MPRI, Fondations mathématiques de la théorie des automates

Olivier Carton, Jean-Éric Pin

Examen du 7 mars 2016, 13h 15. 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.

Soit A un alphabet. Dans ce problème, on s'intéresse aux opérations $L \to LaA^*$, où a est une lettre, et à une variante de cette opération, notée $(LaA^*)_{r,n}$, et décrite plus loin. On note $\sim_L$ la congruence syntactique d'un langage L.

1. Un exemple.

Question 1. Soit $A = \{a, b\}$ et $L = (A^2)^*$. Calculer l'automate minimal de LaA^* (on trouvera 3 états).

Question 2. Calculer le monoïde syntactique M de LaA^* (on trouvera 7 éléments). Donner la liste des éléments de M et sa structure en $\mathcal{J}$-classes. Est-ce que M est commutatif, apériodique, $\mathcal{J}$-trivial, idempotent?

Question 3. Calculer l'ensemble $E(M)$ des idempotents de M et le sous-monoïde de M engendré par $E(M)$. Est-ce que ce monoïde est commutatif, apériodique, $\mathcal{J}$-trivial, idempotent?

2. Propriétés de l'opération $L \to LaA^*$.

Question 4. Soit L un langage de A^* et soient $u, v, z \in A^*$. Montrer que si $z \sim_L zu \sim_L zv$, alors $zu^2 \sim_{LaA^*} zu$ et $zuv \sim_{LaA^*} zvu$.

On rappelle que si G est un groupe fini à n éléments, alors $x^n = 1$ pour tout $x \in G$. On appelle *langage à groupe* un langage dont le monoïde syntactique est un groupe fini.

Question 5. Montrer qu'un langage est à groupe si et seulement si il est reconnu par un automate de permutations (i.e. pour chaque lettre a, l'application $q \to q \cdot a$ est une permutation de l'ensemble des états).

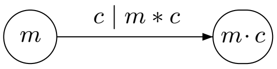
Question 6. Montrer que si L est un langage à groupe, alors les idempotents du monoïde syntactique de LaA^* commutent (on pourra faire une preuve directe ou utiliser la question 4 avec $z = 1$).

3. Une opération sur les langages

Soient L un langage reconnaissable sur l'alphabet A, a une lettre de A et u un mot de A^*. On note $N(u)$ le nombre de factorisations de u de la forme $u = u_0au_1$ avec $u_0 \in L$ et $u_1 \in A^*$.

Question 7. On prend $A = \{a, b\}$ et $L = (ab)^*$. Calculer $N(u)$ lorsque u est l'un des mots suivants: ab, $abab$, $ababab$, $ababbab$.

Question 8. On se propose de construire un transducteur séquentiel qui calcule $N(u)$. On note $\eta : A^* \to M$ le morphisme syntactique de L et on pose $P = \eta(L)$. Si $m \in M$ et $c \in A$, on pose $m \cdot c = m\eta(c)$. On sait alors que l'automate $(M, A, \cdot, 1, P)$ reconnaît le langage L. Comment ajouter à cet automate une fonction de sortie à valeurs dans le monoïde $(\mathbb{N}, +)$, de façon à ce que le transducteur séquentiel résultant calcule $N(u)$? (on notera que le monoïde $(\mathbb{N}, +)$ est isomoorphe au monoïde a^* si vous préférez rester dans le cadre strict des langages).



Autrement dit, quelles valeurs choisir pour chaque $m * c$ pour que $1 * u = N(u)$ pour tout $u \in A^*$?

Question 9. Soient r et n deux entiers tels que $0 \leqslant r < n$. On pose

$$(LaA^*)_{r,n} = \big\{u \in A^* \mid N(u) \equiv r \bmod n\big\}$$

Déduire de la question 8 que $(LaA^*)_{r,n}$ est reconnu par le produit en couronne $\mathbb{Z}/n\mathbb{Z} \circ M$.

Question 10. En déduire que si L est un langage à groupe, alors $(LaA^*)_{r,n}$ est aussi un langage à groupe. (On admettra sans démonstration que le produit en couronne de deux groupes est un groupe).

4. Un peu de topologie.

On note $\mathbf{G}$ la variété des groupes finis et on s'intéresse à $d_{\mathbf{G}}$ définie page 138. On a donc

$$r_{\mathbf{V}}(u, v) = \min\big\{\mathrm{Card}(G) \bigm| G \text{ est un groupe fini qui separe } u \text{ et } v\ \big\}$$

et $d_{\mathbf{V}}(u, v) = 2^{-r_{\mathbf{V}}(u,v)}$ avec les conventions habituelles $\min \emptyset = +\infty$ et $2^{-\infty} = 0$.

Question 11. Soient deux mots distincts. Montrer qu'il existe un automate de permutations fini qui accepte l'un des mots et rejette l'autre. En déduire qu'il existe un groupe fini qui sépare les deux mots.

A partir de là, il est facile de voir (et on admettra) que $d_{\mathbf{G}}$ est une distance ultramétrique sur A^*. On s'intéresse maintenant à la topologie sur A^* définie par $d_{\mathbf{G}}$.

Question 12. Montrer que pour tout mot u de A^*, on a $\lim_{n\to\infty} u^{n!} = 1$. En déduire que si L est une partie fermée de A^*, et si $xu^+y \subseteq L$, alors $xu^*y \subseteq L$.

Question 13. On suppose l'alphabet non vide. Montrer que tout langage fini non vide est fermé mais n'est pas ouvert.

Question 14. Montrer que les langages à groupe de A^* sont à la fois ouvert et fermés dans l'espace métrique $(A^*, d_{\mathbf{G}})$. En déduire que si L est un langage à groupe, les langages $(LaA^*)_{r,n}$ sont ouverts, puis que le langage LaA^* est aussi ouvert.

MPRI, Mathematical foundations of automata theory

Olivier Carton, Jean-Éric Pin

March 7, 2016. Duration: 2h 30.

⋆ ⋆ ⋆

Warning : Clearness, accuracy and concision of the writing will be rewarded.

Let A be an alphabet. In this problem, we are interested in the operations $L \to LaA^*$, where a is a letter, and to a variant of this operation, denoted by $(LaA^*)_{r,n}$ (see below). We denote by $\sim_L$ the syntactic congruence of a language L.

1. An example.

Question 1. Let $A = \{a, b\}$ and $L = (A^2)^*$. Compute the minimal automaton of LaA^* (you should find 3 states).

Question 2. Compute the syntactic monoid M of LaA^* (you should find 7 elements). Give the list of all elements de M and its $\mathcal{J}$-class structure. Is M commutative, aperiodic, $\mathcal{J}$-trivial, idempotent?

Question 3. Compute the set $E(M)$ of idempotents of M and the submonoid of M generated by $E(M)$. Is this monoid commutative, aperiodic, $\mathcal{J}$-trivial, idempotent?

2. Properties of the operation $L \to LaA^*$.

Question 4. Let L be a language of A^* and let $u, v, z \in A^*$. Show that if $z \sim_L zu \sim_L zv$, then $zu^2 \sim_{LaA^*} zu$ and $zuv \sim_{LaA^*} zvu$.

Recall that if G is an n-element finite group, then $x^n = 1$ for all $x \in G$. A *group language* is a language whose syntactic monoid is a finite group.

Question 5. Show that a language is a group language if and only if it is recognised by a permutation automaton (i.e. for each letter a, the map $q \to q \cdot a$ is a permutation of the set of states).

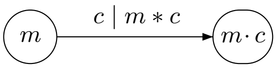
Question 6. Show that if L is a group language, then the idempotents of the syntactic monoid of LaA^* commute (one could either give a direct proof or use question 4 with $z = 1$).

3. An operation on languages

Let L be a recognisable language on the alphabet A, a a letter of A and u a word of A^*. Let $N(u)$ denote the number of factorisations of u of the form $u = u_0au_1$ with $u_0 \in L$ et $u_1 \in A^*$.

Question 7. Let $A = \{a, b\}$ and $L = (ab)^*$. Compute $N(u)$ when u is one of the following words: ab, $abab$, $ababab$, $ababbab$.

Question 8. The purpose of this question is to find a sequential transducer computing $N(u)$. Let $\eta : A^* \to M$ be the syntactic morphism of L and let $P = \eta(L)$. If $m \in M$ and $c \in A$, let us set $m\cdot c = m\eta(c)$. Then we know that the automaton $(M, A, \cdot\,, 1, P)$ recognises the language L. How to supplement this automaton with an output function taking values in the monoid $(\mathbb{N}, +)$, in such a way that the resulting sequential transducer computes $N(u)$? (Note that the monoid $(\mathbb{N}, +)$ is isomorphic to the monoid a^*, if you prefer to stay in the context of language theory).



In other words, which values should one take for each $m * c$ in order to have $1 * u = N(u)$ for all $u \in A^*$?

Question 9. Let r and n be two integers such that $0 \leqslant r < n$. Let us set

$$(LaA^*)_{r,n} = \bigl\{u \in A^* \mid N(u) \equiv r \bmod n\bigr\}$$

Deduce from question 8 that $(LaA^*)_{r,n}$ is recognised by the wreath product $\mathbb{Z}/n\mathbb{Z} \circ M$.

Question 10. Conclude that, if L is a group language, then $(LaA^*)_{r,n}$ also is a group language. (You may assume without proof that the wreath product of two groups is a group).

4. A bit of topology.

Let $\mathbf{G}$ denote the variety of finite groups and let $d_{\mathbf{G}}$ be the metric defined on page 138. Thus one has

$$r_{\mathbf{V}}(u, v) = \min\bigl\{\mathrm{Card}(G) \bigm| G \text{ is a finite group separating } u \text{ and } v\bigr\}$$

and $d_{\mathbf{V}}(u, v) = 2^{-r_{\mathbf{V}}(u,v)}$ with the usual convention $\min \emptyset = +\infty$ and $2^{-\infty} = 0$.

Question 11. Consider two distinct words. Prove that there exists a permutation automaton accepting one of the words and rejecting the other one. Conclude that there is a finite group separating the two words.

From there, it is easy to see (and you don't need to prove it) that $d_{\mathbf{G}}$ is an ultrametric on A^*. We are interested now in the topology on A^* defined by $d_{\mathbf{G}}$.

Question 12. Show that for each $u \in A^*$, one has $\lim_{n\to\infty} u^{n!} = 1$. Conclude that if L is a closed subset of A^*, and if $xu^+y \subseteq L$, the $xu^*y \subseteq L$.

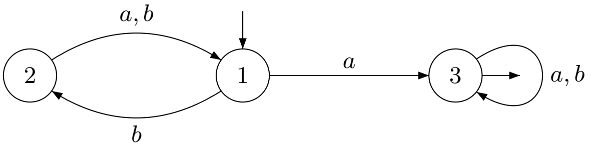
Question 13. Suppose that the alphabet is nonempty. Show that every finite nonempty language is closed but is not open.

Question 14. Show that the group languages of A^* are clopen in the metric space $(A^*, d_{\mathbf{G}})$. Conclude that if L is a group language, then the languages $(LaA^*)_{r,n}$ are open, and then that the language LaA^* is also open.

Solution

1. An example

Question 1. The minimal automaton of $(A^2)^*aA^*$ is represented below



Question 2. The elements of the syntactic monoid of $(A^2)^*aA^*$ are given below

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

Relations: $b^2 = 1$, $a^2 = 0$ and $aba = a$.
$\mathcal{D}$-classes:

$^*1\ \ b$

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

^{*0}

This monoid is not commutative, since $ab \neq ba$. This monoid is not aperiodic, since the identity $x^\omega = x^{\omega+1}$ is not satisfied for $x = b$.

Question 3. One gets $E(M) = \{1, ab, ba, 0\}$. The idempotents of M commute and thus the monoid generated by $E(M)$ is equal to $E(M)$. This monoid is idempotent and hence commutative, aperiodic and $\mathcal{J}$-trivial.

2. Properties of the operation $L \to LaA^*$.

Question 4. Suppose that $z \sim_L zu \sim_L zv$. Let us prove that $zu^2 \sim_{LaA^*} zu$.

If $xzu^2y \in LaA^*$ then $xzu^2y = v_0av_1$ for some $v_0 \in L$ and $v_1 \in A^*$. If $|xzu| \leqslant |v_0|$, then xzu is a prefix of v_0, and $v_0 = xzus$ for some $s \in A^*$ such that $sav_1 = uy$. Since $z \sim_L zu$, it follows that $xzs \in L$ and thus $xzsav_1 = xzuy \in LaA^*$. If now $|v_0| < |xzu|$, then v_0a is a prefix of xzu, that is $xzu = v_0as$ for some s such that $suy = v_1$. Then $xzuy = v_0asv \in LaA^*$. Thus $xzu^2y \in LaA^*$ implies that $xzuy \in LaA^*$.

In the opposite direction, suppose that $xzuy \in LaA^*$. Then $xzuy = v_0av_1$ for some $v_0 \in L$ and $v_1 \in A^*$. If $|xz| \leqslant |v_0|$, then xz is a prefix of v_0, and $v_0 = xzs$ for some $s \in A^*$ such that $sav_1 = uy$. Since $z \sim_L zu$ and $xzs \in L$, it follows that $xzus \in L$ and thus $xzusav_1 = xzuuy \in LaA^*$. If now $|v_0| < |xz|$, then v_0a is a prefix of xz, that is $xz = v_0as$ for some s such that $suy = v_1$. Then $xzu^2y = v_0asu^2y \in LaA^*$. Thus $xzuy \in LaA^*$ implies that $xzu^2y \in LaA^*$ and thus $zu^2 \sim_{LaA^*} zu$.

Let us prove now that $zuv \sim_{LaA^*} zvu$. By symmetry, it suffices to prove that $zuv \leqslant_{LaA^*} zvu$, that is, $xzuvy \in LaA^*$ implies $xzvuy \in LaA^*$. If $xzuvy \in LaA^*$, then $xzuvy = v_0av_1$ for some $v_0 \in L$ and $v_1 \in A^*$. If $|xzuv| \leqslant |v_0|$, then $v_0 = xzuvs$ for some $s \in A^*$ such that $sav_1 = y$. Now since $z \sim_L zu \sim_L zv$, one gets $zuv \sim_L zvu$ and since $xzuvs \in L$, $xzvus \in L$. It follows that $xzvuy = xzvusav_1 \in LaA^*$. If now $|xzu| \leqslant |v_0| < |xzuv|$, then $v_0 = xzus$, $sat = v$ and $ty = v_1$ for some $s, t \in A^*$. Since $z \sim_L zu$ and $xzus \in L$, one gets $xzs \in L$ and thus $xzvuy = xzsatuy \in LaA^*$. Finally, if $|xzuv| \leqslant |v_0|$, then $v_0 = xzuvs$ for some $s \in A^*$ such that $sav_1 = y$. Since $zuv \sim_L zvu$ and $xzuvs \in L$, one gets $xzvus \in L$, and thus $xzvuy = xzvusav_1 \in LaA^*$.

Question 5. If L is recognised by a permutation automaton, then its minimal automaton also is a permutation automaton and its syntactic monoid is a permutation group. Conversely, if L is recognized by a finite group G, then the permutation automaton $(G, A, \cdot, 1, P)$ recognises L.

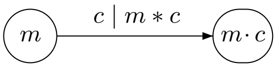
Question 6. Let L be a group language and let G be its syntactic monoid (which is a group). Let u and v be two words of LaA^* such that $u \sim_{LaA^*} u^2$ and $v \sim_{LaA^*} v^2$. Let $n = |G|$. Then $1 \sim_L u^n \sim_L v^n$ and by the previous question applied with $z = 1$, one gets $uv \sim_{LaA^*} vu$. If follows that the idempotents of the syntactic monoid of LaA^* commute.

3. An operation on languages

Let L be a recognisable language on the alphabet A, a a letter of A and u a word of A^*. Let $N(u)$ denote the number of factorisations of u of the form $u = u_0au_1$ with $u_0 \in L$ et $u_1 \in A^*$.

Question 7. Let $A = \{a, b\}$ and $L = (ab)^*$. Then $N(ab) = 1$ since $ab = (1)a(ab)$, $N(abab) = 2$ since $abab = (1)a(bab) = (ab)a(b)$, $N(ababab) = 3$ since $ababab = (1)a(babab) = (ab)a(bab) = (abab)a(b)$ and $N(ababbab) = 2$ since $ababbab = (1)a(babbab) = (ab)a(bbab)$.

Question 8. Let $\eta : A^* \to M$ be the syntactic morphism of L and let $P = \eta(L)$. Let $\mathcal{T} = (M, A, \cdot, *, 1, P)$ be the sequential transducer in which the transitions are of the form



where

$$m * c = \begin{cases} 1 & \text{if } c = a \text{ and } m \in P \\ 0 & \text{otherwise} \end{cases}$$

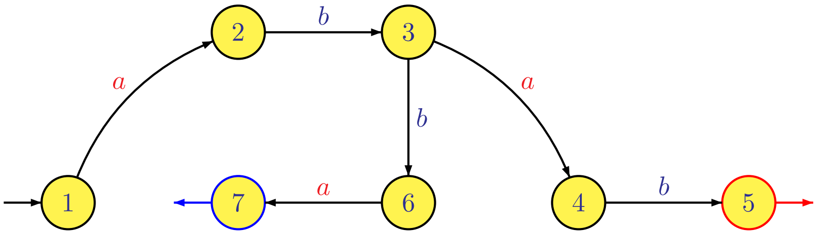
Then $\mathcal{T}$ computes $N(u)$.

Question 9. Let r and n be two integers such that $0 \leqslant r < n$. since $(LaA^*)_{r,n} = \{u \in A^* \mid N(u) \equiv r \bmod n\}$ and since the set $\{k \in \mathbb{N} \mid k \equiv r \bmod n\}$ is a recognisable subset of $\mathbb{N}$ recognized by $\mathbb{Z}/n\mathbb{Z}$, the language $(LaA^*)_{r,n}$ is recognised by the wreath product $\mathbb{Z}/n\mathbb{Z} \circ M$.

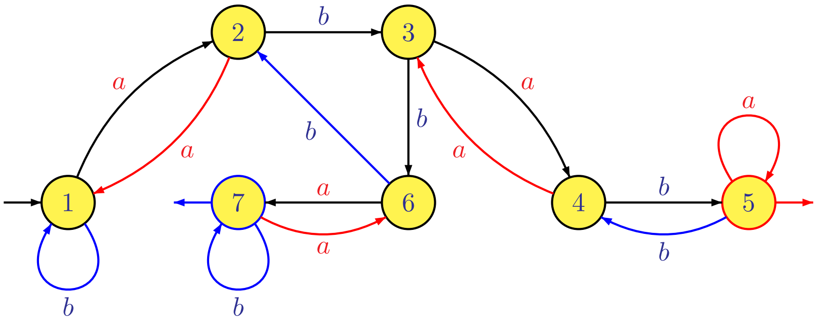
Question 10. In particular, if L is a group language, then M is a group and so is $\mathbb{Z}/n\mathbb{Z} \circ M$. Thus $(LaA^*)_{r,n}$ is also a group language.

4. A bit of topology.

Question 11. The construction is better explained on an example. Let $u = abab$ and $v = abba$. We first compute a deterministic automaton in which the two words arrive to different states, when read from the initial state.



Observe that each letter induces an injective partial function. Let us complete these partial injective partial functions into permutations in some arbitray way. For instance:



The resulting permutation group separates $abab$ and $abba$ since $1 \cdot abab = 5$ and $1 \cdot baba = 7$.

Question 12. Let $u \in A^*$ and let $\pi : A^* \to G$ be a monoid morphism, where G is a group of order $\leqslant n$. Then $|G|$ divides $n!$ and thus $g^{n!} = 1$ for all $g \in G$. It follows that $\pi(u^{n!}) = 1$ and thus $d_{\mathbf{G}}(u^{n!}, 1) < 2^{-n}$. It follows that $\lim_{n\to\infty} u^{n!} = 1$.

Since the product is continuous, one gets $\lim_{n\to\infty} xu^{n!}y = x(\lim_{n\to\infty} u^{n!})y = xy$. It follows that if L is a closed subset of A^* and if $xu^+y \subseteq L$, then $xy \in L$ and thus $xu^*y \subseteq L$.

Question 13. In a metric space, every finite set is closed. However, if F is finite and nonempty, F is not open. Indeed, suppose that F is open. Then F^c is closed. Let $x \in F$ and let u be a nonempty word. Since F is finite, $xun! \in F^c$ for n large enough and since F^c is closed, $\lim_{n\to\infty} xu^{n!} = x$ belongs to F^c. Contradiction.

Question 14. The group languages of A^* are clopen in the metric space $(A^*, d_{\mathbf{G}})$. In particular, the languages $(LaA^*)_{r,n}$ are open, and since $LaA^* = \bigcup_{n>0}(LaA^*)_{1,n}$, the language LaA^* is also open.