Five lectures in the theory of
Weighted Automata and Transducers

Jacques Sakarovitch

IRIF/CNRS–Université Paris Diderot & Télécom-ParisTech

Lectures notes of the Master Parisien de Recherche en Informatique

Course 2.16 — Finite automata based computation models

December 2018 – January 2019

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. The note and the reference sections are still missing, as well as the general index.

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 min or $-\infty$ when the 'addition' is 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

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

That is, φ 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.

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.

cf. EAT, p. 402

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).

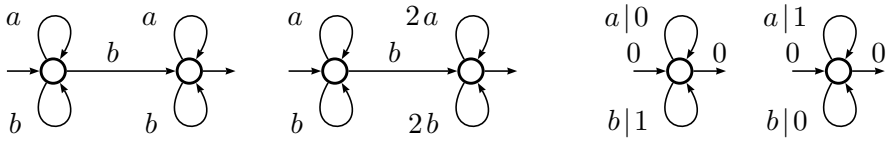


Figure 1: Two $\mathbb{N}$-automata and one $\mathbb{N}\mathsf{min}$-automaton

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

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

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* $\mathbb{K}$ (*cf.* Exercise 16. for more details on the definition of distance on $\mathbb{K}\langle\langle A^*\rangle\rangle$).

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

[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.

to be somewhat harder. The difficulty arises from the fact that we want a sort 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 ε, there exists a finite subset J_ε of I such that, for all finite subsets L of I which contain J_ε, the distance between t and the sum of the t_i for i in L is less than ε; 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:

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

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

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

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

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

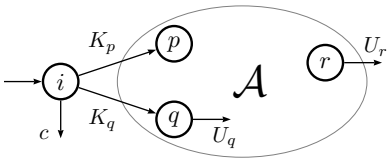
- $\mathcal{A}\cdot\mathcal{A}' = \left\langle \begin{pmatrix} 1 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & K & cK' \\ 0 & F & H \\ 0 & 0 & F' \end{pmatrix}, \begin{pmatrix} cc' \\ V \\ U' \end{pmatrix} \right\rangle$,

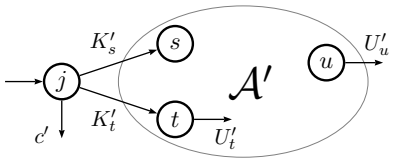
where $H = (U\cdot K')\cdot F'$ and $V = U\,c' + (U\cdot K')\cdot U'$;

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

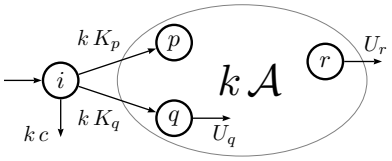
which is defined if and only if c^* *is defined*, and where $G = U\cdot c^*K + F$.

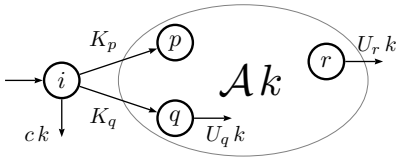
Some figures may help visualize these constructions. Let $\mathcal{A} = \langle\,\{i\}, E, T\,\rangle$ and $\mathcal{A}' = \langle\,\{j\}, E', T'\,\rangle$ be two standard automata drawn as:



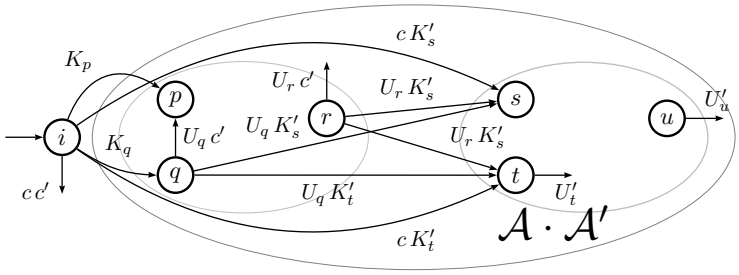


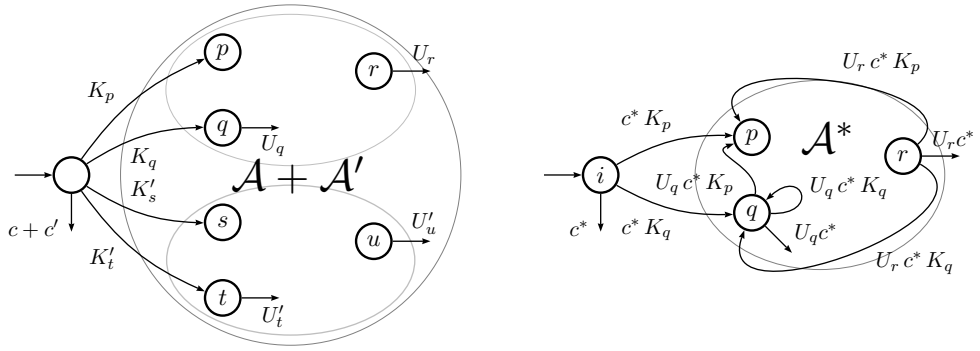
Then $k\,\mathcal{A}$ and $\mathcal{A}\,k$ are drawn as:





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_{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

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

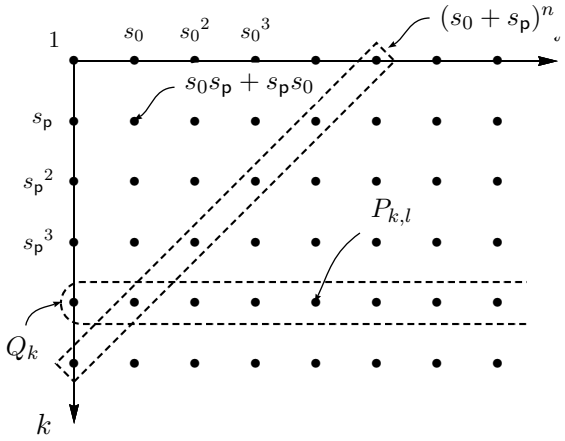


Figure 2: A graphical representation of Proposition 26

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* μ 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 φ, 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 φ 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 α 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 α 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

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

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

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

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 ℓ 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 ℓ 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 $Q'\times R'$.*

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

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:

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

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 1×1 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 $(I\otimes J,\ \mu\otimes\kappa, T\otimes U)$. $\blacksquare$

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 *tensor product* of $\mathbb{K}$-automata.

More precisely, if $\mathcal{A} = \langle I, E, T \rangle$ and $\mathcal{B} = \langle J, F, U \rangle$ are two automata of dimension Q and R respectively, then $\mathcal{A} \otimes \mathcal{B} = \langle I \otimes J, E \otimes F, T \otimes U \rangle$ where

$$(E \otimes F)_{(p,r),(q,s)} = E_{p,q} \odot F_{r,s}$$

for every p, q in Q and every r, s in R.

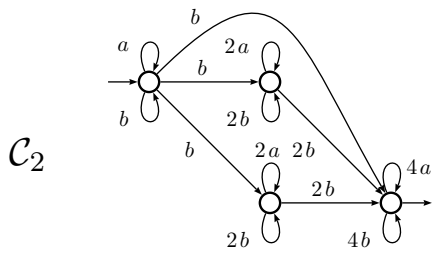


Figure 3: $\mathcal{C}_2 = \mathcal{C}_1 \otimes \mathcal{C}_1$, the tensor product of $\mathcal{C}_1$ by itself

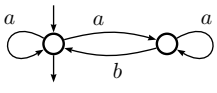
Example 55. The $\mathbb{N}$-automaton $\mathcal{C}_2$ of Fig. 3 is the Hadamard product of the $\mathbb{N}$-automaton $\mathcal{C}_1$ of Fig. 1 by itself. Therefore, for every w in A^*, $\langle\!|\mathcal{C}_2|\!, w\rangle = \overline{w}^2$ holds.

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

(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 $\langle\!|\mathcal{D}_1|\!, w\rangle = |w|_a - |w|_b$, 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 $\mathsf{supp}\,(\langle\!|\mathcal{A}|\!\rangle) \subseteq |\mathsf{supp}\,\mathcal{A}|$ 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^*, $\langle\!|\mathcal{F}_1|\!, w\rangle$ 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 $\langle\!|\mathcal{F}_1|\!, w\rangle = \mathsf{min}\{n_0, n_1, \ldots, n_k\}$.

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.

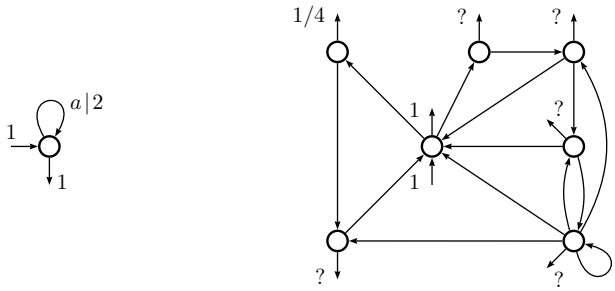


Figure 5: Two $\mathbb{Q}$-automata

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
$$s^{-1}(k) = \{w \in A^* \mid \langle s, w\rangle = k\} \text{ and } s^{-1}(k + \mathbb{N}) = \{w \in A^* \mid \langle s, w\rangle \geqslant k\}$$
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
$$s^{-1}(k) = \{w \in A^* \mid \langle s, w\rangle = k\} \quad \text{and} \quad s^{-1}(k + \mathbb{N}) = \{w \in A^* \mid \langle s, w\rangle \geqslant k\}$$
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
$$u \mapsto (u, 1_{B^*}) \quad \text{and} \quad v \mapsto (1_{A^*}, v) \ ,$$
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 μ 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*
$$s = \sum_{i\in I} r_i \otimes t_i \ .$$

16. **Distance on the semirings of series.**

A *distance* on any set S is a map $\mathbf{d}\colon S\times S \to \mathbb{R}_+$ with the three properties: for all x, y and z in S it holds:

(i) *symmetry*: $\mathbf{d}(x,y) = \mathbf{d}(y,x)$;
(ii) *positivity*: $\mathbf{d}(x,y) = 0 \Leftrightarrow x = y$;
(iii) *triangular inequality*: $\mathbf{d}(x,z) \leqslant \mathbf{d}(x,y) + \mathbf{d}(y,z)$.

If (iii) is replaced by the stronger property:

(iv) *triangular inequality*: $\mathbf{d}(x,z) \leqslant \mathsf{max}(\mathbf{d}(x,y), \mathbf{d}(y,z))$,

then $\mathbf{d}$ is said to be an *ultrametric distance*.

(a) Show that the function defined on S by

$$\forall x,y \in S \qquad \mathbf{d}(x,y) = \begin{cases} 0 & \text{if } x = y \\ 1 & \text{otherwise} \end{cases}$$

is an ultrametric distance. We call it the *discrete distance* on S.

Classically, a sequence $\left(s_n\right)_{n\in\mathbb{N}}$ of elements of S *converges* to s in S for the distance $\mathbf{d}$ if:

$$\forall \varepsilon > 0 \quad \exists N \in \mathbb{N} \quad \forall n > N \qquad \mathbf{d}(s_n, s) < \varepsilon \ .$$

In this way, a distance $\mathbf{d}$ defines a *topology* on S.

(b) Show that if S is equipped with the discrete distance, the only convergent sequences are the ultimately stationnary sequences.

Two distances on S are *equivalent* if the same sequences converge, that is, $\mathbf{d}$ and $\mathbf{d}'$ are equivalent if for any sequence $s = \left(s_n\right)_{n\in\mathbb{N}}$, s converges for $\mathbf{d}$ if and only if it converges for $\mathbf{d}'$.

(c) Show that one can always assume that *a distance is bounded by* 1, that is, if $\mathbf{d}$ is a distance on S, the function $\mathbf{f}$ defined by

$$\forall x,y \in S \qquad \mathbf{f}(x,y) = \mathsf{inf}\{\mathbf{d}(x,y), 1\}$$

is a distance, equivalent to $\mathbf{d}$.

(d) Let $\mathbf{d}$ and $\mathbf{d}$' be two distances on S. Show that if there exist two constant C and D in $\mathbb{R}_+ \setminus \{0\}$ such that

$$\forall x,y \in S \qquad C\,\mathbf{d}(x,y) \leqslant \mathbf{d}'(x,y) \leqslant D\,\mathbf{d}(x,y)$$

then $\mathbf{d}$ and $\mathbf{d}$' are equivalent. Is this condition necessary for $\mathbf{d}$ and $\mathbf{d}$' be equivalent?

Let $\mathbb{K}$ be a semiring. For s and t in $\mathbb{K}\langle\langle A^*\rangle\rangle$, let $\mathbf{e}(s,t)$ be the *gap between s and t*, defined as the minimal length of words on which s and t are different:

$$\mathbf{e}(s,t) = \mathsf{min}\,\{n \in \mathbb{N} \mid \exists w \in A^*, \quad |w| = n \text{ and } \langle s,w\rangle \neq \langle t,w\rangle\} \ .$$

The gap is a generalisation of the notion of *valuation* of a series. The valuation $\mathbf{v}(s)$ of s in $\mathbb{K}\langle\langle A^*\rangle\rangle$ is defined by:

$$\mathbf{v}(s) = \mathbf{e}(s,0) = \mathsf{min}\,\{|w| \mid \langle s,w\rangle \neq 0\} = \mathsf{min}\,\{|w| \mid w \in \mathsf{supp}\ s\} \ .$$

Conversely, and if $\mathbb{K}$ is a *ring*, $\mathbf{e}(s,t) = \mathbf{v}(s-t)$.

(e) Show that the map defined by

$$\forall s,t \in \mathbb{K}\langle\langle A^* \rangle\rangle \qquad \mathbf{d}'(s,t) = 2^{-\mathbf{e}(s,t)} \tag{4.1}$$

is an ultrametric distance on $\mathbb{K}\langle\langle A^* \rangle\rangle$, bounded by 1.

(f) Let $\mathbf{c}$ be a distance on $\mathbb{K}\langle\langle A^* \rangle\rangle$, bounded by 1. Show that the map defined by

$$\forall s,t \in \mathbb{K}\langle\langle A^* \rangle\rangle \qquad \mathbf{d}(s,t) = \frac{1}{2} \sum_{n \in \mathbb{N}} \left(\frac{1}{2^n} \mathsf{max} \{ \mathbf{c}(\langle s,w \rangle, \langle t,w \rangle) \mid |w| = n \} \right) \tag{4.2}$$

is a distance on $\mathbb{K}\langle\langle A^* \rangle\rangle$, bounded by 1.

(g) Show that, whatever the distance $\mathbf{c}$, $\mathbf{d}(s,t) \leqslant \mathbf{d}'(s,t)$ holds.

(h) Show that if $\mathbf{c}$ is the discrete distance, then $\mathbf{d}'(s,t) \leqslant 2\,\mathbf{d}(s,t)$ holds, hence that (4.1) and (4.2) define two equivalent distances on $\mathbb{K}\langle\langle A^* \rangle\rangle$ if $\mathbb{K}$ is equipped with the discrete distance.

(i) Show that the topology defined by $\mathbf{d}$ on $\mathbb{K}\langle\langle A^* \rangle\rangle$ is the topology of the simple convergence.

(j) Show that if $\mathbb{K}$ is a topological semiring, then so are $\mathbb{K}^{Q \times Q}$ (Q finite) and $\mathbb{K}\langle\langle A^* \rangle\rangle$.

(k) Let $\big(s_n\big)_{n \in \mathbb{N}}$ and $\big(t_n\big)_{n \in \mathbb{N}}$ be two sequences of series in the topological semiring $\mathbb{K}\langle\langle A^* \rangle\rangle$. Verify that $\big(s_n + t_n\big)_{n \in \mathbb{N}}$ or $\big(s_n\, t_n\big)_{n \in \mathbb{N}}$ may be convergent sequences, without $\big(s_n\big)_{n \in \mathbb{N}}$ or $\big(t_n\big)_{n \in \mathbb{N}}$ being convergent sequences.

Lecture II

Morphisms of weighted automata

Conjugacy and minimal quotient

In this lecture, we address the problem of finding, given a $\mathbb{K}$-automaton $\mathcal{A}$, a $\mathbb{K}$-automaton $\mathcal{B}$, hopefully of smaller dimension than $\mathcal{A}$, and that inherits the *structure* of $\mathcal{A}$, that is, such that there is a correspondence between the *computations* of $\mathcal{A}$ and those of $\mathcal{B}$. This amounts to describing the *morphisms of* $\mathbb{K}$*-automata*, that is, the mappings between $\mathbb{K}$-automata that *preserve their structure*.

Contents

The classical notion of the minimal automaton of a language, minimal quotient of any determistic that accepts the language is at the same time an example of what we want to generalise and somewhat misleading. Already when it deals with non-deterministic (Boolean) automata, this generalisation requires a *lateralisation* which is not usually associated with the notion morphism and this may may explain it has been given the other name of *bisimulation* in the literature. We call it *Out-morphism* to stress the link with the notion of morphism.

We define Out-morphism in a naive way for non-deterministic Boolean automata and by means of the mathematical notion of *conjugacy* for the general case of weighted automata. This notion being set up, the same theory as the classical

one for complete deterministic automata can be rolled out and it is easily seen that every weighted automaton admits a *minimal quotient* as the image of the coarsest Out-morphism which is computed essentially by the same algorithm.

It is worth to be noted that, at least in the case of Boolean automata, the converse operation is indeed at least as interesting: given $\mathcal{A}$, build $\mathcal{B}$ of which $\mathcal{A}$ is a morphic image, hence larger than $\mathcal{A}$, but whose computations are less entangled, in such a way that it becomes possible, by means of other operations, to distinguish and make choices between these computations. Such constructions are essentially considered (even for general weighted automata) when the computations of $\mathcal{B}$ are in a 1-to-1 corespondence with those of $\mathcal{A}$, that is, when $\mathcal{B}$ is a *covering* of $\mathcal{A}$.

1 Morphisms of Boolean automata

This section is more than a reminder or an appetizer. It introduces at the end the notions of *local properties* of morphisms, that will be instrumental in the study of transducers.

In this section, all automata are Boolean automata. We begin with the presentation of the classical definition and computation of the minimal automaton of a rational language while insisting on the morphism point of view.

1.1 The case of (complete) deterministic automata

A deterministic automaton is denoted by $\mathcal{A} = \langle\, A, Q, i, \delta, T \,\rangle$ rather than by $\mathcal{A} = \langle\, A, Q, I, E, T \,\rangle$, where δ is the *transition function*, that is, a map $\delta\colon Q \times A \to Q$. For every w in A^* and p in Q, we write $p \cdot w = q$ rather than $\delta(p, w) = q$. Since $(p \cdot u) \cdot v = p \cdot u\,v$ the transition function δ defines an *action* of A^* over Q.

The *minimal automaton* of a language L of A^* is defined by means of the *quotient* operation that anticipate the notion of *quotient of a series* (*cf.* Definition III.13): if u is in A^*, the *(left) quotient of L by u* is the language $u^{-1}L = \{v \in A^* \mid u\,v \in L\}$. Let $\mathbf{R}_L$ be the set of quotients of L: $\mathbf{R}_L = \{u^{-1}L \mid u \in A^*\}$; $\mathbf{R}_L$ is finite if and only if L is a rational language.

Since $(u\,v)^{-1}L = v^{-1}(u^{-1}L)$, the (left) quotient is a (right) action of A^* over the set of languages $\mathfrak{P}(A^*)$, which in turn defines a deterministic automaton on any set of languages closed by quotient.

For every rational language L, let us denote by $\mathcal{A}_L$ the finite deterministic automaton $\mathcal{A}_L = \langle\, A, \mathbf{R}_L, \{L\}, \triangleright, T_L \,\rangle$, where $\triangleright$ is another notation for the quotient:

$$L \triangleright u = u^{-1}L \qquad \text{and} \qquad T_L = \left\{ u^{-1}L \;\middle|\; 1_{A^*} \in u^{-1}L \right\} \ .$$

The automaton $\mathcal{A}_L$ accepts L and is called the *minimal automaton* of L, a terminology that is justified by the following.

Let $\mathcal{A} = \langle A, Q, i, \delta, T \rangle$ be a complete deterministic accessible automaton and $L = L(\mathcal{A})$ the language that it accepts. For all p in Q, we write L_p for the language accepted by the automaton obtained from $\mathcal{A}$ by replacing the initial state i by p:

$$L_p = L(\langle A, Q, p, \delta, T \rangle) = \{w \in A^* \mid p \cdot w \in T\} \quad .$$

Definition 1. The *Nerode equivalence* is the relation ν defined on Q by

$$p \equiv q \mod \nu \iff L_p = L_q \quad .$$

Proposition 2. *The Nerode equivalence induces a map* $\varphi\colon Q \to Q/\nu$ *which saturates* T *and such that* $\varphi(p \cdot a) = (\varphi(p)) \cdot a$.

Proposition 2 allows to define a quotient automaton $\mathcal{A}/\nu = \langle A, Q/\nu, [i]_\nu, \delta_\nu, T_\nu \rangle$.

Theorem 3. $\mathcal{A}_L = \mathcal{A}/\nu$.

Theorem 3 tells at the same time that $\mathcal{A}_L$ is the quotient of *every* complete deterministic automaton that accepts L and that it is the complete deterministic automaton that accepts L with the minimal number of states.

Example 4. Figure 1 shows a complete deterministic automaton and its minimal quotient, obtained by merging states.

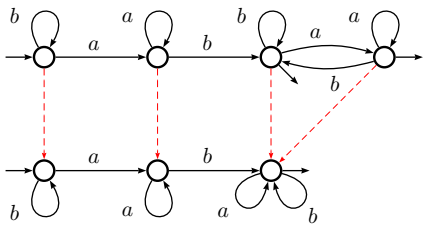


Figure 1: A complete deterministic automaton and its minimal quotient

Proposition 5. *The Nerode equivalence of a finite deterministic automaton is effectively computable by a partition refinement algorithm.*

1.2 The case of general (Boolean) automata

The first definition of morphism for (Boolean) automata follows naturally from the one for deterministic ones. It appears however that it has to be strengthened in order to give rise to the notion of minimal quotient. This new definition of *Out-morphism*, similar to the one of *simulation* for transition systems, applies to any Boolean automaton but is *lateralised* (or *directed*). It is described more systematically in the next subsection.

For the rest of this section, the alphabet is A and fixed, and $\mathcal{A} = \langle Q, I, E, T \rangle$ and $\mathcal{B} = \langle R, J, F, U \rangle$ are two Boolean automata.

Definition 6. A map $\varphi\colon Q \to R$ is a *morphism (of automata)* if:

(i) $\varphi(I) \subseteq J$,

(ii) $\varphi(T) \subseteq U$, and

(iii) for every transition (p, a, q) in E, $(\varphi(p), a, \varphi(q))$ is a transition in F.

If φ is such a morphism, we write $\varphi\colon \mathcal{A} \to \mathcal{B}$.

Proposition 7. *If $\varphi\colon \mathcal{A} \to \mathcal{B}$ is a morphism, then $\big|\mathcal{A}\big| \subseteq \big|\mathcal{B}\big|$.*

The notion of morphism is somewhat weak as shown by the next example.

Example 8. (i) If $\mathcal{U}$ is the one-state automaton which accepts the whole A^*, then the map which sends all states of any automaton $\mathcal{A}$ on the unique state of $\mathcal{U}$ is a morphism.

(ii) If $\mathcal{C} = \mathcal{A} \times \mathcal{B}$, then both projections $\pi_{\mathcal{A}}\colon \mathcal{C} \to \mathcal{A}$ and $\pi_{\mathcal{B}}\colon \mathcal{C} \to \mathcal{B}$ are morphisms.

The reason for the inclusion in Proposition 7 be strict is that not every (successful) computation in $\mathcal{B}$ may be *lifted* into a (successful) computation in $\mathcal{A}$: the morphism φ is said not to be *conformal*. The two sorts of morphisms in Example 8 are not conformal. Figure 2 gives another example of a non-conformal morphism. It shows that the inclusion in Proposition 7 may be strict, even when φ induces a *bijection* between the transitions — which is the strongest possible condition besides being the identity.

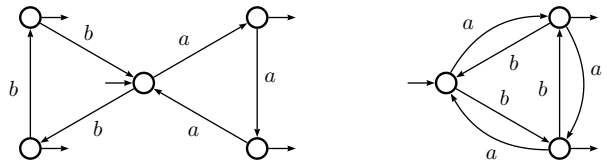


Figure 2: A non-conformal morphism (the morphism is the horizontal projection)

Example 8 (i) shows how weak the notion of automaton morphism can be. In order to have morphisms which really preserve the *structure* of automata (which is supposed to be the role of morphisms) we consider morphisms which meet additional conditions. We first do it 'directly'; in the next subsection, we introduce the more general notion of *local properties* of morphisms that allows to define a richer variety of morphisms.

Definition 9. A map $\varphi\colon Q \to R$ is an *Out-morphism* if:

(o) $\varphi(Q) = R$, that is, if φ is *surjective*,

(i) $\varphi(I) = J$,

(ii) $T = \varphi^{-1}(U)$,

(iii) for every transition (p, a, q) in E, $(\varphi(p), a, \varphi(q))$ is a transition in F,

(iv) for every transition (r, a, s) in F and every p in $\varphi^{-1}(r)$, there exists a q in $\varphi^{-1}(s)$ such that (p, a, q) is a transition in E.

Remark 10. The notion of Out-morphism is *directed* since condition (iv), which consists in a succession of two quantifiers: 'for all..., there exists...', breaks the symmetry between the origin and the destination of the transitions.

Examples 11. (i) If $\mathcal{B}$ is complete and with every state being final, then $\pi_{\mathcal{A}} \colon \mathcal{A} \times \mathcal{B} \to \mathcal{A}$ is an Out-morphism.

(ii) A morphism from a complete deterministic automaton onto an accessible deterministic automaton is an Out-morphism.

Definition 12.
An automaton $\mathcal{B}$ is *a quotient* of $\mathcal{A}$ if there exists an Out-morphism $\varphi \colon \mathcal{A} \to \mathcal{B}$.

Remark 13. The terminology does not make it so clear, but the notion of quotient is *directed* as it derives from the one of Out-morphism. It means somehow that the *true* morphisms are the Out-morphisms.

Proposition 14. *If $\mathcal{B}$ is a quotient of $\mathcal{A}$, then every (successful) computation $\mathcal{B}$ can be lifted into a (successful) computation $\mathcal{A}$.*

Corollary 15. *If $\mathcal{B}$ is a quotient of $\mathcal{A}$, then $\big|\mathcal{A}\big| = \big|\mathcal{B}\big|$.*

These two statements show that we have reached our goal with the notion of Out-morphism. In order to avoid repetition, we postpone to after the definition of local properties of morphisms and a new expression of Out-morphisms, the presentation of results attached to the notion of quotient.

1.3 Local properties of morphisms

We now take more precise definitions for characterising morphisms; we first set up a convention that reduces the notion of automaton morphism to that of a labelled graph morphism (and get rid of conditions (i) and (ii) in Definitions 6 and 9).

1.3.1 Subliminal states

With every automaton $\mathcal{A} = \langle\, Q, I, E, T \,\rangle$, we associate, by a sort of normalisation, an automaton $\mathcal{A}_{\mathsf{n}}$ to which we have added two new states — $i_{\mathcal{A}}$, an initial state, and $t_{\mathcal{A}}$, a final state — and some transitions, *labelled with* 1_{A^*}, which go from $i_{\mathcal{A}}$ to each initial state of $\mathcal{A}$ and from each final state of $\mathcal{A}$ to $t_{\mathcal{A}}$:

$$\mathcal{A}_{\mathsf{n}} = \langle\, Q \cup \{i_{\mathcal{A}}, t_{\mathcal{A}}\}, i_{\mathcal{A}}, E_{\mathsf{n}}, t_{\mathcal{A}} \,\rangle \ ,$$
$$E_{\mathsf{n}} = E \cup \ \{(i_{\mathcal{A}}, 1_{A^*}, i) \mid i \in I\} \ \cup \ \{(t, 1_{A^*}, t_{\mathcal{A}}) \mid t \in T\} \ .$$

These two new states, $i_{\mathcal{A}}$ and $t_{\mathcal{A}}$, are called the (initial and final) *subliminal* states of $\mathcal{A}$. We verify easily that $\mathcal{A}_{\mathsf{n}}$ is equivalent to $\mathcal{A}$. More precisely, there is a bijection between the computations of $\mathcal{A}_{\mathsf{n}}$ and those of $\mathcal{A}$ and, of course, the computations that correspond in this bijection have the same label.

Remark 16. Even though we deal here with Boolean automata only, the definition of $\mathcal{A}_{\mathsf{n}}$ may seem to imply a drastic change in the model of (finite) automata since it allows the empty word to be the label of a transition, transitions that are then called *spontaneous transitions* (or ε-transitions). In full generality, this feature opens the possibility for a word to be the label of an infinite number of computations and raises the (difficult) problem of the *validity* when it comes to weighted automata, a problem which will not be treated in these notes. However, if there is *no circuit* of spontaneous transitions in the automaton, then every word is still the label of a *finite number* of computations, its weight can be computed by the sum in Equation (I.1.1) and the behaviour of the automaton is well-defined. Clearly, the construction of $\mathcal{A}_{\mathsf{n}}$ fall in this case where no circuit of spontaneous transitions is created.

If φ is a map from $\mathcal{A}$ to $\mathcal{B}$, we extend it to a map φ_{n} from $\mathcal{A}_{\mathsf{n}}$ to $\mathcal{B}_{\mathsf{n}}$ by taking $\varphi_{\mathsf{n}}(i_{\mathcal{A}}) = i_{\mathcal{B}}$ and $\varphi_{\mathsf{n}}(t_{\mathcal{A}}) = t_{\mathcal{B}}$. We then verify, just as easily, that $\varphi\colon \mathcal{A} \to \mathcal{B}$ is an automaton morphism if and only if $\varphi_{\mathsf{n}}\colon \mathcal{A}_{\mathsf{n}} \to \mathcal{B}_{\mathsf{n}}$ is a labelled graph morphism.

1.3.2 Outgoing and incoming bouquets

For every state p of $\mathcal{A} = \langle Q, I, E, T \rangle$, we denote by $\mathsf{Out}_{\mathcal{A}}(p)$ the set of transitions in $\mathcal{A}$ *outgoing* from p and by $\mathsf{In}_{\mathcal{A}}(p)$ the set of transitions *arriving* at p:

$$\mathsf{Out}_{\mathcal{A}}(p) = \{e \in E \mid e = (p, a, q)\}\ , \qquad \mathsf{In}_{\mathcal{A}}(p) = \{e \in E \mid e = (q, a, p)\}\ ,$$

and we call these sets the *outgoing bouquet* and the *incoming bouquet at state* p respectively (the automaton $\mathcal{A}$ being understood). These notions are *directed*, of course, and *dual*, that is, $\mathsf{In}_{\mathcal{A}}(p) = \mathsf{Out}_{{}^{\mathsf{t}}\mathcal{A}}(p)$ for every p in Q (with the slight abuse which consists in considering that $\mathcal{A}$ and ${}^{\mathsf{t}}\mathcal{A}$ have the *same* set of transitions). The purpose of the definition of these bouquets is the description of morphism properties based on the remark that if $\varphi\colon \mathcal{A} \to \mathcal{B}$ is a morphism, then, for every p in Q, φ maps $\mathsf{Out}_{\mathcal{A}}(p)$ into $\mathsf{Out}_{\mathcal{B}}(\varphi(p))$ and $\mathsf{In}_{\mathcal{A}}(p)$ into $\mathsf{In}_{\mathcal{B}}(\varphi(p))$.

Definition 17. A morphism $\varphi\colon \mathcal{A} \to \mathcal{B}$ is *Out-surjective* (resp. Out-injective, Out-bijective*) if, for every state p of $\mathcal{A}_{\mathsf{n}}$, the restriction of φ to $\mathsf{Out}_{\mathcal{A}}(p)$ is a surjective (resp. injective, bijective) map into $\mathsf{Out}_{\mathcal{B}}(\varphi(p))$.*

The morphism φ is In-surjective *(resp.* In-injective, In-bijective*) if, for every state p of $\mathcal{A}_{\mathsf{n}}$, the restriction of φ to $\mathsf{In}_{\mathcal{A}}(p)$ is a surjective (resp. injective, bijective) map into $\mathsf{In}_{\mathcal{B}}(\varphi(p))$.*

The 'Out-properties' and the corresponding 'In-properties' are *dual properties*, that is, if $\varphi\colon \mathcal{A} \to \mathcal{B}$ is Out-surjective (resp. Out-injective, Out-bijective) then $\varphi\colon {}^{\mathsf{t}}\mathcal{A} \to {}^{\mathsf{t}}\mathcal{B}$ is In-surjective (resp. In-injective, In-bijective).

Remark 18. Condition (iv) of Definition 9 is another way to express that the morphism φ is Out-surjective.

Remark 19. The conditions of Definition 17 on the outgoing bouquets of the *subliminal initial states* imply that if $\varphi\colon \mathcal{A} \to \mathcal{B}$ is *Out-surjective*, then $\varphi(I) = J$ (condition (i) of Definition 9). Considering the outgoing bouquets of the terminal states — and the transitions toward *subliminal final states* imply that if φ is *Out-surjective*, then $T = \varphi^{-1}(U)$ (condition (ii) of Definition 9). Similarly, if φ is *Out-injective*, then, for every j in J, there exists at most *one* i in I such that $\varphi(i) = j$.

In a dual way, if $\varphi\colon \mathcal{A} \to \mathcal{B}$ is *In-surjective*, then $\varphi(T) = U$ and $I = \varphi^{-1}(J)$ and if φ is *In-injective*, then for every u in U there exists at most *one* t in T such that $\varphi(t) = u$.

Out-surjective morphisms are *conformal*, as expressed by the following statement which is easily verified by induction on the length of paths.

Proposition 20. *Let $\varphi\colon \mathcal{A} \to \mathcal{B}$ be an Out-surjective morphism. For every path d in $\mathcal{B}$ whose source s is in the image of φ and for every p such that $\varphi(p) = s$ there exists at least one path c in $\mathcal{A}$ whose source is p and such that $\varphi(c) = d$.* ∎

Corollary 21. *If $\varphi\colon \mathcal{A} \to \mathcal{B}$ is an Out-surjective morphism, then $|\mathcal{A}| = |\mathcal{B}|$.*

Corollary 22. *If $\varphi\colon \mathcal{A} \to \mathcal{B}$ is an Out-bijective morphism, then φ is a* bijection *between the* successful computations *of $\mathcal{A}$ and those of $\mathcal{B}$.*

Corollary 23. *If $\varphi\colon \mathcal{A} \to \mathcal{B}$ is an Out-surjective morphism and if $\mathcal{B}$ is* accessible *then φ is (globally)* surjective.

1.3.3 Out- and In-morphisms revisited

With Corollary 23, we see that Out-surjective morphisms are 'almost always' surjective (condition (o) of Definition 9). For simplification and conciseness, and in order to avoid special cases, we take that latter property as an hypothesis and set up the following definitions.

Definition 24. A surjective Out-surjective morphism is called an *Out-morphism.*
A surjective In-surjective morphism is called an *In-morphism.*
A surjective Out-bijective morphism is called a *covering.*[1]
A surjective In-bijective morphism is called a *co-covering.*[2]
A surjective Out-injective morphism is called an *immersion.*
A surjective In-injective morphism is called a *co-immersion.*

Remarks 18 and 19 show that Definition 9 and the definition above coincide (for Out-morphisms). We then repeat Definition 12 and Proposition 14.

[1] In French, *revêtement.*
[2] In French, *co-revêtement.*

Definition 25.
An automaton $\mathcal{B}$ is a *quotient* of $\mathcal{A}$ if there exists an Out-morphism $\varphi\colon \mathcal{A} \to \mathcal{B}$.
An automaton $\mathcal{B}$ is a *co-quotient* of $\mathcal{A}$ if there exists an In-morphism $\varphi\colon \mathcal{A} \to \mathcal{B}$.

The automaton $\mathcal{B}$ is *a co-quotient* of $\mathcal{A}$ if ${}^{\mathsf{t}}\mathcal{B}$ is a quotient of ${}^{\mathsf{t}}\mathcal{A}$.

Proposition 26. *If $\mathcal{B}$ is a quotient (resp. a co-quotient) of $\mathcal{A}$, then every (successful) computation $\mathcal{B}$ can be lifted onto a (successful) computation $\mathcal{A}$.*

Remark 27. Proposition 26 implies that if $\mathcal{B}$ is a quotient of $\mathcal{A}$, then $\mathcal{A}$ is a *simulation* of $\mathcal{B}$. The terminology of simulation is very common in several areas close to automata theory but using a different vocabulary (transition systems, coalgebra, *etc.*). Note that the converse statement (if $\mathcal{A}$ is a simulation of $\mathcal{B}$, then $\mathcal{B}$ is a quotient of $\mathcal{A}$) does not hold. See Proposition 31 below.

The notion of quotient allows to extend the one of minimal automata.

Proposition 28. *Every automaton $\mathcal{A}$ has a* minimal quotient *$\mathcal{C}$, which is unique up to an isomorphism, and which is the quotient of any quotient $\mathcal{B}$ of $\mathcal{A}$.*

Remark 29. The minimal quotient of an automaton *is not canonically attached to the accepted language* anymore but depends on the automaton it is computed from.

The dual of Proposition 28 also holds.

Proposition 30. *Every automaton $\mathcal{A}$ has a* minimal co-quotient *$\mathcal{D}$, which is unique up to an isomorphism, and which is the co-quotient of any co-quotient $\mathcal{B}$ of $\mathcal{A}$.*

The minimal quotient or co-quotient of an automaton can be computed by a kind of Moore algorithm that consists in successive refinements of the trivial partition on the set of states. We come back to this question at Section 2.3 in the more genral setting of weighted automata.

Finally, let us note that the notion of quotient allows to give a clean definition of *bisimulation*. (We give it as a statement, assuming that the definition of bisimilarity has been elsewhere.)

Proposition 31. *Two automata $\mathcal{A}$ and $\mathcal{B}$ are bisimilar if and only if they have the same (or isomorphic) minimal quotient.*

1.4 The Schützenberger covering

We begin with an elementary statement.

Proposition 32. *Let $\mathcal{A}$ be an accessible automaton, $\mathcal{B}$ a complete deterministic automaton equivalent to $\mathcal{A}$, and $\mathcal{E}$ the accessible part of $\mathcal{B} \times \mathcal{A}$. Then $\pi_{\mathcal{A}}$, the projection of $\mathcal{B} \times \mathcal{A}$ onto $\mathcal{A}$, is a covering from $\mathcal{E}$ to $\mathcal{A}$.* ∎

Definition 33. Let $\mathcal{A}$ be an accessible automaton and $\widehat{\mathcal{A}}$ its determinisation. The *Schützenberger covering*, or *S-covering*, of $\mathcal{A}$ is the accessible part $\mathcal{S}$ of $\widehat{\mathcal{A}}\times\mathcal{A}$.

Theorem 34. *Let $\mathcal{A}$ be an accessible automaton and $\mathcal{S}$ its Schützenberger covering. Then $\mathcal{S}$ satisfies:*

(i) *$\pi_{\mathcal{A}}$ is a covering from $\mathcal{S}$ to $\mathcal{A}$;*

(ii) *$\pi_{\widehat{\mathcal{A}}}$ is an In-morphism from $\mathcal{S}$ to $\widehat{\mathcal{A}}$.*

Example 35. Figure 3 shows the S-covering of the automaton $\mathcal{A}_1$,

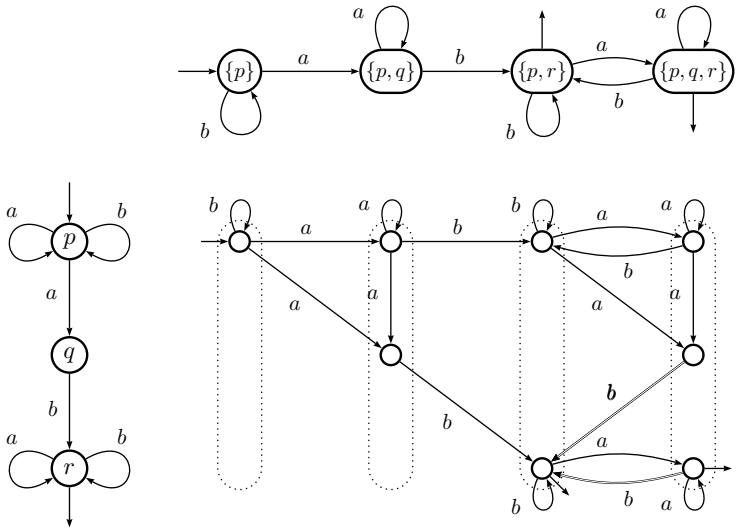


Figure 3: The S-covering of $\mathcal{A}_1$

Proof of Theorem 34. Since $\widehat{\mathcal{A}}$ is a complete deterministic automaton equivalent to $\mathcal{A}$, condition (i) is the instance of Proposition 32 for $\mathcal{B} = \widehat{\mathcal{A}}$ and it remains to prove condition (ii). From the definition of transitions in $\widehat{\mathcal{A}} = \langle\, \mathfrak{P}(Q), \{I\}, F, U \,\rangle$, namely,

$$P \xrightarrow[\widehat{\mathcal{A}}]{a} S \quad \Longleftrightarrow \quad S = \left\{ q \;\middle|\; \exists p \in P \quad p \xrightarrow[\mathcal{A}]{a} q \right\} , \tag{1.1}$$

we first deduce:

Property 36. *The states of $\mathcal{S}$ are the pairs (P, p) where P is a state of $\widehat{\mathcal{A}}$ and p is in P.*

Proof. Let P be a state of $\widehat{\mathcal{A}}$: that is, there exists w in A^* such that

$$P = \left\{ p \;\middle|\; \exists i \in I \quad i \xrightarrow[\mathcal{A}]{w} p \right\} .$$

Thus (P, p) is a state of $\mathcal{S}$; that is, it is accessible in $\widehat{\mathcal{A}}\times\mathcal{A}$ for all p in P. Conversely, if (P, q) is a state of $\mathcal{S}$, there exists w in A^* and i in I such that both $\{I\} \xrightarrow[\widehat{\mathcal{A}}]{w} P$ and $i \xrightarrow[\mathcal{A}]{w} q$, and hence q is in P. ∎

We next deduce by (1.1) that

$$\forall P, S \subseteq Q\,,\ \forall q \in S\,,\ \forall a \in A \qquad P \xrightarrow[\widehat{\mathcal{A}}]{a} S \implies \exists p \in P \qquad p \xrightarrow[\mathcal{A}]{a} q$$
$$\implies \exists p \in P \qquad (P,p) \xrightarrow[\widehat{\mathcal{A}}\times\mathcal{A}]{a} (S,q)$$

since (P,p) is a state of $\mathcal{S}$, which indeed means that $\pi_{\widehat{\mathcal{A}}}\colon \mathcal{S} \to \widehat{\mathcal{A}}$ is an In-surjective labelled graph morphism.

If $P \subseteq Q$ is final in $\widehat{\mathcal{A}}$ there exists at least one t in P which is final in $\mathcal{A}$, hence a state (P,t) which is final in $\mathcal{S}$. On the other hand, I is the unique initial state of $\widehat{\mathcal{A}}$, every i in I is initial in $\mathcal{A}$, hence every state (I,i) is initial in $\mathcal{S}$. Altogether, $\pi_{\widehat{\mathcal{A}}}$ is an In-surjective morphism. ∎

Corollary 37. *For every Boolean automaton $\mathcal{A}$, there exists an automaton $\mathcal{T}$ such that:*

(i) *$\mathcal{T}$ is equivalent to $\mathcal{A}$;*
(ii) *$\mathcal{T}$ is unambiguous;*
(iii) *there exists a morphism $\varphi\colon \mathcal{T} \to \mathcal{A}$.*

Proof. Let $\mathcal{S}$ be the Schützenberger covering of $\mathcal{A}$ and $\pi_{\widehat{\mathcal{A}}}\colon \mathcal{S} \to \widehat{\mathcal{A}}$ the projection of $\mathcal{S}$ on $\widehat{\mathcal{A}}$. Since $\pi_{\widehat{\mathcal{A}}}$ is In-surjective, it is possible, by deleting some transitions in $\mathcal{S}$, to obtain a sub-automaton $\mathcal{T}$ of $\mathcal{S}$ such that $\pi_{\widehat{\mathcal{A}}}\colon \mathcal{T} \to \widehat{\mathcal{A}}$ be *In-bijective*. The projection $\pi_{\widehat{\mathcal{A}}}$ yields a bijection between the successful computations of $\widehat{\mathcal{A}}$ and those of $\mathcal{T}$, hence $\mathcal{T}$ is equivalent to $\widehat{\mathcal{A}}$ and then to $\mathcal{A}$ and $\mathcal{T}$ is unambiguous since so is $\widehat{\mathcal{A}}$. The restriction to $\mathcal{T}$ of the projection $\pi_{\mathcal{A}}$ is a morphism. ∎

It is not a new result that given an automaton $\mathcal{A}$, it is possible to find an unambiguous automaton $\mathcal{T}$ equivalent to $\mathcal{A}$: the determinisation of $\mathcal{A}$ for instance answers the question. That $\mathcal{T}$ can be chosen in a way there is a morphism from $\mathcal{T}$ to $\mathcal{A}$, that is, one can see in $\mathcal{A}$ the computations of $\mathcal{T}$ is a new, and far reaching, property.

2 Morphisms of weighted automata

After the definition of any structure one looks for *morphisms* between objects of that structure, and weighted automata are no exception. Moreover, morphisms of graphs, and therefore of classical Boolean automata, are not less classical, and one waits for their generalisation to weighted automata. Taking into account multiplicity proves however to be not so simple. In the sequel, all automata are supposed to be of finite dimension.[3]

[3] May be it should have been mentioned that the matter developed in Section 1 did not require the automata be finite.

We choose to describe the morphisms of weighted automata via the notion of *conjugacy*, borrowed from the theory of symbolic dynamical systems.

2.1 Conjugacy

Definition 38. A $\mathbb{K}$-automaton $\mathcal{A} = \langle I, E, T \rangle$ *is conjugate to* a $\mathbb{K}$-automaton $\mathcal{B} = \langle J, F, U \rangle$ if there exists a matrix X with entries in $\mathbb{K}$ such that

$$I\,X = J, \qquad E\,X = X\,F, \quad \text{and} \quad T = X\,U.$$

The matrix X is the *transfer matrix* of the conjugacy and we write $\mathcal{A} \overset{X}{\Longrightarrow} \mathcal{B}$.

If $\mathcal{A}$ is conjugate to $\mathcal{B}$, then, for every n, the series of equalities holds:

$$I\,E^n\,T = I\,E^n\,X\,U = I\,E^{n-1}\,X\,F\,U = \ldots = I\,X\,F^n\,U = J\,F^n\,U \ ,$$

from which the following is directly deduced.

Proposition 39. *If $\mathcal{A}$ is conjugate to $\mathcal{B}$, then $\mathcal{A}$ and $\mathcal{B}$ are equivalent.*

Example 40. It is easily checked that the $\mathbb{Z}$-automaton $\mathcal{Y}_1$ of Figure 4 is conjugate to the $\mathbb{Z}$-automaton $\mathcal{Z}_1$ of the same figure with the transfer matrix X_1:

$$X_1 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \ .$$

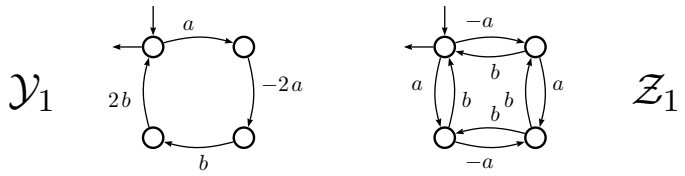


Figure 4: Two conjugate $\mathbb{Z}$-automata

In spite of the idea conveyed by the terminology, the conjugacy relation *is not an equivalence* but a *preorder* relation. Suppose that $\mathcal{A} \overset{X}{\Longrightarrow} \mathcal{C}$ holds; if $\mathcal{C} \overset{Y}{\Longrightarrow} \mathcal{B}$ then $\mathcal{A} \overset{XY}{\Longrightarrow} \mathcal{B}$, but if $\mathcal{B} \overset{Y}{\Longrightarrow} \mathcal{C}$ then $\mathcal{A}$ is not necessarily conjugate to $\mathcal{B}$, and we write $\mathcal{A} \overset{X}{\Longrightarrow} \mathcal{C} \overset{Y}{\Longleftarrow} \mathcal{B}$ or even $\mathcal{A} \overset{X}{\Longrightarrow} \overset{Y}{\Longleftarrow} \mathcal{B}$. This being well understood, we shall speak of "conjugate automata" when the direction does not matter.

If $\mathcal{A} = \langle I, E, T \rangle$ is conjugate to $\mathcal{B} = \langle J, F, U \rangle$ then the same conjugacy relation holds between the matrices of the corresponding representations, that is, if $\mathcal{A} = (I, \mu, T)$ and $\mathcal{B} = (J, \kappa, U)$, then, as above, $I\,X = J$, $T = X\,U$, and

$$\forall a \in A \qquad \mu(a)\,X = X\,\kappa(a) \ . \tag{2.1}$$

Then, the same conjugacy relation holds for the representations of every word, that is:

$$\forall w \in A^* \qquad \mu(w)\,X = X\,\kappa(w) \ . \tag{2.2}$$

2.2 Out-morphisms, In-morphisms

Let $\varphi\colon Q \to R$ be a *surjective* map and X_φ the $Q \times R$-matrix where the (q,r)-th entry is 1 if $\varphi(q) = r$, and 0 otherwise. Since φ is a map, every row of X_φ contains exactly one 1 and since φ is surjective, every column of X_φ contains at least one 1. Such a matrix is called an *amalgamation matrix* in the setting of symbolic dynamics.

Definition 41. Let $\mathcal{A}$ and $\mathcal{B}$ be two $\mathbb{K}$-automata of dimension Q and R respectively. We say that a surjective map $\varphi\colon Q \to R$ is an *Out-morphism* (from $\mathcal{A}$ onto $\mathcal{B}$) if $\mathcal{A}$ is conjugate to $\mathcal{B}$ by X_φ, that is, if $\mathcal{A} \overset{X_\varphi}{\Longrightarrow} \mathcal{B}$, and we write $\varphi\colon \mathcal{A} \to \mathcal{B}$.

We also say that $\mathcal{B}$ is a *quotient* of $\mathcal{A}$, if there exists an Out-morphism $\varphi\colon \mathcal{A} \to \mathcal{B}$.

Remark 42. If $\mathbb{K} = \mathbb{B}$, then Definition 41 coincide with Definition 9.

Again, the notions of Out-morphism and quotient are *lateralised*, or *directed*, since the conjugacy relation is not symmetric. Stated otherwise, and as we see with Proposition 47, it is directed in that it refers not to the transitions of the automaton but to the *outgoing* transitions from the states of the automaton. We then define the *dual* notions of *In-morphism* and *co-quotient*.

Definition 43. With the notation above, a surjective map $\varphi\colon Q \to R$ is an *In-morphism* (from $\mathcal{A}$ onto $\mathcal{B}$) if $\mathcal{B}$ is conjugate to $\mathcal{A}$ by ${}^{\mathsf{t}}X_\varphi$, that is, if $\mathcal{B} \overset{{}^{\mathsf{t}}X_\varphi}{\Longrightarrow} \mathcal{A}$, and we write again $\varphi\colon \mathcal{A} \to \mathcal{B}$.

We say that $\mathcal{B}$ is a *co-quotient* of $\mathcal{A}$, if there exists an In-morphism $\varphi\colon \mathcal{A} \to \mathcal{B}$.

Example 44. Let $\mathcal{C}_2$ be the $\mathbb{N}$-automaton of Figure I.3 and φ_2 the map from $\{j,r,s,u\}$ to $\{i,q,t\}$ such that $j\varphi_2 = i$, $u\varphi_2 = t$ and $r\varphi_2 = s\varphi_2 = q$, then

$$X_{\varphi_2} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

and φ_2 is an Out-morphism from $\mathcal{C}_2$ onto $\mathcal{V}_2$ and an In-morphism from $\mathcal{C}_2$ onto $\mathcal{V}_2'$.

In contrast with this special example, a map $\varphi\colon Q \to R$ is not usually both an Out- and an In-morphism. When necessary we shall write $\varphi\colon \mathcal{A} \overset{\mathsf{Out}}{\to} \mathcal{B}$ and $\varphi\colon \mathcal{A} \overset{\mathsf{In}}{\to} \mathcal{B}$ in order to specify, or to distinguish between, the case.

It directly follows from Definitions 41 and 43 that if $\varphi\colon \mathcal{A} \overset{\mathsf{Out}}{\to} \mathcal{B}$ and $\psi\colon \mathcal{A} \overset{\mathsf{In}}{\to} \mathcal{C}$ are an Out- and an In-morphism respectively, then

$$\mathcal{C} \overset{{}^{\mathsf{t}}X_\psi}{\Longrightarrow} \mathcal{A} \overset{X_\varphi}{\Longrightarrow} \mathcal{B} \qquad \text{hence} \qquad \mathcal{C} \overset{{}^{\mathsf{t}}X_\psi X_\varphi}{\Longrightarrow} \mathcal{B} \ . \tag{2.3}$$

For instance, it holds:

$$\mathcal{V}_2' \overset{\left(\begin{smallmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{smallmatrix}\right)}{\Longrightarrow} \mathcal{V}_2 \ .$$

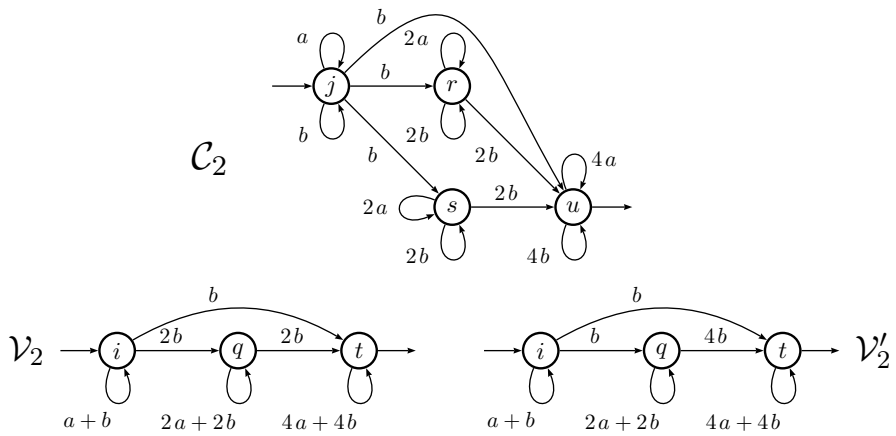


Figure 5: $\mathcal{V}_2$ is a quotient and $\mathcal{V}_2'$ a co-quotient of $\mathcal{C}_2$

The problem of establishing a converse to the implication expressed in (2.3), that is, proving that if two automata $\mathcal{B}$ and $\mathcal{C}$ are conjugate then there exists an automaton $\mathcal{A}$ such that $\mathcal{B}$ is a quotient of $\mathcal{A}$ and $\mathcal{C}$ a co-quotient of $\mathcal{A}$ is out of the scope of these lecture notes (the answer is indeed somewhat more complex). But we can at lest state the following.

Theorem 45. *Let $\mathbb{K} = \mathbb{B}$ or $\mathbb{N}$. If $\mathcal{A}$ and $\mathcal{B}$ are equivalent $\mathbb{K}$-automata, then there exists a $\mathbb{K}$-automaton $\mathcal{C}$ such that $\mathcal{A}$ is a quotient of $\mathcal{C}$ and $\mathcal{B}$ a co-quotient of $\mathcal{C}$.*

For instance, if $\mathcal{A}$ is a Boolean automaton and $\mathcal{B} = \widehat{\mathcal{A}}$, the Schützenberger covering is the automaton $\mathcal{C}$ the existence of which is insured by the theorem.

Remark 46. The entries of the amalgamation matrix X_φ are $0_\mathbb{K}$ or $1_\mathbb{K}$, hence belong to the center of $\mathbb{K}$ and from (2.1) follows that if φ is an In-morphism from $\mathcal{A}$ to $\mathcal{B}$ it holds

$$\forall a \in A \qquad \kappa(a)\ {}^{\mathsf{t}}X_\varphi = {}^{\mathsf{t}}X_\varphi\, \mu(a) \qquad \text{and then} \qquad {}^{\mathsf{t}}\mu(a)\ X_\varphi = X_\varphi\ {}^{\mathsf{t}}\kappa(a) \quad ,$$

which means that in the case where we could speak of the transpose of an automaton, hence essentially when $\mathbb{K}$ is commutative, φ is an In-morphism from $\mathcal{A}$ to $\mathcal{B}$ if φ is an Out-morphism from ${}^{\mathsf{t}}\mathcal{A}$ to ${}^{\mathsf{t}}\mathcal{B}$. This statement makes appear more clearly that In-morphism is the dual notion of Out-morphism. Our definition has the advantage that it does not depend on the one of the transpose of an automaton.

It is to be noted that in the definition of an Out-morphism $\varphi\colon \mathcal{A} \to \mathcal{B}$, *the image is immaterial* and only counts the map equivalence of φ — which is sufficient to determine the matrix X_φ. From any amalgamation matrix X_φ, we construct a matrix Y_φ by transposing X_φ and by cancelling certain of its entries in such a way that Y_φ is row monomial (with exactly one 1 per row); Y_φ is not uniquely determined by φ but also depends on the choice of a 'representative' in each class for the map equivalence of φ. Whatever this choice, the product $Y_\varphi \cdot X_\varphi$ is the identity matrix

of dimension R (as the matrix representing $\varphi \circ \varphi^{-1}$). Easy matrix computations establish the following.

Proposition 47. *Let $\mathcal{A} = \langle I, E, T \rangle$ be a $\mathbb{K}$-automaton of dimension Q. An equivalence φ on Q is an* Out-morphism *if and only if E and T satisfy the two equations*

$$X_\varphi \cdot Y_\varphi \cdot E \cdot X_\varphi = E \cdot X_\varphi \ , \tag{2.4}$$

and

$$X_\varphi \cdot Y_\varphi \cdot T = T \ . \tag{2.5}$$

In this case, the $\mathbb{K}$-automaton $\mathcal{B} = \langle J, F, U \rangle$ defined by the following equations

$$F = Y_\varphi \cdot E \cdot X_\varphi \,, \qquad J = I \cdot X_\varphi \qquad \text{and} \qquad U = Y_\varphi \cdot T \tag{2.6}$$

is the quotient of $\mathcal{A}$ by φ.

Equations 2.4 and 2.5 can be read in the following way: an equivalence φ on Q is an Out-morphism (understood, of $\mathcal{A}$) if for any two states p and p' equivalent modulo φ the *sum* of the labels of the transitions that go from p to *all the states of a whole class* modulo φ is equal to the *sum* of the labels of the transitions that go from p' to the same states *and* if any two entries of T indexed by equivalent states modulo φ are equal, that is (we denote by $[q]_\varphi$ the class of q modulo φ):

$$\forall p, p', q \in Q \qquad p \equiv p' \mod \varphi \implies \begin{cases} \text{(i)} & \displaystyle\sum_{r \in [q]_\varphi} E_{p,r} = \sum_{s \in [q]_\varphi} E_{p',s} \\ \text{(ii)} & T_p = T_{p'} \end{cases} \tag{2.7}$$

Remark 48. It is easy to chek that Definitions 12 and 17 coincide if $\mathbb{K} = \mathbb{B}$.

2.3 Minimal quotient

Theorem 49. *Let $\mathcal{A}$ be a $\mathbb{K}$-automaton of finite dimension. Among all quotients of $\mathcal{A}$ (resp. among all co-quotients of $\mathcal{A}$), there exists one quotient (resp. one co-quotient), unique up to isomorphism and effectively computable from $\mathcal{A}$, which has a minimal number of states and which is a quotient (resp. a co-quotient) of all these $\mathbb{K}$-automata.*

Proof. A surjective map $\varphi\colon Q \to R$ defines an Out-morphism $\varphi\colon \mathcal{A} \to \mathcal{B}$ if and only if Equations (2.4) and (2.5) (which do not involve $\mathcal{B}$) are satisfied.

To prove the existence of a minimal quotient, it suffices to show that if $\varphi\colon Q \to R$ and $\psi\colon Q \to P$ are two maps that define Out-morphisms, the map $\omega\colon Q \to S$ also defines an Out-morphism, where $\omega = \varphi \vee \psi$ is the map whose map equivalence is the upper bound of those of φ and ψ; that is, the finest equivalence which is coarser than the map equivalences of φ and ψ. In other words, there exist $\varphi'\colon R \to S$ and $\psi'\colon P \to S$ such that $\omega = \varphi\varphi' = \psi\psi'$ and each class modulo $\omega = \varphi \vee \psi$ can

be seen at the same time as a union of classes modulo φ and as a union of classes modulo ψ. It follows that

$$E \cdot X_\omega = E \cdot X_\varphi \cdot X_{\varphi'} = E \cdot X_\psi \cdot X_{\psi'} \tag{2.8}$$

and if two states p and r of Q are congruent modulo ω, there exists q such that $\varphi(p) = \varphi(q)$ and $\psi(q) = \psi(r)$ (in fact a sequence of states q_i *etc.*). The rows p and q of $E \cdot X_\varphi$ are equal, and the rows q and r of $E \cdot X_\psi$ are equal, hence, by (2.8), the rows p and r of $E \cdot X_\omega$ are too.

To compute this minimal quotient we can proceed by successive refinements of partitions, exactly as for the computation of the minimal automaton of a language from a deterministic automaton which recognises the language.

In what follows the maps φ_i are identified with their map equivalences; the image is irrelevant. A state r of Q is identified with the row vector of dimension Q, characteristic of r and treated as such. For example, $\varphi(r) = \varphi(s)$ can be written $r \cdot X_\varphi = s \cdot X_\varphi$.

The map φ_0 has the same map equivalence as T; that is,

$$r \cdot X_{\varphi_0} = s \cdot X_{\varphi_0} \Leftrightarrow r \cdot T = s \cdot T \ ,$$

which can also be written

$$X_{\varphi_0} \cdot Y_{\varphi_0} \cdot T = T \ , \tag{2.9}$$

and the same equation holds for every map finer than φ_0. For each i, φ_{i+1} is finer than φ_i and, by definition, r and s are joint in φ_i (that is, $r \cdot X_{\varphi_i} = s \cdot X_{\varphi_i}$) and disjoint in φ_{i+1} if $r \cdot E \cdot X_{\varphi_i} \neq s \cdot E \cdot X_{\varphi_i}$. Let j be the index such that $\varphi_{j+1} = \varphi_j$, that is, such that

$$r \cdot X_{\varphi_j} = s \cdot X_{\varphi_j} \quad \Longrightarrow \quad r \cdot E \cdot X_{\varphi_j} = s \cdot E \cdot X_{\varphi_j} \ , \tag{2.10}$$

which can be rewritten

$$X_{\varphi_j} \cdot Y_{\varphi_j} \cdot E \cdot X_{\varphi_j} = E \cdot X_{\varphi_j} \ . \tag{2.11}$$

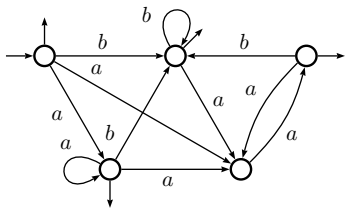
By (2.9) and (2.11), φ_j is an Out-morphism.

Conversely, every Out-morphism ψ satisfies (2.5) and is hence finer than φ_0. Then, for all i, if ψ is finer than φ_i it must also be finer than φ_{i+1}. In fact, if r and s are joint in ψ, it follows that $r \cdot X_\psi = s \cdot X_\psi$ and hence also $r \cdot X_{\varphi_i} = s \cdot X_{\varphi_i}$ since φ_i is coarser than ψ, and hence r and s are joint in φ_{i+1}: ψ is finer than φ_j which is thus the coarsest Out-morphism. ■

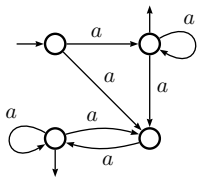
Remark 50. After establishing that the minimal quotient of a $\mathbb{K}$-automaton and the minimal automaton of a language are computed by the *same* algorithm, let us repeat what we already stated in Remark 29: the latter automaton is canonically associated with the language, whereas the former is associated with the $\mathbb{K}$-automaton we started from, and not with its behaviour.

3 Exercises

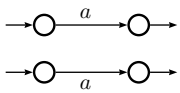
1. Compute the (minimal) quotient of the following $\mathbb{B}$-automaton:



2. Let $\mathcal{D}_1$ be the $\mathbb{B}$-automaton below. Compute the (minimal) quotient of $\mathcal{D}_1$, the co-quotient of $\mathcal{D}_1$, the co-quotient of the quotient of $\mathcal{D}_1$, *etc.*



3. Calculate all the quotients and all the co-quotients of the $\mathbb{N}$-automaton:

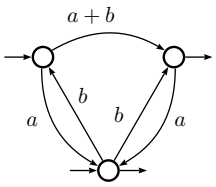


4. <**Coloured Transition Lemma.** Establish the following statement:

Let $\mathcal{A}$ be a (Boolean) automaton on a monoid M the transitions of which are coloured in red or in blue. Then, the set of labels of computations of $\mathcal{A}$ that contain at least one red transition is a rational set (of M).

5. Show that any $\mathbb{Z}$-rational series is the difference of two $\mathbb{N}$-rational series.

6. Construct the Schützenberger covering $\mathcal{S}$ of the following $\mathbb{B}$-automaton $\mathcal{A}$.



How many S-immersions are there in this covering (that is, how many sub-automata $\mathcal{T}$ of $\mathcal{S}$ that are unambiguous and equivalent to $\mathcal{A}$)?

7. Compute the Schützenberger covering of the $\mathbb{B}$-automaton $\mathcal{B}_1$ of the Figure 6.

8. **Quotients and product of automata.** Let $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ be three $\mathbb{K}$-automata on A^*. Show that if $\mathcal{B}$ is a quotient of $\mathcal{A}$, then $\mathcal{B}\otimes\mathcal{C}$ is a quotient of $\mathcal{A}\otimes\mathcal{C}$.

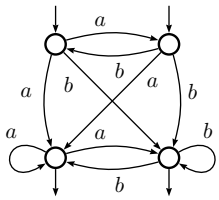


Figure 6: The automaton $\mathcal{B}_1$

9. **Quotients and co-quotients of the $\mathcal{C}_n$.**

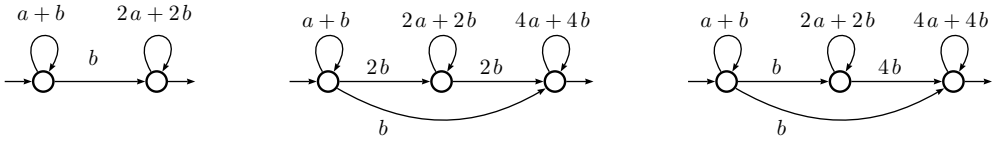
The $\mathbb{N}$-automaton $\mathcal{C}_1$ over $\{a,b\}^*$ shown at Figure 7 (a) associates with every word w the integer $\overline{w}$ the binary representation of which is w when a is replaced by the digit 0 and b by 1.

Let $\mathcal{C}_2$ be the tensorial square of $\mathcal{C}_1$: $\mathcal{C}_2 = \mathcal{C}_1 \otimes \mathcal{C}_1$; $\mathcal{V}_2$, shown at Figure 7 (b), is the minimal quotient of $\mathcal{C}_2$ and $\mathcal{V}_2'$, shown at Figure 7 (c), is the minimal co-quotient of $\mathcal{C}_2$.

(a) Compute the minimal quotient $\mathcal{V}_3$ and the minimal co-quotient $\mathcal{V}_3'$ of $\mathcal{C}_3 = \mathcal{C}_2 \otimes \mathcal{C}_1$.

(b) Compute the minimal co-quotient $\mathcal{V}_4'$ of $\mathcal{C}_4 = \mathcal{C}_3 \otimes \mathcal{C}_1$. Compare with $\mathcal{V}_3'$.

(c) Generalising the above computation, compute the minimal co-quotient $\mathcal{V}_{n+1}'$ of $\mathcal{C}_{n+1} = \mathcal{C}_n \otimes \mathcal{C}_1$, for every n.



(a) $\mathcal{C}_1$ (b) $\mathcal{V}_2$ (c) $\mathcal{V}_2'$

Figure 7: Three $\mathbb{N}$-automata

10. **Conjugacy of an automaton and its determinisation.**

(a) Let $\mathcal{A}_1$ be the (Boolean) automaton of Figure 8 and $\widehat{\mathcal{A}_1}$ its determinisation. Verify that $\widehat{\mathcal{A}_1} \overset{X_1}{\Longrightarrow} \mathcal{A}_1$ holds, with

$$X_1 = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix} .$$

(b) Generalisation. Let $\mathcal{A}$ be a (Boolean) automaton and $\widehat{\mathcal{A}}$ its determinisation. Show that there exists an Boolean matrix X such that $\widehat{\mathcal{A}} \overset{X}{\Longrightarrow} \mathcal{A}$.

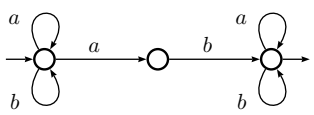


Figure 8: L'automate $\mathcal{A}_1$

11. **Automata with bounded ambiguity and the Schützenberger covering.** In the sequel, $\mathcal{A}$ is a Boolean automaton, $\widehat{\mathcal{A}}$ its determinisation, and $\mathcal{S}$ its Schützenberger covering.

Definition 51. We call *concurrent transition set* of $\mathcal{S}$ a set of transitions which

(i) have the same destination (final extremity),

(ii) are mapped onto the same transition of $\widehat{\mathcal{A}}$.

Two transitions of $\mathcal{S}$ are called *concurrent* if they belong to the same concurrent transition set.

We also set the folllowing definition:

Definition 52. An automaton $\mathcal{A}$ over A^* is of *bounded ambiguity* if there exists an integer k such that every word w in $\big|\mathcal{A}\big|$ is the label of at most k distinct computations. The smallest such k is the *ambiguity degree* of $\mathcal{A}$.

(a) What can be said of an automaton whose Schützenberger covering contains no concurrent transitions?

(b) Show that there exists a computation in $\mathcal{S}$ which contains two transitions of the same concurrent transition set if and only if there exists a concurrrent transition which belongs to a circuit.

(c) Let $p \xrightarrow{a} s$ and $q \xrightarrow{a} s$ be two concurrent transitions of $\mathcal{S}$ and

$$c := \underset{\mathcal{S}}{\longrightarrow} i \xrightarrow[\mathcal{S}]{x} p \xrightarrow[\mathcal{S}]{a} s \xrightarrow[\mathcal{S}]{y} q \xrightarrow[\mathcal{S}]{a} s \xrightarrow[\mathcal{S}]{z} t \underset{\mathcal{S}}{\longrightarrow}$$

a computation of $\mathcal{S}$ where i is an initial state and t a final state. Show that $w = x\,a\,y\,a\,z$ is the label of at least two computations of $\mathcal{A}$.

(d) Prove that an automaton $\mathcal{A}$ is of bounded ambiguity if and only if no concurrent transition of its Schützenberger covering belongs to a circuit.

(e) Check that $\mathcal{B}_1$ of Figure 6 is of bounded ambiguity.

(f) Give a bound on the ambiguity degree of an automaton as a function of the cardinals of the concurrent transition sets of its Schützenberger covering.

Compute that bound in the case of $\mathcal{B}_1$.

(g) Infer from the above the complexity of an algorithm which decide if an automaton is of bounded ambiguity.

Lecture III

Reduction of weighted automata

Controllability and observability

Given a (finite) $\mathbb{K}$-automaton over A^*, we want to build equivalent ones, hopefully of smaller dimension. In this lecture, we base this construction upon the *behaviour* of the automaton, in contrast with the preceding lecture where we have addressed the same question by considering the *structure* of the automaton.

Contents

We first define a process, that we call the *universal determinisation process*, and that yields a possibly infinite automaton, but a *deterministic* one and with no multiplicity on the transitions but in the final function. We then explain how to work in that framework.

In the next section, we introduce the notion of *quotient* of a series, show a characterisation of recognisable series in terms of quotients, and relate the quotients with the previous construction.

Finally, we combine the two approaches to set up the theory for reduction in the case where the multiplicity semiring is a field, or a subsemiring of a field, and turn the theory into a polynomial algorithm.

As a preliminary, let us recall the main result of Lecture I: finite $\mathbb{K}$-automata over A^* and $\mathbb{K}$-representations of A^* are one and the same thing and we use the most convenient form; here, the representation. Before all, let us introduce the notion of *action* that will be ubiquitous in that lecture.

1 Actions and representations

The notion of *action* is central in this lecture. We first define it and then describe how it relates to the one of representation.

Contrary to what we have done in Lecture I (where we have defined $\mathbb{K}$-automata over free monoids and then extended the notion to $\mathbb{K}$-automata over general (graded) monoids), we first define actions of arbitrary monoids and then consider in the following sections actions of free monoids only.

1.1 Action of a monoid on a set

Definition 1. Let S be a set — finite or infinite — and M a monoid. An *action* δ of M on S is a map from $S \times M$ to S, denoted by $s \cdot m$ rather than by $\delta(s, m)$, which meets the two conditions:

$$\begin{aligned} &\forall s \in S && s \cdot 1_M = s \,, \\ &\forall s \in S \,,\ \forall m, m' \in M && (s \cdot m) \cdot m' = s \cdot (m\, m') \ . \end{aligned} \tag{1.1}$$

The *orbit* of an element s of S under the action δ is the subset of S that can be reached from s by the actions of all elements of M, that is, the set $\{s \cdot m \mid m \in M\}$.

When necessary, we write $s \mathbin{\dot{\delta}} m$ in order to differentiate between two different actions of a monoid or from the matrix product symbol.

Examples 2. (i) Permutation groups are examples of monoid actions; even more, one may say that monoid actions are generalisation of permutation groups.

(ii) Every monoid M defines an *action on itself* by multiplication on the right:

$$\forall p, m \in M \qquad p \cdot m = pm \ ;$$

this action is called the *right translation* or the *(right) regular representation* of M over itself. It is denoted (when necessary) by ρ.

(iii) Likewise, every morphism $\alpha\colon M \to N$ defines an action of M on N:

$$\forall n \in N\,,\ \forall m \in M \qquad n \cdot m = n\,(\alpha(m)) \ . \tag{1.2}$$

This map satisfies (1.1) because α is a morphism and multiplication in N is associative. (The regular representation above corresponds to the identity morphism ι from M onto itself.[1])

Right and left actions The actions we have thus defined are *right* actions. We could have defined in a dual manner a *left* action of M on S as a map from $M \times S$ to S that satisfies the conditions:

$$\begin{aligned} &\forall s \in S && 1_M \cdot s = s\,, \\ &\forall s \in S\,,\ \forall m, m' \in M && m' \cdot (m \cdot s) = (m'\,m) \cdot s \ . \end{aligned} \tag{1.3}$$

Actions on structured sets An action of M on S is a morphism from M into the monoid of maps from S into itself. If S has a structure (*e.g.* being a group, a ring, *etc.*), we want an action to be a morphism from M into the monoid of *endomorpisms* of S. In the sequel, S is a $\mathbb{K}$-module and an action of M on S is 'linear':

$$\begin{aligned} &\forall s, t \in S\,,\ \forall m, m' \in M && (s+t) \cdot m = s \cdot m + t \cdot m\,, \\ &\forall k \in \mathbb{K} && (k\,s) \cdot m = k\,(s \cdot m) \ . \end{aligned}$$

Examples 3. (i) A special case of Example 2 (ii) is the regular representation of A^* over itself, which extends by linearity to an action of A^* on the $\mathbb{K}$-*module* $\mathbb{K}\langle A^*\rangle$, and which we call the *right translation* by A^*.

(ii) Any $\mathbb{K}$-representation (or morphism) $\mu\colon M \to \mathbb{K}^{Q\times Q}$ of dimension Q defines an action of M on the *(left)* $\mathbb{K}$-*module* $\mathbb{K}^Q$ (on $\mathbb{K}^{1\times Q}$, indeed), also denoted by μ:

$$\forall x \in \mathbb{K}^Q\,,\ \forall m \in M \qquad x \underset{\mu}{\cdot} m = x \cdot \mu(m) \ . \tag{1.4}$$

Since $\mathbb{K}$ is not supposed to be commutative, it is important to specify that $\mathbb{K}^{1\times Q}$ is a left module.

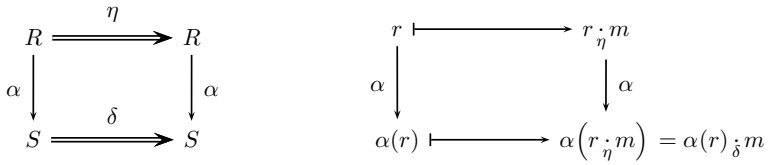
[1] But is not denoted as such, as it is misleading to denote by ι a map which is not the identity.

Action morphisms Let R and S be two structures (in the sequel, they will be $\mathbb{K}$-modules) and suppose that M acts on both R and S, by η and δ respectively.

A morphism $\alpha\colon R \to S$ is an *action morphism* if

$$\forall r \in R\,,\ \forall m \in M \qquad \alpha(r) \mathbin{\dot\delta} m = \alpha\Big(r \mathbin{\dot\eta} m\Big) \ ,$$

that is, if the following diagram is commutative (for every m in M).



1.2 Actions and deterministic automata

If we distinguish an element s_0 in S, that will play the role of an initial state, and a subset T of S, that will play the role of the set of final states, any action δ of M on S defines an automaton $\mathcal{A}_\delta = \langle\, M, S, \{s_0\}, \delta, T\,\rangle$.

If M is a free monoid A^*, $\mathcal{A}_\delta$ is a *deterministic* Boolean automaton. And conversely any (complete) deterministic Boolean automaton $\mathcal{A}$ over A^* determines an action of A^* on the state set of $\mathcal{A}$. If M is not a free monoid, the notion of action is indeed *the* way to generalise the one of deterministic automaton.

exercice to be written

If we replace the subset T by a *function* T from S to $\mathbb{K}$ (the former being a function from S to $\mathbb{B}$) the action δ, together with s_0 and T, now defines a $\mathbb{K}$-automaton, which we call again deterministic in which the weight of every transition is $1_\mathbb{K}$ and the final function is T. The behaviour of $\mathcal{A}_\delta$ is then defined by $\big\langle\big|\mathcal{A}_\delta\big|, m\big\rangle = T(s_0 \cdot m)$ for every m in M.

1.3 Closed sets

In this section, S is a (left) $\mathbb{K}$-module (later, it will be $\mathbb{K}^{1\times Q}$ or $\mathbb{K}\langle\langle A^*\rangle\rangle$). We first take some notations that prove to be (very) convenient.

Notations for submodules Any finite subset G of S induces a morphim

$$\alpha_G\colon \mathbb{K}^G \to S \ ,$$

whose image is $\langle G \rangle$, the sub(-$\mathbb{K}$-)module of S generated by G:

$$\forall x \in \mathbb{K}^G \qquad \alpha_G(x) = \sum_{g\in G} x_g\, g \ .$$

We also write

$$\alpha_G(x) = x \cdot G$$

which implicitely means that G is viewed as a *column-vector* of dimension G of elements of S.

Conversely, let $\beta_G \colon \langle G \rangle \to \mathbb{K}^G$ be a map that performs, for every v in $\langle G \rangle$, *a choice* of a decomposition of v over the elements of G and hence, for every v in $\langle G \rangle$:

$$\alpha_G(\beta_G(v)) = v \ . \tag{1.5}$$

Such a decomposition *is not unique* in general; that is, when G is not a *basis* of $\langle G \rangle$) and $\beta_G(\alpha_G(x))$ and x are not necessarily equal (but $\alpha_G(\beta_G(\alpha_G(x))) = \alpha_G(x)$ holds.

It is natural, even though not necessary, to assume that for every g in G, $\beta_G(g)$ is the vector whose all entries are $0_{\mathbb{K}}$ but the g-th one which is $1_{\mathbb{K}}$. In other words, $\beta_G(G)$ is the identity matrix of dimension G. *Is this alinea useful?*

Definition 4. Let S be a (left) $\mathbb{K}$-module and δ a (right) action of A^* on S. A subset G of S is said to be δ-*closed* if the orbit of G is contained in $\langle G \rangle$, that is, if

$$\forall g \in G \,,\ \forall w \in A^* \qquad g \mathbin{\dot{\delta}} w \in \langle G \rangle \ ,$$

which amounts to say that $\langle G \rangle$ itself is closed, or *stable*, under the action of δ:

$$\forall v \in \langle G \rangle \,,\ \forall w \in A^* \qquad v \mathbin{\dot{\delta}} w \in \langle G \rangle \ .$$

If v is in $\langle G \rangle$, there exists x in $\mathbb{K}^G$ such that $v = \alpha_G(x) = x \cdot G$ and then:

$$v \mathbin{\dot{\delta}} w = x \cdot \left(G \mathbin{\dot{\delta}} w \right) \ .$$

1.4 Closed sets and representations

The core of this section is to show that an action on a finite closed set can be lifted into a representation. We give indeed two versions of this construction: the lifting of actions and the lifting of representations.

1.4.1 Lifting of actions

Proposition 5. *Let S be a (left) $\mathbb{K}$-module, δ a (right) action of A^* on S and G a finite subset of S.*

If G is δ-closed, then there exists a $\mathbb{K}$-representation κ_G (not necessarily unique) of A^* of dimension G such that α_G is an action morphism between the action of A^* on $\mathbb{K}^G$ defined by κ_G and δ, that is, such that the following diagram commutes.

$$\begin{array}{ccc} \mathbb{K}^G & \xrightarrow{\kappa_G} & \mathbb{K}^G \\ {\scriptstyle \alpha_G}\downarrow & & \downarrow{\scriptstyle \alpha_G} \\ S \supseteq \langle G \rangle & \xrightarrow{\delta} & \langle G \rangle \subseteq S \end{array} \tag{1.6}$$

Proof. Let $\beta_G\colon \langle G \rangle \to \mathbb{K}^G$ be a fixed map defined as above. Since G is δ-closed, for every g in G and every a in A, $g \mathbin{\dot\delta} a$ is in $\langle G \rangle$ and hence $\beta_G\Big(g \mathbin{\dot\delta} a\Big)$ is in $\mathbb{K}^G$. Let κ_G be defined by

$$\forall a \in A \qquad \kappa_G(a) = \beta_G\Big(G \mathbin{\dot\delta} a\Big) \quad ,$$

that is, since we see G as a column-vector of dimension G, $\beta_G\Big(G \mathbin{\dot\delta} a\Big)$ is a $G \times G$-matrix (with entries in $\mathbb{K}$) the g-th row of which is $\beta_G\Big(g \mathbin{\dot\delta} a\Big)$. Hence every map β_G defines a representation κ_G, possibly distinct from the others.

If we instanciate diagram (1.6) for x in $\mathbb{K}^G$ and a in A, it comes:

$$\begin{array}{ccc} x & \longmapsto & x \cdot \kappa_G(a) = x \cdot \beta_G\Big(G \mathbin{\dot\delta} a\Big) \\ {\scriptstyle \alpha_G}\downarrow & & \downarrow{\scriptstyle \alpha_G} \\ x \cdot G = \alpha_G(x) & \longmapsto & \alpha_G(x) \mathbin{\dot\delta} a = x \cdot \Big(G \mathbin{\dot\delta} a\Big) \end{array} \tag{1.7}$$

on which we read the following sequence of equalities:

$$\alpha_G(x) \mathbin{\dot\delta} a = (x \cdot G) \mathbin{\dot\delta} a = x \cdot (G \mathbin{\dot\delta} a) \qquad \text{as } \delta \text{ is } \mathbb{K}\text{-linear.} \tag{1.8}$$

$$\begin{aligned} \text{On the other hand} \qquad x \cdot \kappa_G(a) &= x \cdot \beta_G\Big(G \mathbin{\dot\delta} a\Big) && \text{by definition} \\ \alpha_G(x \cdot \kappa_G(a)) = \alpha_G\Big(x \cdot \beta_G\Big(G \mathbin{\dot\delta} a\Big)\Big) &= x \cdot \alpha_G\Big(\beta_G\Big(G \mathbin{\dot\delta} a\Big)\Big) && \text{as } \alpha_G \text{ is } \mathbb{K}\text{-linear,} \\ &= x \cdot (G \mathbin{\dot\delta} a) && \text{by (1.5).} \end{aligned} \tag{1.9}$$

The equality between the right hand-sides of (1.8) and (1.9) expresses that the diagram (1.6) commutes. ∎

1.4.2 Lifting of representations

Let Q be any finite set. We study the preceding case when $S = \mathbb{K}^Q$ and δ is the action on $\mathbb{K}^Q$ defined by a representation $\mu\colon A^* \to \mathbb{K}^{Q \times Q}$.

Let G be a finite subset of $\mathbb{K}^Q$. We denote by M_G the $G \times Q$-matrix (with entries in $\mathbb{K}$) the g-th row of which is the *row-vector* g of $\mathbb{K}^Q$. In this context, to say that G is a column-vector amounts to say that G is in $\Big(\mathbb{K}^{1 \times Q}\Big)^{G \times 1} = \mathbb{K}^{G \times Q}$ that is, that G *is the matrix* $M_G = \alpha_G(\mathtt{Id})$, where $\mathtt{Id}$ is the identity matrix of dimension G.

Proposition 6. *Let $\mathcal{A} = \langle\, I, \mu, T \,\rangle$ be a $\mathbb{K}$-representation of A^* of dimension Q. Any finite subset G of $\mathbb{K}^Q$, that is μ-closed and that decomposes I, determines (not uniquely) a $\mathbb{K}$-representation $\langle\, J, \kappa_G, U \,\rangle$ of dimension G that is conjugate to $\mathcal{A}$ by M_G (hence equivalent to $\mathcal{A}$).*

Proof. Let us come back to diagram (1.6) of Proposition 5:

$$\begin{array}{ccc} \mathbb{K}^G & \xrightarrow{\kappa_G} & \mathbb{K}^G \\ {\scriptstyle \alpha_G}\downarrow & & \downarrow{\scriptstyle \alpha_G} \\ \mathbb{K}^{1\times Q} \supseteq \langle G \rangle & \xrightarrow{\mu} & \langle G \rangle \subseteq \mathbb{K}^{1\times Q} \end{array}$$

If we replace in the proof of Proposition 5 the action δ by the action μ determined by $\mathcal{A}$ and G by M_G, it comes

$$G \mathbin{\dot{\delta}} a = M_G \cdot \mu(a) \qquad \text{and} \qquad \forall x \in \mathbb{K}^G \qquad \alpha_G(x) = x \cdot M_G \ .$$

The above diagram instanciated for $\mathtt{Id}$ and any letter a of A yields

$$\begin{array}{ccc} \mathtt{Id} & \longmapsto & \kappa_G(a) \\ {\scriptstyle \alpha_G}\downarrow & & \downarrow{\scriptstyle \alpha_G} \\ M_G & \longmapsto & M_G \cdot \mu(a) = \kappa_G(a) \cdot M_G \end{array}$$

which shows that for every a in A, $\kappa_G(a)$ is conjugate to $\mu(a)$ by M_G.

Furthermore, to say that G decomposes I, that is, $I \in \langle G \rangle$, implies that there exists J in $\mathbb{K}^G$ such that $I = \alpha_G(J) = J \cdot M_G$. If we write $U = M_G \cdot T$, we have built the $\mathbb{K}$-representation we wanted. ∎

2 Control

Starting from a $\mathbb{K}$-automaton, we paradoxically begin our search for small equivalent automata by the definition and idealistic constuction of two automata that are infinite (in the general case). In this section, we start from the given automaton itself, in the next one from its behaviour. In some sense, we thus begin with the *effective* level as the (finite) automaton (or representation) is effectively given; since the computations may lead to an *infinite* automaton, this effectivity is somewhat relative. Linear algebra will then allow to fold these infinite automata into finite ones, hopefully, and when possible, optimally.

For the rest of this section, $\mathcal{A} = \langle I, \mu, T \rangle$ is a $\mathbb{K}$-representation of A^*, of dimension Q.

2.1 The reachability set

The representation $\mathcal{A}$, the morphism μ indeed, determines an *action* of A^* on $\mathbb{K}^Q$, called μ again, by

$$\forall x \in \mathbb{K}^Q \qquad x \mathbin{\dot{\mu}} w = x \cdot \mu(w) \ .$$

Definition 7. The *reachability set* $\mathbf{R}_\mathcal{A}$ of $\mathcal{A}$ is the orbit of I under the action μ:

$$\mathbf{R}_\mathcal{A} = \{ I \cdot \mu(w) \mid w \in A^* \} \ .$$

This set $\mathbf{R}_\mathcal{A}$ may well be, and in general is, infinite.

Determinisation The set $\mathbf{R}_{\mathcal{A}}$ is closed under the action μ and this action of A^* on $\mathbf{R}_{\mathcal{A}}$ can be seen as a *deterministic automaton*, denoted by $\widehat{\mathcal{A}}$, and called the *determinisation* of $\mathcal{A}$ as it is defined by $\mathcal{A}$:

$$\widehat{\mathcal{A}} = \left\langle A, \mathbf{R}_{\mathcal{A}}, \{I\}, \mu, \widehat{T} \right\rangle \ .$$

Its transitions are defined by: $[I \cdot \mu(w)]_{\mu} a = I \cdot \mu(w) \cdot \mu(a) = I \cdot \mu(w\,a)$ and its final function by: $\widehat{T}(x) = x \cdot T$. The automaton $\widehat{\mathcal{A}}$, a priori infinite, is equivalent to $\mathcal{A}$.

Example 8. Let $\mathcal{A}_2$ and $\mathcal{A}_3$ be the $\mathbb{N}$-automata over a^* of dimension 1 and 2 defined by $\mathcal{A}_2 = \left\langle \begin{pmatrix}1\end{pmatrix}, \begin{pmatrix}2\end{pmatrix}, \begin{pmatrix}1\end{pmatrix} \right\rangle$ and by $\mathcal{A}_3 = \left\langle \begin{pmatrix}1 & 0\end{pmatrix}, \begin{pmatrix}1 & 1\\ 0 & 2\end{pmatrix}, \begin{pmatrix}1\\ 1\end{pmatrix} \right\rangle$ respectively. Their determinisations are shown at Figure 1. The determinisation of $\mathcal{B}_1$ (*cf.* Example I.3) is shown at Figure 2.

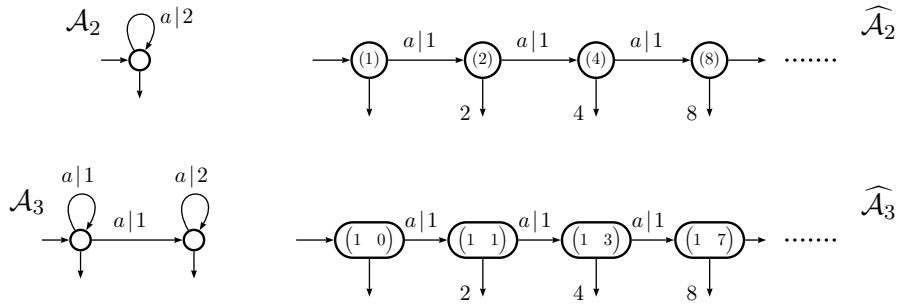


Figure 1: Two (equivalent) $\mathbb{N}$-automata and their (equal) determinisations

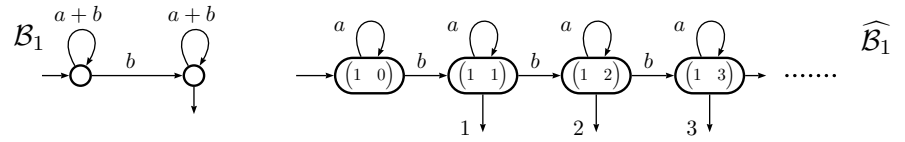


Figure 2: The determinisation of the $\mathbb{N}$-automaton $\mathcal{B}_1$

The Boolean case The use of the word determinisation, as in the case of classical Boolean automata, is not a coincidence: if $\mathbb{K} = \mathbb{B}$ the construction we have described *is* the so-called *subset construction*. Every Boolean vector of $\mathbb{B}^Q$ can be identified with a subset of Q , and conversely. The initial state I *is* the set of initial states of $\mathcal{A}$, and $(I \cdot \mu(w)) \cdot \mu(a)$ *is* the set of states reached by the letter a from the set of states $I \cdot \mu(w)$.

2.2 The state-space

So far, $\mathbb{K}^Q$, and thus $\mathbf{R}_{\mathcal{A}}$, have been considered as sets without any structure. We now bring into play the fact that $\mathbb{K}^Q = \mathbb{K}^{1\times Q}$ is a *(left) module over* $\mathbb{K}$.

Definition 9. We call *state-space* of $\mathcal{A}$ the $\mathbb{K}$-module $\mathbb{K}^Q$.

Definition 10. We call *control morphism* of $\mathcal{A}$ the morphism of $\mathbb{K}$-modules $\Psi_{\mathcal{A}}$:

$$\Psi_{\mathcal{A}} \colon \mathbb{K}\langle A^* \rangle \longrightarrow \mathbb{K}^Q \ ,$$

defined by $\Psi_{\mathcal{A}}(w) = I \cdot \mu(w)$ for every w in A^* and extended to $\mathbb{K}\langle A^* \rangle$ by linearity.

With these definition and notation, it holds:

$$\mathbf{R}_{\mathcal{A}} = \Psi_{\mathcal{A}}(A^*) \qquad \text{and} \qquad \mathsf{Im}\, \Psi_{\mathcal{A}} = \Psi_{\mathcal{A}}(\mathbb{K}\langle A^* \rangle) = \langle \mathbf{R}_{\mathcal{A}} \rangle \ .$$

and the following statement is almost a tautology.

Proposition 11. *The control morphism $\Psi_{\mathcal{A}}$ is an action morphism from the right translation by A^* on $\mathbb{K}\langle A^* \rangle$ to the action μ of A^*.*

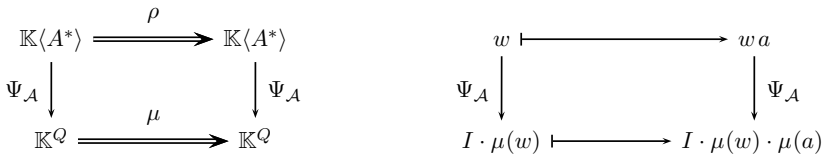


Figure 3: The control morphism is a morphism of actions

Definition 12. A $\mathbb{K}$-representation (or $\mathbb{K}$-automaton) $\mathcal{A}$ is said to be *controllable*[2] if $\Psi_{\mathcal{A}}$ is *surjective*.

The automaton $\mathcal{A}$ is controllable if for every point in the state space, there exists at least one linear combination of input that leads $\mathcal{A}$ to that point.

3 Observation

We now define a third action of A^*, on $\mathbb{K}\langle\langle A^* \rangle\rangle$ this time. It allows to *characterise* recognisable series and to associate with every such series a *minimal* deterministic automaton.

3.1 Quotient of series

The quotient of a series is the generalisation to series of the quotient of a subset of a monoid (of a free monoid in this case).

Definition 13. Let s be in $\mathbb{K}\langle\langle A^* \rangle\rangle$ and w in A^*. The (left) quotient of s by w is the series denoted by $w^{-1}s$ and defined by:

$$w^{-1}s = \sum_{v \in A^*} \langle s, w\, v \rangle\, v \ , \qquad \text{that is,} \qquad \forall v \in A^* \quad \langle w^{-1}s, v \rangle = \langle s, w\, v \rangle \ . \tag{3.1}$$

$$\text{In particular,} \qquad \forall w \in A^* \qquad \langle w^{-1}s, 1_{A^*} \rangle = \langle s, w \rangle \ . \tag{3.2}$$

[2] *commandable* in French.

For every w, the operation $s \mapsto w^{-1}s$ is an *endomorphism* of the $\mathbb{K}$-module $\mathbb{K}\langle\langle A^*\rangle\rangle$: it is *additive*:

$$w^{-1}(s+t) = w^{-1}s + w^{-1}t\,,$$

and *commutes with the exterior multiplications* of $\mathbb{K}$ on $\mathbb{K}\langle\langle A^*\rangle\rangle$:

$$w^{-1}(k\,s) = k\,(w^{-1}s) \qquad \text{and} \qquad w^{-1}(s\,k) = (w^{-1}s)\,k\,.$$

Moreover, it is *continuous*. These three properties ensure that the quotient by w is entirely defined on $\mathbb{K}\langle\langle A^*\rangle\rangle$ by its values on A^* since $\mathbb{K}\langle A^*\rangle$ is dense in $\mathbb{K}\langle\langle A^*\rangle\rangle$ (*cf.* Section I.2.2.1).

The associativity of concatenation implies then that

$$\forall u, v \in A^* \qquad (u\,v)^{-1}s = v^{-1}\left[u^{-1}s\right]\,,$$

that is, thanks to the preceding properties:

Proposition 14. *The (left) quotient is a (right) action of A^* on the (left) $\mathbb{K}$-module $\mathbb{K}\langle\langle A^*\rangle\rangle$.*[3]

The orbit of a series s under the quotient action is denoted by $\mathbf{R}_s$:

$$\mathbf{R}_s = \left\{ w^{-1}\,s \;\middle|\; w \in A^* \right\}\quad .$$

Example 15. Let $s_2 = (\underline{a^*})^2 = \sum_{n\in\mathbb{N}}(n+1)\,a^n$ in $\mathbb{N}\mathrm{Rat}\,a^*$. For every integer k, it holds:

$$(a^k)^{-1}s_2 = \sum_{n\in\mathbb{N}}(k+n+1)\,a^n = s_2 + k\,\underline{a^*}\quad .$$

All quotients of s_2 are distinct and $\mathbf{R}_{s_2} = \{s_2 + k\,\underline{a^*} \mid k \in \mathbb{N}\}$.

Example 15 shows that, in general, and in contrast with the case for (recognisable) languages, the family of quotients of a rational, and thus recognisable, series is not necessarily finite. On the other hand, and despite its simplicity, it exhibits the property that we seek: there are infinitely many quotients, but they can all be expressed as the linear combination of a *finite number* of suitably chosen series.

3.2 The observation morphism

Let again $\mathcal{A} = \langle\, I, \mu, T\,\rangle$ be a $\mathbb{K}$-representation of A^*, of dimension Q.

Definition 16. We call *observation morphism* of $\mathcal{A}$ the morphism of $\mathbb{K}$-modules $\Phi_{\mathcal{A}} \colon \mathbb{K}^Q \longrightarrow \mathbb{K}\langle\langle A^*\rangle\rangle$ defined by:

$$\forall x \in \mathbb{K}^Q \quad \Phi_{\mathcal{A}}(x) = \big|\langle\, x, \mu, T\,\rangle\big| = \sum_{w\in A^*}(x\cdot\mu(w)\cdot T)\,w\quad .$$

[3] In diagrams, the quotient action will be denoted by $\triangleright$.

The definition of quotient (Equation (3.1)) directly implies that if $s = \big|\langle\, I, \mu, T \,\rangle\big|$, then, for every w in A^*, $w^{-1}s = \big|\langle\, I \cdot \mu(w)\,, \mu, T \,\rangle\big|$, that is:

Property 17. *For every w in A^*, and every x in $\mathbb{K}^Q$, $w^{-1}\Phi_{\mathcal{A}}(x) = \Phi_{\mathcal{A}}(x \cdot \mu(w))$.*

In other words:

Proposition 18. *The observation morphism $\Phi_{\mathcal{A}}$ is an action morphism from the action μ of A^* on $\mathbb{K}^Q$ to the quotient action of A^* on $\mathbb{K}\langle\langle A^*\rangle\rangle$.*

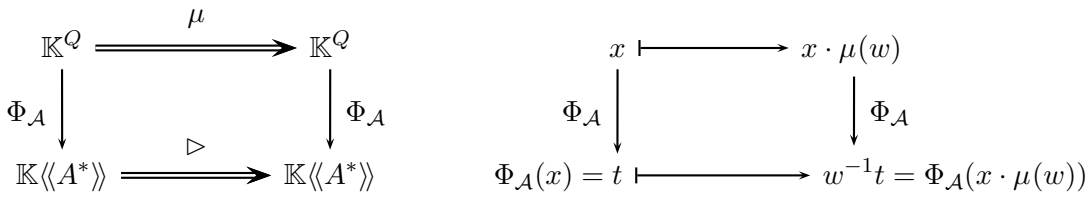


Figure 4: The observation morphism is a morphism of actions

From Property 17 also follows:

Property 19. $\mathbf{R}_s = \Phi_{\mathcal{A}}(\Psi_{\mathcal{A}}(A^*)) = \Phi_{\mathcal{A}}(\mathbf{R}_{\mathcal{A}})$.

Definition 20. A $\mathbb{K}$-representation (or $\mathbb{K}$-automaton) $\mathcal{A}$ is said to be *observable* if $\Phi_{\mathcal{A}}$ is *injective*.

That is, $\mathcal{A}$ is observable if no two distinct starting points in the state-space of $\mathcal{A}$ yield the same behaviour for $\mathcal{A}$.

3.3 The minimal deterministic automaton

The set $\mathbf{R}_s$ is closed under the quotient action and this action of A^* on $\mathbf{R}_s$ can be seen as a *deterministic automaton*, denoted by $\mathcal{A}_s$:

$$\mathcal{A}_s = \langle\, A, \mathbf{R}_s, \{s\}, \triangleright, \mathsf{c} \,\rangle \ .$$

Its transitions are defined by

$$[w^{-1}s] \triangleright a = a^{-1}w^{-1}s = (w\,a)^{-1}s\,,$$

its unique initial state is s, and its final function c maps every state $w^{-1}\,s$ to its *constant term*, that is, $\mathsf{c}(w^{-1}s) = \langle w^{-1}s, 1_{A^*}\rangle = \langle s, w\rangle$.

The automaton $\mathcal{A}_s$, a priori infinite, is equivalent to $\mathcal{A}$: every word w labels a unique path with multiplicity $1_{\mathbb{K}}$ from the initial state s to the state $w^{-1}\,s$ and the final function gives that computation the weight $\langle s, w\rangle$ by definition.

If s is a $\mathbb{B}$-series, that is, if s is a language L, then $\mathcal{A}_L$ is the *minimal automaton* of L. The well-known relation between the determinisation of an automaton and the minimal automaton of the recognised language generalises to series.

Strictly writing, we have defined *Out-morphisms* and *quotients* for finite ($\mathbb{K}$)-automata only but in the same way we have defined *the weight of a word* for finite automata and noted that it could be defined for infinite deterministic automata (Note I.3, p. 6), Out-morphisms and quotients are easily defined also for infinite deterministic $\mathbb{K}$-automata since the definition coincide with the one for finite deterministic automata: two states can be merged if they have the same transitions to the other (merged) states, and give the final function the same value.

Proposition 21. *Let s be a $\mathbb{K}$-recognisable series and $\mathcal{A}$ any finite $\mathbb{K}$-automaton that realises s. Then $\mathcal{A}_s$ is the* minimal quotient *of $\widehat{\mathcal{A}}$.*

Proof. By Property 19, we already know that $\mathbf{R}_s = \Phi_{\mathcal{A}}(\mathbf{R}_{\mathcal{A}})$. Stating that $\Phi_{\mathcal{A}}$ is a morphism of actions is exactly the same thing as saying that $\Phi_{\mathcal{A}}$ is an Out-morphism between the deterministic automata induced by these actions, here, from $\widehat{\mathcal{A}}$ onto $\mathcal{A}_s$.

Conversely, every state of $\mathcal{A}_s$ corresponds to the series that is accepted by this state taken as the initial state. Thus, two *distinct* states of $\mathcal{A}_s$ correspond to *distinct series* and then cannot be mapped by a morphism onto the same state of a proper quotient since they would correspond to the *same* series. ∎

3.4 Stability

The notion of quotient allows us to give an intrinsic characterisation of recognisable series, via the one of stability.

Definition 22. A subset U of $\mathbb{K}\langle\langle A^*\rangle\rangle$ is called *stable* if it is closed under quotient; that is, for every s in U and every w in A^*, $w^{-1}s$ is in U.

Theorem 23 (Fliess–Jacob)**.** *A series of $\mathbb{K}\langle\langle A^*\rangle\rangle$ is $\mathbb{K}$-recognisable if and only if it is contained in a* stable *finitely generated submodule of $\mathbb{K}\langle\langle A^*\rangle\rangle$.*

Proposition 24. *If s is a series realised by $\mathcal{A}$, then $\mathsf{Im}\,\Phi_{\mathcal{A}}$ is a stable (finitely generated) submodule of $\mathbb{K}\langle\langle A^*\rangle\rangle$ that contains s.*

Proof. The submodule $\mathsf{Im}\,\Phi_{\mathcal{A}}$ is finitely generated since $\mathbb{K}^Q$ is, is stable since $\Phi_{\mathcal{A}}$ is a morphism of actions, and contains $s = \Phi_{\mathcal{A}}(I)$. ∎

Proposition 25. *Let U be a stable submodule of $\mathbb{K}\langle\langle A^*\rangle\rangle$ generated by a finite set G. Then, every series in U is realised by a $\mathbb{K}$-representation of dimension G.*

Proof. Proposition 5, applied to the case where $S = \mathbb{K}\langle\langle A^*\rangle\rangle$ and δ is the quotient, yields a $\mathbb{K}$-representation κ_G of dimension G. Every g in G is a series in $\mathbb{K}\langle\langle A^*\rangle\rangle$; let us denote by $T = \langle G, 1_{A^*}\rangle$ the (column) vector whose g-th entry is $\langle g, 1_{A^*}\rangle$.

If a series s is in U, there exists an x in $\mathbb{K}^G$ such that $s = x \cdot G$. By definition of κ_G, for every w in A^*, $w^{-1}s = x \cdot \kappa_G(w) \cdot G$ and then

$$\langle s, w\rangle = \langle w^{-1}s, 1_{A^*}\rangle = \langle x \cdot \kappa_G(w) \cdot G, 1_{A^*}\rangle = x \cdot \kappa_G(w) \cdot \langle G, 1_{A^*}\rangle = x \cdot \kappa_G(w) \cdot T \ .$$

Hence s is realised by $\langle x, \kappa_G, T \rangle$. ∎

Propositions 24 and 25 together prove Theorem 23. ∎

4 Reduction of representations in a field

We now suppose that $\mathbb{K}$ is a *field*, not necessarily commutative, hence a *skew field*, or *division ring*. The preceding considerations about quotients of series will take on, we might say, a new dimension since the ring of series $\mathbb{K}\langle\langle A^* \rangle\rangle$ is not only a $\mathbb{K}$-algebra, but a (left) $\mathbb{K}$-*vector space*, and the *dimension* of subspaces will give us a new invariant.

We use the notion of dimension essentially via two results:

- Every submodule V (called *subspace*) of a vector space is given a dimension $\mathsf{dim}\, V$ and if $V \subseteq V'$, and $\mathsf{dim}\, V = \mathsf{dim}\, V'$ finite, then $V = V'$.
- From every generating set G of a subspace V of finite dimension, one can effectively extract a *basis*, that is, a free generating set of V.

For the rest of this section, $\mathbb{K}$ is a division ring.

4.1 Rank of a series

Definition 26. The *rank* $\mathsf{r}(s)$ of a series s of $\mathbb{K}\langle\langle A^* \rangle\rangle$ is the dimension of the subspace of $\mathbb{K}\langle\langle A^* \rangle\rangle$ generated by $\mathbf{R}_s$ the set of (left) quotients of s:

$$\mathsf{r}(s) = \mathsf{dim}\, \langle \mathbf{R}_s \rangle \ .$$

In this setting, and with no further ado, Theorem 23 becomes — since $\langle \mathbf{R}_s \rangle$ is obviously stable and contains s:

Theorem 27. *A series s of $\mathbb{K}\langle\langle A^* \rangle\rangle$ is recognisable if and only if $\mathsf{r}(s)$ is finite.* ∎

Let $\mathcal{A}$ be a $\mathbb{K}$-representation of dimension n that realises s.[4] From Property 19 follows that $\mathsf{r}(s)$ is smaller than, or equal to, $\mathsf{dim}(\mathsf{Im}\, \Phi_{\mathcal{A}})$ which is smaller than, or equal to, n. Hence the minimal dimension of a representation for s is $\mathsf{r}(s)$.

Definition 28. A representation of a recognisable series s is *reduced* if its dimension is equal to the rank of s.

From Proposition 25 follows that reduced representations do exist since we have:

Property 29. *With every basis of $\langle \mathbf{R}_s \rangle$ is associated a reduced representation of s.*

[4] In this context where we *compare* dimensions, it is more convenient they be integers rather than sets.

Conversely, reduced representations are characterised by the following statement.

Theorem 30. *A $\mathbb{K}$-representation $\mathcal{A}$ is reduced if and only if it is both* controllable *and* observable*, that is, if and only if $\Psi_{\mathcal{A}}$ is surjective, and $\Phi_{\mathcal{A}}$ injective.*

Proof. Let s be the series realised by $\mathcal{A}$. The morphism

$$\Phi_{\mathcal{A}} \circ \Psi_{\mathcal{A}} \colon \mathbb{K}\langle A^* \rangle \longrightarrow \mathbb{K}\langle\langle A^* \rangle\rangle \qquad \text{is such that} \qquad [\Phi_{\mathcal{A}} \circ \Psi_{\mathcal{A}}](w) = w^{-1}s$$

for every w in A^* and $\mathsf{Im}\,[\Phi_{\mathcal{A}} \circ \Psi_{\mathcal{A}}]$ is the subspace $\langle \mathbf{R}_s \rangle$. For the dimension of $\mathsf{Im}\,[\Phi_{\mathcal{A}} \circ \Psi_{\mathcal{A}}]$ be equal to n, the dimension of $\mathcal{A}$, it is necessary, and sufficient, that the dimension of both $\mathsf{Im}\,\Psi_{\mathcal{A}}$ and $\mathsf{Im}\,\Phi_{\mathcal{A}}$ be equal to n. The second equality holds if and only if the dimension of $\mathsf{Ker}\,\Phi_{\mathcal{A}}$ is zero. ∎

4.2 Reduction of a representation

It is not enough to know that reduced representations exist and to characterise them. We want to effectively compute them and, for that purpose, we establish the following.

Theorem 31. *A reduced representation of a recognisable series s is effectively computable from any representation that realises s with a procedure whose complexity is cubic in the dimension of the representation.*

For the rest of this section, let $\mathcal{A}$ be a $\mathbb{K}$-representation of A^* of dimension n, that realises the series $s = |\mathcal{A}|$. Let us first assume that, given $\mathcal{A}$, one can effectively compute a *basis* of the subspace $\mathsf{Im}\,\Psi_{\mathcal{A}}$ (this will be proved in the next subsection, where the complexity of the whole procedure will be established as well).

Proposition 32. *Let G be a basis of the state-space $\mathsf{Im}\,\Psi_{\mathcal{A}}$, of cardinal m. This basis determines a $\mathbb{K}$-representation $\mathcal{A}'$ of dimension m, conjugate to $\mathcal{A}$, and with the properties:*

(i) *$\Psi_{\mathcal{A}'}$ is surjective ($\mathcal{A}'$ is controllable);*

(ii) *if $\Phi_{\mathcal{A}}$ is injective, so is $\Phi_{\mathcal{A}'}$ (if $\mathcal{A}$ is observable, so is $\mathcal{A}'$).*

Proof. By Proposition 6 and with the notation set there, the existence of G, generating set of $\mathsf{Im}\,\Psi_{\mathcal{A}}$ of cardinal m, implies the one of a $\mathbb{K}$-representation $\mathcal{A}' = \langle\, J, \kappa_G, U \,\rangle$ of dimension m which is conjugate to $\mathcal{A} = \langle\, I, \mu, T \,\rangle$ by M_G, that is, such that:

$$I = \alpha_G(J) = J \cdot M_G\,, \quad \forall a \in A \quad \kappa_G(a) \cdot M_G = M_G \cdot \mu(a)\,, \quad U = M_G \cdot T\,.$$

Since G is a *basis*, $\mathsf{dim}(\mathsf{Im}\,\Psi_{\mathcal{A}}) = \mathsf{dim}(\mathbb{K}^G) = m$ and α_G is *injective*. The diagram of Figure 5, that will be shown to commute, helps in understanding the next sequences of equalities.

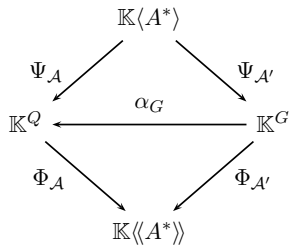


Figure 5: A diagram for Proposition 32

For every w in A^*, it then comes:

$$\begin{aligned}\Psi_{\mathcal{A}}(w) = I \cdot \mu(w) = \alpha_G(J) \cdot \mu(w) = J \cdot M_G \cdot \mu(w) \\ = J \cdot \kappa_G(a) \cdot M_G = \alpha_G(J \cdot \kappa_G(a)) = \alpha_G(\Psi_{\mathcal{A}'}(w)) \quad .\end{aligned}$$

Hence $\Psi_{\mathcal{A}} = \alpha_G \circ \Psi_{\mathcal{A}'}$. Since $\mathsf{dim}(\mathsf{Im}\,\Psi_{\mathcal{A}}) = m$ and α_G is injective, $\mathsf{dim}(\mathsf{Im}\,\Psi_{\mathcal{A}'}) = m$ and $\Psi_{\mathcal{A}'}$ is surjective.

Let x in $\Psi_{\mathcal{A}'}(A^*)$, that is, there exists w in A^* such that $x = \Psi_{\mathcal{A}'}(w)$.
Then $\Phi_{\mathcal{A}'}(x) = \Phi_{\mathcal{A}'}(\Psi_{\mathcal{A}'}(w)) = w^{-1}s$.

On the other hand, $\Phi_{\mathcal{A}}(\alpha_G(\Psi_{\mathcal{A}'}(w))) = \Phi_{\mathcal{A}}(\Psi_{\mathcal{A}}(w)) = w^{-1}s$, and then

$$\Phi_{\mathcal{A}'}(x) = \Phi_{\mathcal{A}}(\alpha_G(x)) \quad . \tag{4.1}$$

Since $\Psi_{\mathcal{A}'}(A^*)$ generates $\mathbb{K}^G$, (4.1) holds on the whole space $\mathbb{K}^G$ and $\Phi_{\mathcal{A}'} = \Phi_{\mathcal{A}} \circ \alpha_G$.
Since α_G is injective, if $\Phi_{\mathcal{A}}$ is injective, so is $\Phi_{\mathcal{A}}$. ∎

We now introduce the *transpose* of the representation $\mathcal{A}$, ${}^{\mathsf{t}}\mathcal{A} = ({}^{\mathsf{t}}T, {}^{\mathsf{t}}\mu, {}^{\mathsf{t}}I)$ where ${}^{\mathsf{t}}\mu(a) = {}^{\mathsf{t}}(\mu(a))$ for every a in A and it comes ${}^{\mathsf{t}}\mu(w) = {}^{\mathsf{t}}(\mu({}^{\mathsf{t}}w))$ for every w in A^*. We then have the following connection between $\mathcal{A}$ and ${}^{\mathsf{t}}\mathcal{A}$.

Remark 33. The use of the transpose of a $\mathbb{K}$-representation is not satisfactory as it is not well-defined when $\mathbb{K}$ is *not commutative*, a case that we want to cover. On the other hand, it is an easy shortcut, as it save the definition of the dual of every notion we have defined so far (state-space, control morphism, *etc.*). It is enough to say that it is legitimate for the case where $\mathbb{K}$ is commutative, which holds in all the forthcoming examples and exercices and that there exists a method to overcome the problem when needed (as in the case of Corollary 43 for instance).

Lemma 34. *If $\Psi_{{}^{\mathsf{t}}\mathcal{A}}$ is surjective, then $\Phi_{\mathcal{A}}$ is injective.*

Proof. If $\Phi_{\mathcal{A}}(x) = 0$ then $x \cdot \mu(w) \cdot T = 0$ for every w in A^* and x belongs to the orthogonal of the subspace generated by the vectors $\{\mu(w) \cdot T \mid w \in A^*\}$ which is of dimension n by hypothesis: thus $x = 0$. ∎

Proof of Theorem 31. Starting from a representation $\mathcal{A}$, we first compute a basis for the state-space of ${}^{\mathsf{t}}\mathcal{A}$ which determines a representation ${}^{\mathsf{t}}\mathcal{A}'$ such that $\Psi_{{}^{\mathsf{t}}\mathcal{A}'}$ is

surjective, and thus by Lemma 34, $\Phi_{\mathcal{A}'}$ is injective. We then compute a basis for the state-space of $\mathcal{A}'$ which determines a representation $\mathcal{A}''$ such that $\Psi_{\mathcal{A}''}$ is surjective and $\Phi_{\mathcal{A}''}$ is injective: $\mathcal{A}''$ is reduced. ∎

The proof of Theorem 31 will be complete when we have proved that basis for the state-spaces are effectively computable (with the ascribed complexity).

4.3 Effective computations

Word basis

Definition 35. We call *word basis* for $\mathcal{A}$ a prefix-closed subset P of A^* such that the set $\Psi_{\mathcal{A}}(P) = \{I \cdot \mu(p) \mid p \in P\}$ is a basis of $\mathsf{Im}\,\Psi_{\mathcal{A}}$.

Proposition 36. *Word basis for $\mathcal{A}$ do exist.*

Proof. If $I = 0$, $\mathsf{Im}\,\Psi_{\mathcal{A}}$ is the null vector space, of dimension 0 and the empty set (which is prefix-closed!) is a word basis. Assuming that I is non-zero, the family of prefix-closed subsets P of A^* such that $\{I \cdot \mu(p) \mid p \in P\}$ is a free subset of $\mathbb{K}^n$ is not empty since it contains at least the singleton $\{1_{A^*}\}$. Every such subset contains at most $k = \dim(\mathsf{Im}\,\Psi_{\mathcal{A}})$ elements and there exist thus maximal elements (for the inclusion order) in that family.

It remains to show that such a maximal element P is a word basis, that is, $\Psi_{\mathcal{A}}(P)$ generates $\mathsf{Im}\,\Psi_{\mathcal{A}}$. By way of contradiction, let w in A^* such that $I \cdot \mu(w)$ does not belong to $\langle \Psi_{\mathcal{A}}(P) \rangle$; the word w factorises in $w = pg$, with p in P, and we choose w in such a way that g is of minimal length. The word g is not empty: $g = ah$, with a in A, and $I \cdot \mu(w) = I \cdot \mu(pa) \cdot \mu(h)$. As P is maximal, $I \cdot \mu(pa)$ belongs to $\langle \Psi_{\mathcal{A}}(P) \rangle$ that is, $I \cdot (pa)\mu = \sum_{p_i \in P} x_i \, (I \cdot \mu(p_i))$. It then follows

$$I \cdot \mu(w) = \left(\sum_{p_i \in P} x_i \, (I \cdot \mu(p_i)) \right) \cdot \mu(h) = \sum_{p_i \in P} x_i \, (I \cdot \mu(p_i h)) \quad .$$

By the minimality of g, every $I \cdot \mu(p_i h)$ belongs to $\langle \Psi_{\mathcal{A}}(P) \rangle$: contradiction. ∎

In the sequel, we do not consider the trivial case $I = 0$ anymore.

If P is a non-empty prefix-closed subset of A^*, the *border of P* is the set:

$$C = PA \setminus P \quad .$$

As an example, the prefix-closed subset $\{1_{A^*}, b, ba\}$ and its border $\{a, bb, baa, bab\}$ are shown in Figure 6.

The following proposition and its proof exhibit the computation underlying Proposition 32.

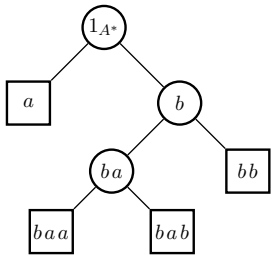


Figure 6: A prefix-closed subset and its border

Proposition 37. *Word basis for $\mathcal{A}$ are effectively computable, with complexity $O(dn^3)$, where d is the cardinal of A.*

Proof. We set $P_0 = \{1_{A^*}\}$ and $C_0 = \emptyset$. The algorithm to compute a word basis P can be written in the following manner.

If $E_k = (P_k A \setminus P_k) \setminus C_k$ is non-empty, choose an arbitrary w in E_k and decide whether $I \cdot \mu(w)$ belongs to $\langle I \cdot P_k \mu \rangle$.

(i) If not, then $P_{k+1} = P_k \cup \{w\}$ and $C_{k+1} = C_k$.

(ii) If so, then $P_{k+1} = P_k$ and $C_{k+1} = C_k \cup \{w\}$.

Set $k = k + 1$ and start again.

The algorithm terminates when E_k is empty and at that moment $C_k = P_k A \setminus P_k$ is the border of P_k. The algorithm must terminate since P_k has at most n elements, so $P_k \cup C_k$ has at most $\|A\| n + 1$ elements and this set grows by 1 at each step of the algorithm.

By construction, P_k is prefix-closed, and each element w of C_k is such that $I \cdot \mu(w)$ belongs to $\langle I \cdot \mu(P_k) \rangle$: when E_k is empty, P_k is maximal. ∎

Gaussian elimination The foregoing proofs all correspond to effective computations, assuming of course that the operations in $\mathbb{K}$ (addition, multiplication, taking the inverse) are effective. All the complexities that follow are calculated assuming that each operation in $\mathbb{K}$ has a fixed constant cost, independent of its operands.[5] Computations in $\mathbb{K}^n$ are based on the *Gaussian elimination* procedure.

Definition 38. A sequence of k vectors $(x^1, x^2, \ldots, x^k)$ of $\mathbb{K}^n$ is an *echelon system* if, for all i in $[k]$:

(i) $x^i{}_i = 1_{\mathbb{K}}$; (ii) $\forall j < i \qquad x^i{}_j = 0_{\mathbb{K}}$.

An echelon system is free and hence $k \leqslant n$. The following proposition is classic, at least for commutative fields, and its proof is not really different for division rings.

Proposition 39 (Gaussian elimination)**.** *Let $\mathbb{K}$ be a skew field and let us view $\mathbb{K}^n$ as a left vector space over $\mathbb{K}$. Let $S = (x^1, x^2, \ldots, x^k)$ be an echelon system and let y be a vector in $\mathbb{K}^n$.*

[5] It is to be acknowledged that this is a completely unrealistic assumption.

(i) *We can decide whether y is in $\langle S \rangle$, the subspace generated by S, and, in this case, compute effectively the coordinates of y in S.*

(ii) *If y is not in $\langle S \rangle$, we can compute effectively y' such that $S' = S \cup \{y'\}$ is echelon and generates the same subspace as $S \cup \{y\}$.*

The complexity of these operations (deciding whether y is in $\langle S \rangle$ and computing the coordinates of either y or y') is $O(kn)$.

From this proposition we deduce the effective nature of the assertions, constructions, and specifications used in the proofs of this section. More precisely:

Corollary 40. *Let S be a finite set of vectors of $\mathbb{K}^n$ and let y be in $\mathbb{K}^n$. We can:*

(i) *decide whether y belongs to $\langle S \rangle$;*

(ii) *extract effectively from S a basis T of $\langle S \rangle$;*

(iii) *compute effectively the coordinates in T of an (explicitly given) vector of $\langle S \rangle$.*

5 Applications of the reduction of recognisable series

5.1 Decidability of the equivalence

Even if a series has not a unique reduced representation (they are all *similar*), the existence of reduced representations implies the decidability of equivalence for automata with weights in a field.

Theorem 41. *The equivalence of recognisable series over A^* with coefficients in a (sub-semiring of a) skew field — and thus of rational series — is decidable, with a procedure which is cubic in the dimension of the representation of the series.*

Proof. Let $\mathbb{K}$ be a sub-semiring of a skew field $\mathbb{F}$. Two series s_1 and s_2 of $\mathbb{K}\mathrm{Rec}\, A^*$ are also in $\mathbb{F}\mathrm{Rec}\, A^*$ and $s_1 = s_2$ holds if and only if $(s_1 - s_2)$ is a series of $\mathbb{F}\mathrm{Rec}\, A^*$ of rank 0, and the rank of $(s_1 - s_2)$ can be computed effectively. ∎

This result, together with the well-known decidability of equivalence of classical Boolean automata, should not let us think that this is the universal status. For instance, the following holds.

Theorem 42 (Krob)**.** *The equivalence of recognisable series over A^* with coefficients in the semiring $\mathbb{M} = \langle \mathbb{N}, \mathsf{min}, + \rangle$ is undecidable.*

Theorem 41 has however far reaching and to some extent 'unexpected' consequences, as the following one, discovered by T. Harju and J. Karhumäki.

Corollary 43. *The equivalence of rational series over $A_1^* \times A_2^* \times \cdots \times A_k^*$ with coefficients in $\mathbb{N}$ is decidable.*

Proof. A series in $\mathbb{N}\mathrm{Rat}\,(A_1^*\times A_2^*\times\cdots\times A_k^*)$ is a series in $[\mathbb{N}\mathrm{Rat}\,(A_2^*\times\cdots\times A_k^*)]\mathrm{Rat}\,A_1^*$. By Theorem 46, the latter family is isomorphic to $[\mathbb{N}\mathrm{Rat}\,(A_2^*\times\cdots\times A_k^*)]\mathrm{Rec}\,A_1^*$ and the decidability of equivalence follows from Theorem 44. ∎

Theorem 44. *$\mathbb{N}\mathrm{Rat}\,(A_2^*\times\cdots\times A_k^*)$ is a sub-semiring of a skew field.*

This result is the direct consequence of a series of classical results in mathematics which we shall not present here.

cf. EAT, Sec. IV.7, p. 616

5.2 From the series to the representation

Another way to exploit Proposition 32, is by 'computing' the coefficients of a *reduced representation* of a recognisable series as a function of the coefficients of the series itself. Going from the series back to the representation does not so much correspond to an effective procedure as it expresses a fundamental property of recognisable series on a field (see an application with Theorem 46).

Proposition 45. *Let $\mathbb{K}$ be a skew field, s a $\mathbb{K}$-recognisable series of rank n, and $\langle I, \mu, T\rangle$ a reduced representation of s. There exist two sets of n words: $P = \{p_1, p_2, \ldots, p_n\}$ and $Q = \{q_1, q_2, \ldots, q_n\}$ (which we can choose to be respectively prefix-closed and suffix-closed) and two $n\times n$ matrices α_P and β_Q such that*

$$\forall w \in A^* \qquad \mu(w) = \alpha_P \cdot (\langle s, p_i\, w\, q_j\rangle) \cdot \beta_Q \ ,$$

where $(\langle s, p_i\, w\, q_j\rangle)$ denote the $n\times n$ matrix whose entry (i,j) is $\langle s, p_i\, w\, q_j\rangle$.

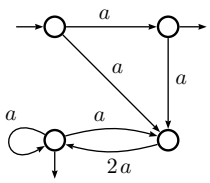
A remarkable application of this result is the following.

Theorem 46. *Let $\mathbb{K}$ be a (skew) field. If s is a $\mathbb{K}$-rational series with a finite image, then $k\,s^{-1}$ is rational for all k in $\mathbb{K}$.*

Proof. Let $\langle I, \mu, T\rangle$ be a reduced representation that recognises s. By Proposition 45, the image $\mu(A^*)$ is a *finite submonoid* of $\mathbb{K}^{Q\times Q}$ if s has a finite image and the conclusion follows. ∎

6 Exercises

1. Compute the reduced representation of the following $\mathbb{N}$-automaton.



2. Let $\mathcal{A}_1$ be the $\mathbb{Q}$-automaton on $\{a\}^*$ shown at Figure 7 (the unique letter a of the alphabet is not shown on the transitions of the figure). Compute a reduced automaton, equivalent to $\mathcal{A}_1$.

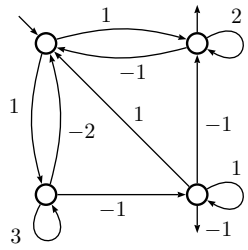


Figure 7: The $\mathbb{Q}$-automaton $\mathcal{A}_1$

3. Consider the minimal (Boolean) automaton of $\{a^n \mid n \equiv 0, 1, 2, 4 \pmod 7\}$ as an automaton with multiplicity in $\mathbb{Z}/2\mathbb{Z}$ and reduce it. Comment.

4. Let $\mathbb{F}$ be a field. Show that two $\mathbb{F}$-recognisable series over A^* are equal if and only if they coincide on all the words of length less than the sum of the dimensions of the representations which realise them.

Show the bound is sharp. [*Hint: consider the following two automata.*]

5. **Discriminating length.** We call the *discriminating length* between two non-equivalent (Boolean) automata $\mathcal{A}$ and $\mathcal{B}$ the length of a shortest word which is accepted by one and not the other. We write $L_{\mathsf{d}}(n, m)$ (resp. $L_{\mathsf{nd}}(n, m)$) for the maximum of the discriminating lengths when $\mathcal{A}$ and $\mathcal{B}$ have respectively n and m states and are deterministic (resp. and are non-deterministic).

(a) With methods relevant to Boolean automata, show that $L_{\mathsf{d}}(n, m) \leqslant n\,m$.

(b) Compute $L_{\mathsf{d}}(n, m)$.

(c) Give an upper bound for $L_{\mathsf{nd}}(n, m)$.

Lecture IV

Transducers (1)

The 2-tape Turing machine model

This part which spans over the last two lectures studies the model of finite automata 'with output' which are usualy called 'transducers'. They can be seen as (finite) automata over a direct product of free monoids as well as (finite) automata over a free monoid with multiplicities in the (rational) subsets of another free monoid, or of a direct product of free monoids. These two models are also equivalent to 'one-way' Turing machines with two or more tapes. In this lecture, we consider the first model only, the second one is subject of the next lecture.

Contents

Automata 'with output' are a very natural, even a necessary, extension of automata that 'read' sequences of symboles. Since the dawn of automata theory (that is, the second half of the fifties), kinds of such automata with output were studied: *Moore machines* in which the sequences of states reached in the course of the reading of a word are observed, *Mealy machines* in which an output letter is associated with every transition. These two models are indeed equivalent up to some adjustment. We start with a model which is strictly more general.

On the other hand, we could define *weighted transducers* - an even more general model, and more in line with the first three lectures. But their study is somewhat more difficult and requires to have the theory of Boolean transducers in background.

1 Definitions

In the sequel, A and B are two alphabets. The set $A^*\times B^*$ of pairs (u,v) with u in A^* and v in B^*, equipped with the product:

$$(u,v)\,(u',v') = (u\,u', v\,v')$$

is a monoid, whose identity element is $(1_{A^*}, 1_{B^*})$, most often denoted by $(1,1)$. The *length* of an element of $A^*\times B^*$ is the sum of the lengths of its components: $|(u,v)| = |u| + |v|$. The monoid $A^*\times B^*$ is *graded* (Definition I.32). Similarly, $A_1^*\times A_2^*\times\cdots\times A_k^*$, the set of k-tuples of words equipped with the componentwise product is a graded monoid.

1.1 Transducers

Definition 1. A *transducer* is an automaton over $A^*\times B^*$ or, more generaly, over $A_1^*\times A_2^*\times\cdots\times A_k^*$, that is, an automaton whose transitions are labelled with k-tuples of words.

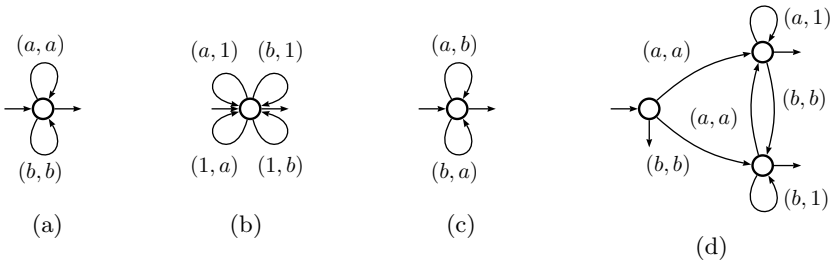
In (almost) all examples, $k = 2$. In the sequel, we also speak of 'pairs' rather than of 'k-tuple', unless stated otherwise.

A transducer is thus implicitely here a *Boolean automaton*[1] which can be denoted by $\mathcal{T} = \langle\, A^*\times B^*, Q, I, E, T\,\rangle$ where, as in the preceding lectures, Q is the state set, I and T are the sets of initial and final states respectively and where $E \subseteq Q\times(A^*\times B^*)\times Q$ is the set of transitions. Figure 1 shows four transducers.

We thus write $p \xrightarrow{(u,v)} q$ for a transition and

$$c = p_0 \xrightarrow{(u_1,v_1)} p_1 \xrightarrow{(u_2,v_2)} p_2\cdots p_{n-1} \xrightarrow{(u_n,v_n)} p_n$$

[1] We could have defined *weighted transducers* but their study is somewhat more complex and we need to know the theory of Boolean transducers first.



(a) (b) (c) (d)

Figure 1: Four transducers

for a computation of $\mathcal{T}$. The label of a computation is the product of the labels of its transitions and we write:

$$c = p_0 \xrightarrow{(u_1u_2\cdots u_n, v_1v_2\cdots v_n)} p_n \quad .$$

A computation is *successful* if its origin is an initial state and if its destination is a final state. A pair of words (u, v) in $A^* \times B^*$ is accepted by $\mathcal{T}$ if it is the label of a successful computation of $\mathcal{T}$. The *behaviour* of $\mathcal{T}$, denoted by $|\mathcal{T}|$, is the the set of pairs of words accepted by $\mathcal{T}$:

$$|\mathcal{T}| = \left\{ (u, v) \in A^* \times B^* \;\middle|\; \exists i \in I\,,\ \exists t \in T \quad i \xrightarrow[\mathcal{T}]{(u,v)} t \right\} \quad .$$

Examples 2. The transducer of Fig. 1 (a) accepts the set of pairs (u, u) where u is any word of $\{a, b\}^*$; the one of Fig. 1 (b) accepts the set of pairs (u, v) where u and v are any words of $\{a, b\}^*$; the one of Fig. 1 (c) accepts the set of pairs (u, v) where u is any word of $\{a, b\}^*$ and where v is obtained from u by replacing the a's by b's and the b's by a's; the one of Fig. 1 (d) accepts the set of pairs (u, v) where u is any word of $\{a, b\}^*$ and where v is obtained from u by replacing every block of a's by a unique a and every block of b's by a unique b.

The behaviour of a transducer is thus a subset of $A^* \times B^*$, that is, what we call a *relation between words*, or a *word relation*.

1.2 Word relations

Relations A relation θ from A^* to B^* is written (with a slight abuse) $\theta\colon A^* \to B^*$ and is *defined by its graph* $\widehat{\theta} \subseteq A^* \times B^*$. By definition, a relation from A^* to B^* associates with every word of A^* a subset of B^*:

$$\forall u \in A^* \quad \theta(u) = \left\{ v \in B^* \;\middle|\; (u, v) \in \widehat{\theta} \right\} \quad .$$

The underlying idea is that the first component of a pair (u, v) is an 'input' and that the second component is the 'output'. This point of view gives distinct roles to the two components of the pair but does not break the symmetry between them.

Inverse Indeed, A^* and B^* play symmetric roles by the way of the graph of θ and the *inverse relation* of θ,

$$\theta^{-1}\colon B^* \to A^* \ ,$$

is defined as the relation from B^* to A^* which has the same graph as θ, modulo the canonical identification between $A^*\times B^*$ and $B^*\times A^*$:

$$\forall v \in B^* \qquad \theta^{-1}(v) = \left\{ u \in A^* \;\middle|\; (u,v) \in \widehat{\theta} \right\} \ .$$

Additivity *By definition*, a relation is extended by *additivity*:

$$\forall L \subseteq A^* \quad \theta(L) = \bigcup_{u\in L} \theta(u)$$

and is thus viewed as an *application from* $\mathfrak{P}(A^*)$ *to* $\mathfrak{P}(B^*)$. Hence, the notion of relation implicitely carries with itself the property of additivity.

A contrario, complementation for instance, which associates with every subset of A^* a subset of A^*, *is not a relation* from A^* to itself.

Complement *By definition*, the *complement* of a relation $\theta\colon A^* \to B^*$ is the relation $\complement\theta\colon A^* \to B^*$ whose graph is the complement in $A^*\times B^*$ of the graph of θ:

$$\widehat{\complement\theta} = \complement\widehat{\theta} \qquad \text{that is,} \qquad \forall u \in A^* \qquad \big[\complement\theta\big](u) = \complement_{B^*}\theta(u) \ .$$

Domain and image If $\theta\colon A^* \to B^*$ is a relation, the *domain* and the *image* of θ are the projections of $\widehat{\theta}$ onto A^* and B^* respectively:

$$\mathsf{Dom}\,\theta = \left\{ u \in A^* \;\middle|\; \exists v \in B^* \quad (u,v) \in \widehat{\theta} \right\} \qquad \text{and}$$
$$\mathsf{Im}\,\theta = \left\{ v \in B^* \;\middle|\; \exists u \in A^* \quad (u,v) \in \widehat{\theta} \right\} \ .$$

Of course, $\mathsf{Dom}\,\theta^{-1} = \mathsf{Im}\,\theta$ and $\mathsf{Im}\,\theta^{-1} = \mathsf{Dom}\,\theta$. It also holds $u \notin \mathsf{Dom}\,\theta$ if and only if $\theta(u) = \emptyset$.

Generalisation to k-ary relations The relations of, or *predicats* on, $A_1^*\times A_2^*\times\cdots\times A_k^*$ — called *k-ary relations* — are defined by their graphs, which are subsets of $A_1^*\times A_2^*\times\cdots\times A_k^*$. What has been said above of *additivity*, inherent to the notion of relation, or of the *complement* of a relation, is naturally extended to k-ary relations. There are many ways[2] of 'currying' a relation of $A_1^*\times A_2^*\times\cdots\times A_k^*$ and the other notions: *domain*, *image*, *inverse* have a meaning only with respect to the way the 'input' and the 'output' components are chosen in the k-tuples elements de $A_1^*\times A_2^*\times\cdots\times A_k^*$. A natural generalisation is the one that could be denoted by $\theta\colon A_1^* \to A_2^*\times\cdots\times A_k^*$, where the input is a word of A_1^* and the output a $(k-1)$-tuple of words in $A_2^*\times\cdots\times A_k^*$.

[2]Properly speaking, 'currying' a function with several arguments consists in transforming it into a one-argument function which returns a function over the rest of the arguments.

1.3 Rational relations

The behaviour of a transducer is a subset of a direct product of free monoids; a transducer thus realises a relation, the one whose graph is the behaviour of this tranducer. For instance, the transducer of Fig. 1 (a) realises the identity function, the one of Fig. 1 (b) the universal relation. The Fundamental Theorem of Finite Automata yields a first characterisation of the relations realised by finite transducers. Let us first recall the definition of *rational subsets*, which holds in any monoid.

Definition 3. Rat $A^* \times B^*$ is the smallest family of subsets of $A^* \times B^*$ which contains the finite subsets and which is closed under the operations of sum, product and star.

Let us recall also that a subset (of a monoid) is rational if and only if it is *denoted* by a rational expression.

Definition 4. A relation $\theta\colon A^* \to B^*$ is *rational* if so is its graph, that is, if $\widehat{\theta} \in \operatorname{Rat} A^* \times B^*$.

From the definition itself follows:

Property 5. *The inverse of a rational relation is a rational relation.*

The Fundamental Theorem of Finite Automata applied to transducers yields:

Theorem 6 (Elgot & Mezei 1965)**.** *$\theta\colon A^* \to B^*$ is a rational relation if and only if $\widehat{\theta} = |\mathcal{T}|$ where $\mathcal{T}$ is a finite transducer over $A^* \times B^*$.*

If the label of every transition of a transducer $\mathcal{T}$ is mapped onto its first (resp. its second) component, one gets an automaton whose transitions are labelled by words — possibly *empty* — and which accepts the domain (resp. the image) of the relation realised by $\mathcal{T}$. This implies the following.

Corollary 7. $\theta\colon A^* \to B^*$ *rel. rat.* $\implies$ $\mathsf{Dom}\,\theta \in \operatorname{Rat} A^*$, $\mathsf{Im}\,\theta \in \operatorname{Rat} B^*$.

2 Working on the model and examples

The converse implication of Theorem 6 can, and must, be made more precise. In order to deal efficiently with transducers, it is convenient to have indeed a more constrained definition that does not diminish the power of the model, and also to be able to enrich it without making it more powerful.

2.1 Normalisation

The alphabet A *freely* 'generates' A^* since every word of A^* is the product of a *unique* sequence of letters of A. The set $(A \times \{1_{B^*}\}) \cup (\{1_{A^*}\} \times B)$ generates $A^* \times B^*$

since every pairs in $A^* \times B^*$ is the product of sequences of elements in $(A \times \{1_{B^*}\}) \cup (\{1_{A^*}\} \times B)$, but these sequences are not unique (in general):

$$(a\,b, b\,a\,b) = (a,1)\,(1,b)\,(1,a)\,(b,1)\,(1,b) = (1,b)\,(a,1)\,(b,1)\,(1,a)\,(1,b) \quad .$$

One can also take $(A \times \{1_{B^*}\}) \cup (\{1_{A^*}\} \times B) \cup (A \times B)$ as generating set of $A^* \times B^*$; it allows to have shorter decomposition sequences:

$$(a\,b, b\,a\,b) = (a,b)\,(b,a)\,(1,b) = (1,b)\,(a,a)\,(b,b) \quad .$$

The automata over A^* are defined with transitions labelled in A and it is known that the model is not more powerful, that is, does not accept more languages, if one allows labels in the whole A^*. For transducers, we follow a reverse process: they are defined with transitions whose labels are taken in the whole $A^* \times B^*$, and one shows that the model is not less powerful, that is, does not accept fewer relations, if the set of authorised labels is constrained.

Definition 8. (i) A transducer over $A^* \times B^*$ whose labels are in

$$(A \times \{1_{B^*}\}) \cup (\{1_{A^*}\} \times B)$$

is called a *normalised transducer*.

(ii) A transducer over $A^* \times B^*$ whose labels are in

$$(A \times \{1_{B^*}\}) \cup (\{1_{A^*}\} \times B) \cup (A \times B)$$

is called a *subnormalised transducer*.

The transducers (a), (c) and (d) of Figure 1 are subnormalised, the transducer (b) is normalised.

Proposition 9.

Every transducer is equivalent to a normalised (or subnormalised) transducer.

Proof. The process for transforming an arbitrary transducer into a normalised (or subnormalised) one is the same as in the case of automata over A^* labelled with words. It starts with the replacement of every transition whose label (u, v) is of length $\ell = |u|+|v|$ greater than 1 by ℓ transitions labelled in $(A \times \{1_{B^*}\}) \cup (\{1_{A^*}\} \times B)$ or by k transitions labelled in $(A \times \{1_{B^*}\}) \cup (\{1_{A^*}\} \times B) \cup (A \times B)$, with k contained between $\mathsf{max}(|u|, |v|)$ and ℓ.

In order to get a normalised, or subnormalised, transducer, it is necessary to eliminate the transitions which are labelled with $(1, 1)$, the identity element of $A^* \times B^*$, and whose presence is not ruled out by Definition 1. This elimination is the result of a classical algorithm which can be described in a slightly more general framework and which is worth to be explicitely given as it will be used later in another construction.

Let M be a monoid. An *automaton over* M is a graph whose transitions are labelled with elements of M. Such an automaton is said to be *proper* if none of its transitions are labelled with the identity element of M.

Theorem 10.

Every finite automaton over M is equivalent to a proper *finite automaton.*

Proof. Let $\mathcal{A} = \langle M, Q, I, E, T \rangle$ be an automaton over M. We write: $E = F \cup S$ where S is the set of *spontaneous* transitions of $\mathcal{A}$, that is, the transitions labelled with 1_M. Without loss of generality, we assume that S is a *transitive* subgraph of $\mathcal{A}$: adding the transitions corresponding to the transitive closure of the set of spontaneous transitions in $\mathcal{A}$ may indeed change the computations of $\mathcal{A}$, but not their labels. A computation of $\mathcal{A}$ is then of the form:

$$c = p_0 \xrightarrow{m_1} p_1 \xrightarrow{m_2} p_2 \cdots p_{n-1} \xrightarrow{m_n} p_n \ ,$$

and, thanks to the hypothesis on S, no two consecutive m_i are both equal to 1_M. Let $\mathcal{B} = \langle M, Q, J, G, T \rangle$ be the automaton defined by:

$$G = F \cup \{(p, m, r) \mid \exists q \in Q \quad (p, m, q) \in F \text{ and } (q, 1_M, r) \in S\} \quad \text{and}$$
$$J = I \cup \{j \mid \exists i \in I \quad (i, 1_M, j) \in S\}$$

which is then easily seen to be equivalent to $\mathcal{A}$. ∎

This construction completes the proof of Proposition 9. ∎

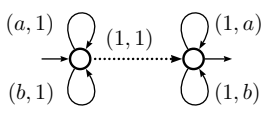
This result is also interesting in that it allows the use of spontaneous transitions for the construction of compact transducers as we see in the next series of examples.

Remark 11. The construction described in the proof of Theorem 10 can be called a *forward closure* as the new transitions are built with the spontaneous transitions that *follow* transitions labelled with elements different from the identity element. Another proper automaton equivalent to $\mathcal{A}$ can obviously be built by means of a dual *backward closure*.

2.2 Examples

Examples 12. (i) **Universal relation, direct product of rational sets**.

The universal relation, that is, the relation whose graph is the whole $A^* \times A^*$, is realised by the transducer of Figure 1 (b). It is also realised by the transducer below, in which every element of $A^* \times A^*$ is *the label of a unique computation* (and which demonstrates the benefit of spontaneous transitions).



If K is a language of A^*, accepted by $\mathcal{A}$, and L a language of B^*, accepted by $\mathcal{B}$, we transform $\mathcal{A}$ into a transducer $\mathcal{A}'$ by replacing the label 'a' of every transition by '$(a,1)$' and $\mathcal{B}$ into a transducer $\mathcal{B}'$ by replacing the label 'b' of every transition by '$(1,b)$' respectively, as shown below.

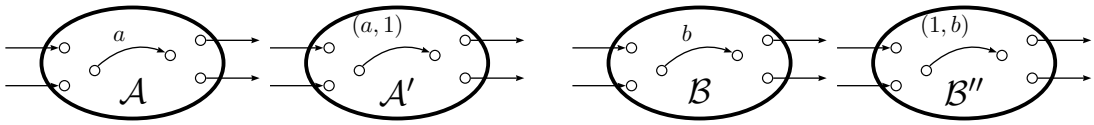


Figure 2: Transformation of automata into transducers

The transducer of Figure 2 realises the relation whose graph is $K \times L$.

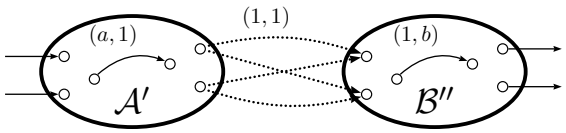
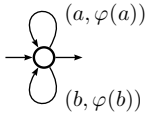


Figure 3: A transducer for $K \times L$

(ii) **Identity, morphisms**. The identity, that is, the relation whose graph is $\{(w,w) \mid w \in A^*\}$, is realised by the transducer of Figure 1 (a). A *morphism* $\varphi\colon A^* \to B^*$ is realised by the transducer below.



(iii) **Intersection with a rational set**. If K is a language of A^*, the *intersection with* K is a relation from A^* into itself, denoted by ι_K, and defined by:

$$\forall w \in A^* \qquad \iota_K(w) = \begin{cases} w & \text{if } w \in K \\ \text{undefined (or } \emptyset) & \text{otherwise.} \end{cases}$$

If K is accepted by $\mathcal{A}$, the relation ι_K is realised by the transducer $\mathcal{A}'''$ obtained from $\mathcal{A}$ by remplacing the label 'a' of every transition by '(a,a)', as shown below.

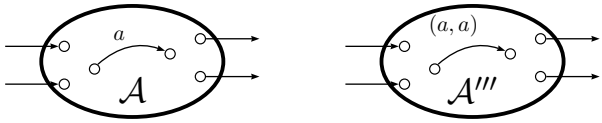
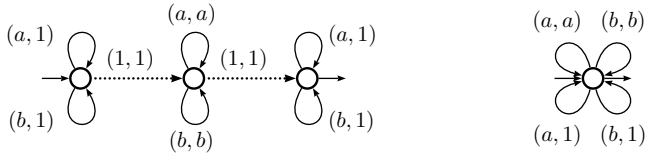


Figure 4: A transducer for ι_K

(iv) **Factors, subwords**. The relation from A^* into itself which associates with every word its *factors* is realised by the transducer shown at Figure 5 (a); the one which associates its *subwords* is realised by the transducer shown at Figure 5 (b).



(a) (b)

Figure 5: Factors and subwords

(vi) **Operations on numbers written in base** p. When a base p is chosen, numbers (non-negative integers) are written[3] on the alphabet $A_p = \{0, 1, \ldots, p-1\}$ and operations on numbers are functions from A_p^*, or $(A_p^*)^2$, or $(A_p^*)^3$, *etc.* into A_p^*. Some are realised by finite transducers. Figure 6 shows the example of the (integer) *division* by a fixed integer k, in the case where $p = 2$ and $k = 3$.

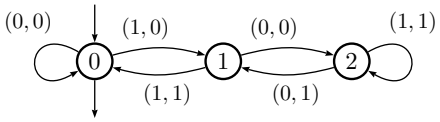


Figure 6: Integer division by 3 of numbers written in binary

2.3 Extension

2.3.1 k-ary transducers

As mentioned in Definition 1, a transducer may be an automaton over a direct product $A_1^* \times A_2^* \times \cdots \times A_k^*$ of k free monoids, not only an automaton over a direct product $A^* \times B^*$ of two free monoids. And as it has been mentioned as well, there are multiple ways of 'currying' a relation over $A_1^* \times A_2^* \times \cdots \times A_k^*$. From a theoretical point of view, it may be interesting to see such a relation as a function from A_1^* into the subsets of $A_2^* \times \cdots \times A_k^*$. From a practical point of view, it is more common to see the first $k-1$ components of a k-tuple as the 'input' and the k-th component as the result, that is, to view the relation as a map from $A_1^* \times A_2^* \times \cdots \times A_{k-1}^*$ into $\mathfrak{P}(A_k^*)$.

Example 13. Product in A^*. The relation $\pi\colon A^* \times A^* \to A^*$ which associates with every pair of words their product: $\pi(u, v) = uv$ for every u, v in A^*, is realised by the transducer below.

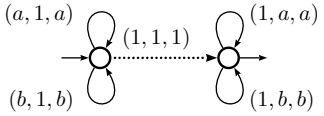


Figure 7: A 3-ary transducer for the product of words

[3]When alphabets of digits are used, the empty word is written ε.

The notions of *normalised* or *subnormalised* k-ary transducers are defined in an obvious manner and every k-ary transducer is equivalent to a normalised or subnormalised one (as is the transducer above if its spontaneous transition is eliminated). It is useful to have this possible extension from 2 to k free monoids in mind but, as already said, we almost exclusively consider transducers over $A^* \times B^*$ in the sequel.

2.3.2 Left transducers, right transducers

A second variation on the model of transducers concerns the *direction of reading*. When we wrote that the *label of a computation is the product of the labels of the transitions* that make this computation (in an automaton or a transducer), it seemed understood that this product be from left to right. This corresponds to the *reading from left to right* in the machine model described in the next subsection. A reverse convention would have been as justified. There are even cases — as is Example 14 below — for which it is more natural.

This problem may be solved by means of *transposition*. The *transpose*, or *mirror image*, of a word $w = a_1\, a_2 \cdots a_n$ is the word ${}^{\mathsf{t}}w = a_n \cdots a_2\, a_1$; the transpose of a pair (u, v) is the pair $({}^{\mathsf{t}}u, {}^{\mathsf{t}}v)$. The transpose of an automaton, or of a transducer, $\mathcal{A} = \langle\, Q, I, E, T \,\rangle$, is the automaton, or the transducer, ${}^{\mathsf{t}}\mathcal{A} = \langle\, Q, T, {}^{\mathsf{t}}E, I \,\rangle$, avec

$$ {}^{\mathsf{t}}E = \{(p, {}^{\mathsf{t}}x, q) \mid (q, x, p) \in E\} \quad . $$

A word w is accepted by $\mathcal{A}$ in a *right-to-left reading* if and only if ${}^{\mathsf{t}}w$ is accepted by ${}^{\mathsf{t}}\mathcal{A}$ in a *left-to-right reading*. A word v belongs to the image of a word u in the relation realised by a transducer $\mathcal{A}$ in a *right-to-left reading* if and only if ${}^{\mathsf{t}}v$ belongs to the image of ${}^{\mathsf{t}}v$ par ${}^{\mathsf{t}}\mathcal{A}$ in a *left-to-right reading*.

In this way, it is seen that the inversion of the reading direction does not change the power of the model and does not bring anything new (as far as we have the transposition operator at hand). In some cases however, it may be simpler, more convenient or natural, to consider transducers *that read from right to left*, for instance when the transposed transducer is *input deterministic* as in Example 14 (what is called *right sequential transducer* in the last lecture).

Example 14. Addition in base 2. Let $A_2 = \{0, 1\}$. The map which associates with every pair (u, v) of words of $A_2 \times A_2$, that are the binary representations of the integers $\overline{u}$ and $\overline{v}$, the binary representation of $\overline{u} + \overline{v}$ is realised by the transducer of Figure 8 when it reads pairs *from right to left* (which is the usual way to perform addition indeed) and with the convention that the two words u and v are *justified on the right* and that the shorter one is padded with a sufficient number of '0' on the left to be of the same length as the longer one and, finally, that a last '0' is added on the left to both words in order to allow a last transition toward the final state (if necessary).

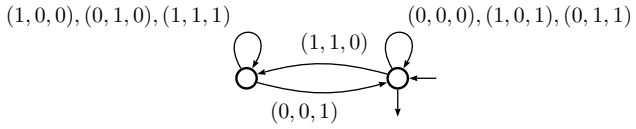


Figure 8: A 3-ary right transducer for the binary addition

2.4 Transducers as machines

Modelling a finite Boolean automaton as a '1-way Turing machine' leads naturally to a generalisation of the model that features 'several tapes'. The machine consists of a finite state control unit and *several* tapes. The control unit is connected to every tape by a *reading head* (*cf.* Figure 9).

At every step of the computation, the control unit 'chooses', according to its state p, a tape on which it 'reads' and, depending on the symbol a read on the tape, jumps in state q and moves the reading head on that tape to the next cell on the right.[4] As reading heads are moved *always in the same direction*, this type of mahine is called *1-way* Turing machine.

At the beginning of a computation, a word is written on each of the k tapes, every reading head stays on the first cell of its tape and the control unit is in a distinguished state called *initial*. After a succession of steps, a computation ends if every reading head has reached on its tape the cell that contains the *end-of-tape* symbol. The computation is successful if at the end of the computation the control unit is in a state called *final*. A k-tuple of words is accepted by the machine if it can be read by a successful computation.

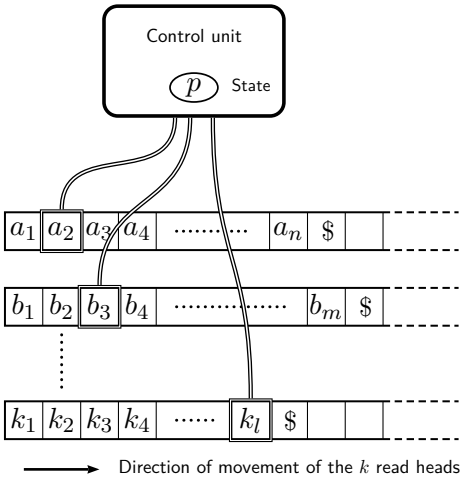


Figure 9: A k-tape 1-way Turing machine

Finite transducers over $A^* \times B^*$ are *strongly equivalent* to 1-way 2-tape Turing

[4]Other computation rules for such a device are possible. For instance, the choice of the read tape and of the destination state may depend not only on the state p but also on the symbols read on all tapes. All such definitions prove to be indeed equivalent.

machines (1W2TTM) in the sense that for every transducer one can build such a machine which is not only equivalent (that is, accepts the same pair of words) but such that there is a bijection between their successful computations and *vice versa*. The generalisation to transducers over $A_1^* \times A_2^* \times \cdots \times A_k^*$ and to 1-way k-tape Turing machines is tedious but conceals no difficulties.

3 Some facts

We review a series of negative results concerning rational sets of direct products of free monoids, hence rational relations. We end with an essential positive result that will be developped in Section 5: the closure by composition of rational relations. In the sequel, we call the finite transducers simply *transducers*.

3.1 Intersection, complement

In contrast with rational *languages*, rational *relations* are not closed under *intersection* and hence under *complement*.

Fact 15. $R, S \in \mathrm{Rat}\, A^* \times B^* \quad \not\Longrightarrow \quad R \cap S \in \mathrm{Rat}\, A^* \times B^*$.

Example 16. The behaviours of transducers[5] of Figure 10 are:

$$\big|\mathcal{V}_1\big| = \{(a^n b^m, c^n) \mid n, m \in \mathbb{N}\} \qquad \text{and} \qquad \big|\mathcal{W}_1\big| = \{(a^n b^m, c^m) \mid n, m \in \mathbb{N}\} \quad .$$

Hence $$\big|\mathcal{V}_1\big| \cap \big|\mathcal{W}_1\big| = \{(a^n b^n, c^n) \mid n \in \mathbb{N}\} \notin \mathrm{Rat}\, \{a, b\}^* \times \{c\}^*$$

since $$\mathsf{Dom}\left(\big|\mathcal{V}_1\big| \cap \big|\mathcal{W}_1\big|\right) = \{a^n b^n \mid n \in \mathbb{N}\} \notin \mathrm{Rat}\, \{a, b\}^* .$$

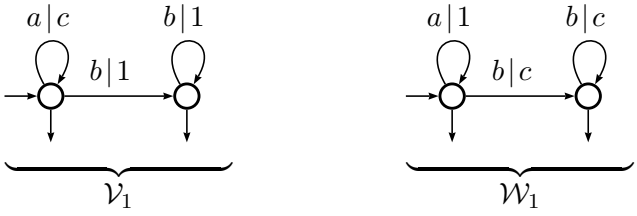


Figure 10: Transducers $\mathcal{V}_1$ and $\mathcal{W}_1$ over $\{a, b\}^* \times \{c\}^*$

Corollary 17. $\mathrm{Rat}\, A^* \times B^*$ *is not closed under complement.*

It holds nevertheless:

Proposition 18. *The complement of the identity is a rational relation.*

The proof reduces to the construction of the transducer of Figure 11 (and to the verification that its behaviour is indeed the complement of the identity).

[5] From this example on, we write $a \,|\, b$ instead of (a, b) for the labels of transitions, in order to lighten notation.

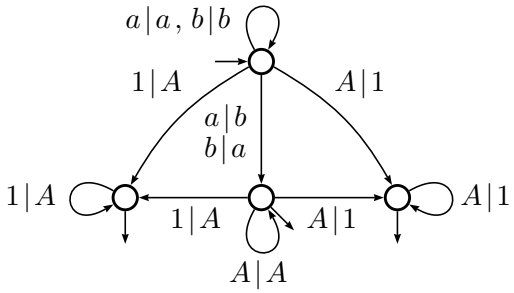


Figure 11: A transducer for the complement of the identity

3.2 Equivalence

A fundamental property of finite automata over a free monoid is that their *equivalence* is decidable, that is, there exists an algorithm which computes whether two such automata accept the same language. This property does not extend to rational relations.

Theorem 19 (Rabin & Scott 1959)**.** *Let* $R, S \in \operatorname{Rat} A^* \times B^*$, $\|A\|, \|B\| \geqslant 2$. *It is undecidable whether* $R \cap S = \emptyset$ *or not.*

It follows:

Theorem 20 (Fischer & Rozenberg 1968)**.**
The equivalence of finite transducers is undecidable.

These two negative results are established in the next section. Their statements leave open the status of the same questions in the cases where $\|A\| \geqslant 2, \|B\| = 1$ on one hand-side and $\|A\| = \|B\| = 1$ on the other. The first case exhibits an interesting separation between the two above statements.

Theorem 21 (Gibbons & Rytter 1986)**.**
Let $R, S \in \operatorname{Rat} \{a, b\}^* \times \{c\}^*$. *It is decidable whether* $R \cap S = \emptyset$ *or not.*

Theorem 22 (Ibarra 1978 – Lisovik 1979)**.**
The equivalence of finite transducers over $\{a, b\}^* \times \{c\}^*$ *is undecidable.*

The second case pertains to a completely different theory. It is noticed that $\{a\}^* \times \{b\}^*$ is isomorphic to $\mathbb{N}^2$, the free commutative monoid with two generators. And it holds:

Theorem 23 (Ginsburg & Spanier 1966)**.**
$\operatorname{Rat} \mathbb{N}^k$ *is an effective Boolean algebra, for every integer* k.

The proof of these last three results exceeds the program of these lectures (*cf.* EAT). We end this negative list with a positive result, not so much for cheering up, but because we need it in the next section for the proof of undecidability results.

3.3 Composition

The composition of functions directly extends to the one of (2-ary) relations.

Definition 24. Let $\theta\colon A^* \to B^*$ et $\sigma\colon B^* \to C^*$ two relations. The *composition* of θ and σ is the relation

$$\sigma \circ \theta\colon A^* \to C^* \qquad \text{defined by} \qquad \forall u \in A^* \qquad [\sigma \circ \theta](u) = \sigma(\theta(u)) \quad .$$

The composition of relations can be defined (or expressed) by means of their graph:

$$\widehat{\sigma \circ \theta} = \left\{ (u, w) \in A^* \times C^* \;\middle|\; \exists v \in B^* \quad (u, v) \in \widehat{\theta} \text{ and } (v, w) \in \widehat{\sigma} \right\} \quad .$$

Theorem 25 (Elgot & Mezei 1965)**.**
The composition of two rational relations is a rational relation.

We come back to, and establish, this fundamental result in Section 5.

Remark 26. In contrast with most of the other notions, composition of 2-ary relations does not generalise in a straightforward manner: one has to specify which component(s) of the first relation are considered to be the 'result' and these components have to be the 'input' of the seconde relation. And Theorem 25 generalises if the result is *a word* only, that is, if there is *only one* component in the result.

For instance, if θ is a rational relation over $A^* \times B^* \times C^*$ and σ a rational relation over $C^* \times D^*$, that is, θ is seen as $\theta\colon A^* \times B^* \to C^*$ and σ as $\sigma\colon C^* \to D^*$, then $\sigma \circ \theta$ is a rational relation. But if π is a rational relation over $B^* \times C^* \times D^*$ seen as $\pi\colon B^* \times C^* \to D^*$ and θ is seen as $\theta\colon A^* \to B^* \times C^*$, then $\pi \circ \theta$ is a relation over $A^* \times D^*$ which is not necessarily a rational relation.

4 Undecidability results

The undecidable property par excellence is the 'halting problem for a Turing machine'. But one can take as a basis any other propery already proved to be undecidable. The one we shall use in the sequel, because it is simpler to state, and easier to deal with in connection with automata, is known as the 'Post Correspondence Problem'.

The Post Correspondence Problem (PCP)

Let B be an alphabet with at least two letters. Given an integer k and two sets of k words of B^*: $\{u_1, u_2, \ldots, u_k\}$ et $\{v_1, v_2, \ldots, v_k\}$, does there exist a sequence of indices $i_1, \ldots, i_p$ in $[k]$ such that

$$u_{i_1} u_{i_2} \cdots u_{i_p} = v_{i_1} v_{i_2} \cdots v_{i_p} \,?$$

Theorem 27 (Post 1946)**.** *(PCP) is recursively undecidable.*

This statement holds for the problem in full generality. If one looks for its status according to the number k that allows to formulate an instance, the situation is more complex. Let (PCP_k) be the above problem in which the integer k is fixed. It is known that (PCP_2) is decidable and, since recently, that (PCP_k) is undecidable for $k \geqslant 5$. The status of (PCP_k) is still open for k equal to 3 or 4.

Translation in the vocabulary of Language and Automata Theory

The reason for our choice is that (PCP) can be easily expressed in terms of *morphisms between free monoids.*

If $U = \{u_1, u_2, \ldots, u_k\}$ is given, we write: $A_k = \{1, 2, \ldots, k\}$, and

$$\tau_U \colon A_k^* \to B^* \qquad \text{for the morphism defined by} \quad \tau_U(i) = u_i \quad \text{for every } i \text{ in } [k] \ .$$

Similarly, if $V = \{v_1, v_2, \ldots, v_k\}$, we write: $\tau_V \colon A_k^* \to B^*$ the morphism defined by $\tau_V(i) = v_i$ for every i in $[k]$. A 'sequence of indices' is a word of A_k^* and (PCP) is rephrased into:

does there exist a word w in A_k^* such that $\tau_U(w) = \tau_V(w)$?

Theorem 27 then becomes:

Theorem 28. *Let θ and $\mu \colon A^* \to B^*$ be two morphisms.*
It is undecidable whether there exists w in A^* such that $\theta(w) = \mu(w)$ or not.

Proof of Theorem 19. Let U and V be two sets of k words of B^* which produce an undecidable instance of (PCP) and $\tau_U \colon A_k^* \to B^*$ and $\tau_V \colon A_k^* \to B^*$ the corresponding morphisms.

To state that it is undecidable whether there exists w in A_k^* such that $\tau_U(w) = \tau_V(w)$ is equivalent as to state that it is undecidable whether

$$\widehat{\tau_U} \bigcap \widehat{\tau_V} = \emptyset \ ,$$

and τ_U and τ_U are rational relations (Example 2(ii)). It remains to show that Theorem 19 holds for an alphabet $A = \{a, b\}$ with two letters only.

Let $\kappa \colon A_k^* \to A^*$ an *injective morphism* (defined, for instance, by $\kappa(i) = a^i b$). By Theorem 25, $\tau_U \circ \kappa^{-1}$ and $\tau_V \circ \kappa^{-1}$ are rational relations and, since κ is injective, it holds:

$$\widehat{\tau_U \circ \kappa^{-1}} \bigcap \widehat{\tau_V \circ \kappa^{-1}} = \emptyset \quad \Longleftrightarrow \quad \widehat{\tau_U} \bigcap \widehat{\tau_V} = \emptyset \ . \qquad \blacksquare$$

Theorem 20 is a direct consequence of the following, more precise, statement.

Theorem 29. *Let $R \in \mathrm{Rat}\, A^* \times B^*$, $\|A\|, \|B\| \geq 2$.*
It is undecidable whether $R = A^* \times B^*$ or not.

We first prove:

Lemma 30. *Let* $\theta\colon A^* \to B^*$ *be a* functional *rational relation. Then* $\complement\theta\colon A^* \to B^*$ *is a rational relation.*

Proof. Let χ be the complement of the identity on B^*, a rational relation by Proposition 18. We have:

$$\widehat{\complement\theta} = [(A^* \setminus \mathsf{Dom}\,\theta)\times B^*] \ \cup \ \widehat{\chi\circ\theta} \ .$$

The first term of the union is rational (Example 2(i)) and so is the second one by Theorem 25. ∎

Proof of Theorem 29. With the notation of the proof of Theorem 19, $\tau_U \circ \kappa^{-1}$ and $\tau_V \circ \kappa^{-1}$ are functional rational relations and it holds:

$$\complement\left(\widehat{\tau_U \circ \kappa^{-1}}\right) \bigcup \complement\left(\widehat{\tau_V \circ \kappa^{-1}}\right) = A^*\times B^* \quad\Longleftrightarrow\quad \widehat{\tau_U \circ \kappa^{-1}} \bigcap \widehat{\tau_V \circ \kappa^{-1}} = \emptyset \ . \ \blacksquare$$

5 Composition and evaluation

The closure by composition of rational relations (Theorem 25) is a fundamental property, as is the consequence we deduce from it: the Evaluation Theorem (Theorem 35).[6] Together, they make of rational relations a powerful tool for the classification of formal languages. But above all, they give its consistency to the model of transducers.

5.1 The Composition Theorem

Theorem 25 (Elgot & Mezei 1965)**.**

$$\theta\colon A^* \to B^* \ , \quad \sigma\colon B^* \to C^* \quad \textit{rat. rel.} \quad\Longrightarrow\quad \sigma\circ\theta\colon A^* \to C^* \quad \textit{rat. rel.}$$

Proof. Let $\mathcal{T} = \langle\, A^*\times B^*, Q\, I, E, T\,\rangle$ and $\mathcal{S} = \langle\, B^*\times C^*, R, J, F, U\,\rangle$ be two *subnormalised* transducers which realise θ and σ respectively. We define a *composition product* of transducers $\mathcal{U} = \mathcal{S}\circ\mathcal{T}$ by:

$$\mathcal{U} = \langle\, A^*\times C^*, Q\times R, I\times J, G, T\times U\,\rangle \qquad \text{with}$$

$$G = \left\{(p,r) \xrightarrow{x|y} (q,s) \;\middle|\; \exists b\in B\,,\ \exists p \xrightarrow{x|b} q \in E\,,\ \exists r \xrightarrow{b|y} s \in F \quad x\in A\cup 1\,,\ y\in C\cup 1\right\}$$
$$\bigcup \left\{(p,r) \xrightarrow{a|1} (q,r) \;\middle|\; \exists p \xrightarrow{a|1} q \in E \qquad \forall r\in R\right\}$$
$$\bigcup \left\{(p,r) \xrightarrow{1|c} (p,s) \;\middle|\; \exists r \xrightarrow{1|c} s \in F \qquad \forall p\in Q\right\} \ .$$

[6] In the next lecture, we proceed in the reverse way: we first establish the Evaluation Theorem from wich we deduce the Composition Theorem.

By induction on the length of the computation, it is verified that:

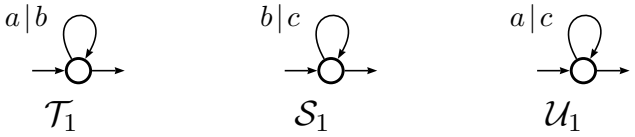
$$(p,r) \xrightarrow[\mathcal{U}]{u|w} (q,s) \quad \text{if and only if} \qquad \exists v \qquad p \xrightarrow[\mathcal{T}]{u|v} q \quad \text{and} \quad r \xrightarrow[\mathcal{S}]{v|w} s \ ,$$

which establish $|\mathcal{U}| = \widehat{\sigma \circ \theta}$. ■

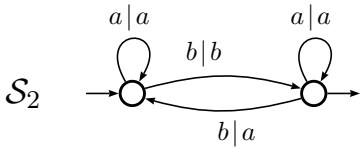
This composition product may yield a transducer $\mathcal{U}$ with some transitions that are labelled by $1|1$ (in the first group of transitions of G, when x *and* y are both equal to 1). These spontaneous transitions are eliminated (by 'backward' or 'forward' closure, for instance) in order to obtain a subnormalised transducer. In the sequel, it will be this[7] subnormalised transducer which will be denoted by $\mathcal{U} = \mathcal{S} \circ \mathcal{T}$.

Example 31 (trivial)**.**

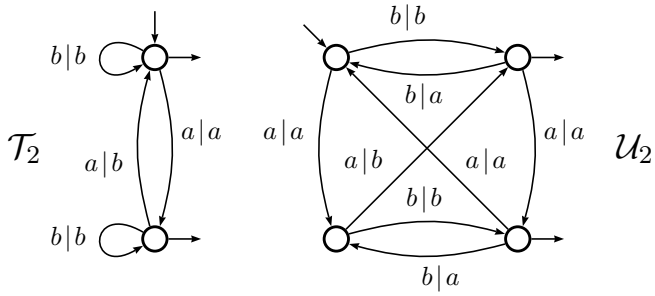
$\mathcal{U}_1 = \mathcal{S}_1 \circ \mathcal{T}_1$



Example 32 (less trivial but still simple)**.**



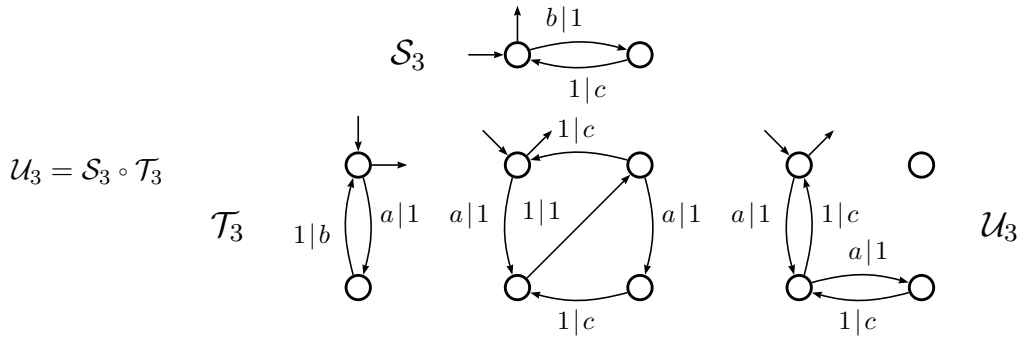
$\mathcal{U}_2 = \mathcal{S}_2 \circ \mathcal{T}_2$



These two examples show 'letter-to-letter transducers'. More general examples will be considered in the exercises.

Remark 33. If we consider the *normalised* transducers $\mathcal{T}_3$ and $\mathcal{S}_3$ that are equivalent to $\mathcal{T}_1$ and $\mathcal{S}_1$ respectively, a spontaneous transition appears in the course of the construction of the composition product $\mathcal{U}_3 = \mathcal{S}_3 \circ \mathcal{T}_3$. It is also important to note that in this case also, the *multiplicity* of computations is *not preserved*. A more elaborate construction (that is, a more sophisticated composition product) allows us to overcome this problem.

[7] A slight abuse, as $\mathcal{S} \circ \mathcal{T}$ is not completely determined since we have left the choice open for the closure.



Remark 34. It is possible to define (finite) transducers on direct products of *non free* monoids and hence *rational relations* between non necessarily free monoids: relations from M to N whose graph is in $\operatorname{Rat} M \times N$.

Two such relations $\theta\colon M \to N$ and $\sigma\colon N \to P$ can be then be composed. The construction of the proof of Theorem 25 still holds true — and the composition is a rational relation — as long as N is a free monoid B^*, but the composition may well be a non rational relation *if N is not a free monoid*.

For instance, let $\theta\colon \{a\}^* \to a^* \times b^*$ be the morphism defined by $\theta(a) = (a, b)$ and $\sigma\colon a^* \times b^* \to \{a, b\}^*$ the relation whose graph is $\widehat{\sigma} = \big((a, 1), a\big)^* \big((1, b), b\big)^*$. Then $\sigma((a^n, b^m)) = a^n b^m$ holds and $\mathsf{Im}\,(\sigma \circ \theta) = \{a^n b^n \mid n \in \mathbb{N}\}$. It follows that $\sigma \circ \theta\colon \{a\}^* \to \{a, b\}^*$ is not a rational relation.

5.2 Two consequences

From Theorem 25 we deduce two important results: the Evaluation Theorem, and a 'restriction theorem'.

Theorem 35.

The image of a rational language by a rational relation is a rational language.

Proof. We want to prove: $\theta\colon A^* \to B^*$ rational relation and K in $\operatorname{Rat} A^*$ imply that $\theta(K)$ is in $\operatorname{Rat} B^*$. The following sequence of equalities holds.

$$\begin{aligned}
\theta(K) &= \bigcup_{v \in K} \theta(v) = \bigcup_{v \in K} \Big\{ w \in B^* \;\Big|\; (v, w) \in \widehat{\theta} \Big\} \\
&= \Big\{ w \in B^* \;\Big|\; \exists v \in K \quad (v, w) \in \widehat{\theta} \Big\} \\
&= \Big\{ w \in B^* \;\Big|\; \exists u \in A^*,\ \exists v \in K \quad (u, v) \in \widehat{\iota} \text{ and } (v, w) \in \widehat{\theta} \Big\} \\
&= \mathsf{Im}\ (\theta \circ \iota)
\end{aligned}$$

■

This result can be seen as a particular case of the following statement which contrasts with Fact 15.

Theorem 36. *Let $\theta\colon A^* \to B^*$ be a rational relation, K a rational language of A^* and L a rational language of B^*. Then $\widehat{\theta} \cap (K \times L)$ is the graph of a rational relation.*

Proof. It is easily verified that $\widehat{\theta} \cap (K \times L) = \widehat{\iota_L \circ \theta \circ \iota_K}$. ∎

Remark 37. Theorem 36 can also be seen as a particular instance of a general result on rational and *recognisable* subsets of non free monoids: the intersection of a rational and of a recognisable subsets of an arbitrary monoid is a rational subset and $K \times L$ is a recognisable subset of $A^* \times B^*$.

6 Exercises

1. **Orders.** The alphabet A is *totally ordered* and this order is denoted by $\leqslant$.

The *lexicographic order*, denoted by $\preccurlyeq$, extends the order on A to an order on A^* and is defined as follows. Let v and w be two words in A^* and u their *longest common prefix*. Then, $v \preccurlyeq w$ if $v = u$ or, if $v = u\,a\,s$, $w = u\,b\,t$ with a and b in A, then $a < b$.

(a) Give a finite transducer over $A^* \times A^*$ which realises $\preccurlyeq$, that is, which asssociates with every word u of A^* the set of words which are equal to or greater than u.

The *radix order* (also called the *genealogical order* or the *short-lex order*), denoted by $\sqsubseteq$, is defined as follows: $v \sqsubseteq w$ if $|v| < |w|$ or $|v| = |w|$ and $v \preccurlyeq w$.

(b) Give a finite transducer over $A^* \times A^*$ which realises $\sqsubseteq$,

For every language L of A^*, we denote by $\mathsf{minlg}\,(L)$ (resp. $\mathsf{Maxlg}\,(L)$) the set of words of L which have no smaller (resp. no greater) words in L of *the same length* in the lexicographic order.

(c) Show that if L is a rational language, so are $\mathsf{minlg}\,(L)$ and $\mathsf{Maxlg}\,(L)$.

2. **Number representation.**

Let $A_2 = \{0, 1\}$ and $A_3 = \{0, 1, 2\}$ be two alphabets of digits.

The alphabet A_3 can be first considered as a non-canonical alphabet for the representation of integers in base 2: $\overline{1\,2} = 4$, $\overline{2\,0\,1} = 9$, *etc.*

Let $\nu_2 \colon A_3^* \to A_2^*$ be the *normalisation* in base 2, that is, the relation which associates with a word of A_3^* the word of A_2^* which represents the same integer in base 2.

(a) Give a transducer which realises ν_2. Comment.

Let $\varphi \colon A_2^* \to A_3^*$ be the function which maps the binary representation of every integer onto its representation in base 3, *e.g.* $\varphi(1\,0\,0\,0) = 2\,2$.

(b) Show that φ is not a rational relation.

3. **Operation on numbers.**

(a) Give a transducer which realises the multiplication by 9 on the integers written in binary representation, that is, the relation $\tau \colon A_2^* \to A_2^*$ such that $\overline{\tau(w)} = 9 \cdot \overline{w}$.

(b) Let $\mu\colon A_2^* \times A_2^* \to A_2^*$ be the relation which realises the multiplication, that is, such that $\mu(u,v) = w$ where $\overline{w} = \overline{u} \cdot \overline{v}$.
Show that μ is not a rational relation.

4. **Map equivalence of a morphism.**

Let $\varphi_1\colon \{a,b,c\}^* \to \{x,y\}^*$ be the morphism defined by:

$$\varphi_1(a) = x\,, \quad \varphi_1(b) = y\,x\,, \quad \varphi_1(c) = x\,y\,.$$

(a) Give a subnormalised transducer which realises φ_1.

(b) Give a subnormalised transducer which realises ${\varphi_1}^{-1}$.

(c) Compute a subnormalised transducer which realises ${\varphi_1}^{-1} \circ \varphi_1$.

5. **Iteration Lemma.** Let $\theta\colon A^* \to B^*$ be a rational relation.

(a) Show that there exists an integer N such that for every pair (u,v) in $\widehat{\theta}$ whose length[8] is greater than N, there exists a factorisation:

$$(u,v) = (s,t)\,(x,y)\,(w,z)$$

such that: (i) $1 \leqslant |x| + |y| \leqslant N$ and (ii) $(u,v) = (s,t)\,(x,y)^*(w,z) \subseteq \widehat{\theta}\,.$

(b) Show that the *mirror* function $\rho\colon A^* \to A^*$:

$$\rho(a_1\,a_2 \cdots a_n) = a_n\,a_{n-1} \cdots a_1 \ ,$$

is not a rational relation.
[*Hint: Let $K = a^*b^*$, $L = b^*a^*$. Consider the relation $\pi = \iota_L \circ \rho \circ \iota_K$ and apply the Iteration Lemma to a pair $(a^N b^N, b^N a^N)$.*]

6. **Conjugacy.** Let $\mathsf{Conj}\colon A^* \to A^*$ be the relation which associates with every word w the set of its *conjugates*: $\mathsf{Conj}(w) = \{v\,u \mid u,v \in A^* \quad u\,v = w\}$.

(a) Show that if L is a rational language, then so is $\mathsf{Conj}(L)$.

(b) Give a transducer which associates with every word w of $\{a,b\}^*$ the word obtained by moving the first letter of w to its end.

(c) Compose this transducer with itself.

(d) Show that Conj is not a rational relation.

[8] The length of a pair is the sum of the lengths of its components.

Lecture V

Transducers (2)

Realisation by representations

In this lecture, we start again the study of transducers, this time by means of their *realisation by representations*. To that end, one term of the direct product plays a particular role and this yields an almost new computation model which proves to be more natural and apt to many specialisations.

After a new presentation of the Composition Theorem in this other framework, the remaining of the lecture presents such specialisations. We first present the *rational uniformisation* property which is obtained by the construction of the Schützenberger immersions applied to the 'real-time' transducers model. We then sketch the properties of *functional rational relations*.

Contents

1 Real-time transducers and representations

We define a new model of transducers, which leads then naturally to a matrix representation of automata that realise rational relations.

1.1 Definitions

The definition of real-time transducers we have in mind requires first a slight shift of Definition IV.1 of transducers towards the one of weighted automata, with the transformation of the notion of *initial* and *final states* into the one of *initial* and *final functions*. In a Boolean automaton $\mathcal{A} = \langle A, Q, I, E, T \rangle$, the subsets $I \subseteq Q$ and $T \subseteq Q$ are transformed into the functions $I\colon Q \to \mathfrak{P}(A^*)$ and $T\colon Q \to \mathfrak{P}(A^*)$ defined by:

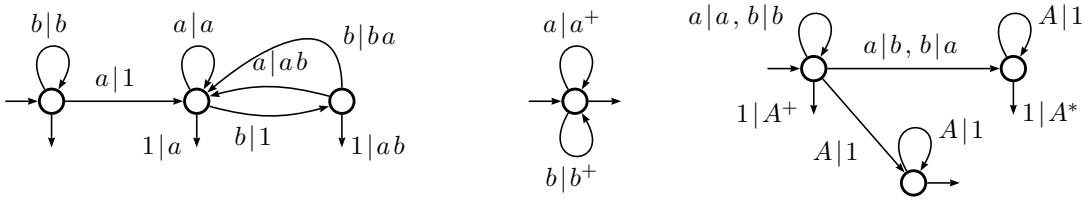
$$I(q) = \begin{cases} 1_{A^*} & \text{if } q \text{ is an initial state} \\ \emptyset & \text{otherwise}, \end{cases} \qquad T(p) = \begin{cases} 1_{A^*} & \text{if } p \text{ is a final state} \\ \emptyset & \text{otherwise}. \end{cases}$$

The definition of the label of a computation is changed accordingly so that the language accepted by the automaton, its *behaviour*, stays unchanged.

Definition 1. A *real-time transducer*[1] on $A^* \times B^*$, $\mathcal{T} = \langle A^* \times B^*, Q, I, E, T \rangle$, is an automaton the transitions of which are labelled by elements of $A \times \mathfrak{P}(B^*)$ and with initial and final functions with values in $\mathfrak{P}(B^*)$, that is, $E \subseteq Q \times A \times \mathfrak{P}(B^*) \times Q$ and $I, T\colon Q \to \mathfrak{P}(B^*)$.

The transducer $\mathcal{T}$ is said to be *finite* if E is finite, if every transition is labelled in $A \times \operatorname{Rat} B^*$ and if I and T are with values in $\operatorname{Rat} B^*$.

Example 2. Figure 1 shows three real-time transducers.



(a) (b) $\mathcal{G}_1$ (c)

Figure 1: Three real-time transducers

More generally, the transitions of $\mathcal{T}$ are thus of the form:

$$p \xrightarrow[\mathcal{T}]{a|K_{a,p,q}} q \qquad \text{with} \qquad a \in A\,,\ K_{a,p,q} \subseteq B^*\ ,$$

[1] In French: *transducteur "temps-réel"*, a terminology that is not completely satisfactory but that I use for lack of a better translation.

from which we deduce the form of the computations of $\mathcal{T}$:

$$c = \xrightarrow{I(p_0)} p_0 \xrightarrow{a_1|K_{a_1,p_0,p_1}} p_1 \xrightarrow{a_2|K_{a_2,p_1,p_2}} p_2 \cdots p_{n-1} \xrightarrow{a_n|K_{a_n,p_{n-1},p_n}} p_n \xrightarrow{T(p_n)} ,$$

and the one of their label:

$$|c| = (1_{A^*}, I(p_0))\,(a_1, K_{a_1,p_0,p_1})\,(a_2, K_{a_2,p_1,p_2}) \cdots (a_n, K_{a_n,p_{n-1},p_n})\,(1_{A^*}, T(p_n)) \ .$$

The relation $\theta\colon A^* \to B^*$ realised by $\mathcal{T}$ is thus, for every $w = a_1a_2\cdots a_n$ in A^*:

$$\theta(w) = \bigcup_{\substack{c \text{ calcul de } \mathcal{T} \\ \pi_{A^*}(|c|)=w}} I(p_0)\,K_{a_1,p_0,p_1}\,K_{a_2,p_1,p_2}\,\cdots\,K_{a_n,p_{n-1},p_n}\,T(p_n) \ .$$

1.2 Realisation of rational relations

Theorem 3. *A relation* $\theta\colon A^* \to B^*$ *is* rational *if and only if it is realised by a* finite *real-time transducer.*

Proof. (i) The condition is sufficient. Let $\mathcal{T}$ be a *finite* real-time transducer. If K is a rational language of B^* accepted by $\mathcal{A}$ (Figure 2 (a)), a transition $p \xrightarrow{a|K} q$ of $\mathcal{T}$ (Figure 2 (b)) is replaced by a set of labelled transitions (Figure 2 (c)), the initial and final functions $I(q) = K$ and $T(p) = K$ by two sets of labelled transitions (Figure 2 (d) and (e)).

The transducer we obtain in this way is easily seen to be equivalent to the tranducer $\mathcal{T}$ we started from. This construction yields spontaneous transitions which are then eliminated by the classical algorithm and the result is then a normalised transducer, still equivalent to $\mathcal{T}$. Figure 3 shows this construction applied to the transducer $\mathcal{G}_1$ of Figure 1 (b).

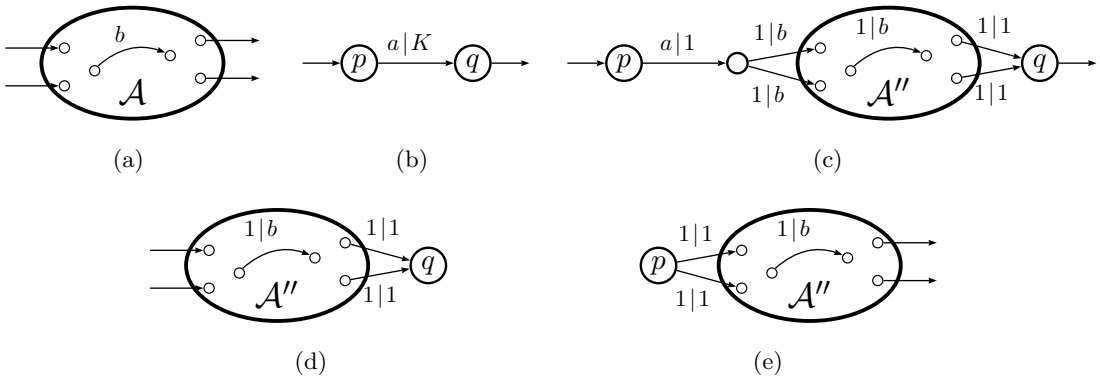


Figure 2: Transforming a real-time transducer into a normalised transducer

(ii) The condition is necessary. Let $\theta\colon A^* \to B^*$ be a rational relation and $\mathcal{T}$ a subnormalised transducer which realises $\widehat{\theta}$. The transducer $\mathcal{T}$ is written under the

matrix form: $\mathcal{T} = \langle\, I, E, T \,\rangle$ where I and T are Boolean vectors of dimension Q and where E is a matrix of dimension $Q \times Q$ the entries of which are subsets of $(A \times \{1_{B^*}\}) \cup (\{1_{A^*}\} \times B)$ — if $\mathcal{T}$ is normalised — or
of $(A \times \{1_{B^*}\}) \cup (\{1_{A^*}\} \times B) \cup (A \times B)$ — if $\mathcal{T}$ is subnormalised. In any case,

$$\widehat{\theta} = \big|\mathcal{T}\big| = I \cdot E^* \cdot T \tag{1.1}$$

holds. In any case also, we can write

$$E = F + G \qquad \text{with} \qquad G \in (\{1_{A^*}\} \times B)^{Q \times Q}$$
$$\text{and} \qquad F \in (A \times \{1_{B^*}\})^{Q \times Q} \qquad \text{or} \qquad F \in \big(\,(A \times \{1_{B^*}\}) \cup (A \times B)\,\big)^{Q \times Q}$$

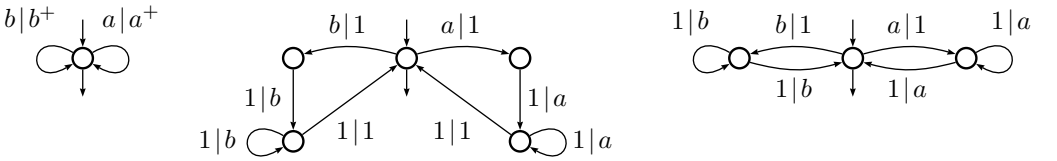
according to whether $\mathcal{T}$ is normalised or subnormalised. Equation (1.1) reads then:

$$\big|\mathcal{T}\big| = I \cdot E^* \cdot T = I \cdot (F + G)^* \cdot T = I \cdot (G^* \cdot F)^* \cdot G^* \cdot T = I \cdot G^* \cdot (F \cdot G^*)^* \cdot T\,.$$

This shows that $\mathcal{T}$ is equivalent to the two transducers $\mathcal{T}'$ and $\mathcal{T}''$ with:

$$\mathcal{T}' = \langle\, I, G^* \cdot F, G^* \cdot T \,\rangle \qquad \text{and} \qquad \mathcal{T}'' = \langle\, I \cdot G^*, F \cdot G^*, T \,\rangle\ .$$

The entries of G^* belong to $(1_{A^*} \times \mathrm{Rat}\, B^*)$, and those of $F \cdot G^*$ belong to $(A \times \mathrm{Rat}\, B^*)$: $\mathcal{T}'$ and $\mathcal{T}''$ are two finite real-time transducers. Example 4 gives these computations for the transducer $\mathcal{G}_2$. ■



(a) $\mathcal{G}_1$ (b) after construction (c) $\mathcal{G}_2$, after closure (and quotient)

Figure 3: A real-time transducer transformed into a normalised transducer

Example 4. The matrix representation of $\mathcal{G}_2$ is:

$$I_2 = \begin{pmatrix} 1 & 0 & 0 \end{pmatrix}, \quad E_2 = \begin{pmatrix} 0 & (a,1) & (b,1) \\ (1,a) & (1,a) & 0 \\ (1,b) & 0 & (1,b) \end{pmatrix}, \quad T_2 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}\ ,$$

from which we compute:

$$F_2 = \begin{pmatrix} 0 & (a,1) & (b,1) \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad G_2 = \begin{pmatrix} 0 & 0 & 0 \\ (1,a) & (1,a) & 0 \\ (1,b) & 0 & (1,b) \end{pmatrix}, \quad G_2^* = \begin{pmatrix} (1,1) & 0 & 0 \\ (1,a^+) & (1,a^*) & 0 \\ (1,b^+) & 0 & (1,b^*) \end{pmatrix}\ .$$

The transducer $\mathcal{G}_2' = \langle\, I_2, G_2^* \cdot F_2, G_2^* \cdot T_2 \,\rangle$ is shown at Figure 4 and it is easily seen that $\mathcal{G}_2'' = \langle\, I_2 \cdot G_2^*, F_2 \cdot G_2^*, T_2 \,\rangle$ is equal to $\mathcal{G}_1$.

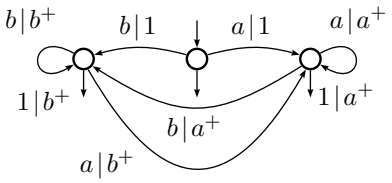


Figure 4: The real-time transducer $\mathcal{G}'_2$

1.3 Representations of rational relations

Definition 5. Let A^* and B^* be two free monoids. A *representation* of A^* into $\operatorname{Rat} B^*$ *of dimension* Q is a triple $\langle\, I, \mu, T \,\rangle$ where $\mu\colon A^* \to (\operatorname{Rat} B^*)^{Q\times Q}$ is a *morphism* (hence entirely defined by the matrices $\mu(a)$ for a in A), and where I and T are respectively row and column vectors in $(\operatorname{Rat} B^*)^Q$.

Theorem 6. *A relation* $\theta\colon A^* \to B^*$ *is* rational *if and only if there exists a representation* $\langle\, I, \mu, T \,\rangle$ *of* A^* *into* $\operatorname{Rat} B^*$ *which realises* θ*, that is, such that*

$$\forall w \in A^* \qquad \theta(w) = I \cdot \mu(w) \cdot T \quad .$$

Proof. If $\mathcal{T} = \langle\, I, E, T \,\rangle$ is a real-time transducer, the matrix E defines the morphism $\mu\colon A^* \to (\operatorname{Rat} B^*)^{Q\times Q}$ by

$$E = \sum_{a\in A} (a, 1)\,(1, \mu(a)) \tag{1.2}$$

that is,

$$\forall a \in A\,,\ \forall p, q \in Q \qquad \mu(a)_{p,q} = \begin{cases} K_{a,p,q} & \text{if } p \xrightarrow[\mathcal{T}]{a|K_{a,p,q}} q \\ \emptyset & \text{otherwise}, \end{cases}$$

under the hypothesis, necessary for the writing $\mathcal{T} = \langle\, I, E, T \,\rangle$, that for every a in A and every pair p, q in Q, there exists at most one transition in $\mathcal{T}$ which goes from p to q and whose first component is a.

Conversely, a morphism $\mu\colon A^* \to (\operatorname{Rat} B^*)^{Q\times Q}$ defines, via the same Equation (1.2), the adjacency matrix of a finite real-time transducer. By induction on the length of the words w, it is then checked that:

$$\forall w \in A^*\,,\ \forall p, q \in Q \qquad \mu(w)_{p,q} = L \quad \Longleftrightarrow \quad L = \bigcup \left\{ H \;\middle|\; p \xrightarrow[\mathcal{T}]{w|H} q \right\}$$

from which it follows that, for every w in A^*, $\theta(w) = I \cdot \mu(w) \cdot T$. ∎

Example 7. The representations of the real-time transducers of Example 4 are the following:

$$\text{(a)} \quad I = \begin{pmatrix} 1 & 0 & 0 \end{pmatrix}, \quad \mu(a) = \begin{pmatrix} 0 & 1 & 0 \\ 0 & a & 0 \\ 0 & ab & 0 \end{pmatrix}, \quad \mu(b) = \begin{pmatrix} b & 0 & 0 \\ 0 & 0 & 1 \\ 0 & ba & 0 \end{pmatrix}, \quad T = \begin{pmatrix} 1 \\ a \\ ab \end{pmatrix};$$

(b) $I = \begin{pmatrix} 1 \end{pmatrix}$, $\mu(a) = \begin{pmatrix} a^+ \end{pmatrix}$, $\mu(b) = \begin{pmatrix} b^+ \end{pmatrix}$, $T = \begin{pmatrix} 1 \end{pmatrix}$;

(c) $I = \begin{pmatrix} 1 & 0 & 0 \end{pmatrix}$, $\mu(a) = \begin{pmatrix} a & b & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$, $\mu(b) = \begin{pmatrix} b & a & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$, $T = \begin{pmatrix} A^+ \\ A^* \\ 1 \end{pmatrix}$.

2 Composition and evaluation theorems

The realisation of rational relations by representations allows us to state a *new* composition theorem. The result in itself is not new, since we proved it already in Lecture IV: *the composition of two rational relations is a rational relation* (Theorem IV.25). But, at the same time, what we present here is more than a new proof. This result is indeed now the *consequence* of another result: *the composition of two representations is a representation*, which is new.

The idea is to generalise the *composition of morphisms* between free monoids to representations. In Lecture IV, we had deduce the 'evaluation theorem' from the closure under composition of rational relations. We now proceed the other way around and establish the evaluation theorem first.

2.1 Evaluation Theorem

In the sequel of the section, $\langle I, \mu, T \rangle$ is a representation of A^* into $\operatorname{Rat} B^*$ of dimension Q, that is:

$$\mu \colon A^* \to (\operatorname{Rat} B^*)^{Q \times Q} \quad \text{is a morphism}, \quad I \in (\operatorname{Rat} B^*)^{1 \times Q} \quad \text{and} \quad T \in (\operatorname{Rat} B^*)^{Q \times 1}.$$

Recall that any application extends additively, that is, if $K \subseteq A^*$, then:

$$\mu(K) = \sum_{w \in K} \mu(w) \quad \text{that is,} \quad \forall p, q \in Q \qquad \mu(K)_{p,q} = \bigcup_{w \in K} \mu(w)_{p,q} \quad .$$

Proposition 8. *If $\mu \colon A^* \to (\operatorname{Rat} B^*)^{Q \times Q}$ is a morphism, then*

$$K \in \operatorname{Rat} A^* \quad \Longrightarrow \quad \mu(K) \in (\operatorname{Rat} B^*)^{Q \times Q} \quad ,$$

that is, for every p and q in Q, $\mu(K)_{p,q}$ belongs to $\operatorname{Rat} B^*$.

From which follows:

Corollary 9. *If $\theta \colon A^* \to B^*$ is a rational relation, then*

$$K \in \operatorname{Rat} A^* \quad \Longrightarrow \quad \theta(K) \in \operatorname{Rat} B^* \quad .$$

Proof. If θ is a rational relation, then θ is realised by a representation $\langle\, I, \mu, T \,\rangle$ and

$$\theta(K) = \bigcup_{w\in K} \theta(w) = \bigcup_{w\in K} I\cdot\mu(w)\cdot T = I\cdot\left(\bigcup_{w\in K}\mu(w)\right)\cdot T = I\cdot\mu(K)\cdot T \ . \quad \blacksquare$$

Proof of Proposition 8. (i) Let us recall first the theorem (see proof of Proposition I.17):

Theorem 10. *Let M be a monoid and E a matrix of dimension $Q\times Q$, the entries of which are in $\mathfrak{P}(M)$. Then, the entries of E^* belong to the* rational closure *of the entries of E.* ■

(ii) Preparation. We check successively the 'closure' by union, product, and star. First, the union:

$$\mu(K)\,,\mu(L)\in(\mathrm{Rat}\,B^*)^{Q\times Q} \quad\Longrightarrow\quad \mu(K\cup L)\in(\mathrm{Rat}\,B^*)^{Q\times Q} \tag{2.1}$$

since, for every p and q in Q, $\mu(K\cup L)_{p,q} = \mu(K)_{p,q}\cup\mu(L)_{p,q}$. Second, the product:

$$\mu(K)\,,\mu(L)\in(\mathrm{Rat}\,B^*)^{Q\times Q} \quad\Longrightarrow\quad \mu(K\,L)\in(\mathrm{Rat}\,B^*)^{Q\times Q} \tag{2.2}$$

since, on the one hand:

$$\begin{aligned}\mu(K\,L) &= \bigcup\{\mu(w)\mid w\in K\,L\} = \bigcup\{\mu(u\,v)\mid u\in K\,,\ v\in L\}\\ &= \bigcup\{\mu(u)\,\mu(v)\mid u\in K\,,\ v\in L\}\\ &= \Big(\bigcup\{\mu(u)\mid u\in K\}\Big)\ \Big(\bigcup\{\mu(v)\mid v\in L\}\Big) = \mu(K)\,\mu(L)\ ,\end{aligned}$$

and, on the other, $(\mathrm{Rat}\,B^*)^{Q\times Q}$ is closed by product since $\mathrm{Rat}\,B^*$ is a semiring. And, finally:

$$\mu(K)\in(\mathrm{Rat}\,B^*)^{Q\times Q} \quad\Longrightarrow\quad \mu(K^*)\in(\mathrm{Rat}\,B^*)^{Q\times Q} \tag{2.3}$$

since, by (2.1) and (2.2)

$$\mu(K^*) = \mu\left(\bigcup_{n\in\mathbb{N}} K^n\right) = \bigcup_{n\in\mathbb{N}}\mu(K^n) = \bigcup_{n\in\mathbb{N}}\big(\mu(K)\big)^n = \big(\mu(K)\big)^*\ ,$$

and Theorem 10 applies.

(iii) Proposition 8 is the consequence of the three equations (2.1), (2.2) and (2.3), by induction on the depth of a rational expression which denotes K. ■

2.2 Composition of representations

Definition 11. Let $\mu\colon A^*\to(\mathrm{Rat}\,B^*)^{Q\times Q}$ and $\nu\colon B^*\to(\mathrm{Rat}\,C^*)^{R\times R}$ be two morphisms. The *composition of μ by ν* is the map $\pi=\nu\circ\mu$ from A^* to $(\mathrm{Rat}\,C^*)^{(Q\times R)\times(Q\times R)}$, defined by the block decomposition:

$$\forall w\in A^* \qquad \pi(w)_{p\times R,q\times R} = \nu\Big(\mu(w)_{p,q}\Big)\ .$$

Remark that this definition is built upon Proposition 8 in the sense that the entries of $\pi(w)$ are *a priori* in $\mathfrak{P}(A^*)$ and it is the proposition that insures that they are in $\operatorname{Rat} C^*$.

Examples 12. (i) Definition 11 coincides with the definition of monoid morphisms composition, in the case where μ and ν are such morphisms.

(ii) If $\mu_1(a) = \begin{pmatrix} 0 & a \\ b & 0 \end{pmatrix}$, $\mu_1(b) = \begin{pmatrix} b & 0 \\ 0 & b \end{pmatrix}$ and $\nu_1(a) = \begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix}$, $\nu_1(b) = \begin{pmatrix} 0 & b \\ a & 0 \end{pmatrix}$, then:

$$\pi_1(a) = \begin{pmatrix} 0 & 0 & a & 0 \\ 0 & 0 & 0 & a \\ 0 & b & 0 & 0 \\ a & 0 & 0 & 0 \end{pmatrix} \quad \text{and} \quad \pi_1(b) = \begin{pmatrix} 0 & b & 0 & 0 \\ a & 0 & 0 & 0 \\ 0 & 0 & 0 & b \\ 0 & 0 & a & 0 \end{pmatrix} .$$

Definition 11 is legitimate thanks to the following proposition:

Proposition 13. *$\pi = \nu \circ \mu \colon A^* \to (\operatorname{Rat} C^*)^{(Q\times R)\times(Q\times R)}$ is a morphism.*

Proof. We want to prove that, for every u and v in A^*, it holds:

$$[\nu \circ \mu](u\,v) = [\nu \circ \mu](u)\,[\nu \circ \mu](v) \ .$$

For every p and q in Q, it holds:

$$\begin{aligned}
([\nu \circ \mu](u\,v))_{p\times R, q\times R} &= \nu\Big(\mu(u\,v)_{p,q}\Big) = \nu\left(\sum_{r\in Q}\Big(\mu(u)_{p,r}\ \mu(v)_{r,q}\Big)\right) \\
&= \sum_{r\in Q} \nu\Big(\mu(u)_{p,r}\ \mu(v)_{r,q}\Big) = \sum_{r\in Q}\Big(\nu\Big(\mu(u)_{p,r}\Big)\ \nu\Big(\mu(v)_{r,q}\Big)\Big) \\
&= \sum_{r\in Q} ([\nu \circ \mu](u)_{p\times R, r\times R}\,[\nu \circ \mu](v)_{r\times R, q\times R}) \\
&= ([\nu \circ \mu](u) \cdot [\nu \circ \mu](v))_{r\times R, q\times R} \ . \qquad \blacksquare
\end{aligned}$$

Theorem 14. *Let $\theta \colon A^* \to B^*$ and $\sigma \colon B^* \to C^*$ be two rational relations, realised by the representations $\langle\, I, \mu, T \,\rangle$ and $\langle\, J, \nu, U \,\rangle$ respectively. Then, $\sigma \circ \theta \colon A^* \to C^*$ is the rational relation realised by the representation $\langle\, K, \pi, V \,\rangle$, with:*

$$\pi = \nu \circ \mu\,, \qquad K = J \cdot \nu(I) \qquad \text{and} \qquad V = \nu(T) \cdot U \ .$$

Proof.

$$\begin{aligned}
\forall w \in A^* \qquad [\sigma \circ \theta](w) &= \sigma(\theta(w)) = \sigma(I \cdot \mu(w) \cdot T) \\
&= J \cdot \nu(I \cdot \mu(w) \cdot T) \cdot U \\
&= (J \cdot \nu(I)) \cdot \nu(\mu(w)) \cdot (\nu(T) \cdot U) \ . \qquad \blacksquare
\end{aligned}$$

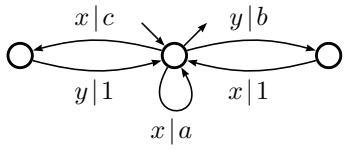
Theorem 14 yields a new method of composition of transducers, by means of the sequence of transformations:

transducer $\rightsquigarrow$ real-time transducer $\rightsquigarrow$ representation $\rightsquigarrow$
composition of representations $\rightsquigarrow$ real-time transducer $\rightsquigarrow$ transducer.

Example 15. The morphism $\varphi_1 \colon \{a,b,c\}^* \to \{x,y\}^*$ defined by:

$$\varphi_1(a) = x\,, \quad \varphi_1(b) = y\,x\,, \quad \varphi_1(c) = x\,y\,.$$

is realised by the representation $\langle\, 1, \varphi_1, 1 \,\rangle$. The relation ${\varphi_1}^{-1} \colon \{x,y\}^* \to \{a,b,c\}^*$ is realised by the real-time transducer below:



and then by the representation $\langle\, J_1, \kappa_1, U_1 \,\rangle$ where:

$$J_1 = \begin{pmatrix} 1 & 0 & 0 \end{pmatrix}, \quad \kappa_1(x) = \begin{pmatrix} a & c & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad \kappa_1(y) = \begin{pmatrix} 0 & 0 & b \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \text{and} \quad U_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} .$$

The relation $\varphi_1^{-1} \circ \varphi_1 \colon A^* \to A^*$ is realised by the representation
$\langle\, J_1, \kappa_1, U_1 \,\rangle \circ \langle\, 1, \varphi_1, 1 \,\rangle = \langle\, J_1, \pi_1, U_1 \,\rangle$ with $\pi_1 = \kappa_1 \circ \varphi_1$, that is: $\pi_1(a) = \kappa_1(x)$,

$$\pi_1(b) = \kappa_1(y\,x) = \begin{pmatrix} b & 0 & 0 \\ a & c & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \text{and} \quad \pi_1(c) = \kappa_1(x\,y) = \begin{pmatrix} c & 0 & a\,b \\ 0 & 0 & 0 \\ 0 & 0 & b \end{pmatrix} ,$$

which corresponds to the real-time transducer below.

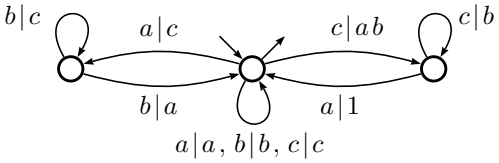


Figure 5: A transducer for $\varphi_1^{-1} \circ \varphi_1$

3 Uniformisation of rational relations

As a first illustration of the realisation of rational relations by representations, we establish a *Rational Uniformisation Theorem*.

3.1 Uniformisation of a relation

The notion of *uniformisation of a relation* (and the terminology) comes from logic (more precisely from descriptive set theory). In these notes, it will be only a definition that allows us to state a result.

If θ is any relation, from a set E into another set F, a function τ from E to F *uniformises* θ (or is *a uniformisation* of θ) if, for every e in the domain of θ, $\tau(e)$ is an element of $\theta(e)$, that is:

$$\mathsf{Dom}\,\tau = \mathsf{Dom}\,\theta \qquad \text{and} \qquad \forall e \in \mathsf{Dom}\,\theta \quad \tau(e) \in \theta(e) \quad .$$

A uniformisation τ of a relation θ consists then in a *choice function*, a choice that is repeated for every element of the domain of θ, in the *selection* of an element $\tau(e)$ in each of the subsets $\theta(e)$ (*cf.* Figure 6).

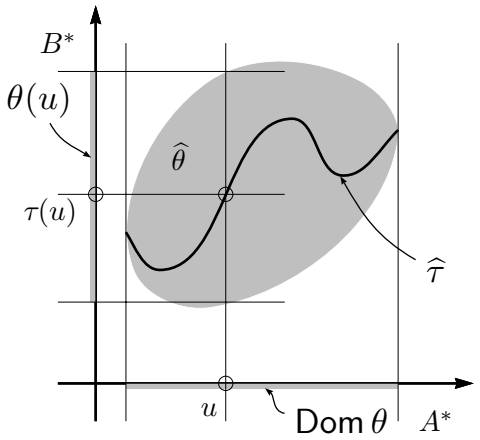


Figure 6: A uniformisation τ of a relation θ

Example 16. If θ is a relation from A^* (from an arbitrary set E indeed) in a free monoid B^*, the *radix uniformisation*[2] is the function θ_{rad} which associates with every element e of the domain of θ the smallest element, in the radix order, of $\theta(e)$.

Along the same lines, we can define the *lexicographic selection* θ_{lex}: it is the function which associates with every element e of the domain of θ the smallest element, in the lexicographic order, of $\theta(e)$, *if it exixts*. As the lexicographic order *is not a well-order*, this minimum element does not necessarily exist and θ_{lex} is not always a uniformisation, that is, it is a function the domain of which may be *strictly contained* in the one of θ.

[2]If radix order is called, a tempting option, *military order* (the oldest in the highest rank), we then get a *military uniformisation*...

3.2 The Rational Uniformisation Theorem

After the definition of uniformisation, the problem consists in knowing whether it is possible, when θ belongs to a certain family of relations, to define *by selecting* for every e an element in $\theta(e)$, a uniformisation τ which belongs to that family of relations, or to another given family of functions. The problem is solved for the family of rational relations by the following result:

Theorem 17 (Eilenberg 1974)**.** *Every rational relation is uniformised by an unambiguous functional rational relation.*

Before proving Theorem 17, let us complete the definition of its statement. A rational relation $\theta\colon A^* \to B^*$ is *unambiguous* if there exists a transducer $\mathcal{T}$ which realises θ and such that every pair (u, v) in $\widehat{\theta} = \big|\mathcal{T}\big|$ is the label of *exactly one* successful computation of $\mathcal{T}$. A rational relation which is *not* unambiguous is called *inherently ambiguous*.

Fact 18. *There exist inherently ambiguous rational relatons.*

Example 19. Let us take the behaviours of transducers $\mathcal{V}_1$ and $\mathcal{W}_1$ of Example IV.2:

$$\big|\mathcal{V}_1\big| = \{(a^n b^m, c^n) \mid n, m \in \mathbb{N}\} \qquad \text{and} \qquad \big|\mathcal{W}_1\big| = \{(a^n b^m, c^m) \mid n, m \in \mathbb{N}\} \quad .$$

Then the rational relation $\big|\mathcal{V}_1\big| \cup \big|\mathcal{W}_1\big|$ is inherently ambiguous.

Proof of Theorem 17. Recall that if $\mathcal{A}$ is an automaton over A^* and $\widehat{\mathcal{A}}$ is its determinisation, the accessible part $\mathcal{S}$ of $\widehat{\mathcal{A}} \times \mathcal{A}$ is a covering of $\mathcal{A}$, called the *Schützenberger covering* or *S-covering* of $\mathcal{A}$, and that the projection $\pi_{\widehat{\mathcal{A}}}$ of $\mathcal{S}$ onto $\widehat{\mathcal{A}}$ is an In-morphism. As a result, and by eliminating some transitions in $\mathcal{S}$, we can construct a sub-automaton $\mathcal{T}$ of $\mathcal{S}$, called an *S-immersion* of $\mathcal{A}$, which is *unambiguous* and *equivalent* to $\mathcal{A}$ (see Corollary II.37).

Let $\mathcal{C}$ be a real-time transducer which realises a relation θ (corresponding to a representation $\langle I, \mu, T \rangle$ of θ), $\mathcal{A}$ its underlying input automaton, and $\mathcal{T}$ an S-immersion in $\mathcal{A}$. Since $\mathcal{T}$ is an immersion in $\mathcal{A}$, each transition (r, a, s) of $\mathcal{T}$ corresponds to a *unique transition* (p, a, q) in $\mathcal{A}$ and hence to a unique transition $(p, (a, \mu(a)_{p,q}), q)$ in $\mathcal{C}$. If we choose, *arbitrarily*, a word w in $\mu(a)_{p,q}$, we construct a transducer $\mathcal{U}$ by replacing every transition (r, a, s) in $\mathcal{T}$ by $(r, (a, w), s)$. Since $\mathcal{T}$ is unambiguous, the relation τ realised by $\mathcal{U}$ is an unambiguous function, and since $\mathcal{T}$ is equivalent to $\mathcal{A}$ it has the same domain as θ, and its graph is contained in that of θ, by the choice of w. ∎

Example 20. — Let θ_2 be the relation from $\{a, b\}^*$ into itself which replace in every word one of its factor ab by the set b^+a (and which is not defined on the words that do not contain such a factor). Figure 7 shows a transducer $\mathcal{E}_2$ which realises θ_2 and whose underlying input automaton is $\mathcal{A}_1$ (on the left, verticaly), its

determinisation $\widehat{\mathcal{A}_1}$ (horizontaly, at the top) and a possible result of the construction decribed in the proof above. □

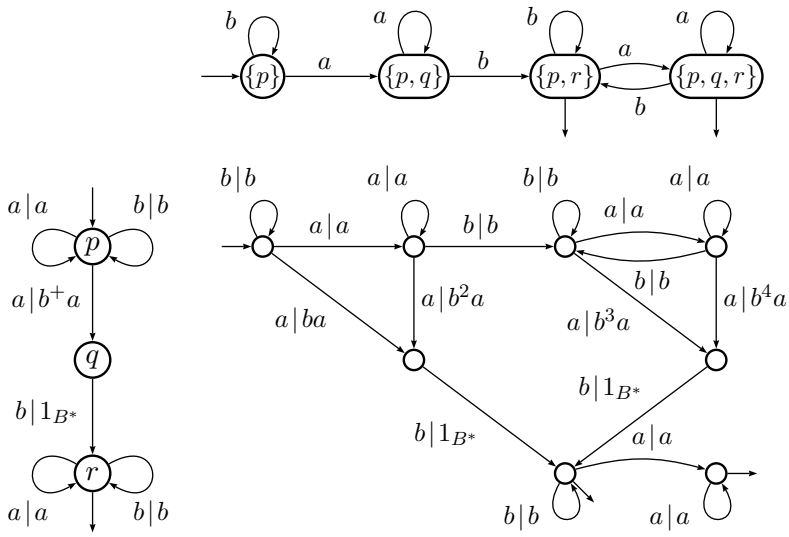


Figure 7: The transducer $\mathcal{E}_2$ and a S-uniformisation of θ_2

Since the only possible uniformisation of a function is the function itself, Theorem 17 implies the following statement (which contrasts with Fact 18).

Corollary 21.

Every functional rational relation is an unambiguous rational relation.

4 Rational and sequential functions

The realisation by representation yields handy criteria for defining or characterising classes of rational relations. Two classes of relations are naturally investigated from this point of view: the one of *functional rational relations* — which we call *rational functions*, in spite of the unfortunate collision with a classical terminology in mathematics and the one of *sequential rational functions* — which we simply call *sequential functions*.

Each of these classes would deserve a full lecture. We just give here the definitions and state the main results.

4.1 Rational functions

Proposition 22. *Let $\theta\colon A^* \to B^*$ be a rational relation realised by a* trim *representation $\langle I, \mu, T \rangle$. If θ is a function, then all non zero entries of the matrices $\mu(a)$ are words (*monomials*).* ■

Theorem 23. *It is decidable whether a rational relation is functional or not (with a quadratic complexity).* ■

Example 24. Figure 8 shows a transducer which realises a functional relation (which is not so obvious at first sight).

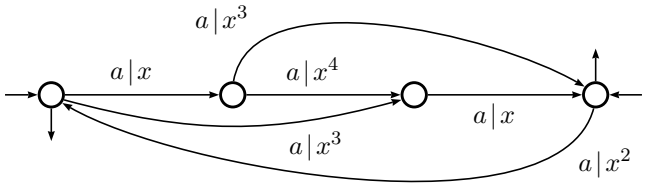


Figure 8: A functional transducer

4.2 Sequential functions

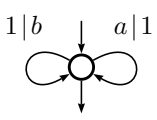
Definition 25. A rational function is *sequential* (resp. *co-sequential*) if it is realised by a *row-monomial* representation (resp. *column-monomial* representation), that is, if the underlying input automaton of a real-time transducer which realises it is *deterministic* (resp. *co-deterministic*).

Theorem 26. *It is decidable whether a rational function is sequential or not (with a polynomial complexity).* ■

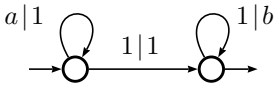
Theorem 27. *Every rational function is the composition of a sequential function by a co-sequential function.* ■

5 Exercises

1. Apply the construction of the proof of Theorem 3 in order to build real-time transducers from the two transducers below which realise the universal relation on $\{a\}^* \times \{b\}^*$.



(a) $\mathcal{U}_1$



(b) $\mathcal{U}_2$

2. Give a realisation by representation of the following relations:

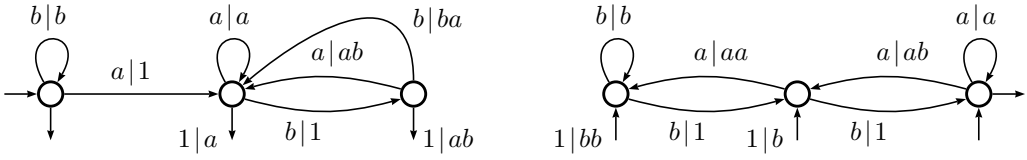
(a) the complement of the identity; (b) the lexicographic order; (c) the radix order.

3. **Finite and infinite components of a rational relation.** Let $\tau\colon A^* \to B^*$ be a relation. The *finite* and *infinite components* τ_{f} and τ_∞ of τ are defined by:

$$\tau_{\mathsf{f}}(w) = \begin{cases} \tau(w) & \text{if } \|\tau(w)\| \text{ is finite} \\ \emptyset & \text{otherwise} \end{cases} \quad \text{et} \quad \tau_\infty(w) = \begin{cases} \emptyset & \text{if } \|\tau(w)\| \text{ is finite} \\ \tau(w) & \text{otherwise} \end{cases}$$

Show that if τ is rational, then τ_{f} and τ_∞ are rational and effectively computable from τ.

4. **Fibonacci reduction.** Give a transducer which realises the composition of the relations realised by the transducers below (the transducer on the left by the transducer on the right).



5. **Choosing the uniformisation.** Let $A = \{a, b, c\}$ be a totally ordered alphabet, where $a < b < c$, and let θ be the rational relation from A^* into itself whose graph is:

$$\widehat{\theta} = (a,a)^*\,(b,1)^*\,(1,b) \cup (a,1)^*\,(b,a)^*\,(1,c) \ .$$

Show that neither the radix uniformisation θ_{rad} nor the lexicographic selection θ_{lex} are rational functions.

6. **Inherently ambiguous rational relation.** Let $\mathcal{V}_1$ and $\mathcal{W}_1$ be the transducers of Example IV.2:

$$\big|\mathcal{V}_1\big| = \{(a^n b^m, c^n) \mid n, m \in \mathbb{N}\} \qquad \text{and} \qquad \big|\mathcal{W}_1\big| = \{(a^n b^m, c^m) \mid n, m \in \mathbb{N}\} \ .$$

Show that the rational relation $\big|\mathcal{V}_1\big| \cup \big|\mathcal{W}_1\big|$ is inherently ambiguous.

Notation Index