Complementation of rational sets on countable scattered linear orderings

Chloé Rispal[1] and Olivier Carton[2]

[1] IGM, Université de Marne-la-Vallée, 5 boulevard Descartes, F-77454 Marne-la-Vallée Cedex 2, France,
chloe.rispal@univ-mlv.fr,
[2] LIAFA, Université Paris 7, 2, place Jussieu, F-75251 Paris cedex 05, France,
Olivier.Carton@liafa.jussieu.fr

Abstract. In a preceding paper (Bruyère and Carton, automata on linear orderings, MFCS'01), automata have been introduced for words indexed by linear orderings. These automata are a generalization of automata for finite, infinite, bi-infinite and even transfinite words studied by Büchi. Kleene's theorem has been generalized to these words. We prove that rational sets of words on countable scattered linear ordering are closed under complementation using an algebraic approach.

1 Introduction

In his seminal paper [12], Kleene showed that automata on finite words and regular expressions have the same expressive power. Since then, this result has been extended to many classes of structures like infinite words [6, 15], bi-infinite words [10, 16], transfinite words [8, 1], traces, trees, pictures...

In [4], automata accepting linear-ordered structures have been introduced with corresponding rational expressions. These linear structures include finite words, infinite, transfinite words and their mirrors. These automata are usual automata on finite words, extended with limit transitions. A Kleene-like theorem was proved for words on countable scattered linear orderings. Recall that an ordering is scattered if it does not contain a dense subordering isomorphic to $\mathbb{Q}$.

For many structures, the class of rational sets is closed under many operations like substitutions, inverse substitutions and boolean operations. As for boolean operations, the closure under union and intersection are almost always easy to get. The closure under complementation is often much more difficult to prove. This property is important both from the practical and the theoretical point of view. It means that the class of rational sets forms an effective boolean algebra. It is used whenever some logic is translated into automata. For instance, in both proofs of the decidability of the monadic second-order theory of the integers by Büchi [7] and the decidability of the monadic second-order theory of the infinite binary tree by Rabin [19], the closure under complementation of automata is the key property.

In [4], the closure under complementation was left as an open problem. In this paper, we solve that problem in a positive way. We show that the complement of a rational set of words on countable scattered linear orderings is also rational.

The classical method to get an automaton for the complement of a set of finite words accepted by an automaton $\mathcal{A}$ is through determinization. It is already non-trivial that the complement of a rational set of infinite words is also rational. The determinization method cannot be easily extended to infinite words. In his seminal paper [7], Büchi used another approach based on a congruence on finite words and Ramsey's theorem. This method is somehow related to our algebraic approach. McNaughton extended the determinization method to infinite words [13] proving that any Büchi automaton is equivalent to a deterministic Muller automaton. Büchi pushed further this method and extended it to transfinite words [8]. It is then very complex. In [3], the algebraic approach was used to give another proof of the closure under complementation for transfinite words. In [9], we have already proved the result for words on countable scattered linear orderings of finite ranks. The determinization method cannot be applied because any automaton is not equivalent to a deterministic one. In that paper, we extended the method used by Büchi in [7] using an additional induction on the rank. Since ranks of countable scattered linear orderings range over all countable ordinals, this approach is not suitable for words on all these orderings. In this paper, we prove the whole result for all countable scattered linear orderings using an algebraic approach. We define a generalization of semigroups, called $\diamond$-semigroups. We show that, when finite, these $\diamond$-semigroups are equivalent to automata. We also show that, by analogy with the case of finite words, a canonical $\diamond$-semigroup, called the syntactic $\diamond$-semigroup, can be associated with any rational set X. It has the property of being the smallest $\diamond$-semigroup recognizing X. A continuation of this paper would be to extend the equivalence between star-free sets, first order logic and aperiodic semigroups [22, 14, 2] and also between rational sets and the monadic second order theory.

Both hypotheses that the orderings are scattered and countable are really necessary. Büchi already pointed out that rational sets of transfinite words of length greater that ω_1 (the least non-countable ordinal) are not closed under complement. It can be proved that the set of words on scattered linear orderings is not rational as a subset of words on all linear orderings although its complement is rational.

Our proof of the complementation closure is effective. Given an automaton $\mathcal{A}$, it gives another automaton $\mathcal{B}$ that accepts words that are not accepted by $\mathcal{A}$. It gives another proof of the decidability of the equivalence of these automata [5].

This paper is organized as follows. Definitions concerning linear orderings and rational sets are first recalled in Sections 2 and 3. Then, Section 4 introduces the algebraic structure of $\diamond$-semigroup. The proof of equivalence between finite $\diamond$-semigroups and automata is sketched in both directions in Sections 5 and 6. Finally, the syntactic $\diamond$-semigroup corresponding to a rational set is defined in Section 7.

2 Words on linear orderings

This section recalls basic definitions on linear orderings but the reader is referred to [21] for a complete introduction. Hausdorff's characterization of countable scattered linear orderings is given and words indexed by linear orderings are introduced.

Let J be a set equipped with an order $<$. The ordering J is *linear* if for any j and k in J, either $j < k$ or $k < j$. Let A be a finite alphabet. A *word* $x = (a_j)_{j\in J}$ indexed by a linear ordering J is a function from J to A. J is called the *length* of x. For instance ω is the length of right-infinite words $a_0a_1...$ and ζ is the length of bi-infinite words $...a_{-1}a_0a_1...$.

2.1 Product of words indexed by linear orderings

For any linear ordering J, we denote by $-J$ the backward linear ordering that is the set J equipped with the reverse ordering. For instance, $-\omega$ is the linear ordering of negative integers.

The sum $J + K$ of two linear orderings is the set $J \cup K$ equipped with the ordering $<$ extending the orderings of J and K by setting $j < k$ for any $j \in J$ and $k \in K$. Formally, the *sum* $\sum_{j\in J} K_j$ is the set of all pairs (k, j) such that $k \in K_j$ equipped with the ordering defined by $(k_1, j_1) < (k_2, j_2)$ if and only if $j_1 < j_2$ or ($j_1 = j_2$ and $k_1 < k_2$ in K_{j_1}).

The sum of linear orderings helps to define the products of words. Let J be a linear ordering and let $(x_j)_{j\in J}$ be words of respective length K_j for any $j \in J$. The word $x = \prod_{j\in J} x_j$ obtained by concatenation of the words x_j with respect to the ordering on J is of length $L = \sum_{j\in J} K_j$. For instance, if for any $j \in \omega$, we denote by $x_j = a^{\omega^j}$, then $x = \prod_{j\in\omega} x_j$ is the word $x = a^{\omega^\omega}$ of length $\sum_{j\in\omega} \omega^j = \omega^\omega$. The sequence $(x_j)_{j\in J}$ of words is called a *J-factorization* of the word $x = \prod_{j\in J} x_j$.

2.2 Scattered linear orderings

A linear ordering J is *dense* if for any j and k in J such that $j < k$, there exists an element i of J such that $j < i < k$. It is *scattered* if it contains no dense subordering. The ordering ω of natural integers and the ordering ζ of relative integers are scattered. More generally, ordinals are scattered orderings. We denote by $\mathcal{N}$ the subclass of finite linear orderings, $\mathcal{O}$ the class of countable ordinals and of $\mathcal{S}$ the class of countable scattered linear orderings. The following characterization of scattered orderings is due to Hausdorff.

Theorem 1. *[Hausdorff [11]] A countable linear ordering J is scattered if and only if J belongs to $\bigcup_{\alpha\in\mathcal{O}} V_\alpha$ where the classes V_α are inductively defined by:*

1. $V_0 = \{\mathbf{0}, \mathbf{1}\}$

2. $V_\alpha = \{\sum_{j\in J} K_j \mid J \in \mathcal{N} \cup \{\omega, -\omega, \zeta\}$ *and* $K_j \in \bigcup_{\beta<\alpha} V_\beta\}$.

where $\mathbf{0}$ *and* $\mathbf{1}$ *are respectively the orderings with zero and one element.*

In order to simplify the proofs, we use slightly different inductive classes: For any $\alpha \in \mathcal{O}$, the class W_α is defined by : $W_\alpha = \{\sum_{j\in J} K_j \mid J \in \mathcal{N}$ and $K_j \in V_\alpha\}$. The inclusions $V_\alpha \subset W_\alpha \subset V_{\alpha+1}$ hold for any ordinal α thus, using Theorem 1, scattered linear orderings can be defined from the classes W_α by: $\mathcal{S} = \bigcup_{\alpha\in\mathcal{O}} W_\alpha$. The *rank* of a linear ordering J is the smallest ordinal α such that $J \in W_\alpha$. We denote by $A^\diamond$ the set of all words over A indexed by countable scattered linear orderings.

3 Rational sets of words on linear orderings

Bruyère and Carton have introduced rational expressions and automata for words indexed by countable scattered linear orderings. They have proved that a set of words is rational if and only if it is accepted by a finite automaton extending Kleene's theorem. This section shortly recalls definitions of rational operations and automata but the reader is referred to [4] for more details.

3.1 Rational expressions

Let A be a finite alphabet. The set $Rat(A^\diamond)$ of rational sets of words over A indexed by countable scattered linear orderings is the smallest set containing $\{a\}$ for any $a \in A$ and closed under the following rational operations defined for any subsets X and Y of $A^\diamond$ by :

$$X + Y = \{z \mid z \in X \cup Y\}$$

$$X \cdot Y = \{x \cdot y \mid x \in X, y \in Y\} \qquad X^* = \{\prod_{j=1}^{n} x_j \mid n \in \mathcal{N}, x_j \in X\}$$

$$X^\omega = \{\prod_{j\in\omega} x_j \mid x_j \in X\} \qquad X^{-\omega} = \{\prod_{j\in-\omega} x_j \mid x_j \in X\}$$

$$X^\# = \{\prod_{j\in\alpha} x_j \mid \alpha \in \mathcal{O}, x_j \in X\} \qquad X^{-\#} = \{\prod_{j\in-\alpha} x_j \mid \alpha \in \mathcal{O}, x_j \in X\}$$

$$X \diamond Y = \{\prod_{j\in J\cup\hat{J}^*} z_j \mid J \in \mathcal{S}\setminus\emptyset, z_j \in X \text{ if } j \in J \text{ and } z_j \in Y \text{ if } j \in \hat{J}^*\} \text{ where}$$

$$\hat{J}^* = \hat{J} \setminus \{(\emptyset, J), (J, \emptyset)\}.$$

The notation $\hat{J}$ is defined in the next section.

3.2 Automata on linear orderings

An automaton on linear orderings is a classical finite automaton with additional limit transitions of the form $P \longrightarrow q$ or $q \longrightarrow P$ where P is a set of states.

Definition 1. *An* automaton $\mathcal{A} = (Q, A, E, I, F)$ on linear orderings *is defined by a finite set of states* Q*, a finite alphabet* A*, a set of transitions* $E \subseteq (Q \times A \times Q) \cup (\mathcal{P}(Q) \times Q) \cup (Q \times \mathcal{P}(Q))$ *and initial and final sets of states* $I \subseteq Q$ *and* $F \subseteq Q$*.*

The definition of paths is based on the notion of cut that we explain now: Let x be a word indexed by an ordering $J \in \mathcal{S}$. To any two-factorization $x = yz$ of x, one can associate a partition of J into two intervals (K, L) such that $|y| = K$ and $|z| = L$. Such a partition is called a *cut* of J. The set $\hat{J} = \{(K, L) | K \cup L = J \wedge \forall k \in K, \forall l \in L, k < l\}$ is the set of cuts of the ordering J. Then, a path labelled x is a function from the set $\hat{J}$ into the set of states. As the set $\hat{J}$ is naturally equipped with the ordering $(K_1, L_1) < (K_2, L_2)$ if and only if $K_1 \subset K_2$, a path labelled by a word of length J is a word over Q of length $\hat{J}$.

Let $\gamma = (q_c)_{c \in \hat{J}}$ be a word of length $\hat{J}$ over Q, the limit sets of states of γ at a given cut c of $\hat{J}$ are defined by:

$$\lim_{c^-} \gamma = \{q \in Q |\ \forall c^{'} < c,\ \exists c^{''} \quad c^{'} < c^{''} < c \text{ and } q = q_{c^{''}}\}$$

$$\lim_{c^+} \gamma = \{q \in Q |\ \forall c^{'} > c,\ \exists c^{''} \quad c < c^{''} < c^{'} \text{ and } q = q_{c^{''}}\}$$

Definition 2. *Let* $\mathcal{A} = (Q, A, E, I, F)$ *be an automaton on linear orderings and let* $x = (a_j)_{j \in J}$ *be a word of length* J *on* A*. A* path γ *of label* x *in* $\mathcal{A}$ *is a word* $\gamma = (q_c)_{c \in \hat{J}}$ *of length* $\hat{J}$ *over* Q *such that for any* $(K, L) \in \hat{J}$*:*

- *If there exists* $l \in L$ *such that* $(K \cup \{l\}, L \setminus \{l\}) \in \hat{J}$ *then* $q_{(K,L)} \xrightarrow{a_l} q_{(K \cup \{l\}, L \setminus \{l\})} \in E$ *else* $\lim_{(K,L)^-} \gamma \rightarrow q_{(K,L)} \in E$*.*
- *If there exists* $k \in K$ *such that* $(K \setminus \{k\}, L \cup \{k\}) \in \hat{J}$ *then* $q_{(K \setminus \{k\}, L \cup \{k\})} \xrightarrow{a_k} q_{(K,L)} \in E$ *else* $q_{(K,L)} \rightarrow \lim_{(K,L)^+} \gamma \in E$*.*

Thus, if a cut has a predecessor or a successor, usual transitions are used, otherwise the path uses limit transitions.

As $\hat{J}$ has the least element $(\emptyset, J)$ and the greatest element $(J, \emptyset)$ for any linear ordering J, a path has always a first and a last state. A word is *accepted* by an automata if it is the label of a path leading from an initial state to a final state. We denote by $p \stackrel{x}{\Longrightarrow} q$ the existence of a path leading from the state p to the state q of label x.

It has been proved in [4] that automata and rational expressions have the same expressive power.

Theorem 2. *[4] A set of words indexed by countable scattered linear orderings is rational if and only if it is accepted by a finite automata.*

4 Algebraic characterization of rational sets

A semigroup is a set S equipped with an associative binary product. The semigroup S in which had been added a neutral element is denoted by S^1. An element

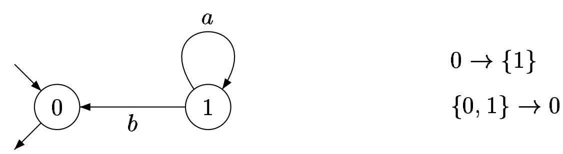

Fig. 1. Automaton on linear orderings accepting the set $(a^{-\omega}b)^{\#}$.

$e \in S$ is an *idempotent* if $e^2 = e$ and the set of idempotents of S is denoted by $E(S)$. A pair $(s, e) \in S \times S$ is *right linked* (respectively left linked) if $e \in E(S)$ and $se = s$ (respectively $es = s$). Two right linked pairs (s_1, e_1) and (s_2, e_2) are *conjugated* if there exists a, $b \in S^1$ such that $e_1 = ab$, $e_2 = ba$, $s_1 a = s_2$ and $s_2 b = s_1$. The conjugacy relation is an equivalence relation on right linked pairs [17].

4.1 $\diamond$-semigroups

The product of semigroups is generalized to recognize sets of words indexed by countable scattered linear orderings. A $\diamond$-semigroup is a generalization of a usual semigroup. The product of a sequence indexed by any scattered ordering is defined.

Definition 3. *A* $\diamond$-semigroup *is a set S equipped with product $\pi : S^\diamond \longrightarrow S$ which maps any word of countable scattered linear length over S to an element of S.*

- *for any element s of S, $\pi(s) = s$.*
- *for any word x over S of countable scattered linear length and for any factorization $x = \prod_{j \in J} x_j$ where $J \in \mathcal{S}$,*

$$\pi(x) = \pi(\prod_{j \in J} \pi(x_j))$$

The latter condition is a generalization of associativity.

For instance, the set $A^\diamond$ equipped with the concatenation is a $\diamond$-semigroup.

Example 1. The set $S = \{0, 1\}$ equipped with the product π defined for any $u \in S^\diamond$ by $\pi(u) = 0$ if u has at least one occurence of the letter 0 and $\pi(u) = 1$ otherwise is a $\diamond$-semigroup.

For any two elements s and t of a $\diamond$-semigroup (S, π), the finite product $\pi(st)$ is merely denoted by st.

A *sub-$\diamond$-semigroup* T of a $\diamond$-semigroup S is a subset of S closed under product. A *morphism of $\diamond$-semigroup* is an application which preserves the product. A *congruence* of $\diamond$-semigroup is an equivalence relation $\sim$ stable under product: If

$s_j \sim t_j$ for any $j \in J$, then $\pi(\prod_{j\in J} s_j) \sim \pi(\prod_{j\in J} t_j)$. The set $S/\!\sim$ is a $\diamond$-semigroup. A $\diamond$-semigroup T is a *quotient* of a $\diamond$-semigroup S if there exists an onto morphism from S to T. A $\diamond$-semigroup T *divides* S if T is the quotient of a sub-$\diamond$-semigroup of S.

4.2 Finite $\diamond$-semigroups

A $\diamond$-semigroup (S, π) is said to be finite if S is finite. Even when S is finite, the function π is not easy to describe because the product of any sequence has to be given. It turns out that the function π can be described using a semigroup structure on S with two additional functions (called τ and $-\tau$) from S to S. This gives a finite description of the function π. The functions τ and $-\tau$ are the counterpart of limit transitions of automata. This finite description is based on the next Lemma which follows directly from Ramsey's Theorem [20].

Let $x = \prod_{i\in\omega} x_i$ an ω-factorization. Another factorization $x = \prod_{i\in\omega} y_i$ is called a *superfactorization* if there is a sequence $(k_i)_{i\in\omega}$ of integers such that $y_0 = x_0 \dots x_{k_0}$ and $y_i = x_{k_{i-1}+1} \dots x_{k_i}$ for all $i \geq 1$.

Lemma 1. *Let $\varphi : A^\diamond \longrightarrow S$ be a morphism into a finite $\diamond$-semigroup. For any factorization $x = \prod_{i\in\omega} x_i$, there exists a superfactorization $x = \Pi_{i\in\omega} y_i$ and a right linked pair $(s, e) \in S \times E(S)$ such that $\varphi(y_0) = s$ and $\varphi(y_i) = e$ for any $i > 0$.*

Such a factorization is called a *ramseyan factorization*, see Theorem 3.2 in [18].

Definition 4. *Let S be a semigroup. A function $\tau : S \longrightarrow S$ (respectively $-\tau : S \longrightarrow S$) is* compatible to the right *with S (respectively* to the left*) if and only if for any s, t in S and any integer n the following properties hold: $s(ts)^\tau = (st)^\tau$ and $(s^n)^\tau = s^\tau$ (respectively $(st)^{-\tau}s = (ts)^{-\tau}$ and $(s^n)^{-\tau} = s^{-\tau}$).*

The product of a finite $\diamond$-semigroup S can be finitely described by functions compatible to the right and to the left with S.

Theorem 3. *Let (S, π) be a finite $\diamond$-semigroup. The binary product defined for any s, t in S by $s \cdot t = \pi(st)$ naturally endows a structure of semigroup and the functions τ and $-\tau$ respectively defined by $s^\tau = \pi(s^\omega)$ and $s^{-\tau} = \pi(s^{-\omega})$ are respectively compatible to the right and to the left with S.*
Conversely, let S be a finite semigroup and let τ and $-\tau$ be functions respectively compatible to the right and to the left with S. Then S can be uniquely endowed with a structure of $\diamond$-semigroup (S, π) such that $s^\tau = \pi(s^\omega)$ and $s^{-\tau} = \pi(s^{-\omega})$.

The first part of the theorem follows directly from the associativity of the product π. Conversely, let S be a finite semigroup and let τ and $-\tau$ be functions respectively compatible to the right and to the left with S. The product of a word $x = (s_j)_{j\in J}$ over S of length $J \in \mathcal{S}$ is defined by induction on $\alpha \in \mathcal{O}$ for any $J \in W_\alpha$ by the following way:

Let $J \in W_0$ and let $x \in S^J$. There exists an integer m and $s_1, \dots, s_m$ in S such that $x = s_1 \dots s_m$. We set $\pi(x) = s_1 \cdot s_2 \dots s_m$.

Let $J \in W_\alpha$ where $\alpha > 1$ and let $x \in S^J$. The linear ordering J can be decomposed as a sum $J = \sum_{i \in I} K_i$ where $I \in \mathcal{N} \cup \{\omega, -\omega\}$ and for all $i \in I$, $K_i \in \bigcup_{\beta < \alpha} W_\beta$. There exists a factorization $x = \prod_{i \in I} x_i$ such that for all $i \in I$, $|x_i| = K_i$.

- $J = \{1, \ldots, m\} \in \mathcal{N}$: we set $\pi(x) = \pi(x_1) \ldots \pi(x_m)$.
- $J = \omega$: There exists a superfactorization $x = \mathit{\Pi}_{i \in \omega} y_i$ and a right linked pair $(s, e) \in S \times E(S)$ such that $\varphi(y_0) = s$ and $\varphi(y_i) = e$ for any $i > 0$. We set $\pi(x) = se^\tau$.
- $J = -\omega$: Symmetrically to the previous case, we set $\pi(x) = e^{-\tau}s$.

Since two linked pairs associated with two factorizations of a word are conjugated [18], it can be proved by induction on α that π is uniquely defined and associative on $S^\diamond$.

Example 2. The $\diamond$-semigroupe $S = \{0, 1\}$ of Example 1 is defined by the finite product $00 = 01 = 10 = 0$ and $11 = 1$ and by the compatible functions τ and $-\tau$ defined by $0^\tau = 0^{-\tau} = 0$ and $1^\tau = 1^\tau = 1$.

4.3 Recognizability

It is well known that rational sets of finite words are exactly those recognized by finite semigroups. This result is generalized for words indexed by countable scattered linear orderings.

Definition 5. *Let S and T be two $\diamond$-semigroups. The $\diamond$-semigroup T* recognizes *a subset X of S if and only if there exists a morphism $\varphi : S \longrightarrow T$ and a subset $P \subseteq T$ such that $X = \varphi^{-1}(P)$. A set $X \subseteq A^\diamond$ is* recognizable *if and only if there exists a* ***finite*** *$\diamond$-semigroup recognizing it.*

Example 3. The set $S = \{0, 1\}$ equipped with the product π defined for any $u \in S^\diamond$ by $\pi(u) = 1$ if $u \in 1^\#$ and $\pi(u) = 0$ otherwise is a $\diamond$-semigroup. It is also defined by the finite product $00 = 01 = 10 = 0$ et $11 = 1$ and by the compatible functions τ and $-\tau$ defined by $0^\tau = 0^{-\tau} = 1^{-\tau} = 0$ and $1^\tau = 1$. Define the morphism of $\diamond$-semigroup $\varphi : A^\diamond \longrightarrow S$ by $\varphi(a) = 1$ for any $a \in A$. The set $A^\#$ is recognizable since $A^\# = \varphi^{-1}(\{1\})$.

For any finite alphabet A, $Rec(A^\diamond)$ denotes the set of subsets of $A^\diamond$ recognizable by a finite $\diamond$-semigroup.

Theorem 4. *A set of words indexed by countable scattered linear orderings is rational iff it is recognizable.*

Example 4. The set $X = (ab)^\diamond$ is recognized by the $\diamond$-semigroup $S = \{s, t, e, f, 0\}$ whose product is defined by $st = e$, $ts = f$, $ee = e$, $ff = f$, $es = s$, $ft = t$, $sf = s$, $te = t$, $e^\tau = e$, $e^{-\tau} = e$, $f^\tau = t$, $f^{-\tau} = s$ where any other product is equal to 0. Defining the morphism $\varphi : A^\diamond \to S$ by $\varphi(a) = s$ and $\varphi(b) = t$, we get $X = \varphi^{-1}(e)$.

If X is recognized by a morphism $\varphi : S \longrightarrow T$, the set $A^\diamond \setminus X$ is also recognized by φ since $A^\diamond \setminus X = \varphi^{-1}(S \setminus P)$. Therefore, we obtain following theorem.

Theorem 5. *Rational sets of words on countable scattered linear orderings are closed under complementation.*

Example 5. The set $X = A^*$ is recognized by the $\diamond$-semigroup $S = \{0, 1\}$ whose product is defined by $11 = 1$, $01 = 10 = 00 = 0$ and by the compatible functions $0^\tau = 0^{-\tau} = 1^\tau = 1^{-\tau} = 0$. Define the morphism $\varphi : A^\diamond \to S$ by $\varphi(a) = 1$ for any $a \in A$. One gets $X = \varphi^{-1}(1)$ and the complement $A^\diamond \setminus X = (A^\diamond)^\omega A^\diamond + A^\diamond (A^\diamond)^{-\omega} = \varphi^{-1}(0)$.

The next section is devoted to sketches of proof of Theorem 4.

5 From $\diamond$-semigroups to automata

Let (S, π) be a finite $\diamond$-semigroup. By Theorem 3, the product π is defined by compatible functions τ and $-\tau$. Let X be a subset of $A^\diamond$ recognized by S. There exists a morphism of $\diamond$-semigroup $\varphi : A^\diamond \longrightarrow S$ and a subset P of S such that $X = \varphi^{-1}(P)$. Since rational sets are closed under finite union, one may suppose that P is a single element $\{p\}$. Let h be the finite substitution which associates to each element s of S the set $\varphi^{-1}(s) \cap A$. Since $X = h(\pi^{-1}(p) \cap \varphi(A)^\diamond)$, it suffices to prove that the set $\pi^{-1}(p)$ of words over S whose product is p, is rational. Recall that the Green's relations are defined from the following preorders:

$$
\begin{aligned}
s \leq_{\mathcal{R}} t &\iff \exists a \in S^1,\ s = ta \\
s \leq_{\mathcal{L}} t &\iff \exists a \in S^1,\ s = at \\
s \leq_{\mathcal{J}} t &\iff \exists a, b \in S^1,\ s = atb
\end{aligned}
$$

For any $\mathcal{K} \in \{\mathcal{R}, \mathcal{L}, \mathcal{J}\}$, $s\mathcal{K}t$ if and only if $s \leq_{\mathcal{K}} t$ and $t \leq_{\mathcal{K}} s$. We also denote by $s <_{\mathcal{K}} t$ iff $s \leq_{\mathcal{K}} t$ and not $t \leq_{\mathcal{K}} s$. Recall that the equivalence relation $\mathcal{D} = \mathcal{RL} = \mathcal{LR}$ is equal to $\mathcal{J}$ when S is finite.

The proof is by induction on the $\mathcal{D}$-class structure of S. For any $\mathcal{D}$-class D of S, denote by:

$$S_D = \{s \in S \mid \forall p \in D, s \geq_{\mathcal{J}} p\} \ \text{ and } \ T_D = \{s \in S \mid \forall p \in D, s >_{\mathcal{J}} p\}$$

We define an automaton on linear orderings accepting words over S_D and computing the product π of its path's labels in both directions.

Let $\mathcal{A}_D = (Q_D, S_D, E_D)$ be the automaton defined by:

$Q_D = S_D^1 \times S_D^1 \times \mathbb{B}$ is the set of states where $\mathbb{B} = \{0, 1\}$

$$E_D = \{(s, rt, b) \xrightarrow{r} (sr, t, b') \mid b \in \mathbb{B}, b' = (r \in D)\}$$

$$\cup\ \{\{(s_i, t_i, b_i)\}_{1 \leq i \leq m} \longrightarrow (s, t, b) \mid b \in \mathbb{B}, \exists 1 \leq i \leq m\ ,\ b_i = 1\ ,$$
$$\exists 1 \leq k \leq m\ ,\ \exists e \in E(D)\ ,\ s_k e = s_k\ ,\ e t_k = t_k\ ,\ s = s_k e^\tau \text{ and } t_k = e^\tau t\}$$

$$\cup\ \{(s, t, b) \longrightarrow \{(s_i, t_i, b_i)\}_{1 \leq i \leq m} \mid b \in \mathbb{B}, \exists 1 \leq i \leq m\ ,\ b_i = 1\ ,$$
$$\exists 1 \leq k \leq m\ ,\ \exists e \in E(D)\ ,\ s_k e = s_k\ ,\ e t_k = t_k\ ,\ t = e^{-\tau} t_k \text{ and } s_k = s e^{-\tau}\}$$

The boolean component of Q_D allows limit transitions only if the label of the path admits a ramseyan factorization associated to an idempotent of D. Since two right linked pairs (s_1, e_1) and (s_2, e_2) of a same $\mathcal{D}$-class are conjugated iff $s_1 \mathcal{R} s_2$, it can be shown by induction on the rank that $\mathcal{A}_D$ computes properly the product π. The words of $S_D^\diamond$ admitting ramseyan factorizations associated to idempotents of $\mathcal{J}$-above $\mathcal{D}$-classes are taken care by a substitution which is rational by induction. Let D be a $\mathcal{D}$-class of S and let f be the rational substitution defined by:

$$\begin{aligned} f : S_D &\longrightarrow Rat({S_D}^\diamond) \\ s &\longrightarrow \begin{cases} \pi^{-1}(s) & \text{if } s \in T_D \\ \{s\} \cup F_s \cup G_s & \text{if } s \in D \end{cases} \end{aligned}$$

where for any $s \in D$,

$$F_s = \bigcup_{\substack{s_1,\ldots,s_m >_{\mathcal{J}} s, \\ s_1 \ldots s_m = s}} \pi^{-1}(s_1) \ldots \pi^{-1}(s_m)$$

$$G_s = \bigcup_{\substack{t,e >_{\mathcal{J}} s, \\ te^\tau = s}} \pi^{-1}(t)\pi^{-1}(e)^\omega \cup \bigcup_{\substack{t,e >_{\mathcal{J}} s, \\ e^{-\tau} t = s}} \pi^{-1}(e)^{-\omega}\pi^{-1}(t).$$

If L_s denotes the set of words recognized by the automaton $\mathcal{A}_D$ with the initial state $\{(1, s, 0)\}$ and the final set of states $\{(s, 1, b) | \ b \in \mathbb{B}\}$ for any $s \in S_D$, it can be proved that for any $p \in D$, $f(L_p) = \pi^{-1}(p)$ using another induction on the rank.

6 From automata to $\diamond$-semigroups

This proof of the converse is adapted from [3]. Let $\mathcal{A} = (Q, A, E, I, F)$ be an automaton on linear orderings accepting a set $X \subseteq A^\diamond$. The *content* of a path is the set of states occurring in the path and $p \overset{x}{\underset{P}{\Longrightarrow}} q$ denotes a path leading from p to q of label x and of content P. Let $T = \mathcal{P}(Q)$ be the set of all subsets of Q and $K = \mathcal{P}(T)$ be the set of subsets of T. The set K is equipped with the following product and union:

$$kk' = \{t \cup t' \mid t \in k, \, t' \in k'\} \text{ and } k + k' = k \cup k'$$

Let S be the set of all $Q \times Q$ matrices whose entries are in K with product defined by:

$$(m \cdot m')_{q,q'} = \bigcup_{p \in Q} m_{q,p} \cdot m'_{p,q'} = \{t \cup t' \mid \exists p \in Q, \, t \in m_{q,p}, \, t' \in m'_{p,q'}\}$$

The semigroup S is finite and by Theorem 3, it suffices to define compatible functions to endow a structure of $\diamond$-semigroup. Define the function τ by :

$$m^\tau_{q,q'} = \{t \cup \{q'\} \mid \exists t' \subset t, \, \exists p \in Q, \, t \in m^\pi_{q,p}, \, t' \in m^\pi_{p,p} \text{ and } t' \longrightarrow q' \in E\}$$

where π is the smallest integer such that m^π is an idempotent matrix. The function $-\tau$ is defined symmetrically and it can be proved that τ and $-\tau$ are functions respectively compatible on the right and left with S. It remains to define a morphism $\varphi : A^\diamond \longrightarrow S$ recognizing X. For each letter a of A, we define the matrix $m_a = \varphi(a)$ corresponding to the edges of $\mathcal{A}$ labelled by a: The entry (q, q') of m_a is equal to $\{\{q, q'\}\}$ if $q \xrightarrow{a} q' \in E$ or $\emptyset$ otherwise. An induction on the rank would show that for all word $x \in A^\diamond$, $\varphi(x) = m$ where the matrix m memorizes the contents of paths labelled by x:

$$m_{q,q'} = \{l \mid q \underset{l}{\overset{x}{\Longrightarrow}} q'\}$$

A word $x \in A^\diamond$ belongs to X iff $\varphi(x)$ has a (i, f)non-empty entry where i and f are respectively initial and final states. Thus X is recognized by S.

7 Syntactic $\diamond$-semigroup

Let X be a recognizable subset of $A^\diamond$. Among all $\diamond$-semigroups recognizing X, there exists one which is minimal in the sense of division. It is called the syntactic $\diamond$-semigroup of X and is the first canonical object associated to rational sets on linear orderings. For any $\diamond$-semigroup (S, π) and any set $P \subseteq S$, the equivalence relation $\sim_P$ is defined for any s,t in S by $s \sim_P t$ iff for any integer m:

$$\forall s_1, s_2, \ldots, s_m, t_1, t_2, \ldots, t_m \in S^1 \ , \forall \theta_1, \theta_2, \ldots, \theta_{m-1} \in \{\omega, -\omega\} \cup \mathcal{N},$$

$$\pi(s_m(\ldots(s_2(s_1 s t_1)^{\theta_1} t_2)^{\theta_2} \ldots)^{\theta_{m-1}} t_m) \in P$$

$$\Longleftrightarrow \pi(s_m(\ldots(s_2(s_1 t t_1)^{\theta_1} t_2)^{\theta_2} \ldots)^{\theta_{m-1}} t_m) \in P$$

The equivalence relation $\sim_P$ is a congruence of $\diamond$-semigroup. If S finite, then and the quotient $S/{\sim_P}$ is an effective $\diamond$-semigroup.

If X is a recognizable subset of $A^\diamond$, then the quotient $A^\diamond/{\sim_X}$ is finite and recognizes X.

Proposition 1. *Let X be a subset of $A^\diamond$. The set X is recognizable if and only if the relation $\sim_X$ is a congruence of $\diamond$-semigroup of finite index.*

For any recognizable subset X of $A^\diamond$, the $\diamond$-semigroup $A^\diamond/\sim_X$ is called the *syntactic semigroup* of X and is denoted by $S(X)$. It is the smallest $\diamond$-semigroup recognizing X in the sense of division.

Proposition 2. *Let X be a recognizable set of $A^\diamond$ and let T be a $\diamond$-semigroup. Then T recognizes X if and only if $S(X)$ divides T.*

In particular, for any recognizable set X, the relation $\sim_X$ is the coarsest congruence such that the quotient $A^\diamond/\sim_X$ recognizes X. From Theorem 4 and Proposition 2, it follows that the syntatic $\diamond$-semigroup of a rational set is finite.

Theorem 6. *A set of words indexed by countable scattered linear orderings is rational iff its syntactic $\diamond$-semigroup is finite.*

References

1. N. Bedon. Finite automata and ordinals. *Theoret. Comput. Sci.*, 156:119–144, 1996.
2. N. Bedon. Star-free sets of words on ordinals. *Inform. Comput.*, 166:93–111, 2001.
3. N. Bedon and O. Carton. An Eilenberg theorem for words on countable ordinals. In Cláudio L. Lucchesi and Arnaldo V. Moura, editors, *Latin'98: Theoretical Informatics*, volume 1380 of *Lect. Notes in Comput. Sci.*, pages 53–64. Springer-Verlag, 1998.
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. IGM report 2001-12.
5. V. Bruyère, O. Carton, and G. Sénizergues. Tree automata and automata on linear orderings. In Tero Harju and Juhani Karhumäki, editors, *WORDS'2003*, volume 27, pages 222–231, 2003. Turku Center for Computer Science.
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 and C. Rispal. Complementation of rational sets on scattered linear orderings of finite rank. In *LATIN'04*, 2004.
10. D. Girault-Beauquier. Bilimites de langages reconnaissables. *Theoret. Comput. Sci.*, 33(2–3):335–342, 1984.
11. F. Hausdorff. Set theory. In *Chelsea*, New York, 1957.
12. 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.
13. R. McNaughton. Testing and generating infinite sequences by a finite automaton. *Inform. Comput.*, 9:521–530, 1966.
14. R. McNaughton and S. Papert. *Counter free automata.* MIT Press, Cambridge, MA, 1971.
15. D. Muller. Infinite sequences and finite machines. In Proc. of Fourth Annual IEEE Symp., editor, *Switching Theory and Logical Design*, pages 3–16, 1963.
16. 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.
17. J.-P. Pécuchet. Etude syntaxique des parties reconnaissables de mots infinis. volume 226, pages 294–303, 1986.
18. D. Perrin and J.-E. Pin. Infinite words. In Elsevier, editor, *Academic Press*, 2003.
19. M. O. Rabin. Decidability of second-order theories and automata on infinite trees. 141:1–35, 1969.
20. F. D. Ramsey. On a problem of formal logic. *Proc. of the London math. soc.*, 30:338–384, 1929.
21. J. G. Rosenstein. *Linear ordering.* Academic Press, New York, 1982.
22. M.-P. Schützenberger. On finite monoids having only trivial subgroups. *Inform. Control*, 8:190–194, 1965.