Automates sur les mots infinis

Arthur Leclaire

Résumé

Etant donné le travail effectué sur les langages rationnels, on peut construire, à l'aide d'une nouvelle opération, dite d'itération infinie, une classe plus grande de langages : les langages ω-rationnels, qui à la différence des précédents, peuvent contenir des mots infinis. Après avoir fixé des définitions, on verra que certains résultats s'étendent naturellement à ces nouveaux langages comme le théorème de Kleene, alors que d'autres, comme l'algorithme de la déterminisation, deviennent plus problématiques.

1 Langages ω-rationnels

Dans cette partie, A désigne un alphabet fini.

On appelle mot infini sur A toute suite infinie $(a_n)_{n\in\mathbb{N}}$ d'éléments de A, que l'on notera simplement par juxtaposition $a_0a_1a_2\dots a_n\dots$. On note A^ω l'ensemble des mots infinis sur A. Rappelons aussi qu'on note A^* l'ensemble des mots finis sur A et $A^+ = A^* \setminus \{\varepsilon\}$. On note alors A^∞ l'ensemble $A^* \cup A^\omega$ des mots (finis ou infinis) sur A. On appelle langage sur A tout sous-ensemble de A^∞.

L'opération de concaténation s'étend alors de la manière suivante : si $u = a_0a_1\dots a_n$ est un mot fini sur A et $v = b_0b_1\dots$ est un mot infini, le produit uv est le mot infini $a_0a_1\dots a_nb_0b_1\dots$. On se gardera bien de concaténer deux mots infinis. On définit aussi le produit d'une infinité de mots non vides : si pour tout entier n, $u_n = a_{n,0}a_{n,1}\dots a_{n,l_n}$ est un mot fini non vide, le produit des mots u_n est le mot infini

$$u_0u_1\dots u_n\dots = a_{0,0}a_{0,1}\dots a_{0,l_0}a_{1,0}a_{1,1}\dots a_{1,l_1}\dots a_{n,0}a_{n,1}\dots a_{n,l_n}\dots$$

Maintenant, si L_1 et L_2 sont deux langages sur A et si L_1 ne contient que des mots finis, on peut définir leur produit

$$L_1L_2 = \{uv \mid u \in L_1 \, et \, v \in L_2\}$$

La notion de périodicité s'étend elle aussi de manière tout à fait intuitive : le mot infini $u = a_0a_1\dots$ est dit périodique s'il existe entier p tel que pour tout entier n, on ait $a_{n+p} = a_n$, et cet entier p est alors appelé une période de u ; ce mot est dit périodique à partir du rang k si le mot $a_ka_{k+1}\dots$ est périodique ; il est dit ultimement périodique s'il est périodique à partir d'un certain rang.

Il nous reste à définir l'opération d'itération infinie : pour tout langage $L \subseteq A^*$ différent de $\varnothing$ et de $\{\varepsilon\}$, on définit le langage

$$L^\omega = \{u_0 u_1 \dots \,|\, \forall n \in \mathbb{N},\, u_n \in L\backslash\{\varepsilon\}\}$$

Par convention, on pose aussi $\varnothing^\omega = \varnothing$ et $\{\varepsilon\}^\omega = \{\varepsilon\}$. Si maintenant $L = \{u\}$ pour un mot fini $u = a_0 a_1 \dots a_n$, L^ω est composé de l'unique mot $uuu \dots = a_0 a_1 \dots a_n a_0 a_1 \dots a_n a_0 a_1 \dots$ que l'on notera u^ω (et qui admet donc pour période $n+1$).

On peut maintenant introduire les langages ω-rationnels sur A :

Définition 1.1. La classe des langages ω-rationnels sur A est le plus petit ensemble $\mathcal{R}at(A^\omega)$ de langages sur A tel que

1. $\varnothing \in \mathcal{R}at(A^\omega)$ et pour tout lettre $a \in A$, $\{a\} \in \mathcal{R}at(A^\omega)$
2. $\mathcal{R}at(A^\omega)$ est stable par union finie
3. Pour tout $L_1 \subseteq A^*$ et tout $L_2 \subseteq A^\infty$, $L_1 \in \mathcal{R}at(A^\omega)$ et $L_2 \in \mathcal{R}at(A^\omega)$ impliquent $L_1 L_2 \in \mathcal{R}at(A^\omega)$
4. Pour tout $L \subseteq A^*$, $L \in \mathcal{R}at(A^\omega)$ implique $L^* \in \mathcal{R}at(A^\omega)$ et $L^\omega \in \mathcal{R}at(A^\omega)$

On voit donc immédiatement que les langages rationnels sont ω-rationnels.

Exemple 1.1. Le langage L des mots infinis sur l'alphabet $\{a, b\}$ n'ayant qu'un nombre fini d'occurences de b est ω-rationnel car il s'écrit

$$L = (a+b)^* a^\omega$$

Son complémentaire dans A^ω qui est l'ensemble des mots infinis sur l'alphabet $\{a, b\}$ ayant un nombre infini d'occurences de b est aussi ω-rationnel car il s'écrit

$$A^\omega \backslash L = (a^* b)^\omega$$

On établit maintenant un théorème donnant une forme générale pour les langages ω-rationnels.

Théorème 1.1. *Un langage est ω-rationnel si et seulement s'il s'écrit comme union finie de langages de la forme XY^ω où X et Y sont des langages rationnels sur A.*

Démonstration. Notons $\mathcal{R}$ l'ensemble des langages sur A s'écrivant comme union finie de langages de la forme XY^ω avec X, Y langages rationnels.

D'abord, on a vu que $\mathcal{R}at(A^\omega)$ contient tous les langages rationnels. Comme $\mathcal{R}at(A^\omega)$ est stable par produit et par itération infinie, on obtient que si X et Y sont deux langages rationnels, $XY^\omega \in \mathcal{R}at(A^\omega)$. Puisqu'il est aussi stable par union finie, il vient que $\mathcal{R} \subseteq \mathcal{R}at(A^\omega)$.

Réciproquement, pour tout langage $X \in \mathcal{R}at(A^\omega)$, on a

$$X = (X \cap A^*) \cup (X \cap A^\omega)$$

et il suffit donc de montrer que

a) $X \cap A^*$ est un langage rationnel
b) $X \cap A^\omega \in \mathcal{R}$

pour obtenir que $X \in \mathcal{R}$. Notons alors $\mathcal{E}$ la classe des langages X vérifiant a) et b). Montrons que $\mathcal{R}at(A^\omega) \subseteq \mathcal{E}$. Il est clair que $\varnothing \in \mathcal{E}$ et que pour toute lettre a, $\{a\} \in \mathcal{E}$. De plus, si $X \in \mathcal{E}$ et $Y \in \mathcal{E}$, le langage

$$(X \cup Y) \cap A^* = (X \cap A^*) \cup (Y \cap A^*)$$

est rationnel et puisque l'on a aussi

$$(X \cup Y) \cap A^\omega = (X \cap A^\omega) \cup (Y \cap A^\omega)$$

$X \cup Y$ vérifie a) et b), et donc $\mathcal{E}$ est stable par union finie. Par ailleurs, si $X \subseteq A^*$ et $Y \subseteq A^\infty$ sont dans $\mathcal{E}$, le langage $X = X \cap A^*$ est rationnel. Par suite, le langage $(XY) \cap A^* = X(Y \cap A^*)$ est rationnel ; et puisque $Y \cap A^\omega \in \mathcal{R}$, en utilisant la distributivité à droite du produit par rapport à l'union, il vient que $(XY) \cap A^\omega = X(Y \cap A^\omega)$ appartient à $\mathcal{R}$. Enfin, si $X \subseteq A^*$ est dans $\mathcal{E}$, on a vu que X est rationnel. Comme $X^* \cap A^* = X^*$ et $X \cap A^\omega = \varnothing$, on obtient que $X^* \in \mathcal{E}$; et en écartant les cas simples $X = \varnothing$ et $X = \{\varepsilon\}$, on a $X^\omega \cap A^* = \varnothing$ et $X^\omega \cap A^\omega = X^\omega$ d'où $X^\omega \in \mathcal{E}$. Ainsi, avec la définition des langages ω-rationnels, on obtient par minimalité de $\mathcal{R}at(A^\omega)$ que $\mathcal{R}at(A^\omega) \subseteq \mathcal{E}$ ce qui achève la preuve. $\square$

2 Automates de Büchi

La définition formelle des automates de Büchi est la même que celle des automates finis reconnaissant les langages rationnels, que l'on désignera ici par automates traditionnels : un automate de Büchi est un quintuplet $\mathcal{A} = (Q, A, E, I, F)$ où Q est un ensemble fini d'états, A un alphabet, E une partie de $Q \times A \times Q$ donnant les transitions, I une partie de Q correspondant aux états initiaux, et F une partie de Q correspondant aux états finaux. On ne reviendra pas sur les autres définitions usuelles (complétude, déterminisme, émondage ...) qui ne diffèrent en rien des automates traditionnels. La notion de chemin infini est l'extension naturelle de la notion de chemin fini : c'est une suite infinie de transitions consécutives, c'est à dire une suite d'éléments de E de la forme $((q_n, a_n, q_{n+1}))_{n \in \mathbb{N}}$. L'étiquette de ce chemin est alors le mot infini $a_0 a_1 \ldots$. Cela dit, il nous est maintenant indispensable de décider d'un mode d'acceptation pour les mots infinis :

Définition 2.1. Soit $\mathcal{A} = (Q, A, E, I, F)$ un automate de Büchi.

Un chemin infini $((q_n, a_n, q_{n+1}))_{n \in \mathbb{N}}$ dans $\mathcal{A}$ est dit acceptant si $q_0 \in I$ et s'il existe une infinité d'entiers n tels que $q_n \in F$, c'est à dire que l'état d'où il part est un état initial de $\mathcal{A}$, et qu'il "visite" l'ensemble des états finaux une infinité de fois.

Un mot infini u sur A est alors dit accepté par $\mathcal{A}$ s'il existe dans $\mathcal{A}$ un chemin acceptant d'étiquette u.

On notera $L^\omega(\mathcal{A})$ l'ensemble des mots infinis acceptés par l'automate de Büchi $\mathcal{A}$.

Enfin, un langage $L \subseteq A^\omega$ est dit reconnaissable s'il existe un automate de Büchi $\mathcal{A}$ tel que $L = L^\omega(\mathcal{A})$.

Avant de donner des exemples, notons qu'un chemin $\mathcal{C}$ dans un automate de Büchi $\mathcal{A} = (Q, A, E, I, F)$ visite F une infinité de fois si et seulement s'il existe un état final f tel que $\mathcal{C}$ passe une infinité de fois par f : on peut appliquer le lemme des tiroirs étant donné qu'on s'est placé dans le cadre où l'ensemble des états est fini.

Exemple 2.1. Le langage accepté par l'automate de Büchi de la figure 1 est l'ensemble $a(a^*b)^\omega$ des mots commençant par a et contenant une infinité d'occurences de b. Cet automate peut être associé à la fonction "prédire la prochaine lettre lue". Il n'est pas déterministe.

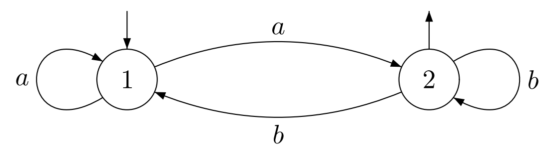

FIG. 1 – Automate de Büchi reconnaissant $a(a^*b)^\omega$

Exemple 2.2. Le langage accepté par l'automate de Büchi de la figure 2 est l'ensemble $\{a,b\}^*a^\omega$ des mots n'ayant qu'un nombre fini d'occurences de b. Notons que le langage rationnel accepté par l'automate traditionnel associé est l'ensemble $\{a,b\}^*$ de tous les mots finis sur $\{a,b\}$.

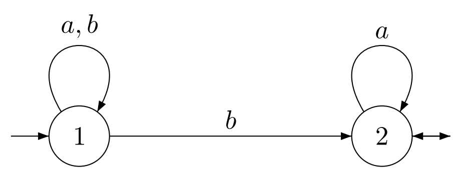

FIG. 2 – Automate de Büchi reconnaissant $\{a,b\}^*a^\omega$

On peut d'ores et déjà énoncer une proposition donnant une contrainte pour qu'un langage soit reconnaissable par un automate de Büchi.

Proposition 2.1. *Tout langage non vide reconnaissable par un automate de Büchi contient un mot ultimement périodique.*

Démonstration. Soit L un langage non vide reconnu par un automate de Büchi $\mathcal{A} = (Q, A, E, I, F)$. Comme L est non vide, il existe un mot $u \in L$ qui est l'étiquette d'un chemin $\mathcal{C}$ acceptant dans $\mathcal{A}$. On a vu qu'il existe un état $f \in F$ par lequel $\mathcal{C}$ passe une infinité de fois. Le mot u se décompose donc en u_0u_1w où u_0 est l'étiquette d'un chemin fini partant d'un état initial de $\mathcal{A}$ et arrivant en f, u_1 l'étiquette d'un chemin fini de longueur strictement positive d'extrémités toutes deux égales à f, et w un mot infini sur A. On voit alors que le mot infini $u_0u_1^\omega$. est accepté par $\mathcal{A}$ et il est bien sûr ultimement périodique. $\square$

De la même manière que pour les automates traditionnels, on peut associer à tout automate de Büchi un autre automate de Büchi reconnaissant le même langage et qui soit émondé, ou complet. Les algorithmes de complétion et d'émondage sont rigoureusement les mêmes et par conséquent préservent encore le déterminisme. Par contre on ne s'intéresse plus à l'idée de normalisation qui est incompatible avec le mode d'acceptation avec lequel on travaille.

Le prochain énoncé donne l'extension du théorème de Kleene pour les langages ω-rationnels.

Théorème 2.2. *Un langage $L \subseteq A^\omega$ est ω-rationnel si et seulement s'il est reconnaissable par un automate de Büchi.*

Démonstration. Dans un premier temps considérons un langage L reconnu par un automate de Büchi $\mathcal{A} = (Q, A, E, I, F)$ et montrons qu'il est ω-rationnel. Pour chaque paire d'états (q, q') de $\mathcal{A}$ on introduit l'automate traditionnel $\mathcal{A}_{q,q'} = (Q, A, E, \{q\}, \{q'\})$, le langage $L^*(\mathcal{A}_{q,q'})$ contenant les mots acceptés par $\mathcal{A}_{q,q'}$ et le langage $L^+(\mathcal{A}_{q,q'}) = L^*(\mathcal{A}_{q,q'}) \setminus \{\varepsilon\}$. Dans ces conditions, on a l'égalité

$$L = \bigcup_{(i,f) \in I \times F} L^*(\mathcal{A}_{i,f})(L^+(\mathcal{A}_{f,f}))^\omega$$

En effet, si L' désigne le membre de droite de l'égalité, l'inclusion $L' \subseteq L$ est évidente et l'inclusion $L \subseteq L'$ provient du fait que si un mot est accepté par $\mathcal{A}$, il est l'étiquette d'un chemin acceptant $\mathcal{C}$ qui passe donc une infinité de fois par un état final $f \in F$. Mais alors, par le théorème de Kleene pour les langages rationnels, pour chaque paire (i, f), les langages $L^*(\mathcal{A}_{i,f})$ et $L^+(\mathcal{A}_{f,f})$ sont rationnels. Comme $I \times F$ est fini, on conclut avec le théorème 1.1 que L est ω-rationnel.

Pour établir la réciproque, on voit par le théorème 1.1 qu'il suffit de montrer que pour tous langages rationnels X, Y, le langage XY^ω est reconnu par un automate de Büchi, et que la classe des langages reconnaissables par un automate de Büchi est stable par union finie. Soient donc X et Y des langages rationnels. Ecartons provisoirement le cas où le mot vide appartient à X. On sait qu'il existe deux automates traditionnels normalisés $\mathcal{A} = (Q, A, E, \{i\}, \{f\})$ et $\mathcal{A}' = (Q, A, E, \{i'\}, \{f'\})$ reconnaissant respectivement les langages $X = X \setminus \{\varepsilon\}$ et $Y \setminus \{\varepsilon\}$ et tels que $Q \cap Q' = \varnothing$. On construit alors un automate de Büchi reconnaissant XY^ω en fusionnant les états f, i' et f' comme indiqué à la figure 3.

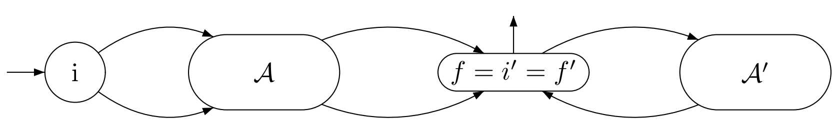

FIG. 3 – Automate de Büchi reconnaissant XY^ω

De manière formelle, on construit l'automate de Büchi

$$\mathcal{B} = ((Q \cup Q') \setminus \{i', f'\}, A, E \cup E_0 \cup E_1 \cup E_2, \{i\}, \{f\})$$

reconnaissant XY^ω en posant

$$E_0 = \{\, (f, a, f) \mid (i', a, f') \in E' \,\}$$

$$E_1 = \{\, (f, a, q) \mid q \in Q' \setminus \{i', f'\} \text{ et } (i', a, q) \in E' \,\}$$

$$E_2 = \{\, (q,a,f) \mid q \in Q' \backslash \{i,f\} \text{ et } (q,a,f') \in E' \,\}$$

Si maintenant le mot vide appartient à X, il suffit de déclarer également l'état f de $\mathcal{B}$ comme état final. Cela montre que XY^ω est reconnaissable par un automate de Büchi, pour tous langages rationnels X et Y.

Pour finir, soient X, Y deux langages reconnaissables par des automates de Büchi, respectivement $\mathcal{A} = (Q, A, E, I, F)$ et $\mathcal{A}' = (Q', A, E', I', F')$. Quitte à renommer les états on peut supposer Q et Q' disjoints. E et E' s'identifient alors à des parties de $(Q \cup Q') \times A \times (Q \cup Q')$. On obtient alors sans difficulté que $X \cup Y$ est reconnu par l'automate de Büchi $(Q \cup Q', A, E \cup E', I \cup I', F \cup F')$. Cela montre que la classe des langages reconnaissables par des automates de Büchi est stable par union finie. □

3 Déterminisation des automates de Büchi

On sait que par un procédé algorithmique, on peut sans problème déterminiser tout automate traditionnel. En ce qui concerne les automates de Büchi, la déterminisation est plus délicate voire même parfois impossible. On verra en effet qu'il existe des langages ω-rationnels non déterministes, c'est à dire qui ne peuvent être reconnus par un automate de Büchi déterministe.

L'étude du problème nécessite l'introduction d'un nouvel opérateur : si L est un langage sur un alphabet A ne contenant que des mots finis, on note $\overrightarrow{L}$ l'ensemble des mots infinis sur A qui ont une infinité de préfixes dans L.

Exemple 3.1. Donnons la valeur de ce nouvel opérateur en quelques langages particuliers :

1. Si $L = a^*b$, alors $\overrightarrow{L} = \varnothing$.
2. Si $L = (ab)^+$, alors $\overrightarrow{L} = (ab)^\omega$.
3. Si L est l'ensemble $(a^*b)^+$ des mots finis sur l'alphabet $\{a, b\}$ terminant par b, alors $\overrightarrow{L}$ est l'ensemble $(a^*b)^\omega$ des mots infinis sur $\{a, b\}$ contenant une infinité d'occurences de b.

Dans ces exemples, on voit qu'à partir d'un langage rationnel L, l'opérateur $\rightarrow$ nous a fourni un langage $\overrightarrow{L}$ ω-rationnel. Plus précisément, on montrera dans la suite qu'à tout langage rationnel $L \subseteq A^+$, on peut associer un automate de Büchi déterministe reconnaissant $\overrightarrow{L}$, et réciproquement.

Exemple 3.2. Le langage $L = (a + b)^*a^\omega$ des mots n'ayant qu'un nombre fini d'occurences de b n'est pas de la forme $\overrightarrow{L_0}$ pour $L_0 \subseteq A^+$. En effet, supposons que ce soit le cas. Le mot $u = ba^\omega$ a donc un préfixe $u_1 = ba^{n_1}$ dans L_0. Le mot $ba^{n_1}ba^\omega$ qui est dans L, a alors un préfixe $u_2 = ba^{n_1}ba^{n_2}$ dans L_0. On construit ainsi par récurrence une suite $(u_k)_{k \geq 1}$ de mots de L_0 et une suite $(n_k)_{k \geq 1}$ d'entiers tels que

$$\forall k \geq 1,\ u_k = ba^{n_1}ba^{n_2} \ldots ba^{n_k}$$

Le mot infini $w = ba^{n_1}ba^{n_2} \ldots$ a alors une infinité de préfixes dans L_0 (les u_k) et est donc dans L ce qui est absurde puisqu'il contient une infinité de b. En utilisant le résultat annoncé plus haut, le langage L ne peut être reconnu par un automate de Büchi déterministe. Cela prouve qu'il existe des langages ω-rationnels ne pouvant être reconnus par un automate de Büchi déterministe.

Proposition 3.1. *Soit $\mathcal{A} = (Q, A, E, \{i\}, F)$ un automate de Büchi déterministe. Alors $L^\omega(\mathcal{A}) = \overrightarrow{L^+(\mathcal{A})}$ (où $L^+(\mathcal{A})$ désigne le langage des mots non vides acceptés par l'automate traditionnel $\mathcal{A}$).*

Démonstration. Si $u \in L^\omega(\mathcal{A})$, u est l'étiquette d'un chemin $((q_n, a_n, q_{n+1}))_{n\in\mathbb{N}}$ tel que $q_0 = i$ et tel qu'il existe une infinité d'entiers n vérifiant $q_n \in F$. On peut donc construire une suite $(n_k)_{k\in\mathbb{N}}$ d'entiers naturels non nuls, strictement croissante, telle que pour tout entier k, $q_{n_k} \in F$. On voit alors que pour tout entier k, le mot $a_0a_1 \dots a_{n_k}$ est un préfixe de u et est accepté par l'automate traditionnel $\mathcal{A}$. Par conséquent, $u \in \overrightarrow{L^+(\mathcal{A})}$.
Réciproquement, si $u \in \overrightarrow{L^+(\mathcal{A})}$, u admet une infinité de préfixes dans $L^+(\mathcal{A})$. Comme $\mathcal{A}$ est déterministe, à chacun de ces préfixes, on peut associer un unique chemin fini dans $\mathcal{A}$ partant de i et arrivant dans un état final. De plus, toujours par déterminisme, si v et v' sont deux de ces préfixes tels que $|v| \leq |v'|$, alors v est un préfixe de v' et le chemin étiqueté par v dans $\mathcal{A}$ est le début du chemin étiqueté par v' dans $\mathcal{A}$. Par conséquent, le mot infini u est bien l'étiquette d'un chemin dans l'automate de Büchi $\mathcal{A}$ et celui-ci est acceptant car il passe dans F après la lecture de chacun des préfixes de u qui sont dans $L^+(\mathcal{A})$. □

On établit maintenant le résultat précédemment annoncé.

Théorème 3.2. *Soit un alphabet A et un langage $X \subseteq A^\omega$.
Le langage X est accepté par un automate de Büchi déterministe si et seulement s'il existe un langage rationnel $L \subseteq A^+$ tel que $X = \overrightarrow{L}$.*

Démonstration. Si X est accepté par un automate de Büchi déterministe $\mathcal{A}$, alors $L^+(\mathcal{A})$ est rationnel et d'après la proposition précédente, $X = \overrightarrow{L^+(\mathcal{A})}$. Réciproquement, s'il existe un langage rationnel $L \subseteq A^+$ tel que $X = \overrightarrow{L}$, on peut choisir un automate (traditionnel) déterministe $\mathcal{A}$ reconnaissant L. Toujours avec la proposition précédente, on obtient que $X = \overrightarrow{L} = \overrightarrow{L^+(\mathcal{A})} = L^\omega(\mathcal{A})$ et donc X est reconnu par l'automate de Büchi déterministe $\mathcal{A}$. □

Pour finir, signalons un résultat très important dans l'étude des langages ω-rationnels : pour tout alphabet A, $\mathcal{R}at(A^\omega)$ est clos par complémentation dans A^ω. Il est possible de le montrer en introduisant les automates de Rabin et de Muller, qui sont des automates sur les mots infinis ayant un mode d'acceptation plus puissant que les automates de Büchi (un théorème de McNaughton montre que tout langage ω-rationnel est reconnaissable pas un automate de Rabin).