j^+} \gamma = \{q \in Q \mid \forall j < k \ \exists i \quad j < i < k \text{ and } q = q_i\}.$$

Thus the left limit at j is the set of states to the left that occur infinitely often infinitely close to j. Note that if j has a predecessor, the limit set $\lim_{j^-} \gamma$ is empty. Conversely, if j has no predecessor and is not the least element of J, $\lim_{j^-} \gamma$ is nonempty since the set Q is finite. Similar results hold for right limit sets. For $c = \min(J)$, we define $\lim_{c^-} \gamma = \varnothing$ and for $c = \max(J)$, we define $\lim_{c^+} \gamma = \varnothing$ (although these two limits are literally defined as equal to Q by the above formulas).

We now come to the definition of a path in an automaton on linear orderings. Roughly speaking, a path is a labeling of each cut of the ordering by a state of the automaton such that local properties are satisfied.

Definition 17 *Let $\mathcal{A}$ be an automaton and let $x = (a_j)_{j \in J}$ be a word of length J. A* path *γ labeled by x is a sequence of states $\gamma = (q_c)_{c \in \hat{J}}$ of length $\hat{J}$ such that*

- *For consecutive cuts c_j^- and c_j^+, $q_{c_j^-} \xrightarrow{a_j} q_{c_j^+}$ is a successor transition.*
- *For any cut c which is not the first cut and which has no predecessor, $\lim_{c^-} \gamma \to q_c$ is a left limit transition.*
- *For each cut c which is not the last cut and which has no successor, $q_c \to \lim_{c^+} \gamma$ is a right limit transition.*

It is worth pointing out that the length of a path labeled by a word x of length J is the ordering $\hat{J}$ of the cuts of J. Since a sequence of states indexed by $\hat{J}$ is actually a function from $\hat{J}$ into Q, we sometimes use a functional notation and the state q_c of a path γ is also denoted by $\gamma(c)$.

By the previous definition, there is a transition entering the state $\gamma(c)$ for each cut c which is not the first cut. This transition is a successor transition if the cut c has a predecessor in $\hat{J}$ and it is a left limit transition otherwise. Similarly, there is a transition leaving $\gamma(c)$ for every cut c which is not the last cut.

Since the ordering $\hat{J}$ has a least and a greatest element, a path always has a first and a last state, which are the images of the first and the last cut. A path is *successful* if its first state is initial and its last state is final. A word is *accepted* or *recognized* by the automaton if it is the label of a successful path. A set of words is *recognizable* if it is the set of words accepted by some automaton.

The notion of path we have introduced for words on orderings coincide with the usual notion of paths considered in the literature for finite words, ω-words and ordinal words. Let x be a finite word $a_1 \dots a_n$. The set of cuts of the finite ordering $\{1, \dots, n\}$ can be identified with $\{1, \dots, n+1\}$ (see Example 3). In our setting, a path labeled by x is then a finite sequence $q_1, \dots, q_{n+1}$ of states such that $q_j \xrightarrow{a_j} q_{j+1}$ is a successor transition for all j in $\{1, \dots, n\}$. This matches the usual definition of a finite path in an automaton [19, p. 5].

Let $x = a_0a_1a_2 \dots$ be an ω-word. The set of cuts of the ordering $J = \omega$ is the ordinal $\omega + 1 = \{0, 1, 2, \dots, \omega\}$ (see Example 4). The pairs of consecutive cuts are the pairs $(j, j+1)$ for $j < \omega$ whereas the cut $c = \omega$ has no predecessor. In our setting, a path γ labeled by x is a sequence $q_0, q_1, q_2, \dots, q_\omega$ of states such that $q_j \xrightarrow{a_j} q_{j+1}$ is a successor transition for all $j < \omega$ and such that $\lim_{\omega^-} \gamma \to q_\omega$ is a left limit transition. Note that $\lim_{\omega^-} \gamma$ is the set of states which occur infinitely many times in γ. This path is successful if and only if q_0 is initial and q_ω is final. Define the family $\mathcal{T}$ of sets of states by

$$\mathcal{T} = \{P \mid \exists q \in F \text{ such that } P \to q \in E\}.$$

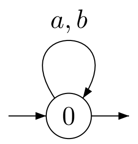
The path γ is then successful if and only if q_0 is initial and the set $\lim_{\omega^-} \gamma$ of states belongs to the family $\mathcal{T}$. This matches the definition of a successful path in a Muller automaton [24, p. 148].

The set of cuts of an ordinal α is the ordinal $\alpha + 1$. Therefore, the notion of path we have introduced coincide for ordinal words with the notion of path considered in [2].

$$\begin{array}{ccccccccccccccccc}
0 & & 1 & 1 & 1 & 2 & 2 & 2 & & 0 & & 1 & 1 & 1 & 2 & 2 & 2 & & 0 \\
\big| & \cdots & | & b & | & b & | & a & | & b & | & b & | & \cdots & \big| & \cdots & | & b & | & b & | & a & | & b & | & b & | & \cdots & \big|
\end{array}$$

Figure 3: A path labeled by $(b^{-\omega}ab^{\omega})^2$

Example 18 Consider the automaton $\mathcal{A}$ of Figure 2 and let x be the word $(b^{-\omega}ab^{\omega})^2$ of length $\mathbb{Z}+\mathbb{Z}$. A successful path γ labeled by x is pictured in Figure 3. This path is made of two copies of the path $01^{-\omega}2^{\omega}0$. A path for the word $(b^{-\omega}ab^{\omega})^{\zeta}$ cannot be made by $\mathbb{Z}$ copies of $01^{-\omega}2^{\omega}0$ because the right limit set of the first state would be $\{0,1,2\}$ and the automaton has no transition of the form $q \to \{0,1,2\}$. The automaton $\mathcal{A}$ recognizes the set $(b^{-\omega}ab^{\omega})^*$.



$\{0\} \to 0$

Figure 4: Automaton of Example 19

Example 19 Consider the automaton $\mathcal{A}$ pictured in Figure 4. This automaton has no right limit transition. It recognizes the words whose length is an ordinal since a linear ordering J is an ordinal if and only if each of its cuts except the last has a successor in $\hat{J}$. The automaton obtained by suppressing the left limit transition of $\mathcal{A}$ recognizes the set of finite words since a linear ordering J is finite if and only if each of its cuts except the first has a predecessor in $\hat{J}$ and each of its cuts except the last a successor in $\hat{J}$.

6 Rational expressions vs automata

In this section, we prove that rational expressions and automata are equivalent. This means that a set of words on linear orderings is rational if and only if it can be recognized by an automaton. This result extends Kleene's well-known theorem for finite words. Kleene's theorem was first extended to words of length ω by Büchi [6] and was later extended to words of ordinal length by Wojciechowski [25]. We have then the following theorem, which is the main result.

Theorem 20 *A set of words on countable scattered linear orderings is rational if and only if it is recognizable.*

The following two examples illustrate the theorem.

Example 21 The automaton pictured in Figure 5 recognizes the set denoted by the rational expression $(A^{\omega})^{-\omega}$. The part of the automaton given by state 1 and the left limit transition $\{1\} \to 2$ recognizes the set A^{ω}. The successor transition from state 2 to state 1 allows the concatenation of two words of A^{ω}. The right limit transition $0 \to \{1,2\}$ leads to a sequence of length $-\omega$ of words of A^{ω}.

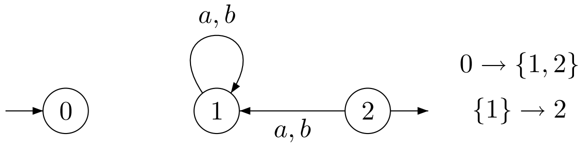


Figure 5: Automaton recognizing $(A^\omega)^{-\omega}$

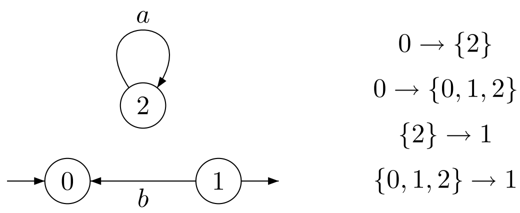


Figure 6: Automaton recognizing $a^\zeta \diamond b$

Example 22 The automaton pictured in Figure 6 recognizes the set denoted by the rational expression $a^\zeta \diamond b$. The part of the automaton given by state 2 and the two limit transitions $0 \to \{2\}$ and $\{2\} \to 1$ accepts the word a^ζ whereas the part given by the successor transition from state 1 to state 0 accepts the word b. Each occurrence of a^ζ (except the first) is preceded by an occurrence of b in the automaton, and each occurrence of a^ζ (except the last) is followed by an occurrence of b. More generally, thanks to the limit transitions $0 \to \{0, 1, 2\}$ and $\{0, 1, 2\} \to 1$, the occurrences of a^ζ are indexed by a linear ordering $J \in \mathcal{S} \setminus \{\varnothing\}$, the occurrences of b are indexed by the ordering $\hat{J}^*$ and they are interleaved according to the ordering $J \cup \hat{J}^*$.

The next two sections are devoted to the proof of the theorem. The first section contains the proof that a rational set is recognizable and the second section contains the proof of the converse.

6.1 From rational expressions to automata

In this section, we prove that any rational set of words is recognized by an automaton. The proof is by induction on the rational expression denoting the set. For each rational operation, we describe a corresponding construction for the automata. The constructions for union, concatenation, and finite iteration are very similar to the classical ones for automata on finite words [19, p. 15].

In order to combine two automata, it is convenient if no transition enters an initial state or leaves a final state. Furthermore, some problems may come from the empty word. To overcome them, any set X may be decomposed as $X = X' + \delta(X)$, where X' is $X - \varepsilon$ and $\delta(X)$ is ε if $\varepsilon \in X$ or $\varnothing$ otherwise.

An automaton is said to be *normalized* if it has a unique initial state i and a unique final state $f \neq i$ and it has no transition which enters i or leaves f. Note that these conditions imply that the states i and f can only occur as the first state and the last state of a path. Therefore transitions of the form $P \to q$ or $q \to P$ where P contains i or f cannot occur in a path. In the sequel, we assume that a normalized automaton does not have transitions of the form $P \to q$ or $q \to P$ where P contains i or f.

The following lemma states that the empty word can be added or removed without changing recognizability, and that a recognizable set which does not contain the empty word can be recognized by a normalized automaton. Note that this condition is necessary since a normalized automaton cannot accept the empty word.

Lemma 23 *Let X be a set of words. The set X is recognizable if and only if $X + \varepsilon$ is recognizable. Furthermore if $\varepsilon \notin X$, then X can be recognized by a normalized automaton.*

Proof Suppose that X is recognized by the automaton $\mathcal{A} = (Q, E, I, F)$. The set $X + \varepsilon$ is recognized by the automaton $\mathcal{A}'$ obtained by adding a new initial and final state i. Let Q' be the set $Q \cup \{i\}$ where i does not belong to Q. The automaton $\mathcal{A}'$ is equal to (Q', E', I', F') where $E' = E$, $I' = I \cup \{i\}$ and $F' = F \cup \{i\}$.

Conversely, suppose that the set $X + \varepsilon$ is recognized by the automaton $\mathcal{A} = (Q, E, I, F)$. Without loss of generality, we may assume $\varepsilon \notin X$. The set X is then recognized by the normalized automaton $\mathcal{A}'$ obtained by modifying $\mathcal{A}$ as follows. Let Q' be the set $Q \cup \{i, f\}$ where i and f are new states that do not belong to Q. Define the set E' of transitions by

$$\begin{aligned} E' = E &\cup \{i \xrightarrow{a} f \mid \exists\, p \xrightarrow{a} q \text{ with } p \in I \text{ and } q \in F\} \\ &\cup \{i \xrightarrow{a} q \mid \exists\, p \xrightarrow{a} q \text{ with } p \in I\} \cup \{i \to P \mid \exists\, p \to P \text{ with } p \in I\} \\ &\cup \{p \xrightarrow{a} f \mid \exists\, p \xrightarrow{a} q \text{ with } q \in F\} \cup \{P \to f \mid \exists\, P \to q \text{ with } q \in F\}. \end{aligned}$$

It is then straightforward to check that the automaton $\mathcal{A}' = (Q', E', \{i\}, \{f\})$ recognizes X. Indeed, a nonempty word x labels a successful path γ in $\mathcal{A}$ if and only if it labels the successful path γ' in $\mathcal{A}'$ obtained by replacing the first and the last state of γ by i and f. $\square$

We claim that if both sets X_1 and X_2 are recognizable, then the sets $X_1 + X_2$, X_1X_2, X_1^*, X_1^ω, $X_1^{-\omega}$, $X_1^\sharp$, $X_1^{-\sharp}$ and $X_1 \diamond X_2$ are also recognizable. For each of these rational operations, we describe a corresponding construction on automata.

Suppose that the sets X_1 and X_2 are recognized by the automata $\mathcal{A}_1 = (Q_1, E_1, I_1, F_1)$ and $\mathcal{A}_2 = (Q_2, E_2, I_2, F_2)$. Without loss of generality, we may assume that Q_1 and Q_2 are disjoint. The set $X_1 + X_2$ is then recognized by the automaton $\mathcal{A}$ obtained by mere juxtaposition of the two automata $\mathcal{A}_1$ and $\mathcal{A}_2$. This automaton $\mathcal{A}$ is equal to $\mathcal{A} = (Q, E, I, F)$ where $Q = Q_1 \cup Q_2$, $E = E_1 \cup E_2$,

$I = I_1 \cup I_2$ and $F = F_1 \cup F_2$. Note that this construction does not require that the automata recognizing X_1 and X_2 be normalized.

In order to define formally the constructions for the other rational operations, we introduce the following notation. Let $\mathcal{A}$ be an automaton (Q, E, I, F) and let q be one of its states. We denote by $E[p \leftarrow q]$ the set of transitions obtained by replacing by p each occurrence of q in each transition of E.

Figure 7: Automata $\mathcal{A}_1$ and $\mathcal{A}_2$

The constructions on automata corresponding to the other rational operations require that the automata be normalized. Define X_1' and X_2' by $X_1' = X_1 - \varepsilon$ and $X_2' = X_2 - \varepsilon$. The sets X_1 and X_2 are equal to $X_1' + \delta(X_1)$ and $X_2' + \delta(X_2)$. By Lemma 23, the sets X_1' and X_2' are recognized by two normalized automata $\mathcal{A}_1 = (Q_1, E_1, \{i_1\}, \{f_1\})$ and $\mathcal{A}_2 = (Q_2, E_2, \{i_2\}, \{f_2\})$ pictured in Figure 7. In the figure, all states which are neither initial nor final are represented by a squared box. Without loss of generality, we may assume that Q_1 and Q_2 are disjoint.

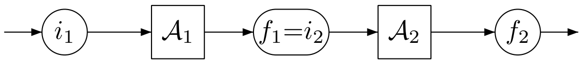


Figure 8: Automaton for $X_1'X_2'$

The set X_1X_2 is equal to $X_1'X_2' + \delta(X_1)X_2' + X_1'\delta(X_2) + \delta(X_1)\delta(X_2)$. Since the class of recognizable sets is closed under union, it suffices to construct an automaton for $X_1'X_2'$. This set is recognized by the automaton $\mathcal{A}$ obtained by first juxtaposing the automata $\mathcal{A}_1$ and $\mathcal{A}_2$ and then merging the final state f_1 of $\mathcal{A}_1$ and the initial state i_2 of $\mathcal{A}_2$ into a state called f_1 which is neither initial nor final. This construction is pictured in Figure 8. The automaton $\mathcal{A}$ is equal to $(Q, E, \{i_1\}, \{f_2\})$ where $Q = Q_1 \cup Q_2 - \{i_2\}$ and $E = E_1 \cup E_2[f_1 \leftarrow i_2]$. Suppose that the words x_1 and x_2 label the successful paths γ_1 and γ_2 in $\mathcal{A}_1$ and $\mathcal{A}_2$. Since i_2 is identified with f_1, the two paths γ_1 and γ_2 can be concatenated to form a successful path in $\mathcal{A}$ labeled by x_1x_2. Conversely, the state $f_1 = i_2$ occurs at least once in any successful path in $\mathcal{A}$ and it occurs exactly once since no transition leaves f_1 in $\mathcal{A}_1$ and no transition enters i_2 in $\mathcal{A}_2$. Therefore, any successful path in $\mathcal{A}$ is the concatenation of successful paths in $\mathcal{A}_1$ and $\mathcal{A}_2$. This proves that $\mathcal{A}$ recognizes the set $X_1'X_2'$.

The set $X_1^* = {X_1'}^*$ is recognized by the automaton $\mathcal{A}$ obtained by merging the initial state i_1 and final state f_1 of $\mathcal{A}_1$ into a state which is both initial and final. This construction is pictured in Figure 9. More formally, the automaton $\mathcal{A}$ is equal to $(Q, E, \{i_1\}, \{i_1\})$ where $Q = Q_1 - \{f_1\}$ and $E = E_1[i_1 \leftarrow f_1]$. Suppose

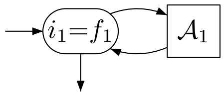


Figure 9: Automaton for X_1^*

that $x_1, \ldots, x_n$ are accepted by $\mathcal{A}_1$. The successful paths $\gamma_1, \ldots, \gamma_n$ in $\mathcal{A}_1$ labeled by $x_1, \ldots, x_n$ can be concatenated to form a successful path in $\mathcal{A}$. Conversely, there are finitely many occurrences of the state $i_1 = f_1$ in a successful path in $\mathcal{A}$ since there is no limit transition $P \to q$ or $q \to P$ with P containing i_1. Therefore, a successful path in $\mathcal{A}$ is the concatenation of a finite number of successful paths in $\mathcal{A}_1$. This proves that $\mathcal{A}$ recognizes the set $X_1^* = {X'_1}^*$.

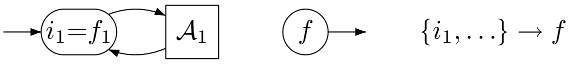


Figure 10: Automaton for ${X'_1}^\omega$

The set X_1^ω is equal to $\delta(X_1){X'_1}^* + {X'_1}^\omega$. Since we have already proved that the class of recognizable sets is closed union and finite iteration, it suffices to construct an automaton for set ${X'_1}^\omega$. This set is recognized by the automaton $\mathcal{A}$ obtained by first merging the initial state i_1 and final state f_1 of $\mathcal{A}_1$ into a state called i_1 which is initial but not final, and then by adding a new final state f together with all left limit transitions $P \to f$ where P contains i_1. The construction is pictured in Figure 10. The automaton $\mathcal{A}$ is equal $(Q, E, \{i_1\}, \{f\})$ where $Q = Q_1 \cup \{f\} - \{f_1\}$ and $E = E_1[i_1 \leftarrow f_1] \cup \{P \to f \mid i_1 \in P\}$. Suppose that the words $x_0, x_1, x_2, \ldots$ are accepted by $\mathcal{A}_1$. The successful paths $\gamma_0, \gamma_1, \gamma_2, \ldots$ in $\mathcal{A}_1$ labeled by $x_0, x_1, x_2, \ldots$ can be concatenated with an additional state f at the end to form a successful path in $\mathcal{A}$. Conversely, there are exactly ω occurrences of the state $i_1 = f_1$ in a successful path in $\mathcal{A}$ since the only transitions entering the final state are the left limit transitions $P \to f$ with $i_1 \in P$. Furthermore, the least upper bound of these occurrences is the last cut since there is no left limit transition $P \to q$ with $i_1 \in P$ and $q \neq f$. Therefore, a successful path in $\mathcal{A}$ is the concatenation of ω successful paths in $\mathcal{A}_1$. This proves that $\mathcal{A}$ recognizes the set ${X'_1}^\omega$.

An automaton recognizing $X_1^{-\omega}$ is similar to the automaton for X^ω.

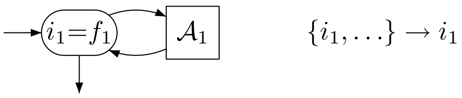


Figure 11: Automaton for $X_1^\sharp$

The set $X_1^\sharp = {X'_1}^\sharp$ is recognized by the automaton $\mathcal{A}$ obtained by first merging the initial state i_1 and the final state f_1 into a state called i_1 which is both initial and final, and then by adding all left limit transitions $P \to i_1$ where P contains i_1. The construction is pictured in Figure 11. The automaton $\mathcal{A}$ is equal to $(Q, E, \{i_1\}, \{i_1\})$ where $Q = Q_1 - \{f_1\}$ and $E = E_1[i_1 \leftarrow f_1] \cup \{P \to i_1 \mid i_1 \in P\}$. Let α be a countable ordinal and suppose that each word x_β is accepted by $\mathcal{A}_1$ for $\beta < \alpha$. Let γ_β be a successful path labeled by x_β. Suppose first that β is not a limit ordinal. Thus the sequence γ of states obtained by concatenating the paths γ_β is a successful path in $\mathcal{A}$ since this automaton has the transitions $P \to i_1$ where $i_1 \in P$. In case β is a limit ordinal, the paths γ_β are concatenated with the additional state i_1 at the end to form a successful path in $\mathcal{A}$. Conversely, since the automaton $\mathcal{A}$ does not have any transition $q \to P$ where $i_1 \in P$, the ordering of the occurrences of the state i_1 in a successful path is well-ordered. Therefore, a successful path in $\mathcal{A}$ is the concatenation of a sequence of successful paths in $\mathcal{A}_1$ indexed by some countable ordinal. This proves that $\mathcal{A}$ recognizes the set ${X_1}^\sharp$.

An automaton recognizing $X_1^{-\sharp}$ is similar to the automaton for $X_1^\sharp$.

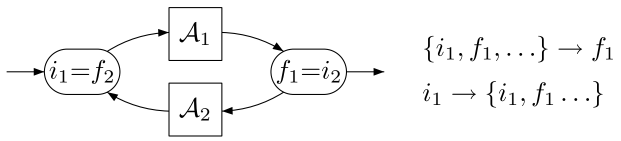
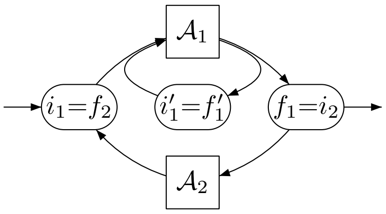


Figure 12: Automaton for $X_1 \diamond X_2$

For the set $X_1 \diamond X_2$, we first suppose that X_1 and X_2 do not contain the empty word. They are then recognized by the normalized automata $\mathcal{A}_1$ and $\mathcal{A}_2$. The set $X_1 \diamond X_2$ is recognized by the automaton $\mathcal{A}$ obtained by juxtaposing the automata $\mathcal{A}_1$ and $\mathcal{A}_2$, by merging the initial state i_1 of $\mathcal{A}_1$ and the final state f_2 of $\mathcal{A}_2$ into an initial state called i_1, by merging the initial state i_2 of $\mathcal{A}_2$ and the final state f_1 of $\mathcal{A}_1$ into a final state called f_1, and by adding all the left limit transitions $P \to f_1$ and all the right limit transitions $i_1 \to P$ where P contains both i_1 and f_1. The construction is pictured in Figure 12. The automaton $\mathcal{A}$ is then equal to $(Q, E, \{i_1\}, \{f_1\})$ where $Q = Q_1 \cup Q_2 - \{i_2, f_2\}$, $E = E_1 \cup E_2[i_1 \leftarrow f_2][f_1 \leftarrow i_2] \cup \{P \to f_1 \mid i_1, f_1 \in P\} \cup \{i_1 \to P \mid i_1, f_1 \in P\}$. Let J be a non empty countable scattered linear ordering. If the words $(x_j)_{j \in J}$ are accepted by $\mathcal{A}_1$ and if the words $(y_c)_{c \in \hat{J}^*}$ are accepted by $\mathcal{A}_2$, the successful paths labeled by these words can be interleaved and concatenated to form a path in $\mathcal{A}$ since this automaton has all the appropriate limit transitions. Conversely, a successful path in $\mathcal{A}$ is made of intervals which either contain states of $\mathcal{A}_1$ or states of $\mathcal{A}_2$. Applying Lemma 8 to the ordering K of these intervals, a path in $\mathcal{A}$ can be decomposed as the concatenation of successful paths $(\gamma_j)_{j \in J}$ in $\mathcal{A}_1$ and successful paths $(\gamma_c)_{c \in \hat{J}^*}$ in $\mathcal{A}_2$. This proves that $\mathcal{A}$ recognizes the set $X_1 \diamond X_2$.

If either X_1 or X_2 contains the empty word, the previous construction must



$\{i_1, f_1 \ldots\} \to f_1, i_1'$
$\{i_1', \ldots\} \to f_1, i_1'$
$i_1, i_1' \to \{i_1, f_1 \ldots\}$
$i_1, i_1' \to \{i_1', \ldots\}$

Figure 13: Automaton for $X_1 \diamond (X_2 + \varepsilon)$

be slightly adapted but it remains essentially the same. We describe it for the case where X_2 contains the empty word but X_1 does not. The other cases are very similar. The construction is as follows. The initial state i_1 of $\mathcal{A}_1$ is duplicated. This means that a new initial state i_1' is added, and new transitions $i_1' \xrightarrow{a} q$ and $i_1' \to P$ are added whenever $i_1 \xrightarrow{a} q$ and $i_1 \to P$ are transitions of $\mathcal{A}_1$. Analogously, the final state f_1 is duplicated by adding a new final state f_1'. The automaton $\mathcal{A}$ recognizing $X_1 \diamond X_2$ is obtained by juxtaposing the automata $\mathcal{A}_1$ and $\mathcal{A}_2$, merging the two states i_1' and f_1', merging the initial state i_1 of $\mathcal{A}_1$ and the final state f_2 of $\mathcal{A}_2$ into an initial state called i_1, merging the initial state i_2 of $\mathcal{A}_2$ and the final state f_1 of $\mathcal{A}_1$ into a final state called f_1, and adding all the left limit transitions $P \to f_1$ and $P \to i_1'$ and all right limit transitions $i_1 \to P$ and $i_1' \to P$ where P either contains i_1' or contains both i_1 and f_1. The construction is pictured in Figure 13.

We have now shown that for each rational operation, there is a corresponding construction on automata. For the base cases of the induction, we note that $\varnothing$ is recognized by the automaton $(\{i, f\}, \varnothing, \{i\}, \{f\})$, and for each letter a, the set $\{a\}$ is recognized by the automaton $(\{i, f\}, \{i \xrightarrow{a} f\}, \{i\}, \{f\})$. Thus, for each rational expression, there is an automaton recognizing the set denoted by that expression.

6.2 From automata to rational expressions

In this section, we prove that for any automaton $\mathcal{A}$, there is a rational expression denoting the set of words recognized by $\mathcal{A}$. We first introduce some notation. Let $\mathcal{A} = (Q, E, I, F)$ be a fixed automaton. The *content* $\mathrm{C}(\gamma)$ of a path γ is the set of states which occur inside γ. It does not take into account the first and the last state of the path. Recall that a path γ labeled by a word of length J is a function from $\hat{J}$ into Q. The content of a path γ is thus formally defined by $\mathrm{C}(\gamma) = \gamma(\hat{J}^*)$. Recall that $\mathrm{Out}(P)$ and $\mathrm{In}(P)$ respectively denote the set of transitions that leave and enter a state in P.

A path γ from state p to state p' having content P and labeled by x is denoted by

$$p \underset{P}{\overset{x}{\rightsquigarrow}} p'.$$

If $x \neq \varepsilon$, the path γ uses a first transition σ which leaves p and a last transition σ'

which enters p'. To emphasize the use of σ and σ', the path γ is then denoted

$$\sigma \underset{P}{\overset{x}{\rightsquigarrow}} \sigma'.$$

In both notations, we may omit the label or the content of the path if they are not relevant.

In the proof, we often decompose a path into several paths. Given a path γ of length $\hat{J}$ and two cuts $c < c'$ of J, the path denoted by $\gamma[c, c']$ is the part of γ from c to c'.

Let P be a subset of Q and let σ and σ' be two transitions of $\mathcal{A}$. We define the sets of words $\Pi^P_{\sigma,\sigma'}$, $\nabla^P_{\sigma,\sigma'}$, $\Delta^P_{\sigma,\sigma'}$, and $\Gamma^P_{\sigma,\sigma'}$ as follows.

$$\begin{aligned}
\Pi^P_{\sigma,\sigma'} &= \{x \mid \exists\gamma \text{ of the form } \sigma \underset{P}{\overset{x}{\rightsquigarrow}} \sigma'\} \\
\nabla^P_{\sigma,\sigma'} &= \{x \mid \exists\gamma \text{ of the form } \sigma \underset{P}{\overset{x}{\rightsquigarrow}} \sigma' \text{ without any transition } P \to r\} \\
\Delta^P_{\sigma,\sigma'} &= \{x \mid \exists\gamma \text{ of the form } \sigma \underset{P}{\overset{x}{\rightsquigarrow}} \sigma' \text{ without any transition } r \to P\} \\
\Gamma^P_{\sigma,\sigma'} &= \{x \mid \exists\gamma \text{ of the form } \sigma \underset{P}{\overset{x}{\rightsquigarrow}} \sigma' \text{ without any transition } r \to P \text{ or } P \to r\}.
\end{aligned}$$

Note that *without any transition* $P \to r$ means that the path γ does not use any left limit transition of the form $P \to r$ for all $r \in Q$ except perhaps for the last transition if σ' is a left limit transition of this form. Thus, the left limit $\lim_{c^-} \gamma$ at any cut c different from the last cut must be a proper subset of P. This fact will be used in the arguments that follow.

Both $\nabla^P_{\sigma,\sigma'}$ and $\Delta^P_{\sigma,\sigma'}$ are subsets of $\Pi^P_{\sigma,\sigma'}$ and the set $\Gamma^P_{\sigma,\sigma'}$ is equal to the intersection $\nabla^P_{\sigma,\sigma'} \cap \Delta^P_{\sigma,\sigma'}$.

The paths considered in the definition of the sets $\Pi^P_{\sigma,\sigma'}$, $\nabla^P_{\sigma,\sigma'}$, $\Delta^P_{\sigma,\sigma'}$, and $\Gamma^P_{\sigma,\sigma'}$ use at least one transition. Therefore, the empty word is not contained in them. Since a path is successful if its first and last states are respectively initial and final, the set of words recognized by the automaton $\mathcal{A}$ is equal to the union

$$\delta(\mathcal{A}) + \bigcup_{P \subseteq Q, \sigma \in \mathrm{Out}(I), \sigma' \in \mathrm{In}(F)} \Pi^P_{\sigma,\sigma'}$$

where $\delta(\mathcal{A})$ is equal to ε if $I \cap F \neq \varnothing$ and to $\varnothing$ otherwise. We claim that every set of the form $\Pi^P_{\sigma,\sigma'}$, $\nabla^P_{\sigma,\sigma'}$, $\Delta^P_{\sigma,\sigma'}$, or $\Gamma^P_{\sigma,\sigma'}$ is rational. The proof is by induction on the cardinality of P.

We first suppose that P is the empty set $\varnothing$. If both transitions σ and σ' are equal to the same successor transition $p \xrightarrow{a} q$, all four sets $\Pi^P_{\sigma,\sigma'}$, $\nabla^P_{\sigma,\sigma'}$, $\Delta^P_{\sigma,\sigma'}$ and $\Gamma^P_{\sigma,\sigma'}$ are equal to the singleton $\{a\}$. Otherwise, they are all empty. In both cases, they are rational. This completes the base case of the induction.

The rest of the proof is devoted to the induction step. We suppose that P is non-empty and that for any $R \subsetneq P$ and any transitions τ and τ', all four sets $\Pi^R_{\tau,\tau'}$, $\nabla^R_{\tau,\tau'}$, $\Delta^R_{\tau,\tau'}$ and $\Gamma^R_{\tau,\tau'}$ are rational. We first prove in Section 6.2.1 that the set $\Gamma^P_{\sigma,\sigma'}$ has a rational expression over the sets $\Pi^R_{\tau,\tau'}$ for $R \subsetneq P$. By the induction hypothesis and by Corollary 14, the set $\Gamma^P_{\sigma,\sigma'}$ is rational. Then we prove in Section 6.2.2 that both sets $\nabla^P_{\sigma,\sigma'}$ and $\Delta^P_{\sigma,\sigma'}$ have rational expressions over the sets $\Gamma^P_{\tau,\tau'}$ studied in Section 6.2.1 and the sets $\Pi^R_{\tau,\tau'}$, $\nabla^R_{\tau,\tau'}$, $\Delta^R_{\tau,\tau'}$ and $\Gamma^R_{\tau,\tau'}$ for $R \subsetneq P$. Again by the induction hypothesis and by Corollary 14, $\nabla^P_{\sigma,\sigma'}$ and $\Delta^P_{\sigma,\sigma'}$ are rational. We finally prove in Section 6.2.3 that the set $\Pi^P_{\sigma,\sigma'}$ has a rational expression over the sets $\Gamma^P_{\tau,\tau'}$, $\nabla^P_{\tau,\tau'}$, $\Delta^P_{\tau,\tau'}$ and the sets $\Pi^R_{\tau,\tau'}$, $\nabla^R_{\tau,\tau'}$, $\Delta^R_{\tau,\tau'}$ and $\Gamma^R_{\tau,\tau'}$ for $R \subsetneq P$. Once again by the induction hypothesis and by Corollary 14, $\Pi^P_{\sigma,\sigma'}$ is rational.

In the following section, the proof uses the Kleene theorem for usual automata accepting finite words and a Kleene like theorem for Büchi automata accepting ω-words. We refer the reader to [19] for finite words and to [20] for ω-words.

6.2.1 $\Gamma^P_{\sigma,\sigma'}$ is rational

Let γ be a path labeled by a word $x = (a_j)_{j\in J}$ in $\Gamma^P_{\sigma,\sigma'}$. Let p and p' be the first and last states of this path. By definition of $\Gamma^P_{\sigma,\sigma'}$, the path γ does not use any limit transition $r \to P$ or $P \to r$ for any $r \in Q$, except possibly for the beginning or ending transition. We show how to decompose the path γ in such a way as to find a rational expression for $\Gamma^P_{\sigma,\sigma'}$ over the sets $\Pi^R_{\tau,\tau'}$ with $R \subsetneq P$. We consider three cases depending on the form of the transitions σ and σ'.

• **σ is different from $p \to P$.** The idea is to decompose the path γ into a sequence of consecutive paths, each of whose content is strictly included in P. The way these consecutive paths are combined to give γ will be described by an automaton. Kleene's theorem for finite words together with its analog for ω-words will allow us to infer from the induction hypotheses that $\Gamma^P_{\sigma,\sigma'}$ is rational.

Consider the cut (K, L) of the ordering $\hat{J}$ defined as follows.

$$K = \{c \in \hat{J} \mid \mathrm{C}(\gamma[c_{\min}, c]) \subsetneq P\}$$
$$L = \{c \in \hat{J} \mid \mathrm{C}(\gamma[c_{\min}, c]) = P\}$$

Note that K is nonempty since the first cut $c_{\min}$ belongs to K and that L is also nonempty since the last cut $c_{\max}$ belongs to L. Since $\hat{J}$ is a complete ordering, K has a least upper bound c_1. By definition of (K, L) and of the content, this cut c_1 must belong to K and it is thus equal to $\max(K)$. Note that c_1 is different from $c_{\min}$ since σ is different from $p \to P$ and that it is different from $c_{\max}$ since $L \neq \varnothing$. Therefore there are transitions τ_1' and τ_1 entering and leaving the state $q_1 = \gamma(c_1)$. Let R_0 and R' be the contents $\mathrm{C}(\gamma[c_{\min}, c_1])$ and $\mathrm{C}(\gamma[c_1, c_{\max}])$. By definition of c_1, one has $R_0 \subsetneq P$ and $R' \subseteq P$. If $R' \subsetneq P$, then x belongs to $\Pi^{R_0}_{\sigma,\tau_1'}\Pi^{R'}_{\tau_1,\sigma'}$. If $R' = P$, then x belongs to $\Pi^{R_0}_{\sigma,\tau_1'}\Gamma^{R'}_{\tau_1,\sigma'}$ and the argument given above for γ can be repeated for the path $\gamma[c_1, c_{\max}]$ since τ_1 is not equal to $p \to P$. We then get a second cut c_2 in the interval $(c_1, c_{\max})$ such that $R_1 = \mathrm{C}(\gamma[c_1, c_2]) \subsetneq P$ and $\mathrm{C}(\gamma[c_2, c_{\max}]) \subseteq P$.

So, we get either a finite sequence or an ω-sequence of distinct cuts $c_1, c_2, \ldots$ corresponding to states $q_1, q_2, \ldots$ with related sets $\Pi^{R_0}_{\sigma,\tau_1'}, \Pi^{R_1}_{\tau_1,\tau_2'}, \ldots$, where $R_k \subsetneq$

P and τ'_k and τ_k are the transitions entering and leaving q_k. Note that by construction

$$R_{k-1} \cup \{q_k\} \cup R_k = P \tag{1}$$

for every positive integer k.

If the sequence of cuts obtained is a finite sequence $c_1, \ldots, c_n$, then the word $x \in \Gamma^P_{\sigma,\sigma'}$ belongs to the finite product

$$\Pi^{R_0}_{\sigma,\tau'_1} \Pi^{R_1}_{\tau_1,\tau'_2} \ldots \Pi^{R_{n-1}}_{\tau_{n-1},\tau'_n} \Pi^{R_n}_{\tau_n,\sigma'}. \tag{2}$$

In that case, the transition σ' must be different from the transition $P \to p'$ because $R_n \subsetneq P$. We claim that if the sequence of cuts is an ω-sequence $c_1, c_2, \ldots$, then x belongs to the ω-product

$$\Pi^{R_0}_{\sigma,\tau'_1} \Pi^{R_1}_{\tau_1,\tau'_2} \Pi^{R_2}_{\tau_2,\tau'_3} \ldots \tag{3}$$

The least upper bound $c = \sup\{c_k \mid k \geq 1\}$ is equal to the last cut of J. Indeed, Equation (1) implies that the left limit $\lim_{c^-} \gamma$ is equal to P. Since the path γ is without any left limit transition $P \to q$ except for the last one, c is the last cut of J and σ' must be the left limit transition $P \to p'$.

Thus, we have proved that if x belongs to $\Gamma^P_{\sigma,\sigma'}$, then it belongs to a product like (2) if σ' is different from $P \to p'$, or like (3) if σ' is equal to $P \to p'$. We consider two cases.

We first suppose that σ' is different from $P \to p'$. Let us now give an automaton $\mathcal{B}$ which describes in a precise way the possible concatenations of the sets $\Pi^{R_k}_{\tau_k,\tau'_{k+1}}$ that can arise in (2). The automaton is over the alphabet $C = \{\Pi^R_{\tau,\tau'} \mid R \subsetneq P, \tau, \tau' \in E\}$. This means that each set $\Pi^R_{\tau,\tau'}$ is considered as a letter.

The states of $\mathcal{B}$ are triplets (R, q, R') where q is a state of $\mathcal{A}$ and R, R' are sets of states of $\mathcal{A}$ such that $R \cup \{q\} \cup R' = P$. The automaton $\mathcal{B}$ also has initial states of the form (p, R') for $p \in P$ and $R' \subsetneq P$ and final states of the form (R, p') for $p' \in P$ and $R \subsetneq P$. Transitions of $\mathcal{B}$ are defined as follows. There is a transition

$$(R_1, q_1, R'_1) \xrightarrow{\Pi^R_{\tau,\tau'}} (R_2, q_2, R'_2)$$

if and only if $R'_1 = R_2 = R$, τ leaves q_1 and τ' enters q_2. There is a transition from an initial state

$$(p, R'_1) \xrightarrow{\Pi^R_{\tau,\tau'}} (R_2, q_2, R'_2)$$

if and only if $R'_1 = R_2 = R$, $\tau = \sigma$ and τ' enters q_2. Analogously, there is a transition to a final state

$$(R_1, q_1, R'_1) \xrightarrow{\Pi^R_{\tau,\tau'}} (R_2, p')$$

if and only if $R'_1 = R_2 = R$, τ leaves q_1 and $\tau' = \sigma'$. Note that the initial states can only occur as the first state of a path and that final states can only occur as the last state of a path. Indeed, there is no transition entering initial states and no transition leaving final states.

This automaton $\mathcal{B}$ is a usual automaton that accepts finite words over C. We claim that it has the following two properties.

i) If a finite sequence $\Pi^{R_0}_{\sigma,\tau'_1}\Pi^{R_1}_{\tau_1,\tau'_2}\dots\Pi^{R_n}_{\tau_n,\sigma'}$ occurs in (2) as the decomposition of a word x in $\Gamma^P_{\sigma,\sigma'}$, it is then accepted by $\mathcal{B}$.

ii) Conversely, if a sequence $\Pi^{R_0}_{\sigma,\tau'_1}\Pi^{R_1}_{\tau_1,\tau'_2}\dots\Pi^{R_n}_{\tau_n,\sigma'}$ is accepted by $\mathcal{B}$, the finite product of the corresponding sets is contained in $\Gamma^P_{\sigma,\sigma'}$.

Statement i) follows directly from Equation (1). Conversely, let w be a word $\Pi^{R_0}_{\sigma,\tau'_1}\Pi^{R_1}_{\tau_1,\tau'_2}\dots\Pi^{R_n}_{\tau_n,\sigma'}$ accepted by $\mathcal{B}$ and let x be a word over A belonging to the product $\Pi^{R_0}_{\sigma,\tau'_1}\Pi^{R_1}_{\tau_1,\tau'_2}\dots\Pi^{R_n}_{\tau_n,\sigma'}$ of the corresponding sets. The word x can be factorized $x = x_0x_1\dots x_n$ where each x_i belongs to $\Pi^{R_i}_{\tau_i,\tau'_{i+1}}$. By definition of $\Pi^{R_i}_{\tau_i,\tau'_{i+1}}$, there is in $\mathcal{A}$, a path $\tau_i \rightsquigarrow \tau_{i+1}$ of content R_i and labeled by x_i. By definition of the transitions of $\mathcal{B}$, these paths can be concatenated in $\mathcal{A}$ to yield a path $\sigma \rightsquigarrow \sigma'$ labeled by x. By definition of the states of $\mathcal{B}$, the content of this path is P. This completes the proof of Statement ii).

Let W be the set of words accepted by $\mathcal{B}$. By Kleene's theorem, W is rational. By the previous two statements, $\Gamma^P_{\sigma,\sigma'}$ has a rational expression over the sets $\Pi^R_{\tau,\tau'}$ for $R \subsetneq P$. By Corollary 14 and by the induction hypothesis, $\Gamma^P_{\sigma,\sigma'}$ is rational. This completes the proof of the case σ' different from $P \to p'$. Note that the rational expression expressing $\Gamma^P_{\sigma,\sigma'}$ over the $\Pi^R_{\tau,\tau'}$ only involves unions, concatenations, and finite iterations since W only contains finite words.

We now suppose that σ' is equal to the transition $P \to p'$. The proof of this case is very similar, but since the products of (3) are ω-products, the automaton $\mathcal{B}$ is replaced by a Büchi automaton $\mathcal{B}'$ that accepts ω-words over the alphabet C. We first recall that a Büchi automaton is a usual automaton with initial and final states. An infinite path in such an automaton is successful if it starts in an initial state and if it goes through a final state infinitely many times. The automaton $\mathcal{B}'$ is almost the same as the automaton $\mathcal{B}$. The only difference is that there are no final states of the form (R, p') and that all states of the form (R, q, R') and (p, R') are final. All infinite paths starting in an initial state of this Büchi automaton are thus successful.

We claim that the automaton $\mathcal{B}'$ has the following two properties

i) If an ω-sequence $\Pi^{R_0}_{\sigma,\tau'_1}\Pi^{R_1}_{\tau_1,\tau'_2}\Pi^{R_2}_{\tau_2,\tau'_3}\dots$ occurs in (3) as the decomposition of a word x in $\Gamma^P_{\sigma,\sigma'}$, it is then accepted by $\mathcal{B}'$.

ii) Conversely, if an ω-sequence $\Pi^{R_0}_{\sigma,\tau'_1}\Pi^{R_1}_{\tau_1,\tau'_2}\Pi^{R_2}_{\tau_2,\tau'_3}\dots$ is accepted by $\mathcal{B}'$, the ω-product of the corresponding sets is contained in $\Gamma^P_{\sigma,\sigma'}$

The proofs of these two statements are exactly the same as those for $\mathcal{B}$. Let W be the set of ω-words accepted by $\mathcal{B}'$. By the Kleene like theorem for Büchi automata, W is rational. By the previous two statements, $\Gamma^P_{\sigma,\sigma'}$ has a rational expression over the sets $\Pi^R_{\tau,\tau'}$ for $R \subsetneq P$. By Corollary 14 and by the induction hypothesis, $\Gamma^P_{\sigma,\sigma'}$ is rational. This completes the proof of the case σ' equal to

$P \to p'$. Note that the rational expression expressing $\Gamma^P_{\sigma,\sigma'}$ over the $\Pi^R_{\tau,\tau'}$ only involves unions, concatenations, finite iterations, and ω iterations since W only contains ω-words.

• **σ' is different from $P \to p'$.** This case is symmetrical to the previous one. The ω iteration is replaced by the $-\omega$ iteration.

• **σ equals $p \to P$ and σ' equals $P \to p'$.** Let c be a cut of J such that $c_{\min} < c < c_{\max}$. The path γ is then decomposed into two paths $\gamma_1 = \gamma[c_{\min}, c]$ and $\gamma_2 = \gamma[c, c_{\max}]$ and the word x is factorized $x = x_1 x_2$ where x_1 and x_2 are the labels of γ_1 and γ_2. Both contents of γ_1 and γ_2 are equal to P because σ equals $p \to P$ and σ' equals $P \to p'$.

Let q be the state $\gamma(c)$ and let τ, τ' be the transitions entering and leaving q at c. By definition of $\Gamma^P_{\sigma,\sigma'}$, both transitions τ and τ' are different from $p \to P$ and $P \to p'$. The word x belongs then to a product of the form $\Gamma^P_{\sigma,\tau'}\Gamma^P_{\tau,\sigma'}$ for some $(\tau', \tau) \in \mathcal{T}$ where

$$\mathcal{T} = \{(\tau', \tau) \mid \exists q \ \ \tau' \in \mathrm{In}(q), \tau \in \mathrm{Out}(q) \text{ and } \tau' \neq P \to q, \tau \neq q \to P\}.$$

It follows that

$$\Gamma^P_{\sigma,\sigma'} \subset \bigcup_{(\tau',\tau)\in\mathcal{T}} \Gamma^P_{\sigma,\tau'}\Gamma^P_{\tau,\sigma'}$$

and since $\Gamma^P_{\sigma,\tau'}\Gamma^P_{\tau,\sigma'} \subset \Gamma^P_{\sigma,\sigma'}$ for each $(\tau', \tau) \in \mathcal{T}$, the relationship is actually one of equality. The set on the right is a finite union of products of terms shown to be rational in the previous two cases and so is itself rational; thus $\Gamma^P_{\sigma,\sigma'}$ is rational.

6.2.2 $\Delta^P_{\sigma,\sigma'}$ and $\nabla^P_{\sigma,\sigma'}$ are rational

We now prove that $\Delta^P_{\sigma,\sigma'}$ is rational. The proof for $\nabla^P_{\sigma,\sigma'}$ is symmetrical. Let γ be a path labeled by a word $x = (a_j)_{j\in J}$ in $\Delta^P_{\sigma,\sigma'}$. Let p and p' be the first and the last state of this path. By definition, this path does not use any right limit transition $r \to P$, except perhaps for σ.

The idea is again to decompose the path γ into a sequence of consecutive paths. The decomposition is performed according to the occurrences of the left limit transitions $P \to r$ in γ. The label of each path of this decomposition belongs to some set $\Gamma^P_{\tau,\tau'}$ whose rationality has been proved in Section 6.2.1.

Consider the subordering K of $\hat{J}$ defined by the occurrences of left limit transitions $P \to r$ in γ, that is

$$K = \{c \in \hat{J} \mid \lim_{c^-} \gamma = P\}. \tag{4}$$

Note that $c_{\min}$ cannot belong to K because $\lim_{c^-_{\min}} \gamma$ is empty by definition. The last cut $c_{\max}$ belongs to K if and only if σ' is the transition $P \to p'$.

If K is empty, then the word x belongs to $\Gamma^P_{\sigma,\sigma'}$. In the rest of this section, we assume that K is not empty. We will consider two cases depending on whether σ is the transition $p \to P$ or not. Before considering these two cases, we state two properties of K that will be useful in both cases.

We first claim that K has a greatest element $\max(K)$. Let $c = \sup(K)$ be the least upper bound of K. The left limit $\lim_{c^-} \gamma$ is then equal to P and thus c belongs to K. If σ' is equal to $P \to p'$, then $\max(K) = c_{\max}$. Otherwise one has $\max(K) < c_{\max}$.

The ordering K is not necessarily an ordinal but it is almost well-ordered. We claim that if it contains an infinite decreasing sequence $c_0 > c_1 > c_2 > \cdots$, then this sequence converges to $c_{\min}$ and the transition σ must be $p \to P$. This implies in particular that for all cut (L, R) of K different from $(\varnothing, K)$, the ordering R is an ordinal.

Suppose indeed that K contains an infinite sequence $c_0 > c_1 > c_2 > \cdots$ and let c be the greatest lower bound of $\{c_i \mid i \geq 0\}$. The right limit $\lim_{c^+} \gamma$ is then equal to P. Since the path does not use any right limit transition $r \to P$, except perhaps for σ, it follows that c is the first cut $c_{\min}$ and that σ is the transition $p \to P$.

We consider two cases depending on the form of transition σ.

• **σ is different from $p \to P$.** By the second property of K, there exists no infinite decreasing sequence in K, that is, K is well-ordered. It follows that K has a least element $\min(K)$ and that each element c of K different from $\max(K)$ has a successor in K that we denote by $c+1$ by a slight abuse of language. Let τ'_c and τ_c be the transitions entering and leaving the state $q_c = \gamma(c)$. By definition of K, the transition τ'_c is equal to $P \to q_c$ for each $c \in K$. Let x_c be the label of the path $\gamma[c, c+1]$. Each word x_c belongs to $\Gamma^P_{\tau_c,\tau'_{c+1}}$. Let y be the label of the path $\gamma' = \gamma[c_{\min}, \min(K)]$. The word y belongs to $\Gamma^P_{\sigma,\tau'_{\min(K)}}$. If the transition σ' is different from $P \to p'$, one has $\max(K) < c_{\max}$. In that case, we also consider the label z of the path $\gamma'' = \gamma[\max(K), c_{\max}]$. The word z belongs to $\Gamma^P_{\tau_{\max(K)},\sigma'}$ if $\mathrm{C}(\gamma'') = P$ and it belongs to $\Pi^R_{\tau_{\max(K)},\sigma'}$ if $\mathrm{C}(\gamma'') = R \subsetneq P$. If σ' is equal to $P \to p'$, the word x is equal to $y \prod_{c<\max(K)} x_c$ and if σ' is different from $P \to p'$, the word x is equal to $y\big(\prod_{c<\max(K)} x_c\big)z$.

Therefore, as K is well-ordered,

$$\Delta^P_{\sigma,\sigma'} \subseteq \Gamma^P_{\sigma,\sigma'} \cup \left(\bigcup_{\tau' \in T'} \Gamma^P_{\sigma,\tau'} \right) \left(\bigcup_{\tau \in T} \bigcup_{\tau' \in T'} \Gamma^P_{\tau,\tau'} \right)^{\sharp} \tag{5}$$

if σ' equals $P \to p'$, and

$$\Delta^P_{\sigma,\sigma'} \subseteq \Gamma^P_{\sigma,\sigma'} \cup \left(\bigcup_{\tau' \in T'} \Gamma^P_{\sigma,\tau'} \right) \left(\bigcup_{\tau \in T} \bigcup_{\tau' \in T'} \Gamma^P_{\tau,\tau'} \right)^{\sharp} \left(\bigcup_{\tau \in T} \Gamma^P_{\tau,\sigma'} \cup \bigcup_{\tau \in T} \bigcup_{R \subsetneq P} \Pi^R_{\tau,\sigma'} \right) \tag{6}$$

if σ' is different from $P \to p'$, where the subsets T and T' of transitions are defined as follows

$$T = \{\tau \mid \exists q \in P \quad \tau \in \mathrm{Out}(q), \tau \neq q \to P \text{ and } P \to q \in E\}, \tag{7}$$
$$T' = \{P \to q \mid q \in P \text{ and } P \to q \in E\}. \tag{8}$$

Let us show that inclusions (5) and (6) are in fact equalities. We first study the set

$$X = \left(\bigcup_{\tau \in T \ \tau' \in T'} \Gamma^P_{\tau,\tau'} \right)^\sharp .$$

We start with a small remark concerning transitions of T'. Let $\tau'_1, \tau'_2 \in T'$ be two transitions of the form $\tau'_1 = P \to q_1$ and $\tau'_2 = P \to q_2$. Let δ be a path ending with τ'_1, that is δ is a sequence of states of the form $\delta' q_1$. Then $\delta' q_2$ is also a path with the same label and the same content as $\delta' q_1$. By this remark, we can conclude that $\Gamma^P_{\tau,\tau'_1} = \Gamma^P_{\tau,\tau'_2}$.

We claim that each word $x \in X$ labels a path $\delta : \tau \rightsquigarrow \tau'$ of content P for some $\tau \in T$ and $\tau' \in T'$, and that δ is without any right limit transition $r \to P$, for all $r \in Q$. Let x be a word of X. It can be factorized $x = \Pi_{\beta<\alpha} x_\beta$ where $\alpha < \omega_1$ and each word x_β belongs to $\Gamma^P_{\tau_\beta,\tau'_\beta}$ for some $\tau_\beta \in T$ and $\tau'_\beta \in T'$. Let δ_β be a path $\tau_\beta \rightsquigarrow \tau'_\beta$ labelled by x_β. Define δ'_β by $\delta_\beta = \delta'_\beta q_\beta$. We claim that $\delta = (\Pi_{\beta<\alpha} \delta'_\beta) q$ where $P \to q \in E$ is a path labelled by x. Indeed, if β is a successor ordinal $\beta' + 1$, the first state of δ_β can replace the last state of $\delta_{\beta'}$ due to the previous remark. If β is a limit ordinal, the left limit of δ at the first state q of δ_β is P and $P \to q$ is a transition by definition of T. Note that the path δ is without any right limit transition $r \to P$, for all $r \in Q$.

Finally a path δ' of label $y \in \Gamma^P_{\sigma,\tau'}$ with $\tau' \in T'$, can be concatenated with the path δ constructed above, due again to the previous remark. For the same reason, the path $\delta'\delta$ can be concatenated with a path δ'' labeled by z in $\Gamma^P_{\tau,\sigma'}$ or in $\Pi^R_{\tau,\sigma'}$ for $R \subsetneq P$. Moreover the resulting path is without any right limit transition $r \to P$, for all $r \in Q$.

Therefore, if σ' equals $P \to p'$, then Equality (5) holds, otherwise Equality (6) holds. This shows that $\Delta^P_{\sigma,\sigma'}$ has a rational expression over the sets $\Gamma^P_{\tau,\tau'}$ and the sets $\Pi^R_{\tau,\tau'}$ for $R \subsetneq P$. By the induction hypothesis and by Corollary 14, $\Delta^P_{\sigma,\sigma'}$ is rational.

• σ **equals** $p \to P$**.** We work again with the path γ with label x as defined at the beginning of Section 6.2.2. Two cases occur depending on the existence of an infinite decreasing sequence in the set K defined by Equation (4).

We first suppose that there is no infinite decreasing sequence in K. The set K is then well-ordered. If $K \neq \varnothing$, it has a least element $c = \min(K)$ which is different from $c_{\min}$. The path γ can be decomposed into the two paths $\gamma_1 = \gamma[c_{\min}, c]$ and $\gamma_2 = \gamma[c, c_{\max}]$. Let x_1, x_2 be the labels of the paths γ_1, γ_2, and let τ', τ be the transitions entering and leaving the state $q = \gamma(c)$. By definition of c, the word x_1 belongs to $\Gamma^P_{\sigma,\tau'}$. The word x_2 belongs to $\Delta^P_{\tau,\sigma'}$ or in $\Pi^R_{\tau,\sigma'}$ with $R \subsetneq P$ according to the content of γ_2. Note that $\Delta^P_{\tau,\sigma'}$ is such that τ is different from $q \to P$ and thus has been proved to be rational in the previous case.

This shows that

$$x \in \bigcup_{\exists q\ \tau' \in \mathrm{In}(q), \tau \in \mathrm{Out}(q), \tau \neq q \to P} \Gamma^P_{\sigma,\tau'}(\Delta^P_{\tau,\sigma'} \cup \bigcup_{R \subsetneq P} \Pi^R_{\tau,\sigma'}) \tag{9}$$

We now suppose that there is an infinite decreasing sequence $c_1 > c_2 > c_3 > \cdots$ in K. We can assume that c_1 is different from $c_{\max}$. Otherwise c_1 is removed from the sequence to get another infinite decreasing sequence with that property. Let q_i be the state $\gamma(c_i)$, and let τ'_i and τ_i be the transitions entering and leaving the state q_i at c_i. Let γ_i be the path $\gamma[c_{i+1}, c_i]$. Recall that we have proved that the sequence $(c_i)_{i \geq 1}$ converges to $c_{\min}$. This shows that the path $\gamma[c_{\min}, c_1]$ is decomposed in a sequence of length $-\omega$ of the paths γ_i. By definition of $\Delta^P_{\sigma,\sigma'}$, each transition τ_i is different from the transition $q_i \to P$. This shows that the label x_i of γ_i belongs to $\Delta^P_{\tau_{i+1},\tau'_i}$ that has already been shown to be rational in the previous case. Let x_0 be label of the path $\gamma_0 = \gamma[c_1, c_{\max}]$. It belongs to $\Delta^P_{\tau_0,\sigma'}$ if $\mathrm{C}(\gamma_0) = P$ and it belongs to $\Pi^R_{\tau_0,\sigma'}$ if $\mathrm{C}(\gamma_0) = R \subsetneq P$. Thus

$$x \in \left(\bigcup_{\tau \in T\ \tau' \in T'} \Delta^P_{\tau,\tau'} \right)^{-\omega} \left(\bigcup_{\tau \in T} \Delta^P_{\tau,\sigma'} \cup \bigcup_{\tau \in T\ R \subsetneq P} \Pi^R_{\tau,\sigma'} \right) \tag{10}$$

where the sets T and T' are defined by Equations (7) and (8).

Combining the two cases (9), (10) with the case $K = \varnothing$, one has

$$\begin{aligned} \Delta^P_{\sigma,\sigma'} \subset \Gamma^P_{\sigma,\sigma'} \quad &\cup \bigcup_{\tau \in T, \tau' \in T'} \Gamma^P_{\sigma,\tau'} \left(\Delta^P_{\tau,\sigma'} \cup \bigcup_{R \subsetneq P} \Pi^R_{\tau,\sigma'} \right) \\ &\cup \left(\bigcup_{\tau \in T\ \tau' \in T'} \Delta^P_{\tau,\tau'} \right)^{-\omega} \left(\bigcup_{\tau \in T} \Delta^P_{\tau,\sigma'} \cup \bigcup_{\tau \in T\ R \subsetneq P} \Pi^R_{\tau,\sigma'} \right) \end{aligned}$$

One can verify that this inclusion is an equality. The sets $\Delta^P_{\tau,\sigma'}$, $\Delta^P_{\tau,\tau'}$ of this equality have been shown to be rational in the previous case. By Corollary 14 and by the induction hypothesis, $\Delta^P_{\sigma,\sigma'}$ is rational.

6.2.3 $\Pi^P_{\sigma,\sigma'}$ is rational

In this last part of the proof, we show $\Pi^P_{\sigma,\sigma'}$ has a rational expression over the sets $\Gamma^P_{\tau,\tau'}$, $\nabla^P_{\tau,\tau'}$, $\Delta^P_{\tau,\tau'}$ and the sets $\Pi^R_{\tau,\tau'}$, $\nabla^R_{\tau,\tau'}$, $\Delta^R_{\tau,\tau'}$ and $\Gamma^R_{\tau,\tau'}$ for $R \subsetneq P$.

The proof is organized as follows. A path of a word in $\Pi^P_{\sigma,\sigma'}$ is decomposed into small intervals using a first condensation $\sim$ (see Section 2.4 for definition). Four types of intervals are obtained by this decomposition. Then some intervals are grouped with a second condensation $\approx$ to reduce the number of type of intervals to two. Lemma 8 if finally used to get a rational expression involving the binary operation $\diamond$.

As before, let γ be a path labeled by a word $x = (a_j)_{j\in J}$ in $\Pi^P_{\sigma,\sigma'}$ and with a content equal to P. Let p and p' be the first and the last state of this path.

Consider the condensation $\sim$ on $\hat{J}$ defined as follows. For all cuts $c_1, c_2 \in \hat{J}$ where $c_1 < c_2$, the relations $c_1 \sim c_2$ and $c_2 \sim c_1$ hold if and only if for every $c \in [c_1, c_2)$, $\lim_{c^+} \gamma \neq P$ and for every $c \in (c_1, c_2]$, $\lim_{c^-} \gamma \neq P$.

Any equivalence class of $\sim$ is an interval. Note that two consecutive cuts are always $\sim$-equivalent. As $\hat{J}$ is complete with a least and a greatest element, every interval can be expressed as one of the four forms (c_1, c_2), $(c_1, c_2]$, $[c_1, c_2)$ and $[c_1, c_2]$. The form of an interval is unique because consecutive cuts are always equivalent. In the sequel we speak about the four *types* $(\,)$, $(\,]$, $[\,)$ and $[\,]$. Recall that an interval is right open (respectively right closed) if it has type $(\,)$ or $[\,)$ (respectively $(\,]$ or $[\,]$).

Let us study the structure of the equivalence classes of $\sim$ and the consecutive elements of the quotient ordering $K = \hat{J}/{\sim}$. Note that K is a complete scattered ordering with a least and a greatest element.

Let $k \in K$ be an equivalence class of $\sim$, with $c_1 = \inf(k)$ and $c_2 = \sup(k)$. For every $c \in (c_1, c_2)$, one has $\lim_{c^+} \gamma \neq P$ and $\lim_{c^-} \gamma \neq P$. If $c_1 = c_2$, that if k is the singleton $\{c_1\}$, then $\lim_{c_1^-} \gamma = \lim_{c_1^+} \gamma = P$. If $c_1 < c_2$, one verifies that k is left open if and only if $\lim_{c_1^+} \gamma = P$. Symmetrically, k is right open if and only if $\lim_{c_2^-} \gamma = P$.

Consider two consecutive elements $k < k'$ of K. As $\hat{J}$ is complete, it is impossible that k is right open and k' is left open. As consecutive cuts are equivalent, it is also impossible that k is right closed and k' is left closed. Moreover, any element k of K which is right open always has a successor k' which is then left closed. Symmetrically, if k is left open, it has a predecessor k' which is right closed.

We now define a condensation $\approx$ on the ordering $K = \hat{J}/{\sim}$. For all $k_1, k_2 \in K$, the relation

$$k_1 \approx k_2$$

holds if and only if all elements of the interval $[k_1, k_2]$ have the same type.

As done for K, we study the quotient ordering $L = K/{\approx}$ which is a complete scattered linear ordering with a least and a greatest element.

If l is an equivalence class of $\approx$ with elements of type $(\,)$, then l is a singleton. Otherwise, let $k_1 < k_2$ be in l. As K is scattered, there are two consecutive elements $k'_1 < k'_2$ in $[k_1, k_2]$ both of type $(\,)$. This is impossible in K. In the same way, any equivalence class of $\approx$ with elements of type $[\,]$ is a singleton. Classes whose elements are of type $(\,]$ or $[\,)$, however need not to be singletons.

Note that any equivalence class l with elements of type $[\,)$ always has a successor and if l is left open, it also has a predecessor. Moreover, recall that an element k of K of the type $[c_1, c_2)$ satisfies $\lim_{c_2^-} \gamma = P$. Thus, the least upper bound $c' = \sup\{c_2 \mid [c_1\, c_2) \in l\}$ also satisfies $\lim_{c'^-} \gamma = P$. If l is left open, the greatest lower bound $c = \inf\{c_1 \mid [c_1\, c_2) \in l\}$ satisfies $\lim_{c^+} \gamma = P$. In the same way, any class l with elements of type $(\,]$ always has a predecessor and if l is right open, it also has a successor. By similar reasoning, any class l reduced to a singleton of type $(\,)$ has a predecessor and a successor.

Consider now two consecutive elements $l < l'$ of L. Since $K/{\approx}$ is complete, it is impossible that l is right open and l' is left open. Let $k = \sup(l)$ and $k' = \inf(l')$. Assume first that l and l' are respectively right and left closed. Then k and k' are consecutive elements of K, for which we already know the possible configurations. Moreover, k and k' have different types since they are in different classes. Assume now that l is right open and that l' is left closed. If follows that $k = k'$ and the type of k is different from the type of the elements of l. As l is right open, two types are possible for its elements: $(\,]$ or $[\,)$. If the type is $[\,)$, one checks that k has type $[\,]$ due to the properties seen for K. If the type of the elements of l is $(\,]$, then k has either type $[\,]$ or $[\,)$. The last case when l is right closed and l' is left open is symmetrical.

Let us go further. We consider L as a collection of intervals of $\hat{J}$ with partition $\hat{J}$, by composing the two condensations $\sim$ and $\approx$. To avoid any confusion, when L is seen as the quotient ordering over K, that is $L = K/{\approx}$, an equivalence class is described as an interval composed of elements k of K. When L is seen as the quotient ordering over $\hat{J}$, a class is described as an interval of elements c of $\hat{J}$. A class l is then seen as the interval $\bigcup_{k \in l} k$ of $\hat{J}$.

Let us detail the different cases. We begin by considering the case where l is a singleton whose sole member k has type $(\,)$ or $[\,]$. Seen over $\hat{J}$, we respectively get $l = (c, c')$ or $l = [c, c']$. Let τ be the transition leaving $\gamma(c)$ and let τ' be the transition entering $\gamma(c')$. Therefore, if $l = (c, c')$, the label y of the path $\gamma[c, c']$ belongs to

$$\Gamma^P_{\tau,\tau'} \quad \text{with} \quad \tau = \gamma(c) \to P \text{ and } \tau' = P \to \gamma(c') \tag{11}$$

and if $l = [c, c']$, then either $y = \varepsilon$ (when $c = c'$) or y belongs to

$$Z_{\tau,\tau'} = \Gamma^P_{\tau,\tau'} \cup \bigcup_{R \subsetneq P} \Pi^R_{\tau,\tau'} \quad \text{with} \quad \tau \neq \gamma(c) \to P \text{ and } \tau' \neq P \to \gamma(c'). \tag{12}$$

Note that there is a union in the definition of $Z_{\tau,\tau'}$ depending on the context either equal to P or being a strict subset of P.

We next consider those classes l whose elements have type $[\,)$. Suppose that l is left open. We have seen that over $\hat{J}$, l is an interval $l = (c, c')$ such that $\lim_{c^+} \gamma = P$ and $\lim_{c'^-} \gamma = P$. The path $\gamma[c, c']$ is without any right limit transition $r \to P$. Let τ be the transition leaving $\gamma(c)$ and let τ' be the transition entering $\gamma(c')$. Then the label y of the path $\gamma[c, c']$ belongs to

$$\Delta^P_{\tau,\tau'} \quad \text{with} \quad \tau = \gamma(c) \to P \text{ and } \tau' = P \to \gamma(c'). \tag{13}$$

A similar description holds for classes l with elements of type $(\,]$, which are right open. With the same notation, we have $l = (c, c')$ and the label y of the path $\gamma[c, c']$ belongs to

$$\nabla^P_{\tau,\tau'} \quad \text{with} \quad \tau = \gamma(c) \to P \text{ and } \tau' = P \to \gamma(c'). \tag{14}$$

Two cases have yet to be considered: the case of left closed classes l with elements of type $[\,)$ and the symmetrical case of right closed classes l with elements of type ( ]. In the former case, l is an interval $l = [c, c')$ over $\hat{J}$ such that $\lim_{c^+} \gamma \neq P$ and $\lim_{c'^-} \gamma = P$. The path $\gamma[c, c']$ is again without any right limit transition $r \to P$ and with the same notation as before, y belongs to

$$\Delta^P_{\tau,\tau'} \quad \text{with} \quad \tau \neq \gamma(c) \to P \text{ and } \tau' = P \to \gamma(c'). \tag{15}$$

In the latter case, a symmetric description holds with $l = (c, c']$ and y belongs to

$$\nabla^P_{\tau,\tau'} \quad \text{with} \quad \tau = \gamma(c) \to P \text{ and } \tau' \neq P \to \gamma(c'). \tag{16}$$

We now define two sets M and M' such that $M \cap M' = \varnothing$ and such that $M \cup M'$ is a collection of intervals which partition $\hat{J}$. We define M as the subordering of L composed of all the elements l of L which are left and right open intervals (c, c') over $\hat{J}$. These intervals are exactly those paths whose labels are described in (11), (13) and (14).

The set M' is composed of some elements of L, and by pairs or triples of consecutive elements of L being merged into one interval. It is defined as follows. Consider an element $l = [c_1, c_2)$ of L whose label is described in (15). Recall that l has a successor l' which is necessarily a singleton k of type [ ]. Such a class l' has its label described in (12). The class l' is equal to an interval $l' = [c_2, c_3]$. Analogously, if $l'' = (c_3, c_4]$ is an element of L whose label is described in (16), it has a predecessor $l' = [c_2, c_3]$ whose label is described in (12).

Let $l' = [c_2, c_3]$ be an interval whose label is given in (12). If it has a predecessor $[c_1, c_2)$ with a label in (15) and a successor $(c_3, c_4]$ with a label in (16), then the interval $[c_1, c_4]$ obtained by fusion of the three intervals $[c_1, c_2)$, $[c_2, c_3]$ and $(c_3, c_4]$ belongs to M'. Otherwise, if l' has only a predecessor $[c_1, c_2)$ (respectively a successor $(c_3, c_4]$), then the interval $[c_1, c_3]$ obtained by fusion of $[c_1, c_2)$ and $[c_2, c_3]$ (respectively $[c_2, c_4]$ obtained by fusion of $[c_2, c_3]$ and $(c_3, c_4]$) belongs to M'. Otherwise, l' itself belongs to M'. Note that M' contains only intervals of $\hat{J}$ that are both left and right closed. Note also that every element l that, seen over $\hat{J}$, is either left or right closed (or both) becomes part of some fusion which is contained in M'.

Let us prove that the set $M \cup M'$ satisfies the hypotheses of Lemma 8. It is a complete and scattered ordering with a least and a greatest element since it is a condensation of L. If $m = (c, c')$ belongs to M, it has a successor since the cut c' belongs to an interval m' which is left closed and belongs to M'. Similarly, m has a predecessor. Take $m < m'$ two consecutive elements of $M \cup M'$. If $m \in M$, we just proved that $m' \in M'$. If $m \in M'$, it is impossible that $m' \in M'$ due to the possible configurations of consecutive elements in L. Hence, Lemma 8 holds and M' is isomorphic to $\hat{M}$.

For every $m = (c, c')$ in M, the label of the path $\gamma[c, c']$ belongs to (see (11), (13) and (14))

$$X_{\tau,\tau'} = \Gamma^P_{\tau,\tau'} \cup \Delta^P_{\tau,\tau'} \cup \nabla^P_{\tau,\tau'} \quad \text{with} \quad \tau = \gamma(c) \to P \text{ and } \tau' = P \to \gamma(c').$$

Consider $m = [c, c']$ in $\hat{M}$. Let y be the label of path $\gamma[c, c']$ (see (12), (15) and (16) and the definition of M'). Suppose first that $c \neq c'$, that is $y \neq \varepsilon$. Let

τ be the transition leaving $\gamma(c)$ and let τ' be the transition entering $\gamma(c')$. The label y belongs to

$$\begin{aligned} Y_{\tau,\tau'} = & \bigcup_{\substack{(\tau_1',\tau_1)\in\mathcal{T}_1\\(\tau_2',\tau_2)\in\mathcal{T}_2}} \Delta^P_{\tau,\tau_1'} Z_{\tau_1,\tau_2'} \nabla^P_{\tau_2,\tau'} \quad \cup \quad \bigcup_{(\tau_3',\tau_3)\in\mathcal{T}_3} \Delta^P_{\tau,\tau_3'} \nabla^P_{\tau_3,\tau'} \\ & \cup \bigcup_{(\tau_1',\tau_1)\in\mathcal{T}_1} \Delta^P_{\tau,\tau_1'} Z_{\tau_1,\tau'} \quad \cup \quad \Delta^P_{\tau,\tau'} \\ & \cup \bigcup_{(\tau_2',\tau_2)\in\mathcal{T}_2} Z_{\tau,\tau_2'} \nabla^P_{\tau_2,\tau'} \quad \cup \quad \nabla^P_{\tau,\tau'} \\ & \cup \quad Z_{\tau,\tau'} \end{aligned}$$

where the sets $\mathcal{T}_1$, $\mathcal{T}_2$ and $\mathcal{T}_3$ are defined by

$$\begin{aligned} \mathcal{T}_1 &= \{(\tau',\tau) \mid \exists q \in P \ \ \tau' = P \to q, \tau \in \mathrm{Out}(q) \text{ and } \tau \neq q \to P\} \\ \mathcal{T}_2 &= \{(\tau',\tau) \mid \exists q \in P \ \ \tau' \in \mathrm{In}(q), \tau' \neq P \to q \text{ and } \tau = q \to P\} \\ \mathcal{T}_3 &= \{(\tau',\tau) \mid \exists q \in P \ \ \tau' = P \to q \text{ and } \tau = q \to P\}. \end{aligned}$$

Note that if m is not the first element of $M \cup \hat{M}$, then τ belongs to $T_1 = \{\tau \mid \exists\tau' \ \ (\tau',\tau) \in \mathcal{T}_1\}$ and if m is not the last element of $M \cup \hat{M}$, then τ' belongs to $T_2' = \{\tau' \mid \exists\tau \ \ (\tau',\tau) \in \mathcal{T}_2\}$. Otherwise, the set $Y_{\tau,\tau'}$ is such that $\tau = \sigma$ or $\tau' = \sigma'$. Moreover, when $\tau = \sigma$ and $\tau' = \sigma'$, the definition of $Z_{\tau,\tau'}$ given by Equation (12) must be slightly changed into $Z_{\sigma,\sigma'} = \Gamma^P_{\sigma,\sigma'}$ due to the content equal to P.

Let us study the elements $m = [c, c']$ of $\hat{M}$ such that $c = c'$. If $\hat{M}^*$ contains an element $[c, c]$, then $M \neq \varnothing$ and there are in $\mathcal{A}$ two transitions $P \to q$ and $q \to P$ for some state q. Analogously, if the first element of $\hat{M}$ is equal to $[c, c]$, then $M \neq \varnothing$ and the first transition σ is equal to $p \to P$ (similarly for the last element of $\hat{M}$).

We now come back to the label $x \in \Pi^P_{\sigma,\sigma'}$ of the path γ. If $M = \varnothing$, then $\hat{M}$ is a singleton whose sole element is an interval $[c, c']$ with $c \neq c'$. It follows that $x \in Y_{\sigma,\sigma'}$. If $M \neq \varnothing$, we decompose x thanks to the rational operation $\diamond$ used on $M \cup \hat{M}^*$. Different cases have to be considered depending on the transitions σ and σ'. Define the two sets X and Y by

$$X = \bigcup_{\substack{q\in P\\ \tau = q\to P\\ \tau' = P \to q'}} X_{\tau,\tau'} \qquad \text{and} \qquad Y = \bigcup_{\substack{\tau\in T_1\\ \tau'\in T_2'}} Y_{\tau,\tau'}.$$

Add to Y the empty word if there are in $\mathcal{A}$ two transitions $P \to q$ and $q \to P$ for some state $q \in P$. Define the set $Y_\sigma = \bigcup_{\tau'\in T_2'} Y_{\sigma,\tau'}$ if $\sigma \neq p \to P$ and $Y_\sigma = \varepsilon$ if $\sigma = p \to P$. Define also the set $Y_{\sigma'} = \bigcup_{\tau\in T_1} Y_{\tau,\sigma'}$ if $\sigma' \neq P \to p'$ and $Y_{\sigma'} = \varepsilon$ if $\sigma' = P \to p'$. Then the label x belongs to $Y_\sigma(X \diamond Y)Y_{\sigma'}$ showing the inclusion

$$\Pi^P_{\sigma,\sigma'} \subseteq Y_{\sigma,\sigma'} \cup Y_\sigma(X \diamond Y)Y_{\sigma'}.$$

It can be verified that the other inclusion holds. Clearly, the set $Y_{\sigma,\sigma'}$ is included in $\Pi^P_{\sigma,\sigma'}$. One checks that the right limit transitions $q \to P$ and the left limit transitions $P \to q'$ involved in the operation $\diamond$ are well managed thanks to the conditions imposed by T_1 and T'_2.

Therefore, $\Pi^P_{\sigma,\sigma'}$ is expressed as a rational expression on the sets $\nabla^P_{\tau,\tau'}$, $\Delta^P_{\tau,\tau'}$, $\Gamma^P_{\tau,\tau'}$ and $\Pi^R_{\tau,\tau'}$ with $R \subsetneq P$. This completes the proof.

7 Conclusion

In this paper, we have introduced automata and rational expressions for words on linear orderings. We have proved that for words on countable scattered linear orderings, these two notions are equivalent. This result extends the usual Kleene's theorem for finite words.

We mention some open problems. A natural generalization of the result would be to remove the restrictions on the orderings, first considering words on countable linear orderings and then words on all linear orderings. Automata that we have introduced are suitable for all linear orderings. It seems however that new rational operations are then needed. An operation like the η-shuffle introduced in [16] is necessary.

Automata on infinite words were introduced by Büchi to prove the decidability of the monadic second-order theory of the positive integers [6]. Since then, automata and logics have been shown to have strong connections [24]. The next step is to investigate the connections between logics and the automata that we have introduced. Such a study has to begin with the closure of the class of recognizable sets under the boolean operations. On the other hand, it is known that the monadic second-order theory of all linear orderings is undecidable [22].

Since the submission of this article, new related results have been obtained. Very recently, Carton and Rispal have proved that the class of recognizable sets introduced in this article is closed under complementation [11, 10]. The emptiness problem for automata on scattered countable linear orderings has been shown to be decidable with a polynomial time complexity in [9]; the equivalence problem has been proved to be decidable in [5].

Acknowledgments

We would like to thank the anonymous referees for their highly relevant comments and very helpful suggestions about the preliminary version of this paper.

References

[1] E. Asarin, P. Caspi, and O. Maler. A Kleene theorem for timed automata. In *Proceedings, Twelth Annual IEEE Symposium on Logic in Computer Science*, pages 160–171, 1997.

[2] N. Bedon and O. Carton. An Eilenberg theorem for words on countable ordinals. In C. L. Lucchesi and A. V. Moura, editors, *Latin'98: Theoretical Informatics*, volume 1380 of *Lect. Notes in Comput. Sci.*, pages 53–64. Springer-Verlag, 1998.

[3] B. Bérard and C. Picaronny. Accepting Zeno words without making time stand still. In *Mathematical Foundations of Computer Science 1997*, volume 1295 of *Lect. Notes in Comput. Sci.*, pages 149–158, 1997.

[4] V. Bruyère and O. Carton. Automata on linear orderings. In J. Sgall, A. Pultr, and P. Kolman, editors, *MFCS'2001*, volume 2136 of *Lect. Notes in Comput. Sci.*, pages 236–247, 2001.

[5] V. Bruyère, O. Carton, and G. Sénizergues. Tree automata and automata on linear orderings. In T. Harju and J. Karhumäki, editors, *WORDS'2003*, volume 27, pages 222–231. Turku Center for Computer Science, 2003.

[6] J. R. Büchi. Weak second-order arithmetic and finite automata. *Z. Math. Logik und grundl. Math.*, 6:66–92, 1960.

[7] J. R. Büchi. On a decision method in the restricted second-order arithmetic. In *Proc. Int. Congress Logic, Methodology and Philosophy of science, Berkeley 1960*, pages 1–11. Stanford University Press, 1962.

[8] J. R. Büchi. Transfinite automata recursions and weak second order theory of ordinals. In *Proc. Int. Congress Logic, Methodology, and Philosophy of Science, Jerusalem 1964*, pages 2–23. North Holland, 1965.

[9] O. Carton. Accessibility in automata on scattered linear orderings. In K. Diks and W. Rytter, editors, *MFCS'2002*, volume 2420 of *Lect. Notes in Comput. Sci.*, pages 155–164, 2002.

[10] O. Carton and C. Rispal. Complementation of rational sets on countable scattered linear orderings. In *DLT'2004*, volume 3340 of *Lect. Notes in Comput. Sci.*, pages 381–392, 2004.

[11] O. Carton and C. Rispal. Complementation of rational sets on scattered linear orderings of finite rank. In *LATIN'2004*, volume 2976 of *Lect. Notes in Comput. Sci.*, pages 292–301, 2004.

[12] Y. Choueka. Finite automata, definable sets, and regular expressions over ω^n-tapes. *J. Comput. System Sci.*, 17(1):81–97, 1978.

[13] B. Courcelle. Frontiers of infinite trees. *RAIRO Theoretical informatics*, 12(4):319–337, 1978.

[14] D. Girault-Beauquier. Bilimites de langages reconnaissables. *Theoret. Comput. Sci.*, 33(2–3):335–342, 1984.

[15] M. R. Hansen, P. K. Pandya, and Z. Chaochen. Finite divergence. *Theoret. Comput. Sci.*, 138(1):113–139, 1995.

[16] S. Heilbrunner. An algorithm for the solution of fixed-point equations for infinite words. *RAIRO Theoretical informatics*, 14(2):131–141, 1980.

[17] S. C. Kleene. Representation of events in nerve nets and finite automata. In C.E. Shannon, editor, *Automata studies*, pages 3–41. Princeton university Press, Princeton, 1956.

[18] M. Nivat and D. Perrin. Ensembles reconnaissables de mots bi-infinis. In *Proceedings of the Fourteenth Annual ACM Symposium on Theory of Computing*, pages 47–59, 1982.

[19] D. Perrin. Finite automata. In J. van Leeuwen, editor, *Handbook of Theoretical Computer Science*, volume B, chapter 1, pages 1–57. Elsevier, 1990.

[20] D. Perrin and J.-É. Pin. *Infinite Words*. Elsevier, 2004.

[21] J. G. Rosenstein. *Linear Orderings*. Academic Press, New York, 1982.

[22] S. Shelah. The monadic theory of order. *Annals of Mathematics*, 102:379–419, 1975.

[23] W. Thomas. On frontiers of regular sets. *RAIRO Theoretical informatics*, 20:371–381, 1986.

[24] W. Thomas. Automata on infinite objects. In J. van Leeuwen, editor, *Handbook of Theoretical Computer Science*, volume B, chapter 4, pages 133–191. Elsevier, 1990.

[25] J. Wojciechowski. Finite automata on transfinite sequences and regular expressions. *Fundamenta informaticæ*, 8(3-4):379–396, 1985.