MPRI, Fondations mathématiques de la théorie des automates

Olivier Carton, Jean-Éric Pin

Examen du 9 mars 2011. Durée: 2h 30, notes de cours autorisées

⋆⋆⋆

Avertissement : On attachera une grande importance à la clarté, à la précision et à la concision de la rédaction.

1. Un exemple

On considère sur l'alphabet $A = \{a, b\}$ le langage $L = A^*abbA^*$, dont voici l'automate minimal:

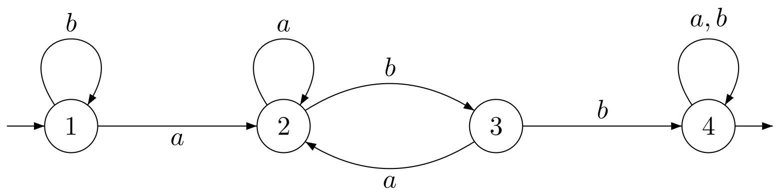


Figure 1: L'automate minimal de A^*abbA^*.

Question 1. Calculer le monoïde syntactique M de L (on trouvera 10 éléments). Donner la liste des idempotents de M et sa structure en $\mathcal{J}$-classes. Est-ce que M est commutatif? apériodique? Justifier chacune de vos réponses.

Question 2. Pour chaque idempotent $e \neq 1$, calculer le semigroupe eMe (qui est l'ensemble des éléments de la forme ese avec $s \in M$). Que peut-on dire de tous ses semigroupes?

2. Langages de la forme A^*uA^*, où u est un mot non vide.

Question 3. Soit $L = A^*uA^*$ et soit $\eta : A^* \to M$ son morphisme syntactique. Montrer que l'élément $\eta(u)$ est un zéro de M, que l'on notera 0 par la suite. Montrer que $L = \eta^{-1}(0)$.

Question 4. Soit x un mot de A^+ tel que $x \sim_L x^2$. Montrer que pour tout $y, z \in A^*$, on a $xyx \leqslant_L x$, $xyxyx \sim_L xyx$ et $xyxzx \sim_L xzxyx$.

Question 5. Montrer que L vérifie les équations suivantes:

Pour tout $x \in A^+$, pour tout $y \in A^*$

(1) $x^\omega y x^\omega y x^\omega = x^\omega y x^\omega$ (locament idempotent)

(2) $x^\omega y x^\omega z x^\omega = x^\omega z x^\omega y x^\omega$ (localement commutatif)

(3) $x^\omega y x^\omega \leqslant x^\omega$ (maximum local)

Pour tout $x, y, s \in A^*$

(4) $s(xy)^\omega x \leftrightarrow s(xy)^\omega$ et $y(xy)^\omega s \leftrightarrow (xy)^\omega s$

Pour tout $x, y \in A^+$, pour tout $r, s \in A^*$

(5) $x^\omega r y^\omega s x^\omega \leftrightarrow y^\omega s x^\omega r y^\omega$

Question 6. Soit L un langage de A^* et soit $\eta : A^* \to M$ son morphisme syntactique. On pose $P = \eta(L)$. Montrer que si L vérifie l'équation (4), alors P *sature* les $\mathcal{J}$-classes de M (i.e. si $s \in P$ et $s \mathrel{\mathcal{J}} t$, alors $t \in P$).

3. Langages localement testables

On définit, pour chaque entier k, une relation $\sim_k$ sur A^* par $u \sim_k v$ si et seulement si

(a) u et v ont les mêmes préfixes de longueur $< k$,

(b) u and v ont les mêmes suffixes de longueur $< k$,

(c) u et v ont les mêmes facteurs de longueur k (sans tenir compte des multiplicités).

Par exemple $ababab cbcb \sim_3 ababcbcbcb$.

Question 7. Montrer que $\sim_k$ est une congruence d'indice fini.

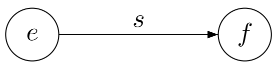
On dit qu'un langage est *k-testable* (kT) s'il est union de classes d'équivalences de $\sim_k$. Il est *localement testable* (LT) s'il est k-testable pour au moins un entier k.

Question 8. Montrer qu'un langage est LT si et seulement s'il est combinaison booléenne de langages de la forme pA^*, A^*uA^* ou A^*s avec $p, u, s \in A^+$.

Question 9. Montrer qu'un langage LT vérifie les équations (1) et (2) ci-dessus.

4. Une caractérisation algébrique

Soit L un langage rationnel de A^+, soit S son semigroupe (pas monoïde!) syntactique et soit $\pi : A^+ \to S$ son morphisme syntactique. On note $G(S)$ le graphe orienté dont les sommets sont les idempotents de S et les arcs sont de la forme



où $e, f \in E(S)$ et s est un élément de S tel que $es = s = sf$. Par définition, *l'étiquette* du chemin de $G(S)$

est le produit $s_1 s_2 \cdots s_n$.

Question 10. Dessiner $G(S)$ lorsque S est le semigroupe syntactique du langage considéré dans la première partie (donc $S = M - \{1\}$).

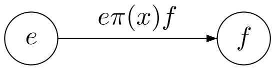
On dit que S vérifie la *condition des chemins* si deux chemins de $G(S)$ issus du même sommet, arrivant sur le même sommet et traversant les mêmes arcs (sans tenir compte de la multiplicité), ont les mêmes étiquettes.

On pose $k = |S| + 1$ et on fixe un ordre total sur les idempotents de S: $e_1 < e_2 < \ldots < e_n$. On dit qu'un mot $p \in A^+$ est *stabilisé* par un idempotent e de S si $\pi(p)e = \pi(p)$.

Question 11. Montrer que tout mot u de longueur $k - 1$ admet un préfixe stabilisé par un idempotent. On note $p(u)$ le plus court de ces préfixes et on l'appelle le *préfixe critique* de u. Le plus petit idempotent e stabilisant $p(u)$ est appelé l'*idempotent critique* de u.

Question 12. Soit w un mot de longueur k. On note a sa première lettre, u son préfixe de longueur $k-1$ et v son suffixe de longueur $k-1$. On a donc $av = w$. Montrer qu'il existe un unique mot $x \in A^*$ tel que $p(u)x = ap(v)$.

Soit e $[f]$ l'idempotent critique de u ($[v]$). On associe à w l'arc de $G(S)$



Question 13. Plus généralement, on associe à un mot w la suite des arcs obtenus à partir de la suite de ses facteurs de longueur k, pris de gauche à droite. Montrer qu'on associe ainsi à w un chemin de $G(S)$.

Question 14. Montrer que si S vérifie la condition des chemins, et si $w \sim_k w'$ alors $\pi(w) = \pi(w')$. En déduire que si S vérifie la conditions des chemins, alors L est LT.

Question 15. (Vraiment difficile). Montrer que si S vérifie les équations (1) et (2), alors il vérifie la conditions des chemins.

MPRI, Fondations mathématiques de la théorie des automates

Olivier Carton, Jean-Éric Pin

March 9, 2011. Duration: 2h 30.

⋆ ⋆ ⋆

Warning : Clearness, accuracy and concision of the writing will be rewarded. Parts 2, 3 and 4 are independent.

1. An example

Consider the alphabet $A = \{a, b\}$ and the language $L = A^*abbA^*$, whose minimal automaton is given below:

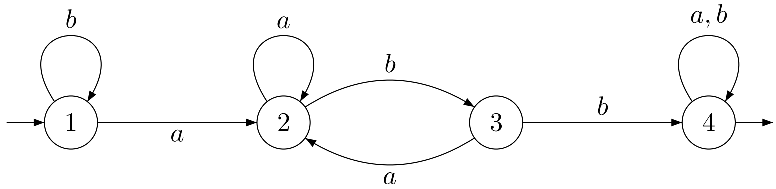


Figure 2: The minimal automaton of A^*abbA^*.

Question 1. Compute the syntactic monoid M of L (you should find 10 elements). Give the list of the idempotents of M and its $\mathcal{J}$-class structure. Is M commutative? aperiodic? Justify your answers.

Question 2. For each idempotent $e \neq 1$, compute the semigroup eMe (which is the set of elements of the form ese with $s \in M$). What can be said about these semigroups?

2. Languages A^*uA^*, where u is a nonempty word

Question 3. Let $L = A^*uA^*$ and let $\eta : A^* \to M$ be its syntactic morphism. Show that $\eta(u)$ is a zero of M, which will be denoted by 0. Show that $L = \eta^{-1}(0)$.

Question 4. Let x be a word of A^+ such that $x \sim_L x^2$. Show that for all $y, z \in A^*$, one has $xyx \leqslant_L x$, $xyxyx \sim_L xyx$ and $xyxzx \sim_L xzxyx$.

Question 5. Show that L satisfies the following equations:

For all $x \in A^+$, for all $y \in A^*$

(1) $x^\omega y x^\omega y x^\omega = x^\omega y x^\omega$ (locally idempotent)

(2) $x^\omega y x^\omega z x^\omega = x^\omega z x^\omega y x^\omega$ (locally commutative)

(3) $x^\omega y x^\omega \leqslant x^\omega$ (local maximum)

For all $x, y, s \in A^*$

(4) $s(xy)^\omega x \leftrightarrow s(xy)^\omega$ and $y(xy)^\omega s \leftrightarrow (xy)^\omega s$

(5) $x^\omega r y^\omega s x^\omega \leftrightarrow y^\omega s x^\omega r y^\omega$

Question 6. Let L be a language of A^* and let $\eta : A^* \to M$ be its syntactic morphism. Let $P = \eta(L)$. Show that if L satisfies Equation (4), then P *satures* the $\mathcal{J}$-classes of M (that is, if $s \in P$ and $s \mathrel{\mathcal{J}} t$, then $t \in P$).

3. Locally testable languages

For each positive integer k, let $\sim_k$ be the relation on A^* defined by $u \sim_k v$ if and only if

(a) u and v have the same prefixes of length $< k$,
(b) u and v have the same suffixes of length $< k$,
(c) u and v have the same factors of length k (without counting multiplicities).

For instance, $abababcbcb \sim_3 ababcbcbcb$.

Question 7. Show that $\sim_k$ is a congruence of finite index.

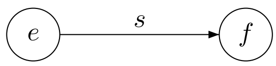
A language is *k-testable* (kT) if it is union of $\sim_k$-classes. It is *locally testable* (LT) if it is k-testable for some k.

Question 8. Show that a language is LT if and only if it is a Boolean combination of languages of the form pA^*, A^*uA^* or A^*s with $p, u, s \in A^+$.

Question 9. Show that a LT language satisfies the equations (1) and (2).

4. An algebraic characterisation.

Let L be a regular language of A^+, let S be its syntactic semigroup (not monoid!) and let $\pi : A^+ \to S$ its syntactic morphism. Let us denote by $G(S)$ the directed graph whose vertices are the idempotents of S and the edges are of the form



where $e, f \in E(S)$ and s is an element of S such that $es = s = sf$. By definition, the *label* of the path

of $G(S)$ is the product $s_1 s_2 \cdots s_n$.

Question 10. Draw $G(S)$ when S is the syntactic semigroup of the language considered in the first part of the problem (thus $S = M - \{1\}$).

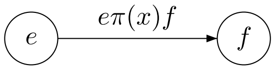
A semigroup S satisfies the *path condition* if two paths of $G(S)$ with the same origin and the same end, and containing the same edges (without counting multiplicities), have the same labels.

Let $k = |S| + 1$ and let us fix a total order on the idempotents of S: $e_1 < e_2 < \ldots < e_n$. A word $p \in A^+$ is said to be *stabilised* by an idempotent e of S if $\pi(p)e = \pi(p)$.

Question 11. Show that every word u of length $k - 1$ has a prefix which is stabilised by some idempotent. We denote by $p(u)$ the shortest of these prefixes and call it the *critical prefix* of u. The smallest idempotent e stabilising $p(u)$ is called the *critical idempotent* of u.

Question 12. Let w be a word of length k. Let a be its first letter, u be its prefix of length $k-1$ and v its suffix of length $k-1$. Thus we have $av = w$. Show that there exists a unique word $x \in A^*$ such that $p(u)x = ap(v)$.

Let e [f] be the critical idempotent of u ([v]). One associates to w the edge of $G(S)$



Question 13. More generally, one associates to a word w the sequence of edges associates to the sequence of its factors of length k, read from left to right. Show that this sequence defines a path of $G(S)$.

Question 14. Show that if S satisfies the path condition, and if $w \sim_k w'$ then $\pi(w) = \pi(w')$. Use this result to prove that if S satisfies the path condition, then L is LT.

Question 15. (Really difficult). Show that if S satisfies satisfies the equations (1) et (2), then it satisfies the path condition.

Solution

An example

Question 1. The syntactic monoid of L is generated by the following generators:

	1	2	3	4	5
a	4	3	3	4	0
b	2	2	5	5	2

Elements:

	1	2	3	4
$* \; 1$	1	2	3	4
$* \; a$	2	2	2	4
b	1	3	4	4
$* \; ab$	3	3	3	4
$* \; ba$	2	2	4	4
$* \; b^2$	1	4	4	4
$* \; ab^2$	4	4	4	4
bab	3	3	4	4
b^2a	2	4	4	4
b^2ab	3	4	4	4

Note that ab^2 is a zero of M. Thus we set $ab^2 = 0$. The other relations defining M are:

$$a^2 = a \qquad aba = a \qquad b^3 = b^2$$

Idempotents:

$$E(S) = \{1, a, ab, ba, b^2, ab^2\}$$

$\mathcal{D}$-classes:

^{*1}

b

$^*b^2$

$^*a$	$^*ab$
$^*ba$	bab

b^2a	b^2ab

$^*ab^2$

M is not commutative, since $ab \neq ba$. It is aperiodic since it is $\mathcal{H}$-trivial.

Question 2. For each idempotent $e \neq 1$, one has $eSe = \{e, 0\}$. All these monoids are idempotent and commutative.

2. Languages A^*uA^*, where u is a nonempty word

Question 3. For all $x, y \in A^*$, one has $xuy \in L$. Therefore $u \sim_L xuy$ and hence $\eta(x)\eta(u)\eta(y) = \eta(u)$. Since η is surjective, it follows that $\eta(u)$ is a zero of M

Question 4. Let x be a word of A^+ such that $x \sim_L x^2$ and let $n = |u|$. Let $s, t \in A^*$. If $sxt \in L$, then $sx^nt \in L$ since $x \sim_L x^n$. Thus u is a factor of one of the words sx^n or x^nt. In both cases, one gets $sx^nyx^nt \in L$ and thus $sxyxt \in L$ since $x \sim_L x^n$. Thus $xyx \leqslant_L x$.

This relation gives also $xyxyx \leqslant_L xyx$. We claim that $xyx \leqslant_L xyxyx$. If $sxyxyxt \in L$, then $sx^nyx^nyx^nt \in L$. Thus u is a factor of one of the words sx^n, x^nyx^n or x^nt. In all cases, one gets $sx^nyx^nt \in L$ and finally $sxyxt \in L$, which proves the claim.

For the last equation, it suffices by symmetry to prove that $xyxzx \leqslant_L xzxyx$, or equivalently, that $x^nyx^nzx^n \leqslant_L x^nzx^nyx^n$. If $sx^nzx^nyx^nt \in L$, then u is a factor of one of the words sx^n, x^nyx^n, x^nzx^n or x^nt. In all cases, one gets $sx^nyx^nzx^nt \in L$ and finally $sxzxyxt \in L$, which concludes the proof.

Question 5. Let S be the ordered syntactic semigroup of L. The previous question shows that if e is an idempotent of S and $s, t \in S$, then $ese \leqslant e$, $esese = ese$ and $esete = etese$. This gives immediately the equations (1), (2) and (3).

Equation (4). Observing that for all $x \in A^*$, $x^n \sim_L x^{n+1}$, it suffices to prove that $s(xy)^{n+1} \leftrightarrow s(xy)^{n+1}x$. The result is obvious if $x = 1$, so we may assume that $x \in A^+$. If u is a factor of $s(xy)^{n+1}$, then it is also a factor of $s(xy)^{n+1}x$. If u is a factor of $s(xy)^{n+1}x$, then it is a factor of one of the words $s(xy)^{n+1}$ or $(yx)^n$. Since $(yx)^n$ is itself a factor of $s(xy)^{n+1}$, we have proved that $s(xy)^\omega x \leftrightarrow s(xy)^\omega$. A similar proof would show that $y(xy)^\omega s \leftrightarrow (xy)^\omega s$.

Equation (5). By symmetry, it suffices to prove that $x^nry^nsx^n \to y^nsx^nry^n$. If u is a factor of $x^nry^nsx^n$, then it is a factor of one of the words x^nry^n or y^nsx^n. In both cases it is a factor of $y^nsx^nry^n$.

Question 6. We claim that P saturates the $\mathcal{R}$-classes. Let $s, t \in S$ be such that $s \mathrel{\mathcal{R}} t$ and suppose that $s \in P$. Then $t = sx$ and $s = ty$ for some $x, y \in S$. It follows that $s(xy) = ty = s$ and thus $s(xy)^\omega = s$. Thus $s(xy)^\omega \in P$. Since L satisfies the equation $s(xy)^\omega x \leftrightarrow s(xy)^\omega$, one also gets $s(xy)^\omega x \in P$. Thus $t \in P$, which proves the claim.

A symmetric argument would show that P saturates the $\mathcal{L}$ classes and since $\mathcal{J} = \mathcal{D}$, P saturates the $\mathcal{J}$-classes.

3. Locally testable languages

Question 7. It is clear that $\sim_k$ is an equivalence relation. It is also a congruence since if $u \sim_k v$ and a is a letter, then $au \sim_k av$ and $ua \sim_k va$. Finally, $\sim_k$ has finite index since the equivalence classes depend only of the following parameters: the prefixes of length $< k$, the suffixes of length $< k$ and the factors of length k. In each case there are only finitely many possible choices.

Question 8. Let $k = |p|$. If $u \in pA^*$ and $u \sim_k v$ then p is a prefix of v and thus $v \in pA^*$. It follows that pA^* is LT. A similar argument would show that A^*uA^* and A^*s are LT.

Let $x \in A^+$. If $|x| < k$, the $\sim_k$-class of x is $\{x\}$, which can be written as $xA^* - \bigcup_{a \in A} A^*xaA^*$. If $|x| \geqslant k$, let p [s] be its prefix [suffix] of length $k - 1$ and let F be the set of its factors of length

k. Then the $\sim_k$-class of x is the set

$$pA^* \cap A^*s \cap \Big(\bigcap_{u \in F} A^*uA^* \setminus \bigcup_{u \in A^k - F} A^*uA^*\Big)$$

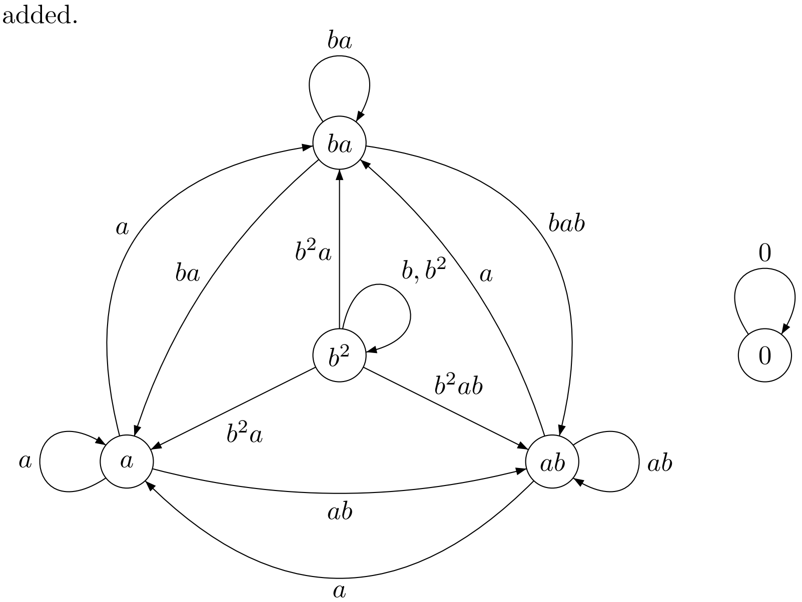
In all cases it is a Boolean combination of languages of the form pA^*, A^*uA^* or A^*s.

Question 9. We have already seen that a language of the form A^*uA^* satisfies the equations (1) and (2). The languages of the form pA^* or A^*s (see the notes). It follows that all LT languages satisfy these equations.

4. An algebraic characterisation.

Question 10.

The graph $G(S)$ is partially represented in the picture below. The edges of the form $(e, 0, f)$ should be added.



Question 11. See Proposition II.6.4.

Question 12. The words $p(u)$ and $ap(v)$ are both prefixes of w. But $ap(v)$ cannot be strictly shorter than $p(u)$, since the critical idempotent of v stabilises $p(v)$ and hence $ap(v)$. Thus $|p(u)| \geqslant |ap(v)|$ and there exists a unique word $x \in A^*$ such that $p(u)x = ap(v)$.

Question 13. In the edge (e, s, f) generated by a factor of length k, the idempotent e $[f]$ depends only on the prefix [suffix] of length $k - 1 = |S|$. Thus the sequence defines a path of $G(S)$.

Question 14. Suppose that $w \sim_k w'$. If $|w| < k$ or $|w'| < k$, then $w = w'$ and the result is obvious. Otherwise, let p $[s]$ be the common prefix [suffix] of length $k - 1$ of w. Let e $[f]$ be the critical idempotent of p $[s]$. The paths defined by w and w' have same origin, same end, and go through the same edges. By the path condition they have the same label x. Now $\pi(w) = \pi(p)exf\pi(s) = \pi(w')$.

Thus if S satisfies the path condition, L is LT.