ANNALES

DE

L'INSTITUT FOURIER

Julien CASSAIGNE & Nataliya CHEKHOVA

Fonctions de récurrence des suites d'Arnoux-Rauzy et réponse à une question de Morse et Hedlund

Tome 56, nº 7 (2006), p. 2249-2270.

<http://aif.cedram.org/item?id=AIF_2006__56_7_2249_0>

Ann. Inst. Fourier, Grenoble
56, 7 (2006) 2249-2270

FONCTIONS DE RÉCURRENCE DES SUITES D'ARNOUX-RAUZY ET RÉPONSE À UNE QUESTION DE MORSE ET HEDLUND

par Julien CASSAIGNE & Nataliya CHEKHOVA

RÉSUMÉ. — La fonction de récurrence $R(n)$ d'une suite symbolique compte au bout de combien de temps on voit tous les mots de longueur n. Nous la calculons explicitement pour les suites d'Arnoux-Rauzy, définies par des conditions combinatoires qui en font une généralisation naturelle des suites sturmiennes. Puis nous répondons à une question de Morse et Hedlund (1940) en montrant que $\frac{R(n)}{n}$ ne peut avoir une limite finie pour aucune suite non ultimement périodique.

ABSTRACT. — The recurrence function $R(n)$ of a symbolic sequence counts how long one has to wait to see every word of length n. We compute it explicitly for the Arnoux-Rauzy sequences, which are defined by combinatorial conditions making them a natural generalization of the Sturmian sequences. We then answer a question of Morse and Hedlund (1940) by showing that $\frac{R(n)}{n}$ cannot have a finite limit for any non-eventually periodic sequence.

Étant donnée une suite symbolique sur un alphabet fini, la *fonction de récurrence* compte la longueur minimale $R(n)$ telle que tout facteur de longueur $R(n)$ de la suite contienne tout facteur de longueur n. Cette fonction, qui est finie pour les suites *uniformément récurrentes*, est un des éléments qui mesurent le désordre, ou le caractère aléatoire (ou non) d'une suite. Elle a été introduite en 1940 par Morse et Hedlund [8], qui la calculent pour des suites particulières, les *suites sturmiennes*, codages naturels des rotations du tore $\mathbb{T}^1$; le calcul de la fonction de récurrence d'une suite sturmienne fait intervenir le développement en fraction continue de l'angle de la rotation associée. À la fin de leur article, Hedlund et Morse posent la question du calcul de $R(n)$ pour d'autres suites, et demandent s'il existe

Mots-clés : dynamique symbolique, combinatoire des mots, mot infini, fonction de récurrence, suite d'Arnoux-Rauzy, graphe de Rauzy, facteur bispécial, mot singulier, mot de retour.
Classification math. : 37B20, 37B10, 68R15.

une suite symbolique, non ultimement périodique, pour laquelle $\frac{R(n)}{n}$ a une limite finie.

La fonction de récurrence a fait l'objet d'études récentes par Cassaigne [5] et Durand, Host et Skau [7], montrant en particulier l'intérêt, sur le plan de la dynamique symbolique, des suites pour lesquelles $R(n)$ est à croissance linéaire. Toutefois, il semble difficile de calculer explicitement $R(n)$ pour une suite donnée.

Dans cet article, nous calculons explicitement la fonction de récurrence pour des généralisations naturelles des suites sturmiennes, les suites introduites dans [2] et dites depuis *suites d'Arnoux-Rauzy*. Celles-ci sont définies par certaines conditions combinatoires, et sont naturellement associées à des couples d'irrationnels pour lesquels elles fournissent un algorithme d'approximation simultanée ; nous déterminons ainsi lesquelles sont à récurrence linéaire, ce qui en fait l'analogue des suites sturmiennes associées aux irrationnels à quotients partiels bornés.

En utilisant des techniques développées par Alessandri [1] et Cassaigne [3, 4], nous montrons ensuite que pour aucune suite non ultimement périodique $\frac{R(n)}{n}$ n'a de limite finie (ou bien cette limite est infinie, ou bien cette quantité n'a pas de limite), répondant négativement à la question de Morse et Hedlund.

1. Préliminaires

1.1. Dynamique symbolique et langages formels

Soit Λ un alphabet fini ; on considère les suites unilatérales $u = (u_n)_{n\in\mathbb{N}}$ à valeurs dans Λ.

Un *mot* de longueur $|w| = h$ est une suite finie $w = w_1 \dots w_h$ d'éléments de Λ. La concaténation de deux mots v et w se note vw. Les notions de *préfixe* et *suffixe* sont définies naturellement. Le mot $w = w_1 \dots w_h$ *apparaît* au *rang* i dans la suite $u = (u_n)$ ou le mot $u = u_0 \dots u_s$ si $u_i = w_1$, …, $u_{i+h-1} = w_h$. Le mot w est *facteur* de u s'il apparaît dans u à au moins un rang.

Si w apparaît dans un mot ou une suite aux rangs i et $j > i$, la *distance* entre ces deux apparitions est le nombre $j - i$. Si ces deux apparitions sont *consécutives*, c'est-à-dire si w n'apparaît pas aux rangs $i+1$, …, $j-1$, cette distance est appelée *lacune* ou *temps de retour* de w dans u.

Une suite u est dite *récurrente* si chacun de ses facteurs y apparaît à une infinité de rangs. Si de plus chaque facteur de u y apparaît avec des lacunes bornées, la suite u est dite *minimale* ou *uniformément récurrente*.

Le *langage* $F_n(u)$ est l'ensemble des facteurs de longueur n de u ; le langage $F(u)$ est l'ensemble des facteurs de u.

La *complexité* d'une suite u est la fonction qui à tout entier $n \geqslant 0$ associe le nombre $p(n) = \#F_n(u)$ de facteurs de longueur n. On a toujours $p(0) = 1$.

Nous rappelons qu'une *suite sturmienne* est une suite vérifiant $p(n) = n+1$ pour tout $n \geqslant 0$.

DÉFINITION 1.1. — *Une suite $u = (u_n)_{n\in\mathbb{N}}$ sur l'alphabet $\Lambda = \{0, 1, 2\}$ est une suite d'Arnoux-Rauzy si elle a les quatre propriétés suivantes :*
- *elle est minimale ;*
- *elle vérifie $p(n) = 2n+1$ pour tout $n \geqslant 0$;*
- *chaque mot de $F_n(u)$ est préfixe d'un seul mot de $F_{n+1}(u)$, sauf un qui est préfixe de trois mots ;*
- *chaque mot de $F_n(u)$ est suffixe d'un seul mot de $F_{n+1}(u)$, sauf un qui est suffixe de trois mots.*

1.2. Description du graphe de Rauzy

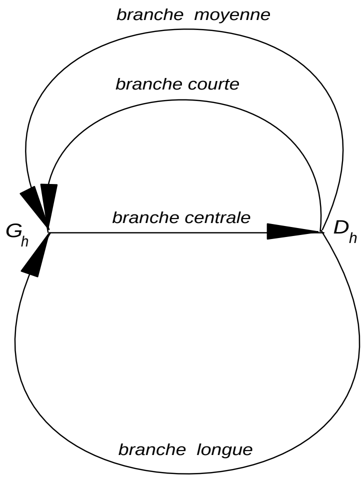
DÉFINITION 1.2. — *Pour une suite u et un entier h, on définit le* graphe de Rauzy *d'ordre h de u (ou graphe des mots de longueur h de u), Γ_h, de la manière suivante : les sommets sont les éléments de $F_h(u)$, et il y a une arête de w vers w' si w et w' apparaissent successivement dans u, c'est-à-dire si $w = av$ et $w' = vb$ avec avb facteur de u, pour des lettres a et b et un mot v de $F_{h-1}(u)$; on étiquette cette arête par $avb \in F_{h+1}(u)$, et l'ensemble des arêtes est identifié à $F_{h+1}(u)$.*

Considérons maintenant une suite d'Arnoux-Rauzy u ; ses graphes de Rauzy sont décrits dans [2]. Le graphe Γ_h comporte un unique sommet D_h *triprolongeable à droite* (trois arêtes D_h0, D_h1 et D_h2 le quittent) et un unique sommet G_h *triprolongeable à gauche* (trois arêtes $0G_h$, $1G_h$ et $2G_h$ y arrivent) ; D_h et G_h peuvent être confondus. Pour tout autre point, il y a une seule arête entrante et une seule arête sortante.

On partitionne les sommets et les arêtes de Γ_h en quatre *branches* ; une branche est un ensemble de sommets et d'arêtes, et sa *longueur* est le nombre d'arêtes qu'elle contient.
- La *branche centrale* comprend le sommet G_h, l'unique arête sortant de G_h, le sommet vers lequel va cette arête, l'unique arête sortant de ce sommet, et ainsi de suite jusqu'au sommet D_h inclus ; si $D_h = G_h$, la branche centrale est réduite au sommet $D_h = G_h$.

– Pour $i = 0, 1, 2$, la branche i comprend l'arête $D_h i$, le sommet vers lequel va cette arête, l'unique arête sortant de ce sommet, et ainsi de suite jusqu'au sommet G_h exclu ; si l'arête $D_h i$ va vers G_h, la branche i est réduite à l'arête $D_h i$.

Nous verrons que, sauf pour les premières valeurs de h, les longueurs des trois branches 0, 1, 2 sont toujours différentes, et nous les appellerons respectivement *branche courte*, *branche moyenne* et *branche longue* (voir figure) (pour les premières valeurs de h, on choisira arbitrairement les appellations si des branches sont de même longueur). Le *circuit court* C_h (resp. *moyen* M_h et *long* L_h) part de G_h et est composé de la branche centrale puis de la branche courte (resp. moyenne et longue). Pour éviter toute confusion avec la longueur des mots, nous noterons $\|C_h\|$, $\|M_h\|$ et $\|L_h\|$ la longueur de ces circuits.



L'évolution des graphes Γ_h quand h varie est décrite dans [2]. Les sommets de Γ_{h+1} sont les arêtes de Γ_h ; si $D_h \neq G_h$, les sommets d'une branche de Γ_{h+1} sont les arêtes de la même branche de Γ_h (il y a *fente* d'une arête). De ce fait, la longueur de la branche centrale diminue de 1 quand on passe de h à $h+1$.

Donc, pour une infinité de h, la branche centrale de Γ_h est de longueur 0, c'est-à-dire réduite à un sommet ; on a donc $G_h = D_h$, et on dit qu'il y a *éclatement*.

Les différents types d'éclatement sont décrits dans [2], voir en particulier la figure 4, mais nous introduisons ici une nouvelle terminologie ; nous disons qu'il y a un éclatement *renversant* (noté E_L) si les sommets de la branche centrale de Γ_{h+1} sont les arêtes de la branche longue de Γ_h ; il y a un éclatement *moyen* (E_M) si les sommets de la branche centrale de Γ_{h+1} sont les arêtes de la branche moyenne de Γ_h ; il y a un éclatement *court* (E_C) si les sommets de la branche centrale de Γ_{h+1} sont les arêtes de la branche courte de Γ_h ; dans les trois cas, la branche courte de Γ_{h+1} est réduite à une arête.

Si l'on énumère successivement les mots de longueur h apparaissant dans la suite u au rang i pour $i = 0, \dots n, \dots$, on obtient un chemin infini γ_h dans le graphe Γ_h, composé d'une succession de circuits courts, moyens et longs (le premier étant éventuellement tronqué). Aussi γ_h peut être vu comme une suite à valeurs dans l'alphabet $\Lambda_h = \{C_h, M_h, L_h\}$. Si en h il y a fente, γ_h se déduit de γ_{h+1} en remplaçant C_{h+1} par C_h, M_{h+1} par M_h et L_{h+1} par L_h ; s'il y a éclatement court, γ_h se déduit de γ_{h+1} en remplaçant C_{h+1} par C_h, M_{h+1} par C_hM_h et L_{h+1} par C_hL_h (et en effaçant éventuellement le C_h initial) ; s'il y a éclatement moyen, γ_h se déduit de γ_{h+1} en remplaçant C_{h+1} par M_h, M_{h+1} par M_hC_h et L_{h+1} par M_hL_h (et en effaçant éventuellement le M_h initial) ; s'il y a éclatement renversant, γ_h se déduit de γ_{h+1} en remplaçant C_{h+1} par L_h, M_{h+1} par L_hC_h et L_{h+1} par L_hM_h (et en effaçant éventuellement le L_h initial).

Dans toute la suite, si, par abus de langage, nous parlons d'un *mot* de la branche centrale (resp. courte, moyenne, longue), il s'agira toujours d'un *sommet* de cette branche.

On note e_p l'ordre du graphe où se produit le p-ième éclatement, avec par convention $e_0 = 0$ et $e_1 = 1$. On note c_p, m_p, ℓ_p la longueur des circuits court, moyen et long (respectivement) *après* le p-ième éclatement, c'est-à-dire

$$c_p = \|C_{e_p+1}\| = \|C_{e_{p+1}}\|,$$
$$m_p = \|M_{e_p+1}\| = \|M_{e_{p+1}}\|$$

et

$$\ell_p = \|L_{e_p+1}\| = \|L_{e_{p+1}}\|.$$

On a alors $e_{p+1} = e_p + c_p$. On note λ_i le numéro du i-ème éclatement renversant, avec par convention $\lambda_0 = 0$. Le premier éclatement non court est toujours considéré comme renversant (ce qui est cohérent d'après le lemme 1.4 ci-dessous) et il s'agit donc du λ_1-ième éclatement.

La suite des éclatements (renversants, courts ou moyens) détermine explicitement, à une permutation des lettres près, le langage de la suite d'Arnoux-Rauzy concernée, et détermine, au moyen d'un algorithme d'approximation simultanée, un couple d'irrationnels (α, β) ; voir [2] pour plus de détails.

Une suite d'Arnoux-Rauzy particulière est la *suite de Tribonacci*, voir par exemple [10] et [6], définie comme le *point fixe* de la substitution $0 \to 01$, $1 \to 02$, $2 \to 0$. Dans ce cas particulier, tous les éclatements sont des éclatements renversants E_L.

1.3. Longueurs des circuits

LEMME 1.3. — *Dans Γ_1, les longueurs des circuits sont $c_0 = 1$, $m_0 = 2$ et $\ell_0 = 2$. Ensuite; les longueurs des circuits court, moyen et long sont inchangées entre Γ_h et Γ_{h+1} si en h il y a fente. Si en $h = e_p$ il y a éclatement :*

– *si l'éclatement est un E_L, on a :*

$$c_p = \ell_{p-1},\ m_p = \ell_{p-1} + c_{p-1} \ \text{ et } \ \ell_p = \ell_{p-1} + m_{p-1}\,,$$

– *si l'éclatement est un E_M, on a :*

$$c_p = m_{p-1},\ m_p = m_{p-1} + c_{p-1} \ \text{ et } \ \ell_p = \ell_{p-1} + m_{p-1}\,,$$

– *si l'éclatement est un E_C, on a :*

$$c_p = c_{p-1},\ m_p = m_{p-1} + c_{p-1} \ \text{ et } \ \ell_p = \ell_{p-1} + c_{p-1}\,.$$

Démonstration. — Immédiate d'après la description précédente. □

LEMME 1.4. — *Pour les suites d'Arnoux-Rauzy,*

– *les trois circuits de Γ_h ont des longueurs différentes sauf pour les premières valeurs de h, $h \leqslant \lambda_1$,*

– *il y a une infinité d'éclatements renversants.*

Démonstration. — Tant que les éclatements sont des E_C, les deux circuits les plus longs ont même longueur : pour $0 \leqslant p < \lambda_1$, on a $e_p = p$, $c_p = 1$ et $m_p = \ell_p = p + 2$. Dès qu'il y a un éclatement autre que E_C, ces longueurs deviennent différentes : $e_{\lambda_1} = \lambda_1$, $c_{\lambda_1} = \lambda_1 + 1$, $m_{\lambda_1} = \lambda_1 + 2$, $\ell_{\lambda_1} = 2\lambda_1 + 2$. Et en vertu des formules du lemme 1.3, dès que les longueurs des trois circuits deviennent différentes, elles le resteront après tous les éclatements ultérieurs. Notons qu'il n'est pas possible que tous les éclatements soient des E_C, car si c'était le cas la suite serait en fait triviale, $0000\ldots$, $1111\ldots$ ou $2222\ldots$, et ce ne sont pas des suites d'Arnoux-Rauzy.

Si, pour $h > h_0$, on n'a plus que des éclatements E_M ou E_C, on vérifie que les mots de la branche longue de Γ_{h_0} n'apparaissent plus ultérieurement que dans les circuits longs, donc apparaissent dans u avec des lacunes supérieures à la longueur des autres circuits, ce qui contredit la minimalité. □

2. Fonction de récurrence

Définition 2.1. — *La fonction de récurrence d'une suite u est la fonction, notée $R(n)$, qui à un entier $n \geqslant 1$ associe le plus petit entier r (ou $+\infty$, si un tel r n'existe pas) tel que tout mot de $F_n(u)$ apparaisse dans tout mot de $F_r(u)$.*

Un nombre t est un *temps de retour* d'un mot w dans u si t est la distance entre deux apparitions consécutives de w dans u. On appelle *temps de retour maximal*, noté $t(w)$, d'un mot w dans u, la distance maximale entre deux apparitions consécutives de w dans u ($t(w) = +\infty$ si cette distance est non bornée), et on note $r(n)$ le maximum des $t(w)$ pour tous les mots $w \in F_n(u)$. Si u est minimale, $r(n)$ et $t(w)$ sont toujours finis.

On appelle *mot singulier* tout facteur $w = xvy$ de u, tel que x et y soient des lettres et qu'il existe $x' \neq x$, $y' \neq y$ telles que $x'vy$ et xvy' apparaissent dans u. Par convention, tout mot de longueur 1 est aussi singulier.

L'article [5] précise une méthode générale, essentiellement due à Morse et Hedlund [8], pour calculer la fonction de récurrence d'une suite. La première partie du lemme suivant figure déjà dans la thèse de Mouline [9] ; la démonstration que nous donnons provient de [5], propositions 2 et 3.

Lemme 2.2. — *Les fonctions R et r sont liées, pour tout $n \geqslant 1$, par*

$$R(n) = r(n) + n - 1 .$$

Pour $n \geqslant 1$,

$$r(n) = \max_{w \in S_n(u)} t(w) ,$$

où $S_n(u)$ est l'ensemble des mots singuliers de u de longueur inférieure ou égale à n.

Démonstration [5]. — Soit w un facteur de longueur n de u tel que $t(w) = r(n)$. Si $m \geqslant r(n) + n - 1$, tout mot de $F_m(u)$ doit contenir tout mot de $F_n(u)$; et, si $m < r(n) + n - 1$, on trouve un mot de $F_m(u)$ qui ne contient pas w, d'où le premier résultat.

On pose $r_0(n) = \max_{w \in S_n(u)} t(w)$; comme le temps de retour maximal d'un mot est au moins égal à celui de ses préfixes, $r_0(n) \leqslant r(n)$. Si

$$r(n) > r(n-1),$$

soit $w = xvy$ de longueur n tel que $r(n) = t(w)$. Comme

$$t(xv) \leqslant r(n-1) < r(n) = t(w),$$

il existe des apparitions de xv suivies d'une autre lettre que y, et de même il existe des apparitions de vy précédées d'une autre lettre que x ; w est donc un mot singulier et donc $r(n) \leqslant r_0(n)$. Il en est de même si $n = 1$. Enfin, si $r(n) = r(n-1)$, on choisit le plus petit $n' \geqslant 1$ tel que

$$r(n) = r(n'),$$

et

$$r(n) = r(n') = r_0(n') \leqslant r_0(n),$$

d'où le deuxième résultat. □

Lemme 2.3. — *Soit u une suite d'Arnoux-Rauzy. Alors tout mot singulier de u autre que les lettres est de longueur $e_{p-1} + 2$ pour un certain $p \geqslant 1$, et son temps de retour maximal r_p est :*

- *si le p-ième éclatement est un E_C, $r_p = \ell_p$,*
- *si le p-ième éclatement est un E_L ou un E_M et si le $a(p)$-ième éclatement, défini comme le premier éclatement E_L ou E_M après le p-ième éclatement, est un E_M, $r_p = \ell_{a(p)}$,*
- *si le p-ième éclatement est un E_L ou un E_M et si le $a(p)$-ième éclatement est un E_L, $r_p = \ell_{b(p)}$, où le $b(p)$-ième éclatement est le premier éclatement E_L après le $a(p)$-ième éclatement.*

Les mots singuliers de longueur 1, c'est-à-dire les lettres de l'alphabet, ont pour temps de retour maximal 2, ℓ_{λ_1} et $r_0 = \ell_{\lambda_2}$.

Démonstration. — Par définition, w est un mot singulier de longueur $n \geqslant 2$ si et seulement s'il est l'unique arête d'une branche (autre que la branche centrale) de Γ_{n-1}, ce qui ne peut se produire que juste après un éclatement : il existe $p \geqslant 1$ tel que $n-1 = e_{p-1}+1$. Comme seule la branche courte de $\Gamma_{e_{p-1}+1}$ est de longueur 1, il n'y a qu'un seul mot singulier de longueur n.

Dans Γ_{n-1}, le mot singulier w est une arête du circuit court $C = C_{n-1}$, et n'est une arête d'aucun des autres circuits, $M = M_{n-1}$ et $L = L_{n-1}$.

La suite u correspond à un chemin infini $\gamma = \gamma_{n-1}$ dans le graphe Γ_{n-1} ; ce chemin est une suite de circuits C, L et M, et on peut donc le considérer comme une suite infinie à valeurs dans l'alphabet $\Lambda' = \{C, L, M\}$; deux apparitions consécutives de w dans u correspondent à deux apparitions consécutives de C dans γ. Nous allons donc étudier ces apparitions dans la suite γ.

Si l'éclatement suivant, l'éclatement p, est un E_C, alors le chemin γ est une concaténation de mots (sur Λ') C, CM, CL ; on trouve C dans tous ces mots, et on voit toujours C après CM ou CL ; le plus long temps de retour possible pour le mot w est donc la longueur du circuit CL, soit

$$t(w) = \|CL\| = c_{p-1} + \ell_{p-1} = \ell_p.$$

Si l'éclatement p est un E_L ou E_M, γ est une concaténation respectivement de $C' = L$, $M' = LC$, $L' = LM$ ou de $C' = M$, $M' = MC$, $L' = ML$. Dans les deux cas, deux apparitions consécutives de w dans u correspondent à deux apparitions consécutives de C dans γ, donc de M'.

Le prochain éclatement E_L ou E_M est le $a(p)$-ième éclatement ; soit $h = e_{a(p)}$. La suite u correspond à un chemin infini γ_h dans Γ_h. Comme tous les éclatements entre le p-ième et le $a(p)$-ième sont des E_C, γ se déduit de γ_h en remplaçant C_h par C', M_h par $C'^{a(p)-p-1}M'$ et L_h par $C'^{a(p)-p-1}L'$, le premier mot pouvant être tronqué à gauche. Deux apparitions consécutives de w dans u correspondent à deux apparitions consécutives de M' dans γ, donc à deux apparitions consécutives de M_h dans γ_h.

Si l'éclatement $a(p)$ est un E_M, γ_h est une concaténation de mots M_h, M_hC_h, M_hL_h ; on trouve M_h dans tous ces mots, et on voit toujours M_h après M_hC_h ou M_hL_h ; le plus long temps de retour possible pour le mot w est donc la longueur du circuit M_hL_h, soit $t(w) = \|M_hL_h\| = \ell_{a(p)}$.

Si l'éclatement $a(p)$ est un E_L, γ_h est une concaténation de $C'' = L_h$, $M'' = L_hC_h$, $L'' = L_hM_h$, et deux apparitions consécutives de w dans u correspondent à deux apparitions consécutives de M_h dans γ_h, donc de L''. On regarde alors l'éclatement $b(p)$, le premier éclatement renversant après le $a(p)$-ième éclatement (cet éclatement existe grâce au lemme 1.4), et le chemin γ_k avec $k = e_{b(p)}$. Les éclatements précédents étant des E_C ou E_M, γ_h se déduit de γ_k en remplaçant C_k, M_k et L_k par des mots C''', M''' et L''' respectivement (le premier mot pouvant être tronqué à gauche), où la seule occurrence de L'' et donc de M_h est dans L'''. Deux apparitions consécutives de w dans u correspondent donc à deux apparitions consécutives de L_k dans γ_k. Puisque l'éclatement $b(p)$ est renversant, γ_k est une concaténation de L_k, L_kC_k, L_kM_k, et on trouve toujours L_k après L_kC_k ou L_kM_k ; le plus long temps de retour possible pour le mot w est donc $t(w) = \|L_kM_k\| = \ell_{b(p)}$.

Parmi les lettres de l'alphabet, l'une est triprolongeable à gauche et à droite : par exemple $0 = G_1 = D_1$. La suite u se factorise alors sur

$\{0, 01, 02\}$, et on a donc $t(0) = 2$. Après un certain nombre d'éclatements courts, on arrive à Γ_{λ_1} dont les circuits sont étiquetés par 0, $0^{\lambda_1}1$, $0^{\lambda_1}2$; l'éclatement suivant est renversant, et donne (en prenant $0^{\lambda_1}1$ comme circuit long) les circuits $0^{\lambda_1}1$, $0^{\lambda_1}10$, $0^{\lambda_1}10^{\lambda_1}2$, donc $t(1) = 2\lambda_1 + 2 = \ell_{\lambda_1}$. Enfin, pour que 2 apparaisse dans tous les circuits, il faut attendre l'éclatement renversant suivant, et $t(2) = \ell_{\lambda_2}$. □

Proposition 2.4. — *Soient u une suite d'Arnoux-Rauzy et $R(n)$ sa fonction de récurrence. On note e_p la longueur des mots au moment du p-ième éclatement, ℓ_p la longueur du circuit long après cet éclatement, λ_i le numéro du i-ème éclatement renversant, et μ_i le numéro du dernier éclatement moyen ou renversant précédant l'éclatement λ_i (avec par convention $e_{-1} = -1$, $e_0 = 0$ et $\lambda_0 = \mu_1 = 0$). Alors, pour tout $i \geqslant 1$, et pour tout n compris entre $e_{\mu_i-1} + 2$ et $e_{\mu_{i+1}-1} + 1$, on a $R(n) = n - 1 + \ell_{\lambda_{i+1}}$.*

Démonstration. — Montrons pour commencer que seuls les mots singuliers correspondant au troisième cas du lemme 2.3, et ceux de longueur 1, ont une influence sur la fonction de récurrence ; ces mots singuliers sont précisément ceux dont la longueur est de la forme $e_{\mu_i-1} + 2$ (ce qui inclut les lettres, pour $i = 1$).

Soient w un mot singulier qui n'est pas de ce type, et $p \geqslant 1$ tel que $|w| = e_{p-1} + 2$. Le temps de retour de w est, d'après le lemme 2.3, soit ℓ_p, soit $\ell_{a(p)}$, donc $t(w) \leqslant \ell_{a(p)}$ puisque la suite (ℓ_i) des longueurs des circuits longs est strictement croissante. Si $p < \lambda_1$, alors w est dans le premier cas du lemme 2.3 et $t(w) = \ell_p < \ell_{\lambda_2} = t(w')$, où w' est la lettre qui maximise le temps de retour.

Si $p \geqslant \lambda_1$, soit $p' \leqslant p$ maximal tel que le p'-ième éclatement soit renversant (éventuellement $p' = p$, et en tout cas $p' \geqslant \lambda_1$). Si $p' = \lambda_1$, alors $a(p) < \lambda_2$ et on a $t(w) \leqslant \ell_{a(p)} < \ell_{\lambda_2} = t(w')$, où w' est la lettre qui maximise le temps de retour, comme ci-dessus. Si $p' > \lambda_1$, soit q le numéro du dernier éclatement E_M ou E_L précédant le p'-ième éclatement, et w' le mot singulier de longueur $e_{q-1} + 2$: ce mot singulier w' est dans le troisième cas du lemme 2.3, et $b(q) \geqslant a(p)$, donc $t(w) \leqslant \ell_{a(p)} \leqslant \ell_{b(q)} = t(w')$.

On a donc, pour tout $n \geqslant 1$,

$$r(n) \;=\; \max_{w \in S'_n} t(w)\,,$$

où S'_n est l'ensemble des mots singuliers de longueur au plus n qui sont soit des lettres, soit dans le troisième cas du lemme 2.3.

Si $e_{\mu_i-1} + 2 \leqslant n \leqslant e_{\mu_{i+1}-1} + 1$, on a $\max_{w \in S'_n} t(w) = \ell_{b(\mu_i)} = \ell\lambda_{i+1}$, d'où le résultat. □

Ces formules permettent de calculer explicitement la fonction $R(n)$ pour une suite d'Arnoux-Rauzy définie par sa suite d'éclatements ; par exemple, on retrouve immédiatement le cas particulier de Tribonacci [6].

La proposition précédente implique :

Corollaire 2.5. — *Pour une suite d'Arnoux-Rauzy,*

$$\limsup_{n\to+\infty} \frac{R(n)}{n} < +\infty$$

si et seulement si le nombre d'éclatements entre deux E_L consécutifs, $\lambda_{i+1}-\lambda_i$, est borné.

Démonstration. — Supposons d'abord que $\lambda_{i+1} - \lambda_i \leqslant A$ pour tout $i \geqslant 0$. On a alors $\lambda_{i+1} - \mu_i \leqslant 2A$. Pour tout p, $e_{p+1} = e_p + c_p$ avec $c_p \leqslant \ell p - 1 = \|L_{e_p}\| \leqslant 2e_p$ (la somme des longueurs des branches de Γ_{e_p} est $2e_p + 2$, puisque u est de complexité $2n+1$, donc la branche longue est au plus de longueur $2e_p$), donc $e_{p+1} \leqslant 3e_p$, et $e_{\lambda_{i+1}+1}/e_{\mu_i-1} \leqslant 3^{2A+2}$. Si $e_{\mu_i-1} + 2 \leqslant n \leqslant e_{\mu_{i+1}-1} + 1$, $R(n) = n - 1 + \ell\lambda_{i+1}$ par la proposition 2.4, donc $R(n)/n \leqslant 1 + \ell\lambda_{i+1}/e_{\mu_i-1} \leqslant 1 + 2e_{\lambda_{i+1}+1}/e_{\mu_i-1} \leqslant 1 + 2.3^{2A+2}$.

Dans le cas contraire, pour tout A fixé, il existe i tel que $\lambda_{i+1}-\lambda_i > A$: le λ_i-ème éclatement, renversant, est suivi de A éclatements non renversants. Soit $n = e_{\mu_i-1} + 2$: par la proposition 2.4, $R(n) = n - 1 + \ell\lambda_{i+1}$. Comme l'éclatement λ_i est renversant, on a, par le lemme 1.3, $c_{\lambda_i} = \ell\lambda_i - 1 = \|L_{e_{\lambda_i}}\| \geqslant (2e_{\lambda_i} + 4)/3$ (la somme des longueurs des branches de $\Gamma_{e_{\lambda_i}}$ est $2e_{\lambda_i} + 2$, et la branche centrale est de longueur 0 tandis que la branche longue est strictement plus longue que les deux autres). Après les A éclatements suivants, on a $\ell\lambda_i + A \geqslant \ell\lambda_i + Ac_{\lambda_i}$ (avec égalité si tous ces éclatements sont courts), donc $\ell\lambda_i + A \geqslant (2e_{\lambda_i} + 4)A/3$. Ainsi

$$\frac{R(n)}{n} \geqslant \frac{\ell\lambda_{i+1}}{e_{\mu_i-1}+2} \geqslant \frac{\ell\lambda_i + A}{e_{\lambda_i}+2} \geqslant \frac{2A}{3} \ :$$

$R(n)/n$ n'est pas borné. □

En adoptant la terminologie de [7], conçue par analogie avec le cas des suites sturmiennes, on pourra appeler *suite d'Arnoux-Rauzy à quotients partiels bornés* toute suite d'Arnoux-Rauzy telle que le nombre d'éclatements entre deux E_L consécutifs est borné.

3. Réponse à une question de Morse et Hedlund

Le résultat suivant ne concerne pas particulièrement les suites d'Arnoux-Rauzy, mais répond à une question datant de 1940 [8] : existe-t-il des suites non triviales pour lesquelles $\frac{R(n)}{n}$ a une limite finie ? Pour cela, *nous supposons d'abord que l'alphabet* Λ *a deux lettres* (qu'on notera 0 et 1), et montrons une série de lemmes, dont nous déduirons le résultat pour les suites sur deux lettres. Nous l'étendrons ensuite à k lettres par un procédé de codage.

Un facteur v de u est dit *spécial* à droite (resp. à gauche) s'il existe des lettres x et y ($x \neq y$) telles que les mots vx et vy (resp. xv et yv) sont aussi facteurs de u.

Pour tout n, on note $s(n)$ le nombre de facteurs spéciaux à droite de u de longueur n, qui est aussi le nombre de facteurs spéciaux à gauche de longueur n si la suite u est récurrente (ce qui est toujours vrai des suites qui nous intéressent ici). On a alors $s(n) = p(n+1) - p(n)$, voir [4].

Lemme 3.1. — *Si* $\limsup \frac{R(n)}{n} < +\infty$*, pour tout* n *la suite* u *a un nombre borné par une constante* k *de facteurs spéciaux à droite de longueur* n*.*

Démonstration. — Si l'hypothèse est vérifiée, alors il existe a tel que $R(n) \leqslant an$; comme tous les $p(n)$ facteurs de longueur n de u apparaissent dans un même mot de longueur $R(n)$, on a $p(n) \leqslant R(n) - n + 1 \leqslant an$; en vertu d'un résultat de Cassaigne [3], la quantité $s(n) = p(n+1) - p(n)$ est donc bornée par une constante k ; or $s(n)$ n'est autre que le nombre de facteurs spéciaux à droite de longueur n. □

Définition 3.2. —

- *On appelle mots de retour d'un facteur* w *de* u *les mots* $u_{s+q} \dots u_{t+q-1}$*, où* $u_s \dots u_{s+q-1}$ *et* $u_t \dots u_{t+q-1}$ *sont deux apparitions consécutives de* w*.*
- *On appelle chemin élémentaire d'ordre* n *un mot* $c' = u_{s+n} \dots u_{t+n-1}$*, où* $u_s \dots u_{s+n-1} = d'$ *est spécial à droite,* $u_t \dots u_{t+n-1} = d''$ *est spécial à droite, et il n'y a pas d'apparition de facteur spécial à droite de longueur* n *entre* $s+1$ *et* $t-1$*.*

On peut aussi voir un chemin élémentaire comme mot de retour de l'ensemble des facteurs spéciaux à droite.

Lemme 3.3. — *Le nombre de chemins élémentaires est exactement* $2s(n)$*.*

Démonstration. — On sait qu'il y a exactement $s(n)$ facteurs spéciaux à droite, et de chacun partent exactement deux chemins élémentaires. □

Les lemmes 3.4 et 3.5 sont démontrés indépendamment (avec des meilleures constantes) dans [7].

LEMME 3.4. — *Si* $\limsup \frac{R(n)}{n} < +\infty$ *et si la suite* u *n'est pas ultimement périodique, il existe* $0 < b < a$ *tels que, si* w *est un facteur de longueur* n*, tous ses temps de retour sont compris entre* bn *et* an*.*

Démonstration. — On sait déjà que $R(n) \leqslant an$ pour tout $n \geqslant 1$ et donc $r(n) \leqslant an$, où $a = \sup_n \frac{R(n)}{n}$. Il nous reste à minorer le plus petit temps de retour. Comme u est non ultimement périodique, $p(n) \geqslant n+1$ pour tout n, d'après [8]. Soit $m = \left\lfloor \frac{n}{a} \right\rfloor$.

Il y a au moins $m+1$ facteurs de longueur m, dont au plus k sont spéciaux à droite. D'après le lemme 3.3, il y a donc au plus $2k$ chemins élémentaires. Soit $(w_0, \dots, w_\ell)$ le plus long de ces chemins. $w_1, \dots, w_\ell$ sont alors ℓ facteurs différents de longueur m tels que

- $\ell \geqslant \frac{m+1}{2k}$,
- la j-ème lettre de w_{i+1} est la $(j+1)$-ième lettre de w_i, pour $1 \leqslant j \leqslant m-1$, $1 \leqslant i \leqslant \ell - 1$,
- les w_i ne sont pas spéciaux à droite pour $1 \leqslant i \leqslant \ell - 1$, tandis que w_ℓ est spécial à droite.

Par conséquent, toute apparition de w_1 est suivie d'apparitions de $w_2, \dots$, w_ℓ, et tout temps de retour de w_1 est au moins ℓ.

Mais tout mot de longueur n contient w_1, car $R(\left\lfloor \frac{n}{a} \right\rfloor) \leqslant a \left\lfloor \frac{n}{a} \right\rfloor \leqslant n$; donc tout mot de longueur n a tous ses temps de retour supérieurs à $\ell \geqslant \frac{m+1}{2k}$. Comme $m = \left\lfloor \frac{n}{a} \right\rfloor > \frac{n}{a} - 1$, on a $\ell > \frac{n}{2ka}$; on peut donc prendre $b = \frac{1}{2ka}$. □

LEMME 3.5. — *On suppose que* $\limsup \frac{R(n)}{n} < +\infty$ *et que la suite* u *n'est pas ultimement périodique. Alors le nombre de mots de retour différents de tout facteur spécial à droite est borné par une constante* B*.*

Démonstration. — Soient d un facteur spécial à droite de longueur n et c un mot de retour de d. On coupe c en chemins élémentaires. On a vu dans le lemme 3.3 qu'il y a exactement $2s(n)$ chemins élémentaires et par le lemme 3.1, on a $2s(n) \leqslant 2k$. Nous qualifions de *longs* ceux dont la longueur est au moins $\frac{bn}{k}$, de *courts* les autres. Le chemin c est une concaténation de chemins élémentaires $c'_1 \dots c'_l$, comprenant $\ell 1$ chemins longs et au plus $\ell 1 + 1$ plages de chemins courts consécutifs. On a $\ell 1 \leqslant \frac{ak}{b}$, sinon c aurait une longueur supérieure à an ; chaque plage de chemins courts comprend au plus k chemins, sinon deux de ces chemins partiraient du même facteur spécial à droite, qui aurait de ce fait un temps de retour inférieur à bn. Donc c est une concaténation d'au plus $A = \frac{ak}{b} + (\frac{ak}{b} + 1)k$ chemins élémentaires, et on ne peut construire ainsi qu'au plus $B = (2k+1)^A$ tels mots. □

Les deux lemmes suivants (lemmes 3.6 et 3.7) concernent toutes les suites symboliques récurrentes sur deux lettres, sans condition sur $R(n)$. Le premier se trouve dans [4], et sa démonstration est immédiate ; le second se trouve dans [3] et [1], et nous reproduisons intégralement sa démonstration. Nous rappelons que v est un *facteur bispécial* (c'est-à-dire spécial dans les deux directions à la fois) s'il existe des lettres $x \neq x'$ et $y \neq y'$ telles que vy, vy', xv et $x'v$ sont facteurs de u. Le mot v est un facteur bispécial *strict* si xvy, $x'vy$, xvy', $x'vy'$ sont facteurs, *ordinaire* si xvy, xvy' et $x'vy'$ sont facteurs, *faible* si seuls xvy et $x'vy'$ sont facteurs. L'ensemble des facteurs spéciaux à gauche forme un arbre infini, avec une arête de v à vx si x est une lettre et vx est spécial à gauche. Dans cet arbre, les facteurs bispéciaux stricts indiquent les embranchements et les facteurs bispéciaux faibles les extrémités des branches.

Pour tout n, on note $bs(n)$ le nombre de facteurs bispéciaux stricts de longueur n, et $bf(n)$ le nombre de facteurs bispéciaux faibles de longueur n.

Lemme 3.6. — *Pour tout n, $s(n+1) - s(n) = bs(n) - bf(n)$.*

Démonstration. — $s(n)$ augmente de un quand il y a un embranchement dans l'arbre des facteurs spéciaux à droite, et diminue de un quand il y a une extrémité. □

Lemme 3.7. — *Soient $n_1 < n_2$ deux entiers. Le nombre de facteurs bispéciaux stricts dont la longueur est comprise entre n_1 et $n_2 - 1$ est majoré par*

$$\sum_{n=n_1}^{n_2-1} bs(n) \leqslant s(n_1)\left(s(n_1) + s(n_2) + 2(p(n_2+1) - p(n_1+1))\frac{s(n_1)}{n_1}\right).$$

Démonstration [3, 1]. — L'ensemble des facteurs spéciaux à gauche dont la longueur est comprise entre n_1 et n_2 est une forêt de $s(n_1)$ arbres binaires finis dont les branches ont une longueur inférieure ou égale à $n_2 - n_1$ (on met une arête de w vers w' si $w' = wa$ où a est une lettre). Cette forêt contient

$$\sum_{n=n_1}^{n_2} s(n) = p(n_2+1) - p(n_1)$$

nœuds et

$$\sum_{n=n_1+1}^{n_2} s(n) = p(n_2+1) - p(n_1+1)$$

arêtes. Elle comporte $d = \sum_{n=n_1}^{n_2-1} bs(n)$ embranchements ; il y a $s(n_2)$ branches de longueur $n_2 - n_1$ et $h = \sum_{n=n_1}^{n_2-1} bf(n)$ branches qui se terminent avant d'atteindre cette longueur. De plus, d'après le lemme 3.6, on a $s(n_2) - s(n_1) = d - h$.

Nous allons construire une projection φ de cette forêt sur le graphe Γ_{n_1} comme suit : on envoie tout facteur spécial à gauche de longueur comprise entre n_1 et n_2 sur son suffixe de longueur n_1, et si une arête va de m à ma, on la projette sur l'arête de Γ_{n_1} qui va de $\varphi(m)$ à $\varphi(ma)$.

Pour tout sommet v (resp. toute arête e) de Γ_{n_1}, on note $\pi(v)$ (resp. $\pi(e)$) le nombre de nœuds (resp. d'arêtes) qui se projettent sur v (resp. e) ; cette quantité est appelée *poids* de v (resp. de e). La somme des poids des arêtes entrantes (resp. sortantes) de v sera notée $\pi_e(v)$ (resp. $\pi_s(v)$).

Pour tout sommet v de Γ_{n_1}, on note de plus :

- $c(v) = 1$ si v est facteur spécial à gauche et $c(v) = 0$ sinon,
- $t(v)$ le nombre de facteurs spéciaux à gauche de longueur n_2 qui ont v pour suffixe (c'est-à-dire le nombre des branches de la forêt de longueur $n_2 - n_1$ dont la projection se termine en v),
- $d(v)$ le nombre de facteurs bispéciaux stricts de longueur comprise entre n_1 et $n_2 - 1$ qui se projettent en v,
- $h(v)$ le nombre de facteurs bispéciaux faibles de longueur comprise entre n_1 et $n_2 - 1$ qui se projettent en v.

Remarquons que $h(v)$ et $d(v)$ ne sont non nuls que si v est spécial à droite, et alors $h(v) \leqslant d(v) + 1$ puisque l'ensemble des facteurs spéciaux à droite de longueur comprise entre n_1 et n_2 qui ont v pour suffixe forme un arbre binaire comportant $d(v)$ embranchements et au moins $h(v)$ feuilles.

Pour tout sommet v de Γ_{n_1}, on a :

$$\pi_s(v) - \pi_e(v) = c(v) + d(v) - h(v) - t(v)\,. \tag{3.1}$$

En effet, les seuls facteurs spéciaux à gauche qui se projettent en v et qui ne sont pas assortis d'exactement une arête entrante et une arête sortante sont :

- v lui-même s'il est facteur spécial à gauche puisque le nœud correspondant n'a pas d'arête entrante,
- ceux qui sont facteurs bispéciaux stricts puisqu'ils ont deux arêtes sortantes,
- ceux qui sont facteurs bispéciaux faibles et ceux qui sont de longueur maximale qui n'ont pas d'arêtes sortantes.

La combinaison de ces effets donne (3.1).

Soit maintenant S un sous-graphe de Γ_{n_1}, c'est-à-dire un graphe constitué de certains sommets de Γ_{n_1} et de certaines arêtes reliant ces sommets.

Une arête qui n'est pas dans S mais dont l'origine (resp. l'arrivée) est un sommet de S sera dite sortante de S (resp. entrante dans S). Notons alors :

- $\pi_e(S)$ le poids total des arêtes entrantes dans S,
- $\pi_s(S)$ le poids total des arêtes sortantes de S,
- $c(S)$ le nombre de facteurs spéciaux à gauche contenus dans S,
- $d(S)$ le nombre de facteurs bispéciaux stricts qui se projettent sur un sommet de S,
- $h(S)$ le nombre de facteurs bispéciaux faibles qui se projettent sur un sommet de S,
- $t(S)$ le nombre de branches dont la projection se termine dans S.

En sommant l'égalité précédente sur tous les sommets de S, on obtient :

$$\pi_s(S) - \pi_e(S) = c(S) + d(S) - h(S) - t(S)\,.$$

Mais comme $h(S) \leqslant d(S) + s(n_1)$ (il y a au plus $s(n_1)$ facteurs spéciaux à droite dans S), que $t(S) \leqslant s(n_2)$ (il n'y a que $s(n_2)$ branches de longueur maximale), et que $c(S) \geqslant 0$, il vient que :

$$\pi_e(S) - \pi_s(S) \leqslant s(n_1) + s(n_2)\,. \tag{3.2}$$

Les d facteurs bispéciaux stricts se projettent sur les $s(n_1)$ facteurs spéciaux à droite de longueur n_1. Il y a donc au moins un facteur spécial à droite v de longueur n_1 sur lequel se projettent au moins $\frac{d}{s(n_1)}$ facteurs bispéciaux stricts. Notons e_1 et e_2 les deux arêtes sortantes de v dans Γ_{n_1}. Leurs poids sont au moins égaux à $\frac{d}{s(n_1)}$.

Soit

$$f = \frac{1}{s(n_1)}\left(\frac{d}{s(n_1)} - s(n_1) - s(n_2)\right)\,.$$

Pour $i = 1, 2$, considérons le sous-graphe S_i constitué de tous les sommets qui peuvent être atteints à partir de v en empruntant d'abord e_i, puis des arêtes de poids au moins f, qui seront les arêtes de S_i.

Supposons que v n'est pas dans S_i (ce qui implique $f > 0$). Notons q le nombre de sommets de degré sortant 2 contenus dans S_i ; comme v n'est pas dans S_i, on a $q < s(n_1)$. Puisque S_i est connexe par construction, le nombre d'arêtes sortantes de S_i est au plus $q + 1$. Et ces arêtes sont de poids strictement inférieurs à f puisqu'elles ne sont pas dans S_i. Le poids total des arêtes sortantes vérifie donc $\pi_s(S_i) < (q+1)f \leqslant s(n_1)f$. Mais si v n'est pas dans S_i, e_i est une arête entrante dans S_i et donc $\pi_e(S_i) \geqslant \frac{d}{s(n_1)}$. Il vient alors que :

$$\pi_e(S_i) - \pi_s(S_i) > \frac{d}{s(n_1)} - s(n_1)f = s(n_1) + s(n_2)$$

ce qui est exclu par (3.2).

Donc v appartient à S_1 et à S_2. Il existe donc deux boucles allant de v à lui-même qui n'empruntent que des arêtes de poids au moins f ; une commençant par e_1 et l'autre par e_2. Soient $\ell 1$ et $\ell 2$ les longueurs de ces boucles. Les deux boucles correspondent à deux mots w_1 et w_2 de longueurs respectives $\ell 1$ et $\ell 2$ tels que vw_1 et vw_2 ont tous les deux v comme suffixe. De plus, w_1 et w_2 ne commencent pas par la même lettre (puisqu'une boucle commence par e_1 et l'autre par e_2). Les mots vw_1w_2 et vw_2w_1 ont pour suffixe v, et comme $|w_1w_2| = |w_2w_1|$ et $w_1w_2 \neq w_2w_1$, il vient que $|w_1w_2| > |v|$, c'est-à-dire que $\ell 1 + \ell 2 > n_1$. La réunion des deux boucles contient au moins $\max(\ell 1, \ell 2)$ arêtes, donc au moins $\frac{n_1}{2}$ arêtes. Il y a donc dans Γ_{n_1} au moins $\frac{n_1}{2}$ arêtes de poids au moins f. Le poids total du graphe étant $p(n_2+1)-p(n_1+1)$ (c'est le nombre d'arêtes de la forêt qu'on projette sur Γ_{n_1}), il vient donc :

$$p(n_2+1) - p(n_1+1) \geqslant \frac{n_1}{2} f\,,$$

c'est-à-dire :

$$p(n_2+1) - p(n_1+1) \geqslant \frac{n_1}{2s(n_1)} \left(\frac{d}{s(n_1)} - s(n_1) - s(n_2) \right)$$

d'où le résultat. □

Lemme 3.8. — *On suppose que* $\limsup \frac{R(n)}{n} < +\infty$ *et que la suite* u *n'est pas ultimement périodique. Soit* k *défini dans le lemme 3.1. Alors, pour tout* n*, et tout* $x \geqslant 1$*, le nombre de facteurs bispéciaux stricts de longueur comprise entre* n *et* xn *est au plus* $2xk^3$ *et le nombre de facteurs bispéciaux faibles de longueur comprise entre* n *et* xn *est au plus* $2xk^3 + k$*.*

Démonstration. — Le résultat est trivial si $k = 1$ ou $n = 1$. On suppose donc que $k \geqslant 2$ et $n \geqslant 2$. Pour les bispéciaux stricts, on applique le lemme 3.7 avec $n_1 = n$ et $n_2 = \lfloor xn \rfloor + 1$, sachant que $s(n) \leqslant k$ pour tout n et donc $p(n_2+1) - p(n_1+1) \leqslant k(n_2 - n_1)$:

$$\begin{aligned}\sum_{i=n_1}^{n_2-1} bs(i) \leqslant k \left(2k + 2k(n_2 - n_1)\frac{k}{n_1} \right) \\ \leqslant 2k^2 \left(1 + k(x - 1 + \frac{1}{n}) \right) \leqslant 2k^3 x\,.\end{aligned}$$

On applique ensuite le lemme 3.6 pour les bispéciaux faibles. □

Lemme 3.9. — *Soient* e *un facteur bispécial ordinaire de longueur* n *et* $x \in \Lambda$ *tel que* xe *est encore spécial à droite. Soient* $c_1, \dots, c_r$ *les mots de retour de* e*,* r *étant majoré par la constante* B *du lemme 3.5,* $C_1, \dots, C_{r'}$

les mots de retour de xe, avec $r' \leqslant B$, et a et b les constantes du lemme 3.4. Alors :

– *ou bien*

$$\sum_{k=1}^{r'} |C_k| \geqslant \sum_{i=1}^{r} |c_i| + bn\,,$$

– *ou bien*

$$\sum_{k=1}^{r'} |C_k| = \sum_{i=1}^{r} |c_i|$$

et il existe un bispécial faible g, de longueur inférieure à $(a+1)n$, et qui contient e comme préfixe.

Démonstration. — Soit e un facteur bispécial ordinaire, alors e est précédé de x ou x', suivi de y ou y', mais seuls xey, $x'ey$ et xey' sont facteurs de u. Les mots de retour de xe sont des concaténations de mots de retour de e, et chaque mot de retour de e apparaît dans au moins un mot de retour de xe.

Si un mot de retour de e apparaît deux fois dans les mots de retour de xe, comme les autres mots de retour de e apparaissent au moins une fois, on a

$$\sum_{k=1}^{r'} |C_k| \geqslant \sum_{i=1}^{r} |c_i| + bn$$

puisque $|c_i| \geqslant bn$ par le lemme 3.4.

Supposons que chaque mot de retour de e n'apparaît qu'une fois dans les mots de retour de xe de sorte que $\sum_{k=1}^{r'} |C_k| = \sum_{i=1}^{r} |c_i|$. Soit c un mot de retour de e ; si xec est facteur de u, c apparaît au début d'un mot de retour de xe ; si $x'ec$ est facteur, c apparaît dans un mot de retour de xe, mais précédé d'au moins un autre mot de retour de e. Comme c n'apparaît qu'une fois dans les mots de retour de xe, les mots xec et $x'ec$ ne peuvent être tous les deux facteurs, donc ec n'est pas spécial à gauche. Soit d de longueur maximale tel que ed soit spécial à gauche : un tel d existe puisque e est spécial à gauche, et d ne peut avoir aucun mot de retour c de e comme préfixe, donc d est un préfixe propre d'un tel mot de retour et $|d| < an$. Le mot ed est alors bispécial faible.

□

Lemme 3.10. — *Si $\limsup \frac{R(n)}{n} < +\infty$, il existe des constantes Q et Q' telles que pour tout $n \geqslant 1$ il existe deux entiers $n \leqslant m < m' \leqslant 2n$ vérifiant*

– $m' - m \geqslant \frac{n}{Q'}$,
– *le nombre de mots singuliers de longueur strictement comprise entre m et m' est au plus Q.*

Démonstration. — Si u est ultimement périodique, u a un nombre fini de mots singuliers et le résultat est donc trivial. On suppose donc que u n'est pas ultimement périodique. Par définition, si w est un mot singulier, $w = xvy$, où v est un facteur bispécial strict ou ordinaire.

Par le lemme 3.8, le nombre N_1 de facteurs bispéciaux stricts dont la longueur est comprise entre $n-1$ et $2n-2$ est borné par $4k^3$, où k est la constante définie dans le lemme 3.1, et le nombre N_2 de facteurs bispéciaux faibles dont la longueur est comprise entre $n-1$ et $2n-2$ par $4k^3+k$ et donc $N_2+N_1 \leqslant 8k^3+k$. En posant $Q' = 8k^3+k+1$, on peut donc trouver m et m' tels que $n \leqslant m < m' \leqslant 2n$, $m'-m \geqslant \frac{n}{Q'}$ et tels qu'il n'y ait aucun facteur bispécial faible ou strict de longueur strictement comprise entre $m-2$ et $m'-2$ (notons que si $n \leqslant Q'$, c'est trivialement vrai).

Soit P le nombre de facteurs bispéciaux ordinaires dont la longueur est strictement comprise entre $m-2$ et $m'-2$; on va majorer P.

Soient $d_1, \ldots, d_{k'}$ les facteurs spéciaux à droite de longueur $m'-2$, on a $k' \leqslant k$. Pour tout $p \in [m-1, m'-3]$, les facteurs spéciaux à droite de longueur p sont les suffixes de longueur p de $d_1, \ldots, d_{k'}$. En effet, sinon il existe $p \in [m-1, m'-3]$ et un mot w de longueur p tel que w est spécial à droite, mais ni $0w$ ni $1w$ ne sont spéciaux à droite, et $0w$ et $1w$ ne peuvent donc apparaître que suivis respectivement de x et x' avec $x \neq x'$; donc $0wx$ et $1wx'$ sont facteurs, mais pas $1wx$ ni $0wx'$, et w est bispécial faible, ce qui contredit le choix de m et m'. Les facteurs bispéciaux étant spéciaux à droite, il existe donc au moins $\frac{P}{k}$ facteurs bispéciaux ordinaires qui sont des suffixes d'un même d_{i_0}. Soit e_j le suffixe de longueur $m-2+j$, $1 \leqslant j \leqslant m'-m$, de ce d_{i_0}.

Soient $c_1, \ldots, c_r$ les mots de retour de e_j. Si e_j n'est pas bispécial, e_j est toujours précédé d'une même lettre x, et les mots de retour de xe_j sont encore $c_1, \ldots, c_r$. Si e_j est bispécial, il ne peut être qu'ordinaire par construction de m et m', donc on est dans un des deux cas du lemme 3.9.

Si on est dans le deuxième cas, c'est-à-dire si la somme des temps de retour de e_j est égale à celle des temps de retour de xe_j, son extension spéciale à droite, alors il existe un bispécial faible g_j dont la longueur est inférieure à $(a+1)(2n-2)$ et qui contient e_j comme préfixe. Pour un g donné, soient $j_1, j_2, \ldots, j_h$ ($j_1 < j_2 < \ldots < j_h$) les j pour lesquels g_j existe et $g_j = g$. On a donc $g_{j_1} = g_{j_2} = \ldots = g_{j_h} = g$. On a une occurrence de e_{j_1} comme suffixe de chaque e_{j_i}, qui est préfixe de g, et donc h occurrences distinctes de e_{j_1} à l'intérieur de g. Mais les mots de retour de e_{j_1} ont une longueur supérieure ou égale à $b(n-1)$, donc la longueur de g est comprise entre $(n-1)+(h-1)b(n-1)$ et $(a+1)(2n-2)$, d'où $h \leqslant 1 + \frac{2a+1}{b}$.

Le nombre de e_j pour lesquels on se trouve dans le deuxième cas du lemme 3.9 est donc égal au plus à $\frac{2a+b+1}{b}M$, où M est le nombre de bispéciaux faibles de longueur comprise entre $(n-1)$ et $(a+1)(2n-2)$; par le lemme 3.8, $M \leqslant 2(2a+2)k^3+k$, donc le nombre de tels e_j est au plus

$$K = \frac{2a+b+1}{b}\left((4a+4)k^3+k\right),$$

indépendamment de n.

Mais, chaque fois qu'on est dans le premier cas, la somme des longueurs des mots de retour augmente d'au moins $b(n-1)$. Si on part des mots de retour de e_1, qui sont de longueur au moins $b(n-1)$, on applique le processus ci-dessus pour arriver aux mots de retour de $e_{m'-m} = d_{i_0}$, après être passé par au moins $\frac{P}{k}$ facteurs bispéciaux ordinaires.

La somme des longueurs des mots de retour de d_{i_0} est alors au moins $\left(\frac{P}{k}-K\right)b(n-1)$ et l'un d'eux a donc une longueur d'au moins $\left(\frac{P}{k}-K\right)\frac{b(n-1)}{B}$, et d'au plus $a(2n-2)$; on a donc $P \leqslant Q = \frac{2aBk}{b}+kK$.

À chaque facteur bispécial ordinaire de longueur strictement comprise entre $m-2$ et $m'-2$ correspond exactement un mot singulier de longueur strictement comprise entre m et m' ; comme il n'y a pas de facteurs bispéciaux stricts de longueur strictement comprise entre $m-2$ et $m'-2$, on obtient ainsi tous les mots singuliers, et leur nombre est donc majoré par Q. □

Théorème 3.11. — *Si u est non ultimement périodique, $\frac{R(n)}{n}$ ne peut avoir une limite finie.*

Démonstration. —

Cas des alphabets à deux lettres.

Supposons que $\frac{R(n)}{n}$ a une limite finie. Si u est non ultimement périodique, il y a une infinité de mots singuliers, car il y a une infinité de facteurs bispéciaux stricts ou ordinaires.

Soient h_p les longueurs successives des mots singuliers de u, et r_p le plus grand temps de retour de tous les mots singuliers de longueur inférieure ou égale à h_p. D'après le lemme 2.2, $r(h_p) = r(h_{p+1}-1) = r_p$. Donc $\frac{R(n)}{n} = \frac{r_p+h_p-1}{h_p}$ pour $n = h_p$, $\frac{R(n)}{n} = \frac{r_p+h_{p+1}-2}{h_{p+1}-1}$ pour $n = h_{p+1}-1$; $\frac{r_p}{h_p}$ et $\frac{r_p}{h_{p+1}}$ ont donc même limite, finie et non nulle par le lemme 3.4. Donc $\frac{h_{p+1}}{h_p} \to 1$ quand $p \to +\infty$.

Mais le lemme 3.10 permet, pour tout n fixé, de trouver

$$n \leqslant m < m' \leqslant 2n, \quad m'-m \geqslant \frac{n}{Q'},$$

tels qu'il y ait au plus Q mots singuliers de longueur strictement comprise entre m et m' ; il existe donc (parmi ces mots singuliers, plus le mot singulier précédent et le mot singulier suivant) deux mots singuliers consécutifs dont les longueurs diffèrent d'au moins $\frac{n}{(Q+1)Q'}$; on a donc trouvé un p tel que $h_p \leqslant 2n$ et $h_{p+1} - h_p \geqslant \frac{n}{(Q+1)Q'}$, donc $\frac{h_{p+1}}{h_p} \geqslant 1 + \frac{1}{2(Q+1)Q'}$; comme $h_{p+1} \geqslant n$, cette situation se produit pour une infinité de p, ce qui contredit le paragraphe précédent.

Cas des alphabets plus grands.

Nous allons utiliser les techniques développées dans [3].

Soient maintenant Λ un alphabet de taille quelconque s et u une suite sur Λ. On va coder u de la manière suivante.

Supposons que les éléments de Λ sont numérotés de c_0 à c_{s-1} et soit φ le morphisme de $\Lambda^\star$ dans $\Lambda'^\star$ avec $\Lambda' = \{a, b\}$, où $\Lambda^\star$ est l'ensemble des mot finis sur l'alphabet Λ, défini par $\varphi(c_i) = a^{i+1}b^{s-i}$ pour tout i compris entre 0 et $s-1$.

Pour $n \geqslant s+1$, on a

$$(s+1)R_u\left(\left\lfloor \frac{n}{s+1} \right\rfloor\right) - s \leqslant R_{\varphi(u)}(n) \leqslant (s+1)R_u\left(\left\lceil \frac{n+s}{s+1} \right\rceil\right) + s\,.$$

En effet, tout facteur de longueur n de $\varphi(u)$ est facteur d'un mot de la forme $\varphi(v)$, où v est un facteur de longueur $\lceil \frac{n+s}{s+1} \rceil$ de u, qui apparaît donc dans tout facteur de longueur $R_u(\lceil \frac{n+s}{s+1} \rceil)$ de u, d'où la majoration. Inversement, soient v un facteur de u de longueur $\lfloor \frac{n}{s+1} \rfloor$ et w un facteur de u de longueur $R_u(\lfloor \frac{n}{s+1} \rfloor) - 1$ tel que w ne contient pas v : alors $\varphi(w)$ est un facteur de $\varphi(u)$ de longueur $(s+1)R_u(\lfloor \frac{n}{s+1} \rfloor) - s - 1$ qui ne contient pas $\varphi(v)$, car, par construction de φ, $\varphi(v)$ n'apparaît dans $\varphi(u)$ qu'à des positions multiples de $s+1$, d'où la minoration.

D'après cet encadrement, $\frac{R_u(n)}{n}$ a une limite finie si et seulement si $\frac{R_{\varphi(u)}(n)}{n}$ a une limite finie (et ces limites sont alors égales). On a ainsi ramené le cas général à celui de l'alphabet binaire. □

BIBLIOGRAPHIE

[1] P. ALESSANDRI, « Codages de rotations et basses complexités », Thèse, Université Aix-Marseille II, 1996.

[2] P. ARNOUX & G. RAUZY, « Représentation géométrique de suites de complexité $2n+1$ », *Bull. Soc. Math. France* **119** (1991), p. 199-215.

[3] J. CASSAIGNE, « Special factors of sequences with linear subword complexity », *Developments in Language Theory (Magdeburg, 1995)* (1996), p. 25-34, World Scientific.

[4] ———, « Complexité et facteurs spéciaux », *Bull. Belg. Math. Soc.* **4** (1997), p. 67-88.

[5] ———, « Limit values of the recurrence quotient of Sturmian sequences », *Theoret. Comp. Sci.* **218** (1999), p. 3-12.

[6] N. CHEKHOVA, P. HUBERT & A. MESSAOUDI, « Propriétés combinatoires, ergodiques et arithmétiques de la substitution de Tribonacci », *J. Théorie Nombres Bordeaux* **13** (2001), p. 371-394.

[7] F. DURAND, B. HOST & C. SKAU, « Substitutional dynamical Bratteli diagrams and dimension groups », *Ergodic Theory Dynam. Systems* **19** (1999), p. 953-993.

[8] M. MORSE & G. A. HEDLUND, « Symbolic dynamics II. Sturmian trajectories », *Amer. J. Math.* **62** (1940), p. 1-42.

[9] J. MOULINE, « Contribution à l'étude de la complexité des suites substitutives », Thèse, Université de Provence, 1989.

[10] G. RAUZY, « Nombres algébriques et substitutions », *Bull. Soc. Math. France* **110** (1982), p. 147-178.

Julien CASSAIGNE
Institut de mathématiques de Luminy
163 avenue de Luminy
Case 907
13288 Marseille Cedex 9 (France)
cassaigne@iml.univ-mrs.fr

Nataliya CHEKHOVA
Université de Tours
Faculté des sciences et techniques
Laboratoire de mathématiques et physique théorique
Parc de Grandmont
37200 Tours (France)
natalya@univ-tours.fr