Birecurrent sets

Francesco Dolce[1], Dominique Perrin[2],
Antonio Restivo[3], Christophe Reutenauer[1], Giuseppina Rindone[2]

[1] Université du Québec à Montréal, LaCIM,
[2] Université Paris Est, LIGM,
[3] Università di Palermo

Abstract

A set is called recurrent if its minimal automaton is strongly connected and birecurrent if it is recurrent as well as its reversal. We prove a series of results concerning birecurrent sets. It is already known that any birecurrent set is completely reducible (that is, such that the minimal representation of its characteristic series is completely reducible). The main result of this paper characterizes completely reducible sets as linear combinations of birecurrent sets

Contents

1 Introduction

A recurrent set of words is such that its minimal automaton is strongly connected. A birecurrent set is a recurrent set such that its reversal is also recurrent. The recurrent (resp. birecurrent) sets contains the submonoids generated by prefix (resp. bifix) codes. The automata recognizing birecurrent are a generalization of well known families of automata such as group automata. There are more general sets and the general form of birecurrent sets does seem easy to describe.

Our interest in birecurrent sets is motivated by their relation with completely reducible sets put in evidence in [7]. Indeed, it is shown in [7] that the syntactic representation of the characteristic series of a birecurrent set is completely reducible. This generalizes the result of Reutenauer [10] which proves that the complete reducibility holds for the submonoid generated by a bifix code.

Our main result characterizes completely reducible sets as linear combinations of birecurrent sets (Theorem 4.3.4). We also prove a number of other results concerning birecurrent sets, and in particular birecurrent sets of finite type, which are a generalization of the submonoids generated by finite bifix codes.

The paper is organized as follows.

We first give in Section 2 a number of definitions concerning words, automata and formal series.

We introduce recurrent and birecurrent sets in Section 3. In Section 3.2 we prove a result which characterizes the minimal automata of birecurrent sets (Theorem 3.2.1). This result extends a property proved in [7] and allows to construct directly birecurrent automata by choosing an appropriate set of terminal states of a strongly connected deterministic automaton.

In Section 3.3, we define birecurrent sets of finite type. A recurrent set S is of the form $S = X^*P$ where X is a prefix code and P a set of proper prefixes of X. Thus a birecurrent set S has the form X^*P, as above, as well as $S = QY^*$ where Y is a suffix code and Q a set of proper suffixes of Y. We say that S is of finite type if X and Y are finite. Thus a birecurrent set of finite type is a generalization of a submonoid generated by a finite bifix code (which corresponds to the case where P and Q are reduced to the empty word). We prove a result allowing one to build birecurrent sets of finite type (Theorem 3.4.2).

In Section 3.6, we define the degree and the index of a dense birecurrent set. We prove that the density of a dense birecurrent set with respect to a positive Bernoulli distribution is the inverse of its index (Theorem 3.6.2).

In Section 3.7, we prove that if a recognizable maximal prefix code is indecomposable, then either it is synchronized, or it is the left root of a dense birecurrent set (Theorem 3.7.1). We relate this result with an old conjecture of Schützenberger.

In Section 4, we come to the connection with complete reducibility.

We start in Section 4.1 with an introduction to the linear representations of formal series. In Section 4.2, we prove a statement (Theorem 4.2.1) which characterizes completely reducible sets by a property of their syntactic monoid. We derive from this result several corollaries and, in particular, the main result of [7] asserting that a birecurrent set is completely reducible.

We then present in Section 4.3 our main result which characterizes completely reducible sets as linear combinations of birecurrent sets (Theorem 4.3.4).

In Section 5, we consider unambiguous automata. We characterize the unambiguous automata recognizing recurrent sets (Theorem 5.2.1). We derive as a corollary a characterization of unambiguous automata recognizing birecurrent sets (Corollary 5.2.2). We also give examples of such automata with an iteration of the construction of Section 3.4, using an argument originally developped in [15].

Acknowledgements The authors wish to thank for their help several colleagues, including Clelia De Felice, Andrew Ryzhikov and the anonymous referee.

2 Preliminaries

We recall briefly some terminology about words, automata and, in some more detail, formal series. We refer to [2] or [11] for undefined terms.

2.1 Words

Let A be a finite alphabet and let A^* be the free monoid over A. The elements of A^* are called *words* and the subsets of A^* *formal languages.* We denote by ε the empty word.

The *reversal* of a word $w = a_1 \cdots a_n$ is the word $\tilde{w} = a_n \cdots a_1$. By extension, the reversal of $X \subset A^*$ is the set $\tilde{X} = \{\tilde{w} \mid w \in X\}$.

For $u \in A^*$ and $X \subset A^*$, we denote $u^{-1}X = \{v \in A^* \mid uv \in X\}$. Symmetrically, we denote $Xv^{-1} = \{u \in A^* \mid uv \in X\}$.

A set $S \subset A^*$ is said to be *dense* in A^* if for any word $v \in A^*$, there are words $u, w \in A^*$ such that $uvw \in S$. A set which is not dense is said to be *thin.*

A set $S \subset A^*$ is said to be *right dense* in A^* if for any $v \in A^*$, there is $w \in A^*$ such that $vw \in S$.

2.2 Automata

We denote by $\mathcal{A} = (Q, I, T)$ an automaton with Q as set of states, I as set of initial states and T as set of terminal states, given by a set of edges which are triples $(p, a, q) \in Q \times A \times Q$. The automaton is said to be finite if Q is finite.

The set *recognized* by the automaton $\mathcal{A} = (Q, I, T)$ is the set of labels of paths from I to T. A set is *recognizable* if it can be recognized by a finite automaton.

The automaton $\mathcal{A} = (Q, I, T)$ is *deterministic* if $I = \{i\}$ and for each $p \in Q$ and $a \in A$ there is at most one edge (p, a, q). For $p \in Q$ and $a \in A$, we denote by $p \cdot a$ the unique state q such that there is an edge from p to q labeled a. Thus, the set recognized by $\mathcal{A}$ is the set of words w such that $i \cdot w \in T$. We denote $\mathcal{A} = (Q, i, T)$ and (Q, i, t) if $T = \{t\}$.

The *minimal automaton* of a set $X \subset A^*$ is the deterministic automaton having for states the nonempty sets $u^{-1}X$ for $u \in A^*$, with X as initial state and the family $\{u^{-1}X \mid u \in X\}$ as terminal states.

An automaton $\mathcal{A} = (Q, I, T)$ is *trim* if for any $q \in Q$ there are words u, v such that there is a path from I to q labeled u and a path from q to T labeled v.

A deterministic automaton $\mathcal{A} = (Q, i, T)$ on the alphabet A is *complete* if for any $q \in Q$ and any $a \in A$ one has $q \cdot a \neq \emptyset$. A set $S \subset A^*$ is right dense if and only if its minimal automaton is complete.

For a deterministic automaton $\mathcal{A} = (Q, i, T)$ and a word w, we denote by $\varphi_{\mathcal{A}}(w)$ the map $q \mapsto q \cdot w$ from Q into Q. The monoid $M_{\mathcal{A}} = \varphi_{\mathcal{A}}(A^*)$ is the *transition monoid* of the automaton $\mathcal{A}$. For $m \in M_{\mathcal{A}}$, denote $pm = q$ if the image of p by m is q, or equivalently if $m = \varphi_{\mathcal{A}}(w)$ and $p \cdot w = q$. Similarly, for $S \subset Q$, we denote $mS = \{p \in Q \mid pm \in S\}$.

The *rank* of a word w is the rank of the map $\varphi_{\mathcal{A}}(w)$, that is the cardinality of the set $\{p \cdot w \mid p \in Q\}$. When $\mathcal{A}$ is a strongly connected finite automaton, the image by $\varphi_{\mathcal{A}}$ of the set of words of minimal nonzero rank is, together with 0, the unique 0-minimal ideal J of the monoid $\varphi_{\mathcal{A}}(A^*)$. It is the union of all 0-minimal right (resp. left) ideals. It is formed of 0 and a regular $\mathcal{D}$-class (see [2, Corollary 1.12.10]). Each $\mathcal{H}$-class of J which is a group is a transitive permutation group on the common image of its elements (see [2, Theorem 9.3.10]).

Let $\mathcal{A} = (Q, i, T)$ be a deterministic automaton. The *domain* of a word w, denoted $\mathrm{Dom}(w)$, is the set $w \cdot Q$, that is, the set of $q \in Q$ such that $q \cdot w \neq \emptyset$. The *kernel* of a word w is the partial equivalence on $\mathrm{Dom}(w)$ defined by $p \equiv q$ if $p \cdot w = q \cdot w$.

The *degree* of a strongly connected deterministic automaton $\mathcal{A}$, denoted $d(\mathcal{A})$, is the minimal nonzero rank of all words. The automaton is *synchronized* if its degree is 1.

Example 1 Let $\mathcal{A}$ be the automaton represented in Figure 2.1. The word b is synchronizing since it has rank 1. The 0-minimal ideal of the monoid $\varphi_{\mathcal{A}}(A^*)$ is represented in Figure 2.2. We represent the $\mathcal{D}$-class by an array in which the

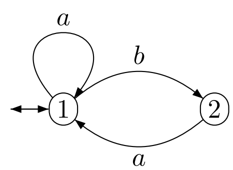

Figure 2.1: A synchronized automaton.

	1	2
$1,2$	$^{*}a$	$^{*}ab$
1	$^{*}ba$	b

Figure 2.2: The 0-minimal ideal of $\varphi_{\mathcal{A}}(A^*)$ in Example 1.

rows are the $\mathcal{R}$-classes and the columns are the $\mathcal{L}$-classes. An $\mathcal{H}$-class containing an idempotent is indicated by a $*$.

2.3 Prefix and bifix codes

A set $X \subset A^+$ is a *prefix code* if no word in X is a proper prefix of another word in X. The set X is a *bifix code* if X and $\tilde{X}$ are prefix codes.

Let $X \subset A^+$ be a prefix code. The *literal automaton* of X^* is the deterministic automaton $\mathcal{A} = (Q, \varepsilon, \varepsilon)$ where Q is the set of proper prefixes of X and for $q \in Q$ and $a \in A$, one has

$$q \cdot a = \begin{cases} qa & \text{if } qa \in Q \\ \varepsilon & \text{if } qa \in X \\ \emptyset & \text{otherwise.} \end{cases}$$

It is a strongly connected deterministic automaton recognizing X^*.

The *degree* $d(X)$ of a prefix code X is the degree of the minimal automaton of X^*. The prefix code X is said to be *synchronized* if $d(X) = 1$.

Example 2 The automaton of Example 1 is both the minimal and the literal automaton of X^* where X is the synchronized prefix code $\{a, ba\}$.

2.4 Formal series

Let K be a field. A *formal series* on the alphabet A with coefficients in K is a map $\sigma : A^* \to K$. We denote by $K\langle\langle A\rangle\rangle$ the algebra of these series. For $w \in A^*$, we denote by (σ, w) the value of σ on w. We denote by $K\langle A\rangle$ the corresponding ring of polynomials, which are the series with a finite number of nonzero coefficients.

We denote by $\underline{S}$ the characteristic series of a set $S \subset A^*$.

Let $n \geq 0$ be an integer. Let λ be a row n-vector, let μ be a morphism from A^* into the monoid of $n \times n$-matrices and let γ be a column n-vector, all with coefficients in K. The triple (λ, μ, γ) is called a *linear representation*. It is said to *recognize* a series σ if for every $w \in A^*$, one has

$$(\sigma, w) = \lambda\mu(w)\gamma. \tag{2.1}$$

Equivalently, one can define a linear representation as a triple (λ, μ, γ) formed, for some vector space V, of a vector $\lambda \in V$, a morphism $\mu : A^* \to \mathrm{End}(V)$ and a linear form γ on V such that Equation (2.1) holds for every $w \in A^*$. The vector λ is called the *initial vector*. The endomorphism $\mu(w)$ and the linear form γ act on the right of their argument and thus the expression $\lambda\mu(w)\gamma$ in Equation (2.1) is read parenthesized from left to right as $(\lambda\mu(w))\gamma$. It can also be parenthesized from right to left considering $\mu(w)$ as the endomorphism of the dual V' of V defined by the formula $\ell(\mu(w)c) = (\ell\mu(w))c$.

3 Birecurrent sets

In this section, we define birecurrent sets and prove some elementary properties. We prove a characterization (Theorem 3.2.1) which is used in the next section. A recognizable set is *recurrent* if its minimal automaton is strongly connected.

Clearly, a recognizable set is recurrent if and only if it recognized by a strongly connected deterministic automaton. Indeed, if $\mathcal{A}$ is a strongly connected deterministic automaton recognizing X, then the minimal automaton of X is also strongly connected.

As a simple and well known example, for any recognizable prefix code X, the submonoid X^* generated by X is recurrent.

A recognizable set is *birecurrent* if X and its reversal $\tilde{X}$ are both recurrent.

Thus, for example, the submonoid X^* generated by a bifix code is a birecurrent set.

As a related example, let us consider a reversible automaton, that is a deterministic automaton such that every letter defines an injective map on the set of states. This class of automata has been considered frequently (see for example [1], [9] or [8]). The reversal of a reversible automaton $\mathcal{A}$ is still deterministic and thus the set recognized by a strongly connected reversible automaton is birecurrent.

The two examples of submonoids generated by bifix codes and of sets recognized by strongly connected reversible automata partially overlap. Indeed, if $\mathcal{A} = (Q, i, i)$ is a reversible automaton with a unique terminal state equal to the initial state, the set recognized by $\mathcal{A}$ is a submonoid generated by a bifix code. However, there are many examples of submonoids generated by bifix codes which are not recognizable by a reversible automaton (see for instance the bifix code of Example 11).

We will see that there are many other examples of birecurrent sets.

We begin by recalling some basic and well-known facts concerning the reversal of an automaton.

3.1 Reversal of an automaton

Let $\mathcal{A} = (Q, I, T)$ be an automaton. The *reversal* of $\mathcal{A}$ is the automaton $\tilde{A} = (Q, T, I)$ obtained by reversing the edges of $\mathcal{A}$ and exchanging I and T. Clearly, the reversal of $\mathcal{A}$ recognizes the reversal $\tilde{X}$ of the set X recognized by $\mathcal{A}$.

For an automaton $\mathcal{A} = (Q, I, T)$, we denote by $\mathcal{A}^\delta$ the *determinization* of $\mathcal{A}$. Its states are the nonempty sets $I \cdot w = \{q \in Q \mid i \xrightarrow{w} q \text{ for some } i \in I\}$. The initial state is I and the set of terminal states is the family of states U of $\mathcal{A}^\delta$ such that $U \cap T \neq \emptyset$.

The *deterministic reversal* of $\mathcal{A}$ is the automaton $\tilde{\mathcal{A}}^\delta$ obtained by determinization of the reversal of $\mathcal{A}$. We denote by $\tilde{Q}$ the set of states of $\tilde{\mathcal{A}}^\delta$. Thus $\tilde{Q}$ is the family of nonempty sets of the form

$$w \cdot T = \{q \in Q \mid q \xrightarrow{w} t \text{ for some } t \in T\}.$$

The set of terminal states of $\tilde{\mathcal{A}}^\delta$ is $\{U \in \tilde{Q} \mid I \cap U \neq \emptyset\}$. Clearly, $\tilde{\mathcal{A}}^\delta$ recognizes the reversal of the set recognized by $\mathcal{A}$.

The following statement is well known (see [5] p. 48).

Proposition 3.1.1 *If $\mathcal{A}$ is a trim deterministic automaton recognizing X, then $\tilde{\mathcal{A}}^\delta$ is the minimal automaton of $\tilde{X}$.*

Proof Since $\mathcal{A}$ is trim, for any word w, one has $w \cdot T \neq \emptyset$ if and only if $Xw^{-1} \neq \emptyset$. Moreover, for any $w, w' \in A^*$, one has

$$w \cdot T = w' \cdot T \Leftrightarrow Xw^{-1} = Xw'^{-1}$$

as one may easily verify. Since the nonempty sets Xw^{-1} are the reversals of the states of the minimal automaton of $\tilde{X}$, the map $w \cdot T \mapsto \tilde{w}^{-1}\tilde{X}$ is a bijection which identifies $\tilde{\mathcal{A}}^\delta$ with the minimal automaton of $\tilde{X}$. ∎

It follows from Proposition 3.1.1 that if $\mathcal{A}$ is the minimal automaton of a recognizable set X, then X is birecurrent if and only if $\mathcal{A}$ and $\tilde{\mathcal{A}}^\delta$ are strongly connected.

3.2 A characterization of birecurrent sets

The following statement characterizes birecurrent sets. One direction appears as [7, Proposition 5.7].

We say that a set $S \subset Q$ is *saturated* by a word w if S is a union of classes of the kernel of w. Note that S is saturated by w if and only if $S = w \cdot U$ for some $U \subset Q$.

Theorem 3.2.1 *Let S be a recurrent set and let $\mathcal{A} = (Q, i, T)$ be its minimal automaton. Then S is birecurrent if and only if T is saturated by a word of minimal nonzero rank.*

Proof Let us first show that the condition is necessary. Let v be a word of nonzero minimal rank such that $i \cdot v \in T$. Since $v \cdot T \neq \emptyset$ and since $\tilde{\mathcal{A}}^{\delta}$ is strongly connected, there is a word u such that $(uv) \cdot T = T$. Thus T is saturated by uv.

Conversely, set $\varphi = \varphi_{\mathcal{A}}$, $M = M_{\mathcal{A}}$ and let J be the 0-minimal ideal of M. Assume that T is saturated by a word x of minimal nonzero rank. Then $\varphi(x) \in J$. Let u be a word such that $u \cdot T \neq \emptyset$. We have to show that there is a word w such that $(wu) \cdot T = T$. This will prove that T is accessible from $u \cdot T$ in $\tilde{\mathcal{A}}$, which implies that $\tilde{\mathcal{A}}^{\delta}$ is strongly connected.

Since $u \cdot T \neq \emptyset$, we have $\varphi(ux) \neq 0$. Since $\varphi(x) \in J$, the left ideal generated by $\varphi(x)$ is 0-minimal. Thus $\varphi(ux)$ is in the $\mathcal{L}$-class of x. Let v be a word such that $\varphi(vux) = \varphi(x)$. Since $\varphi((vu)^n x) = \varphi(x)$ for all $n \geq 1$, the idempotent e which is a power of $\varphi(vu)$ is such that $e\varphi(x) = \varphi(x)$. Since T is saturated by x, there is a set U such that $T = x \cdot U$. Then $eT = e(x \cdot U) = e\varphi(x)U = \varphi(x)U = T$. Finally, let m be such that $e = \varphi(vu)^m$ and let $w = (vu)^{m-1}v$. Then $(wu) \cdot T = eT = T$ and the proof is complete. ∎

Note that if the minimal automaton $\mathcal{A} = (Q, i, T)$ of a birecurrent set S is complete and synchronized, then $\mathcal{A}$ is the trivial automaton with only one state and $S = A^*$. This can of course be proved directly, but it follows easily from the characterization above. Indeed, in this case, T is saturated by a word of minimal rank if and only if $T = Q$, which implies the conclusion $S = A^*$.

Theorem 3.2.1 gives a method to find birecurrents sets starting with a stongly connected automaton with an uspecified set of terminal states. Provided one knows a word x of minimal nonzero rank, one may choose as set of terminal states a set saturated by x and finally minimize the resulting automaton.

However, this does not give a substantially more efficient algorithm than the computation of the deterministic reversal. Indeed, it is shown in [12] that deciding whether a partial automaton is strongly connected as well as its deterministic reversal is PSPACE-complete.

The following examples illustrate Theorem 3.2.1.

Example 3 The automaton represented in Figure 3.1 on the left and its deterministic reversal on the right are both strongly connected. In agreement with

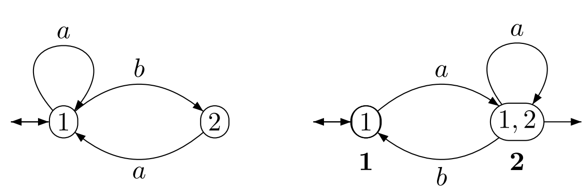

Figure 3.1: A deterministic automaton and its deterministic reversal.

Theorem 3.2.1, the set $T = \{1\}$ of terminal states of the first automaton is the preimage of b, which has rank 1.

We give a second example with a minimal rank larger than 1.

Example 4 Consider the complete deterministic automaton $\mathcal{A}$ given in Figure 3.2 on the left with its deterministic reversal represented on the right. The

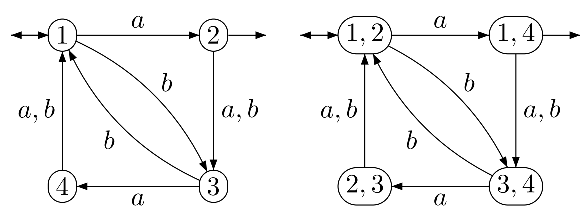

Figure 3.2: The automata $\mathcal{A}$ and $\tilde{\mathcal{A}}^\delta$ in Example 4.

0-minimal ideal of the monoid $\varphi_{\mathcal{A}}(A^*)$ is represented in Figure 3.3. We can check that the set $T = \{1, 2\}$ is a class of the kernel of b, which is a word of minimal rank equal to 2.

	$1/3$	$2/4$
$1, 2/3, 4$	$^*\ b$	$^*\ ba$
$1, 4/2, 3$	$^*\ ab$	$^*\ aba$

Figure 3.3: The 0-minimal ideal of $\varphi_{\mathcal{A}}(A^*)$ in Example 4.

In the last example, we show a simple case in which T is a union of several classes of the kernel of a word of minimal nonzero rank.

Example 5 Let $\mathcal{A}$ be the deterministic automaton on $Q = \{1, 2, 3\}$ and with $A = \{a\}$ where a is the circular permutation (123). Choosing $i = 1$ and $T = \{1, 2\}$, we obtain for $\tilde{\mathcal{A}}^\delta$ the automaton on the set of states $\{\{1, 2\}, \{1, 3\}, \{2, 3\}\}$ on which a acts again as a circular permutation. Thus $\tilde{\mathcal{A}}^\delta$ is strongly connected (see Figure 3.4). Actually, this holds for any strongly connected group automaton.

3.3 Birecurrent sets of finite type

In this section, we define birecurrent sets of finite type and in the next section, we prove a statement which allows to build an infinite family of such sets (Theorem 3.4.2). We first prove the following elementary statement concerning recurrent sets.

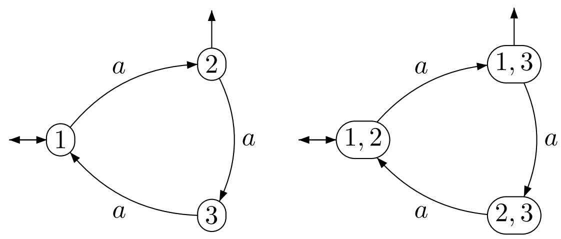

Figure 3.4: The automata $\mathcal{A}$ and $\tilde{\mathcal{A}}^\delta$ in Example 5.

Proposition 3.3.1 *A set $S \subset A^*$ is recurrent if and only if there is a recognizable prefix code X and a recognizable set P of proper prefixes of X such that $S = X^*P$.*

Proof Assume first that S is recurrent. Let $\mathcal{A} = (Q, i, T)$ be the minimal automaton of S. Let X be the prefix code generating the submonoid recognized by the automaton (Q, i, i) and let P be the set of proper prefixes p of X such that $i \cdot p \in T$. Then $S = X^*P$.

Conversely, let $\mathcal{A} = (Q, \varepsilon, \varepsilon)$ be the literal automaton of X^*. Recall from Section 2 that Q is the set of proper prefixes of X and that $\mathcal{A}$ recognizes X^*. Then S is recognized by the automaton (Q, ε, P) which is a strongly connected automaton. Moreover, since X and P are recognizable, S is recognizable. Thus S is recurrent. ∎

A pair (X, P) as above is called a *decomposition* of the recurrent set S. The prefix code X obtained as above using the minimal automaton of S is called the *left root* of S.

Note that a recurrent set has in general several decompositions. However, the left root X of S is such that for any such pair (X', P'), we have $X' \subset X^*$, that is the submonoid X^* is maximal for this property. Indeed, let $(Q', \varepsilon, \varepsilon)$ be the literal automaton of X'^*. Then $\mathcal{A}' = (Q', \varepsilon, P')$ recognizes S and thus there is a reduction from $\mathcal{A}'$ onto the minimal automaton (Q, i, P) of S which sends ε to i. The image by this reduction of a path in $\mathcal{A}'$ from ε to ε is a path in $\mathcal{A}$ from i to i and thus $X' \subset X^*$.

Proposition 3.3.2 *The left root of a recurrent set S is finite if and only if S has a decomposition (X, P) with X finite.*

Proof Let (X, P) and (X', P') be two decompositions of a recurrent set S with X the left root of S. Assume that X' is finite.

Any word of X is a prefix of a word in X'. Indeed, let $x \in X$ and let p be some element of P. Then $xp = x'p'$ with $x' \in X'^*$ and $p' \in P'$. Thus x is a prefix of a word in X'^*. Since $X' \subset X^*$, this implies that x is a prefix of a word in X'. This shows that the length of x is bounded by the maximal length of the words in X'. Therefore X is finite.

The other implication is clear. ∎

For a birecurrent set S, we have

$$S = X^* P = QY^* \tag{3.1}$$

where $\tilde{Y}$ is the left root of $\tilde{S}$ and Q is a set of proper suffixes of Y. The suffix code Y is called the *right root* of S. When S is the submonoid generated by a bifix code X, then X is the left root of S and $\tilde{X}$ is its right root.

We say that a birecurrent set S is of *finite type* if its left and right roots are finite.

For example, for every finite bifix code X, the set $S = X^*$ is a birecurrent set of finite type.

On the contrary, the birecurrent set of Example 3 is not of finite type (see Example 8).

We give two examples of birecurrent sets of finite type. The first one is Example 4. Recall from Section 2 that $\underline{Y}$ denotes the characteristic series of a set $Y \subset A^*$. Note that when (X, P) is a decomposition of a recurrent set S, we have $\underline{S} = \underline{X}^* \underline{P}$. Indeed, since X is a prefix code, we have $\underline{M} = (\underline{X})^*$ for the submonoid $M = X^*$ and $\underline{S} = \underline{M}\ \underline{P}$ because the product (M, P) is unambiguous (see [2]).

Example 6 The birecurrent set S of Example 4 is of finite type. Indeed, its left root is the finite prefix code $X = Z^2$ with $Z = aA \cup b$ and its right root is the finite suffix code $\tilde{X}$. One has $S = X^* P = P\tilde{X}^*$ with $P = \{\varepsilon, a\}$. One has actually $S = \tilde{S}$. This can be checked by comparing the two automata of Figure 3.2 or directly by observing that $\underline{Z} = a^2 + (1 + a)b$ implies that $\underline{Z}(1 + a) = (1 + a)\underline{\tilde{Z}}$, whence $X^* P = P\tilde{X}^*$ since $\underline{P} = 1 + a$.

The second example is Example 3.6.13 in [2].

Example 7 Let $\mathcal{A} = (Q, i, T)$ be the automaton given by Table 3.1 with $i = 1$ and $T = \{1, 6\}$. The minimal rank is equal to 3 as one may check by computing

	1	2	3	4	5	6	7	8	9
a	2	3	1	1	3	8	9	3	1
b	4	6	7	5	1	4	1	5	1

Table 3.1: The transitions of the automaton $\mathcal{A}$ in Example 7.

the minimal images which are $\{1, 2, 3\}$, $\{4, 6, 7\}$, $\{1, 4, 5\}$ and $\{1, 8, 9\}$. The set $\{1, 6\}$ is a class of the kernel of a^2 which has image $\{1, 2, 3\}$ and thus minimal rank. The automaton $\tilde{\mathcal{A}}^\delta$ is thus strongly connected by Theorem 3.2.1 and the set S recognized by $\mathcal{A}$ is birecurrent. The left root of S is the finite maximal prefix code represented in Figure 3.5 as the set of leaves of a binary tree. In this figure, as in the following ones, we represent words on the alphabet $\{a, b\}$ by nodes of a binary tree. The edges going up are labeled a, those going down are labeled b.

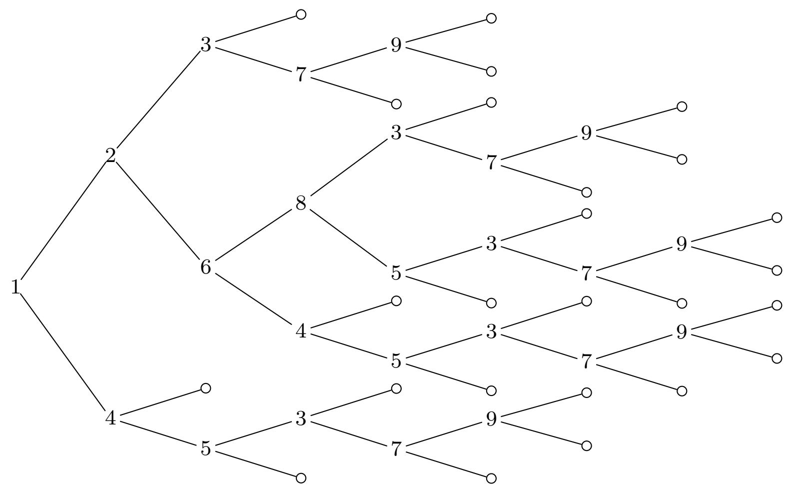

Figure 3.5: The left root of the set S i Example 7.

1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
$1,6$	$3,4,9$	$2,5,7,9$	$2,5,7,8$	$1,7$	$3,4,8$	$3,5,7,9$	$2,5,6,8$	$2,4,8$

Table 3.2: The states of $\tilde{\mathcal{A}}^\delta$ in Example 7.

The deterministic reversal of $\mathcal{A}$ has the set of states represented in Table 3.2.

The transitions of $\tilde{\mathcal{A}}^\delta$ are given in Table 3.3. The reversal $\tilde{Y}$ of the right

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
a	**2**	**4**	**5**	**1**	**2**	**8**	**4**	**1**	**1**
b	**3**	**1**	**6**	**6**	**7**	**1**	**6**	**9**	**1**

Table 3.3: The transitions of the automaton $\tilde{\mathcal{A}}^\delta$ in Example 7.

root of S is represented in Figure 3.6.

Since $\tilde{Y}$ is finite, S is a birecurrent set of finite type.

We give now two examples of birecurrent sets such that the left root is finite but the right root is infinite and thus which are not of finite type.

Example 8 Consider again the set S recognized by the automaton of Figure 3.1 on the left with its deterministic reversal on the right (Example 3). The left root of S is $\{a, ba\}$ and thus it is finite. However, the right root of S is ba^+ and

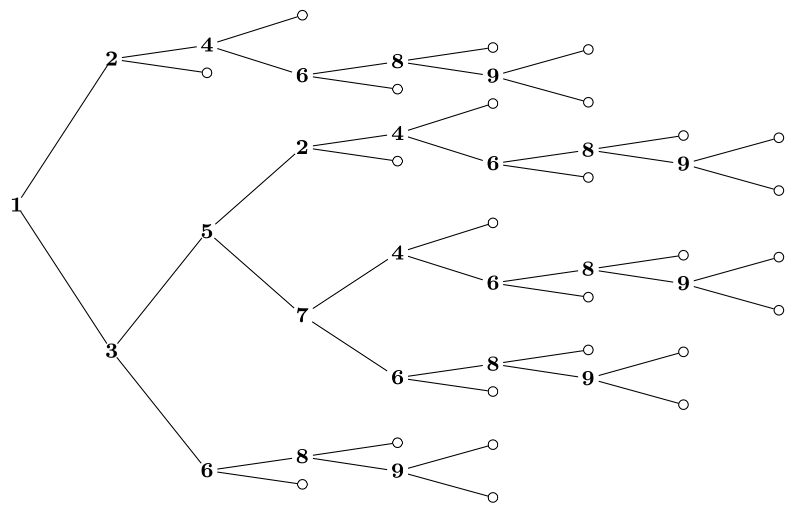

Figure 3.6: The reversal of the right root of the set S in Example 7.

it is infinite. Note that the two factorizations of S correspond to the identity

$$\{a, ba\}^* = a^*(ba^+)^*$$

which is itself a particular case of the *sumstar* identity $(x + y)^* = x^*(yx^*)^*$.

In the second example, the birecurrent set S is dense.

Example 9 Consider the finite maximal prefix code $X = \{aa, aba, abb, b\}^2$. The minimal automaton of X^* is given by its transitions in Table 3.4 on the left with 1 as initial and terminal state.

	1	2	3	4	5	6
a	2	3	4	1	3	1
b	3	5	1	6	3	1

	1	**2**	**3**	**4**	**5**	**6**
a	**2**	**4**	**1**	**6**	**4**	**1**
b	**3**	**4**	**5**	**5**	**3**	**1**

Table 3.4: The transitions of $\mathcal{A}$ and $\tilde{\mathcal{A}}^\delta$ in Example 9.

The word ab^2 has minimal rank 2 and its kernel is $\{1, 2, 5\}, \{3, 4, 6\}$. Keeping the same initial state and taking $T = \{1, 2, 5\}$, we obtain a deterministic automaton $\mathcal{A}$ recognizing the set $S = X^*P$ with $P = \{\varepsilon, a, ab\}$. The set of states of $\tilde{\mathcal{A}}^\delta$ is given in Table 3.5. Its transitions are given in Table 3.4 on the right. The right root of S is the maximal suffix code $\tilde{Y}$ where Y^* is the submonoid recognized by $\tilde{\mathcal{A}}^\delta$ with $\mathbf{1}$ as initial and terminal state. Since, in $\tilde{\mathcal{A}}^\delta$, there is a loop $\mathbf{3} \xrightarrow{b} \mathbf{5} \xrightarrow{b} \mathbf{3}$, the code Y is infinite.

1	**2**	**3**	**4**	**5**	**6**
$1, 2, 5$	$1, 4, 6$	$2, 3, 6$	$3, 4, 6$	$1, 4, 5$	$2, 3, 5$

Table 3.5: The states of $\tilde{\mathcal{A}}^\delta$ in Example 9.

3.4 A construction of birecurrent sets of finite type

Examples 6 and 7 are particular cases of a general construction that we now describe. Let $Z \subset A^+$ be a set of words. A word $x \in Z$ is said to be a *pure square* for Z if $x = w^2$ for some $w \in A^+$ and if $Z \cap wA^* \cap A^*w = \{x\}$. The following result is [2, Exercise 14.1.9 (a)].

Proposition 3.4.1 *If Z is a finite maximal prefix code and if w^2 is a pure square for Z, then, denoting $G = Zw^{-1}$ and $D = w^{-1}Z$, the expression*

$$(1 + w)(\underline{Z} - 1 + (\underline{G} - 1)w(\underline{D} - 1)) + 1 \tag{3.2}$$

is the characteristic polynomial of a finite maximal prefix code. Moreover, the expression

$$(\underline{Z} - 1 + (\underline{G} - 1)w(\underline{D} - 1))(1 + w) + 1 \tag{3.3}$$

is the characteristic polynomial of a finite maximal code.

The prefix code defined by Equation (3.2) is denoted $\delta_w(Z)$. The maximal code defined by Equation (3.3) is denoted $\gamma_w(Z)$. If Z is bifix, then $\gamma_w(Z)$ is a suffix code since $\gamma_w(Z)$ is the reversal of $\delta_{\tilde{w}}(\tilde{Z})$.

Theorem 3.4.2 *Let $Z \subset A^+$ be a finite maximal bifix code and let $x = w^2$ be a pure square for Z. The set $S = \delta_w(Z)^*\{\varepsilon, w\}$ is a dense birecurrent set of finite type.*

Proof Set $X = \delta_w(Z)$ and $Y = \gamma_w(Z)$. Observe that, by defintion of $\delta_w(Z)$ and $\gamma_w(Z)$,

$$\begin{aligned}(1 + w)(\underline{Y} - 1) &= (1 + w)(\underline{Z} - 1 + (\underline{G} - 1)w(\underline{D} - 1))(1 + w)\\ &= (\underline{X} - 1)(1 + w).\end{aligned}$$

Thus, multiplying on the left by $\underline{X}^*$ and on the right by $\underline{Y}^*$, we obtain $\underline{X}^*(1 + w) = (1 + w)\underline{Y}^*$. This shows that $S = X^*\{\varepsilon, w\} = \{\varepsilon, w\}Y^*$.

Since $S = X^*\{\varepsilon, w\}$, the set S is recurrent by Proposition 3.3.1 and its left root is finite by Proposition 3.3.2.

Similarly, since $S = \{\varepsilon, w\}Y^*$, the set S is birecurrent. Since Y finite, the right root of S is also finite. Thus S is birecurrent of finite type.

Since X is a maximal prefix code, the submonoid X^* is right dense. Thus S is right dense. ∎

Observe that if R is the set of proper prefixes of Z and U the set of proper prefixes of D, the characteristic polynomial of the set P of proper prefixes of $X = \delta_w(Z)$ is

$$\underline{P} = (1+w)(\underline{R} + (\underline{G}-1)w\underline{U}). \tag{3.4}$$

Indeed, since Z, X and D are maximal prefix codes, we have $\underline{Z} - 1 = \underline{R}(\underline{A}-1)$, $\underline{X} - 1 = \underline{P}(\underline{A}-1)$ and $\underline{D} - 1 = \underline{U}(\underline{A}-1)$. Thus Equations (3.2) and (3.4) are equivalent.

In the next two examples, we show that Examples 6 and 7 can be obtained by the contruction decribed in Theorem 3.4.2.

Example 10 Let S be the birecurrent set of Example 4. We have seen in Example 6 that $S = X^*P$ with $X = (aA \cup b)^2$ and $P = \{\varepsilon, a\}$. We obtain X as in Theorem 3.4.2 using $Z = A^2$ and $x = a^2$. Indeed, we have $G = D = A$ and thus Equation (3.4) gives

$$(1+a)(1+a+b+(a+b-1)a) = (1+a)(1+Aa+b) = (1+aA+b)(1+a)$$

which is the characteristic polynomial of the set of proper prefixes of X.

Example 11 Let S be the birecurrent set of Example 7. We start from the finite maximal bifix code

$$Z = \{aaa, aab, abaa, abab, abb, ba, bbaa, bbab, bbb\}$$

which is represented in Figure 3.7 on the left with its reversal on the right.

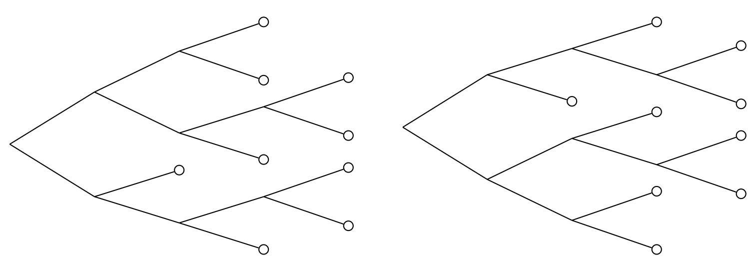

Figure 3.7: The bifix code Z and its reversal $\widetilde{Z}$ in Example 11.

The word $x = w^2$ with $w = ab$ is a pure square for Z. We have $G = \{a, ab, bb\}$ and $D = \{aa, ab, b\}$.

The set R of proper prefixes of Z is $R = \{\varepsilon, a, b, aa, ab, bb, aba, bba\}$ and the set U of proper prefixes of D is $U = \{\varepsilon, a\}$.

The set R is represented in Figure 3.8 on the left, the set $R \setminus wU$ is represented in the middle, and the set $Q = (R \setminus wU) \cup GwU$ on the right (the white nodes being not in the set).

Finally, the set $P = \{\varepsilon, w\}Q$ is represented in Figure 3.9 (with the nodes of Q in black and those of wQ in red). By Equation (3.4), it is the set of proper prefixes of the maximal prefix code $X = \delta_w(Z)$ represented in Figure 3.5.

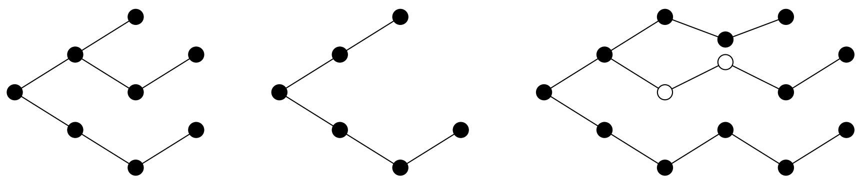

Figure 3.8: The sets R, $R \setminus wU$ and $Q = (R \setminus wU) \cup GwU$ in Example 11.

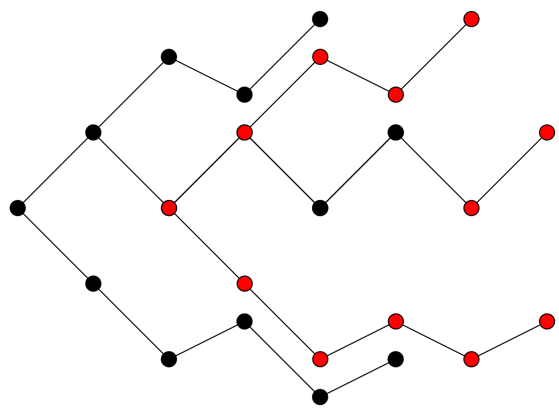

Figure 3.9: The set $P = \{\varepsilon, w\}Q$ in Example 11.

Theorem 3.4.2 gives an infinite family of examples of birecurrent sets of finite type which are not submonoids generated by a bifix code. Indeed, for any even $n \geq 1$, the word a^n is a pure square for the bifix code A^n.

3.5 Multiple factorizations of noncommutative polynomials

Let S be a dense birecurrent set of finite type. Then $S = X^*P = QY^*$ where X (resp. Y) is a finite maximal prefix code (resp. a finite maximal suffix code). Since all products are non ambiguous, we have the equality

$$(1 - \underline{X})\underline{Q} = \underline{P}(1 - \underline{Y}). \tag{3.5}$$

Let J be the set of proper prefixes of X and let K be the set of proper suffixes of Y. Then

$$1 - \underline{X} = \underline{J}(1 - \underline{A}), \quad 1 - \underline{Y} = (1 - \underline{A})\underline{K}. \tag{3.6}$$

Combining Equations (3.5) and (3.6), we obtain

$$\underline{J}(1 - \underline{A})\underline{Q} = \underline{P}(1 - \underline{A})\underline{K}. \tag{3.7}$$

We conjecture that, with the above notation there exist sets $M, N \subset A^*$ such that

$$\underline{J} = \underline{P}\,\underline{M}\,, \quad \underline{K} = \underline{N}\,\underline{Q}\,, \quad \underline{M}\,(1-\underline{A}\,) = (1-\underline{A}\,)\underline{N} \tag{3.8}$$

This is true when S is generated by a maximal bifix code and when it is obtained by the construction of Section 3.4. Indeed, in this case, we have $\underline{P} = \underline{Q} = 1+w$ and by Equation (3.4)

$$\underline{J} = (1+w)(\underline{P_Z} + (\underline{G} - 1)w\underline{P_D}\,)$$

and

$$\underline{K} = (1+w)(\underline{S_Z} + (\underline{S_G} - 1)w(\underline{D} - 1)).$$

where P_Z, P_D denote the set of proper prefixes of Z, D and S_Z, S_G denote the set of proper suffixes of Z, G. Thus $\underline{J} = \underline{P}\,\underline{M}$ with $\underline{M} = \underline{P_Z} + (\underline{G} - 1)w\underline{P_D}$ and $\underline{K} = \underline{N}\,\underline{Q}$ with $\underline{N} = \underline{S_Z} + (\underline{S_G} - 1)w(\underline{D} - 1)$. Moreover $\underline{M}\,(\underline{A} - 1) = \underline{Z} - 1 + (\underline{G} - 1)w(\underline{D} - 1) = (1-\underline{A}\,)\underline{N}\,$.

A weak form of this conjecture is proved in [4, Theorem 3.1].

3.6 Degree of a dense birecurrent set

Let S be a dense birecurrent set. We define its *degree* as the degree of its minimal automaton. Recall from Section 2 that the degree $d(X)$ of a prefix code X is, by definition, the degree of the minimal automaton of X^*. Thus the degree of a dense birecurrent set is the degree of its left root (this is true even if the minimal automaton of X^* is a quotient of the minimal automaton of S, see [2, Proposition 9.5.2]).

Thus, when $S = X^*$ with X a maximal bifix code, the degree of S is the degree of X.

Note that the degree of the right root may be different from the degree of the left root, and thus of the degree of the reversal, as shown in the following example.

Example 12 Let S be the dense birecurrent set recognized by the automaton $\mathcal{A}$ represented in Figure 3.10 on the left. The automaton $\tilde{\mathcal{A}}^\delta$ is represented on the right. The degree of S is 4 since the transition monoid of $\mathcal{A}$ is the symmetric group $\mathfrak{S}_4$ on 4 elements. But the degree of the right root of S is 6 since the transition monoid of the automaton $\tilde{\mathcal{A}}^\delta$ is a representation of $\mathfrak{S}_4$ on 6 elements.

Let S be a dense birecurrent set of degree d and let $\mathcal{A} = (Q, i, T)$ be its minimal automaton. By Theorem 3.2.1, the set T is saturated by a word w of minimal rank. The rank of w is by definition the degree d of S. Let $k \geq 1$ be the number of classes of the kernel of w whose union is T. We define the *index* of a dense birecurrent set, denoted $i(S)$, as the rational number d/k.

Since $1 \leq k \leq d$, we have $1 \leq i(S) \leq d$.

Note that the index does not depend on the choice of the word w. Indeed, since S is dense, any minimal image (that is the image of a word of minimal

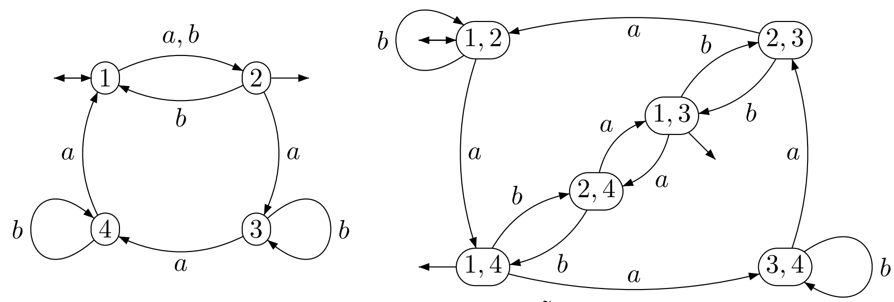

Figure 3.10: The automata $\mathcal{A}$ and $\tilde{\mathcal{A}}^\delta$ in Example 12.

rank) is a set of representatives of any kernel of a word of minimal rank and thus k is the number of elements of T in each minimal image.

When $S = X^*$ with X a maximal bifix code, the index of S is the degree of X.

Example 13 Let S be the dense birecurrent set of Example 12. Since the degree of S is 4, since $\varphi_{\mathcal{A}}(A^*)$ is a group and since $\mathrm{Card}(T) = 2$, we have $i(S) = 2$.

Example 14 Let S be the birecurrent set of Example 5. Then $i(S) = 3/2$.

Proposition 3.6.1 *Let S be a dense birecurrent set and let $\mathcal{A}$ be the minimal automaton of S. Then for every $\mathcal{H}$-class H of the minimal ideal of $\varphi_{\mathcal{A}}(A^*)$, one has*

$$i(S) = \mathrm{Card}(H)/\,\mathrm{Card}(H \cap \varphi_{\mathcal{A}}(S)). \tag{3.9}$$

Proof Set $\mathcal{A} = (Q, i, T)$ and let d be the degree of S. Let x be a word of rank d which saturates T. Let H be the $\mathcal{H}$-class of x and let k be the number of classes of the kernel of x whose union is T. Let I be the common image of the elements of H. Since I is a system of representatives of the kernel of x, it contains k elements of T. Let j be the element of I such that $ih = jh$ for every $h \in H$. Then, for every $h \in H$, one has $ih \in T$ if and only if $jh \in T$. Thus the set $H \cap \varphi_{\mathcal{A}}(S)$ is a union of k cosets of the subgroup of H fixing j. Thus $\mathrm{Card}(H \cap \varphi_{\mathcal{A}}(S)) = k\,\mathrm{Card}(H)/d = \mathrm{Card}(H)/i(S)$ ■

Note that, as a consequence, the index of a dense birecurrent set and of its reversal are the same (contrary to the degree). Indeed, the monoids $\varphi_{\mathcal{A}}(A^*)$ and $\varphi_{\tilde{\mathcal{A}}}(A^*)$ are antiisomorphic and antiisomorphic groups are isomorphic.

By Equation (3.9), the index of a birecurrent set is a measure of its size. We will make this idea more precise using probabilities. We begin with some definitions on Bernoulli distributions (see [2] for more details).

Let π be a positive Bernoulli distribution on A^*. By definition, π is a morphism from A^* into the interval $]0, 1]$ such that $\sum_{a \in A} \pi(a) = 1$. For $S \subset A^*$, we denote by $\pi(S)$ the (possibly infinite) sum $\sum_{x \in S} \pi(x)$.

For any code, one has $\pi(X) \le 1$ and when X is a thin maximal code, one has $\pi(X) = 1$ (see [2, Theorem 2.5.19]). Moreover, if X is prefix, then we have $\lambda(X) = \pi(P)$ where $\lambda(X) = \sum_{x \in X} |x|\pi(x)$ is the *average length* of X and where P is the set of proper prefixes of X (see [2, Corollary 3.7.13]).

The *density* of a recognizable set S, denoted $\delta(S)$ is the Cesaro limit of the numbers $\pi(S \cap A^n)$. Thus $\delta(S) = \lim \frac{1}{n} \sum_{0 \le i < n} \pi(S \cap A^i)$. By [2, Theorem 13.2.9], when X is a thin maximal code, we have $\delta(X^*) = 1/\lambda(X)$.

The following result shows that the density of a dense birecurrent set is a rational number, independent of the Bernoulli ditribution π. This surprising property was first put in evidence for recognizable bifix codes in [14].

Theorem 3.6.2 *The density of a dense birecurrent set is the inverse of its index.*

Proof Let S be a dense birecurrent set and let $\mathcal{A}$ be its minimal automaton. Set $\varphi = \varphi_A$ and $M = \varphi(A^*)$, and let K be the minimal ideal of M. Let $\nu = \delta\varphi^{-1}$ where δ denotes the density. Then, by [2, Theorem 13.4.7], ν is a probability measure on the family of subsets of M. Moreover, $\nu(K) = 1$ and for every $m \in K$, one has

$$\nu(m) = \frac{\nu(H)}{\mathrm{Card}(H)} \tag{3.10}$$

where H is the $\mathcal{H}$-class of m. Let $\mathcal{F}$ denote the family of $\mathcal{H}$-classes of K. Then, using Proposition 3.6.1,

$$\begin{aligned}
\delta(S) &= \sum_{m \in \varphi(S) \cap K} \nu(m) \\
&= \sum_{H \in \mathcal{F}} \mathrm{Card}(H \cap \varphi(S)) \frac{\nu(H)}{\mathrm{Card}(H)} \\
&= \frac{1}{i(S)} \sum_{H \in \mathcal{F}} \nu(H) = \frac{\nu(K)}{i(S)} = \frac{1}{i(S)}.
\end{aligned}$$

■

Theorem 3.6.2 implies the following result for a dense birecurrent set of finite type.

Corollary 3.6.3 *Let $S = X^*P$ be a dense birecurrent set of finite type with left root X. Then, for any positive Bernoulli distribution π,*

$$\lambda(X) = i(S)\pi(P).$$

Proof Since $S = X^*P$, we have $\delta(S) = \delta(X^*)\pi(P)$ by [2, Proposition 13.2.5]. Since X is a finite maximal prefix code, we have $\delta(X^*) = 1/\lambda(X)$ by [2, Theorem 13.2.9]. Thus, by Theorem 3.6.2 we have $\lambda(X) = \pi(P)/\delta(S) = i(S)\pi(P)$. ■

Note that when Equation (3.8) holds, we have $\pi(M) = i(S)$. The fact that $\pi(M)$ is a rational number independant of π is itself a consequence of the equation

$\underline{M}(1-\underline{A}) = (1-\underline{A})\underline{N}$ which implies by left Euclidean division $\underline{M} = \alpha + (1-\underline{A})u$ for polynomial u and some scalar α and thus $\pi(M) = \alpha$.

By a well-known result, for every $d \geq 1$ and any finite alphabet, there is only a finite number of finite maximal bifix codes of degree d on this alphabet (see [2, Theorem 6.5.2]). There is no similar property for dense birecurrent sets of finite type, as shown by the following example.

Example 15 For $n \geq 2$, let $Z = \{a^n, a^{n-1}b, \ldots, ab, b\}$, let $X = Z^2$ and let $P = \{\varepsilon, a, \ldots, a^{n-1}\}$. Then $S = X^*P$ is a birecurent set (the case $n = 2$ is Example 4). Indeed, we have $\underline{Z}\,\underline{P} = \underline{P}\,\underline{\tilde{Z}}$ and thus $\tilde{S} = \underline{P}\,\underline{\tilde{X}}^*$. The degree of S is 2 for every $n \geq 1$ because $(ZP \cup P)b \subset X \cup Z$ and thus the rank of b is 2.

3.7 Indecomposable prefix codes

In this section, we relate birecurrent sets with the notion of decomposition of prefix codes.

A prefix code $X \subset A^*$ is *indecomposable* if $X \subset Z^*$, with $Z \subset A^*$ a prefix code, implies $Z = A$ or $Z = X$. Otherwise X is said to be *decomposable* over Z(see [2] for a more detailed presentation).

If X is decomposable over Z, let $\beta : B \to Z$ be a bijection of Z with an alphabet B, extended to a morphism from B^* onto Z^*. Then $Y = \beta^{-1}(X)$ is a prefix code on the alphabet B. We denote $X = Y \circ_\beta Z$. The prefix code X is thin maximal, if and only if Y and Z are also thin maximal prefix codes. Moreover, one has $d(X) = d(Y)d(Z)$ [2, Proposition 11.1.2]. In particular, X is synchronized (that is, of degree 1) if and only if Y and Z are synchronised.

We will prove the following result. It singles out two basic building blocks for recognizable maximal prefix codes: synchronized ones on the one hand, and left roots of dense birecurrent sets on the other. Note that no nontrivial prefix code (that is, distinct of the alphabet) can belong to both families.

Theorem 3.7.1 *Let X be a recognizable maximal prefix code. If X is indecomposable, either X is synchronized or it is the left root of a dense birecurrent set.*

To prove Theorem 3.7.1, we introduce the following notion, which plays a role in the solution of the road coloring problem (see [2, Lemma 10.4.3]).

Let $\mathcal{A} = (Q, i, T)$ be a finite deterministic automaton. A pair of states $p, q \in Q$ is called *synchronizable* if there is a word v such that $p \cdot v = q \cdot v$. It is called *strongly synchronizable* if for every $u \in A^*$ the pair $p \cdot u, q \cdot u$ is synchronizable.

We note that the equivalence ρ on Q defined by $p \equiv q \bmod \rho$ if p, q are strongly synchronizable is a *stable equivalence*. This means that if $p \equiv q \bmod \rho$, then $p \cdot u \equiv q \cdot u \bmod \rho$ for every word u.

Proposition 3.7.2 *Let $\mathcal{A}$ be a strongly connected and complete finite deterministic automaton. Two states p, q of $\mathcal{A}$ are strongly synchronizable if and only if $p \cdot x = q \cdot x$ for every word x of minimal rank.*

Proof The condition is necessary. Indeed, let p, q be strongly synchronizable and let x be a word of minimal rank. Let y be a word such that $p \cdot xy = q \cdot xy$. Since $\varphi_{\mathcal{A}}(x)$ generates a minimal right ideal, there is a word z such that $\varphi_{\mathcal{A}}(xyz) = \varphi_{\mathcal{A}}(x)$. Thus $p \cdot x = p \cdot xyz = q \cdot xyz = q \cdot x$.

Conversely, assume that the condition is satisfied. Let x be a word of minimal rank. For every word u, the word ux has minimal rank and thus $p \cdot ux = q \cdot ux$. Thus p, q are strongly synchronizable. ■

Let X be a recognizable maximal prefix code and let $\mathcal{A} = (Q, i, i)$ be the minimal automaton of X^*. If ρ is a stable equivalence on Q, then the set $U = \{u \in A^* \mid i \cdot u \equiv i \bmod \rho\}$ is a submonoid generated by a prefix code Z with $X \subset Z^*$. If $X = Z$, then ρ must be the equality since $\mathcal{A}$ is minimal. Thus, if X is indecomposable, ρ must be the equality.

Proof of Theorem 3.7.1. Assume that X is not synchronized. Let $\mathcal{A} = (Q, i, i)$ be the minimal automaton of X^*. Let x be a word of minimal rank and let T be a class of the kernel of x. Let $\mathcal{A}' = (Q, i, T)$. The set S recognized by $\mathcal{A}'$ is birecurrent by Theorem 3.2.1. Let us verify that $\mathcal{A}'$ is minimal. This will imply our conclusion since then X is the left root of the birecurrent set S.

We first observe that since X is indecomposable, two strongly synchronizable states are equal. Indeed, this follows from the observation made above that the equivalence on Q defined by $p \equiv q$ if p, q are strongly synchronizable is a stable equivalence.

Let p, q be two states such that for every word w, $p \cdot w \in T$ if and only if $q \cdot w \in T$. Let u be a word. Since $\mathcal{A}$ is strongly connected, there is some word v such that $p \cdot uv \in T$. Then $q \cdot uv \in T$ and thus $p \cdot uvx = q \cdot uvx$. This shows that p, q are strongly synchronizable. Since X is indecomposable, this implies $p = q$. Thus $\mathcal{A}'$ is minimal. ■

Theorem 3.7.1 is related with an old conjecture of Schützenberger asserting that if a finite maximal prefix code is indecomposable, either it is synchronized or it is bifix. The conjecture is not true as shown by the left root X of the birecurrent set S of Example 7 (which appeared originally in [6]). In fact X is indecomposable (see below), is not bifix and not synchronized since its degree is 3.

Example 16 Let us show that the left root X of the birecurrent set of Example 7 is indecomposable and, more generally, that if Z is a finite maximal bifix code of prime degree $d \geq 3$ and w is a pure square for Z, the prefix code X given by Equation (3.2) is indecomposable (this is already proved in [6], but we reproduce the proof for convenience).

We first observe that $a^d \in X$ for all $a \in A$. Indeed, one has $a^d \in Z$ as for any finite maximal bifix code of degree d [2, Proposition 6.5.1]. Next, since $d \geq 3$, it is odd and thus we cannot have $w \in a^*$.

Let T be a prefix code such that $X \subset T^*$. Then, since d is prime, one has for any $a \in A$, either $a \in T$ or $a^d \in T$. Set $X = Y \circ_\beta T$. Since $d(X) = d(Y)d(T)$, either Y is synchronized or T is synchronized. We show that $T = A$ or $T = X$.

Assume first that T is synchronized. Then Y has degree d. If $a^d \in T$ for some $a \in A$, then Y is synchronized (some power of $b = \beta^{-1}(a^d)$ is a synchronizing word). This is impossible since $d > 1$. Therefore $a \in T$ for every $a \in A$ and thus $T = A$.

Assume next that Y is synchronized. Then, since T has degree d, $a^d \in T$ for every $a \in A$. Fix some $a \in A$. By inspection of Equation 3.2, the suffixes of X which are in a^* are of length at most d and the only proper prefixes p of X such that $pa^d \in X$ are ε and w. This implies that the only proper prefixes of X which possibly belong to T^* are ε and w. Indeed, if $p \in T^*$ and $a^n \in X$, then $a^n \in T^*$ and thus n is multiple of d. But if $w \in T^*$, we have $D \subset T^*$ since $w^3 D \subset X$ which is impossible since the integer n such that $a^n \in D$ is strictly less than d. This forces $T = X$ and shows that X is indecomposable.

Note that the statement is not true for $d = 2$ since the code X of Example 10 is decomposable.

We formulate the following open problem, as an attempt to replace the conjecture of Schützenberger by a weaker statement: if the prefix code in Theorem 3.7.1 is additionnaly finite, can one prove that either it is synchronized or it is the left root of a dense birecurrent set of finite type?

4 Complete reducibility

In this section, we develop the link between birecurrent sets and completely reducible sets. We begin with an introduction to formal series (see [3] or [13] for a more detailed presentation).

4.1 Recognizable series

We consider formal series with coefficients in the field $K = \mathbb{Q}$. Recall from Section 2 that we denote by $K\langle\langle A\rangle\rangle$ the ring of formal series with coefficients in K and noncommutative variables in A.

A series is *recognizable* (or equivalently, by Schützenberger's theorem, *rational*) if there is a linear representation over a finite dimensional space recognizing it. There is a unique linear representation of minimal dimension recognizing a given recognizable series (up to the choice of a basis if the representation is given in matrix terms). One can compute it following three steps.

1. Start from any linear representation (λ, μ, γ) recognizing σ.
2. Take the representation $(\lambda', \mu', \gamma')$ obtained by taking $\lambda' = \lambda$ and by restricting μ and γ to the subspace generated by the vectors $\lambda\mu(w)$ for $w \in A^*$.
3. Take the representation $(\lambda'', \mu'', \gamma'')$ obtained by taking $\gamma'' = \gamma'$ and by restricting μ' and λ' to the subspace generated by the vectors $\mu'(w)\gamma'$ for $w \in A^*$.

EXAMPLE 17 Let $A = \{a\}$. The linear representation

$$\lambda = [1\ 0], \quad \mu(a) = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}, \quad \gamma = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

recognizes the characteristic series of the set a^+. It is minimal. Choosing $\gamma = [1\ 1]^t$, the representation recognizes a^*. It is not minimal because $\mu(a)\gamma = \gamma$. It is easy to see that the minimal representation of a^* is of dimension 1.

The computation of the minimal representation of the characteristic series of a set $S \subset A^*$ is closely related to the computation of the minimal automaton of S and of its reversal. It can be described as follows.

Consider the vector space K^Q as containing Q, identifying $q \in Q$ with its characteristic function. Let (λ, μ, γ) be the following linear representation on the space $V = K^Q$. Set $\lambda = i$. For $w \in A^*$, define $\mu(w)$ as the endomorphism of V such that $q\mu(w) = q \cdot w$. Finally set $\gamma = \underline{T}$ where $\underline{T}$ is the linear form on V which is the characteristic function of T. Then (λ, μ, γ) obviously recognizes $\underline{S}$ and we have performed the first step of the algorithm given above to compute the minimal representation. Step 2 does not change the representation since each state of the minimal automaton is accessible from the initial state. To perform Step 3, we compute the set $\tilde{Q}$ of states of the reversal $\tilde{\mathcal{A}}^\delta$ of $\mathcal{A}$. For $w \in A^*$, the vector $\mu(w)\underline{T}$ is precisely $\underline{w \cdot T}$. Thus, the minimal representation is the restriction of μ to the subspace of K^Q generated by the vectors $\underline{U}$ for $U \in \tilde{Q}$ (considered as a column vector on which each $\mu(w)$ acts on the left).

We illustrate this algorithm in the following examples. In the first one, the representation given by the minimal automaton is the minimal one.

EXAMPLE 18 The minimal automaton $\mathcal{A}$ of $S = a^+$ is represented in Figure 4.1. The linear representation built from the automaton $\mathcal{A}$ is that of Example 17.

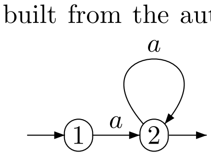

Figure 4.1: The minimal automaton of a^+ in Example 18.

The states of $\tilde{\mathcal{A}}^\delta$ are $\{2\}$ and $\{1, 2\}$. Their characteristic functions are linearly independent. Thus Step 3 does not modify the representation, which is minimal.

In the second example, the representation given by the minimal automaton is not minimal.

EXAMPLE 19 Let X be the bifix code represented in Figure 3.7 on the right (it is the unique maximal bifix code of degree 3 with internal part ab). The minimal automaton $\mathcal{A} = (Q, 1, 1)$ of $S = X^*$ is defined by its transitions given in Table 4.1.

The set $\tilde{Q}$ of states of the reversal automaton $\tilde{\mathcal{A}}^\delta$ are given in Table 4.2.

	1	2	3	4	5
a	2	3	1	3	1
b	4	1	5	5	1

Table 4.1: The transitions of the minimal automaton $\mathcal{A}$ in Example 19.

$\tilde{Q}$	**1**	**2**	**3**	**4**	**5**
	1	2, 4	3, 5	2, 5	3, 4

Table 4.2: The states of the automaton $\tilde{\mathcal{A}}^\delta$ in Example 19.

The vector space generated by the corresponding characteristic vectors has dimension 4 because $\underline{\mathbf{2}} + \underline{\mathbf{3}} = \underline{\mathbf{4}} + \underline{\mathbf{5}}$. Choosing the basis formed of $\underline{\mathbf{1}}, \underline{\mathbf{2}}, \underline{\mathbf{3}}, \underline{\mathbf{4}}$ the minimal representation (λ, μ, γ) of $\sigma = \underline{S}$ is

$$\lambda = [1\ 0\ 0\ 0], \quad \mu(a) = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \mu(b) = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & -1 & -1 \end{bmatrix}, \quad \gamma = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Let us verify for example the value of the first column of $\mu(a)$. We have $\mu(a)\underline{\mathbf{1}} = \underline{a \cdot \{1\}} = \underline{\{3,5\}} = \underline{\mathbf{3}}$. Similarly, the last column of $\mu(b)$ is computed as $\underline{\mu(b)\mathbf{4}} = \underline{b \cdot \{2,5\}} = \underline{\{3,4\}} = \underline{\mathbf{5}} = \underline{\mathbf{2}} + \underline{\mathbf{3}} - \underline{\mathbf{4}}$.

Let $\mu : A^* \to \mathrm{End}(V)$ be a morphism. We let the endomorphisms of V act on the left of the vectors in V. A subspace W of V is said to be *invariant* with respect to μ if for every $x \in W$ and $w \in A^*$, one has $x\mu(w) \in W$. The vector space V is said to be *irreducible* with respect to μ if $V \neq 0$ and if its only invariant subspaces are 0 and V. Finally, it is said to be *completely reducible* with respect to μ if $V = \oplus_{i=1}^n V_i$ where each V_i is an invariant irreducible subspace of V. Let $\mu_i(w)$ be the restriction of $\mu(w)$ to V_i. The representation μ is the direct sum of the representations μ_i which are called the *irreducible components* of μ.

We also say that a linear representation (λ, μ, γ) is completely reducible if the underlying vector space is completely reducible with respect to μ.

Example 20 The linear representation of Example 17 is completely reducible. Indeed, the subspaces generated respectively by $[-1\ 1]$ and by $[0\ 1]$ are invariant and obviously irreducible. In the basis formed by these vectors, the representation takes the following equivalent form.

$$\lambda' = [-1\ 1], \quad \mu'(a) = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \quad \gamma' = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

For $u \in A^*$, we denote by $\sigma \cdot u$ the series defined by $(\sigma \cdot u, v) = (\sigma, uv)$ for every $v \in A^*$.

The *syntactic space* of a series σ, denoted V_σ is the vector space generated by the series $\sigma \cdot u$ for all $u \in A^*$. The *syntactic representation* of σ is the linear representation $\psi_\sigma : K\langle A\rangle \to \mathrm{End}(V_\sigma)$ defined for $\tau \in V_\sigma$ and $u \in A^*$ by

$$\tau\psi_\sigma(u) = \tau \cdot u.$$

The linear representation defined by the triple $(\lambda, \psi_\sigma, \gamma)$ with $\lambda = \sigma$ and γ defined by $\tau\gamma = (\tau, \varepsilon)$ recognizes σ. Indeed, one has for every $w \in A^*$,

$$\lambda\psi_\sigma(w)\gamma = (\sigma \cdot w, \varepsilon) = (\sigma, w).$$

It can be shown that the syntactic representation of a recognizable series is its minimal representation (see [3]).

If a series has a completely reducible representation, then its syntactic representation is also completely reducible.

Example 21 The syntactic space of the series $\sigma = \underline{a^+}$ has dimension 2 and a basis is formed by σ and $\sigma \cdot a = \underline{a^*}$. The corresponding linear representation is given in Example 17.

4.2 Completely reducible sets

A set $S \subset A^*$ is *completely reducible* over K if the syntactic representation of the series $\underline{S}$ is completely reducible. Equivalently, its syntactic algebra is semi-simple [3, Chapter 12].

Example 22 The set a^* is completely reducible since its syntactic space has dimension 1. The set $S = a^+$ is also completely reducible since its syntactic representation is completely reducible by Examples 20 and 21.

We give a second example, in which S is the submonoid generated by a finite bifix code.

Example 23 Consider again the set S of Example 19. We have found a minimal representation of dimension 4. This representation is not irreducible because the space generated by $\underline{\mathbf{1}} + \underline{\mathbf{2}} + \underline{\mathbf{3}}$ is invariant by μ (with $\mu(w)$ acting on the left). The space generated by $\underline{\mathbf{2}} - \underline{\mathbf{1}}$, $\underline{\mathbf{3}} - \underline{\mathbf{1}}$ and $\underline{\mathbf{4}} - \underline{\mathbf{1}}$ is a stable complement and the representation takes in the basis $\underline{\mathbf{1}} + \underline{\mathbf{2}} + \underline{\mathbf{3}}, \underline{\mathbf{2}} - \underline{\mathbf{1}}, \underline{\mathbf{3}} - \underline{\mathbf{1}}, \underline{\mathbf{4}} - \underline{\mathbf{1}}$ the form of a direct sum of two representations μ'_1 of dimension 1 and μ'_2 of dimension 3. In this basis, the vector λ becomes $\lambda' = [1\ \ -1\ \ -1\ \ -1]$ (its components are the values of the linear form defined by λ on each vector of the basis), and $\mu(a)$, $\mu(b)$, γ become

$$\mu'(a) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & -1 & -1 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \mu'(b) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & -1 & -2 & -2 \end{bmatrix}, \quad \gamma' = \begin{bmatrix} 1/3 \\ -1/3 \\ -1/3 \\ 0 \end{bmatrix}$$

The value of γ' results from the computation of $\gamma = \underline{\mathbf{1}} = \frac{1}{3}(\underline{\mathbf{1}} + \underline{\mathbf{2}} + \underline{\mathbf{3}}) - \frac{1}{3}(\underline{\mathbf{2}} - \underline{\mathbf{1}}) - \frac{1}{3}(\underline{\mathbf{3}} - \underline{\mathbf{1}})$. The representations μ'_1 and μ'_2 are irreducible. This is obvious for μ'_1. For μ'_2 it can be proved directly by verifying that the matrices $\mu'_2(w)$ generate the algebra of 3×3-matrices over K or deduced as a consequence of [7, Theorem 2.2]).

Let $S \subset A^*$ be a recognizable set, let $\sigma = \underline{S}$ and let $\mathcal{A} = (Q, i, T)$ be its minimal automaton. Set $\varphi = \varphi_{\mathcal{A}}$, $\psi = \psi_\sigma$ and $M = \psi(A^*)$. By [2, Proposition 14.7.1] for all $u, v \in A^*$, one has

$$\varphi(u) = \varphi(v) \Leftrightarrow \psi(u) = \psi(v). \tag{4.1}$$

In particular, M and $\varphi(A^*)$ are isomorphic.

An element of $M = \psi(A^*)$ is a linear map from V_σ into itself and, as such, it has a kernel and an image which are subspaces of V_σ.

The *eventual kernel* of S, denoted $EK(S)$, is the intersection of the kernels of all elements of minimal nonzero rank of M.

Symmetrically, the *eventual range* of S, denoted $ER(S)$, is the subspace spanned by the images of all elements of minimal nonzero rank of M.

Both $EK(S)$ and $ER(S)$ are invariant subspaces of V_σ. Indeed, let $x \in EK(S)$ and let $w \in A^*$. For any $m \in M$ of minimal nonzero rank, $\psi(w)m$ is either zero or has minimal nonzero rank . In both cases $x\psi(w)m = 0$. Thus $x \in EK(S)$. Similarly, let $x \in ER(S)$ and let $w \in A^*$. Since $x \in ER(S)$, we have $x = \sum_{m \in J} x_m m$ where J denotes the set of elements of minimal nonzero rank in M and $x_m \in V_\sigma$. Then $x\psi(w) = \sum_{m \in J} v_m m\psi(w)$ and thus $x\psi(w) \in ER(S)$.

Theorem 4.2.1 *A recurrent set is completely reducible if and only if $EK(S) = \{0\}$.*

We first prove the following statement.

Proposition 4.2.2 *Let S be a recognizable set. If S is completely reducible, then $EK(S) \cap ER(S) = \{0\}$.*

Proof Set $\sigma = \underline{S}$ and $V = V_\sigma$. Let J be the set of elements of $M = \psi(A^*)$ of nonzero minimal rank. Set $W = EK(S) \cap ER(S)$. Since W is invariant, and since S is completely reducible, there is a complement W' of W which is an invariant subspace of V.

Let $v \in W$. Then, as for any element of $ER(S)$, we have $v = \sum_{m \in J} \alpha_m(\sigma m)$ for some $\alpha_m \in K$. Since $V = W \oplus W'$, there exist $\rho \in W$ and $\rho' \in W'$ such that $\sigma = \rho + \rho'$. Then

$$\begin{aligned} v &= \sum_{m \in J} \alpha_m(\sigma m) = \sum_{m \in J} \alpha_m(\rho m) + \sum_{m \in J} \alpha_m(\rho' m) \\ &= \sum_{m \in J} \alpha_m(\rho' m). \end{aligned}$$

Indeed, since $\rho \in W$, we have $\rho m = 0$ for every $m \in J$. This implies that $v \in W'$, and finally $v = 0$. Therefore we conclude that $EK(S) \cap ER(S) = \{0\}$. ■

Proof of Theorem 4.2.1. Let $\mathcal{A} = (Q, i, T)$ be the minimal automaton of the recurrent set S. Set $\varphi = \varphi_{\mathcal{A}}$, $\sigma = \underline{S}$, $\psi = \psi_\sigma$ and $M = \psi(A^*)$.

Assume first that the eventual kernel of S is 0.

It is enough to prove the property under the hypothesis that S contains the empty word. Indeed, Let S' be the set recognized by the automaton $\mathcal{A}' = (Q, t, T)$ for some $t \in T$. Then S' is a set recognized by a strongly connected deterministic automaton and S' contains the empty word. Since $\mathcal{A}$ is strongly connected, S is a residual of S' and S' is a residual of S. Thus the syntactic representations of S and S' only differ by the choice of the initial vector. Thus S is completely reducible whenever S' is and the eventual kernel of S' is also 0.

Since $\mathcal{A}$ is strongly connected, the monoid $\varphi(A^*)$ has a unique 0-minimal ideal K which is a regular $\mathcal{D}$-class plus 0 (see Section 3). Let $x \in A^*$ be such that $\varphi(x)$ is an idempotent of K. We may assume that $i \in Q \cdot x$. Indeed, since $\mathcal{A}$ is strongly connected, the state i is, as any state of Q, in the image of some word w of minimal nonzero rank. Since $K \setminus 0$ is regular $\mathcal{D}$-class, the $\mathcal{L}$-class of w contains an idempotent which has the same image as w. Since $\varphi(x)$ is idempotent, $i \in Q \cdot x$ implies that $i \cdot x = i$. Since $M = \psi(A^*)$ is isomorphic to $\varphi(A^*)$, $e = \psi(x)$ is an idempotent of M.

Set $V = V_\sigma$. We verify that the conditions of [7, Corollary 1] are satisfied by V and e.

(i) The subspace Ve is completely reducible with respect to the restriction of ψ to xA^*x. This is true by Maschke's Theorem asserting that any linear representation of a finite group on a field with characteristic zero is completely reducible. Indeed, $\psi(xA^*x)$ is a finite group or a finite group union zero.

(ii) The space V is generated by VeM. Indeed $i \cdot x = i$ implies $\lambda e = \lambda$. Thus VeM contains $\lambda eM = \lambda M$ which generates V.

(iii) We have $\{v \in V \mid vMe = 0\} = 0$. Indeed, $vme = 0$ if and only if v is in the kernel of me. Since K is 0-minimal, we have $K = MeM$ and thus the kernel of any element of minimal rank is equal to the kernel of some me. Hence we have $vMe = 0$ if and only if v belongs to the intersection of the kernels of the elements of K.

Thus we can conclude that ψ is completely reducible.

Conversely, since S is recurrent, we have $ER(S) = V$ and thus $EK(S) = \{0\}$ by Proposition 4.2.2. ■

We obtain as a corollary [7, Theorem 5.2], which is a generalization of the result of [10] asserting that the submonoid generated by a bifix code is completely reducible.

Corollary 4.2.3 *Any recognizable birecurrent set is completely reducible.*

Proof Let $\mathcal{A} = (Q, i, T)$ be the minimal automaton of the set S. Set $V = K^Q$ and let (λ, μ, γ) be the linear representation associated to the automaton $\mathcal{A}$. Thus $\gamma = \underline{T}$. The minimal representation of $\underline{S}$ is, as seen before, the restriction μ' of μ to the subspace V' of column vectors generated by the vectors $\mu(w)\gamma$. We consider μ' as a representation acting on the right on a supplementary subspace W of the ortogonal of V'.

By Theorem 3.2.1, T is saturated by a word x of minimal nonzero rank. Let K be the 0-minimal ideal of the monoid $M = \mu(A^*)$. Then $\mu(x) \in K$. Let e be an idempotent of K in the $\mathcal{R}$-class of $\mu(x)$ and let y be such that $\mu(y) = e$. Then $\varphi_{\mathcal{A}}(y)$ is an idempotent. Since $\mu(x)$ and $\mu(y)$ belong to the same $\mathcal{R}$-class, x and y have the same kernel. Thus T is a union of classes of the kernel of y, which implies $y \cdot T = T$ and thus $e\gamma = \gamma$.

Let $v \in W$ be such that $vm = 0$ for every $m \in K$. Then, for any $m \in M$, we have $vm\gamma = vme\gamma$ since $e\gamma = \gamma$. But since $me \in K \cup 0$, we have $vme\gamma = 0$ and thus $vm\gamma = 0$ for all $m \in M$. This shows that v is orthogonal to V', and thus that $v = 0$.

We conclude that $EK(S) = 0$ and thus that S is completely reducible by Theorem 4.2.1. ∎

Note the following precision on Theorem 4.2.1. Let M be a finite monoid having a unique 0-minimal ideal J. Then all $\mathcal{H}$-classes of J which are groups are isomorphic. The *Suschkevitch group* of M is any of them (see [2] or [11]).

Proposition 4.2.4 *Let S be a recurrent completely reducible set, let $\sigma = \underline{S}$ and let $\psi = \psi_\sigma$. The number of irreducible constituents of ψ is equal to the number of irreducible contituents of its restriction to the Suschkevitch group of $\psi(A^*)$.*

Proof Let $x \in A^*$ be a word such that $e = \psi(x)$ is an idempotent of the 0-minimal ideal of the monoid $\psi(A^*)$. Set $V = V_\sigma$. The restriction of ψ to group $G = \psi(xA^*x)$ is a representation of G on Ve. Moreover, by [7, Corollary 1], if $V = \oplus_{i=1}^m V_i$ is a decomposition of V in irreducible subspaces, then $Ve = \oplus_{i=1}^m V_i e$ is a decomposition of Ve in irreducible subspaces. Thus the number of irreducible components of ψ is equal to the number of irreducible components of the restriction of ψ to G. ∎

Note that, in particular, if the Suschkevitch group of $\psi(A^*)$ is trivial, then ψ is irreducible.

The following example shows that the hypothesis of Theorem 4.2.1 can be satisfied by a set which is not birecurrent.

Example 24 Consider again the strongly connected automaton $\mathcal{A}$ of Figure 3.1 on the left with its deterministic reversal on the right (Example 3). We change the automaton $\mathcal{A}$ into an automaton $\mathcal{A}'$ by choosing this time $T' = \{2\}$ as set of terminal states. The automata $\mathcal{A}'$ and $\tilde{\mathcal{A}'}^{\delta}$ are represented in Figure 4.2. Let

S' be the set recognized by $\mathcal{A}'$. Since $\tilde{\mathcal{A}'}^{\delta}$ is not strongly connected, S' is not birecurrent.

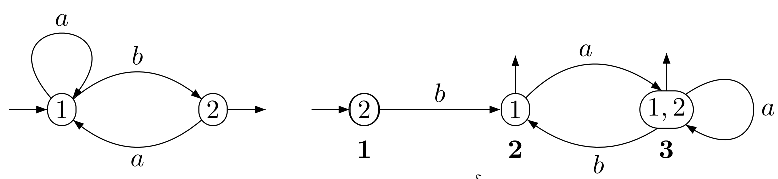

Figure 4.2: The automata $\mathcal{A}'$ and $\tilde{\mathcal{A}'}^{\delta}$ in Example 24.

	1	2
1, 2	* a	* ab
1	* ba	b

	2	**1**
1, 2	* a	* ab
2	* ba	b

Figure 4.3: The 0-minimal ideals of $\mathcal{A}$ and $\tilde{\mathcal{A}}^{\delta}$ in Example 3.

The 0-minimal ideal of the monoid $\varphi_{\mathcal{A}}(A^*) = \varphi_{\mathcal{A}'}(A^*)$ is represented in Figure 4.3 on the left and the set Q' of states of $\tilde{\mathcal{A}'}^{\delta}$ in Table 4.3. Except $\{2\}$,

$\tilde{Q}'$	**1**	**2**	**3**
	2	1	1, 2

Table 4.3: The states of the automaton $\tilde{\mathcal{A}'}^{\delta}$ in Example 24.

all states of $\tilde{\mathcal{A}'}^{\delta}$ are classes of the kernel of an element of minimal rank (see Figure 4.3). The space generated by the characteristic functions of these states has dimension 2, as for the same space corresponding to $\mathcal{A}'$. Indeed $T = \{1\}$ is an element of $\tilde{Q}'$ and $\underline{T'}$ is in the space generated by the characteristic functions of the states of $\mathcal{A}$ since $\underline{\{2\}} = \underline{\{1,2\}} - \underline{\{1\}}$. Thus S' is completely reducible althought it is not birecurrent.

We now prove a second corollary of Theorem 4.2.1 (actually of Proposition 4.2.2). It shows that a completely reducible set which is dense satisfies a property which is well-known for the submonoid generated by a maximal bifix code.

Corollary 4.2.5 *Let S be a set recognized by a strongly connected unambiguous finite automaton $\mathcal{A}$. If S is completely reducible and dense, then $\varphi_{\mathcal{A}}(S)$ meets every $\mathcal{H}$-class of the minimal ideal of the monoid $\varphi_{\mathcal{A}}(A^*)$.*

Proof We use the same notation as in the proof of Theorem 4.2.1. Since $\mathcal{A}$ is strongly connected and S is dense, the monoid $\varphi(A^*)$ does not contain 0 and has a unique minimal ideal which is a union of groups. Moreover $\varphi(S)$ meets the minimal ideal of $\varphi(A^*)$. Since $M = \psi(A^*)$ is the image of $\varphi(A^*)$ by a morphism, it has also a unique minimal ideal which meets $\psi(S)$. Let J be the minimal ideal of the monoid $M = \psi(A^*)$.

For a subset H of M, denote $\underline{H} = \sum_{m \in H} m$.

Note that for any $\mathcal{H}$-class H of J, $(\sigma\underline{H}, \varepsilon) = \mathrm{Card}(H \cap \psi(S))$. Indeed, let $m \in H$ and let $x \in A^*$ be such that $\psi(x) = m$. Then $(\sigma m, \varepsilon) = (\sigma\psi(x), \varepsilon) = (\sigma, x)$ which is 1 if $x \in S$ and 0 otherwise.

Let H, K be two $\mathcal{H}$-classes of J of the monoid M contained in the same $\mathcal{R}$-class R. Then the vector $\sigma\underline{H} - \sigma\underline{K}$ belongs to the eventual range $ER(S)$.

Moreover, it belongs to the eventual kernel $EK(S)$. Indeed, for every $m \in J$, we have $Hm = Km$ since both are equal to the $\mathcal{H}$-class $R \cap Mm$. Thus, by Theorem 4.2.1, we have $\sigma\underline{H} = \sigma\underline{K}$.

By Proposition 4.2.2, this forces $\sigma\underline{H} - \sigma\underline{K} = 0$ for all $\mathcal{H}$-classes H, K included in the same $\mathcal{R}$-class.

Since $\tilde{S}$ satisfies the same hypotheses as S (using the automaton $\tilde{\mathcal{A}}$), the conclusion follows. ∎

We give below an example illustrating Corollary 4.2.5

Example 25 Consider again the deterministic automaton $\mathcal{A} = (Q, i, T)$ of Example 9. We change the automaton $\mathcal{A}$ into an automaton $\mathcal{A}' = (Q, i, T')$ by choosing this time $T' = \{1, 2, 6\}$ as set of terminal states instead of $T = \{1, 2, 5\}$. Let S' be the set recognized by $\mathcal{A}'$. The minimal ideal of the monoid $\varphi_{\mathcal{A}}(A^*) = \varphi_{\mathcal{A}'}(A^*)$ is represented in Figure 4.4. The set $\tilde{Q}'$ of states of the automaton

	1/3	2/4	5/6
1, 4, 5/2, 3, 6	* b^2	* b^2a	* b^2ab
1, 2, 5/3, 4, 6	* ab^2	*	*
1, 4, 6/2, 3, 5	* a^2b^2	*	*

Figure 4.4: The minimal ideal in Example 25.

$\tilde{\mathcal{A}'}^\delta$ is represented in Table 4.4. Except $\{1, 2, 6\}$, all states of $\tilde{\mathcal{A}'}^\delta$ are classes

$\tilde{Q}'$	**1**	**2**	**3**	**4**	**5**	**6**	**7**
	1, 2, 6	1, 4, 6	3, 4, 6	2, 3, 5	1, 4, 5	2, 3, 6	1, 2, 5

Table 4.4: The states of the automaton $\tilde{\mathcal{A}'}^\delta$.

of the kernel of an element of minimal rank (see Figure 4.4). The space generated by the characteristic functions of these states has dimension 4, as for the same space corresponding to $\mathcal{A}'$. Indeed $T = \{1, 2, 5\}$ is an element of $\tilde{Q}'$ and $\underline{T'}$ is in the space generated by the characteristic functions of the states of $\mathcal{A}$ since $\underline{\{1, 2, 6\}} = \underline{\{1, 2, 5\}} - \underline{\{1, 4, 5\}} + \underline{\{1, 4, 6\}}$. Thus S' is completely reducible althought it is not birecurrent. To see that S' is not birecurrent, one may either use the fact that T' is not saturated by any word of minimal rank and apply Theorem 3.2.1 or compute directly $\tilde{\mathcal{A}'}^{\delta}$ (see Figure 4.5). Note that

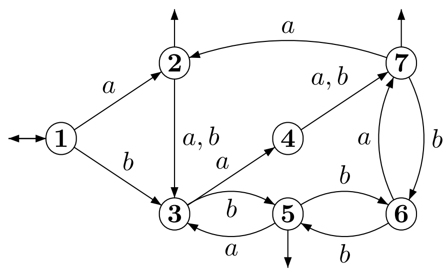

Figure 4.5: The automaton $\tilde{\mathcal{A}}'$

the hypothesis of Theorem 4.2.1 is satisfied because it is already satisfied by the minimal representation of $\underline{S}$, which is the same as that of $\underline{S'}$ except for the vectors $\gamma = \underline{T}$ and $\gamma' = \underline{T'}$.

We finally deduce from Corollary 4.2.5 the following result, originally proved in [10] (see also [2, Theorem 14.7.7]).

Corollary 4.2.6 *Let X be a recognizable maximal code. If X^* is completely reducible, then X is bifix.*

Proof Let $\mathcal{A}$ be the minimal automaton of X^*. Since X is a recognizable code, it is thin by [2, Proposition 2.5.20]. Since X is a thin maximal code, X^* is dense by [2, Theorem 2.5.5]. As for any recognizable code, there is a trim unambiguous finite automaton $\mathcal{A} = (Q, i, i)$ recognizing X^* (see [2, Proposition 4.1.2]). Since $\mathcal{A}$ is trim with a unique initial state equal to the unique terminal state, it is strongly connected. By Corollary 4.2.5, $\varphi_{\mathcal{A}}(X^*)$ meets every $\mathcal{H}$-class of the minimal ideal of $\varphi_{\mathcal{A}}(A^*)$. This implies that X^* is recurrent and thus X is prefix by [2, Proposition 3.3.11]. Symmetrically, X is suffix, whence the conclusion. ∎

4.3 A characterization of completely reducible sets

In this section, we consider rational series over $\mathbb{Q}$ whose set of coefficients is finite. It is known that the minimal representation (λ, μ, γ) of such a series satisfies the following finiteness property: the matrix monoid $\mu(A^*)$ is finite. This follows from a theorem of Schützenberger, see [3, Corollary 2.3].

Moreover, such a series is a linear combination over $\mathbb{Q}$ of characteristic series of rational languages. This follows from the fact that for any $\alpha \in \mathbb{Q}$, the set

$\{w \in A^* \mid (S,w) = \alpha\}$ is rational. For this, see [3, Theorem 2.10], another theorem of Schützenberger.

For a rational series S whose set of coefficients is finite, we may construct a special kind of deterministic automaton, called an automaton *with scalar output function*. It is a deterministic automaton $\mathcal{A} = (Q, i, \tau)$ where $\tau : Q \to \mathbb{Q}$ is a mapping called the *terminal function*, which recognizes S in the following sense. For each word w, one has $(S, w) = \tau(i \cdot w)$. In other words, one reads w on the automaton, starting from the initial state i and one reaches a state $q = i \cdot w$. The coefficient (S, w) of w in S is $\tau(q)$. If there is no path from i labeled w, we let $(S, w) = 0$.

The notions of trim automaton and of minimal automaton extend easily to these automata.

For a word $u \in A^*$ and a series S, we denote here $u^{-1}S$ the series defined by $(u^{-1}S, v) = (S, uv)$ (this series is denoted $S \cdot u$ in Section 4.1) and symmetrically Su^{-1} the series defined by $(Su^{-1}, v) = (S, vu)$.

There is also a Nerode criterium for these series. Indeed, a series S is rational (and has a finite set of coefficients) if and only if the set $\{u^{-1}S \mid u \in A^*\}$ is finite.

We say that a rational series with a finite number of coefficients is *recurrent* if its minimal automaton with scalar output function is strongly connected.

Clearly, a series S is recurrent if and only if it is recognized by a strongly connected automaton with scalar output function. Moreover, if S is recurrent, then so are $u^{-1}S$ and Su^{-1} for any word u, since they are recognized by the minimal automaton with output function of S.

It is called *birecurrent* if S and $\tilde{S}$ are both recurrent, where $\tilde{S}$ is the series such that $(\tilde{S}, w) = (S, \tilde{w})$ for all $w \in A^*$.

Proposition 4.3.1 *Let S be a birecurrent series. Then S is a linear combination over $\mathbb{Q}$ of characteristic series of birecurrent sets.*

Proof Let $I = \{\alpha \in \mathbb{Q} \setminus 0 \mid (S, w) = \alpha \text{ for some } w \in A^*\}$. Then I is finite. It is enough to show that for any $\alpha \in I$, the set $L = \{w \in A^* \mid (S, w) = \alpha\}$ is recurrent.

Let (Q, i, τ) be a strongly connected automaton with scalar output function recognizing S. Let (Q, i, T) be the automaton defined by $T = \{q \in Q \mid \tau(q) = \alpha\}$. Then this automaton recognizes L. It is deterministic and strongly connected and thus L is recurrent. ■

Proposition 4.3.2 *Let S be a completely reducible series. Then S is a linear combination over $\mathbb{Q}$ of birecurrent series.*

Proof Let (λ, μ, γ) be a minimal representation of S. Since the set of coefficients of S is finite, we know that $\mu(A^*)$ is a finite monoid.

Since the representation is completely reducible, it is isomorphic to a direct sum of representations $(\lambda_i, \mu_i, \gamma_i)$ which are irreducible over $\mathbb{Q}$.

It follows that S is the corresponding sum of irreducible series. We may therefore assume that (λ, μ, γ) is irreducible. Note that $\mu(A^*)$ is finite.

The set $R(\lambda) = \{\lambda\mu(u) \mid u \in A^*\}$ is finite. One obtains a right action of A^* on this set by $\lambda\mu(u) \cdot w = \lambda\mu(uw)$. There exists therefore a word u such that the set $R(\lambda\mu(u))$ is a minimal invariant subset for this action.

Similarly, A^* acts on the left on the finite set $L(\gamma) = \{\mu(v)\gamma \mid v \in A^*\}$ by $w \cdot \mu(v)\gamma = \mu(wv)\gamma$. There exists similarly a word v such that the set $L(\mu(v)\gamma)$ is minimal for this action.

Consider the linear representation $(\lambda\mu(u), \mu, \mu(v)\gamma)$. It recognizes a rational series, call it T, whose set of coefficients is finite since $\mu(A^*)$ is finite.

For this series T, we may construct the following deterministic automaton with scalar output function. Its set of states is $R(\lambda\mu(u))$, its initial state is $\lambda\mu(u)$ and its terminal function is defined by $\tau(\lambda') = \lambda'\mu(v)\gamma$ for any state $\lambda' \in R(\lambda\mu(u))$.

Clearly this automaton recognizes T. Since its set of states is a minimal invariant subset for the right action of A^*, this automaton is strongly connected. Thus T is recurrent.

By symmetry, $\tilde{T}$ is also recurrent. Hence T is birecurrent.

Turning back to the representation (λ, μ, γ), there exists, since it is irreducible, polynomials $X, Y \in \mathbb{Q}\langle A\rangle$ such that $\lambda\mu(u)\mu(X) = \lambda$ and $\mu(Y)\mu(v)\gamma = \gamma$ where the monoid morphism $\mu : A^* \to \mathbb{Q}^{n\times n}$ is extended to an algebra morphism still denoted μ from $\mathbb{Q}\langle A\rangle$ into $\mathbb{Q}^{n\times n}$.

Finally, for any word w, one has

$$\begin{aligned}
(S, w) = &= \lambda\mu(w)\gamma \\
&= \lambda\mu(u)\mu(X)\mu(w)\mu(Y)\mu(v)\gamma \\
&= \lambda\mu(u)\left(\sum_{x\in A^*}(X, x)\mu(x)\right)\mu(w)\left(\sum_{y\in A^*}(Y, y)\mu(y)\right)\mu(v)\gamma \\
&= \sum_{x,y}(X, x)(Y, y)\lambda\mu(u)\mu(x)\mu(w)\mu(y)\mu(v)\gamma
\end{aligned}$$

Thus, since $T = \sum_w \lambda\mu(u)\mu(w)\mu(v)\gamma$, we have $x^{-1}Ty^{-1} = \sum_w \lambda\mu(u)\mu(x)\mu(w)\mu(y)\mu(v)\gamma$, and therefore $S = \sum_{x,y}(X, x)(Y, y)x^{-1}Ty^{-1}$. Since T is birecurrent, each $x^{-1}Ty^{-1}$ is birecurrent. Hence S is a linear combination of birecurrent series. ■

Proposition 4.3.3 *Let $S \in \mathbb{Q}\langle\langle A\rangle\rangle$ be a rational series whose set of coefficients is finite. Then S is completely reducible if and only if it is a linear combination over $\mathbb{Q}$ of characteristic series of birecurrent sets.*

Proof Suppose that S is completely reducible. Then the conclusion follows by Propositions 4.3.1 and 4.3.2.

Conversely, suppose that S is a linear combination over $\mathbb{Q}$ of characteristic series of birecurrent sets. By Corollary 4.2.3, each such series is completely reducible. By [7, Proposition 4.1], a linear combination over $\mathbb{Q}$ of completely reducible series is completely reducible. Thus S is completely reducible. ■

We thus obtain as a main result of this section.

Theorem 4.3.4 *A language is completely reducible if and only if its characteristic series is $\mathbb{Q}$-linear combination of characteristic series of birecurrent languages.*

The following example shows that the linear combination need not have coefficients in $\mathbb{Z}$.

Example 26 Let $\mathcal{A} = (Q, i, T)$ be the automaton represented in Figure 4.6 on the left with its reversal on the right.

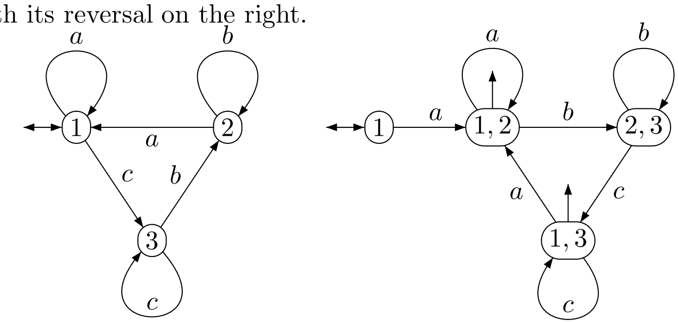

Figure 4.6: A finite automaton and its deterministic reversal.

Let X be the set recognized by $\mathcal{A}$. Since $\mathcal{A}$ is strongly connected, X is recurrent. Since $\tilde{A}^\delta$ is not strongly connected, X is not birecurrent. Let X_i be the set recognized using, instead of $T = \{1\}$, the set of terminal states T_i for $1 \leq i \leq 3$ with $T_1 = \{2, 3\}$, $T_2 = \{1, 3\}$ and $T_3 = \{1, 2\}$. Since these sets are saturated respectively by b, c and a, which have rank 1, the sets X_i are birecurrent (this can also be seen easily in Figure 4.6). Since $\underline{T} = \underline{T_3} - \underline{T_1} - \underline{T_2}$, we have $X = \frac{1}{2}(X_3 - X_1 + X_2)$. Note that, by Proposition 4.2.4, the linear representation (λ, μ, γ) associated to the automaton $\mathcal{A}$ is irreducible. Indeed, this representation is clearly minimal and since the letters are of rank 1, the Suschkevitch group of the monoid $\mu(A^*)$ is trivial. Thus we obtain a decomposition of an irreducible set X as a $\mathbb{Q}$-linear combination of birecurrent sets. We conjecture that X cannot be obtained as a linear combination of birecurrent sets with coefficients in $\mathbb{Z}$.

5 Unambiguous automata

In this section, we generalize some of the results concerning birecurrent sets using unambiguous automata, which are nondeterministic automata closely linked with linear representation of series. For an introduction to this class of automata, see [2].

5.1 Unambiguous automata and linear representations

An automaton $\mathcal{A} = (Q, I, T)$ is *unambiguous* if for any word $w \in A^*$ there is at most one path from a state $i \in I$ to a state $t \in T$ labeled w.

A trim unambiguous automaton has the following property (used in [2] as a definition of unambiguous automata). For every word $w \in A^*$ and every pair $(p, q) \in Q$ of states there is at most one path from p to q labeled w. Indeed, since $\mathcal{A}$ is trim, there is a path $i \xrightarrow{u} p$ with $i \in I$ and a path $q \xrightarrow{v} t$ with $t \in T$. Since there is at most one path from i to t labeled uwv, there is at most one path from p to q labeled w.

Clearly a deterministic automaton is unambiguous.

Example 27 The automaton $\mathcal{A} = (Q, 1, 1)$ represented in Figure 5.1 is unambiguous. One can check this by computing the automaton of pairs and by

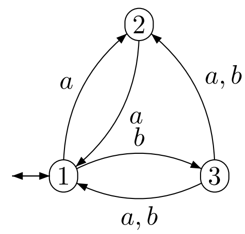

Figure 5.1: An unambiguous automaton.

checking that there is no path from a pair (p, p) to a pair (q, q) using a pair (r, s) with $r \neq s$.

To every unambiguous automaton $\mathcal{A} = (Q, I, T)$ recognizing a set S, we may associate a linear representation recognizing the characteristic series $\underline{S}$ of the set S. We consider, as in Section 2 a field K and the vector space $V = K^Q$. Consider the representation (λ, μ, γ) where $\lambda \in V$ is the characteristic function $\underline{I}$ of I considered as a row vector, $\gamma = \underline{T}$ considered as a column vector and for $a \in A$, $\mu(a)$ is the $Q \times Q$-matrix $\mu_{\mathcal{A}}(a)$ defined by

$$(p, \mu_{\mathcal{A}}(a), q) = \begin{cases} 1 & \text{if } p \xrightarrow{a} q \\ 0 & \text{otherwise.} \end{cases}$$

Then μ extends to a morphism from A^* into $\mathrm{End}(V)$ and $\lambda\mu(w)\gamma = (\underline{S}\,, w)$ for every $w \in A^*$. When $\mathcal{A}$ is trim, all the matrices $\mu_{\mathcal{A}}(w)$ for $w \in A^*$ have coefficients 0 or 1.

Example 28 The linear representation corresponding to the unambiguous automaton of Figure 5.1 is

$$\lambda = [1\ 0\ 0], \quad \mu(a) = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \end{bmatrix}, \quad \mu(b) = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 1 & 0 \end{bmatrix}, \quad \gamma = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}.$$

For a set $U \subset Q$ of states and a word w, we denote

$$\begin{aligned} U \cdot w &= \{q \in Q \mid u \xrightarrow{w} q \text{ for some } u \in U\}, \\ w \cdot U &= \{q \in Q \mid q \xrightarrow{w} u \text{ for some } u \in U\}. \end{aligned}$$

Let $\mathcal{A}$ be an unambiguous automaton. Recall that we denote by $\mathcal{A}^\delta$ the determinization of $\mathcal{A}$. Thus the states of $\mathcal{A}^\delta$ are the nonempty sets $I \cdot w$ for $w \in A^*$. The initial state is I and the set of terminal states is the set of $U = I \cdot w$ such that $U \cap T \neq \emptyset$. Note that since $\mathcal{A}$ is unambiguous, $U \cap T$ contains at most one element and that $\mathcal{A}$ and $\mathcal{A}^\delta$ recognize the same set of words.

When $\mathcal{A}$ is trim, the action of A^* on the states of $\mathcal{A}^\delta$ is the same as the right multiplication by the matrices $\mu_{\mathcal{A}}(w)$ on the characteristic vectors. Indeed, one has $U = I \cdot w$ if and only if $\underline{U} = \underline{I}\, \mu_{\mathcal{A}}(w)$.

Also recall that we denote by $\tilde{\mathcal{A}}^\delta$ the deterministic reversal of the automaton $\mathcal{A}$.

Example 29 Let $\mathcal{A}$ be the unambiguous automaton of Figure 5.1. The automaton $\mathcal{A}^\delta$ is represented in Figure 5.2 on the left and the automaton $\tilde{\mathcal{A}}^\delta$ on the right.

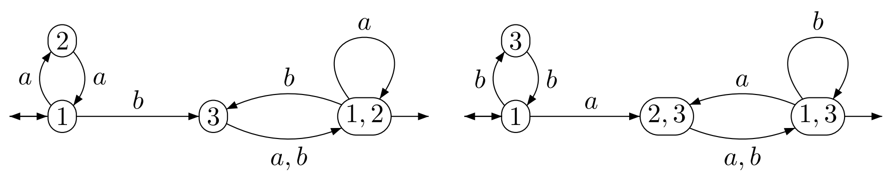

Figure 5.2: The automata $\mathcal{A}^\delta$ and $\tilde{\mathcal{A}}^\delta$.

The *rank* of a word w with respect to an unambiguous automaton $\mathcal{A}$ is the rank of the linear map $\mu_{\mathcal{A}}(w)$. This definition is consistent with the one given for a deterministic automaton in Section 3. Indeed, when $\mathcal{A}$ is deterministic, the matrix $\mu_{\mathcal{A}}(w)$ is the matrix of a partial map from Q into itself and the rank of $\mu_{\mathcal{A}}(w)$ is equal to the rank of the map. The definition of rank for an unambiguous automaton given in [2] is different but equivalent to this one (see [2, Exercise 9.3.2]).

Let $\mathcal{A} = (Q, I, T)$ be a trim unambiguous automaton. Then the monoid $\mu_{\mathcal{A}}(A^*)$ is formed of $\{0, 1\}$-matrices and thus it is finite. As for a deterministic automaton, the set of elements of minimal nonzero rank of the monoid $M = \mu_{\mathcal{A}}(A^*)$ is the unique 0-minimal ideal J of M. It is the union of all 0-minimal right (resp. left) ideals. It is formed of a regular $\mathcal{D}$-class, plus possibly 0. Each $\mathcal{R}$-class of $J \setminus \{0\}$ is formed of elements which have the same set of rows. Each $\mathcal{H}$-class of $J \setminus \{0\}$ which is a group is a transitive permutation group on the common set of rows of its elements.

Example 30 Let $\mathcal{A}$ be the automaton represented in Figure 5.1. The minimal ideal J of the monoid $M = \mu_{\mathcal{A}}(A^*)$ is represented in Figure 5.3. There is no

	3	1, 2
2, 3	$^*ab$	$^*aba$
1, 3	$^*bab$	$^*ba$

Figure 5.3: The minimal ideal of M.

word of rank 0 and the minimal rank is 1. For example, we have

$$\mu_{\mathcal{A}}(a) = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \end{bmatrix}, \quad \mu_{\mathcal{A}}(ab) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix},$$

the first one being of rank 2 and the second of rank 1. We represent for each $\mathcal{R}$-class of J the set of rows of the elements (representing a row as a set of states) and for each $\mathcal{L}$-class its set of columns.

5.2 Unambiguous automata and birecurrent sets of finite type

The following result is a generalization of Theorem 3.2.1 which gives a sufficient condition for the set recognized by an unambiguous automaton to be recurrent. The proof is quite similar.

Theorem 5.2.1 *Let $\mathcal{A} = (Q, I, T)$ be a strongly connected unambiguous finite automaton. The automaton $\mathcal{A}^\delta$ is strongly connected if and only if there is a word x of minimal nonzero rank such that $I \cdot w = I$.*

Proof Assume first that $\mathcal{A}^\delta$ is strongly connected. Let x be a word of minimal nonzero rank. Since x has nonzero rank there exist p, q such that $p \xrightarrow{x} q$. Since $\mathcal{A}$ is strongly connected, there is a word u such that $i \xrightarrow{u} p$ for some $i \in I$. Then $y = ux$ is a word of minimal nonzero rank such that $I \cdot y \neq \emptyset$. Thus $I \cdot y$ is a state of $\mathcal{A}^\delta$. Since $\mathcal{A}^\delta$ is strongly connected, there is a word z such that $I \cdot yz = I$. Since yz has minimal nonzero rank, the conclusion follows.

Conversely, assume that $I = I \cdot w$ with w of minimal nonzero rank. Consider a state U of $\mathcal{A}^\delta$. Then $U = I \cdot u$ for some $u \in A^*$ such that $I \cdot u \neq \emptyset$. Then $I \cdot wu = I \cdot u$ and thus wu is a word of minimal nonzero rank. Set $\mu = \mu_{\mathcal{A}}$. Since the right ideal generated by $\mu(w)$ is 0-minimal, there is a word $v \in A^*$ such that $\mu(wuv) = \mu(w)$. Let $e = \mu(uv)^n$ with $n \geq 1$ be the idempotent which is a power of $\mu(uv)$. Since $\mu(wuv) = \mu(w)$, we have $\mu(w)e = \mu(w)$. Then $\underline{I}e = \underline{I \cdot w}e = \underline{I}\mu(w)e = \underline{I}\mu(w) = \underline{I}$. Set $x = v(uv)^{n-1}$. Then $\underline{I}\mu(ux) = \underline{I}e = \underline{I}$ and thus $I \cdot ux = I$. This shows that I and $I \cdot u$ belong to the same strongly connected component and thus that $\mathcal{A}^\delta$ is strongly connected. ■

A symmetric result holds for the automaton $\tilde{\mathcal{A}}^\delta$, which is strongly conected if and only if there is a word x of minimal nonzero rank such that $x \cdot T = T$.

Theorem 3.2.1 follows easily from the symmetric version of Theorem 5.2.1. In fact, assume that $\mathcal{A}$ is deterministic. A set U is saturated by a word of minimal nonzero rank if and only if there is a word x of minimal nonzero rank such that $x \cdot U = U$. Indeed, the condition is sufficient. Conversely, if $U = x \cdot V$ for some $V \subset Q$ and some x of minimal nonzero rank, let y be a word such that $\varphi_{\mathcal{A}}(y)$ is a nonzero idempotent in the right ideal generated by $\varphi_{\mathcal{A}}(x)$. Then $\varphi_{\mathcal{A}}(yx) = \varphi_{\mathcal{A}}(x)$ and thus $y \cdot U = yx \cdot V = x \cdot V = U$.

Corollary 5.2.2 *Let $\mathcal{A} = (Q, I, T)$ be a strongly connected unambiguous finite automaton recognizing a set S. If there are words x, y of 0-minimal rank such that $I \cdot x = I$ and $y \cdot T = T$, then S is birecurrent.*

Proof By Theorem 5.2.1, since $I \cdot x = I$, the automaton $\mathcal{A}^\delta$ is strongly connected and conversely. Symmetrically, since $y \cdot T = T$, the automaton $\tilde{\mathcal{A}}^\delta$ is strongly connected and conversely. ∎

Iterating the construction of Section 3.4, one obtains examples of birecurrent sets defined by unambiguous automata as in Corollary 5.2.2. The iteration relies on the following result from [15] (see [2, Exercise 14.1.9]) where we use again the notation δ_w and γ_w introduced after Proposition 3.4.1.

Proposition 5.2.3 *Let Z be a finite maximal prefix code and let w^2 be a pure square for Z. Then w^4 is a pure square for $X = \delta_w(Z)$ and for $Y = \gamma_w(Z)$. The sets $G' = X(w^2)^{-1}$ and $D' = (w^2)^{-1}X$ satisfy*

$$\underline{G'} - 1 = (1 + w)(\underline{G} - 1), \quad \underline{D'} - 1 = (1 + w)(\underline{D} - 1)$$

and the sets $G'' = Y(w^2)^{-1}$ and $D'' = (w^2)^{-1}Y$ satisfy

$$\underline{G''} - 1 = (\underline{G} - 1)(1 + w), \quad \underline{D''} - 1 = (\underline{D} - 1)(1 + w).$$

We deduce from Proposition 5.2.3 the following result.

Theorem 5.2.4 *Let Z be a finite maximal bifix code, let w be a pure square for Z and let $U = \gamma_{w^2}(\delta_w(Z))$. Then $\{\varepsilon, w^2\}U^*\{\varepsilon, w\}$ is a birecurrent set of finite type.*

Proof Set $X = \delta_w(Z)$, $Y = \gamma_w(Z)$, $L = \delta_{w^2}(X)$ and $R = \gamma_{w^2}(Y)$. Since Z is a finite maximal prefix code, X and L are finite maximal prefix codes. Symmetrically, since Z is a finite maximal suffix code, Y and R are finite maximal suffix codes.

Let G, D, G', D' be as in Proposition 5.2.3. We have by Equation (3.3)

$$\underline{U} - 1 = (\underline{X} - 1 + (\underline{G'} - 1)w^2(\underline{D'} - 1))(1 + w^2)$$

and by Equation (3.2)

$$\underline{L} - 1 = (1 + w^2)(\underline{X} - 1 + (\underline{G'} - 1)w^2(\underline{D'} - 1)).$$

Thus

$$(1 + w^2)(\underline{U} - 1) = (\underline{L} - 1)(1 + w^2).$$

Therefore $(1 + w^2)\underline{U}^*(1 + w) = \underline{L}^*(1 + w^2)(1 + w)$. This shows that $V = \{\varepsilon, w^2\}U^*\{\varepsilon, w\}$ is recurrent with a finite left root. Similarly

$$\underline{R} - 1 = (\underline{Y} - 1 + (\underline{G''} - 1)w^2(\underline{D''} - 1))(1 + w^2).$$

Since $(\underline{X} - 1)(1 + w) = (1 + w)(\underline{Y} - 1)$ and $(\underline{G'} - 1)w^2(\underline{D'} - 1)(1 + w) = (1 + w)(\underline{G''} - 1)w^2(\underline{D''} - 1)$, we have

$$(1 + w)(\underline{R} - 1) = (\underline{X} - 1 + (\underline{G'} - 1)w^2(\underline{D'} - 1))(1 + w)(1 + w^2)$$

showing that

$$(1 + w^2)(1 + w)(\underline{R} - 1) = (\underline{L} - 1)(1 + w)(1 + w^2).$$

This implies that V is birecurrent and that its right root is finite. ∎

Example 31 Let $\mathcal{A} = (Q, I, T)$ be the nondeterministic automaton with transitions given in Table 5.1 with $I = \{1, 14\}$ and $T = \{1, 4\}$. One may verify

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
a	2	6, 10	4	2	6		8		10	1, 14	12		14		16		1
b	3	1, 14	7, 11	5	7, 11	7		9		11	1, 14	13		15		17	
a^2	6, 10	1, 14	2	6, 10					1, 14	2							2

Table 5.1: The transitions of the automaton $\mathcal{A}$.

that this automaton is unambiguous. The word a^2 has rank 3 since the matrix $\mu_{\mathcal{A}}(a^2)$ has 3 distinct nonzero rows which are the characteristic vectors of $\{1, 14\}$, $\{6, 10\}$ and $\{2\}$.

This rank is minimal as one can check by computing the 9 images of this 3-element set by the action of the letters. Since $\underline{I}$ is a row of $\mu(a^2)$, the hypothesis of Theorem 5.2.1 are satisfied. The automaton $\mathcal{A}^\delta$ is strongly connected and has also 17 states represented in Table 5.2. The transitions of $\mathcal{A}^\delta$ are represented

1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
1, 14	2	3, 15	6, 10	4, 16	7, 11	5, 17	8, 12	1, 6	9, 13	3, 7	10, 14

13	**14**	**15**	**16**	**17**
4, 8	11, 15	5, 9	12, 16	13, 17

Table 5.2: The states of the automaton $\mathcal{A}^\delta$.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**
a	**2**	**4**	**5**	**1**	**2**	**8**	**9**	**1**	**2**	**12**	**13**	**1**	**2**	**1**	**4**	**1**	**1**
b	**3**	**1**	**6**	**6**	**7**	**1**	**6**	**10**	**11**	**1**	**6**	**14**	**15**	**6**	**6**	**17**	**1**

Table 5.3: The transitions of the automaton $\mathcal{A}^\delta$.

in Table 5.3. The set $T = \{1, 4\}$ is a column of $\mu_{\mathcal{A}}(a^2)$ (the columns of index 6 and 10). Thus, by the symmetric of Theorem 5.2.1, the automaton $\tilde{\mathcal{A}}^\delta$ is also strongly connected. Thus the set S recognized by $\mathcal{A}$ is birecurrent.

The set S is an instance of Theorem 5.2.4. To see this, we start with the finite maximal bifix code $\tilde{Z}$ represented in Figure 3.7 on the right. We choose the word $x = (ba)^2$ which is a pure square for $\tilde{Z}$. Then $\tilde{Y} = \delta_w(\tilde{Z})$ is the maximal prefix code represented in Figure 3.6. The word $w^2 = (ba)^4$ is a pure square for $\tilde{Y}$ and the set $X = \gamma_{w^2}$ is a maximal code generating the set recognized by the automaton $\mathcal{A}$ with 1 as initial and terminal state.

References

[1] Dana Angluin. Inference of reversible languages. *J. Assoc. Comput. Mach.*, 29(3):741–765, 1982.

[2] Jean Berstel, Dominique Perrin, and Christophe Reutenauer. *Codes and Automata*. Cambridge University Press, 2009.

[3] Jean Berstel and Christophe Reutenauer. *Noncommutative rational series with applications*, volume 137 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 2011.

[4] Véronique Bruyère and Clelia De Felice. Synchronization and decomposability for a family of codes. *Internat. J. Algebra Comput.*, 2(4):367–393, 1992.

[5] Samuel Eilenberg. *Automata, languages, and machines. Vol. A*. Academic Press [A subsidiary of Harcourt Brace Jovanovich, Publishers], New York, 1974. Pure and Applied Mathematics, Vol. 58.

[6] Dominique Perrin. Codes asynchrones. *Bull. Soc. Math. France*, 105(4):385–404, 1977.

[7] Dominique Perrin. Completely reducible sets. *Internat. J. Algebra Comput.*, 23(4):915–941, 2013.

[8] Jean-Eric Pin. On reversible automata. In *LATIN '92 (São Paulo, 1992)*, Lecture Notes in Comput. Sci.

[9] Christophe Reutenauer. Une topologie du monoïde libre. *Semigroup Forum*, 18(1):33–49, 1979.

[10] Christophe Reutenauer. Semisimplicity of the algebra associated to a biprefix code. *Semigroup Forum*, 23:327–342, 1981.

[11] John Rhodes and Benjamin Steinberg. *The q-theory of finite semigroups.* Springer Monographs in Mathematics. Springer, New York, 2009.

[12] Andrew Ryzhikov. On automata recognizing birecurrent sets. 2017. (`arxiv.org/abs/1711.0106`).

[13] Jacques Sakarovitch. *Elements of automata theory.* Cambridge University Press, Cambridge, 2009. Translated from the 2003 French original by Reuben Thomas.

[14] M. P. Schützenberger. On a special class of recurrent events. *Ann. Math. Statist.*, 32:1201–1213, 1961.

[15] Max Vincent. Construction de codes indécomposables. *RAIRO Inform. Théor.*, 19(2):165–178, 1985.