On the supports of recognizable series over a field and a single letter alphabet
Guillaume Chapuy, Ines Klimann

To cite this version:

HAL Id: hal-00663147
https://hal-univ-diderot.archives-ouvertes.fr/hal-00663147
Submitted on 26 Jan 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
On the supports of recognizable series over a field and a single letter alphabet

Guillaume Chapuy\textsuperscript{a}, Ines Klimann\textsuperscript{a}

\textsuperscript{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 \cite{Kirsten2012}. 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 \cite{KirstenQuaas2012}.

1. Introduction

A series over a semiring $\mathcal{S}$ and an alphabet $\Sigma$ is a function $S : \Sigma^* \rightarrow \mathcal{S}$ which maps a word on $\Sigma$ to an element of $\mathcal{S}$. 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 \cite{Kirsten2012}. An SR-semiring is a semiring $\mathcal{S}$ such that the support of any recognizable series over $\mathcal{S}$ and any finite alphabet $\Sigma$ 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 $\Sigma$ to any fixed finite alphabet of at least two letters. More precisely, given a finite alphabet $\Sigma$ of at least two letters, a semiring $\mathcal{S}$ is an SR-semiring if and only if for every recognizable series $S : \Sigma^* \rightarrow \mathcal{S}$ the support of $S$ is a recognizable language. In \cite[p.500]{Kirsten2012} Kirsten asked whether one can even restrict $\Sigma$ to a single letter alphabet: Given a single letter alphabet $\Sigma$ and a semiring $\mathcal{S}$, if the support of every recognizable series $\Sigma^* \rightarrow \mathcal{S}$ is a recognizable language, is $\mathcal{S}$ 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 $(S, +, 0, 1)$ consists of a set $S$ equipped with two binary operations: product $\cdot$ (or simply denoted by juxtaposition) and sum $+$ such that $(S, +, 0)$ is a commutative monoid, $(S, \cdot, 1)$ is a monoid and product distributes over sum. If no confusion arises we will denote it by $S$.

Matrices over a semiring $S$ can be equipped with usual matrix multiplication.

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

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

- $Q$ is a finite set of states,
- $M \in S^{Q \times Q}$ is a $Q \times Q$-matrix of transitions,
- $I \in S^Q$ is the initial vector,
- $F \in S^Q$ 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 $A$, also called its behaviour, is the series

$$|A| : \{a\}^* \rightarrow S : a^n \mapsto IM^nF.$$
A series is recognizable if it is the behaviour of an automaton. The support \( \text{Supp}(S) \) 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 \( S \) and a single letter alphabet \( \{a\} \). It is well-known that there exists a normalized WFA \( \mathcal{A}' \) such that \( (|A|, w) = (|A'|, w) \) 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 \( M^n \):

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

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

\[
M^{|Q|} = a_{|Q|-1}M^{|Q|-1} + \cdots + a_1M + a_0I_{|Q|}, \quad \text{for some } (a_0, \ldots, a_{|Q|-1}) \in S^{|Q|},
\]

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

\[
M^k = a_{|Q|-1}M^{k-1} + \cdots + a_1M^{k-|Q|+1} + a_0M^{k-|Q|}.
\]

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

\[
u_k = a_{|Q|-1}u_{k-1} + \cdots + a_1u_{k-|Q|+1} + a_0u_{k-|Q|}.
\]

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

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

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.](image)

We have for each $n$:

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

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

- If $n$ is a power of $p$, say $p^\alpha$, it is written $10^\alpha$ 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 divisible 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 $1d_1 \ldots d_\beta e_1 \ldots e_\gamma$ in base $p$ and choosing for $k$ the number $10^\beta$ in base $p$, $\binom{n}{k}$ is not divisible 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.


