MPRI, Fondations mathématiques de la théorie des automates

Olivier Carton, Jean-Éric Pin

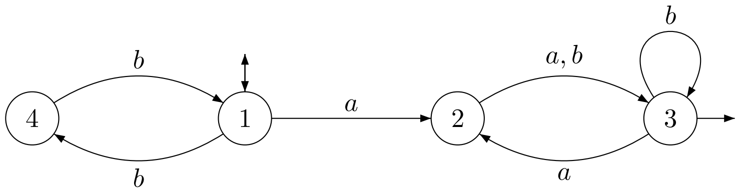
Examen du 14 mars 2013. Durée: 2h 30, notes de cours autorisées

⋆⋆⋆

Avertissement : On attachera une grande importance à la clarté, à la précision et à la concision de la rédaction. Les deux parties sont indépendantes.

1. Étude d'un langage

On considère sur l'alphabet $A = \{a, b\}$ le langage $L = (aa + ab + abb + aab + bb)^*$, dont voici l'automate minimal incomplet:



Question 1. Calculer le monoïde syntactique de L (on trouvera 9 éléments).

Question 2. Quels sont les idempotents de M?

Question 3. Déterminer la structure en $\mathcal{D}$-classes de M.

Question 4. Le monoïde M est-il apériodique? $\mathcal{R}$-trivial? $\mathcal{L}$-trivial?

Question 5. Parmi ces identités profinies, quelles sont celles qui sont satisfaites par M? (Justifier vos réponses).

(1) $x^3 = x$
(2) $(xy)^\omega(yx)^\omega(xy)^\omega = (xy)^\omega$
(3) $(xy)^\omega(yx)(xy)^\omega = (xy)^\omega$

2. Langages bègues

Soit A un alphabet. On dit qu'un langage L de A^* est *bègue* si, pour toute lettre $a \in A$, et pour tout $x, y \in A^*$, on a $xay \in L$ si et seulement si $xaay \in L$. Cette condition s'exprime aussi en disant que pour toute lettre $a \in A$, $a \sim_L a^2$. Le but du problème est d'étudier la classe $\mathcal{C}$ des langages qui sont à la fois bègues et testables par morceaux.

On appelle *bègue-élémentaire* un langage de la forme

$$A^*a_1A^*a_2 \cdots A^*a_kA^*$$

où les a_i sont des lettres telles que $a_i \neq a_{i+1}$, pour $1 \leqslant i \leqslant k-1$.

Question 6. Montrer que tout langage bègue-élémentaire est bègue.

Question 7. Montrer que toute combinaison booléenne de langages bègues-élémentaires appartient à $\mathcal{C}$.

Question 8. [Plus difficile] Montrer que, réciproquement, tout langage de $\mathcal{C}$ est combinaison booléenne de langages bègues-élémentaires.

Question 9. Donner un ensemble d'équations profinies définissant la classe $\mathcal{C}$.

Question 10. Dire, en le justifiant, si $\mathcal{C}$ est fermée pour chacune des opérations suivantes: intersection finie, union finie, quotients, inverses de morphismes augmentant la longueur, respectivement diminuant la longueur.

On considère la signature $\{\mathbf{a} \mid a \in A\} \cup \{\leqslant\}$, où le symbole $\leqslant$ est interprété comme la relation d'ordre habituelle sur les entiers et chaque symbole **a** a son interprétation usuelle. On prendra garde que cette signature diffère de la signature $\{\mathbf{a} \mid a \in A\} \cup \{<, =\}$ utilisée habituellement.

On s'intéresse aux fragments du premier ordre $\Sigma_1[\leqslant]$ et $\mathcal{B}\Sigma_1[\leqslant]$.

Question 11. Donner une formule de $\Sigma_1[\leqslant]$ définissant le langage $A^*aA^*bA^*aA^*$ $(A = \{a, b, c\})$.

Question 12. Montrer que tout langage bègue-élémentaire peut être défini par une formule de $\Sigma_1[\leqslant]$.

Question 13. Montrer qu'un langage est défini par une formule de $\mathcal{B}\Sigma_1[\leqslant]$ si et seulement si il appartient à $\mathcal{C}$.

Question 14. Donner un algorithme pour décider si un langage reconnaissable, donné par son automate minimal, est définissable par une formule de $\mathcal{B}\Sigma_1[\leqslant]$. Evaluer la complexité de votre algorithme.

MPRI, Fondations mathématiques de la théorie des automates

Olivier Carton, Jean-Éric Pin

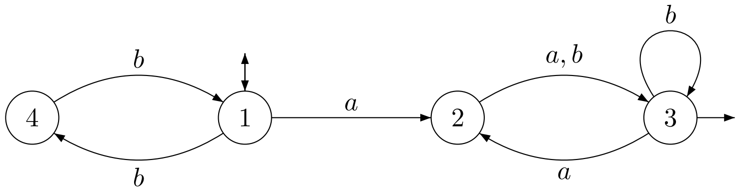
March 13, 2013. Duration: 2h 30.

⋆ ⋆ ⋆

Warning : Clearness, accuracy and concision of the writing will be rewarded. The two parts are independent one from the other.

1. Study of a language

Consider the alphabet $A = \{a, b\}$ and the language $L = (aa + ab + abb + aab + bb)^*$, whose minimal incomplete automaton is given below:



Question 1. Compute the syntactic monoid M (you should find 9 elements).

Question 2. Find the idempotents of M?

Question 3. Give the $\mathcal{D}$-class structure of M.

Question 4. Is M aperiodic? $\mathcal{R}$-trivial? $\mathcal{L}$-trivial?

Question 5. Among these profinite identities, which ones are satisfied by M? (justify your answers)

(1) $x^3 = x$
(2) $(xy)^\omega(yx)^\omega(xy)^\omega = (xy)^\omega$
(3) $(xy)^\omega(yx)(xy)^\omega = (xy)^\omega$

2. Stutter-invariant languages

Let A be an alphabet. A language L of A^* is said to be *stutter-invariant* if, for each letter $a \in A$ and for all $x, y \in A^*$, one has $xay \in L$ if and only if $xaay \in L$. This is equivalent to saying that, for each letter $a \in A$, $a \sim_L a^2$. The aim of this problem is to study the class $\mathcal{C}$ of all languages that are both stutter-invariant and piecewise testable.

A language of the form

$$A^*a_1A^*a_2 \cdots A^*a_kA^*$$

where the a_i are letters such that $a_i \neq a_{i+1}$, for $1 \leqslant i \leqslant k-1$, is said to be *elementary stutter-invariant*

Question 6. Prove that any langage elementary stutter-invariant is stutter-invariant.

Question 7. Prove that any Boolean combination of elementary stutter-invariant languages belongs to $\mathcal{C}$.

Question 8. [More difficult] Prove that, conversely, each language of $\mathcal{C}$ is a Boolean combination of elementary stutter-invariant languages.

Question 9. Give a set of profinite equations defining the class $\mathcal{C}$.

Question 10. State whether $\mathcal{C}$ is closed under each of the following operations (justify your answer): finite intersection, finite union, quotients, inverses of length-increasing morphisms, inverses of length-decreasing morphisms.

Consider the signature $\{\mathbf{a} \mid a \in A\} \cup \{\leqslant\}$, where the symbol $\leqslant$ is interpreted as the usual order relation on integers and each symbol $\mathbf{a}$ has his usual interpretation. Be aware that this signature differs from the usual signature $\{\mathbf{a} \mid a \in A\} \cup \{<, =\}$.

We are interested in the first order fragments $\Sigma_1[\leqslant]$ et $\mathcal{B}\Sigma_1[\leqslant]$.

Question 11. Give a $\Sigma_1[\leqslant]$-formula defining the language $A^*aA^*bA^*aA^*$ ($A = \{a, b, c\}$).

Question 12. Prove that each elementary stutter-invariant language can be defined by a $\Sigma_1[\leqslant]$ formula.

Question 13. Prove that a language can be defined by a $\mathcal{B}\Sigma_1[\leqslant]$-formula if and only if it belongs to $\mathcal{C}$.

Question 14. Give an algorithm to decide whteher a recognizable language, given by its minimal automaton, can be defined by a $\mathcal{B}\Sigma_1[\leqslant]$ formula. Analyse the complexity of your algorithm.

Solution

1. Study of a language

Here is the syntactic monoid of L

	1	2	3	4
$* \; 1$	1	2	3	4
a	2	3	2	0
b	4	3	3	1
$* \; a^2$	3	2	3	0
$* \; ab$	3	3	3	0
$* \; ba$	0	2	2	2
$* \; b^2$	1	3	3	4
$* \; aba$	2	2	2	0
$* \; ba^2$	0	3	3	3

Relations:

$$a^3 = a \qquad a^2b = ab \qquad ab^2 = ab \qquad bab = ba^2 \qquad b^2a = aba \qquad b^3 = b \qquad aba^2 = ab$$

Idempotents:

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

$\mathcal{D}$-class structure:

^{*1}

$^*a^2$	a

$^*b^2$	b

$^*ab$	$^*aba$
$^*ba^2$	$^*ba$

This monoid is not aperiodic, since the identity $x^\omega = x^{\omega+1}$ is not satisfied for $x = a$. This monoid is neither $\mathcal{R}$-trivial nor $\mathcal{L}$-trivial.

This monoid satisfies the identities $x^3 = x$ (all elements are regular and all groups have order 1 or 2). It also satisfies the identity $(xy)^\omega(yx)^\omega(xy)^\omega = (xy)^\omega$. Indeed, $(xy)^\omega$ and $(yx)^\omega$ are two conjugated idempotents e and f. If e and f belong to the minimal ideal, then $efe = e$. Otherwise, $e = f$ and the result is trivial. The identity $(xy)^\omega(yx)(xy)^\omega = (xy)^\omega$ is not satisfied for $x = a$ and $y = 1$.

2. Stutter-invariant languages

This problem was inspired by the article by M. Kufleitner and A. Lauser, Lattices of Logical Fragments over Words, *CoRR* **abs/1202.3355** (2012).

Question 6. Let $L = A^*a_1A^*a_2 \cdots A^*a_kA^*$ be an elementary stutter-invariant language. Then L is by construction piecewise testable. Let a be a letter and $x, y \in A^*$. If $xay \in L$, then $xa^2y \in L$ since xay is a subword of xa^2y. Conversely, if $xa^2y \in L$, then $xa^2y = u_0a_1u_1 \cdots a_ku_k$ for some words $u_0, ..., u_k \in A^*$. Since two consecutive a_i are distinct, one of the occurrences of the two letters a is inside one of the words u_i. It follows that $xay \in L$. Therefore $a \sim_L aa$ and hence L is stutter-invariant.

Question 7. Any Boolean combination of elementary stutter-invariant languages is by definition piecewise testable. Note that the stutter-invariant languages are characterized by the equations $a = a^2$, for all $a \in A$. Therefore, they are closed under Boolean operations. In particular, a Boolean combination of elementary stutter-invariant languages is stutter-invariant.

Question 8. Let L be a language of $\mathcal{C}$. Since L is piecewise testable it can be written as

$$\bigcup_{1 \leqslant i \leqslant s} S_i \setminus \bigcup_{1 \leqslant i \leqslant t} T_i$$

where S_i and T_i are languages of the form

$$A^*a_1A^*a_2 \cdots A^*a_kA^*.$$

Let us rewrite such a language as

$$K = (A^*a_1)^{e_1}(A^*a_2)^{e_2} \cdots (A^*a_n)^{e_n}A^*$$

where $e_1, ..., e_n$ are positive integers and $a_i \neq a_{i+1}$ for $1 \leqslant i \leqslant n - 1$. Let

$$R(K) = A^*a_1A^*a_2 \cdots A^*a_nA^*$$

By construction, $R(K)$ is stutter-invariant and $K \subseteq R(K)$. We claim that

$$L = \bigcup_{1 \leqslant i \leqslant s} R(S_i) \setminus \bigcup_{1 \leqslant i \leqslant t} R(T_i) \tag{1}$$

Let R be the right member of (1). We first prove the inclusion $L \subseteq R$. Let $u \in L$. Then, for $1 \leqslant i \leqslant s$, $u \in S_i$ and hence $u \in R(S_i)$ since $S_i \subseteq R(S_i)$. Suppose that $u \in R(T_i)$ for some i. Let $R(T_i) = A^*a_1A^*a_2 \cdots A^*a_nA^*$. Then $u = u_0a_1u_1 \cdots a_nu_n$ for some $u_0, ..., u_n \in A^*$. Therefore there exist positive integers e_i such that the word $v = u_0a_1^{e_1}u_1 \cdots a_n^{e_n}u_n$ belongs to T_i. Thus, $v \notin L$ and, by stutter-invariance of L, u does not belong to L. Thus $L \subseteq R$.

Let now $u \in R$. Then u belongs to some $R(S_i)$ and $u \notin \bigcup_{1 \leqslant i \leqslant t} R(T_i)$. Let $R(S_i) = A^*a_1A^*a_2 \cdots A^*a_nA^*$ where $a_i \neq a_{i+1}$ for $1 \leqslant i \leqslant n - 1$ and let $u = u_0a_1u_1 \cdots a_nu_n$. Then there exist positive integers e_i such that the word $v = u_0a_1^{e_1}u_1 \cdots a_n^{e_n}u_n$ belongs to S_i. By stutter-invariance of $R(S_i)$, we get $v \notin \bigcup_{1 \leqslant i \leqslant t} R(T_i)$. In particular, $v \notin \bigcup_{1 \leqslant i \leqslant t} T_i$ and thus $v \in L$. By stutter-invariance of L, we get $u \in L$.

Question 9. The class $\mathcal{C}$ is defined by the identities defining the piecewise testable languages, for instance $(xy)^\omega x = (xy)^\omega = y(xy)^\omega$ for all $x, y \in A^*$ and by the equations $a^2 = a$ for all $a \in A$.

Question 10. It follows that $\mathcal{C}$ is closed under Boolean operations and quotients. It is also closed under inverses of length-decreasing morphisms.

However, $\mathcal{C}$ is not closed under inverses of length-increasing morphisms. Consider for instance the morphism $\varphi : a^* \to A^*$ (where $A = \{a, b\}$) defined by $\varphi(a) = ab$. Then $L = A^*aA^*bA^*aA^*bA^*$ belongs to $\mathcal{C}$, but $\varphi^{-1}(L) = a^*aa^*aa^*$ does not belong to $\mathcal{C}$.

Question 11. One has $A^*aA^*bA^*aA^* = L(\varphi)$ with $\varphi = \exists x\ \exists y\ \exists z\ (x \leqslant y) \wedge (y \leqslant z) \wedge \mathbf{a}x \wedge \mathbf{b}y \wedge \mathbf{a}z$.

Question 12. More generally each elementary stutter-invariant language can be defined by a $\Sigma_1[\leqslant]$ formula. Indeed, let $L = A^*a_1A^*a_2 \cdots A^*a_kA^*$ where $a_i \neq a_{i+1}$, for $1 \leqslant i \leqslant k-1$. Then $L = L(\varphi)$, where $\varphi = \exists x_1 \ \cdots \exists x_k \ (x_1 \leqslant x_2) \wedge (x_2 \leqslant x_3) \wedge \cdots (x_{k-1} \leqslant x_k) \wedge \mathbf{a_1} x_1 \wedge \cdots \wedge \mathbf{a_k} a_k$. The trick is that since $a_i \neq a_{i+1}$ for $1 \leqslant i \leqslant k-1$, all inequalities are strict.

Question 13. It follows that each language of $\mathcal{C}$ can be defined by a $\mathcal{B}\Sigma_1[\leqslant]$-formula. To prove the opposite direction, it suffices to show that each $\Sigma_1[\leqslant]$-formula φ defines a language of $\mathcal{C}$. Since φ is also in $\Sigma_1[<, =]$, it defines a piecsewise testable language. It remains only to show that $L(\varphi)$ is stutter-invariant.

Question 14. Let $(Q, A, \cdot)$ be the minimal automaton of L can be done in time $O(|Q||A|)$. It suffices to check, for each letter $a \in A$ and each state q, whether $q \cdot a = q \cdot a^2$