SUITES STURMIENNES

Nicolas Chevallier 2011

1 Complexité d'une suite

Définition 1 *Soit $\mathcal{A}$ un alphabet fini et $u \in \mathcal{A}^{\mathbf{N}}$. La complexité de la suite u est l'application $p_u : \mathbf{N}^* \to \mathbf{N}$ définie par*

$$p_u(n) = \operatorname{card} \mathcal{L}_n(u)$$

où $\mathcal{L}_n(u)$ est l'ensemble des facteurs de longueur n de u.

Exemples. 1. Soit $I_1 \cup I_2$ une partition du tore $\mathbf{T}^1$ et $T : x \in \mathbf{T}^1 \to x + \alpha \in \mathbf{T}^1$ une translation. Si $u = \pi(x)$ est le codage du point x défini par T et la partition alors $p_u(n) \leq 2n$.
2. $\mathbf{T}^1 = [0, \frac{1}{2}[\cup[\frac{1}{2}, 1[$ et $T : x \in \mathbf{T}^1 \to 2x \in \mathbf{T}^1$. D'après Birkhoff presque tout x à un codage $\pi(x)$ de complexité 2^n : tout mot w de longueur n définit un intervalle I_w de longueur $2^{-|w|}$ tel que pour tout $x \in I_w$ on a $\pi(x)_i = w_{i+1}$, $i = 0, ..., n-1$. D'après le théorème de Birkhoff presque tout point passe par I_w et il y a un nombre dnb de tels intervalles.
Exercice. Trouver une suite explicite $u \in \{0,1\}^{\mathbf{N}}$ telle que $p_n(u) = 2^n$.

Définition 2 ***Graphe de Rauzy d'une suite****. Soit $\mathcal{A}$ un alphabet fini, $u \in \mathcal{A}^{\mathbf{N}}$ et $n \geq 1$. Le graphe de Rauzy $G_n = (S_n, A_n)$ est le graphe orienté dont l'ensemble des sommets est $S_n = \mathcal{L}_n(u)$ et dont l'ensemble des arêtes est $A_n = \mathcal{L}_{n+1}(u)$. Pour chaque arête $a \in A_n$, l'origine de a est $o(a) = a_1...a_n$ et l'extrémité terminal est $t(u) = a_2...a_n$.*

Deux mots de longueur n sont liés par une arête ssi on peut les trouver consécutivement dans la suite u. Si on ne tient pas compte de l'orientation les graphes de Rauzy sont connexes et si tout préfixe de u réapparaît alors les graphes de Rauzy sont fortement connexes.

Proposition 1 *Soit $u \in \{0,1\}^{\mathbf{N}}$. Si il existe un entier $n \geq 1$ tel que $p_u(n) \leq n$ alors u est ultimement périodique (périodique à partir d'un certain rang).*

Dem. Soit n_0 le plus petit entier tel que $p_u(n) \leq n$. Si $n_0 = 1$, il n'y a qu'une lettre et u est constante. Supposons donc $n_0 > 1$. Considérons le graphe G_{n_0-1}. Par définition de n_0, il a au moins n_0 sommets et au plus n_0 arêtes. Comme chaque sommet est l'origine d'au moins une arête, chaque sommet est l'origine d'une arête exactement. Cela définit une application successeur s sur l'ensemble des sommets. Soit x_0 le préfixe de longueur $n_0 - 1$ de u. Soit $k_1 < k_2$ les plus petits entiers tels que $s^{k_1}(x_0) = s^{k_2}(x_0)$. On vérifie que la suite u est périodique à partir de k_1 avec période $k_2 - k_1$. $\square$

Lemme 1 *Soit $u \in \{0,1\}^{\mathbf{N}}$. Si u est ultimement périodique alors la complexité de u est bornée.*

Dem. : exercice.
Exercice. Soit $u \in \{0,1\}^{\mathbf{N}}$. Montrer que s'il existe $n_0 \in \mathbf{N}^*$ tel que $p_u(n_0) = p_u(n_0+1)$ alors u est ultimement périodique de période $\leq p_u(n_0)$.

2 Suites sturmiennes

Définition 3 *Une suite $u \in \{0,1\}^{\mathbf{N}}$ est sturmienne si $p_u(n) = n+1$ pour tout $n \geq 1$.*

Lemme 2 *Une suite sturmienne est récurrente.*

Dem. Si un facteur w d'une suite sturmienne u n'apparaissait qu'un nombre fini de fois alors $p_{S^n u}(|w|) \leq |w|$ pour n assez grand et donc la suite $S^n u$ serait ultimement périodique ce qui est impossible car u ne l'est pas. $\square$

Lemme 3 *Le langage $\mathcal{L}_2(u)$ d'une suite sturmienne ne peut pas contenir simultanément les mots* 00 *et* 11.

Dem. Si une suite sturmienne contient les mots 00 et 11 alors elle doit contenir aussi les mots 01 et 10 pour passer du mot 00 au mot 11 et inversement. Il y a donc 4 mots de longueur 2. □

Définition 4 *Une suite sturmienne u est de type* 0 *ou* 1 *suivant que* 00 *ou* 11 *est un facteur de u.*

Lemme 4 *Soit $u \in \{0,1\}^{\mathbf{N}}$ une suite sturmienne. Pour chaque $n \in \mathbf{N}^*$, le langage $\mathcal{L}_n(u)$ contient exactement un mot ayant deux prolongements à droite et un mot ayant deux prolongements à gauche (facteur bi-spécial ou bi-prolongeable droit et bi-spécial gauche).*

Dem. : exercice.

Lemme 5 *Une suite sturmienne ne peut pas contenir des plages arbitrairement longues d'une seule lettre.*

Dem. Soit u une suite sturmienne de type 0. Comme u est récurrente, 1 apparait au moins deux fois, u admet au moins un facteur du type 10^p1. Considérons un tel facteur avec p minimal et supposons que u contienne des plages de 0 de longueur arbitraire. En considérant la première apparition de ce facteur après une plage 0 de longueur $\geq p+1$, on voit que $0^{p+1}10^p1$ est facteur de u. Après ce facteur, 0^{p+1} doit à nouveau apparaître. Donc il existe v tel que $w = 0^{p+1}10^p1v0^{p+1}$ soit un facteur de u sans que 0^{p+1} soit un facteur de $v0^p$. Par conséquent, soit v est le mot vide, soit v se termine par un 1 ; dans les deux cas 0^{p+1} est précédé d'un 1. Notons $n = |w|$ et considérons $w' = 0^{n-2}1$ qui est un facteur de u.

Tous les suffixes $w_{\geq k} = w_k ... w_n$, $k = 2, ..., n-p-1$ de w et les suffixes $w'_{\geq k} = w'_k ... w'_n$, $k = 1, ..., n-p-2$ ont une longueur $\leq n-1$. Tous ces mots $w_{\geq k}$ et $w'_{\geq k}$ sont donc des préfixes de mots $\in \mathcal{L}_{n-1}(u)$. On vérifie que tous ces mots sont distincts :
- la première apparition du facteur 0^{p+1} change de position pour chaque $w_{\geq k}$, donc les prolongements des $w_{\geq k}$ sont distincts,
- la plage de 0 qui débute chaque $w'_{\geq k}$ à une longueur décroissante, donc les prolongements des $w'_{\geq k}$ sont distincts,
- cette plage 0 a un longueur $\geq p+1$ et 0^{p+1} n'est un préfixe d'aucun $w_{\geq k}$, donc aucun prolongement d'un $w_{\geq k}$ ne peut être égal à un prolongement d'un $w'_{\geq k}$.

Or le nombre de total des $w_{\geq k}$ et des $w'_{\geq k}$ est $N = n-p-2+n-p-2$ et $n = |w| \geq 3p+4$ donc $N \geq n + (3p+4) - (2p+4) = n+p$. Finalement, comme $p \geq 1$, N est strictement supérieur au nombre de mots de longueur $n-1$ d'une suite sturmienne. □

Lemme 6 *Si $u \in \{0,1\}^{\mathbf{N}}$ est une suite sturmienne alors l'une au moins des suites $1u$ ou $0u$ a le même langage que u. Elle est donc sturmienne.*

Dem. Pour chaque $n \geq 1$, notons u^n le préfixe de u de longueur n. Comme u est récurrente, il existe $k_n \geq 1$ tel que u^n soit aussi le préfixe $S^{k_n}u$. Par construction, la suite $S^{k_n}u$ converge vers u lorsque $n \to \infty$. Soit v une valeur d'adhérence de la suite $(S^{k_n-1}u)_n$. On a $Sv = u$ et donc $\mathcal{L}(u) \subset \mathcal{L}(v)$. De plus $\mathcal{L}(v) \subset \mathcal{L}(u)$ car $v \in \overline{\mathcal{O}(u)}$ donc v est sturmienne. □

Définition 5 ***équilibre.*** *Une suite $u \in \{0,1\}^{\mathbf{N}}$ est équilibrée si pour tout couple de facteurs (v,w) de u de même longueur, $||v|_0 - |w|_0| \leq 1$.*

Remarque. $||v|_0 - |w|_0| \leq 1 \Leftrightarrow ||v|_1 - |w|_1| \leq 1$.

Lemme 7 *Soit $u \in \{0,1\}^{\mathbf{N}}$. Si il existe $n \in \mathbf{N}^*$ tel que $\forall v, w \in \mathcal{L}_n(u),\ |v|_1 = |w|_1$ alors u est périodique.*

Dem. : exercice.

Théorème 1 *Une suite $u \in \{0,1\}^{\mathbf{N}}$ est sturmienne ssi elle est équilibrée et non ultimement périodique.*

Dem. 1.

Lemme 8 *Soit $v \in \{0,1\}^{\mathbf{N}}$ une suite qui ne soit pas équilibrée. Alors il existe $w \in \mathcal{L}(u)$ tel que*
- $0w0$ et $1w1 \in \mathcal{L}(u)$
- $\forall m, m' \in \mathcal{L}(u),\ |m| = |m'|$ et $||m|_1 - |m'|_1| \geq 2 \Rightarrow |m| \geq |w| + 2$.

Dem. Soit a et b deux mots de $\mathcal{L}(u)$ de même longueur n tels que $||a|_1 - |b|_1| \geq 2$. On peut supposer que n est minimal. Notons $m_{\leq k}$ le préfixe de longueur k du mot m. Posons $d_0 = 0$ et pour $i = 1, ..., n,$, $d_i = |a_{\leq i}|_1 - |b_{\leq i}|_1$. On vérifie que $d_{i+1} - d_i = -1, 0$ ou 1. Comme $|d_n| \geq 2$, il existe k minimal tel que $|d_k| \geq 2$ et par minimalité de n on a $k = n$. Si $a_1 \neq a_n$ ou si $b_1 \neq b_n$ alors l'un des 4 couples de mots $(a_i....a_j, b_p...b_q)$ avec $(i,j), (p,q) \in \{(1, n-1),\ (2, n)\}$, est deséquilibré. La minimalité implique donc que $a_1 = a_n$ et $b_1 = b_n$. L'égalité $a_1 = b_1$ contredit encore la minimalité donc $a_1 \neq b_1$. Supposons par exemple $a_1 = 1$ et $b_1 = 0$. Si $n = 2$ le lemme est prouvé. Supposons $n \geq 3$. Pour finir la preuve du lemme il suffit de prouver par récurrence sur k que $a_k = b_k$, $k = 2, ..., n-1$. Si $a_2 = 1$ et $b_2 = 0$ alors le couple a_1a_2, b_1b_2 est déséquilibré et si $a_2 = 0$ et $b_2 = 1$ alors c'est le couple $a_3...a_n$, $b_3...b_n$ qui l'est. On a donc par minimalité de n, $a_2 = b_2$. Supposons $a_i = b_i$, $i = 2, ..., k$. Si $a_{k+1} = 1$ et $b_{k+1} = 0$ alors le couple $a_1...a_{k+1}$, $b_1...b_{k+1}$ est déséquilibré et si $a_{k+1} = 0$ et $b_{k+1} = 1$ alors le couple $a_{k+1}...a_n$, $b_{k+1}...b_n$ l'est. Par conséquent $a_{k+1} = b_{k+1}$. $\square$

Soit u une suite sturmienne. Supposons que u ne soit pas équilibrée. Nous allons en déduire une contradiction en montrant que u est ultimement périodique. Considérons un mot w du langage de u associé à u par le lemme précédent. Comme les mots 00 et 11 ne font pas tous deux parti du langage de u, w n'est pas le mot vide. Montrons par récurrence sur k que $w_k = w_{n+1-k}$, $k = 1, ..., [\frac{n}{2}]$ où $n = |w|$. Autrement dit w est un palindrome. Supposons la suite u de type 1. Si w_1 ou $w_n = 0$ alors $00 \in \mathcal{L}(u)$, par conséquent $w_1 = 1 = w_n$. Supposons $w_i = w_{n+1-i}$ pour $i = 1, ..., k$. Si $w_{k+1} = 1$ et $w_{n+1-(k+1)} = 0$ alors le couple $1w_1...w_k1$ et $0w_{n+1-k}...w_n0$ est déséquilibré et si $w_k = 0$ et $w_{n+1-(k+1)} = 1$ alors le couple $0w_1...w_k0$, $1w_{n+1-k}...w_n1$ est déséquilibré. Ce qui contredit la minimalité de n. Donc w est un palindrome.

Le mot w est bi-prolongeable à droite (et à gauche). Par unicité des facteurs bi-prolongeables, w est un suffixe du facteur bi-prolongeable à droite de $\mathcal{L}_{n+1}(u)$, donc $0w$ ou $1w$ est un facteur bi-prolongeable à droite et un seul de ces deux facteurs l'est. Supposons que ce soit $0w$. Appelons i un rang d'apparition de $1w1$ dans u : $u = u_0...u_{i-1}1w1...$.

Montrons que le mot $0w$ n'est pas un facteur de $v = u_i...u_{i+2n+1}$. Comme $|v| = 2n+2 < |1w1| + |0w| = 2n+3$, si le mot $0w$ est un facteur de v, alors il chevauche le préfixe $1w1$ de v. Cela veut dire que 0 qui est la première lettre de $0w$ est l'une des lettres w_k de w. On a donc $0w_1...w_{n+1-k} = w_k...w_n1$. Mais cela entraîne que $w_k = 0$ et $w_{n+1-k} = 1$ ce qui contredit le fait que w soit un palindrome.

Le mot v contient $2n+2-n = n+2$ facteurs de longueur $n+1$. Deux de ces facteurs sont égaux car $\mathcal{L}_{n+1}(u)$ contient $n+2$ éléments dont $0w$ qui n'est pas un facteur de v. Un facteur m apparait donc deux fois dans v sans qu'aucun facteur intermédiaire ne soit $0w$. Or $0w$ est l'unique facteur bi-prolongeable à droite, la suite u est donc périodique après la première occurence de m.

2. Supposons u équilibrée. Utilisons le lemme

Lemme 9 *Soit* $(\mathcal{L}_n)_{n\geq 0}$ *une suite telle que*
- $\forall n \geq 0,\ \mathcal{L}_n \subset \{0,1\}^n$,
- $\forall n \geq 1,\ \forall w \in \mathcal{L}_n$, *les facteurs de* w *de longueur* $n-1$ *appartiennent à* $\mathcal{L}_{n-1}$.
- $\forall w, w' \in \mathcal{L}_n,\ ||w|_0 - |w|_0| \leq 1$.
Alors pour tout $n \in \mathbf{N}$, *il existe au plus un mot* $w \in \mathcal{L}_n$ *tel que* $w0$ *et* $w1 \in \mathcal{L}_{n+1}$ *(bi-prolongeable à droite).*

Dem du lemme. Récurrence sur n. $\mathcal{L}_0 = \{\text{mot vide}\}$.
Supposons le résultat vrai pour n.
Cas 1. La deuxième hypothèse du lemme montre que si aucun élément de $\mathcal{L}_n$ n'est bi-prolongeable à droite alors aucun mot de $\mathcal{L}_{n+1}$ ne peut l'être.
Cas 2. Supposons qu'il y ait un unique mot $m \in \mathcal{L}_n$ bi-prolongeable à droite. Soit $w \in \mathcal{L}_{n+1}$ bi-prolongeable. On a $w = am'$ où $m' \in \mathcal{L}_n$. Les mots $am'0$ et $am'1 \in \mathcal{L}_{n+2}$ donc les mots $m'0$ et $m'1 \in \mathcal{L}_{n+1}$. Par unicité, $m' = m$.
Envisageons les deux cas : ou bien $0m$ et $1m \in \mathcal{L}_{n+1}$ et sont bi-prolongeables ou bien seul l'un de ces deux mots est bi-prolongeable. Dans le premier cas les 4 mots $0m0$, $0m1$, $1m0$, $1m1$ appartiennent à $\mathcal{L}_{n+2}$. Mais $0m0$ et $1m1$ ne peuvent appartenir à $\mathcal{L}_{n+2}$ sans contredire l'équilibre. Le deuxième cas est donc le seul possible. $\square$

Fin de la démonstration de du théorème. Si u est équilibrée, la suite $\mathcal{L}_n(u)$ vérifie les hypothèses du lemme. Donc pour chaque n, $\mathcal{L}_n(u)$ contient au plus un facteur bi-prolongeable à droite. En considérant l'application de $\mathcal{L}_{n+1}(u)$ dans $\mathcal{L}_n(u)$ qui à un élément de $\mathcal{L}_{n+1}(u)$ associe son préfixe, on voit que pour chaque n, $p_u(n+1) \leq p_u(n)+1$. Comme $p_u(1) \leq 2$, $p_u(n) \leq n+1$ pour tout n. $\square$

3 Fréquence des lettres.

Proposition 2 *Si $u \in \{0,1\}^{\mathbf{N}}$ est une suite équilibrée, alors la limite*

$$\lim_{n\to\infty} \frac{1}{n+1} |u_0...u_n|_1$$

existe.

Dem. Pour chaque entier $p \geq 1$, notons a_p le nombre minimal de 1 d'un facteur de longueur p de u. Fixons p et pour chaque n, effectuons la division de $n+1$ par p : $n+1 = qp+r$ où $0 \leq r < p$. En décomposant $u_0...u_n$ en mots de longueur p + un "reste", on obtient

$$qa_p \leq |u_0...u_n|_1 \leq q(a_p+1)+r \leq qa_p+q+p.$$

Donc

$$\frac{qa_p}{qp+r} \leq \frac{1}{n+1}|u_0...u_n|_1 \leq \frac{qa_p+q+p}{qp+r}.$$

Comme $\frac{qa_p}{qp+r} \geq \frac{qa_p}{qp+p} = \frac{a_p}{p}\frac{1}{1+1/q} \geq \frac{a_p}{p}(1-\frac{1}{q})$ et $\frac{qa_p+q+p}{qp+r} \leq \frac{qa_p+q+p}{qp} \leq \frac{a_p}{p}+\frac{1}{q}+\frac{1}{p}$, on obtient pour $n+1 \geq p^2$,

$$\frac{a_p}{p} - \frac{1}{p} \leq \frac{a_p}{p} - \frac{a_p}{pq} \leq \frac{1}{n+1}|u_0...u_n|_1 \leq \frac{a_p}{p}+\frac{2}{p}.$$

Par conséquent, la suite $(\frac{1}{n+1}|u_0...u_n|_1)_{n\geq 0}$ est de Cauchy. $\square$

Proposition 3 *Si $u \in \{0,1\}^n$ est sturmienne alors alors la limite*

$$\lim_{n\to\infty} \frac{1}{n+1} |u_0...u_n|_1$$

existe et n'appartient pas à $\mathbf{Q}$.

Dem. Comme u est sturmienne, elle est équilibrée, la fréquence limite existe donc. Supposons que la limite soit un rationnel $\frac{a}{b}$. Avec les mêmes notations que dans la proposition précedente on obtient, $ka_n \leq a_{kn} < a_{kn}+1 \leq k(a_n+1)$. Donc si n divise n' alors

$$\frac{a_n}{n} \leq \frac{a_{n'}}{n'} < \frac{a_{n'}+1}{n'} \leq \frac{a_n+1}{n}.$$

En faisant tendre k vers l'infini, on obtient $\frac{a_n}{n} \leq \frac{a}{b} \leq \frac{a_n}{n}+\frac{1}{n}$. En prenant $n=b$ on voit que $\frac{a}{b} = \frac{a_b}{b}$ ou $\frac{a_b+1}{b}$.

Supposons que $\frac{a}{b} = \frac{a_b}{b}$. Comme la suite $(\frac{a_{kb}}{kb})_k$ est croissante, $\geq \frac{a}{b}$ et converge vers $\frac{a}{b}$, elle est égale à $\frac{a}{b}$ pour tout k.

Comme la suite u n'est pas ultimement périodique, il existe $w \in \mathcal{L}_b(u)$ tel que $|w|_1 = a_b+1$. Comme u est récurrente, ce mot w apparait une une infinité de fois, donc il apparait dans deux positions congruentes modulo b. On peut donc trouver un mot m de longueur kb commençant et finissant par w. On a donc $|m|_1 \geq ka_b+2$. Mais cela contredit le fait que $\frac{a_b}{b} = \frac{a_{kb}}{kb}$ car $a_{kb}+1 = ka_b+1 < |m_1|$.

L'autre cas : $\frac{a}{b} = \frac{a_b+1}{b}$ se traite en inversant les rôles de 0 et de 1. $\square$

4 Suites sturmiennes associées aux rotations

Proposition 4 *Soit $\alpha \in]0,1[\backslash\mathbf{Q}$. Considérons les codages π et π' définis par la translation $\tau : x \in [0,1[\to x+\alpha \bmod 1$ et les intervalles $I_0 = [0,1-\alpha[$ et $I_1 = [1-\alpha,1[$ où les intervalles $I'_0 =]0,1-\alpha]$ et $I'_1 =]1-\alpha,1]$ (codages sturmiens). Alors pour tout $x \in [0,1[$, $\pi(x)$ est sturmienne.*

Dem. Soit $w \in \{0,1\}^n$ et $I_w = \cap_{i=1}^n \tau^{-i+1} I_{w_i}$. Dans le tore $\mathbf{T}^1$ considérons les intervalles semi-ouverts $J_0, ..., J_n$ définis par les points $-k\alpha$, $k = 0, ..., n$. Soit x et y deux points du même intervalle J_k. On peut supposer que l'intervalle $[x, y]$ orienté positivement est inclus dans J_k. Pour tout $i = 0, ..., n-1$, l'intervalle $]\tau^i x, \tau^i y]$ ne contient ni 0 ni $-\alpha$, sinon $]x, y]$ contiendrait $-i\alpha$ ou $-(i+1)\alpha$. Par conséquent, soit $\tau^i x$ et $\tau^i y$ sont soit tous deux dans I_0 soit tous deux dans I_1. Ainsi chaque intervalle J_k est inclus dans un I_w pour un certain w de longueur n. Il y a donc au plus $n+1$ mots $w \in \{0,1\}^n$ tels que $I_w \neq \emptyset$. Par conséquent, pour tout $x \in \mathbf{T}^1$, $p_{\pi(x)}(n) \leq n+1$. Il reste à voir que $\pi(x)$ n'est pas ultimement périodique. Supposons le contraire : $\pi(x) = vw^\infty$. Pour tout n, on doit avoir $\tau^{|v|+n|w|} x \in I_w$ mais ceci est impossible car $I_w \subset I_0$ ou I_1 et comme $\alpha \notin \mathbf{Q}$, la suite $(T^{|v|+n|w|} x)_{n \geq 0}$ est dense. $\square$

Remarque. Le théorème de Birkhoff montre immédiatement que les suites sturmiènnes construites par la proposition précédente ont une fréquence de 1 égale à α.

Problème : réciproque ?

Lemme 10 *Soit $u, v \in \{0,1\}^{\mathbf{N}}$ deux suites équilibrées ayant même fréquence α de 1. Notons pour chaque $n \geq 1$, n_u et n_v le nombre minimum de 1 de $\mathcal{L}_n(u)$ et $\mathcal{L}_n(v)$. Si $\alpha \notin \mathbf{Q}$ alors $n_u = n_v$ pour tout $n \geq 1$.*

Dem. Supposons qu'il existe $b \geq 1$ tel que $b_u < b_v$. Pour chaque $n \in \mathbf{N}$, soit U^n le préfixe de longueur nb de u et V^n le préfixe de longueur nb de u. On a

$$\frac{|U^n|_1}{nb} \leq \frac{n(b_u+1)}{nb} \leq \frac{nb_v}{nb} \leq \frac{|V^n|}{nb}$$

donc

$$\alpha = \lim_{n\to\infty} \frac{|U^n|_1}{nb} \leq \frac{b_v}{b} \leq \lim_{n\to\infty} \frac{|V^n|_1}{nb} = \alpha.$$

Par conséquent $\alpha = \frac{b_v}{b} \in \mathbf{Q}$. $\square$

Corollaire 1 *Deux suites sturmiennes de même fréquence ont même langage.*

Dem. Soit u et v deux suites sturmiennes de même fréquence α. D'après le lemme précédent, pour chaque $n \geq 1$, $\mathcal{L}_n = \mathcal{L}_n(u) \cup \mathcal{L}_n(v)$ vérifie les hypothèses du lemme 7. Par conséquent, $\mathcal{L}_n$ contient au plus un facteur bi-prolongeable à droite et ainsi $\operatorname{card} \mathcal{L}_{n+1} \leq \operatorname{card} \mathcal{L}_n + 1$. Il en découle que $\operatorname{card} \mathcal{L}_n \leq n+1 = \operatorname{card} \mathcal{L}_n(u) = \operatorname{card} \mathcal{L}_n(v)$ donc $\mathcal{L}_n(u) = \mathcal{L}_n(v)$. $\square$

Exercice. Trouver une nouvelle démonstration du fait qu'une suite sturmienne ne peut pas admettre des plages arbitrairement longues d'une seule lettre.

Corollaire 2 *Les suites sturmiennes sont uniformément récurrente.*

Dem. Soit u sturmienne de fréquence α et $v = \pi(x)$ le codage d'un point quelconque x associé à α et aux intervalles $I_0 = [0, 1-\alpha[$ et $I_1 = [1-\alpha, 1[$. D'après le corollaire précédent, il suffit de montrer que v est uniformément récurrente. Soit $w \in \mathcal{L}(v)$ et $I_w = \cap_{i=0}^{|w|-1} \tau^{-i}(I_{w_i})$. I_w contient un élément de la forme $\tau^n(x)$ donc I_w est non vide. Comme les intervalles I_0 et I_1 sont semi-ouvert de même sens, l'intérieur J_w de I_w est non vide. Par minimalité de τ, il existe N tel que $\cup_{n=0}^N \tau^{-n} J_w = \mathbf{T}^1$. Donc $\forall y \in \mathbf{T}^1, \exists n \leq N$, $\tau^n(y) \in J_w$ et par conséquent, $\forall W \in \mathcal{L}(u)$ tel que $|W| \geq N + |w|$, w est un facteur de W. $\square$

Théorème 2 *(Hedlund-Morse 1938 et Coven-Hedlund 1973) Soit $u \in \{0,1\}^{\mathbf{N}}$ une suite sturmienne. Alors il existe un nombre irrationnel $\alpha \in]0,1[$ et $x \in [0,1[$ tel que $u = \pi(x)$ ou $\pi'(x)$, où π et π' sont les codages définis par la translation $\tau : x \in [0,1[\to x + \alpha \bmod 1$ et les intervalles $I_0 = [0, 1-\alpha[$ et $I_1 = [1-\alpha, 1[$ ou les intervalles $I'_0 =]0, 1-\alpha]$ et $I'_1 =]1-\alpha, 1]$.*

Dem. Soit u une suite sturmienne de fréquence α et $v = \pi(x)$ le codage associé à α d'un point quelconque x. On sait que u et v ont le même langage donc il existe une suite $(n_k)_k$ tel que $S^{n_k} v \to u$. A extraction près on peut supposer que la suite $(\tau^{n_k}(x))_k$ converge vers y. Si n_k ne tend vers l'infini lorsque $k \to \infty$ alors il existe p tel que $u = S^p v$ et donc $u = \pi(\tau^p(x))$ ce qui prouve le théorème dans ce cas. Supposons donc $n_k \to \infty$ et montrons que $u = \pi(y)$ ou $\pi'(y)$.

Plaçons nous dans le tore $\mathbf{T}^1$. Soit $i \in \mathbf{N}$ tel que $\tau^i(y) \notin \{0, -\alpha\}$. Pour un tel i, $\tau^i(y) \in I^\circ_{\pi(y)_i} = I'^\circ_{\pi(y)_i}$. Donc pour k assez grand, $\tau^i(\tau^{n_k}(x))$ est dans $I^\circ_{\pi(y)_i} = I'^\circ_{\pi(y)_i}$. Par conséquent, pour k assez grand,

$$u_i = S^{n_k}(v)_i = \pi(\tau^{n_k}(x))_i = \pi(y)_i = \pi'(y)_i.$$

On en déduit que si $y \notin -\mathbf{N}\alpha$ alors pour tout $i \in \mathbf{N}$, $\tau^i(y) \notin \{0, -\alpha\}$, et donc $u = \pi(y) = \pi'(y)$. Il reste à examiner le cas $y = -p\alpha$ avec $p \in \mathbf{N}$.

Cas 1 : $p = 0$. Comme α est irrationnel, pour tout $i \geq 1$, $y + i\alpha = i\alpha \notin \{0, -\alpha\}$ donc $u_i = \pi(y)_i = \pi'(y)_i$. Pour $i = 0$ on a soit $u_0 = 0 = \pi(y)_0$ soit $u_0 = 1 = \pi'(y)_0$.

Cas 2 : $p \geq 1$, $u_{p-1} = 1$ et $\alpha < \frac{1}{2}$. On a $\tau^{p-1}(y) = -\alpha \in I_1 = I_{u_{p-1}}$. De plus l'hypothèse $\alpha < \frac{1}{2}$ entraîne que u est de type 0 donc $u_p = 0$, donc $\tau^p(y) = 0 \in I_0 = I_{u_p}$. Pour tous les entiers $i \neq p - 1$ ou p, on sait déja que $\pi(y)_i = u_i$. Donc $u = \pi(y)$.

Cas 3 : $p \geq 1$, $u_{p-1} = 1$ et $\alpha > \frac{1}{2}$. On a $\tau^{p-1}(y) = -\alpha \in I_1 = I_{u_{p-1}}$. De plus le nombre maximum de 1 consécutifs dans v est inférieur ou égal à

$$a = [\frac{|I_1|}{d(\alpha, \mathbf{Z})}] + 1$$

et l'hypothèse $\alpha > \frac{1}{2}$ entraîne que $d(\alpha, \mathbf{Z}) = 1 - \alpha$ donc $a = [\frac{\alpha}{1-\alpha}] + 1$. Comme u et v ont le même langage, ce maximum est aussi valable pour u. Supposons $p \geq a$. Pour $k = 1, ..., a - 1$, les points $\tau^{p-1-k}(y) = y + (p - 1 - k)\alpha = -\alpha - k\alpha$ appartiennent à l'intérieur de I_1. On en déduit que $u_{p-1-k} = \pi(y)_{p-1-k}$ pour $k = 1, ..., a - 1$. Comme par hypothèse $u_{p-1} = 1$, l'égalité $u_p = 1$ entraînerait $u_{p-1-k} = 1$ pour $k = -1, 0, ..., a-1$ ce qui donnerait une plage de 1 de longueur $a+1$ dans u. Donc $u_p = 0 = \pi(y)_p$ et $u = \pi(y)$.

Si $p < a$, on se ramène à $p \geq a$ en utilisant que $n_k \to \infty$. En effet, avec une nouvelle extraction on peut supposer que $S^{n_k - a}v$ converge aussi vers une suite u', on a alors $u = \lim_{k\to\infty} S^a(S^{n_k-a}v) = S^a u'$. Comme le langage de la suite u' est inclus dans celui de la suite v, les plages de 1 de la suite u' ont une longueur $\leq a$. De plus la suite $(\tau^{n_k-a}(x))_k$ converge vers $y' = y - a\alpha = -(a+p)\alpha = -p'\alpha$. Le raisonnement avec $p \geq a$ s'applique donc à u' et y' et on otient $u' = \pi(y')$ ce qui donne $u = S^a u' = \pi(\tau^a(y')) = \pi(y)$.

Cas 4 : $p \geq 1$, $u_{p-1} = 0$ et $\alpha < \frac{1}{2}$. On a $\tau^{p-1}(y) = -\alpha \in I'_0$. Le nombre maximum de 0 consécutif est $a = [\frac{1-\alpha}{\alpha}] + 1$. Comme u et v ont le même langage, ce maximum est aussi valable pour u. Supposons $p \geq a$. Pour $k = 1, ..., a-1$, les points $\tau^{p-1-k}(y) = y + (p-1-k)\alpha = -\alpha - k\alpha$ appartiennent à l'intérieur de I'_0. On conclut comme dans le cas précédent que $u = \pi'(y)$.

Cas 5 : $p \geq 1$, $u_{p-1} = 0$ et $\alpha > \frac{1}{2}$. On a $\tau^{p-1}(y) = -\alpha \in I'_0$. La suite u est de type 1 donc $u_p = 1$ et $\tau^p(y) = 0 \in I'_{u_p}$. □

5 Théorème de Rauzy

5.1 Induction

Soit τ_0 et τ_1 les deux substitutions définies sur $\{0, 1\}$ par

$$\tau_0(0) = 0,\ \tau_0(1) = 01,\ \tau_1(0) = 10,\ \tau_1(1) = 1.$$

Proposition 5 *Soit $\alpha, \beta > 0$, $I = [-\alpha, \beta[$, $I_0 = [-\alpha, -\alpha + \beta[$, $I_1 = [-\alpha + \beta, \beta[$, $a = [\frac{\beta}{\alpha}]$, $\gamma = \beta - a\alpha$, $J = [-\alpha, \gamma[$, $J_0 = [-\alpha, -\alpha + \gamma[$, $J_1 = [-\alpha + \gamma, \gamma[$. Considérons les deux applications $r : I \to I$ et $r' : J \to J$ définies par*

$$r(x) = \begin{cases} x + \alpha,\ x \in I_0 \\ x - \beta,\ x \in I_1 \end{cases},\ r'(x) = \begin{cases} x + \alpha,\ x \in J_0 \\ x - \gamma,\ x \in J_1 \end{cases}.$$

1. r' est l'induite de r sur J.
2. Soit π et π' les codages définies par les applications r et r' et les partitions I_0, I_1 et J_0, J_1. Pour tout $y \in J$ on a $\pi(y) = \tau_0^a(\pi'(y))$.
3. Si on définit $a = [\frac{\alpha}{\beta}]$, $\gamma = \alpha - a\beta$, $J = [-\gamma, \beta[$, $J_0 = [-\gamma, -\gamma + \beta[$, $J_1 = [-\gamma + \beta, \beta[$ et $r : I \to I$ et $r' : J \to J$ par

$$r(x) = \begin{cases} x + \alpha,\ x \in I_0 \\ x - \beta,\ x \in I_1 \end{cases},\ r'(x) = \begin{cases} x + \gamma,\ x \in J_0 \\ x - \beta,\ x \in J_1 \end{cases}$$

alors r' est l'induite de r et pour tout $y \in J$, $\pi(y) = \tau_1^a(\pi'(y))$.

Remarque. r et r' sont des translations modulo $\alpha+\beta$ et $\alpha+\gamma$.

Dém. 1. Lorsque $a=0$, on a $r=r'$ et il n'y a rien à prouver. Supposons donc $a>0$ *i.e.* $\beta\geq\alpha$. Pour chaque $y\in J$, notons $t(y)$ le temps de retour dans J et r_1 la transformation induite :

$$t(y)=\min\{n\geq 1 : r^n(y)\in J\},\ r_1(y)=r^{t(y)}(y).$$

Pour $y\in J_0$, $r(y)=y+\alpha\in[0,\gamma[\subset J$, donc $t(y)=1$ et $r_1(y)=y+\alpha$.

Pour $y\in J_1$ et $1\leq l\leq a$,

$$y+l\alpha<\gamma+l\alpha\leq\beta,$$

donc $r^l(y)=y+l\alpha\in[\gamma,\beta[$. Par conséquent $t(y)>a$. De plus

$$\begin{aligned}y+a\alpha-\beta&<\gamma+a\alpha-\beta=0,\\ y+a\alpha-\beta&\geq-\alpha+\gamma+a\alpha-\beta=-\alpha,\end{aligned}$$

donc $r^{a+1}(y)\in[-\alpha,0[\subset J$. Par conséquent, $t(y)=a+1$ et $r_1(y)=y+a\alpha-\beta=y-\gamma$. Par conséquent $r_1=r'$.

2. Déterminons la relation entre les codages π et π'. Soit $y\in J$, $v=\pi(y)$ et $v'=\pi'(y)$. Appelons $t_k(y)=t_{k-1}(y)+t(r_1^{k-1}(y))$ le $k^{ème}$ temps de retour dans J ($t_0=0$ et $t_1=t$). Par définition, $r^n(y)\in J$ ssi il existe k tel que $n=t_k(y)$.

Si $a=0$, alors $\pi=\pi'$. Supposons donc $a\geq 1$. Dans ce cas $J\subset I_0$.

Si $v'_k=0$ alors $r'^k(y)\in J_0$ et d'après 1, $t_{k+1}(y)=t_k(y)+1$. Par conséquent, $\mathbf{N}\cap[t_k(y),t_{k+1}(y)[=\{t_k(y)\}$ et $v_{t_k(y)}=0$.

Si $v'_k=1$ alors $r'^k(y)\in J_1$ et d'après 1, $t_{k+1}(y)=t_k(y)+a+1$. Par conséquent, $\mathbf{N}\cap[t_k(y),t_{k+1}(y)[=\{t_k(y),...,t_k(y)+a\}$. Comme précédement, si $1\leq l\leq a-1$,

$$r^{t_k(y)}(y)+l\alpha<\gamma+l\alpha\leq\gamma+a\alpha-\alpha=\beta-\alpha,$$

donc $r^{t_k(y)+l}(y)\in I_0$. De même,

$$\beta-\alpha=\gamma+a\alpha-\alpha\leq r^{t_k(y)}(y)+a\alpha<\gamma+a\alpha=\beta$$

donc $r^{t_k(y)+a}(y)\in I_1$. Par conséquent, $v^{t_k(y)+l}(y)=0$, $l=0,...,a-1$ et $v^{t_k(y)+a}(y)=1$.

Conclusion $v=\tau_0^{a_1}(v^{'})$.

3. Il suffit d'intervertir les rôles de α et β. □

On déduit immédiatement de la proposition le corollaire suivant en utilisant les relations de récurrences vérifiées par le développement en fraction continue.

Corollaire 3 *Soit* $\alpha_0\in[0,1]$ *un nombre irrationnel. Définissons les suites* $(\delta_n)_{n\geq-1}$, $(a_n)_{n\geq1}$, $(I_n)_{n\geq0}$, $(I_{n,0})_{n\geq0}$, $(I_{n,1})_{n\geq0}$ *et* $(r_n)_{n\geq0}$ *par*

- $\delta_{-1}=1$, $\delta_0=\alpha_0$, *pour* $n\geq1$, $a_n=[\frac{\delta_{n-2}}{\delta_{n-1}}]$ *et* $\delta_n=\delta_{n-2}-a_n\delta_{n-1}$,

- si n est pair, $I_n=[-\delta_n,\delta_{n-1}[$, $I_{n,0}=[-\delta_n,-\delta_n+\delta_{n-1}[$, $I_{n,1}=[-\delta_n+\delta_{n-1},\delta_{n-1}[$, *et* $r_n:I_n\to I_n$ *définie par*

$$r_n(x)=\begin{cases}x+\delta_n,\ x\in I_{n,0}\\ x-\delta_{n-1},\ x\in I_{n,1}\end{cases}$$

- si n est impair, $I_n=[-\delta_{n-1},\delta_n[$, $I_{n,0}=[-\delta_{n-1},-\delta_{n-1}+\delta_n[$, $I_{n,1}=[-\delta_{n-1}+\delta_n,\delta_n[$, *et* $r_n:I_n\to I_n$ *définie par*

$$r_n(x)=\begin{cases}x+\delta_{n-1},\ x\in I_{n,0}\\ x-\delta_n,\ x\in I_{n,1}\end{cases}.$$

Notons enfin π^n *le codage associé à* r_n *et à la partition* $I_{n,0}$, $I_{n,1}$.

Pour tout $y\in I_{n+1}$, *on a* $\pi^0(y)=\tau_0^{a_1}\circ\tau_1^{a_2}\circ...\circ\tau_{n \bmod 2}^{a_{n+1}}(\pi^{n+1}(y))$. *De plus* $\alpha_0=[0;a_1,a_2,...,a_n,...]$.

Remarque. Attention si on normalise r_0 par longueur de $I_0=1$ on obtient la translation $x\to x+\frac{\alpha_0}{|I_0|}=x+\frac{\alpha_0}{1+\alpha_0}$. Le développement en fraction continue de $\alpha=\frac{\alpha_0}{1+\alpha_0}$ est $[0;a_1+1,a_2,...]$.

5.2 “Désubstitution” d’une suite sturmienne

Lemme 11 *Soit $i \neq j \in \{0,1\}$. Si $u \in \{0,1\}^{\mathbf{N}}$ est une suite ne contenant pas le facteur jj et si $u_0 = i$ alors il existe une unique suite u' telle que $u = \tau_i(u')$.*

Dém. Supposons $i = 0$ (démonstration identique dans le cas $i = 1$). Démontrons par récurrence sur la longueur que pour chaque mot v débutant par 0 et ne contenant pas le facteur 11, il existe un unique mot v' tel que $\tau_0(v') = v$. Si $|v| = 1$ ou 2, c’est évident. Soit $n \geq 2$. Supposons le résultat vrai pour toutes les longueurs $\leq n$ et soit v un mot de longueur $n+1$ débutant par 0 et ne contenant pas le facteur 11. Si $v = 00w$ alors l’hypothèse de récurrence appliquée à $0w$ montre qu’il existe w' tel que $0w = \tau_0(w')$ et donc $v = \tau_0(v')$ avec $v' = 0w'$. L’unicité de v' vient de l’hypothèse de récurrence et du fait que v' doit débuter par un 0. Si $v = 01w$ alors la première lettre de w est un 0 et $v = \tau(1)0x$. L’hypothèse de récurrence appliqué à $0x$ implique l’existence de v' tel que $\tau(v') = v$. L’unicité de v' vient de l’hypothèse de récurrence et du fait que v' débute par un 1.

Supposons que $u_0 = 0$. D’aprés ce qui précède pour chaque préfixe v de u, il existe v'unique tel que $v = \tau_0(v')$. Si $v = xy$ est un préfixe de u tel que y débute par un 0 alors par unicité $v' = x'y'$. Ce qui montre que l’ensemble des v' tels que la lettre suivant v soit un 0 est l’ensemble des préfixes d’un mot infini u' tel que $\tau_0(u') = u$. Pour l’unicité considérons un préfixe v de u tel que la lettre suivant v soit un 0. Si v' est le préfixe de longueur maximale de u' tel que $\tau_0(v')$ soit un préfixe de v alors $|\tau_0(v')| \leq |v| < |\tau_0(v'a)| = |\tau_0(v')| + |\tau_0(a)|$ où a est la lettre qui suit v'. Par conséquent, si $|\tau_0(v')| < |v|$ alors $|\tau_0(a)| \geq 2$ et donc $a = 1$. Mais cela contredit le fait que v soit suivi d’un 0. Ainsi v' est l’unique suite telle que $\tau_0(v') = v$ et l’unicité de u' en découle. $\square$

Lemme 12 *Soit u et $u' \in \{0,1\}^{\mathbf{N}}$ deux suites telles que $\tau_i(u') = u$. Si u est sturmienne alors u' l’est aussi.*

Dém. Supposons $i = 0$, u est alors sturmienne de type 0. Supposons que u' ne soit pas sturmienne. Si u' était périodique à partir d’un certain rang, u le serait aussi ; donc u' n’est pas équilibrée. Il existe donc $w \in \mathcal{L}(u')$ tel que $0w0$ et $1w1 \in \mathcal{L}(u')$. Soit $a \in \{0,1\}$ tel que $0w0a \in \mathcal{L}(u')$. Les mots $\tau_0(0w0a) = 0\tau_0(w)0\tau_0(a)$ et $\tau_0(1w1) = 01\tau_0(w)01$ appartiennent à $\mathcal{L}(u)$. Comme $\tau_0(a)$ débute par un 0, $0\tau_0(w)00$ et $1\tau_0(w)01$ appartiennent à $\mathcal{L}(u)$ ce qui contredit l’équilibre de u. $\square$

Exercice Etudier la réciproque.

Notations. Notons $\mathcal{S}$ l’ensemble des suites sturmiennes et $\phi : \mathcal{S} \to \{0,1\}^{\mathbf{N}}$ l’application définie pour tout $u \in \mathcal{S}$ de type i, par

$$\phi(u) = u'$$

où u' est l’unique suite telle que $\tau_i(u') = u$ lorsque $u_0 = i$, et u' est l’unique suite telle que $\tau_i(u') = iu$ lorsque $u_0 \neq i$.

Corollaire 4 *Si $u \in \{0,1\}^{\mathbf{N}}$ est une suite sturmienne alors $\phi(u)$ est sturmienne.*

Démonstration. Supposons que u soit de type 0. Si $u_0 = 0$ alors $u = \tau_0(u')$ et on utilise le lemme précédent. Si $u_0 = 1$ alors $0u$ est sturmienne d’après le lemme de prolongement des suites sturmiennes et on applique le lemme à $0u$. $\square$

5.3 Théorème de Rauzy

Théorème 3 *(Rauzy, publié dans Arnoux-Rauzy 1991) Soit u une suite sturmienne. Pour chaque $n \geq 1$, notons i_n le type de $u^{n-1} = \phi^{n-1}(u)$ et posons $b_n = 0$ si $u_0^{n-1} = i_n$ et $b_n = 1$ sinon, de tel sorte que :*

$$u = S^{b_1} \circ \tau_{i_1} \circ ... \circ S^{b_n} \circ \tau_{i_n}(u^n).$$

1. La suite des types $i_1 i_2 ... i_n ...$ est une suite de 0 et de 1 qui n’est pas constante à partir d’un certain rang. Ecrivons cette suite sous la forme $i_1 i_2 ... i_n ... = 0^{a_1} 1^{a_2} 0^{a_3} 1^{a_4} ...$ où $a_i \geq 1$ pour $i \geq 2$ (la suite ne commence pas nécessairement par un 0).

2. Posons $\alpha = [0; a_1 + 1, a_2, a_3, ...]$ (développement en fraction continue). Notons π et π' les applications codages définis par la translation $\tau : x \in [0,1[\to x + \alpha \bmod 1$ et les intervalles $I_0 =$

$[0, 1-\alpha[$ et $I_1 = [1-\alpha, 1[$ ou les intervalles $I'_0 =]0, 1-\alpha]$ et $I'_1 =]1-\alpha, 1]$. Alors $\pi(\alpha) = \lim_{n\to\infty} \tau_0^{a_1} \circ \tau_1^{a_2} \circ ... \circ \tau_{n-1 \bmod 2}^{a_n}(\pi^n(0)_0)$.
3. u et $\pi(\alpha)$ (et $\pi'(\alpha)$) ont le même langage et
4. (Morse-Hedlund nouvelle démonstration). Il existe $x \in [0, 1[$ tel que $u = \pi(x)$ ou $\pi'(x)$.

Dém. 1. Supposons $i_n = 0$ pour tous les $n \geq n_0$ (on raisonne de la même manière avec $i_n = 1$). La suite u^n étant sturmienne la distance entre deux 1 consécutifs est bornée par une constante N_n car les plages de 0 sont bornées. Par définition de τ_0, pour $n \geq n_0$, on a $N_n = N_{n+1} + 1$. La suite $(N_n)_{n\geq n_0}$ est donc strictement décroissante. Mais lorsque $N_n \leq 1$ la suite u^n est de type 1 ce qui contredit l'hypothèse $i_n = 0$ pour tout $n \geq n_0$.

Faisons d'abord la démonstration dans le cas $a_1 > 0$. Le cas $a_1 = 0$ se ramène au cas $a_1 > 0$, voir la fin de la démonstration. Soit $\alpha_0 = [0; a_1, a_2, ...]$ et $\alpha = \frac{\alpha_0}{1+\alpha_0} = [0; a_1 + 1, a_2, ...]$. Utilisons les notations du corollaire sur l'induction. Considérons l'application $f : [0, 1[\to [-\alpha_0, 1[$ qui à x associe $-\alpha_0 + (1+\alpha_0)x$. Comme $f(1-\alpha) = -\alpha_0 + (1+\alpha_0)(1 - \frac{\alpha_0}{1+\alpha_0}) = -\alpha_0 + 1$, on a $f(I_0) = I_{0,0}$ et $f(I_1) = I_{0,1}$, et on vérifie que f conjugue les translations $r(x) = x + \alpha \bmod 1$ et r_0.

2. Comme pour tout $n \geq 0$, $0 \in I_n$, d'après le corollaire sur l'induction, $\pi^0(0) = \tau_0^{a_1} \circ \tau_1^{a_2} \circ ... \circ \tau_{n-1 \bmod 2}^{a_n}(\pi^n(0))$. Or $f(\alpha) = 0$, donc $\pi(\alpha) = \pi^0(0)$ (Idem avec des primes). Comme la longueur du mot $\tau_0^{a_1} \circ \tau_1^{a_2} \circ ... \circ \tau_{n-1 \bmod 2}^{a_n}(\pi^n(0)_0)$ tend vers ∞, $\pi(\alpha) = \lim_{n\to\infty} \tau_0^{a_1} \circ \tau_1^{a_2} \circ ... \circ \tau_{n-1 \bmod 2}^{a_n}(\pi^n(0)_0)$.

3. Comme pour chaque l, $\mathrm{card}\, \mathcal{L}_l(u) = \mathrm{card}\, \mathcal{L}_l(\pi(\alpha)) = l + 1$, il suffit de montrer que $\mathcal{L}(u) \subset \mathcal{L}(\pi(\alpha))$. Soit $l \geq 1$. Comme la suite u est uniformément récurrente, il existe un entier L tel que tous les mots de longueur l sont facteurs de tout mot de longueur L. Il existe un entier k tel que la longueur de $\tau_0^{a_1} \circ \tau_1^{a_2} \circ ... \circ \tau_{k-1 \bmod 2}^{a_k}(\pi^k(0)_0)$ soit $\geq L$. Soit $n = a_1 + a_2 + ... + a_k$. Choisissons $m > B = b_1 + ... + b_n$ tel que $u^n_m = \pi^k(0)_0$. Remarquons que si τ est une substitution et $b \geq 1$ est un entier, alors pour tout mot infini v, $\tau \circ S^b(v)$ est un suffixe de $S^b \circ \tau(v)$ car une substitution n'efface pas de lettre. Le mot $\tau_{i_1} \circ ... \circ \tau_{i_n}(S^B(u^n))$ est donc un suffixe de $u = S^{b_1} \circ \tau_{i_1} \circ ... \circ S^{b_n} \circ \tau_{i_n}(u^n)$, par conséquent $\tau_0^{a_1} \circ \tau_1^{a_2} \circ ... \circ \tau_{k-1 \bmod 2}^{a_k}(\pi^k(0)_0) = \tau_{i_1} \circ ... \circ \tau_{i_n}(u^n_m)$ est un facteur de u. Comme ce facteur a une longueur $\geq L$, il contient tous les mots de $\mathcal{L}_l(u)$ et donc $\mathcal{L}_l(u) \subset \mathcal{L}_l(\pi(\alpha))$.

4. Remarquons que pour tout $y \in [0, 1[$, un mot w de longueur n appartient à $\mathcal{L}(\pi(y))$ ssi $I_w = \cap_{i=0}^{n-1} r^{-i}(I_{w_i}) \neq \emptyset$ où r est la translation de α modulo 1. Par conséquent, pour chaque préfixe w^n de longueur n de u, $I_{w^n} \neq \emptyset$. Choisissons pour chaque n un point $x_n \in I_{w^n}$. Comme la suite $(I_{w^n})_n$ décroît et que la longueur de I_{w^n} tend vers 0, x_n converge vers un x. Par définition, $\pi(x_n)$ converge vers u. Si $x \in \cap_{n\geq 0} I_{w^n}$, alors $u = \pi(x)$. Si $x \notin \cap_{n\geq 0} I_{w^n}$, alors il existe un entier $n \geq 0$ tel que x soit une extrémité droite de $r^{-n}(I_{w^n_n}) = r^{-n}(I_{u_n})$. Notons encore n le minimum de ces entiers. Vérifions que $u = \pi'(x)$ où π' est le codage associé aux intervalles $I'_0 =]0, 1-\alpha]$, $I'_1 =]1-\alpha, 1]$.

Cas 1. $u_n = 0$. x est donc l'extrémité droite de $r^{-n}(I_0)$ donc $x = (1-\alpha) - n\alpha \bmod 1$. $r^k(x)$ n'est une extrémité de I_0 ou I_1 que pour $k = n$ et $n+1$. Par conséquent, pour tout $k \neq n, n+1$, $r^k(x)$ appartient à l'intérieur de I_{u_k} et donc $u_k = \pi(x)_k$. Si $u_{n+1} = 1$ alors $\pi'(x) = u$. Si $u_{n+1} = 0$ alors u est de type 0 et $\alpha < \frac{1}{2}$. Le nombre maximum de 0 consécutifs d'un codage $\pi(y)$ est $a = [\frac{1-\alpha}{\alpha}] + 1 = [\frac{1}{\alpha}]$. On sait que $r^{n+1}(x) = 0 \bmod 1$ et donc $u_{n+2} = ... = u_{n+a} = 0$. Donc $u_n = ... = u_{n+a} = 0$ mais ceci est impossible car u aurait $a + 1$ 0 consécutifs.

Cas 2. $u_n = 1$. x est donc l'extrémité droite de $r^{-n}(I_1)$ donc $x = -n\alpha \bmod 1$. $r^k(x)$ n'est une extrémité de I_0 ou I_1 que pour $k = n-1$ et n. Par conséquent, pour tout $k \neq n-1, n$, $r^k(x)$ appartient à l'intérieur de I_{u_k}. Donc si $n = 0$, $u = \pi'(x)$. Supposons donc $n \geq 1$. Si $u_{n-1} = 0$ alors $\pi'(x) = u$. Si $u_{n-1} = 1$ alors est de type 1 et $\alpha > \frac{1}{2}$. Le nombre maximum de 1 consécutifs d'un codage $\pi(y)$ est $a = [\frac{1}{1-\alpha}]$. On sait que $r^n(x) = 0 \bmod 1$ et donc $u_{n+1} = ... = u_{n+a} = 1$. Donc $u_{n-1} = ... = u_{n+a} = 1$ mais ceci est impossible car u aurait $a + 1$ 1 consécutifs.

Il reste le cas $a_1 = 0$. Considérons la substitution σ qui intervertit 0 et 1 et la suite $\sigma(u) = v$, nous obtenons une nouvelle suite sturmienne v. On a alors $v = S^{b'_1} \circ \tau_{i'_1} \circ ... \circ S^{b'_n} \circ \tau_{i'_n}(v^n)$ où $v^n = \phi^n(v)$. L'inversion des lettres 0 et 1 implique que pour tout $k \geq 1$, i'_k et i_k sont aussi inversés $(0 \rightleftarrows 1)$ et que $b'_k = b_k$. Par conséquent, $a'_k = a_{k+1}$. Posons donc $\beta = [0; a_2 + 1, a_3, a_4, ...]$ et notons ρ et ρ' les codages sturmiens associés à β. D'après le cas $a_1 > 0$, les suites v, $\rho(\beta)$ et $\rho'(\beta)$ ont le même langage et il existe y tel que $v = \rho(y)$ ou $\rho'(y)$. On remarque que $\beta + \alpha = 1$. Considérons donc l'application $g : x \in [0, 1[\to 1 - x \in]0, 1]$, on a $g(I_0) =]1-\beta, 1]$ et $g(I_1) =]0, 1-\beta]$

et inversement. Par conséquent, pour tout $z \in \mathbf{T}^1$, $\sigma(\rho(z)) = \pi'(g(z))$ et $\sigma(\rho'(z)) = \pi(g(z))$ et le théorème dans le cas $a_1 = 0$, découle du cas $a_1 \neq 0$. $\square$