Recurrent and birecurrent sets

Dominique Perrin
(with Francesco Dolce, Antonio Restivo, Christophe Reutenauer and Giuseppina Rindone)

January 24, 2017
A new class of recognizable sets

A recognizable set $X$ is \textit{recurrent} if its minimal automaton is strongly connected. A recognizable set $X$ is \textit{birecurrent} if $X$ and its mirror image $\tilde{X}$ are recurrent.

Birecurrent sets

- Form an interesting class of sets and automata, never systematically studied before.
- Are a generalization of submonoids generated by bifix codes.
- Are linked with completely reducible formal series.
Summary of results

- A new proof of Schützenberger theorem on codes with bounded deciphering delay.
- A characterization allowing to build birecurrent sets easily.
- A class of birecurrent sets of finite type generalizing the submonoids generated by finite bifix codes.
- A characterization of completely reducible sets in terms of birecurrent sets.
Deciphering delay

A set $X \subset A^*$ is said to have finite deciphering delay if there exists an integer $d \geq 0$ such that the following holds. If $xu$ is a prefix of $x'y'$ with $x, x' \in X$, $u$ a prefix of a word in $X^*$ and $y' \in X^*$, then $|u| \geq d$ implies $x = x'$. The least integer $d$ such that this holds is called the deciphering delay of $X$. A prefix code has deciphering delay 0.
Example

The code $X = \{a, ab, abc\}$ has deciphering delay 1. Its minimal automaton is the strongly connected automaton represented below.

The submonoid $X^*$ is recurrent.
A theorem of Schützenberger

**Theorem (Schützenberger, 1966)**

A finite maximal code $X$ having finite deciphering delay is prefix.

Our new proof shows that $X^*$ is recurrent using the following lemma.

**Lemma**

Let $X$ be a finite code and let $A$ be the minimal automaton of $X^*$. If $X$ has finite deciphering delay, then all the cycles of $A$ are contained in the same strongly connected component.
Deterministic reversals

Let $A = (Q, I, T)$ be an automaton. The reversal of $A$ is the automaton $\tilde{A} = (Q, T, I)$ obtained by reversing the edges of $A$ and exchanging $I, T$. For an automaton $A = (Q, I, T)$, we denote by $A^\delta$ the result of the determinization of $A$.

The deterministic reversal of a deterministic automaton $A$ is the automaton $\tilde{A}^\delta$ obtained by determinization of the reversal of $A$.

The following statement is well known.

**Proposition**

*If $A$ is a trim deterministic automaton recognizing $X$, then $\tilde{A}^\delta$ is the minimal automaton of $\tilde{X}$.***
A deterministic finite automaton $\mathcal{A}$ is **biconnected** if $\mathcal{A}$ and $\tilde{\mathcal{A}}^\delta$ are strongly connected. Thus a set is birecurrent if and only if its minimal automaton is biconnected.

Examples:

- **group automata**
- **reversible** automata which are such that the maps defined by all words are injective
- the class of strongly connected automata $\mathcal{A} = (Q, i, i)$ satisfying the following condition. For any $q \in Q$ and any $w \in A^*$ such that $q \cdot w \neq \emptyset$,

  $$i \cdot w = q \cdot w \Rightarrow q = i.$$  

This is the class of automata recognizing the submonoid generated by a **bifix code**.
Transition monoids

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 $\mathcal{A}$. The rank of a word $w$ is the rank of the map $\varphi_{\mathcal{A}}(w)$. 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 the unique 0-minimal ideal of the monoid $M_{\mathcal{A}}$. 
Let $A = (Q, i, T)$ be an automaton. The domain of a word $w$, denoted $\text{Dom}(w)$, 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 $\text{Dom}(w)$ defined by $p \equiv q$ if $p \cdot w = q \cdot w$.

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

Let $A$ be the automaton represented below.

![Automaton diagram]

The word $b$ is synchronizing since it has rank 1. The 0-minimal ideal of the monoid $M_A$ is represented below.

![Matrix representation]
Characterization of biconnected automata

We say that a set $S \subseteq Q$ is **saturated** by a word $w$ if $S$ is a union of classes of the kernel of $w$.

**Theorem**

A strongly connected deterministic finite automaton $A = (Q, i, T)$ is biconnected if and only if $T$ is saturated by a word of minimal nonzero rank.
Note that if $\mathcal{A} = (Q, i, T)$ is complete and synchronized, then $\mathcal{A}$ is biconnected if and only if $T = Q$. In this case $\tilde{\mathcal{A}}^\delta$ has just one state. Indeed, all words of minimal rank are constant functions and thus have a kernel with one class equal to $Q$.

We do not know if the equivalent conditions of the theorem can be verified in polynomial time.
Example 1

The automaton represented below on the left and its deterministic reversal on the right are both strongly connected and thus they are biconnected.

In agreement with the theorem, the set $T = \{1\}$ of terminal states of the first automaton is the preimage of $b$, which has rank 1.

The 0-minimal ideals of $A$ and $\tilde{A}^\delta$:
Example 2

The 0-minimal ideal of the monoid $M_A$ is represented below.

The diagram shows the structure of the monoid $M_A$ with transitions labeled by elements $a$ and $b$. The starred elements indicate transitions that are not directly visible in the diagram but are inferred from the structure.

A table is also provided, listing the transitions for elements $1, 2, 3$ and sub-elements $1, 2, 3$. The table includes elements such as $b$, $ba$, $ba^2$, $ab$, $aba$, and $a^2b$, reflecting the structure and transitions within the monoid.
Example 3

Consider the complete deterministic automaton $A$ given below on the left with its reversal represented on the right.
The 0-minimal ideal of the monoid $M_\mathcal{A}$ is represented below. 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.

![Diagram]

Dominique Perrin (with Francesco Dolce, Antonio Restivo, Christophe Reutenauer and Giuseppina Rindone)
Example 4

Let $\mathcal{A}$ be the automaton represented below on the left with its reversal on the right.

![Automaton Diagram]

The conditions of the Theorem are satisfied because $T$ is a class of the kernel of $b$, which is of minimal rank 2. However, the reversal of $\tilde{\mathcal{A}}$ is not equal to $\mathcal{A}$ because $\mathcal{A}$ is not minimal.
The 0-minimal ideal of the monoid $M = \varphi_A(A^*)$ is represented below.

\[
\begin{array}{|c|c|c|c|}
\hline
1/4 & 1/2 & 2/3 & 3/4 \\
\hline
* b & * ba & * ba^2 & * ba^3 \\
\hline
\end{array}
\]

1, 3/2, 4

It is formed of a single $\mathcal{R}$-class and the corresponding kernel is the equivalence giving the minimal automaton. The automaton $\tilde{\mathcal{A}}^\delta$ is itself its own reversal.
Birecurrent sets of finite type

Let $S$ be a recurrent set and let $A = (Q, i, T)$ be its minimal automaton. The prefix code $X$ such that $X^*$ is recognized by $(Q, i, i)$ is called the left root of $S$. When $S$ is birecurrent, it has a left root and a right root $Y$ where $\tilde{Y}$ is the left root of $\tilde{S}$. We say that a birecurrent set $S$ is of finite type if its left and right roots are finite. 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. For example, for every finite bifix code $X$, the set $S = X^*$ is a birecurrent set of finite type.
The birecurrent set $S$ of Example 3 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 actually $S = \tilde{S}$. 
Example 5

Let $X$ be the maximal prefix code represented below and let $S$ be the recurrent set obtained with $i = 1$ and $T = \{1, 6\}$. 

![Graph showing the maximal prefix code and the recurrent set.](image)
The minimal rank is equal to 3 as one may check by computing 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 $A$ is thus biconnected and the set $S$ recognized by $A$ is birecurrent.
The deterministic reversal of $\mathcal{A}$ has the set of states represented below.

\[
\begin{array}{cccccccccc}
1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\
1, 6 & 3, 4, 9 & 2, 5, 7, 9 & 2, 5, 7, 8 & 1, 7 & 3, 4, 8 & 3, 5, 7, 9 & 2, 5, 6, 8 & 2, 4, 8 \\
\end{array}
\]

The transitions of $\tilde{\mathcal{A}}^\delta$ are given below.

\[
\begin{array}{cccccccccc}
1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\
\hline
a & 2 & 4 & 5 & 1 & 2 & 8 & 4 & 1 & 1 \\
b & 3 & 1 & 6 & 6 & 7 & 1 & 6 & 9 & 1 \\
\end{array}
\]
The reversal $\tilde{Y}$ of the right root of $S$ is represented below.

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

Examples 3 and 4 are particular cases of a general construction that we now describe. It is a modification of the internal transformation of bifix codes. Let $Z$ be a maximal bifix code and let $w$ be a proper prefix and suffix of $Z$. Set $G = Zw^{-1}$ and $D = w^{-1}Z$. Assume that $Gw \cap wD = \emptyset$. Then

$$Z + (G - 1)w(D - 1)$$

is the characteristic polynomial of a maximal bifix code obtained by internal transformation from $Z$.

**Theorem (Cesari, 1974)**

Any finite maximal bifix code is obtained by internal transformations starting from $A^n$ for some $n \geq 1$. 
Example

The two maximal bifix codes obtained starting from $A^3$ are represented below.
Pure squares

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 $Z \cap wA^* \cap A^*w = \{x\}$.

**Proposition**

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

$$ (1 + w)(Z - 1 + (G - 1)w(D - 1)) + 1 $$

(2)

is the characteristic polynomial of a finite maximal prefix code.

The prefix code defined by Equation (2) is denoted $\delta_w(Z)$. 
Theorem

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.

The right root of $S$ is the suffix code with characteristic polynomial

$$
(Z - 1 + (G - 1)w(D - 1))(1 + w) + 1
$$

(3)
Let $S$ be the birecurrent set of Example 3. We have seen in Example 3 bis that $S = X^*P$ with $X = (aA \cup b)^2$ and $P = \{\varepsilon, a\}$. We obtain $X$ as in the Theorem using $Z = A^2$ and $x = a^2$. Indeed, we have $G = D = A$ and thus the set of proper prefixes of $\delta_a(Z)$ is

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

which is the characteristic polynomial of the set of proper prefixes of $X$. 
Example 4 bis

Let $S$ be the birecurrent set of Example 4. We start from the finite maximal bifix code

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

which is represented below on the left with its reversal on the right.

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

Dominique Perrin (with Francesco Dolce, Antonio Restivo, Christophe Reutenauer and Giuseppina Rindone)
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 below 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 below (with the nodes of $Q$ in black and those of $wQ$ in red). It is the set of proper prefixes of the maximal prefix code $X = \delta_w(Z)$.
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

$$(1 - X)Q = P(1 - Y).$$

(4)

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

$$1 - X = J(1 - A), \quad 1 - Y = (1 - A)K.$$  

(5)

Combining Equations (4) and (5), we obtain

$$J(1 - A)Q = P(1 - A)K.$$  

(6)
Conjecture 1

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

$$J = PM, \quad K = NQ, \quad M(1 - A) = (1 - A)N.$$  \hspace{1cm} (7)

This is true when $S$ is generated by a maximal bifix code and when it is obtained by the construction described before.
The continuant polynomials \( p_n = p(a_1, \ldots, a_n) \) in the noncommutative variables \( a_1, a_2, \ldots \) are as follows. Set \( p_0 = 1 \) and \( p_1 = a_1 \). Then for \( n \geq 0 \), we have the recurrence relation

\[
p_{n+2} = p_{n+1}a_{n+2} + p_n.
\]

For \( n \geq 1 \), one has the Wedderburn relation

\[
p(a_1, \ldots, a_n)p(a_{n-1}, \ldots, a_1) = p(a_1, \ldots, a_{n-1})p(a_n, \ldots, a_1).
\]
Let $B$ be a polynomial of degree 1 (like $1 - A$). For $p = \alpha + Bu$, denote $\bar{p} = \alpha + uB$.

**Theorem**

Let $J, K, P, Q$ be polynomials in $K\langle A \rangle$. One has $JBQ = PBK$ if and only if, for some $n \geq 1$, some polynomials $x_i = \alpha_i + Bu_i$ with $\alpha_i \in K$ and $u_i \in K\langle A \rangle$, and some polynomials $U, V$, one has

$$J = Up(x_1, \ldots, x_n), \quad Q = p(\bar{x}_{n-1}, \ldots, \bar{x}_1)V,$$

and

$$P = Up(x_1, \ldots, x_{n-1}), \quad K = p(\bar{x}_n, \ldots, \bar{x}_1)V,$$
For $n = 1$, we have $J = P x_1$ and $K = \bar{x}_1 Q$. This corresponds to Equation (7) with $x_1 = M$ and $\bar{x}_1 = N$. Thus conjecture 1 holds in this case.
Indecomposable 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 $Z$.

**Theorem**

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.

This theorem 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 4. In fact $X$ is indecomposable, is not bifix and not synchronized.
Conjecture 2

We formulate the daring conjecture that if the prefix code in the theorem is additionally finite, then either it is synchronized or it is the left root of a dense birecurrent set of finite type.
Formal series

Let $K = \mathbb{Q}$ be the field of rational numbers. A formal series on the alphabet $A$ with coefficients in $K$ is a map $\sigma : A^* \rightarrow K$. We denote by $K\langle\langle A\rangle\rangle$ the algebra of these series. For $w \in A^*$, we denote by $(\sigma, w)$ the value of $\sigma$ on $w$.

We consider the characteristic series $S$ of a set $S \subset A^*$ as a series with coefficients in $K$. 
Let $n \geq 0$ be an integer. Let $\lambda$ be a row $n$-vector, let $\mu$ be a morphism from $A^*$ into the monoid of $n \times n$-matrices and let $\gamma$ be a column $n$-vector, all with coefficients in $K$. The triple $(\lambda, \mu, \gamma)$ is called a linear representation. It is said to recognize a series $\sigma$ if for every $w \in A^*$, one has
\[
(\sigma, w) = \lambda \mu(w) \gamma.
\] (10)
A series is **recognizable** 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).
Example 5

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 $\mu(a)\gamma = \gamma$. The minimal representation of $a^*$ is of dimension 1.
Example 6

Let $X$ be 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 below.

|   | 1  | 2  | 3  | 4  | 5  |
|---|----|----|----|----|----|
| $a$| 2  | 3  | 1  | 3  | 1  |
| $b$| 4  | 1  | 5  | 5  | 1  |

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

|   | 1  | 2  | 3  | 4  | 5  |
|---|----|----|----|----|----|
| $\tilde{Q}$| 1  | 2,4 | 3,5 | 2,5 | 3,4 |
The vector space generated by the corresponding characteristic vectors has dimension 4 because $2 + 3 = 4 + 5$. Choosing the basis formed of $1, 2, 3, 4$ the minimal representation $(\lambda, \mu, \gamma)$ of $\sigma = S$ is

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

Let us verify for example the value of the first column of $\mu(a)$. We have $\mu(a)1 = a \cdot \{1\} = \{3, 5\} = \overline{3}$. Similarly, the last column of $\mu(b)$ is computed as $\mu(b)4 = b \cdot \{2, 5\} = \{3, 4\} = \overline{5} = \overline{2} + \overline{3} - \overline{4}$. 
Let $\mu : A^* \to \text{End}(V)$ be a morphism. A subspace $W$ of $V$ is said to be invariant with respect to $\mu$ 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 $\mu$ 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 $\mu$ if $V = \bigoplus_{i=1}^{n} V_i$ where each $V_i$ is an invariant irreducible subspace of $V$. 
Example

The linear representation of Example 5 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.

\[
\lambda' = [-1, 1], \quad \mu'(a) = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \quad \gamma' = [1].
\]
A set $S \subset A^*$ is completely reducible over $K$ if the minimal representation of the series $S$ is completely reducible.
Example

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.
Consider again the set $S$ of Example 6. We have found a minimal representation of dimension 4. This representation is not irreducible because the space generated by $1 + 2 + 3$ is invariant by $\mu$ (with $\mu(w)$ acting on the left). The space generated by $2 - 1, 3 - 1$ and $4 - 1$ is a stable complement and the representation takes in the basis $1 + 2 + 3, 2 - 1, 3 - 1, 4 - 1$ the form of a direct sum of two representations $\mu'_1$ of dimension 1 and $\mu'_2$ of dimension 3. In this basis, the vector $\lambda$ becomes $\lambda' = [1 - 1 - 1 - 1]$ (its components are the values of the linear form defined by $\lambda$ on each vector of the basis), and $\mu(a), \mu(b), \gamma$ become

$$
\mu'(a) = \begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & -1 & -1 & -1 \\
0 & 0 & 0 & 0
\end{bmatrix}, \quad
\mu'(b) = \begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 \\
0 & 0 & 1 & 1 \\
0 & -1 & -2 & -2
\end{bmatrix}, \quad
\gamma' = \begin{bmatrix}
1/3 \\
-1/3 \\
-1/3 \\
0
\end{bmatrix}
$$
Let $S \subset A^*$ be a recognizable set, let $\sigma = S$ and let $A = (Q, i, T)$ be its minimal automaton. Set $\varphi = \varphi_A$, $\psi = \psi_\sigma$ and $M = \psi(A^*)$. For all $u, v \in A^*$, one has

$$\varphi(u) = \varphi(v) \iff \psi(u) = \psi(v).$$

(11)

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

An element of $M = \psi(A^*)$ is a linear map from $V_S$ into $V_S$ and, as such, it has a kernel which is a subspace of $V_S$. The eventual kernel of $S$ is the intersection of the kernels of all elements of the 0-minimal ideal of $M$.

**Theorem**

A recurrent set $S$ is completely reducible if and only if its eventual kernel is 0.
We obtain as a corollary, which is a generalization of the result of Reutenauer (1981) asserting that the submonoid generated by a bifix code is completely reducible.

**Corollary**

*Any recognizable birecurrent set is completely reducible.*
A characterization of completely reducible sets

**Theorem**

A language is completely reducible if and only if its characteristic series is \(\mathbb{Q}\)-linear combination of characteristic series of birecurrent languages.