Ensembles inévitables et classes de conjugaison

Jean-Marc Champarnaud   Georges Hansel

Résumé

Un ensemble X de mots sur un alphabet A est dit inévitable si tout mot infini sur A admet au moins un facteur dans X. Le cardinal d'un ensemble inévitable de mots de longueur k sur un alphabet A est supérieur ou égal au nombre de classes de conjugaison des mots de longueur k sur A. Nous donnons la construction d'un ensemble inévitable de mots de longueur k dont le cardinal est égal au nombre de classes de conjugaison.

Abstract

A set of words X is said to be unavoidable on a given alphabet A if every infinite word on A has a factor in X. The cardinality of an unavoidable set of words of length k on an alphabet A is not less than the number of conjugacy classes of words of length k on A. We show how to construct an unavoidable set of words of length k the cardinality of which is equal to the number of conjugacy classes.

1 Introduction

Un ensemble X de mots sur un alphabet A est dit inévitable si tout mot suffisamment long sur A admet au moins un facteur dans X [8].[1] Il est facile de voir qu'un ensemble inévitable de mots de longueur k sur un alphabet A doit contenir au moins un élément de chaque classe de conjugaison des mots de longueur k sur

[1] Une notion d'inévitabilité différente de celle-ci a été introduite dans [1].

1991 *Mathematics Subject Classification :* 68R15.

Key words and phrases : Mots de Lyndon, classes de conjugaison, ensembles inévitables.

A. Nous montrons que pour tout entier k, il existe un ensemble inévitable de mots de longueur k dont le cardinal est égal au nombre de classes de conjugaison et nous donnons la construction d'un tel ensemble. Ce résultat avait été conjecturé et vérifié jusqu'à l'ordre 7 par P. Higgins et C. Saker [5], et vérifié jusqu'à l'ordre 9 par G. Han et D. Perrin [4]. Une autre preuve, non constructive, a été également donnée par les auteurs en collaboration avec D. Perrin dans [2].

2 Notations, définitions, rappels

Soit A un alphabet fini muni de la relation d'ordre $<$. L'ensemble A^* des mots sur l'alphabet A est muni de l'ordre lexicographique induit par l'ordre sur A. La longueur d'un mot $u \in A^*$ est notée $|u|$. Soit un mot $u \in A^*$; un *conjugué* de u est un mot v qui se factorise en un produit $v = v_1 v_2$ avec $v_2 v_1 = u$. La conjugaison est une relation d'équivalence sur l'ensemble des mots. Un *mot de Lyndon* est un mot *minimal (pour l'ordre $<$) dans sa classe de conjugaison* et qui, de plus, est *primitif*, c'est-à-dire n'est la puissance d'aucun autre. Un mot minimal est, soit un mot de Lyndon s'il est primitif, soit une puissance d'un mot de Lyndon (uniquement déterminé). Dans ce qui suit, l'alphabet est l'ensemble à deux lettres $A = \{a, b\}$ mais les constructions se généralisent aisément à toute taille d'alphabet. Dans notre cas, tout mot minimal commence par a et se termine par b à l'exception des mots de la forme a^n et b^n. Pour les notions les plus courantes de combinatoire des mots, on pourra se reporter à l'un des ouvrages de M. Lothaire [6, 7].

Le résultat classique suivant est constamment utilisé dans la suite.

Proposition 1. *Soit w un mot primitif non vide. Les conditions suivantes sont équivalentes :*

a) w est un mot de Lyndon ;

b) le mot w est strictement inférieur pour l'ordre lexicographique à chacun de ses suffixes propres.

3 Ensembles inévitables

Soit A un alphabet fini. Un ensemble $I \subset A^*$ de mots sur l'alphabet A est *inévitable* si tout mot suffisamment long (ou infini, ou doublement infini) admet nécessairement pour facteur au moins un élément de I.

Exemple. Soit $A = \{a, b\}$. L'ensemble $U = \{a, b^{10}\}$ est inévitable car tout mot de longueur 10 admet nécessairement pour facteur, soit la lettre a, soit le mot b^{10}. En revanche, l'ensemble $V = \{aa, b^{10}\}$ n'est pas inévitable : le mot infini $ababababab\ldots$ n'admet aucun facteur dans V.

La proposition suivante est immédiate.

Proposition 2. *Soit A un alphabet fini et I un ensemble inévitable de mots ayant tous la même longueur k. Alors le cardinal de l'ensemble I est nécessairement supérieur ou égal au nombre de classes de conjugaison des mots de longueur k.*

Démonstration. Soit u un mot quelconque de longueur k. L'ensemble des facteurs de longueur k du mot infini $uuuuuu\ldots$ est l'ensemble des éléments de la classe de conjugaison de u. L'ensemble inévitable I doit donc contenir au moins un élément de cette classe. ∎

Exemple. Posons $u = abbab$. Un mot u^m, $m \geq 2$, admet pour facteurs de longueur cinq les éléments de l'ensemble $\{abbab, babba, ababb, babab, bbaba\}$; donc tout ensemble inévitable I contient au moins l'un de ces cinq mots.

Sur $A = \{a, b\}$, si k est un nombre premier, tous les mots de longueur k sont primitifs à l'exception des mots a^k et b^k. Toutes les classes de conjugaison des autres mots contiennent donc k éléments. Le nombre s_k de classes de conjugaison vérifie donc l'égalité

$$s_k = \frac{2^k - 2}{k} + 2$$

Plus généralement, on montre que le nombre $s_k(p)$ de classes d'équivalence des mots de longueur k sur un alphabet à p lettres est donné par la formule

$$s_k(p) = \frac{1}{k} \sum_{d|k} \phi(k/d) p^d$$

où $\phi(n)$ est le nombre des entiers positifs inférieurs à n et premiers avec n, le nombre 1 étant également compté (la *fonction d'Euler*).

Soit un ensemble de mots $X \subset A^*$; un mot u est *interdit* par X si tout mot infini qui n'admet aucun facteur dans X ne peut avoir le mot u comme facteur.

Exemple. Soient $A = \{a, b\}$ et $X = \{babb, aababa\}$. Le mot $aabab$ est interdit par X. En effet il ne peut se prolonger que par a ou b, et dans les deux cas apparaît un facteur appartenant à X.

Si un mot u est interdit, tous les mots de l'idéal A^*uA^* sont également interdits. Un ensemble est inévitable si tout mot est interdit, c'est-à-dire si l'idéal des mots interdits est A^+. Il est connu [7] que la détermination du générateur de l'idéal des mots interdits par un ensemble de mots X peut s'effectuer au moyen de l'algorithme suivant (que nous décrivons dans le cas $A = \{a, b\}$ mais qui se généralise simplement) :

1) Si X contient deux mots u et v avec u facteur de v, retirer le mot v.

2) Si X contient un mot de la forme xya et le mot yb, réduire xya en xy (on peut évidemment échanger les rôles de a et b).

2′) Si X contient un mot de la forme axy et le mot bx, réduire axy en xy.

Répéter les opérations précédentes jusqu'à ce qu'elles ne soient plus possibles.

L'ensemble X est inévitable si l'algorithme précédent réduit X au mot vide.

4 Un ensemble inévitable minimal

Dans cette section, nous donnons la construction d'un ensemble inévitable de mots de longueur k en indiquant, pour chaque mot minimal de longueur k, le mot choisi dans sa classe de conjugaison. On note M_k l'ensemble des mots minimaux de longueur k. Pour tout mot v de longueur inférieure ou égale à k, on note $M_k(v)$ l'ensemble des mots minimaux de longueur k dont v est préfixe. Nous allons commencer par définir une partition de M_k en ensembles $M_k(v)$.

Définitions. Soit $A = \{a, b\}$.

1) Soit $w = va \in A^*$ un mot se terminant par a ; le mot $w' = vb$ est appelé le *successeur* de w et est noté $S(w)$.

2) Soit $w = vab^i \in A^*$ un mot se terminant par une suite de b ; le mot $w' = va$ est appelé le *réduit* de w et est noté $R(w)$.

3) Soit $w \in A^*$ un mot de longueur inférieure ou égale à k ; le *périodisé* de w est la plus grande puissance entière w^i de w dont la longueur est inférieure ou égale à k et est noté $P(w)$.

Nous définissons par récurrence deux suites finies de mots (u_n) et (v_n) de la façon suivante :

$$u_0 = a^k \qquad v_0 = P(S(u_0)) = a^{k-1}b$$

Ayant défini u_{n-1} et v_{n-1}, on pose

$$u_n = R(v_{n-1}) \qquad v_n = P(S(u_n))$$

La construction s'arrête lorsque $v_n = b^k$. On note N_k l'ensemble des entiers n tels que u_n et v_n sont définis.

Exemple. Le tableau suivant donne les suites u_n et v_n pour $k = 7$.

aaaaaaa	aaaaaab
aaaaaa	aaaaab
aaaaa	aaaab
aaaa	aaab
aaa	aabaab
aabaa	aabab
aaba	aabb
aa	ababab
ababa	ababb
aba	abbabb
abba	abbb
a	bbbbbbb

Proposition 3. *Les ensembles $M_k(v_n)$ sont disjoints et on a*

$$M_k = \bigcup_n M_k(v_n) \cup \{a^k\}$$

Démonstration. 1) Par construction la suite de mots (v_n) est croissante et majorée par b^k. Plus précisément, soit $m < n$ et notons v'_m (respectivement v'_n) le préfixe de v_m (respectivement de v_n) de longueur $\inf(|v_m|, |v_n|)$; on montre facilement que $v'_m < v'_n$. Les ensembles $M_k(v_n)$ sont donc disjoints.

2) Soit v un mot minimal de longueur k quelconque et supposons que

$$v_m \leq v < v_{m+1} \tag{1}$$

Montrons que v admet v_m pour préfixe.

Le mot v_m se factorise sous la forme

$$v_m = a^{i_1}b^{j_1} \dots a^{i_p}b^{j_p} \tag{2}$$

Comme $v_{m+1} = P(S(R(v_m)))$, on a pour un certain entier positif n

$$v_{m+1} = (a^{i_1} b^{j_1} \dots a^{i_p - 1} b)^n \tag{3}$$

Il résulte déjà immédiatement de 1, 2 et 3 que v admet pour préfixe le mot

$$w = a^{i_1} b^{j_1} \dots a^{i_p - 1}$$

Montrons que v ne peut admettre wb pour préfixe. Si tel était le cas, comme $v < v_{m+1}$ et comme $|v| = k \geq |v_{m+1}|$, il existerait des mots x, y et z tels que

$$v = wbxyz \quad \text{avec} \quad y < a^{i_1} b^{j_1} \dots a^{i_p - 1} b = wb \tag{4}$$

Par conséquent v ne serait pas un mot minimal (puisqu'on aurait $yzwbx < wbxyz = v$).

Donc v admet wa pour préfixe et par conséquent, en vertu de 1 et 2, et comme $|v| > |v_m|$, v admet v_m pour préfixe. ∎

Définitions. La partition $(M_k(v_n))_{n \in N_k}$ est appelée la *partition standard de* $M_k - \{a^k\}$.

Pour tout n, posons

$$L_k(v_n) = \{v \in A^* | v_n v \in M_k(v_n)\}$$

On a donc $M_k(v_n) = v_n L_k(v_n)$. Posons

$$I_k = \{a^k\} \cup \bigcup_n L_k(v_n) v_n$$

Par construction, l'ensemble I_k contient un élément et un seul dans chaque classe de conjugaison des mots de longueur k. Le but de la section suivante est de montrer que I_k est un ensemble inévitable.

Exemple. Le tableau suivant donne pour $k = 7$ les ensembles $M = M_k$, V la suite des v_n et $I = I_k$.

M	V	I
aaaaaaa		aaaaaaa
aaaaaab	aaaaaab	aaaaaab
aaaaabb	aaaaab	baaaaab
aaaabab	aaaab	abaaaab
aaaabbb		bbaaaab
aaabaab	aaab	aabaaab
aaababb		abbaaab
aaabbab		babaaab
aaabbbb		bbbaaab
aabaabb	aabaab	baabaab
aababab	aabab	abaabab
aababbb		bbaabab
aabbabb	aabb	abbaabb
aabbbab		babaabb

aabbbbb		bbbaabb
abababb	ababab	bababab
ababbbb		bbababb
abbabbb	abbabb	babbabb
abbbbbb	abbb	bbbabbb
bbbbbbb	bbbbbbb	bbbbbbb

5 I_k est un ensemble inévitable

Nous utilisons dans la suite deux résultats [3] dont nous donnons la démonstration pour être complets.

Proposition 4. *Soient u et v deux mots de Lyndon avec $v > u$. Alors le produit uv est un mot de Lyndon.*

Démonstration. Le mot uv est primitif car, dans le cas contraire, au moins l'un des deux mots u et v admettrait un préfixe qui soit également suffixe, ce qui est impossible pour un mot de Lyndon. En vertu de la proposition 1, il suffit donc de montrer que pour tout suffixe propre s de uv, on a $s > uv$.

a) Supposons d'abord que $s = v$ et montrons que $v > uv$. Si $|u| \geq |v|$, l'inégalité $v > u$ entraîne $v > uv$ (propriété de l'ordre lexicographique). Si $|v| > |u|$, posons $v = u'v'$, où $|u'| = |u|$. L'inégalité $v > u$ implique que $u' \geq u$. De plus, comme v est un mot de Lyndon, on a $v' > v$ et donc

$$v = u'v' \geq uv' > uv$$

b) Soit s un suffixe propre de uv différent de v. Si $|s| < |v|$, comme v est un mot de Lyndon et en tenant compte de a), on a $s > v > uv$. Si, au contraire $|s| > |v|$, le mot s se factorise sous la forme $s = s'v$, où s' est un suffixe propre de u. Comme u est un mot de Lyndon, on a $s' > u$ et par conséquent $s = s'v > uv$.

Donc, dans tous les cas, on a $s > uv$ et par conséquent uv est un mot de Lyndon.■

Proposition 5. *Soit v un mot de Lyndon et soit $w = v^p$ une puissance de v. Si w admet une factorisation de la forme $w = uxu$, alors les mots u et x sont des puissances de v.*

Démonstration. Si u est une puissance de v, soit $u = v^m$, alors $x = v^{p-2m}$. Supposons donc que u n'est pas puissance de v. Remarquons que l'on a nécessairement $|u| > |v|$. Dans le cas contraire, v admettrait une factorisation $v = u'yu'$. Comme v est un mot de Lyndon, on aurait $u'u'y > u'yu'$, ce qui implique $u'y > yu'$, soit encore $v = u'yu' > yu'u'$, ce qui est absurde.

Il existe donc deux mots g et h tels que $v = gh$, g est suffixe propre de u, h préfixe de x ou x préfixe de h, et tels que $uxu = (gh)^p$, $u = (gh)^m g$, $xu = (hg)^n h$ (et $n + m + 1 = p$). Comme uxu est puissance d'un mot de Lyndon, c'est le mot minimal de sa classe de conjugaison. On a donc

$$\begin{aligned} uux &\geq uxu \\ xuu &\geq uxu \end{aligned}$$

Ces inégalités entraînent l'égalité $ux = xu$. On a donc

$$(gh)^m g(hg)^n h = uxu = xuu = (hg)^n h(gh)^m g$$

Soit encore

$$(gh)^p = (hg)^p$$

Il en résulte que $gh = hg$ ce qui est absurde puisque $gh = v$ est un mot de Lyndon. ■

Définition. Soit $n \in N_k$; on dit qu'un mot $x \in A^*$ *a la propriété* U_n si pour toute factorisation de x de la forme $x = x'bx''$, le mot $x'a$ admet un facteur dans l'ensemble $U_n = \{u_0, u_1, \ldots, u_n\}$ (c'est-à-dire $x'a = w_1ww_2$ où $w \in U_n$).

Proposition 6. *Pour tout $n \in N_k$, le mot v_n a la propriété U_n.*

Démonstration. On procède par récurrence. Le mot $v_0 = a^{k-1}b$ admet évidemment la propriété U_0 puisque $u_0 = a^k$. Supposons que la proposition soit vérifiée jusqu'à l'ordre n. Comme v_n a la propriété U_n, il en est de même de $u_{n+1} = R(v_n)$; donc le mot $S(R(v_n))$ a la propriété U_{n+1} et il en est encore de même de $v_{n+1} = P(S(R(v_n)))$. ■

Proposition 7. *Soit v un mot de Lyndon (autre que b). Alors $S(R(v))$ est un mot de Lyndon et $S(R(v)) > v$.*

Démonstration. Si v se factorise sous la forme $v = a^ib^j$, on a $S(R(v)) = a^{i-1}b$ et le résultat est vérifié. Sinon le mot v se factorise (de façon unique) sous la forme

$$v = wba^ib^j$$

et on a

$$S(R(v)) = wba^{i-1}b$$

Soit x un suffixe propre de $S(R(v))$ et montrons que $x > S(R(v))$.

Si $|x| \leq i$, on a

$$x \geq a^{i-1}b$$

Comme v est un mot de Lyndon, le mot wb admet un préfixe de la forme a^mb^n avec nécessairement $i \leq m$, ce qui entraîne $a^{i-1}b > a^mb^n$ et donc aussi $a^{i-1}b > S(R(v))$ et donc $x > S(R(v))$.

Supposons maintenant $|x| > i$. Le mot x se factorise en $x = yba^{i-1}b$, où yb est suffixe propre de wb. Soit z le plus long préfixe de yb qui est également préfixe de wb.

Si $|z| < |yb|$, comme v est un mot de Lyndon, wb admet pour préfixe za et yb admet pour préfixe zb (si wb admettait pour préfixe zb et yb admettait pour préfixe za, on aurait $yba^ib^j < wba^ib^j = v$, ce qui est impossible). Par conséquent $yb > wb$ et donc

$$x = yba^{i-1}b > wba^{i-1}b = S(R(v))$$

Supposons enfin $z = yb$. Alors v se factorise sous la forme $v = yby'ba^ib^j$ et on a $yba^ib^j = za^ib^j > v$. Il en résulte que $y' = a^my''$ où $m \geq i$. Donc

$$S(R(v)) = wba^{i-1}b = yba^my''ba^{i-1}b$$

D'où

$$x = yba^{i-1}b > S(R(v))$$

Par conséquent $S(R(v))$ est un mot de Lyndon. Enfin l'inégalité $S(R(v)) > v$ est évidente. ■

Proposition 8. *Les mots v_n, $n \in N_k$, sont des puissances de mots de Lyndon.*

Démonstration. Le mot $v_0 = a^{k-1}b$ est un mot de Lyndon. Supposons que la propriété soit vérifiée jusqu'au rang n et soit w le mot de Lyndon tel que $v_n = w^p$. On a

$$S(R(v_n)) = w^{p-1}S(R(w))$$

D'après la proposition 7, le mot $S(R(w))$ est un mot de Lyndon et $S(R(w)) > w$. Alors il résulte de la proposition 4 employée de manière récurrente que $S(R(v_n)) = w^{p-1}S(R(w))$ est un mot de Lyndon. Donc $v_{n+1} = P(S(R(v_n)))$ est une puissance de mots de Lyndon. ∎

Proposition 9. *Soit $n \in N_k$ et soit un mot w non vide vérifiant les inégalités*

$$\begin{aligned} |w| &\leq k - |v_n| \\ w &\leq v_n \end{aligned}$$

Alors le mot wv_n admet un facteur dans U_n.

Démonstration. Le mot v_n admet une factorisation de la forme

$$v_n = a^{i_1}b^{j_1} \dots a^{i_p}b^{j_p}$$

D'après la proposition 6, le mot a^{i_1+1} admet un facteur dans U_n. Par conséquent si le mot w est de la forme $w = w'a$, le mot wv_n admet a^{i_1+1} pour facteur et donc a un facteur dans U_n. On peut donc supposer que w est de la forme $w = w'b$. D'autre part l'inégalité $w \leq v_n$ implique que w est de la forme $w = aw''$. Par conséquent w se factorise sous la forme

$$w = a^{k_1}b^{l_1} \dots a^{k_q}b^{l_q}$$

1) Supposons tout d'abord que w n'est pas préfixe de v_n. Il existe alors un premier indice m tel que

$$a^{k_m}b^{l_m} < a^{i_m}b^{j_m}$$

et donc au moins l'une des deux conditions $k_m > i_m$ ou $l_m < j_m$ est remplie.

Supposons que $k_m > i_m$; il résulte de la proposition 6 que le mot $a^{i_1}b^{j_1} \dots a^{i_m+1}$ admet un facteur dans U_n et il en est donc de même de wv_n (et même ici de w).

Supposons que $k_m = i_m$ et $l_m < j_m$; il résulte de la proposition 6 que le mot $x = a^{i_1}b^{j_1} \dots a^{i_m}b^{l_m}a$ admet un facteur dans U_n et il en est donc de même de wv_n (soit x est préfixe de w, soit $w = a^{i_1}b^{j_1} \dots a^{i_m}b^{l_m}$).

2) Supposons maintenant que w est préfixe de v_n. Soit m le plus grand entier tel que le mot v_n admette une factorisation de la forme

$$v_n = w^m w'$$

Observons que w' ne peut être le mot vide. Sinon l'égalité

$$w^m = v_n = P(S(R(v_{n-1})))$$

et le théorème du défaut [6] impliqueraient que w est une puissance du mot de Lyndon $S(R(v_{n-1}))$ et cela contredirait la construction de v_n (puissance maximum de $S(R(v_{n-1}))$ de longueur inférieure ou égale à k).

Il est également impossible que w' soit un préfixe propre de w. En effet cela impliquerait que le mot v_n se factorise sous la forme $v_n = w'xw'$, ce qui, d'après la proposition 5, entraînerait que w' est une puissance de $S(R(v_{n-1}))$ ce qui contredit à nouveau la construction de v_n.

Comme v_n est un mot minimal dans sa classe et comme w' n'est pas puissance de $S(R(v_{n-1}))$, on a l'inégalité

$$w'w^m > w^m w'$$

Soit $l = \inf(|w|, |w'|)$ et soient x et x' les préfixes de longueur l de w et w' respectivement. Il résulte de ce qui précède que x et x' diffèrent et donc l'inégalité ci-dessus entraîne que $x' > x$. Soit y leur plus long préfixe commun. Les mots x et x' admettent alors des factorisations de la forme

$$\begin{aligned} x &= yaz \\ x' &= ybz' \end{aligned}$$

Le mot $w^m yb$ est préfixe de v_n et donc, d'après la proposition 6, le mot $w^m ya$ admet un facteur dans U_n. Mais le mot $w^m ya$ est préfixe de w^{m+1} et donc aussi de $wv_n = ww^m w'$. Donc wv_n admet un facteur dans U_n. ■

Proposition 10. *Soient $n \in N_k$ et $w \in A^*$ tel que $|v_n| + |w| = k$. Les conditions suivantes sont équivalentes :*

1) *Tout suffixe x non vide de w vérifie l'inégalité $x > v_n$;*

2) *Le mot $v_n w$ est minimal dans sa classe de conjugaison (c'est-à-dire appartient à l'ensemble $M_k(v_n)$).*

Démonstration. 1) $\Rightarrow$ 2) Soit x un suffixe propre non vide de $v_n w$.

Si $|x| \leq |w|$, on a par hypothèse $x > v_n$; comme par construction de v_n, on a $|x| \leq |w| < k/2 < |v_n|$, cela entraîne $x > v_n w$.

Supposons maintenant $|x| > |w|$. Soit $v_n = v^p$ la factorisation de v_n comme puissance du mot de Lyndon v (proposition 8). Le mot x se factorise sous la forme $x = x'v^i w$ où x' est suffixe du mot de Lyndon v. Trois cas sont possibles :

$p = 1$ et donc x' est suffixe propre du mot de Lyndon v_n. Par conséquent $x' > v_n$ et donc aussi $x' > v_n w$.

$p > 1$ et x' est suffixe propre du mot de Lyndon v. Par conséquent $x' > v$ et donc aussi $x' > v^p w = v_n w$.

$p > 1$ et x' est le mot vide. Le mot x se factorise sous la forme $x = v^q w$ avec $q < p$. Par construction de v_n, on a $|w| < |v|$. Comme par hypothèse $w > v_n$, on a aussi $w > v$ et donc $x = v^q w > v^p w = v_n w$.

Dans tous les cas, on a $x > v_n w$ et par conséquent $v_n w$ est minimal dans sa classe de conjugaison.

2) $\Rightarrow$ 1) Soit x un suffixe de w. Comme par construction de v_n, on a $|w| < |v_n|$, le mot $v_n w$ se factorise sous la forme $v_n w = x' v_n' w$, avec $|x'| = |x|$. Comme $v_n w$ est minimal dans sa classe, on a nécessairement $x' \leq x$. Montrons que l'on ne peut avoir $x = x'$.

Supposons donc que $x = x'$. Le mot $v_n w$ est une puissance s^p de mot de Lyndon et se factorise sous la forme $v_n w = xzx$. Il résulte de la proposition 5 que $x = s^q$

pour un certain entier q. Par ailleurs le mot v_n est également une puissance t^r de mot de Lyndon (Proposition 8). Il résulte de la construction de v_n que

$$|s| \leq |x| \leq |w| < |t|$$

Puisque t est un mot de Lyndon, il est primitif. Il existe donc un entier r tel que

$$|s^r| < |t| < |s^{r+1}|$$

Par suite, il existe un préfixe s' de s qui est en même temps suffixe de t. Mais s' est en même temps préfixe de v_n donc de t, ce qui est absurde. ■

Proposition 11. *Pour tout $n \in N_k$, les mots v_n et u_{n+1} sont interdits par l'ensemble $V_n = U_n \bigcup L_k(v_n)v_n$.*

Démonstration. Montrons d'abord que le mot v_n est interdit par V_n. Soit w un mot de longueur $k - |v_n|$. Si pour tout suffixe x de w, on a $x > v_n$, alors, en vertu de la proposition 10, le mot $v_n w \in M_k(v_n)$ et donc $wv_n \in L_k(v_n)v_n$. Si au contraire il existe un suffixe x de w tel que $x \leq v_n$, alors, en vertu de la proposition 9, le mot xv_n admet un facteur dans U_n et il en est à plus forte raison de même de wv_n. Donc dans tous les cas, le mot wv_n admet un facteur dans V_n et, par conséquent, le mot v_n est interdit par l'ensemble V_n.

Le mot v_n se factorise sous la forme $v_n = v'a^p b^q$. Le mot v_n et (en vertu de la proposition 6) le mot $v'a^p b^{q-1}a$ sont tous deux interdits par l'ensemble V_n. Donc le mot $v'a^p b^{q-1}$ est interdit par l'ensemble V_n. En raisonnant par récurrence, on obtient que le mot $u_{n+1} = R(v_n) = v'a^p$ est interdit par l'ensemble V_n. ■

Théorème 1. *Soient $A = \{a, b\}$ et M_k l'ensemble des mots de longueur k sur l'alphabet A qui sont minimaux dans leur classe de conjugaison. Soit $(v_n L_k(v_n))_{n \in N_k}$ la partition standard de $M_k - \{a^k\}$. Alors l'ensemble $I_k = \{a^k\} \cup \bigcup_n L_k(v_n)v_n$ est un ensemble inévitable.*

Démonstration. En raisonnant par récurrence, il résulte de la proposition 11 que, pour tout $n \in N_k$, le mot u_{n+1} est interdit par l'ensemble

$$W_n = \{a^k\} \cup \bigcup_{i=0}^{n} L_k(v_i)v_i$$

Par conséquent, l'ensemble I_k interdit les mots a et b^k. Il est donc inévitable. ■

6 Complément

Dans cette section nous indiquons un algorithme permettant de déterminer simplement le préfixe d'un mot minimal qui appartient à la suite $(v_n)_{n \in N_k}$ (rappelons que pour tout mot minimal w, il existe un et un seul élément de la suite (v_n) qui est préfixe de w).

Proposition 12. *Soit w un mot de longueur k qui est minimal dans sa classe. Soit x un préfixe de Lyndon de w de longueur inférieure ou égale à $k/2$.*

1) *Si y est un préfixe non vide de x tel que $P(y)$ est préfixe de w, alors $y = x$.*

2) *Aucun préfixe z de $P(x)$ de longueur supérieure à $|x|$ n'est un mot de Lyndon.*

Démonstration. 1) Supposons que y soit un préfixe propre de x. Il existe $p \geq 1$ tel que

$$|y|^p < |x| < |y|^{p+1}$$

Il existe donc un préfixe non vide y' de y tel que

$$x = y^p y'$$

Le mot y' est à la fois préfixe et suffixe de x, ce qui est absurde puisque x est un mot de Lyndon. Donc $y = x$.

2) Une démonstration semblable à celle de 1) montre que z admet un préfixe propre qui est en même temps suffixe. ■

Proposition 13. *Soit w un mot de longueur k qui est minimal dans sa classe de conjugaison. Soit v l'élément de la suite (v_n) qui est préfixe de w. Soit $(w_1, w_2, \ldots, w_p)$ la suite rangée par longueur croissante des préfixes de Lyndon de w et soit m le plus petit entier tel que $P(w_m)$ est préfixe de w. Alors $v = P(w_m)$.*

Démonstration. Soit $(n_1, n_2, \ldots, n_q)$ la suite des entiers tels que $|S(R(v_{n_i-1}))| \leq k/2$, de sorte que v_{n_i} n'est pas primitif.

Si le mot w admet un préfixe de la forme $a^n b$, avec $n \geq k/2$, alors, par construction de la suite (v_n), on a $v = a^n b = w_1$ et il n'y a rien à démontrer. Sinon il existe un entier i tel que

$$v_{n_i} \leq v < v_{n_{i+1}}$$

1) Supposons d'abord que $v = v_{n_i}$. Le mot v est de la forme

$$v = S(R(v_{n_i-1}))^p$$

avec $p > 1$. D'après la proposition 12-1, si x est un préfixe propre (en particulier de Lyndon) de $S(R(v_{n_i-1}))$, le mot $P(x)$ ne peut être préfixe de w. Donc $S(R(v_{n_i-1}))$ est le plus court préfixe w_m de Lyndon de w tel que $P(w_m)$ est préfixe de w.

2) Supposons maintenant que

$$v_{n_i} < v < v_{n_{i+1}}$$

Le mot v est alors un mot de Lyndon. Donc il existe un entier m tel que $v = w_m = P(w_m)$. Par construction de la suite (v_n), le mot v se factorise sous la forme $v = v'b$ où v' est le plus long préfixe de w commun à v et v_{n_i}.

Supposons qu'il existe $k < m$ tel que $P(w_k)$ soit préfixe de w. Le mot w_k est préfixe de v' donc de v_{n_i}. D'après la proposition 12-2, on ne peut avoir $|w_k| > |S(R(v_{n_i-1}))|$. On ne peut avoir $|w_k| = |S(R(v_{n_i-1}))|$ car $P(w_k) = v_{n_i}$ n'est pas préfixe de w. Enfin une démonstration semblable à celle de 12-1 montre que l'on ne peut avoir $|w_k| < |S(R(v_{n_i-1}))|$.

Donc le mot v lui-même est bien le plus court préfixe de Lyndon w_m de w tel que $P(w_m)$ est préfixe de w. ∎

Exemple. Le tableau suivant indique pour chaque mot minimal de longueur 8 l'élément de la suite v_n qui en est préfixe.

aaaaaaaa	aabaab'bb
aaaaaaab'	aabab'abb
aaaaaab'b	aabab'bab
aaaaab'ab	aabab'bbb
aaaaab'bb	aabbaabb'
aaaab'aab	aabbab'ab
aaaab'abb	aabbab'bb
aaaab'bab	aabbb'abb
aaaab'bbb	aabbb'bab
aaabaaab'	aabbb'bbb
aaabaab'b	abababab'
aaabab'ab	abababb'b
aaabab'bb	ababb'abb
aaabb'aab	ababb'bbb
aaabb'abb	abbab'bbb
aaabb'bab	abbbabbb'
aaabb'bbb	abbbb'bbb
aabaab'ab	bbbbbbbb'

Références

[1] D. R. Bean, A. Ehrenfeucht, G. F. Mc Nulty, Avoidable patterns in strings of symbols, *Pacific J. Math.* **85** (1979), 261-294.

[2] J.-M. Champarnaud, G. Hansel and D. Perrin, Unavoidable sets of constant length, *Report IGM 2002–06*, Université de Marne-la-Vallée, *Report LIFAR 2002–03*, Université de Rouen, 2002.

[3] J.-P. Duval, Factorizing words over an ordered alphabet, *J. Algorithms*, 4(4) :363–381, 1983.

[4] G. Han and D. Perrin, Ensembles inévitables, *Séminaire Lotharingien de Combinatoire*, 2002.

[5] P. M. Higgins and C. J. Saker, Unavoidable sets of words of uniform length, *Technical report*, University of Essex, 2001.

[6] M. Lothaire, *Combinatorics on Words*, vol. 17 of *Encyclopedia of Mathematics and Applications*, Addison-Wesley, Reading, Mass., 1983. Reprinted in the *Cambridge Mathematical Library*, Cambridge University Press, Cambridge, UK, 1997.

[7] M. Lothaire, *Algebraic Combinatorics on Words*, vol. 90 of *Encyclopedia of Mathematics and Applications*, Cambridge University Press, Cambridge, UK, 2002.

[8] M.-P. Schützenberger, On the synchronizing properties of certain prefix codes, *Inform. and Control*, 7 :23–36, 1964.

LIFAR, Université de Rouen,
76821 Mont-Saint-Aignan Cedex
jmc@dir.univ-rouen.fr