# Five lectures in the theory of
# **Weighted Automata and Transducers**

*Jacques Sakarovitch*

IRIF/CNRS–Université Paris Diderot & Telecom-ParisTech

December 2018 – January 2019

Lectures notes of the Master Parisien de Recherche en Informatique

Course 2.16 — Finite automata based computation models

# Contents



These lectures notes are intended to be as self-contained as possible. However, many complements — sometimes in a slightly different setting, as my point of view has evolved — are to be found in my book *Elements of Automata Theory* (Cambridge University Press, 2009). References to this work are indicated in marginal notes.

Every lecture ends with an exercise section.

# Lecture I

# The model of weighted automata

## Rationality and recognisability

This chapter is aimed at

(i) introducing, or recalling, the notions of *weighted automata* and of *representations*, that are the subject of the lectures to come,

(ii) giving the proof of their equivalence which is considered here as a basic property,

(iii) and *fixing the terminology and notation.*

### Contents

# 1 The model of $\mathbb{K}$-automata

For sake of simplicity, we first restrict ourselves to automata over a free monoid $A^*$; the generalisation to automata over other monoids, at least over *graded* ones (*cf.*Section 2.4), is straightforward.

Automata with multiplicity or weighted automata are perfectly synonymous. The latter is preferred, at least in English, for its conciseness. In French, 'automate à poids' is, as are neckties of the same kind, rather inelegant. Let us mention that *weight* is often attached to 'numerical' multiplicity in the literature but we do not restrict ourselves to this case here.

## 1.1 Weight semirings

**Semirings.** A *semiring* $\mathbb{K}$ is a structure with both an *addition* and a *multiplication*, with the usual distributivity laws. More precisely:

- **SA1.** $\mathbb{K}$ is a *commutative* monoid for addition, written $+$, whose neutral element, called the *zero* of $\mathbb{K}$, is written $0_{\mathbb{K}}$ (or 0).
- **SA2.** $\mathbb{K}$ is a monoid (not necessarily commutative) for multiplication, written by a dot, or more often by simple juxtaposition, whose neutral element, called the *identity* of $\mathbb{K}$, is written $1_{\mathbb{K}}$ (or 1).
- **SA3.** The multiplication distributes left and right over the addition; that is,

$$\forall i,j,k \in \mathbb{K} \qquad i\cdot(j+k) = (i\cdot j)+(i\cdot k) \quad \text{and} \quad (i+j)\cdot k = (i\cdot k)+(j\cdot k) \ .$$

- **SA4.** The neutral element for addition is a zero for multiplication (which justifies the terminology):

$$\forall k \in \mathbb{K} \qquad k\cdot 0_{\mathbb{K}} = 0_{\mathbb{K}}\cdot k = 0_{\mathbb{K}} \ .$$

  If $1_{\mathbb{K}} = 0_{\mathbb{K}}$ , then $\mathbb{K}$ is reduced to this single element. In the sequel, we assume that $1_{\mathbb{K}} \neq 0_{\mathbb{K}}$ .

A semiring is *commutative* when its multiplication is a commutative operation.

The semiring structure is the most rudimentary one such that *matrices* with entries in that structure can be multiplied with the usual laws. On the other hand, if $\mathbb{K}$ is a semiring, then $\mathbb{K}^{Q\times Q}$, the set of *square matrices* of dimension $Q$ with entries in $\mathbb{K}$ and equipped with the usual addition and multiplication, is a *semiring*.

*Remark* 1. We use *sets* rather than integers as a *dimension* for vectors and matrices. The easiness in writing it brings — which puts the emphasis on the fact that listing values in a vector or a matrix is rather about *indexing* these values than comparing their rank — proves to be very convenient.

**The semirings we use.** We shall be concerned mostly with the following four classes of weight semirings:

- First, the *Boolean semiring* $\mathbb{B}$, which indeed means 'no weight'.
- Second, the classical *semirings of numbers*:
  $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}_+$, $\mathbb{Q}$, $\mathbb{R}_+$, $\mathbb{R}$,
  that is, the non-negative integers, the integers, the non-negative rationals, the rationals, the non-negative reals, and the reals.
- Third, the so-called *tropical semirings*:
  $\mathbb{N}\mathsf{min} = \langle\, \mathbb{N} \cup \{+\infty\}, \mathsf{min}, + \,\rangle$ , $\mathbb{N}\mathsf{max} = \langle\, \mathbb{N} \cup \{-\infty\}, \mathsf{max}, + \,\rangle$ ,
  $\mathbb{Z}\mathsf{max} = \langle\, \mathbb{Z} \cup \{-\infty\}, \mathsf{max}, + \,\rangle$ , $\mathbb{Q}\mathsf{max} = \langle\, \mathbb{Q}_+ \cup \{-\infty\}, \mathsf{max}, + \,\rangle$ , *etc.*
  For all these semirings, the *identity* $1_{\mathbb{K}}$ is the number 0; the *zero* $0_{\mathbb{K}}$ is either $+\infty$ when the 'addition' is $\mathsf{min}$ or $-\infty$ when the 'addition' is $\mathsf{max}$.
- and finally the *semirings of subsets* and of *series*:
  - $\langle\, \mathfrak{P}(A^*), \cup, \cdot \,\rangle$ , the semiring of subsets of the free monoid,
  - its subsemiring of rational languages $\operatorname{Rat} A^*$ ,
  - $\mathbb{K}\langle\langle A^* \rangle\rangle$ , the semiring of series[1] over $A^*$ with multiplicity in $\mathbb{K}$, *etc.*
- And, of course, the semirings of (square) matrices with entries in all the above semirings.

In the sequel, $\mathbb{K}$ denotes a semiring.

**Morphisms.** If $\mathbb{K}$ and $\mathbb{L}$ are semirings, a map $\varphi\colon \mathbb{K} \to \mathbb{L}$ is a *morphism of semirings* if

$$\forall k, l \in \mathbb{K} \qquad \begin{cases} \varphi(k+l) = \varphi(k) + \varphi(l) & \text{and} \quad \varphi(0_{\mathbb{K}}) = 0_{\mathbb{L}} \ , \\ \varphi(kl) = \varphi(k)\,\varphi(l) & \text{and} \quad \varphi(1_{\mathbb{K}}) = 1_{\mathbb{L}} \ . \end{cases}$$

That is, $\varphi$ is a morphism of monoids for both the additive and multiplicative structures of $\mathbb{K}$ and $\mathbb{L}$.

A semiring $\mathbb{K}$ is *positive* if both the sum and the product of any two non-zero elements of $\mathbb{K}$ are non-zero; in other words, if the *support map* $\sigma\colon \mathbb{K} \to \mathbb{B}$ such that $\sigma(k) = 1_{\mathbb{B}}$ for all $k \neq 0_{\mathbb{K}}$ (and $\sigma(0_{\mathbb{K}}) = 0_{\mathbb{B}}$) is a morphism of semirings. The semirings $\mathbb{N}$, $\mathbb{Z}\mathsf{min}$ or $\mathbb{Z}\mathsf{max}$ (!), $\mathbb{Q}_+$ and $\operatorname{Rat} A^*$ are positive, while $\mathbb{Z}$, $\mathbb{Q}$ and $\mathbb{R}$ are not.

**Exercises** See Exer. 1. to 2., p. 29.

[1] that will be defined below.

## 1.2 The graph definition of $\mathbb{K}$-automata

A classical, or *Boolean*, automaton $\mathcal{A}$ is a labelled directed graph, denoted[2] as a 5-tuple $\mathcal{A} = \langle A, Q, I, E, T \rangle$, where $A$ is the (input) alphabet, $Q$ the set of *states*, $I$ and $T$ the sets of *initial* and *final* states, and $E \subseteq Q \times A \times Q$ is the set of *transitions* of $\mathcal{A}$.

An *automaton over $A^*$ with weight in $\mathbb{K}$*, or *$\mathbb{K}$-automaton over $A^*$* is a generalisation of the former: it is a *labelled directed graph*. We develop and complete this definition below. In the next section, we build on the identification of a graph with its *incidence matrix* and the proofs will be performed systematically with matrix computations. The essence of an automaton however remains that of a graph and the behaviour of an automaton is defined in the language of graphs. We also continue to use the graph representation and its vocabulary to aid intuition.

*cf.EAT, p. 402*

We take here a definition of automata that is restricted compared to the one taken in EAT. It fits our needs for the developments we want to present and we lose nothing as the more general definition is proved to be equivalent to the restricted one, when it makes sense. We thus save the task of proving this equivalence and, more important, of tackling the problem of characterising when this general definition makes sense. On the other hand, we have to prove the equivalence with automata 'with spontaneous transitions' which will make for the general definition. This happens to be somewhat subtle and difficult and will not be considered in these lectures.

### The definition of $\mathbb{K}$-automata.

**Definition 2.** A $\mathbb{K}$-automaton over $A^*$ is a *labelled directed graph* together with *two maps* from its set of vertices to $\mathbb{K}$. Its vertices are called *states*; its edges, called *transitions*, are associated with *weighted labels*, that are pairs $(a, k)$, with $k$ in $\mathbb{K}$ and $a$ in $A$, also written $k\,a$ or $a\,|\,k$ depending on the context.

We denote a $\mathbb{K}$-automaton over $A^*$ by $\mathcal{A} = \langle \mathbb{K}, A, Q, I, E, T \rangle$ where:

- $\mathbb{K}$ is the weight semiring and $A$ is the alphabet which generates $A^*$.
- $Q$ is the set of *states* of $\mathcal{A}$, also called the *dimension* of $\mathcal{A}$.
- $I$ and $T$ are respectively the *initial* and *final* functions, functions from $Q$ into $\mathbb{K}$, that is, elements of $\mathbb{K}^Q$ , and
- $E \subseteq Q \times A \times \mathbb{K} \times Q$, the set of *weighted transitions*, is the graph of *a partial function* from $Q \times A \times Q$ into $\mathbb{K} \setminus \{0_{\mathbb{K}}\}$.

[2] The notation in EAT is $\mathcal{A} = \langle Q, A, E, I, T \rangle$. It has been changed in order to be consistent with the use of the Awali platform.

Let $e = (p, x, k, q)$ be a transition of $\mathcal{A}$:

- the *source* of $e$, written $\iota(e)$, is $p$ and the *destination* of $e$, written $\tau(e)$, is $q$,
- the *label* of $e$, written $\ell(e)$, is $x$,
- the *weight* of $e$, written $\mathbf{w}(e)$, is $k$, and
- the *weighted label*, *w-label* for short, of $e$, written $\mathbf{wl}(e)$, is the *monomial* $k\,x$.

The assumption that $E$ is a partial function implies that two distinct transitions cannot have the same source, destination, and label, the one that it is a partial function into $\mathbb{K} \setminus \{0_{\mathbb{K}}\}$ implies that the weight of a transition cannot be equal to $0_{\mathbb{K}}$.

A state $p$ is said to be *initial* (resp. *final*) if $I(p)$ (resp. $T(p)$) is different from $0_{\mathbb{K}}$, that is, if $p$ is in the *support* of the function $I$ (resp. $T$).

Figure 1 shows two $\mathbb{N}$-automata, $\mathcal{B}_1$ (left) and $\mathcal{C}_1$ (middle) and one $\mathbb{N}\mathsf{min}$-automaton $\mathcal{M}_1$ (right).

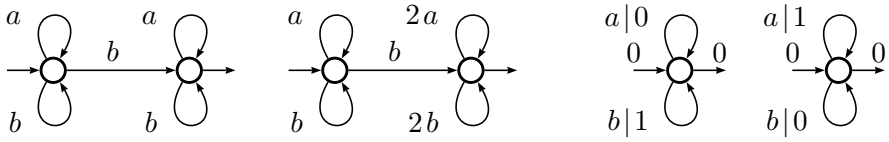



Figure 1: Two $\mathbb{N}$-automata and one $\mathbb{N}\mathsf{min}$-automaton

One reads on this figure conventions commonly taken when drawing weighted automata. For *classical semirings of numbers*, the *multiplicative* identity element $1_{\mathbb{K}}$ remains implicit, hence incoming (resp. outgoing) arrows without label indicate that the initial (resp. final) map gives the corresponding state the value $1_{\mathbb{K}}$, and accordingly a transition without weight is supposed to be given the weight $1_{\mathbb{K}}$. For *tropical* semirings, the *multiplicative* identity which is the number 0 is explicitely written, and so is the weight 1 which is just another element of the weight semiring. In this case also, a monomial $k\,a$ is often written as $a\,|\,k$.

The automaton $\mathcal{A}$ is *finite* if the set $E$ is finite, which is equivalent, when the alphabet $A$ is finite, to the condition that $Q$ be finite. Every automaton we consider in this lecture (but not in this course) is finite.

Most often, the weight semiring $\mathbb{K}$ is understood from the context and we simply write $\mathcal{A} = \langle\, A, Q, I, E, T \,\rangle$. In the sequel, $\mathcal{A}$ denotes a $\mathbb{K}$-automaton.

**Paths and computations.** Since $\mathcal{A}$ is a *graph*, a *path* in $\mathcal{A}$ is a sequence of transitions such that the destination of every transition is the source of the next one; it can be written as:

$$d_1 = e_1 e_2 \cdots e_n \qquad \text{or as} \qquad d_1 = p_0 \xrightarrow{k_1\,x_1} p_1 \xrightarrow{k_2\,x_2} p_2 \cdots \xrightarrow{k_n\,x_n} p_n \ .$$

The *label*, respectively the *weight* and the *w-label*, of a path $d$, is the *product* of the labels, respectively of the weights and of the w-labels, of the transitions of $d$. For instance,

$$\ell(d_1) = x_1\, x_2 \cdots x_n\,, \qquad \mathbf{w}(d_1) = k_1\, k_2 \cdots k_n \ ,$$
$$\text{and} \quad \mathbf{wl}(d_1) = (k_1\, k_2 \cdots k_n)\, x_1\, x_2 \cdots x_n \ .$$

A *computation* in $\mathcal{A}$ is a path together with the values of the initial and final functions at the ends of the path. For instance, the computation corresponding to the above path $d_1$ is $c_1 = (I(p_0), d_1, T(p_n))$ and the label, the weight and the weighted label of $c_1$ are

$$\ell(c_1) = \ell(d_1)\,, \quad \mathbf{w}(c_1) = I(p_0)\,\mathbf{w}(d_1)\,T(p_n) \quad \text{and} \quad \mathbf{wl}(c_1) = I(p_0)\,\mathbf{wl}(d_1)\,T(p_n) \ .$$

The *length* of a path $d$, or of a computation $c$, is the number of transitions it contains and is denoted by $|d|$ (or $|c|$). For instance, $|c_1| = |d_1| = n$. The weighted label of a computation associated with a path that does not start at an initial state or end at a final state is hence equal to $0_{\mathbb{K}}$.

The set of computations of an automaton $\mathcal{A}$ is denoted by $\mathsf{C}_{\mathcal{A}}$. (The seemingly tetrapylotomic distinction between *path* and *computation* will be used later on — in Lemma 7 for instance — but may be forgotten in most cases.)

**The weight of a word and the behaviour of a $\mathbb{K}$-automaton.** The *weight*, or *multiplicity*, of a word $w$ in $\mathcal{A}$ is the *sum* of the weights of the computations in $\mathcal{A}$ whose word label is $w$. Hence the automaton $\mathcal{A}$ associates with every *word* in $A^*$ a value in $\mathbb{K}$, that is, defines a *map* from $A^*$ to $\mathbb{K}$ that we denote by $\big|\mathcal{A}\big|$:

$$\forall w \in A^* \qquad \big|\mathcal{A}\big|(w) = \sum_{c \in \mathsf{C}_{\mathcal{A}},\, \ell(c) = w} \mathbf{w}(c) \ . \tag{1.1}$$

This sum (1.1) is well-defined if $w$ is the word label of a *finite* number only of computations in $\mathcal{A}$. With the definition we have taken for automata, this condition holds for every $w$ in $A^*$ when $Q$ is finite: a word of length $n$ is the label of a computation of length $n$ and there are only a finite number of those in $\mathcal{A}$.[3]

This function $\big|\mathcal{A}\big|\colon A^* \to \mathbb{K}$ is said to be *realised by* $\mathcal{A}$ and is called the *behaviour* of $\mathcal{A}$. It is the natural generalisation of the *language* accepted by a Boolean automaton: the latter can be seen as an application from $A^*$ to $\mathbb{B}$ that maps a word $w$ to $1_{\mathbb{B}}$ or $0_{\mathbb{B}}$ according to whether $w$ belongs or not to the language.

**Example 3. (Automata of Figure 1).** A simple calculation yields the behaviour of $\mathcal{B}_1$: for every $w$ in $\{a, b\}^*$, $\big|\mathcal{B}_1\big|(w) = |w|_b$ holds.

[3] Another case where the weight of every word is well-defined even when $Q$ is infinite is when the *structure* of $\mathcal{A}$ insures that every word is the label of at most a finite number of computations, *e.g.*when $\mathcal{A}$ is *deterministic* or *sequential*, a case that will be considered in Lecture III.

It is as simple to determine that $\big|\mathcal{M}_1\big|(w) = \mathsf{min}\{|w|_a, |w|_b\}$ for every $w$ in $\{a,b\}^*$.

If we use the convention that each word $w$ of $\{a,b\}^*$ is considered as a number written in binary, interpreting $a$ as the digit 0 and $b$ as the digit 1, and if we write $\overline{w}$ for the integer represented by the word $w$, it is easy to verify that $\overline{w}$ is *computed* by $\mathcal{C}_1$, in the sense that $\big|\mathcal{C}_1\big|(w) = \overline{w}$ , for every $w$ in $\{a,b\}^*$.

Before going further, we take a number of notation and definitions concerning these maps from $A^*$ into $\mathbb{K}$.

### 1.3 Series over $A^*$ with coefficients in $\mathbb{K}$

For any set $E$, the set of maps from $E$ to $\mathbb{K}$ is usually written $\mathbb{K}^E$ and canonically inherits from $\mathbb{K}$ a structure of semiring when equipped with *pointwise* addition and multiplication.

When $E$ is a monoid $A^*$, we equip $\mathbb{K}^{A^*}$ with another multiplication which derives from the *monoid structure* of $A^*$ and we thus use different notation and terminology for these maps together with this other semiring structure.

Any map from $A^*$ to $\mathbb{K}$ is a *formal power series* over $A^*$ with coefficients in $\mathbb{K}$ — abbreviated as $\mathbb{K}$-series over $A^*$, or even as *series* if there is ambiguity neither on $\mathbb{K}$ nor on $A^*$. The set of these series is written $\mathbb{K}\langle\langle A^*\rangle\rangle$ . If $s$ is a series, the image of an element $w$ of $A^*$ by $s$ is written $\langle s, w\rangle$ rather than $s(w)$ or $(w)s$ and is called the *coefficient of $w$ in $s$*.

For all $s$ and $t$ in $\mathbb{K}\langle\langle A^*\rangle\rangle$, and all $k$ in $\mathbb{K}$, the following operations are defined:

(i) the (left and right) *'exterior' multiplications*:

$$k\,s \quad \text{and} \quad s\,k \quad \text{by} \quad \forall w \in A^* \quad \langle k\,s, w\rangle = k\langle s, w\rangle \quad \text{and} \quad \langle s\,k, w\rangle = \langle s, w\rangle k$$

(ii) the pointwise *addition*:

$$s + t \quad \text{by} \quad \forall w \in A^* \quad \langle s + t, w\rangle = \langle s, w\rangle + \langle t, w\rangle$$

(iii) and the *Cauchy product*:

$$s\,t \quad \text{by} \quad \forall w \in A^* \quad \langle s\,t, w\rangle = \sum_{\substack{u,v\in A^*\\ uv=w}} \langle s, u\rangle\langle t, v\rangle \ . \tag{1.2}$$

Addition makes $\mathbb{K}\langle\langle A^*\rangle\rangle$ a commutative monoid; together with the two exterior multiplications, it makes $\mathbb{K}\langle\langle A^*\rangle\rangle$ a left, and right, *module* over $\mathbb{K}$.

For every $w$ in $A^*$, the number of factorisations $u\,v = w$ is finite, hence the sum in (1.2) is well-defined, and so is the Cauchy product of two series $s$ and $t$ in $\mathbb{K}\langle\langle A^*\rangle\rangle$. This product, together with the pointwise addition, makes $\mathbb{K}\langle\langle A^*\rangle\rangle$ a semiring and, together with the exterior multiplications, a left, and right, *algebra* over $\mathbb{K}$.

With these notations and definitions, the *behaviour* $\big|\mathcal{A}\big|$ of $\mathcal{A}$ is a series of $\mathbb{K}\langle\langle A^*\rangle\rangle$, the *coefficient* of $w$ in $\big|\mathcal{A}\big|$ is $\big\langle\big|\mathcal{A}\big|, w\big\rangle$ and Example 3 is rewritten as $\big\langle\big|\mathcal{B}_1\big|, w\big\rangle = |w|_b$ , $\big\langle\big|\mathcal{C}_1\big|, w\big\rangle = \overline{w}$ and $\big\langle\big|\mathcal{M}_1\big|, w\big\rangle = \mathsf{min}\{|w|_a, |w|_b\}$ , for every $w$ in $\{a,b\}^*$.

**Lemma 4.** *Let $\mathbb{Q}$ be a finite set. The semiring of square matrices of dimension $Q$ with entries in $\mathbb{K}\langle\langle A^*\rangle\rangle$ is isomorphic to that of series over $A^*$ with coefficient in $\mathbb{K}^{Q\times Q}$, that is, $\mathbb{K}\langle\langle A^*\rangle\rangle^{Q\times Q} \cong \mathbb{K}^{Q\times Q}\langle\langle A^*\rangle\rangle$ .* ■

**Support of a series – polynomials – characteristic series.** The *support* of a series $s$, written $\mathsf{supp}\, s$, is the subset of words in $A^*$ whose coefficient in $s$ is not $0_{\mathbb{K}}$. For instance, $\mathsf{supp}\,|\mathcal{B}_1| = A^*bA^*$, and $\mathsf{supp}\,|\mathcal{M}_1| = A^*$ (since 0 *is not the zero* of $\mathbb{N}\mathsf{min}$).

A series with finite support is a *polynomial*; the set of polynomials over $A^*$ with coefficients in $\mathbb{K}$ is written $\mathbb{K}\langle A^*\rangle$. It is a *sub-algebra* of $\mathbb{K}\langle\langle A^*\rangle\rangle$.

Conversely, if $L$ is a language of $A^*$, $\underline{L}$ denotes the *characteristic* series of $L$ in $\mathbb{N}\langle\langle A^*\rangle\rangle$ or, more generally, in $\mathbb{K}\langle\langle A^*\rangle\rangle$, for any $\mathbb{K}$ given by the context:

$$\forall w \in A^* \qquad \langle \underline{L}, w\rangle = \begin{cases} 1_{\mathbb{K}} & \text{if } w \in L \\ 0_{\mathbb{K}} & \text{otherwise}. \end{cases}$$

Accordingly, a series is said *to be characteristic* if it is equal to the characteric series of its own support.

**Support of an automaton – characteristic automata.** A *Boolean* automaton is exactly a $\mathbb{B}$-automaton and will usually be denoted as such to avoid ambiguity.

Every $\mathbb{K}$-automaton $\mathcal{A}$ can be transformed into a $\mathbb{B}$-automaton, called the *support* of $\mathcal{A}$, denoted by $\mathsf{supp}\,\mathcal{A}$, and obtained by replacing every non-zero (non $0_{\mathbb{K}}$) weight on transitions by $1 = 1_{\mathbb{B}}$. Of course, $\mathsf{supp}\,(|\mathcal{A}|)$ may be strictly contained in $|\mathsf{supp}\,\mathcal{A}|$. The equality $\mathsf{supp}\,(|\mathcal{A}|) = |\mathsf{supp}\,\mathcal{A}|$ holds if $\mathbb{K}$ is positive.

If the weight of all transitions of a $\mathbb{K}$-automaton $\mathcal{A}$, as well as the non-zero values of the initial and final functions, are equal to $1_{\mathbb{K}}$ — as it is the case for $\mathcal{B}_1$ for instance — then $\mathcal{A}$ is said to be *characteristic*.

Given a Boolean automaton $\mathcal{A}$ and a semiring $\mathbb{K}$ (usually it is $\mathbb{N}$), $\underline{\mathcal{A}}$ denotes the characteristic $\mathbb{K}$-automaton the support of which is $\mathcal{A}$. Of course, $|\underline{\mathcal{A}}|$ is not equal to $|\mathcal{A}|$, which is a characteristic series. More precisely, if $\mathcal{A}$ is a Boolean automaton over $A^*$, then, for every $w$ of $A^*$, $\langle |\underline{\mathcal{A}}|, w\rangle$ is the number of successful computations labelled by $w$ in $\mathcal{A}$, that is, the *degree of ambiguity* of $w$ in $\mathcal{A}$.

**Exercises** See Exer. 3. to 6., p. 29.

# 2 Rationality

We give a first characterisation of the behaviour of finite weighted automata. It is not the one which will be most important for us, not the one on which we build

the developments to come in the next two lectures. It is of interest though for three reasons; first because it is the generalisation of the characterisation that is most common when dealing with classical Boolean automata; second because it is the one that holds also for (weighted) automata on *non free monoids*, third because it paves the way to the second characterisation we are aiming at.

## 2.1 The matrix description of $\mathbb{K}$-automata

Graphs can be defined by their *incidence matrix*; we extend this description to automata.

We write the *set* $E$ as a *square matrix* of dimension $Q$: every entry $E_{p,q}$ is the sum of the weighted labels of all transitions in $\mathcal{A}$ from $p$ to $q$, thus a *linear combination of letters in $A$ with coefficients in* $\mathbb{K}$, hence in $\mathbb{K}\langle A^*\rangle$, and can indeed be seen as the label of a unique transition that goes from $p$ to $q$. Along the same line, we see $I$ as a *row-vector*[4] and $T$ as a *column-vector* in $\mathbb{K}^Q$ and the $\mathbb{K}$-automaton $\mathcal{A}$ is then written as $\mathcal{A} = \langle\, I, E, T \,\rangle$.

*Remark* 5. Writing $\underline{E}$ rather than $E$ for the incidence matrix would be more correct as it would mark the distinction between the *set* of transitions and the *matrix* that is derived from it. Such a distinction has proved to be necessary when studying the *validity* of weighted automata with spontaneous transitions (transitions whose label is the empty word, a case which is ruled out in the model we study here) but we shall not study this question in these lectures. On the contrary, we shall deal with the set of transitions of an automaton almost exclusively under the form of the incidence matrix, for which we choose the simpler and lighter notation.

**Example 6** (Example 3 cont.)**.** The $\mathbb{N}$-automaton $\mathcal{B}_1$ over $\{a,b\}^*$ shown in Figure 1 (left) may be written as

$$\mathcal{B}_1 = \left\langle \begin{pmatrix} 1 & 0 \end{pmatrix}, \begin{pmatrix} a+b & b \\ 0 & a+b \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\rangle ,$$

whereas the $\mathbb{N}$-automaton $\mathcal{C}_1$ shown in Figure 1 (middle) is written as

$$\mathcal{C}_1 = \left\langle \begin{pmatrix} 1 & 0 \end{pmatrix}, \begin{pmatrix} a+b & b \\ 0 & 2\,a+2\,b \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\rangle .$$

The $\mathbb{N}\mathsf{min}$-automaton $\mathcal{M}_1$ shown in Figure 1 (right) is written as

$$\mathcal{M}_1 = \left\langle \begin{pmatrix} 0 & 0 \end{pmatrix}, \begin{pmatrix} 0\,a+1\,b & +\infty \\ +\infty & 1\,a+0\,b \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \end{pmatrix} \right\rangle ,$$

[4] I recently became aware that in linear algebra treatises all vectors are column-vectors by definition and a row-vector is the *transpose* of a column-vector. It seems to me that having both possibilities is handier and I stay with my habit, at least in these lecture notes.

The description of the transitions of an automaton by a matrix is justified by the fact that a walk over a graph corresponds to a matrix multiplication. This is expressed by the following statement.

**Lemma 7.** *Let $\mathcal{A} = \langle I, E, T \rangle$ be a $\mathbb{K}$-automaton over $A^*$ of finite dimension. For every integer $n$, $E^n$ is the matrix of the sums of the weighted labels of paths of length $n$.*

*Proof.* By induction on $n$. The assertion is true for $n = 1$ (and also for $n = 0$ by convention). The definition of the $(n+1)$th power of $E$ is given by the equation:

$$\forall p, q \in Q \qquad (E^{n+1})_{p,q} = \sum_{r \in Q} (E^n)_{p,r} \, E_{r,q} \ .$$

Every path of length $n+1$ is the concatenation of a path of length $n$ with a path of length 1, that is, a single transition. We can therefore write[5]

$$\left\{ c \;\middle|\; c \;:=\; p \xrightarrow[\mathcal{A}]{} q \,, \quad |c| = n+1 \right\} = \bigcup_{r \in Q} \left\{ (d, e) \;\middle|\; d \;:=\; p \xrightarrow[\mathcal{A}]{} r \,, \quad |d| = n \,, \quad e \;:=\; r \xrightarrow[\mathcal{A}]{} q \in E \right\} \ ,$$

and hence

$$\begin{aligned} \sum \left( \mathbf{wl}(c) \;\middle|\; c \;:=\; p \xrightarrow[\mathcal{A}]{} q \,, \quad |c| = n+1 \right) &= \sum_{r \in Q} \left( \mathbf{wl}(d)\,\mathbf{wl}(e) \;\middle|\; d \;:=\; p \xrightarrow[\mathcal{A}]{} r \,, \quad |d| = n \,, \quad e \;:=\; r \xrightarrow[\mathcal{A}]{} q \in E \right) \\ &= \sum_{r \in Q} \left( \sum \left( \mathbf{wl}(d) \;\middle|\; d \;:=\; p \xrightarrow[\mathcal{A}]{} r \,, \quad |d| = n \right) \right) E_{r,q} \ . \end{aligned}$$

As $\sum \left( \mathbf{wl}(d) \;\middle|\; d \;:=\; p \xrightarrow[\mathcal{A}]{} r \,, \quad |d| = n \right) = (E^n)_{p,r}$ by the induction hypothesis, the lemma is proved. ∎

Since every word $w$ of $A^*$ appears in the support of the entries of at most the only power $E^n$ where $n = |w|$, the sum $\sum_{n \in \mathbb{N}} E^n$ is well-defined as we shall see in the next subsection and it holds:

**Corollary 8.** *Let $\mathcal{A} = \langle I, E, T \rangle$ be a $\mathbb{K}$-automaton of finite dimension. Then:*

$$\left| \mathcal{A} \right| \;=\; \sum_{n \in \mathbb{N}} (I \cdot E^n \cdot T) \;=\; I \cdot \left( \sum_{n \in \mathbb{N}} E^n \right) \cdot T \ . \qquad \blacksquare$$

[5] Recall that the length of a path $c$ is written $|c|$.

## 2.2 Rational series

As hinted by Corollary 8, the characterisation of the behaviour of (finite) weighted automata implies the definition of *infinite sums* of series. There are essentially two ways for tackling this problem: the axiomatic approach and the topological one. The axiomatic approach consists in imposing a set of properties to an operation called *star*. But the star in the weight semirings we have listed above and that we want to be able to deal with will not meet these properties. We are thus bound to take the topological way, which is not a bad solution anyway.

### 2.2.1 The topological way

**Topological semirings.** Defining a topology on a set is the way to define the notions of limit (or convergence) and, then, of *infinite sums*. Since $\mathbb{K}\langle\langle A^*\rangle\rangle = \mathbb{K}^{A^*}$ is the *set of maps* from $A^*$ to $\mathbb{K}$, it is naturally equipped with the *product topology* of the topology on $\mathbb{K}$, which is also the *simple convergence* topology, that is, if $(s_n)_{n\in\mathbb{N}}$ is a sequence of series

$$s_n \text{ converges to } s \text{ if and only if } \quad \text{for all } w \text{ in } A^*, \quad \langle s_n, w\rangle \text{ converges to } \langle s, w\rangle \ .$$

The semirings we consider are equipped with a *topology defined by a distance* — a more intuitive notion than an abstract definition of the topology — whether it is the *discrete topology* (in the cases of $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Z}\mathsf{min}$, *etc.*) or a more classical one (in the cases of $\mathbb{Q}$, $\mathbb{R}$, *etc.*). Since $A^*$ is countable, the product topology on $\mathbb{K}\langle\langle A^*\rangle\rangle$ is also defined by a distance. If $\mathbf{c}$ is a distance on $\mathbb{K}$ (bounded by 1), the map defined by

$$\forall s,t \in \mathbb{K}\langle\langle A^*\rangle\rangle \qquad \mathbf{d}(s,t) = \frac{1}{2}\sum_{n\in\mathbb{N}}\left(\frac{1}{2^n}\mathsf{max}\left\{\, \mathbf{c}(\langle s,w\rangle, \langle t,w\rangle) \;\middle|\; |m| = n \right\}\right) \tag{2.1}$$

is a distance on $\mathbb{K}\langle\langle A^*\rangle\rangle$ that defines the simple convergence topology. In any case, *the origin of the topology on* $\mathbb{K}\langle\langle A^*\rangle\rangle$ *is the topology on* $\mathbb{K}$.

A semiring $\mathbb{K}$ is a *topological semiring* if not only the set $\mathbb{K}$ is equipped with a topology but if moreover both the addition and the multiplication are *continuous operations* with respect to that topology. If $\mathbb{K}$ is a topological semiring, so is $\mathbb{K}\langle\langle A^*\rangle\rangle$.

**Summable families.** Let $\mathbb{T}$ be a semiring[6] equipped with a distance $\mathbf{d}$ which makes it a topological semiring. We thus know precisely what means that an infinite sequence $(t_n)_{n\in\mathbb{N}}$ converges to a limit $t$ when $n$ tends to infinity. We must now give an equally precise meaning to the sum of an infinite family $(t_i)_{i\in I}$ and it turns out to be somewhat harder. The difficulty arises from the fact that we want a sort

[6] We temporarily change the symbol we use for a semiring on purpose: $\mathbb{T}$ not only plays the role of $\mathbb{K}$ but also of $\mathbb{K}\langle\langle A^*\rangle\rangle$ in this paragraph and the following.

of *associativity–commutativity* extended 'to infinity' and hence to ensure that the result and its existence does not depend on an arbitrary order put on the set $I$ of indices.

We shall therefore define an 'absolute' method of summability, and a family will be described as 'summable' if we can find an increasing sequence of finite sets of indices, a sort of 'kernels', such that not only do partial sums on these sets tend to a limit but above all that any sum on a finite set containing one of these kernels stays close to this limit. More precisely:

**Definition 9.** A family $(t_i)_{i\in I}$ of elements of $\mathbb{T}$ indexed by an arbitrary set $I$ is called *summable* if there exists $t$ in $\mathbb{T}$ such that, for all positive $\varepsilon$, there exists a finite subset $J_\varepsilon$ of $I$ such that, for all finite subsets $L$ of $I$ which contain $J_\varepsilon$, the distance between $t$ and the sum of the $t_i$ for $i$ in $L$ is less than $\varepsilon$; that is:

$$\exists t \in \mathbb{T}\,,\ \forall \varepsilon > 0\,,$$
$$\exists J_\varepsilon \text{ finite}\,,\ J_\varepsilon \subset I\,,\ \forall L \text{ finite}\,,\ J_\varepsilon \subseteq L \subset I \quad \mathbf{d}\left(\sum_{i\in L} t_i, t\right) \leqslant \varepsilon\,.$$

The element $t$ thus defined is *unique* and is called the *sum* of the family $(t_i)_{i\in I}$.

The sum just defined is equal to the usual sum if $I$ is finite, and we write:

$$t = \sum_{i\in I} t_i \quad .$$

We say that a family of series $(s_i)_{i\in I}$ is *locally finite* if for every $w$ in $A^*$ there is only a finite number of indices $i$ such that $\langle s_i, w\rangle$ is different from $0_{\mathbb{K}}$.

**Property 10.** *A locally finite family of power series is summable.* ∎

This simple property is a good example of what the topological structure placed on $\mathbb{K}\langle\langle A^*\rangle\rangle$ brings in. That we can *define a sum* for a locally finite family of series is trivial: pointwise addition is defined for every $w$, independently of any assumption about $\mathbb{K}$. To say that the family is *summable* adds extra information: it ensures that partial sums *converge* to the result of pointwise addition.

For every series $s$, the family of series $\{\langle s, w\rangle\, w \mid w \in A^*\}$, where $w$ is identified with its characteristic series, is locally finite, and we have

$$s = \sum_{w\in A^*} \langle s, w\rangle\, w \quad ,$$

which is the usual notation for series and which is thus justified. We also deduce from this notation that $\mathbb{K}\langle A^*\rangle$ is *dense* in $\mathbb{K}\langle\langle A^*\rangle\rangle$. Along the same line, the sum in Corollary 8 is locally finite since for each pair of indices $(p, q)$, the supports of all $\left(E^n_{p,q}\right)_{n\in\mathbb{N}}$ are pairwise disjoint, hence the sum is well-defined.

Property 10 extends beyond locally finite families and generalises to a proposition which links the summability of a family of series and that of families of coefficients.

**Property 11.** *A family* $(s_i)_{i\in I}$ *of series of* $\mathbb{K}\langle\langle A^*\rangle\rangle$ *is summable with sum* $s$ *if and only if, for every* $w$ *in* $A^*$*, the family* $(\langle s_i, w\rangle)_{i\in I}$ *of elements of* $\mathbb{K}$ *is summable with sum* $\langle s, w\rangle$ *.* ■

### 2.2.2 The star operation.

Let $t$ be an element of a topological semiring $\mathbb{T}$; it is possible for the family $(t^n)_{n\in\mathbb{N}}$ to be, or not to be, summable. If it is summable, we call its sum the 'star of $t$' and write it $t^*$:

$$t^* = \sum_{n\in\mathbb{N}} t^n \ .$$

Whether $t^*$ is defined or not depends on $t$, on $\mathbb{T}$, on the distance on $\mathbb{T}$, or on a combination of all these elements. For example, $(0_\mathbb{T})^* = 1_\mathbb{T}$ is defined for all $\mathbb{T}$ and any topology on $\mathbb{T}$; if $\mathbb{T} = \mathbb{Q}$, we have $(\frac{1}{2})^* = 2$ if $\mathbb{Q}$ is equipped with the natural topology, or undefined if the chosen topology is the discrete topology, while $1^*$ is not defined in either case.

#### The U identity

**Lemma 12.** *Let* $\mathbb{T}$ *be a topological semiring and* $t$ *an element of* $\mathbb{T}$ *whose star is defined. We have the double equality*

$$t^* = 1_\mathbb{T} + t\,t^* = 1_\mathbb{T} + t^*t \ . \qquad (\mathbf{U})$$

*Proof.* We obviously have $t^{\leqslant n} = 1_\mathbb{T} + t\,t^{<n} = 1_\mathbb{T} + t^{<n}t$. As $\mathsf{lim}\, t^{<n} = \mathsf{lim}\, t^{\leqslant n} = t^*$, and as *addition and multiplication are continuous operations* on $\mathbb{T}$, we obtain (**U**) by taking the limit of each side of the above equation. ■

*Remark* 13. If $\mathbb{T}$ is a topological *ring*, and if the star of $t$ is defined, (**U**) can be written $t^* - t\,t^* = t^* - t^*t = 1$ or $(1-t)\,t^* = t^*(1-t) = 1$ and so $t^*$ is the *inverse* of $1-t$. Hence the classic identity

$$t^* = \frac{1}{1-t} = 1 + t + t^2 + \cdots \qquad (2.2)$$

is justified in full generality. It also means that forming the star can be considered as a substitute of taking the inverse in a poor structure that has no inverse.

**Star of proper series** By analogy with polynomials and series in one variable, we call *constant term* of a series $s$ of $\mathbb{K}\langle\langle A^*\rangle\rangle$ the coefficient in $s$ of the empty word, the neutral element of $A^*$. A series is called *proper* if its constant term is zero. The sum of two proper series is a proper series; the product of a proper series with any other series is a proper series. If $s$ is proper, the family $(s^n)_{n\in\mathbb{N}}$ is locally finite and thus *the star of a proper series of* $\mathbb{K}\langle\langle A^*\rangle\rangle$ *is defined.*

In view of further developments, we take the following definition and notation. Let $s$ be a series of $\mathbb{K}\langle\langle A^*\rangle\rangle$; the *proper part* of $s$ is the proper series that coincides with $s$ for all the elements $w$ of $A^*$ other than $1_{A^*}$. It is convenient to write $s_0 = \mathsf{c}(s)$ for the constant term of $s$, and $s_\mathsf{p}$ for the proper part of $s$:

$$\mathsf{c}(s_\mathsf{p}) = \langle s_\mathsf{p}, 1_{A^*}\rangle = 0_\mathbb{K} \qquad \text{and} \qquad \forall\, w \in A^* \setminus 1_{A^*} \qquad \langle s_\mathsf{p}, w\rangle = \langle s, w\rangle\,,$$

and we write $s = s_0 + s_\mathsf{p}$ (rather than $s = s_0\, 1_{A^*} + s_\mathsf{p}$).

### 2.2.3 The set of rational series.

The *(*$\mathbb{K}$*-)rational operations* on $\mathbb{K}\langle\langle A^*\rangle\rangle$ are:

(i) the $\mathbb{K}$-algebra operations, that is:

- the left and right *exterior multiplications* by elements of $\mathbb{K}$;
- the (pointwise) *addition*;
- the (Cauchy) *product*;

(ii) the *star* operation, which is not defined everywhere.

Point (ii) leads us to tighten the notion of closure: a subset $\mathcal{E}$ of $\mathbb{K}\langle\langle A^*\rangle\rangle$ is *closed under star* if for every series $s$ in $\mathcal{E}$ such that $s^*$ is defined, then $s^*$ belongs to $\mathcal{E}$.

A subset of $\mathbb{K}\langle\langle A^*\rangle\rangle$ is *rationally closed* if it is closed under the rational operations; that is, if it is a sub-algebra of $\mathbb{K}\langle\langle A^*\rangle\rangle$ closed under the star operation. The intersection of any family of rationally closed subsets is rationally closed and thus the *rational closure* of a set $\mathcal{E}$ is the *smallest* rationally closed subset which contains $\mathcal{E}$, written $\mathbb{K}\mathrm{Rat}\,\mathcal{E}$.

**Definition 14.** A series of $\mathbb{K}\langle\langle A^*\rangle\rangle$ is $\mathbb{K}$-*rational* if it belongs to the rational closure of $\mathbb{K}\langle A^*\rangle$, the set of polynomials on $A^*$ with coefficients in $\mathbb{K}$. The set of $\mathbb{K}$-rational series (over $A^*$ with coefficients in $\mathbb{K}$) is written $\mathbb{K}\mathrm{Rat}\,A^*$.

If the monoid $A^*$ is implied by the context, we shall say $\mathbb{K}$-*rational series*, or just *rational series*, if $\mathbb{K}$ is also understood.

**Example 15.** (i) Let $A^*$ be the one-generator free monoid $\{x\}^*$ and $\mathbb{K}$ be a field $\mathbb{F}$. Then $\mathbb{F}\mathrm{Rat}\,x^*$ is exactly the set of series developments of ($\mathbb{F}$-)rational functions (that is, quotients of two polynomials) and this is where the name *rational* — rather the more common *regular* (for expressions and languages) — comes from.

(ii) If $\mathbb{K} = \mathbb{B}$, we simply write $\mathrm{Rat}\,A^*$ for $\mathbb{B}\mathrm{Rat}\,A^*$ and its elements are the *rational languages* (or rational subsets) of $A^*$.

## 2.3 The Fundamental Theorem of Finite Automata

We have then defined all notions that are necessary to establish a first characterisation of the behaviour of finite weighted automata. *Almost* all, indeed. The missing one is that of *strong semiring* which we will explained later. It insures that the semiring is 'regular enough' to allow a 'natural' computation for the star of a non proper series. All the semirings that we have mentioned above are strong and this hypothesis is not really restrictive. However, we have to include it in the following statement, for sake of correctness.

**Theorem 16.** *Let $\mathbb{K}$ be a strong semiring. A series of $\mathbb{K}\langle\langle A^*\rangle\rangle$ is rational if and only if it is the behaviour of some finite $\mathbb{K}$-automaton over $A^*$.*

The qualificative *fundamental* we give to this theorem — as well as the differentiation from the statement usually called 'Kleene Theorem' — is justified by the fact that the same statement holds for series over other monoids than free ones, over *graded monoids* as we shall see below, and even over others that will not be considered here. In less formal words, this statement amounts to say that, under mild and natural assumptions, the descriptive or computational power of *finite graphs* is exactly the same as the one of the *star operator* (in presence of algebra operations of course).

Theorem 16 sttates the equality of two families of series. Its proof consists in showing two inclusions.

### 2.3.1 Behaviours of finite weighted automata are rational series

**Proposition 17.** *The behaviour of a finite $\mathbb{K}$-automaton over $A^*$ is a rational series of $\mathbb{K}\langle\langle A^*\rangle\rangle$.*

The proof of Proposition 17 is based on a fundamental property.

**Lemma 18** (Arden)**.** *Let $s$ and $t$ be two series of $\mathbb{K}\langle\langle A^*\rangle\rangle$; if $s$ is a* proper *series, each of the equations*

$$X = sX + t \tag{2.3}$$

*and*

$$X = Xs + t \tag{2.4}$$

*has a* unique solution*: the series $s^*t$ and $ts^*$ respectively.*

*Proof.* In (**U**), we replace $t$ by $s$ and multiply on the left (resp. on the right) by $t$ and we obtain that $s^*t$ (resp. $ts^*$) is a solution of (2.3) (resp. of (2.4)). Conversely, if $u$ is a solution of the equation $X = t + sX$ ,

$$u = t + su \quad \Longrightarrow \quad u = t + st + s^2u = \cdots = s^{<n}t + s^nu$$

holds for all integers $n$. Since $s$ is proper, and multiplication continuous, $\lim s^n = \lim s^n u = 0$ holds, from which follows $u = \lim(s^{<n}t) = (\lim s^{<n})\,t = s^*t$. ∎

From which we deduce:

**Proposition 19.** *Let $s$ and $t$ be two proper series of $\mathbb{K}\langle\langle A^*\rangle\rangle$; the following equalities (or identities) hold:*

$$(s+t)^* = s^*(t\,s^*)^* = (s^*t)^*s^*\,, \tag{\mathbf{S}}$$

$$(s\,t)^* = 1 + s\,(t\,s)^*t\,, \tag{\mathbf{P}}$$

$$\forall n \in \mathbb{N} \qquad s^* = s^{<n}(s^n)^* \quad . \tag{\mathbf{Z}_n}$$

The identity (**S**) is called the *sum-star identity*, (**P**) the *product-star identity.*

*Remark* 20. It follows by Lemma 4 that a square matrix $m$ of dimension $Q$ with elements in $\mathbb{K}\langle\langle A^*\rangle\rangle$ is a proper series of $\mathbb{K}^{Q\times Q}\langle\langle A^*\rangle\rangle$ if all its elements are proper series; (we say in this case that $m$ is proper), and hence that the identities **S**, **P** and $\mathbf{Z}_n$ are satisfied by proper matrices.

*Proof of Proposition 17.* Let $\mathcal{A} = \langle\, I, E, T\,\rangle$ be an automaton whose behaviour is thus defined and equal to $|\mathcal{A}| = I \cdot E^* \cdot T$. This part then amounts to prove that the entries of the star of a proper matrix $E$ belong to the rational closure of the entries of $E$, a classical statement established in general in different setting.

We write $|\mathcal{A}| = I \cdot V$ with $V = E^* \cdot T$. Since $E$ is proper and by Lemmas 4 and 18, $V$ is *the unique solution* of

$$X = E \cdot X + T \tag{2.5}$$

and we have to prove that all entries of the vector $V$ belong to the rational closure of the entries of $E$. Lemma 18 already states that the property holds if $\mathcal{A}$ is of dimension 1. For $\mathcal{A}$ of dimension $Q$, we write (2.5) as a system of $\|Q\|$ equations:

$$\forall p \in Q \qquad V_p = \sum_{q\in Q} E_{p,q}V_q + T_p \quad . \tag{2.6}$$

We choose (arbitrarily) one element $q$ in $Q$ and by Lemma 18 again it comes:

$$V_q = E^*_{q,q}\left[\sum_{p\in Q\setminus\{q\}} E_{q,p}V_p + T_q\right] \quad ,$$

an expression for $V_q$ that can be substituted in every other equations of the system (2.6), giving a new system

$$\forall p \in Q \setminus \{q\} \qquad V_p = \sum_{r\in Q\setminus\{q\}} \left[E_{p,r} + E_{p,q}E^*_{q,q}E_{q,r}\right] V_r + E_{p,q}E^*_{q,q}T_q + T_p \quad .$$

And the property is proved by induction hypothesis. ∎

### 2.3.2 Rational series are behaviours of finite weighted automata

The converse of Proposition 17 reads as follow.

**Proposition 21.** *If $\mathbb{K}$ is a strong semiring and if $s$ is in $\mathbb{K}\mathrm{Rat}\, A^*$, there exixts a finite $\mathbb{K}$-automaton over $A^*$ whose behaviour is equal to $s$.*

We prove indeed that the family of behaviours of finite $\mathbb{K}$-automata over $A^*$ contains the polynomials (the characterisitic series of every letter indeed) and is closed under the exterior multiplication, the sum, the product, and, under the assumption of strongness of $\mathbb{K}$, under star. It follows from Definition 14 that this family contains $\mathbb{K}\mathrm{Rat}\, A^*$.

In order to establish the closure properties, it is convenient to define a restricted class of automata, called the *standard* automata.

#### Standard automata

**Definition 22.** A $\mathbb{K}$-automaton $\mathcal{A} = \langle I, E, T \rangle$ is *standard* if the initial vector $I$ has a single non-zero entry $i$, equal to $1_{\mathbb{K}}$, and if this unique initial state $i$ is not the destination of any transition whose label is non-zero.

In matrix terms, a standard automaton $\mathcal{A}$ can be written

$$\mathcal{A} = \left\langle \begin{pmatrix} 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & K \\ 0 & F \end{pmatrix}, \begin{pmatrix} c \\ U \end{pmatrix} \right\rangle , \tag{2.7}$$

since the entries of the $i$-th column of $E$ are (sums of the) weighted labels of the transitions the destination of which is $i$. The definition does not forbid the initial state $i$ from also being final; that is, the scalar $c$ is not necessarily zero. This value $c$ is the *constant term* of $|\mathcal{A}|$. the following does not participate to the proof of Proposition 21 but tells that standard automata are not 'too special'.

**Proposition 23.** *Every automaton $\mathcal{A}$ is equivalent to a* standard *automaton whose weighted labels are linear combinations of the weighted labels of $\mathcal{A}$.* ∎

We now define *operations* on standard automata that are parallel to the *rational operations*. Let $\mathcal{A}$ (as in (2.7)) and $\mathcal{A}'$ (with obvious translation) be two standard automata; the following standard $\mathbb{K}$-automata are defined:

- $$k\mathcal{A} = \left\langle \begin{pmatrix} 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & kK \\ 0 & F \end{pmatrix}, \begin{pmatrix} kc \\ U \end{pmatrix} \right\rangle$$

  and $$\mathcal{A}k = \left\langle \begin{pmatrix} 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & K \\ 0 & F \end{pmatrix}, \begin{pmatrix} ck \\ Uk \end{pmatrix} \right\rangle ;$$

- $\mathcal{A}+\mathcal{A}' = \left\langle \begin{pmatrix} 1 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & K & K' \\ 0 & F & 0 \\ 0 & 0 & F' \end{pmatrix}, \begin{pmatrix} c+c' \\ U \\ U' \end{pmatrix} \right\rangle$ ;

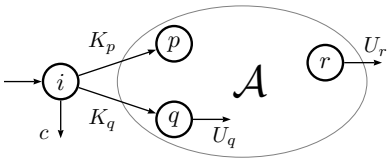
- $\mathcal{A}\cdot\mathcal{A}' = \left\langle \begin{pmatrix} 1 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & K & cK' \\ 0 & F & H \\ 0 & 0 & F' \end{pmatrix}, \begin{pmatrix} cc' \\ V \\ U' \end{pmatrix} \right\rangle$ ,

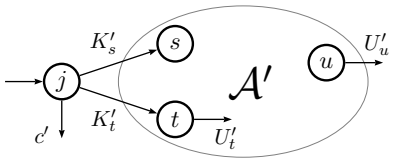
where $H = (U\cdot K')\cdot F'$ and $V = U\,c' + (U\cdot K')\cdot U'$ ;

- $\mathcal{A}^* = \left\langle \begin{pmatrix} 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & c^*K \\ 0 & G \end{pmatrix}, \begin{pmatrix} c^* \\ U\,c^* \end{pmatrix} \right\rangle$ ,

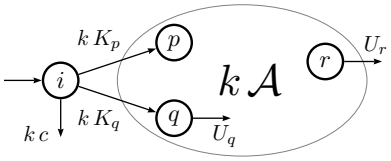
*which is defined if and only if* $c^*$ *is defined*, and where $G = U\cdot c^*K + F$ .

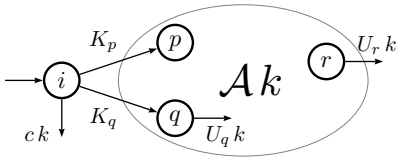
Some figures may help visualize these constructions. Let $\mathcal{A} = \langle\,\{i\}, E, T\,\rangle$ and $\mathcal{A}' = \langle\,\{j\}, E', T'\,\rangle$ be two standard automata drawn as:



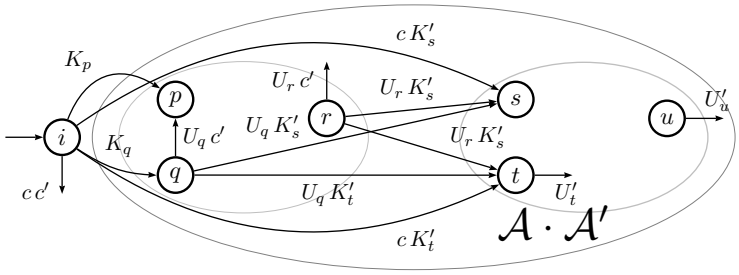



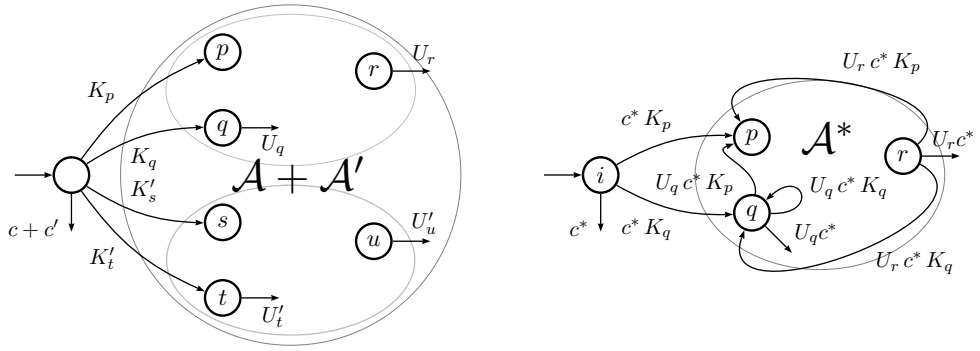

Then $k\,\mathcal{A}$ and $\mathcal{A}\,k$ are drawn as:




and $\mathcal{A}\cdot\mathcal{A}'$ , $\mathcal{A}+\mathcal{A}'$ , and $\mathcal{A}^*$ are respectively drawn as:






Straightforward computations show

**Proposition 24.**

$$\big|k\,\mathcal{A}\big| = k\,\big|\mathcal{A}\big|,\ \big|\mathcal{A}\,k\big| = \big|\mathcal{A}\big|\,k\ ,\ \big|\mathcal{A}+\mathcal{A}'\big| = \big|\mathcal{A}\big| + \big|\mathcal{A}'\big|,\ \textit{and}\ \big|\mathcal{A}\cdot\mathcal{A}'\big| = \big|\mathcal{A}\big|\big|\mathcal{A}'\big|.$$ ■

As expected, the case of the star operator is somewhat more complex. The automaton $\mathcal{A}^*$ is defined if and only if $c^*$ is defined; let $\big|\mathcal{A}\big|_{\mathsf{p}}$ be the *proper part* of the series $\big|\mathcal{A}\big|$. Then we have:

**Proposition 25.**

$$\big|\mathcal{A}^*\big| = c^*\,\big(\big|\mathcal{A}\big|_{\mathsf{p}}\,c^*\big)^*\ .$$ ■

The last step being given by the following which will be established in the next subsection after the definition of strong semirings.

**Proposition 26.** *Let $\mathbb{K}$ be a strong topological semiring and $A^*$ a free monoid. Let $s$ be a series of $\mathbb{K}\langle\langle A^*\rangle\rangle$, $s_0$ its constant term and $s_{\mathsf{p}}$ its proper part. Then $s^*$ is defined if and only if $s_0^*$ is defined and in this case we have*

$$s^* = (s_0^*\,s_{\mathsf{p}})^*\,s_0^* = s_0^*(s_{\mathsf{p}}\,s_0^*)^*\quad . \tag{2.8}$$

**Corollary 27.** *If $\mathbb{K}$ is a strong topological semiring, then $\big|\mathcal{A}^*\big| = \big|\mathcal{A}\big|^*$.* ■

*Proof of Proposition 21.* A trivial construction shows that the family of behaviours of standard automata contains the characteristic series of any letter of $A$, Proposition 24 that it contains the polynomials, Proposition 24 and Corollary 27 that it is rationally closed, and hence contains $\mathbb{K}\mathrm{Rat}\,A^*$. ■

**Strong semirings** As stated by Proposition 26, strong semirings give a framework in which the question whether *the star of an arbitrary series, not necessarily proper, is defined or not* can be given an answer and, when defined, how the star can be computed.

**Definition 28.** A topological semiring is *strong* if the product of two summable families is a summable family; that is, if the two families $(s_i)_{i\in I}$ and $(t_j)_{j\in J}$ are summable with sum $s$ and $t$ respectively, then the family $\{s_i\,t_j \mid (i,j)\in I\times J\}$ is summable with sum $s\,t$.

All the semirings which we shall consider are strong: semirings equipped with the discrete topology, the sub-semirings of $\mathbb{C}^n$ (equipped with the natural topology), and the positive semirings. We then easily verify:

**Property 29.** *The semirings of matrices and the semirings of series on $A^*$, with coefficients in a strong semiring, are strong.* ■

*Remark* 30. The notion of strong semiring has been introduced in *EAT* in order to have a sufficient condition for the proof of Proposition 26. Since then, the question was open whether there exist semirings that are not strong, although the answer was likely to be positive. An example of a non strong semiring has been given very recently by my colleague David Madore. The question whether there exist semirings in which (2.8) does not hold is still open.

*Remark* 31. Along the line of Remark 13, it holds that if $\mathbb{K}$ is a ring, a series of $\mathbb{K}\langle\langle A^*\rangle\rangle$ is invertible if, and only, if its constant term is invertible.

*Proof of Proposition 26.* The condition is necessary since $\langle s^n, 1_{A^*}\rangle = {s_0}^n$ and, if $s^*$ is defined, the coefficients of $1_{A^*}$ in $(s_n)_{n\in\mathbb{N}}$ form a summable family.

Conversely, assume that $(s_0^n)_{n\in\mathbb{N}}$ is summable, with sum $s_0^*$. For all pairs of integers $k$ and $l$, set

$$P_{k,l} = \sum_{\substack{i_0,i_1,\ldots,i_k\in\mathbb{N}\\ i_0+i_1+\cdots+i_k=l}} s_0^{i_0}\, s_\mathsf{p}\, s_0^{i_1}\, s_\mathsf{p} \cdots s_0^{i_{k-1}}\, s_\mathsf{p}\, s_0^{i_k} \quad .$$

By convention, set $P_{0,l} = s_0^l$ and $P_{k,0} = s_\mathsf{p}^k$. We verify by inspection that, for all integers $n$

$$s^n = (s_0 + s_\mathsf{p})^n = \sum_{l=0}^{l=n} P_{n-l,l} \quad . \tag{2.9}$$

By induction on $k$, we will show that the family

$$F_k = \{s_0^{i_0}\, s_\mathsf{p}\, s_0^{i_1}\, s_\mathsf{p} \cdots s_0^{i_{k-1}}\, s_\mathsf{p}\, s_0^{i_k} \mid i_0, i_1, \ldots, i_k \in \mathbb{N}\}$$

is summable in $\mathbb{K}A^*$, with sum

$$Q_k = (s_0^*\, s_\mathsf{p})^k\, s_0^* = s_0^*\, (s_\mathsf{p}\, s_0^*)^k \quad .$$

In fact, the hypothesis on $s_0$ ensures the property for $k = 0$, and also that the family $G = \{{s_0}^n\, s_\mathsf{p} \mid n \in \mathbb{N}\}$ is summable in $\mathbb{K}\langle\langle A^*\rangle\rangle$, with sum ${s_0}^* s_\mathsf{p}$. The family $F_{k+1}$ is the product of the families $G$ and $F_k$ and the assumption that $\mathbb{K}$, and hence $\mathbb{K}\langle\langle A^*\rangle\rangle$, is strong gives us the conclusion.

Hence we deduce that, for each $k$, the family $\{P_{k,l} \mid l \in \mathbb{N}\}$ is summable, with sum $Q_k$. The family $\{Q_k \mid k \in \mathbb{N}\}$ is locally finite, hence summable, with sum

$$t = \sum_{k=0}^{\infty} Q_k = (s_0^*\, s_\mathsf{p})^*\, s_0^* = s_0^*\, (s_\mathsf{p}\, s_0^*)^* \quad .$$

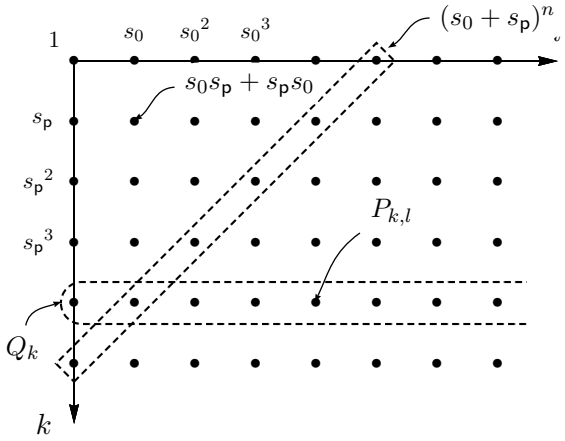



Figure 2: A graphical representation of Proposition 26

We can now easily finish the proof by showing that the 'doubly indexed' family $\{P_{k,l} \mid k, l \in \mathbb{N}\}$ is summable, with sum $t$. Equation (2.9), then ensure that the family $(s^n)_{n\in\mathbb{N}}$ is summable with sum $t$. ■

In the same spirit as Remark 20, we note that (2.8) holds for every matrix $m$ such that the star of its matrix of constant terms is defined. A particularly interesting case of this is where the matrix of constant terms is a strict upper triangular matrix.

## 2.4 Generalisation to graded monoids

**Graded monoids.** For the Cauchy product be always defined on $\mathbb{K}\langle\langle M\rangle\rangle$, independently of $\mathbb{K}$, it is necessary (and sufficient) that, for every $m$ in $M$, the set of pairs $(u, v)$ such that $u\,v = m$ is finite – we will say that $m$ is *finitely decomposable*.

The construction of series over $A^*$, which generalises that of series of one variable, shows that it is from the *length* of words in $A^*$ that we build a topology on $\mathbb{K}\langle\langle A^*\rangle\rangle$. The existence of an *additive length* is the main assumption that we shall make about $M$.

**Definition 32.** Let $M$ be a monoid. A function $\varphi\colon M \to \mathbb{N}$ is a *length* on $M$ if:

(i) $\varphi(m)$ is *strictly* positive for all $m$ other than $1_M$ ;

(ii) $\forall m, n \in M \qquad \varphi(m\,n) \leqslant \varphi(m) + \varphi(n)$ .

We shall say that a length is a *gradation* if it is *additive*; that is, if:

(iii) $\forall m, n \in M \qquad \varphi(m\,n) = \varphi(m) + \varphi(n)$ ;

and that $M$ is *graded* if it is equipped with a gradation.

**Example 33.** (i) Every free monoid is graded.

(ii) *Every cartesian product of free monoids*, in particular, every free commutative monoid, and, more generally, every trace (or free partially commutative) monoid is graded.

The definition implies that $\varphi(1_M) = 0$ and that a finite monoid, more generally a monoid that contains an idempotent other than the identity (for example, a zero), cannot be equipped with a gradation. Any group, finite or infinite, is not a graded monoid.

**Proposition 34.** *In a finitely generated graded monoid, the number of elements whose length is less than an arbitrary given integer $n$ is finite.*

In other words, every element of a graded monoid $M$ can only be written in a finite number of different ways as the product of elements of $M$ other than $1_M$. We can deduce in particular:

**Corollary 35.** *In a finitely generated graded monoid, every element is finitely decomposable.* ∎

Note that a finite monoid is not graded, but that every element is nonetheless finitely decomposable. From Corollary 35, we deduce the proposition aimed at by Definition 32:

**Proposition 36.** *Let $M$ be a finitely generated graded monoid and $\mathbb{K}$ a semiring. Then $\mathbb{K}\langle\langle M\rangle\rangle$, equipped with the Cauchy product, is a semiring and a (left and right) algebra*[7] *over $\mathbb{K}$.* ∎

**The Fundamental Theorem of Finite Automata (bis).** After Proposition 36, the whole theory developed in Sec. 2.2 and 2.3 can be repeated, *mutatis mutandis*, while replacing the free monoid $A^*$ with any graded monoid $M$. In particular, we state:

**Definition 37.** A series of $\mathbb{K}\langle\langle M\rangle\rangle$ is $\mathbb{K}$-*rational* if it belongs to the rational closure of $\mathbb{K}\langle M\rangle$, the set of polynomials on $M$ with coefficients in $\mathbb{K}$. The set of $\mathbb{K}$-rational series (over $M$ with coefficients in $\mathbb{K}$) is written $\mathbb{K}\mathrm{Rat}\, M$.

**Example 38.** (i) The series $s = \sum_{n\in\mathbb{N}}(n+1)\,(a^n, b^n) = \big((a,b)^*\big)^2$ belongs to $\mathbb{N}\mathrm{Rat}\,(\{a\}^*\times\{b\}^*)$.

(ii) If $R \in \mathrm{Rat}\, A^*$ and $S \in \mathrm{Rat}\, B^*$, then $R\times S \in \mathrm{Rat}\,(A^*\times B^*)$.

And it holds:

**Theorem 39.** *Let $\mathbb{K}$ be a strong semiring and $M$ a graded monoid. A series of $\mathbb{K}\langle\langle M\rangle\rangle$ is rational if and only if it is the behaviour of some finite $\mathbb{K}$-automaton over $M$.*

[7] If $\mathbb{K}$ is a ring, $\mathbb{K}\langle\langle M\rangle\rangle$ is even what is classically called a *graded algebra*, which is the origin of the terminology chosen for graded monoids.

# 3 Recognisability

The second characterisation of the behaviour of finite weighted automata as *series realised by representations* will be central in many developments to come in these lectures. In contrast with the preceding one, it holds for series over a free monoid only.

## 3.1 $\mathbb{K}$-representations and $\mathbb{K}$-recognisable series

A $\mathbb{K}$-*representation of* $A^*$ *of dimension* $Q$ is a *morphism* $\mu$ from $A^*$ to the (multiplicative) monoid of square matrices of dimension $Q$ with entries in $\mathbb{K}$. By definition, indeed, for the multiplication of matrices be well-defined, the dimension $Q$ is *finite*. A $\mathbb{K}$-representation of $A^*$ (of dimension $Q$) is also the name we give to a *triple* $\langle I, \mu, T \rangle$ where, as before,

$$\mu \colon A^* \longrightarrow \mathbb{K}^{Q \times Q}$$

is a morphism and where $I$ and $T$ are two vectors:

$$I \in \mathbb{K}^{1 \times Q} \qquad \text{and} \qquad T \in \mathbb{K}^{Q \times 1} \ ;$$

that is, $I$ is a *row* vector and $T$ a *column* vector, of dimension $Q$, with entries in $\mathbb{K}$. Such a representation defines a map from $A^*$ to $\mathbb{K}$ by

$$\forall w \in A^* \qquad w \longmapsto I \cdot \mu(w) \cdot T \ ;$$

that is, the *series* $s$:

$$s = \sum_{w \in A^*} (I \cdot \mu(w) \cdot T)\, w \ .$$

The series $s$ of $\mathbb{K}\langle\langle A^* \rangle\rangle$ is *realised*, or *recognised*, by the representation $\langle I, \mu, T \rangle$. We also say that $\langle I, \mu, T \rangle$ *realises*, or *recognises*, the series $s$.

**Definition 40.** A series of $\mathbb{K}\langle\langle A^* \rangle\rangle$ is $\mathbb{K}$-*recognisable* if it is recognised by a $\mathbb{K}$-representation. The set of $\mathbb{K}$-recognisable series over $A^*$ is written $\mathbb{K}\mathrm{Rec}\, A^*$.

**Example 41** (Example 3 cont.)**.** Let $\langle I_1, \mu_1, T_1 \rangle$ be the representation defined by

$$\mu_1(a) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} , \quad \mu_1(b) = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} , \quad I_1 = \begin{pmatrix} 1 & 0 \end{pmatrix} \quad \text{and} \quad T_1 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \ .$$

For all $w$ in $\{a, b\}^*$, $I_1 \cdot \mu_1(w) \cdot T_1 = |w|_b$ holds, hence the series $t_1 = \sum_{w \in A^*} |w|_b\, w$ is $\mathbb{N}$-recognisable.

**Proposition 42.** *Every finite linear combination, with coefficients in* $\mathbb{K}$*, of* $\mathbb{K}$*-recognisable series over* $A^*$ *is a* $\mathbb{K}$*-recognisable series.*

*Proof.* Let $s$ and $t$ be two $\mathbb{K}$-recognisable series over $A^*$, respectively recognised by the $\mathbb{K}$-representations $\langle I, \mu, T \rangle$ and $\langle J, \kappa, U \rangle$. For all $k$ in $\mathbb{K}$ the series $ks$ is recognised by the representation $\langle kI, \mu, T \rangle$, the series $sk$ by the representation $\langle I, \mu, Tk \rangle$, and the series $s + t$ by the representation $\langle K, \pi, V \rangle$ defined by the following block-decomposition:

$$K = \begin{pmatrix} I & J \end{pmatrix}, \quad \pi(w) = \begin{pmatrix} \mu(w) & 0 \\ 0 & \kappa(w) \end{pmatrix}, \quad V = \begin{pmatrix} T \\ U \end{pmatrix} . \qquad \blacksquare$$

Every morphism of semirings $\varphi\colon \mathbb{K} \to \mathbb{L}$ extends to a morphism from $\mathbb{K}\langle\langle A^* \rangle\rangle$ to $\mathbb{L}\langle\langle A^* \rangle\rangle$, still denoted by $\varphi$, by the pointwise map: for every $s$ in $\mathbb{K}\langle\langle A^* \rangle\rangle$, $\varphi(s)$ is defined by $\langle \varphi(s), w \rangle = \varphi(\langle s, w \rangle)$ for every $w$ in $A^*$. If $\langle I, \mu, T \rangle$ is a representation of the series $s$ of $\mathbb{K}\langle\langle A^* \rangle\rangle$, then $\langle \varphi(I), \varphi \circ \mu, \varphi(T) \rangle$ is a representation of $\varphi(s)$. It then follows:

**Proposition 43.** *Let $\varphi\colon \mathbb{K} \to \mathbb{L}$ be a morphism of semirings. The image under $\varphi$ of a $\mathbb{K}$-recognisable series over $A^*$ is an $\mathbb{L}$-recognisable series over $A^*$.* $\blacksquare$

**Consistency with the classical definition of recognisable sets.** For $\mathbb{K} = \mathbb{B}$, Definition 40 coincides indeed with the definition of the *recognisable subsets* of a monoid *as the sets that are saturated by a congruence of finite index.*

If $s$ is a $\mathbb{B}$-recognisable series over $A^*$, realised by the representation $\langle I, \mu, T \rangle$, then $\mu\colon A^* \to \mathbb{B}^{Q \times Q}$ is a morphism from $A^*$ *to a finite monoid.* The series $s$ of $\mathbb{B}\langle\langle A^* \rangle\rangle$, $s = \sum_{w \in A^*} (I \cdot \mu(w) \cdot T)\, w$ çan be seen as the *subset* $s = \mu^{-1}(P)$ of $A^*$ where $P = \left\{ p \in \mathbb{B}^{Q \times Q} \;\middle|\; I \cdot p \cdot T = 1_{\mathbb{B}} \right\}$.

Conversely, a morphism $\alpha$ from $A^*$ into a finite monoid $N$ is a morphism from $A^*$ into the monoid of Boolean matrices of dimension $N$ (the representation of $N$ by right translations over itself) and the $\mathbb{B}$-representation that realises any subset recognised by $\alpha$ easily follows.

## 3.2 The key lemma

The specificity of the *free monoid* in terms of representation is expressed in the following statement.

**Lemma 44.** *Let $\mathbb{K}$ be a semiring and $A$ a finite alphabet. Let $Q$ be a finite set and $\mu\colon A^* \to \mathbb{K}^{Q \times Q}$ a morphism. We set*

$$X = \sum_{a \in A} \mu(a)\, a \ .$$

*Then, for every $w$ in $A^*$, $\langle X^*, w \rangle = \mu(w)$ holds.*

*Proof.* The matrix $X$ is a proper series of $\mathbb{K}^{Q\times Q}\langle\langle A^*\rangle\rangle$ and hence $X^*$ is defined. We first prove, by induction on the integer $n$, that

$$X^n = \sum_{w\in A^n} \mu(w)\, w \ ,$$

an equality trivially verified for $n=0$, and true by definition for $n=1$. It follows that

$$X^{n+1} = X^n\cdot X = \Big(\sum_{w\in A^n}\mu(w)\,w\Big)\cdot\Big(\sum_{a\in A}\mu(a)\,a\Big) = \sum_{(w,a)\in A^n\times A}(\mu(w)\cdot\mu(a))\,wa$$
$$= \sum_{(w,a)\in A^n\times A}\mu(wa)\,wa = \sum_{v\in A^{n+1}}\mu(v)\,v \ ,$$

since, for each integer $n$, $A^{n+1}$ is in bijection with $A^n\times A$ as $A^*$ is freely generated by $A$. For the same reason, $A^*$ is the *disjoint* union of the $A^n$, for $n$ in $\mathbb{N}$, and it follows that, for every $w$ in $A^*$:

$$\langle X^*, w\rangle = \langle X^{|w|}, w\rangle = \mu(w) \ . \qquad \blacksquare$$

**Example 45** (Example 3 cont.)**.** Take $\mathbb{K}=\mathbb{N}$ and $A^*=\{a,b\}^*$. Then

$$\begin{pmatrix} a+b & b\\ 0 & a+b\end{pmatrix} = \mu_1(a)\,a+\mu_1(b)\,b \quad \text{with} \quad \mu_1(a)=\begin{pmatrix}1&0\\0&1\end{pmatrix},\quad \mu_1(b)=\begin{pmatrix}1&1\\0&1\end{pmatrix}.$$

## 3.3 The Kleene–Schützenberger Theorem

We can now get to our main point: finite $\mathbb{K}$-automata over $A^*$ and $\mathbb{K}$-representations of $A^*$ are one and a same thing when $A$ is finite. We state this under the classical form but we are really interested by the transformations of automata into representations and conversely.

**Theorem 46** (Kleene–Schützenberger)**.** *Let $\mathbb{K}$ be a strong semiring, and $A$ a finite alphabet. A series of $\mathbb{K}\langle\langle A^*\rangle\rangle$ is $\mathbb{K}$-rational if and only if it is $\mathbb{K}$-recognisable. That is:*

$$\mathbb{K}\mathrm{Rec}\,A^* = \mathbb{K}\mathrm{Rat}\,A^* \ .$$

*Proof.* We prove the two inclusions, one at a time:

$$\mathbb{K}\mathrm{Rec}\,A^*\subseteq\mathbb{K}\mathrm{Rat}\,A^* \qquad\text{and}\qquad \mathbb{K}\mathrm{Rat}\,A^*\subseteq\mathbb{K}\mathrm{Rec}\,A^* \ . \tag{3.1}$$

Each of the inclusions is proved in the form of a property and is obtained from the Fundamental Theorem together with the freeness of $A^*$ and the finiteness of $A$ by means of the key Lemma 44.

**Property 47.** *If $A$ is finite, $\mathbb{K}$-recognisable series on $A^*$ are $\mathbb{K}$-rational.*

*Proof.* Let $\langle I, \mu, T \rangle$ be a representation which recognises a series $s$; that is, $\langle s, w \rangle = I \cdot \mu(w) \cdot T$, for every $w$ in $A^*$. Let $\langle I, X, T \rangle$ be the automaton defined by

$$X = \sum_{a \in A} \mu(a)\, a \ .$$

By Lemma 44, we have

$$s = \sum_{w \in A^*} (I \cdot \mu(w) \cdot T)\, w = I \cdot \left( \sum_{w \in A^*} (\mu(w))\, w \right) \cdot T = I \cdot X^* \cdot T \ .$$

The series $s$ is the behaviour of the $\mathbb{K}$-automaton $\langle I, X, T \rangle$. Since $A$ *is finite* this automaton is finite and, by the Fundamental Theorem, $s$ belongs to $\mathbb{K}\mathrm{Rat}\, A^*$. ∎

**Property 48.** *If $\mathbb{K}$ is a strong semiring, $\mathbb{K}$-rational series on $A^*$ are $\mathbb{K}$-recognisable.*

*Proof.* By Theorem 16, a $\mathbb{K}$-rational series $s$ is the behaviour of a finite $\mathbb{K}$-automaton $\langle I, X, T \rangle$ and the entries of $X$ are finite linear combinations of elements of $A$ (and those of $I$ and $T$ are scalar). We can therefore write $X = \sum_{a \in A} \mu(a)\, a$ where $\mu(a)$ is the matrix of coefficients of the letter $a$ in $X$. By Lemma 44, we have

$$\forall w \in A^* \qquad \langle s, w \rangle = \langle I \cdot X^* \cdot T, w \rangle = I \cdot \mu(w) \cdot T \,,$$

and the series $s$ is recognised by the *representation* $\langle I, \mu, T \rangle$. ∎

The two inclusions (3.1) prove the theorem. ∎

On the basis of Theorem 46, we write an automaton $\mathcal{A}$ over $A^*$ indifferently as $\mathcal{A} = \langle I, E, T \rangle$ or as a representation $\mathcal{A} = \langle I, \mu, T \rangle$ with $E = \sum_{a \in A} \mu(a)\, a$.

**Example 49.** (i) **Generating function**. Let $L$ be a language of $A^*$. The *generating function* $g_L$ of $L$ is the series over one variable (written $z$ in general):

$$g_L = \sum_{n \in \mathbb{N}} a_n z^n \ ,$$

such that, for every $n$ in $\mathbb{N}$, $a_n$ is the *number of words of $L$ of length $n$*.

Let $\mathcal{A} = \langle I, \mu, T \rangle$ be an *unambiguous (Boolean) automaton* of dimension $Q$ and $L = |\mathcal{A}|$ the language accepted by $\mathcal{A}$, that is, the behaviour of the ($\mathbb{N}$-)characteristic automaton of $\mathcal{A}$ is a characteristic series: $|\underline{\mathcal{A}}| = \underline{L}$. Let $\pi$ be the $Q \times Q$-matrix with entries in $\mathbb{N}$ defined by:

$$\pi = \sum_{a \in A} \underline{\mu(a)} \ .$$

Then, $\langle I, \pi, T \rangle$ is a representation of $g_L$, that is, for every $n$ in $\mathbb{N}$, $a_n = I \cdot \pi^n \cdot T$.

(ii) **Probabilistic automata**. A $P \times Q$-matrix with entries in $\mathbb{R}$ (or in $\mathbb{Q}$) is said to be *stochatic* if all entries are non negative and if the sum of all entries of every

row is equal to 1. An automaton over $A^*$, $\mathcal{A} = \langle I, \mu, T \rangle$ is said to be *probabilistic* if $I$ and $\mu(a)$, for every $a$ in $A$, are stochastic and if $T$ has 0-1 entries.

For every $w$ in $A^*$, $\langle\!|\mathcal{A}|\!, w\rangle = I \cdot \mu(w) \cdot T$ can be interpreted as the *probability of acceptance* of $w$ by $\mathcal{A}$. Together with a probabilistic automaton $\mathcal{A}$, any $\eta$ in $\mathbb{R}$, $0 \leq \eta < 1$, defines the langage

$$L(\mathcal{A}, \eta) = \{w \in A^* \mid \langle\!|\mathcal{A}|\!, w\rangle \geq \eta\}$$

and such a language is called a *stochastic language*. The family of stochastic languages strctly contains the one of rational languages.

**Computation of coefficients.** The description of automata as representations leads to an efficient solution to the problem of computing the coefficient $\langle s, w \rangle$ of a rational series $s$. Suppose that $s$ is given by a finite automaton $\mathcal{A} = \langle I, E, T \rangle$ or, which is the same, by a representation $\mathcal{A} = \langle I, \mu, T \rangle$ of dimension $n$.

Then, $\langle s, w \rangle = I \cdot \mu(w) \cdot T$ and the computation of $\mu(w)$ would cost $\mathrm{O}(\ell n^3)$ where $\ell$ is the length of $w$; the last step to get $I \cdot \mu(w) \cdot T$ would add another $\mathrm{O}(n^2)$. But a smarter solution is possible. The computation of the succession of the $\ell$ vectors $I \cdot \mu(u)$ of $\mathbb{K}^n$ for all prefixes $u$ of $w$ would cost $\mathrm{O}(\ell n^2)$ with a final overhead of $\mathrm{O}(n)$ in order to get the result.

In the Boolean case, this method of computation for testing whether a word is accepted or not by a non-deterministic automaton is known as the *lazy* or *on-the-fly* determinisation.

## 3.4 The Hadamard product

The *Hadamard product* of series $s$ and $t$, denoted by $s \odot t$, is indeed the product of maps into a monoid:

$$\forall w \in A^* \qquad \langle s \odot t, w \rangle = \langle s, w \rangle \langle t, w \rangle \ .$$

The Hadamard product is defined on general series but it is its effect on recognisable series which interests us, and we first define a product on *representations*.

**Tensor product of $\mathbb{K}$-representations.** Let $X$ be a matrix of dimension $P \times P'$ and $Y$ a matrix of dimension $R \times R'$ (with entries in the same semiring $\mathbb{K}$); the *tensor product* of $X$ by $Y$, written $X \otimes Y$, is a matrix of dimension $(P \times R) \times (P' \times R')$ defined by

$$\forall p \in P\,,\ \forall p' \in P'\,,\ \forall r \in R\,,\ \forall r' \in R' \qquad X \otimes Y_{(p,r),(p',r')} = X_{p,p'} Y_{r,r'} \ .$$

If $\mathbb{K}$ is *commutative*, the tensor product is also commutative, and we keep this hypothesis in this subsection. The next statement, a classical equation in matrix calculus, is a matter of an easy verification.

**Lemma 50.** *Let $\mathbb{K}$ be a commutative semiring. Let $X$, $Y$, $U$ and $V$ be four matrices with entries in $\mathbb{K}$, respectively of dimension $P\times Q$, $P'\times Q'$, $Q\times R$ and $Q'\times R'$.*

$$(X\otimes Y)\cdot(U\otimes V)=(X\cdot U)\otimes(Y\cdot V)\ .$$ ∎

It then follows:

**Proposition 51** (Tensor product of representations)**.** *Let $\mathbb{K}$ be a commutative semiring. Let $\mu\colon A^*\to\mathbb{K}^{Q\times Q}$ and $\kappa\colon A^*\to\mathbb{K}^{R\times R}$ be two representations. The map $\mu\otimes\kappa$, defined for all $(u,v)$ in $A^*\times A^*$ by*

$$[\,\mu\otimes\kappa\,](u,v)=\mu\,(u)\otimes\kappa\,(v)$$

*is a representation of $A^*\times A^*$ in $\mathbb{K}^{(Q\times R)\times(Q\times R)}$.*

*Proof.* For all $(u,v)$ and $(u',v')$ in $A^*\times A^*$, we have:

$$\begin{aligned}([\,\mu\otimes\kappa\,](u,v))\cdot([\,\mu\otimes\kappa\,](u',v')) &= (\mu\,(u)\otimes\kappa\,(v))\cdot(\mu\,(u')\otimes\kappa\,(v'))\\ &= (\mu\,(u)\cdot\mu\,(u'))\otimes(\kappa\,(v)\cdot\kappa\,(v'))\\ &= \mu\,(u\,u')\otimes\kappa\,(v\,v')=[\,\mu\otimes\kappa\,](u\,u',v\,v')\ .\end{aligned}$$ ∎

**Hadamard product of recognisable series.** The Hadamard product is to series what intersection is to sets, which really makes sense only if the semiring of coefficients is commutative.

**Theorem 52** (Schützenberger)**.** *Let $\mathbb{K}$ be a commutative semiring. Then $\mathbb{K}\mathrm{Rec}\,A^*$ is closed under Hadamard product.*

*Proof.* Let $s$ realised by $\langle\, I,\mu,T\,\rangle$ and $t$ realised by $\langle\, J,\kappa,U\,\rangle$ be two series in $\mathbb{K}\mathrm{Rec}\,A^*$. Since the map $w\mapsto(w,w)$ is a morphism from $A^*$ to $A^*\times A^*$, Proposition 51 implies that the map $w\mapsto\mu\,(w)\otimes\kappa\,(w)$ is also a morphism, and we also write it $\mu\otimes\kappa$.

By definition we have, for all $w$ in $A^*$,

$$\langle s\odot t,w\rangle=(I\cdot\mu\,(w)\cdot T)\,(J\cdot\kappa\,(w)\cdot U)=(I\cdot\mu\,(w)\cdot T)\otimes(J\cdot\kappa\,(w)\cdot U)$$

the second equality expressing the product of two elements of $\mathbb{K}$ as the tensor product of two $1\times1$ matrices. Lemma 50 (applied three times) yields:

$$\langle s\odot t,w\rangle=(I\otimes J)\cdot(\mu\,(w)\otimes\kappa\,(w))\cdot(T\otimes U)=(I\otimes J)\cdot([\,\mu\otimes\kappa](w))\cdot(T\otimes U)\ .$$

Since $\mathbb{K}$ is commutative, $\mu\otimes\kappa$ is a $\mathbb{K}$-representation, and $s\odot t$ is recognisable and realised by $(I\otimes J,\,\mu\otimes\kappa,T\otimes U)$. ∎

*Remark* 53. Lemma 50, Proposition 51 and then the proof of Theorem 52 hold indeed under the weaker hypothesis that every entry of one representation commutes with every entry of the other. It is the case in particular when one of the series is *characteristic* or, more prececisely, when one of the series is realised by a *characteristic representation*, with obvious meaning. This setting will also be the one of transducers and relations — automata and series over direct products of free monoids — and their composition (see Exercise 15. and Lect. V).

*Remark* 54. As a consequence of Theorem 46, *the Hadamard product of two* $\mathbb{K}$*-rational series on* $A^*$ *is a* $\mathbb{K}$*-rational series* (if $\mathbb{K}$ is a commutative semiring, or if one is characteristic). Moreover, the tensor product of representations of $A^*$ translates directly into a construction on $\mathbb{K}$-automata over $A^*$ whose labels are linear combinations of letters of $A$, which is the natural generalisation of the Cartesian product of Boolean automata, and which we can call the *Hadamard product* of $\mathbb{K}$-automata.

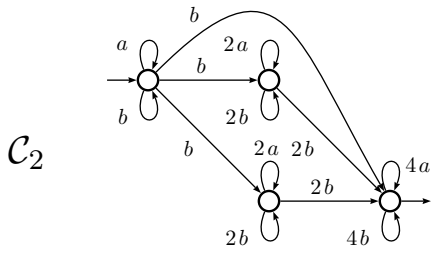



Figure 3: $\mathcal{C}_2$, the Hadamard product of $\mathcal{C}_1$ by itself

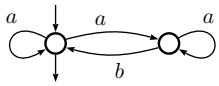
**Example 55.** The $\mathbb{N}$-automaton $\mathcal{C}_2$ of Fig. 3 is the Hadamard product of the $\mathbb{N}$-automaton $\mathcal{C}_1$ of Fig. 1 by itself. Therefore, for every $w$ in $A^*$, $\langle\!|\mathcal{C}_2|\!, w\rangle = \overline{w}^2$ holds.

# 4 Exercises

1. **Semiring structure.** Is $\mathbb{M} = \langle\, \mathbb{N}, \mathsf{max}, +, 0, 0 \,\rangle$ a semiring?

2. **Positive semiring.** Give an example of a semiring in which the sum of any two non-zero elements is non-zero but which *is not* positive. [*Hint: consider a sub-semiring of* $\mathbb{N}^{2\times 2}$.]

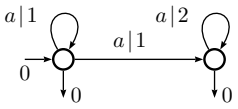
3. **Example of $\mathbb{N}$-automaton.** (a) Compute the coefficient of $a^3ba^2ba$ in the series realised by the $\mathbb{N}$-automaton:



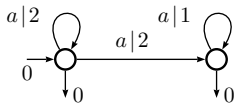

(b) Give the general formula for the coefficient of every word of $A^*$.

4. **Examples of $\mathbb{N}$min, $\mathbb{N}$max-automata.** Let $\mathcal{E}_1$ be the $\mathbb{N}$min-automaton over $\{a\}^*$ shown in Fig. 4 (a) and $\mathcal{E}_2$ the $\mathbb{N}$max-automaton shown in the same figure. Similarly, let $\mathcal{E}_3$ and $\mathcal{E}_4$ be the $\mathbb{N}$min and $\mathbb{N}$max-automata shown in Fig. 4 (b).

Give a formula for $\langle\!|\mathcal{E}_1|\!\rangle, a^n\rangle$, $\langle\!|\mathcal{E}_2|\!\rangle, a^n\rangle$, $\langle\!|\mathcal{E}_3|\!\rangle, a^n\rangle$, and $\langle\!|\mathcal{E}_4|\!\rangle, a^n\rangle$.




(a) The automata $\mathcal{E}_1$ and $\mathcal{E}_2$




(b) The automata $\mathcal{E}_3$ and $\mathcal{E}_4$

Figure 4: Four 'tropical' automata

5. **A $\mathbb{Z}$-automaton.** Build a $\mathbb{Z}$-automaton $\mathcal{D}_1$ such that $\langle\!|\mathcal{D}_1|\!\rangle, w\rangle = |w|_a - |w|_b$ , for every $w$ in $A^*$.

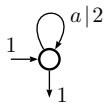
6. **Support of $\mathbb{Z}$-automata.** Give an example of a $\mathbb{Z}$-automaton $\mathcal{A}$ such that the inclusion $\mathsf{supp}\,(|\mathcal{A}|) \subseteq |\mathsf{supp}\,\mathcal{A}|$ is strict.

7. **Automata construction.** Let $\underline{a^*}$ be the characteristic $\mathbb{N}$-series of $a^*$ : $\underline{a^*} = \sum_{n\in\mathbb{N}} a^n$ . Give an 'automatic' proof (that is, by means of automata constructions) for:

$$(\underline{a^*})^2 = \sum_{n\in\mathbb{N}} (n+1)\, a^n \quad .$$

8. **Shortest run and $\mathbb{N}$min-automata.** Build a $\mathbb{N}$min-automaton $\mathcal{F}_1$ such that, for every $w$ in $A^*$, $\langle\!|\mathcal{F}_1|\!\rangle, w\rangle$ is the minimal length of runs of '$a$' 's in $w$, that is, if $w = a^{n_0}\, b\, a^{n_1}\, b \cdots a^{n_{k-1}}\, b\, a^{n_k}$, then $\langle\!|\mathcal{F}_1|\!\rangle, w\rangle = \mathsf{min}\{n_0, n_1, \ldots, n_k\}$.

9. **Identification of a $\mathbb{Q}$-automaton.** Show that the final function of the $\mathbb{Q}$-automaton $\mathcal{Q}_2$ over $\{a\}^*$ depicted on the right in Figure 5 (where every transition is labelled by $a\,|\,1$) can be specified in such a way the result is equivalent to $\mathcal{Q}_1$ depicted on the left.



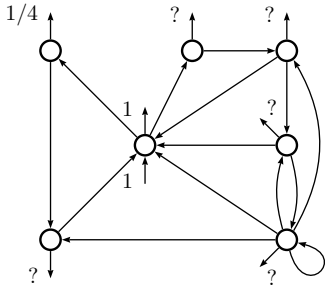




Figure 5: Two $\mathbb{Q}$-automata

10. **Ambiguous automata.** Show that it is decidable whether a Boolean automaton is unambiguous or not. [*Hint: Note that this is* not *a result nor a proof on weighted automata but on Boolean automata. It is put here in view of Example 49.* ]

11. **Representation with finite image.** Let $s$ be a $\mathbb{K}$-recognisable series of $A^*$, realised by a representation $\langle I, \mu, T \rangle$ of dimension $Q$. Show that if $\mu(A^*)$ is a finite submonoid of $\mathbb{K}^{Q\times Q}$, then, for every $k$ in $\mathbb{K}$ the set $s^{-1}(k) = \{w \in A^* \mid \langle s, w\rangle = k\}$ is a recognisable language of $A^*$.

12. **Support of $\mathbb{Z}$-rational series.** (a) Give an example of a $\mathbb{Z}$-rational series over $A^*$ whose support is not a recognisable language of $A^*$.

(b) Give an example of a $\mathbb{Z}$-rational series over $A^*$ which is an $\mathbb{N}$-series (that is, all coefficients are non-negative) and which is not an $\mathbb{N}$-rational series over $A^*$.

13. **Support of $\mathbb{Z}$-rational series.** (a) Prove that the support of an $\mathbb{N}$-rational series over $A^*$ is a recognisable language of $A^*$.

(b) Let $s$ be in $\mathbb{N}\mathrm{Rec}\, A^*$. Prove that for any $k$ in $\mathbb{N}$, the sets
$$s^{-1}(k) = \{w \in A^* \mid \langle s, w\rangle = k\} \quad \text{and} \quad s^{-1}(k + \mathbb{N}) = \{w \in A^* \mid \langle s, w\rangle \geqslant k\}$$
are recognisable languages of $A^*$.

(c) Give an example of a $\mathbb{Z}$-rational series $s$ over $A^*$ such that there exists an integer $z$ such that $s^{-1}(z)$ is not a recognisable language of $A^*$.

14. **Support of $\mathbb{Z}\mathsf{min}$-rational series.** (a) Let $s$ be a $\mathbb{N}\mathsf{min}$-rational series over $A^*$. Prove that for any $k$ in $\mathbb{N}$, the sets
$$s^{-1}(k) = \{w \in A^* \mid \langle s, w\rangle = k\} \qquad \text{and} \qquad s^{-1}(k + \mathbb{N}) = \{w \in A^* \mid \langle s, w\rangle \geqslant k\}$$
are recognisable languages of $A^*$.

(b) Give an example of a $\mathbb{Z}\mathsf{min}$-rational series $s$ over $A^*$ such that there exists an integer $z$ such that $s^{-1}(z)$ is not a recognisable language of $A^*$.

15. **Recognisable series in direct product of free monoids.** Let $\mathbb{K}$ be a commutative semiring. The two semirings $\mathbb{K}\langle\langle A^*\rangle\rangle$ and $\mathbb{K}\langle\langle B^*\rangle\rangle$ are canonically subalgebras of $\mathbb{K}\langle\langle A^* \times B^*\rangle\rangle$; the injection is induced by
$$u \mapsto (u, 1_{B^*}) \qquad \text{and} \qquad v \mapsto (1_{A^*}, v) \ ,$$
for all $u$ in $A^*$ and all $v$ in $B^*$. Modulo this identification, a product $(k\,u)\,(h\,v)$ is written $k\,h\,(u, v)$ and the extension by linearity of this notation gives the following definition.

**Definition 56.** Let $s$ be in $\mathbb{K}\langle\langle A^*\rangle\rangle$ and $t$ be in $\mathbb{K}\langle\langle B^*\rangle\rangle$. The *tensor product* of $s$ and $t$, written $s \otimes t$, is the series of $\mathbb{K}\langle\langle A^* \times B^*\rangle\rangle$ defined by:
$$\forall (u, v) \in A^* \times B^* \qquad \langle s \otimes t, (u, v)\rangle = \langle s, u\rangle \langle t, v\rangle \ .$$

On the other hand, $\mathbb{K}$-*recognisable series* over a non-free monoid $M$ are defined, exactly as the $\mathbb{K}$-recognisable series over a free monoid, as the series realised by a $\mathbb{K}$-representation $\langle I, \mu, T \rangle$, where $\mu$ is a morphism from $M$ into $\mathbb{K}^{Q\times Q}$.

Establish:

**Proposition 57.** *A series $s$ of $\mathbb{K}\langle\langle A^* \times B^*\rangle\rangle$ is recognisable if and only if there exists a finite family $\{r_i\}_{i\in I}$ of series of $\mathbb{K}\mathrm{Rec}\, A^*$ and a finite family $\{t_i\}_{i\in I}$ of series of $\mathbb{K}\mathrm{Rec}\, B^*$ such that*
$$s = \sum_{i\in I} r_i \otimes t_i \ .$$

# Notation Index

# General Index