On the supports of recognizable series over a field and a single letter alphabet

Guillaume Chapuy, Ines Klimann

On the supports of recognizable series over a field and a single letter alphabet

Guillaume Chapuy[a], Ines Klimann[a]

[a] *LIAFA - CNRS UMR 7089 & Université Paris Diderot-Paris 7*
Case 7014 - 75205 Paris Cedex 13

Abstract

We prove that the support of a recognizable series over a field of characteristic zero and a single letter alphabet is recognizable. This provides an answer to a question of Kirsten [4]. Then we give an example of a recognizable series over a field of prime characteristic and a single letter alphabet whose support is not recognizable which provides an answer to a question of Kirsten and Quaas [5].

1. Introduction

A *series* over a semiring $\mathbb{S}$ and an alphabet Σ is a function

$$S : \Sigma^* \to \mathbb{S}$$

which maps a word on Σ to an element of $\mathbb{S}$. The series is *recognizable* if it is the behaviour of a weighted finite automaton. Its *support* is the set of all words which are not mapped to zero.

Recently Kirsten introduced the notion of SR-semirings and characterized them algebraically [4]. An *SR-semiring* is a semiring $\mathbb{S}$ such that the support of any recognizable series over $\mathbb{S}$ and any finite alphabet Σ is a recognizable language. Many semirings are SR-semirings; typical examples of non-SR-semirings are $\mathbb{Z}$ and $\mathbb{Q}$.

Besides an algebraic characterization, Kirsten showed that in the definition of an SR-semiring, one could equivalently restrict the alphabet Σ to any fixed finite alphabet of at least two letters. More precisely, given a finite alphabet Σ of at least two letters, a semiring $\mathbb{S}$ is an SR-semiring if and only if for every recognizable series $S : \Sigma^* \to \mathbb{S}$ the support of S is a recognizable language. In [4, p.500] Kirsten asked whether one can even restrict Σ to a single letter alphabet: Given a single letter alphabet Σ and a semiring $\mathbb{S}$, if the support of every recognizable series $\Sigma^* \to \mathbb{S}$ is a recognizable language, is $\mathbb{S}$ necessarily an SR-semiring?

In this note we negatively answer Kirsten's question, by showing that the support of a recognizable series over $\mathbb{Q}$ and a single letter alphabet is always recognizable. In fact we prove this property for series over any field of characteristic zero. We also show that this property cannot be extended to fields of

characteristic $p \neq 0$: We construct a recognizable series over a field of prime characteristic and a single letter alphabet whose support is not recognizable. This provides an example of a non-SR-semiring which does not include $\mathbb{Z}$, answering a question of Kirsten and Quaas [5].

In Section 2 we present basic definitions on series and weighted finite automata. In Section 3 we prove that the support of a recognizable series over a field of characteristic zero and a single letter alphabet is recognizable. Finally in Section 4 we give an example of a recognizable series over a field of prime characteristic and a single letter alphabet whose support is not recognizable.

2. Preliminaries

Basic notions on (recognizable) languages and finite automata are supposed to be known. The reader can refer to [8] for details. Remind in particular that recognizable languages over a single letter alphabet $\{a\}$ are exactly the sets a^N where $N \subseteq \mathbb{N}$ is ultimately periodic - possibly finite (this is immediate considering deterministic automata).

A *monoid* is a set with a binary associative operation and an identity element for this operation. A monoid whose operation is commutative is called a *commutative monoid*. A *semiring* $(\mathbb{S}, +, \cdot, 0, 1)$ consists of a set $\mathbb{S}$ equipped with two binary operations: product $\cdot$ (or simply denoted by juxtaposition) and sum $+$ such that $(\mathbb{S}, +, 0)$ is a commutative monoid, $(\mathbb{S}, \cdot, 1)$ is a monoid and product distributes over sum. If no confusion arises we will denote it by $\mathbb{S}$.

Matrices over a semiring $\mathbb{S}$ can be equipped with usual matrix multiplication.

A *series* S over a semiring $\mathbb{S}$ and an alphabet Σ is a map from Σ^* into $\mathbb{S}$. The image of a word w is its *coefficient* and is denoted by (S, w).

A *weighted finite automaton* (WFA) $\mathcal{A}$ over a semiring $\mathbb{S}$ and a single letter alphabet $\{a\}$ is a tuple (Q, M, I, F) where

- Q is a finite *set of states*,
- $M \in \mathbb{S}^{Q \times Q}$ is a $Q \times Q$*-matrix of transitions*,
- $I \in \mathbb{S}^Q$ is the *initial vector*,
- $F \in \mathbb{S}^Q$ is the *final vector*.

We say that a state is *initial* (*resp. final*) if its coefficient in I (*resp.* F) is not zero, this coefficient is called the *initial* (*resp. final*) *weight* of the state. The automaton is *normalized* if it has a single initial state i with weight 1, a single final state f with weight 1 and the i-th column and the f-th line of M have only 0-inputs (speaking in graphical terms: no transitions are leading to the initial state and no transitions are leaving the final state).

The *series recognized* by $\mathcal{A}$, also called its *behaviour*, is the series

$$|\mathcal{A}| : \{a\}^* \to \mathbb{S} : a^n \mapsto {}^t I M^n F.$$

A series is *recognizable* if it is the behaviour of an automaton. The *support* $\mathrm{Supp}(S)$ of a series S is the language of all words which are not mapped to zero.

Note that the notions of weighted finite automata and series recognized by such automata can be defined in the same way for any finite alphabet, see [2, 3, 8] for more details.

3. Main result in characteristic zero

This section is devoted to the proof of our main result which is given by the following proposition:

Proposition 1. *The support of a recognizable series over a field of characteristic zero and a single letter alphabet is recognizable.*

Let $\mathcal{A} = (Q, M, I, F)$ be a WFA over a field of characteristic zero $\mathbb{S}$ and a single letter alphabet $\{a\}$. It is well-known that there exists a normalized WFA $\mathcal{A}'$ such that $(|\mathcal{A}|, w) = (|\mathcal{A}'|, w)$ for every non-empty word w [3, Th. 2.11]. Therefore we can suppose that $\mathcal{A}$ is normalized without loss of generality.

Let us denote respectively by i and f the initial state and the final state of $\mathcal{A}$: the coefficient of the word a^n in the behaviour of $\mathcal{A}$ is the (i, f)-input of the matrix M^n:

$$(|\mathcal{A}|, a^n) = (M^n)_{i,f}.$$

We denote by $(u_n)_{n\in\mathbb{N}}$ the sequence $((M^n)_{i,f})_{n\in\mathbb{N}}$.

Lemma 1. *The sequence $(u_n)_{n\in\mathbb{N}}$ is the solution of a linear recursion with constant coefficients.*

Proof. The classical Cayley-Hamilton Theorem states that every square matrix over a commutative ring (and hence over a field) satisfies its own characteristic equation [1]. A direct consequence of this theorem is that any square matrix is the zero of a non-trivial polynomial equation with leading coefficient 1. As a result, we have

$$M^{|Q|} = \alpha_{|Q|-1}M^{|Q|-1} + \cdots + \alpha_1 M + \alpha_0 I_{|Q|}, \qquad \text{for some } (\alpha_0, \ldots, \alpha_{|Q|-1}) \in \mathbb{S}^{|Q|},$$

where $I_{|Q|}$ is the identity $|Q| \times |Q|$-matrix and hence for all $k \geq |Q|$:

$$M^k = \alpha_{|Q|-1}M^{k-1} + \cdots + \alpha_1 M^{k-|Q|+1} + \alpha_0 M^{k-|Q|}.$$

Taking the (i, f)-inputs of all these matrices we thus obtain

$$u_k = \alpha_{|Q|-1}u_{k-1} + \cdots + \alpha_1 u_{k-|Q|+1} + \alpha_0 u_{k-|Q|}.$$

□

We can now apply a result of Lech [6, 7]: in a field of characteristic zero, if the values of the terms of a sequence solution of a linear recursion equation with constant coefficients are infinitely often zero, then they are equal to zero ultimately periodically. Since recognizable languages over a single letter alphabet are the ultimately periodical languages, Proposition 1 is a direct consequence of Lemma 1.

4. Counter-example in characteristic p

Lech proved by giving an example that his result on sequences cannot be extended to fields of prime characteristic [6]. In this section, we adapt this example to prove that neither can be Proposition 1.

Remember that a field has zero or prime characteristic. Denote by $\mathbb{F}_p$ the cyclic field of prime size p and consider $\mathbb{F}_p(X)$ the field of fractions of the polynomial ring in one indeterminate $\mathbb{F}_p[X]$, *i.e.* the smallest field in which $\mathbb{F}_p[X]$ can be embedded: it has characteristic p.

The series $S : \{a\}^* \to \mathbb{F}_p(X)$ given by

$$a^n \mapsto (S, a^n) = (1+X)^n - 1 - X^n$$

is recognized by the automaton of Figure 1. We denote by L the complement of its support.

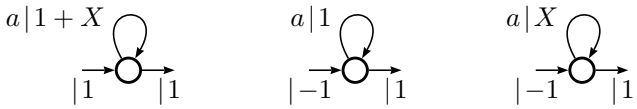

Figure 1: An automaton over $\mathbb{F}_p(X)$ and $\{a\}$ whose behaviour has non-recognizable support.

We have for each n:

$$(S, a^n) = (1+X)^n - 1 - X^n = \sum_{k=1}^{n-1} \binom{n}{k} X^k. \tag{1}$$

Now, the greatest power of p dividing $\binom{n}{k}$ is the number of carries by substracting k to n in base p (see [9]). Therefore we have:

- If n is a power of p, say p^α, it is written 10^α in base p and any k ($1 \leq k \leq n-1$) has at least one 1-digit under one 0-digit of n and so $\binom{n}{k}$ is dividable by p. Hence the right member of Equation (1) is 0 since the field has characteristic p.
- If n is not a power of p, it is written $1d_1 \dots d_\beta 1 e_1 \dots e_\gamma$ in base p and choosing for k the number 10^γ in base p, $\binom{n}{k}$ is not dividable by p. Therefore the coefficient of X^k in Equation (1) is not 0, so the right-hand side of (1) is not the null polynomial.

Hence L is the non-recognizable language of powers of p and the support of S is not recognizable.

Note that this example shows that $\mathbb{F}_p(X)$ is a non-SR-semiring. It is an example of a non-SR-semiring which does not include $\mathbb{Z}$, answering the question of the existence of such semirings [5].

Acknowledgements. We thank an anonymous referee whose fruitful comments improved the presentation of our results.

[1] Michael F. Atiyah and Ian G. MacDonald. *Introduction to Commutative Algebra*. Addison-Wesley, Reading, Massachusetts, 1969.

[2] Jean Berstel and Christophe Reutenauer. *Noncommutative Rational Series with Applications*. Encyclopedia of Mathematics and its Applications. Cambridge University Press, 2010.

[3] Manfred Droste, Werner Kuich, and Heiko Vogler. *Handbook of Weighted Automata*. Springer Publishing Company, Incorporated, 1st edition, 2009.

[4] Daniel Kirsten. An algebraic characterization of semirings for which the support of every recognizable series is recognizable. In *MFCS*, pages 489–500, 2009.

[5] Daniel Kirsten and Karin Quaas. Recognizability of the support of recognizable series over the semiring of the integers is undecidable. *Inf. Process. Lett.*, 111:500–502, April 2011.

[6] Christer Lech. A note on recurring series. *Arkiv för Matematik*, 2:417–421, 1953.

[7] Richard A. Mollin. *Number theory and applications*. Kluwer Academic publ., 1989.

[8] Jacques Sakarovitch. *Elements of Automata Theory*. Cambridge University Press, New York, NY, USA, 2009.

[9] André Weil. *Collected papers – Ernst Eduard Kummer*, volume 1 (Contributions to number theory). Springer-Verlag, 1975.