MPRI, Fondations mathématiques de la théorie des automates

Olivier Carton, Jean-Éric Pin

Examen du 3 février 2008. 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.

On rappelle qu'un langage L de A^* est *commutatif* si, pour tout $a, b \in A$ et $x, y \in A^*$, $xaby \in L$ entraîne $xbay \in L$, ce qui revient à dire que son monoïde syntactique est commutatif.

Le *mélange* de deux mots u et v est le langage $u \sqcup\!\sqcup\, v$ formé des mots $u_1v_1u_2v_2 \ \cdots \ u_kv_k$ où $k \geqslant 0$ et les u_i et les v_i sont des mots A^* tels que $u_1u_2 \cdots u_k = u$ et $v_1v_2 \cdots v_k = v$. Par exemple,

$$ab \sqcup\!\sqcup\, ba = \{abab, abba, baba, baab\}.$$

Par extension, le mélange de deux langages K et L est le langage

$$K \sqcup\!\sqcup\, L = \bigcup_{u \in K, v \in L} u \sqcup\!\sqcup\, v$$

On admettra sans démonstration que le mélange est une opération commutative et associative, distributive par rapport à l'union.

Mélange et langages commutatifs

Si M est un monoïde, on note $\mathcal{P}(M)$ le monoïde des parties de M, muni du produit suivant: si X et Y sont des parties de M, $XY = \{xy \mid x \in X \text{ et } y \in Y\}$.

Question 1. Soient $\eta_1 : A^* \to M_1$ et $\eta_2 : A^* \to M_2$ les morphismes syntactiques de deux langages L_1 and L_2. Soit $\mu : A^* \to \mathcal{P}(M_1 \times M_2)$ le morphisme défini, pour chaque $a \in A$, par $\mu(a) = \{(\eta_1(a), 1), (1, \eta_2(a))\}$. Montrer que μ reconnaît $L_1 \sqcup\!\sqcup\, L_2$.

Question 2. En déduire que si L_1 et L_2 sont reconnaissables, $L_1 \sqcup\!\sqcup\, L_2$ l'est également et que si L_1 et L_2 sont des langages commutatifs, $L_1 \sqcup\!\sqcup\, L_2$ l'est également.

On note $[u]$ la fermeture commutative d'un mot u. Par exemple,

$$[abab] = \{aabb, abab, abba, baab, baba, bbaa\}.$$

Question 3. Soint $u \in A^*$ et $B \subseteq A$. Montrer que $B^* \sqcup\!\sqcup\, [u]$ est l'ensemble des mots v tels que $|v|_a \geqslant |u|_a$ si $a \in B$ et $|v|_a = |u|_a$ si $a \notin B$. En déduire que les langages de la forme $B^* \sqcup\!\sqcup\, [u]$ sont à la fois sans-étoile et commutatifs.[1]

Question 4. Montrer qu'un langage est à la fois sans-étoile et commutatif si et seulement si il est union finie de langages de la forme $B^* \sqcup\!\sqcup\, [u]$, où $u \in A^*$ et $B \subseteq A$.

[1] La proposition 2.8 du chapitre 9 donne une description de ces langages.

Question 5. Montrer que l'ensemble des langages sans-étoile et commutatifs de A^* forme la plus petite algèbre de Boole de langages de A^* fermée par les opérations $L \mapsto L \text{ Ш } a$, pour chaque lettre a.

Mélange et langages non commutatifs

On note $\mathcal{C}$ la plus petite algèbre de Boole de langages $\mathcal{L}$ de A^* telle que

(1) $\mathcal{L}$ contient tous les langages de la forme $\{ab\}$, où a et b sont deux lettres distinctes de A,

(2) $\mathcal{L}$ est fermée par les opérations $L \mapsto L \text{ Ш } a$, pour chaque lettre a de A.

La question 5 montre que $\mathcal{C}$ contient aussi tous les langages sans-étoile et commutatifs de A^*.

Question 6. Montrer que $\mathcal{C}$ contient les langages de la forme $\{abb\}$, où a et b sont deux lettres de A.

Question 7. Démontrer que $\mathcal{C}$ contient tous les langages de la forme $\{u\}$ où u est un mot. En déduire que $\mathcal{C}$ contient tous les langages finis.

Un langage L est dit *valable* s'il existe un langage sans étoile commutatif C tel que la différence symétrique $L \bigtriangleup C$ soit finie.

Question 8. Démontrer que $\mathcal{C}$ est l'ensemble des langages valables.

Question 9. Vérifier que les langages valables vérifient les trois équations $x^\omega = x^{\omega+1}$, $x^\omega y = y x^\omega$ et $x^\omega yz = x^\omega zy$, où x est un mot non vide de A^* et y et z sont des mots quelconques de A^*.

On peut démontrer que ces équations caractérisent les langages valables.

Mélange et produit

Dans cette partie, A désigne un alphabet contenant au moins une lettre, a une lettre de A et L_1 et L_2 deux langages rationnels de A^*.

Question 10. Démontrer que si L_1 et L_2 satisfont l'équation $a^{\omega+1} = a^\omega$, alors le langage $L_1 L_2$ satisfait la même équation.

Question 11. Démontrer que si L_1 et L_2 satisfont l'équation $a^{\omega+1} = a^\omega$, alors le langage $L_1 \text{ Ш } L_2$ satisfait la même équation.

Question 12. Donner un exemple de langage ne satisfaisant pas l'équation $a^{\omega+1} = a^\omega$.

Question 13. La plus petite algèbre de Boole de langages fermée par produit et par mélange est elle égale à l'ensemble de tous les langages rationnels?

MPRI, Fondations mathématiques de la théorie des automates

Olivier Carton, Jean-Éric Pin

February 3, 2009. Duration: 2h 30.

⋆ ⋆ ⋆

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

Recall that a language L of A^* is *commutative* if, for all $a, b \in A$ and $x, y \in A^*$, $xaby \in L$ implies $xbay \in L$, which amounts to saying that the syntactic monoid of L is commutative.

The *shuffle* of two words u and v is the language $u ш v$ consisting of all words $u_1v_1 \cdots u_2v_2 \cdots u_kv_k$ where $k \geqslant 0$ and the u_i and the v_i are words of A^* such that $u_1u_2 \cdots u_k = u$ and $v_1v_2 \cdots v_k = v$. For instance,

$$ab ш ba = \{abab, abba, baba, baab\}$$

By extension, the shuffle of two languages K and L is the language

$$K ш L = \bigcup_{u \in K, v \in L} u ш v$$

It is known that the shuffle is a commutative and associative operation, which is also distributive over union.

Shuffle and commutative languages

Given a monoid M, $\mathcal{P}(M)$ denotes the monoid of subsets of M, equipped with the following product: if X and Y are subsets of M, $XY = \{xy \mid x \in X \text{ et } y \in Y\}$.

Question 1. Let $\eta_1 : A^* \to M_1$ and $\eta_2 : A^* \to M_2$ be the syntactic morphisms of the languages L_1 and L_2. Let $\mu : A^* \to \mathcal{P}(M_1 \times M_2)$ be the morphism defined, for each letter $a \in A$, by $\mu(a) = \{(\eta_1(a), 1), (1, \eta_2(a))\}$. Show that μ recognizes $L_1 ш L_2$.

Question 2. Deduce from the previous question that if L_1 and L_2 are regular, $L_1 ш L_2$ is also regular and that if L_1 and L_2 are commutative languages, $L_1 ш L_2$ is also a commutative language.

Let us denote by $[u]$ the commutative closure of a word u. For instance,

$$[abab] = \{aabb, abab, abba, baab, baba, bbaa\}.$$

Question 3. Let $u \in A^*$ and $B \subseteq A$. Show that $B^* ш [u]$ is the set of all words v such that $|v|_a \geqslant |u|_a$ if $a \in B$ and $|v|_a = |u|_a$ if $a \notin B$. Deduce that the languages of the form $B^* ш [u]$ are star-free and commutative.[2]

Question 4. Show that a language is star-free and commutative if and only if it is a finite union of languages of the form $B^* ш [u]$, where $u \in A^*$ and $B \subseteq A$.

[2] Proposition 2.8 in Chapter 9 gives a description of these languages.

Question 5. Show that the set of star-free and commutative languages of A^* forms the least [i.e. smallest] Boolean algebra of languages of A^* closed under the operations $L \mapsto L \sqcup\!\sqcup\, a$, for each letter a.

Shuffle and noncommutative languages

We denote by $\mathcal{C}$ the least Boolean algebra $\mathcal{L}$ of languages of A^* such that

(1) $\mathcal{L}$ contains all the languages of the form $\{ab\}$, where a and b are two distinct letters of A,

(2) $\mathcal{L}$ is closed under the operations $L \mapsto L \sqcup\!\sqcup\, a$, for each lettre a of A.

Question 5 shows that $\mathcal{C}$ also contains the commutative star-free languages of A^*.

Question 6. Show that $\mathcal{C}$ contains the languages of the form $\{abb\}$, where a and b are two letters of A.

Question 7. Prove that $\mathcal{C}$ contains all languages of the form $\{u\}$ where u is a word. Deduce from this fact that $\mathcal{C}$ contains all finite languages.

A language L is said to be *good* if there exists a commutative star-free language C such that the symmetric difference $L \triangle C$ is finite.

Question 8. Show that $\mathcal{C}$ is the set of good languages.

Question 9. Show that every good languages satisfies the three equations $x^\omega = x^{\omega+1}$, $x^\omega y = yx^\omega$ and $x^\omega yz = x^\omega zy$, where x is a nonempty word of A^* and y and z are arbitrary words of A^*.

One can show that these equations characterise good languages.

Shuffle and product

In this section, A denotes an alphabet containing at least one letter, a denotes a letter of A and L_1 and L_2 are two regular languages of A^*.

Question 10. Show that if L_1 and L_2 satisfy the equation $a^{\omega+1} = a^\omega$, then the language L_1L_2 satisfies the same equation.

Question 11. Show that if L_1 and L_2 satisfy the equation $a^{\omega+1} = a^\omega$, then the language $L_1 \sqcup\!\sqcup\, L_2$ satisfies the same equation.

Question 12. Give an example of a language which does not satisfy the equation $a^{\omega+1} = a^\omega$.

Question 13. Does every regular language belong to the least Boolean algebra of languages closed under product and shuffle?

Solution

Shuffle and commutative languages

Question 1. For each word $u \in A^*$, one has $\mu(u) = \{(\eta_1(u_1), \eta_1(u_2)) \mid u \in u_1 ⧢ u_2\}$. Suppose that $u \in L_1 ⧢ L_2$ and that $\mu(v) = \mu(u)$. Then there exist two words $u_1 \in L_1$ and $u_2 \in L_2$ such that $u \in u_1 ⧢ u_2$, and there exist two words $v_1, v_2 \in A^*$ such that $v \in v_1 ⧢ v_2$, $\eta_1(u_1) = \eta_1(v_1)$ and $\eta_2(u_2) = \eta_2(v_2)$. It follows that $v_1 \in L_1$ and $v_2 \in L_2$ and $v \in L_1 ⧢ L_2$. thus μ recognizes $L_1 ⧢ L_2$.

Question 2. In particular, if L_1 and L_2 are regular, M_1 and M_2 are finite and thus $\mathcal{P}(M_1 \times M_2)$ is also finite. Therefose $L_1 ⧢ L_2$ is regular. If L_1 and L_2 are commutative, then M_1 and M_2 are commutative and $\mathcal{P}(M_1 \times M_2)$ is also commutative. Consequently, $L_1 ⧢ L_2$ is commutative.

Question 3. Since B^* and $[u]$ are commutative languages, the language $B^* ⧢ [u]$ is commutative. Further, $B^* ⧢ [u]$ is a finite union of languages of the form $B^* ⧢ v$ with $v \in [u]$. If $v = a_1 \cdots a_n$, $B^* ⧢ v = B^*a_1B^*a_2 \ \cdots B^*a_nB^*$. Since B^* is star-free, $B^* ⧢ v$ is also star-free. It follows that every finite union of languages of the form $[u] ⧢ B^*$ is commutative and star-free.

Question 4. Let L be a commutative and star-free language. Let $\varphi : A^* \to M$ be its syntactic morphism, let $P = \varphi(L)$ and let N be the exponent of M. Since $L = \bigcup_{m \in P} \varphi^{-1}(m)$, it suffices to prove the result for $L = \varphi^{-1}(m)$, for some $m \in M$. We claim that $L = \bigcup_{u \in F}[u] ⧢ B^*$, where

$$\begin{aligned}
&B = \{a \in A \mid m\varphi(a) = m\} \text{ and}\\
&F = \{u \in A^* \mid |u| \leqslant N|A|,\ \varphi(u) = m \text{ and for all subwords } v \text{ of } u,\ \varphi(v) \neq m\}.
\end{aligned}$$

If $u \in F$ and $w \in [u] ⧢ B^*$, then $w \in u' ⧢ v$ for some $u' \in [u]$ and some $v \in B^*$. Since M is commutative, it follows that $\varphi(w) = \varphi(u)\varphi(v) = m\varphi(v) = m$. Thus $u \in L$. Conversely, let $w \in L$ and let u be a minimal subword of w in L. By construction, $\varphi(u) = m$ and for all subwords v of u, $\varphi(v) \neq m$. Further, if $|u| \geqslant N|A|$, then $|u|_a > N$ for some letter $a \in A$. Therefore, u can be written as u_1au_2 for some words u_1, u_2 such that $|u_1u_2|_a \geqslant N$. Since M is commutative and $\varphi(a^N) = \varphi(a^{N+1})$, it follows that $\varphi(u_1u_2) = \varphi(u)$, a contradiction with the definition of u. Thus $|u| \leqslant N|A|$ and $u \in F$.

Let v be the unique word such that $w \in u ⧢ v$. Since M is commutative, $\varphi(w) = \varphi(u)\varphi(v)$, that is $m = m\varphi(v)$. Since M is aperiodic and commutative, it is $\mathcal{J}$-trivial and thus $m\varphi(a) = m$ for each letter a of v. In other words, $v \in B^*$ and $w \in [u] ⧢ B^*$.

Question 5. Let $\mathcal{F}$ be the least Boolean algebra of languages of A^* closed under the operations $L \mapsto L ⧢ a$, for each letter a.

Let $a \in A$. Then $A^* \in \mathcal{F}$ and thus $A^* ⧢ a = A^*aA^* \in \mathcal{F}$. Since $\mathcal{F}$ is a Boolean algebra it contains all the languages of the form B^* (see Proposition 2.5, Chapter 9). Further, if $u = a_1 \cdots a_n$, then $[u] = a_1 ⧢ \ \cdots \ a_n$. It follows by induction that $B^* ⧢ [u] \in \mathcal{F}$. Thus, by Question 5, $\mathcal{F}$ contains all star-free and commutative languages.

The same argument shows that the star-free and commutative languages forms a Boolean algebra closed the operations $L \mapsto L ⧢ a$, for each letter a.

Shuffle and noncommutative languages

Question 6. If $a = b$, $\{abb\}$ is a commutative finite (and hence star-free) language. If $a \neq b$, one has $abb = ((ab \text{ Ш } b) \cap (bb \text{ Ш } a)) \setminus (ba \text{ Ш } b)$. Now, $\{bb\}$ is a commutaive language and hence belongs to $\mathcal{C}$. The languages $\{ab\}$ and $\{ba\}$ are also in $\mathcal{L}$ by definition. The result follows, since $\mathcal{C}$ is closed under Boolean operations and shuffle by a letter.

Question 7. It suffices to show that $\mathcal{C}$ contains the languages of the form $\{u\}$, for each word u.

Let $n = |u| - 1$ and $E = \{(v, a) \in A^n \times A \mid u \in v \text{ Ш } a\}$. The result will follow from the formula

$$(*) \qquad \{u\} = \Big(\bigcap_{(v,a) \in E} v \text{ Ш } a \Big) \setminus \Big(\bigcup_{(v,a) \in (A^n \times A) \setminus E} v \text{ Ш } a \Big)$$

Let L be the right hand side of $(*)$. It is clear that $u \in L$. Suppose that L contains another word w. Then $|w| = |u|$ and, for every $(v, a) \in E$, $u \in v \text{ Ш } a$ if and only if $w \in v \text{ Ш } a$. Let f be the largest common prefix of u and w. Assuming $u \neq w$, one can write $u = fau'$ and $w = fbw'$, for some $u', w' \in A^*$, $a, b \in A$ and $a \neq b$. We claim that f is the empty word. Otherwise, let c be a letter of f and let $f = f_1cf_2$. Let us assume that $c \neq a$ (the case $c \neq b$ would be symmetric). Then $u \in f_1f_2au' \text{ Ш } c$ and thus $w = f_1cf_2bw' \in f_1f_2au' \text{ Ш } c$. This means that c has to be inserted in the word f_1f_2au' to produce f_1cf_2bw'. Since $a \neq b$, this insertion cannot occur inside the prefix f_1f_2a. Therefore $f_1f_2a = f_1cf_2$, a contradiction, since $|f_1f_2a|_a > |f_1cf_2|_a$.

Thus the largest common prefix of u and w is the empty word, and by a symmetric argument, their largest common suffix is also the empty word. Let c be the first letter of u'. Then $u' = cx$ for some word $x \in A^*$. It follows that $u \in ax \text{ Ш } c$ and thus $w \in ax \text{ Ш } c$. Since the first letter of w is b, it means that $c = b$ and $w = bax$. It follows that x is a common suffix of u and w and thus x is the empty word. Therefore $u = ab$ and $w = ba$, a contradiction, since $|u| \geqslant 3$.

Question 8. Show that $\mathcal{C}$ is the set of good languages.

Question 9. Show that every good languages satisfies the three equations $x^\omega = x^{\omega+1}$, $x^\omega y = yx^\omega$ and $x^\omega yz = x^\omega zy$, where x is a nonempty word of A^* and y and z are arbitrary words of A^*.

One can show that these equations characterise good languages.

Shuffle and product

Question 10. Let L_1 and L_2 be languages of A^* satisfying the identity $a^{\omega+1} = a^\omega$ and let L be their product. Let n be the lcm of the exponents of the languages L_1, L_2 and L. It suffices to prove that $a^{n+1} \sim_L a^n$. Suppose that $xa^ny \in L$. Since $a^n \sim_L a^{2n}$, one has $xa^{2n}y \in L$ and thus $xa^{2n}y = u_1u_2$ for some $u_1 \in L_1$ and $u_2 \in L_2$. It follows that one of the words u_1 or u_2 contains a^n as a factor. Since the two cases are symmetrical, we may assume that $u_1 = xa^nz$ for some $z \in A^*$. It follows that $xa^{n+1}z \in L_1$, since L_1 satisfies the identity $a^{\omega+1} = a^\omega$. Thus $xa^{2n+1}y \in L$ and finally $xa^{n+1}y \in L$ since $a^{2n} \sim_L a^n$. Therefore L satisfies the equation $a^{\omega+1} \leqslant a^\omega$. The opposite direction is similar.

Question 11. Let L_1 and L_2 be languages of A^* satisfying the equation $a^{\omega+1} = a^\omega$ and let $L = L_1 \text{ Ш } L_2$. Let n be the lcm of the exponents of the languages L_1, L_2 and L.

Suppose that $xa^ny \in L$. Since $a^n \sim_L a^{2n}$, one has $xa^{2n}y \in L$ and thus $xa^{2n}y \in u_1 \text{ Ш } u_2$ for some $u_1 \in L_1$ and $u_2 \in L_2$. It follows that one of the words u_1 or u_2 contains a^n as a factor. If, for instance $u_1 = xa^nz$ for some $z \in A^*$, then $xa^{n+1}z \in L_1$ since L_1 satisfies the identity $a^{\omega+1} = a^\omega$. It follows that $xa^{2n+1}y \in L$ and finally $xa^{n+1}y \in L$ since $a^{2n} \sim_L a^n$. Thus L satisfies the identity $a^{\omega+1} \leqslant a^\omega$. The opposite direction is similar.

Question 12. The language $(aa)^*$ does not satisfy the equation $a^{\omega+1} = a^\omega$.

Question 13. Every language of the least Boolean algebra of languages closed under product and shuffle satisfies the equation $a^{\omega+1} = a^\omega$. Therefore, $(aa)^*$ does not belong to this Boolean algebra.