ANNALES

DE

L'INSTITUT FOURIER

Marion LE GONIDEC

Sur la complexité de mots infinis engendrés par des q-automates dénombrables

Tome 56, n° 7 (2006), p. 2463-2491.

<http://aif.cedram.org/item?id=AIF_2006__56_7_2463_0>

Ann. Inst. Fourier, Grenoble
56, 7 (2006) 2463-2491

SUR LA COMPLEXITÉ DE MOTS INFINIS ENGENDRÉS PAR DES q-AUTOMATES DÉNOMBRABLES

par Marion LE GONIDEC

RÉSUMÉ. — On étudie, dans cet article, les propriétés combinatoires de mots engendrés à l'aide de q-automates déterministes dénombrables de degré borné, ou de manière équivalente, engendrés par des substitutions de longueur constante uniformément bornées sur un alphabet dénombrable. En particulier, on montre que la complexité de tels mots est au plus polynomiale et que, sur plusieurs exemples, elle est au plus de l'ordre de grandeur de $n(\log n)^p$.

ABSTRACT. — This article deals with combinatorial properties of infinite words generated by countable and deterministic q-automata of bounded degree. Those words can also be viewed as projections of fixed point of some substitutions of constant length on countable alphabet. We show that complexity function of such words is at most polynomial and, on some examples, the order of growth of this function is at most $n(\log n)^p$.

1. Introduction

De nombreuses études concernent les mots infinis engendrés par des substitutions sur un alphabet fini ou par des automates finis. Les livres [18], [15], [17] et [2] offrent un large panorama des propriétés dynamiques, arithmétiques et combinatoires de ces mots.

Les propriétés combinatoires de ces mots infinis, notamment leur complexité, ont été plus particulièrement étudiées les livres [9] et [10] et dans [5], [6], [14] et [13]. Nous nous proposons d'étendre ces travaux à une nouvelle classe de mots infinis engendrés à l'aide d'algorithmes simples : des substitutions de longueur constante q sur un alphabet dénombrable ou, de

Mots-clés : mots infinis, complexité, substitutions, automates .
Classification math. : 68R15, 11B85.

manière équivalente, des q-automates dont l'ensemble des états est dénombrable. Ces objets ont été définis dans [11].

1.1. Notations concernant les suites symboliques

Soit E un ensemble fini ou dénombrable qui est appelé *alphabet*.

Pour tout entier n positif, on note les éléments de E^n sous forme concaténée : un élément m de E^n se note $m = m_0 m_1 \cdots m_{n-1}$. On appelle ces objets des *mots de longueur* n et la longueur d'un mot m donné est notée $|m|$. On note l'ensemble des mots finis $E^* = \cup_{n \geqslant 0} E^n$. L'ensemble E^* muni de l'opération de concaténation est un monoïde.

On note E^ω l'ensemble des suites indexées par $\mathbb{N}$ à valeurs dans E. Les suites à valeurs dans E seront notées sous forme concaténée et appelées indifféremment *suites* ou *mots infinis* : $m \in E^\omega$ s'écrit $m = m_0 m_1 \cdots m_n \cdots$ avec, pour tout n, $m_n \in E$.

Un mot fini w de longueur n qui apparaît dans un mot infini m, c'est-à-dire qui vérifie pour un certain entier k, $w = m_k m_{k+1} \cdots m_{k+n-1}$, est appelé *facteur* ou *sous-mot de m de longueur n*. Pour tout entier $n > 0$, on note $F_n(m)$ l'*ensemble des facteurs de longueur n de m* et $L(m)$ le *langage du mot m*, c'est-à-dire l'ensemble des mots que l'on peut «lire» dans le mot m :

$$L(m) = \{w \in E^* | \exists (p,s) \in E^* \times E^\omega, m = pws\}.$$

Dans la suite de cet article, on désigne par q un nombre entier supérieur à 2 et par $\overline{n}^q = n_l \cdots n_1 n_0$, la représentation en base q de l'entier $n = \sum_{i=0}^{l} n_i q^i$ où $n_i \in \{0, 1, \ldots, q-1\}$ et $n_l \neq 0$.

On donnera les définitions précises concernant les *substitutions* et les *q-automates* dans le paragraphe 2. Pour plus de détails, voir [17] ou [7].

1.2. Complexité d'un mot infini

Pour un mot $m = m_0 m_1 \cdots m_n \cdots$ fixé à valeurs dans un alphabet fini A, on appelle *fonction de complexité* du mot m, notée p_m, la fonction suivante :

$$\begin{aligned} p_m : \quad \mathbb{N} &\to \mathbb{N} \\ n &\mapsto \mathrm{Card}\,(F_n(m)). \end{aligned}$$

Comme $F_{n+n'}(m) \subset F_n(m) F_{n'}(m)$ pour tous entiers n et n', la fonction de complexité vérifie $p_m(n+n') \leqslant p_m(n) p_m(n')$ et en particulier, la suite

$\left(\frac{\log p_m(n)}{n \operatorname{Card}(A)}\right)_{n\geqslant 1}$ converge. On peut montrer que l'entropie topologique du système dynamique associé à m peut être exprimée à l'aide de la complexité par la formule suivante : $h(m) = \lim_{n\to\infty} \frac{\log p_m(n)}{n \operatorname{Card}(A)}$. Voir, par exemple [18].

On a toujours $0 \leqslant h(m) \leqslant 1$, avec l'idée que plus le langage d'un mot est riche, plus il est «compliqué» et son entropie topologique se rapproche de 1.

On peut classer les mots infinis sur un alphabet fini A par leur complexité. Par exemple, les mots ultimement périodiques sont les mots de complexité bornée. Les mots dits sturmiens sont les mots infinis m de complexité $p_m(n) = n + 1$; ce sont les mots non ultimement périodiques de complexité minimale. Voir, par exemple [17].

Cependant, la détermination d'une formule exacte pour la fonction de complexité d'un mot infini donné reste un problème délicat qui a fait l'objet de nombreux travaux. Voir par exemple [1] ou [17] pour un survol.

Pour certaines classes de mots infinis, les ordres de grandeur possibles de la fonction de complexité ont été déterminés. C'est le cas des mots substitutifs à valeurs dans un alphabet fini A, dont les ordres possibles ont été donnés par J.-J. Pansiot [14] :

THÉORÈME 1.1. — *Pour tout mot infini m engendré par une substitution sur un alphabet fini A, il existe une fonction $f(n)$ égale à 1, n, $n\log n$, $n\log\log n$ ou n^2 telle que :*

$$\exists (c,C) \in \left(\mathbb{R}_+^*\right)^2,\ cf(n) \leqslant p_m(n) \leqslant Cf(n).$$

En particulier, lorsque m est q-automatique et non ultimement périodique, $f(n) = n$.

Ce résultat affine deux théorèmes plus anciens : le premier est dû à A. Cobham [5] :

THÉORÈME 1.2. — *Si m est un mot engendré par un q-automate fini à k états, alors :*

$$\forall n > 0, p_m(n) \leqslant qk^2 n.$$

Le second, dû à A. Ehrenfeucht, K. P. Lee et G. Rozenberg [6] indique que la complexité d'un mot m engendré par une substitution sur un alphabet fini vérifie $p_m(n) = O(n^2)$.

L'idée de cet article est d'obtenir un résultat analogue au théorème 1.2 pour une classe de suites proches des suites automatiques, à savoir des suites engendrées à l'aide d'automates dénombrables.

2. Substitutions et q-automates

Les substitutions et les automates sont des algorithmes simples qui permettent d'engendrer des mots infinis. Même si nous nous limiterons dans la suite à un type de substitutions et d'automates particuliers, les définitions sont données dans un cadre général.

Nous allons exposer notre problématique sous les deux angles. Cependant, le point du vue substitutif sera utilisé de manière privilégiée dans les démonstrations. Le point de vue automatique, beaucoup plus visualisable, permet d'avoir un outil de compréhension supplémentaire.

2.1. Vision substitutive

DÉFINITION 2.1. — *Une* substitution *ou* morphisme $\overline{\sigma}$ *sur un alphabet dénombrable E est une application de E dans E^*, ce que l'on peut écrire de la manière suivante :*

$$\begin{array}{rccl} \overline{\sigma}: & E & \to & E^* \\ & e & \mapsto & \overline{\sigma}_0(e)\overline{\sigma}_1(e)\cdots\overline{\sigma}_{|\overline{\sigma}(e)|-1}(e) \end{array}$$

où, pour tout $i \in \{0,1,\ldots,|\overline{\sigma}(e)|-1\}$, $\overline{\sigma}_i$ est une application de E dans E.

L'application $\overline{\sigma}$ est ensuite étendue au monoïde E^* et à E^ω par concaténation. Lorsque $|\overline{\sigma}(e)|$ est constant sur E, on dit que $\overline{\sigma}$ est une substitution de longueur constante.

Le terme de «substitution» provient du fait que l'action de $\overline{\sigma}$ sur un mot fini ou infini consiste à substituer chaque lettre par son mot-image par $\overline{\sigma}$.

Par la suite, nous allons nous intéresser aux propriétés des mots infinis à valeurs dans E qui sont *point fixes de substitutions de longueur constante*, c'est-à-dire aux mots infinis $\overline{m}$ qui vérifient $\overline{\sigma}(\overline{m}) = \overline{m}$, pour des substitutions $\overline{\sigma}$ de longueur constante et aux projections lettre-à-lettre sur un autre alphabet de ces points fixes.

Pour cela, il est nécessaire de garantir l'existence de mots infinis points fixes de nos substitutions. On supposera donc qu'il existe au moins un élément e_0 de E tel que $\overline{\sigma}(e_0) \in e_0E^*$. Cette condition est nécessaire et suffisante pour qu'une substitution de longueur constante $q \geqslant 2$ admette au moins un point fixe qui ne soit pas le mot vide.

2.2. Vision automatique

DÉFINITION 2.2. — *Un* q-automate déterministe $\mathcal{A}$ *est un quintuplet du type $\mathcal{A} = (E, e_0, \phi, A, \Pi)$, où :*

– *E est un alphabet fini ou dénombrable, appelé* ensemble des états,
– *e_0 est un élément de E, appelé* état initial,
– *ϕ est une application de $E \times \{0, 1, \ldots, q-1\}$ dans E appelée* fonction de transition,
– *A est un alphabet fini ou dénombrable,*
– *Π est une fonction de E vers A donnée, appelée* projection de sortie.

L'automate $\mathcal{A}$ sera qualifié de fini *ou* dénombrable *selon que l'ensemble des états E est fini ou dénombrable.*

On représente en général un automate par un graphe étiqueté, dont l'ensemble des sommets est E et l'ensemble des arcs orientés est donné par $\{(e, e', i) \in E^2 \times \{0, 1, \ldots, q-1\} | \phi(e, i) = e'\}$, les éléments (e, e', i) étant symbolisés par des flèches de e vers e' étiquetées par i.

Pour tout (e, i) de $E \times \{0, 1, \ldots, q-1\}$, on pose $\phi(e, i) = e \cdot i$ et on prolonge ϕ en une application de $E \times \mathbb{N}$ dans E de la manière suivante : si $\overline{n}^q = n_l \cdots n_1 n_0$ est la représentation en base q de $n \in \mathbb{N}$, l'action de n sur un élément e de E est donnée par l'action successive de ses chiffres :

$$e \cdot n = (((e \cdot n_l) \cdots) \cdot n_1) \cdot n_0.$$

On appelle *mot de sortie de l'automate $\mathcal{A}$* le mot $\mu(\mathcal{A})$, défini par $\mu(\mathcal{A})_n = \Pi(e_0 \cdot n)$ pour tout entier $n \geqslant 0$.

2.3. Correspondance entre substitutions et q-automates

Il existe une correspondance entre les q-automates déterministes dont l'ensemble des états est E et les substitutions de longueur constante q sur l'alphabet E (voir [11]). Cette correspondance est analogue à celle décrite par Cobham [5] dans le cas où E est un alphabet fini.

En effet, si $\overline{\sigma}$ est une substitution sur un alphabet E, définie comme au paragraphe 2.1, il existe un graphe étiqueté $\mathcal{G}(\overline{\sigma})$, associé naturellement à $\overline{\sigma}$, dont l'ensemble des sommets est E et l'ensemble des arcs orientés et étiquetés est donné par l'ensemble $\{(e, e', i) \in E^2 \times \{0, 1, \ldots, q-1\}, \overline{\sigma}_i(e) = e'\}$. Les arcs étiquetés sont représentés sur le graphe par des flèches de e vers e' indexée par i.

Lorsque la substitution $\overline{\sigma}$ est de longueur constante, le graphe $\mathcal{G}_{\overline{\sigma}}$ associé à $\overline{\sigma}$ est exactement le graphe associé à l'automate $\mathcal{A}_{\overline{\sigma}} = (E, e_0, \phi_{\overline{\sigma}}, E, Id_E)$, où :
– l'ensemble des états est E,
– l'état initial est e_0,

– la fonction de transition $\phi_{\overline{\sigma}}$ est donnée par : $\forall i \in \{0, 1 \ldots, q-1\}$, $\forall e \in E$, $\phi_{\overline{\sigma}}(e, i) = \overline{\sigma}_i(e)$.

De plus, si $\overline{\sigma}$ admet un point fixe $\overline{m}$ contenu dans $e_0 E^\omega$, $\overline{m}$ est exactement le mot de sortie de l'automate $\mathcal{A}_{\overline{\sigma}}$:

$$\overline{m} = \overline{\sigma}(\overline{m}) = \mu(\mathcal{A}_{\overline{\sigma}}).$$

On retrouvera d'ailleurs ce résultat comme corollaire de la proposition 3.1. L'étude de ces mots peut donc être envisagée sous les deux éclairages différents : comme point fixe de substitution de longueur constante q ou comme mot de sortie d'un q-automate.

Notation 2.3. — Soient E et A deux alphabets finis ou dénombrables. Si $\overline{\sigma}$ est une substitution donnée sur l'alphabet E et Π est une projection lettre-à-lettre de E vers A, on note donc $(E, e_0, \phi_{\overline{\sigma}}, A, \Pi)$ l'automate associé à $\overline{\sigma}$.

2.4. Mots infinis engendrés par un q-automate dénombrable

Définition 2.4. — *On dit qu'une projection lettre-à-lettre* Π *d'un alphabet dénombrable* E *vers un alphabet fini* A *est* admissible *s'il existe un sous-ensemble fini* F *de* E *tel que la restriction de* Π *à* $E \setminus F$ *est constante.*

Définition 2.5. — *Soit* $\overline{\sigma}$ *une substitution de longueur constante* q *à valeurs dans un alphabet* E *qui admet un point fixe* $\overline{m}$ *dans* $e_0 E^\omega$.

On dit qu'un mot infini m *à valeurs dans* A *est* engendré par le q-automate dénombrable $(E, e_0, \phi_{\overline{\sigma}}, A, \Pi)$ *si* A *est un alphabet fini,* Π *est une projection admissible et* m *est la projection par* Π *du point fixe* $\overline{m}$ *de* $\overline{\sigma}$.

L'objet de notre travail est de majorer la fonction de complexité des mots infinis m engendrés par des q-automates dénombrables.

2.5. Pourquoi introduire ces objets ?

Les automates, les langages reconnus par un automate, les mots infinis qu'ils engendrent et les langages associés à ces mots sont des objets qui apparaissent en mathématiques et en informatique dans plusieurs domaines : théorie des langages [19] , logique [16], théorie des codes [3], arithmétique et théorie des nombres [2], [17], mais aussi en physique théorique, en biologie et neurobiologie, etc... Ce sont notamment des outils utiles pour la

recherche de motifs dans une séquence, les modélisations de systèmes de communications ou le codage de systèmes dynamiques.

Du point de vue de la théorie des nombres, ces objets sont utiles car on peut associer à un mot infini des ensembles d'entiers de la manière suivante : pour m mot infini à valeurs dans un alphabet fini A et a une lettre fixée de A, on appelle *support de* a l'ensemble $[m]_a$ défini par $[m]_a = \{n \in \mathbb{N}, m_n = a\}$.

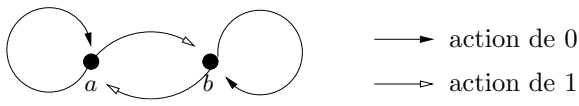
La nature du mot m induit des propriétés combinatoires et arithmétiques sur les ensembles $[m]_a$ pour $a \in A$. Voir [2], [11], [18] et le chapitre 4 de [17].

En particulier, il existe une correspondance entre les mots infinis binaires définis par un q-automate fini et certains sous-ensembles de $\mathbb{N}$, dits ensembles q-reconnaissables (voir par exemple [2]).

Étant donné un q-automate fini fixé, le mot m indicateur de l'ensemble R des entiers dont la représentation en base q est reconnaissable par l'automate est un mot q-automatique.

La complexité de m quantifie alors, en quelque sorte, le caractère aléatoire de la répartition des entiers de R.

Par exemple, l'ensemble des entiers qui ont un nombre pair de 1 dans leur écriture binaire est 2-reconnaissable : le mot de sortie m du 2-automate suivant est exactement le mot indicateur de cet ensemble ($m_n = 1$ si n a un nombre pair de 1 dans son écriture binaire et $m_n = 0$ sinon).



La fonction de sortie est donnée par $\Pi(a) = 1$ et $\Pi(b) = 0$.

Le mot de sortie m est le mot de Thue-Morse.

Ce mot infini peut aussi être vu projection par Π du point fixe infini commençant par a de la substitution sur $\{a, b\}$ suivante :

$$\begin{aligned} \overline{\sigma} : \quad a &\mapsto ab, \\ b &\mapsto ba. \end{aligned}$$

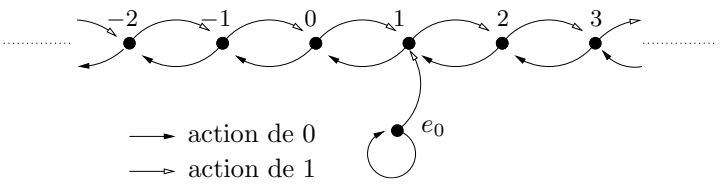
La complexité du mot m est donnée par [4] : $p_m(1) = 2$, $p_m(2) = 4$ et pour tout $n \geqslant 3$, si $n = 2^r + q + 1$ avec $r \geqslant 0$ et $q \in [1, 2^r]$ alors :

$$p_m(n) = \begin{cases} 62^{r-1} + 4q & \text{si } 1 \leqslant q \leqslant 2^{r-1}, \\ 82^{r-1} + 2q & \text{si } 2^{r-1} + 1 \leqslant q \leqslant 2^r. \end{cases}$$

Les q-automates finis sont cependant inutilisables pour décrire des sous-ensembles de $\mathbb{N}$ définis par des propriétés de leur écriture en base q où interviennent certains processus de comptage.

Par exemple, si on considère l'ensemble des entiers $\mathcal{S}$ qui ont autant de 0 que de 1 dans leur écriture binaire, le mot indicateur de cet ensemble, c'est-à-dire le mot infini m tel que $m_n = 1$ si n appartient à $\mathcal{S}$ et $m_n = 0$ sinon, n'est pas q-automatique.

Cependant, ce mot est le mot de sortie de l'automate $(\mathbb{Z} \cup \{e_0\}, e_0, \phi, \{0,1\}, \Pi)$, dont le graphe est le suivant :



la fonction de sortie étant donnée par :

$$\begin{array}{rccc} \Pi : & \mathbb{Z} \cup \{e_0\} & \to & \{0,1\} \\ & 0 & \mapsto & 1 \\ & e \neq 0 & \mapsto & 0. \end{array}$$

On peut aussi voir le mot indicateur de cet ensemble comme la projection par Π du point fixe de la substitution suivante :

$$\begin{array}{rccc} \overline{\sigma} : & \mathbb{Z} \cup \{e_0\} & \to & \mathbb{Z}^* \cup \{e_0 1\} \\ & e & \mapsto & (e-1)(e+1) \\ & e_0 & \mapsto & e_0 1. \end{array}$$

S. Ferenzci a étudié les propriétés dynamiques de cette substitution dans [8].

L'application $\overline{\sigma}$ s'étend au monoïde $(\mathbb{Z} \cup \{e_0\})^*$ par concaténation. On a rajouté ici un élément e_0 pour que la substitution $\overline{\sigma}$ puisse avoir un point fixe $\overline{m}$, contenu dans $e_0\mathbb{Z}^\omega$:

$$\overline{m} = e_0 102(-1)113(-2)0020224(-3)(-1)\cdots$$

La mot indicateur de $\mathcal{S}$ est donné par la projection $m = \Phi_{\{0\}}(\overline{m})$:

$$m = 0010000001101000\cdots$$

On montre, au paragraphe 4.1, que la complexité de ce mot infini est au plus de l'ordre de $n \log_2^2 n$. On trouvera dans [12] et [11] l'étude de certaines propriétés arithmétiques et statistiques de l'ensemble $\mathcal{S}$.

3. Propriétés d'un point fixe de substitution de longueur constante q sur un alphabet dénombrable

Soient E un alphabet dénombrable et q fonctions $\overline{\sigma}_i : E \to E$ avec $i = 0, \ldots, q-1$ et soit $\overline{\sigma}$ la substitution définie, comme au paragraphe 2.1, sur E par $\overline{\sigma}(e) = \overline{\sigma}_0(e)\overline{\sigma}_1(e)\cdots\overline{\sigma}_{q-1}(e)$.

La substitution $\overline{\sigma}$ a un point fixe $\overline{m}$ dans e_0E^ω dont on se propose d'étudier ici les propriétés.

PROPOSITION 3.1. — *Pour un élément e donné de $F_1(\overline{m})$ et un entier $k \geqslant 1$, le mot $\overline{\sigma}^k(e) = u_0u_1\cdots u_{q^k-1}$ vérifie :*

$$\forall n \in \{0, 1, \ldots, q^k - 1\},\ u_n = \overline{\sigma}_{n_0} \circ \overline{\sigma}_{n_1} \circ \cdots \circ \overline{\sigma}_{n_l} \circ \overline{\sigma}_0^{k-(l+1)}(e),$$

où $\overline{n}^q = n_l \cdots n_1 n_0$ est la représentation de n en base q.

Démonstration. — On remarquera avant toute chose que, $\overline{\sigma}$ étant une substitution de longueur constante, pour tout e appartenant à E et tout entier $k \geqslant 0$, le mot $\overline{\sigma}^k(e)$ est de longueur q^k.

Cette proposition se démontre par récurrence sur k.

L'initialisation au rang $k = 1$ découle de la définition de $\overline{\sigma}$, car $\overline{\sigma}(e) = \overline{\sigma}_0(e)\overline{\sigma}_1(e)\cdots\overline{\sigma}_{q-1}(e)$.

Supposons que pour $k \geqslant 1$ fixé, quel que soit l'élément e' de E, le mot $\overline{\sigma}^k(e') = u_0u_1\cdots u_{q^k-1}$ vérifie : $\forall n \in \{0, 1, \ldots, q^k - 1\},\ u_n = \overline{\sigma}_{n_0} \circ \overline{\sigma}_{n_1} \circ \cdots \circ \overline{\sigma}_{n_l} \circ \overline{\sigma}_0^{k-(l+1)}(e')$,

Soit e un élément de E. On pose $w = \overline{\sigma}^{k+1}(e)$. Comme $\overline{\sigma}$ est une substitution, $w = \overline{\sigma}^k\left(\overline{\sigma}(e)\right)$, on a donc l'égalité $w = \overline{\sigma}^k\left(\overline{\sigma}_0(e)\overline{\sigma}_1(e)\cdots\overline{\sigma}_{q-1}(e)\right) = \overline{\sigma}^k\left(\overline{\sigma}_0(e)\right)\overline{\sigma}^k\left(\overline{\sigma}_1(e)\right)\cdots\overline{\sigma}^k\left(\overline{\sigma}_{q-1}(e)\right)$.

Soit $n \in \{0, 1, \ldots, q^{k+1} - 1\}$ et $\overline{n}^q = n_l \cdots n_1 n_0$ sa représentation en base q.

Si $l < k$, c'est-à-dire si $n \in [0, q^k - 1]$, w_n est en fait la n-ième lettre de $\overline{\sigma}^k\left(\overline{\sigma}_0(e)\right)$ et donc, grâce à l'hypothèse de récurrence, on obtient

$$\begin{aligned} w_n &= \overline{\sigma}_{n_0} \circ \overline{\sigma}_{n_1} \circ \cdots \circ \overline{\sigma}_{n_l} \circ \overline{\sigma}_0^{k-(l+1)}\left(\overline{\sigma}_0(e)\right) \\ &= \overline{\sigma}_{n_0} \circ \overline{\sigma}_{n_1} \circ \cdots \circ \overline{\sigma}_{n_l} \circ \overline{\sigma}_0^{k+1-(l+1)}(e). \end{aligned}$$

Si $l = k$, alors w_n est la $(n - n_k q^k)$-ième lettre de $\overline{\sigma}^k\left(\overline{\sigma}_{n_k}(e)\right)$ et en utilisant une nouvelle fois l'hypothèse de récurrence, on obtient

$$w_n = \overline{\sigma}_{n_0} \circ \overline{\sigma}_{n_1} \circ \cdots \circ \overline{\sigma}_{n_{k-1}}\left(\overline{\sigma}_{n_k}(e)\right) = \overline{\sigma}_{n_0} \circ \overline{\sigma}_{n_1} \circ \cdots \circ \overline{\sigma}_{n_{k-1}} \circ \overline{\sigma}_{n_k}(e).$$

On a donc bien, pour tout $n \in \{0, 1, \ldots, q^{k+1} - 1\}$,

$$w_n = \overline{\sigma}_{n_0} \circ \overline{\sigma}_{n_1} \circ \cdots \circ \overline{\sigma}_{n_l} \circ \overline{\sigma}_0^{k+1-(l+1)}(e).$$

L'hypothèse de récurrence passe donc bien au rang $k+1$, ce qui permet de valider le résultat de la proposition pour tout entier k. □

COROLLAIRE 3.2. — *Soient $\overline{m}$ le point fixe d'une substitution $\overline{\sigma}$ de longueur constante q contenu dans e_0E^ω et $\mu(\mathcal{A})$ le mot de sortie de l'automate $\mathcal{A} = (E, e_0, \phi_{\overline{\sigma}}, E, Id_E)$, défini au paragraphe 2.3.*
On a l'égalité $\overline{m} = \mu(\mathcal{A})$.

Démonstration. — Il suffit de remarquer que, puisque $\overline{m} = \overline{\sigma}(\overline{m})$ et $\overline{m}$ est contenu dans e_0E^ω, $e_0 = \overline{\sigma}(e_0)$ et $\overline{m}$ est aussi contenu dans $\overline{\sigma}^k(e_0)E^\omega$, et cela, pour tout entier k.

Soit n un entier et soit k un entier tel que $n \leqslant q^k - 1$. Soit $\overline{n}^q = n_l \cdots n_1 n_0$ la représentation de n en base q. La lettre $\overline{m}_n$ est donc la n-ième lettre de $\overline{\sigma}^k(e_0)$. D'après la proposition 3.1, on en déduit que $\overline{m}_n = \overline{\sigma}_{n_0} \circ \overline{\sigma}_{n_1} \circ \cdots \circ \overline{\sigma}_{n_l} \circ \overline{\sigma}_0^{k-(l+1)}(e_0)$. Comme $\overline{\sigma}_0(e_0) = e_0$, on obtient $\overline{m}_n = \overline{\sigma}_{n_0} \circ \overline{\sigma}_{n_1} \circ \cdots \circ \overline{\sigma}_{n_l}(e_0)$.

D'après la définition de la fonction de transition de $\mathcal{A}$, on a donc $\overline{m}_n = e_0 \cdot n$, ce qui permet de conclure que $\overline{m}_n = \mu(\mathcal{A})_n$ pour tout entier n. □

PROPOSITION 3.3. — *Soit $(e_1, \ldots, e_{q-1}) \in E^{q-1}$ tel que $\overline{\sigma}(e_0) = e_0e_1 \cdots e_{q-1}$.*
Le mot x_1x_2 est dans $F_2(\overline{m})$ si et seulement si un des 4 cas suivants se produit :

1. $x_1x_2 = e_0e_1$,
2. $\exists k \geqslant 0, \exists i \in \{1, \ldots, q-2\}$, $x_1 = \overline{\sigma}_{q-1}^k(e_i)$ *et* $x_2 = \overline{\sigma}_0^k(e_{i+1})$,
3. $\exists k \geqslant 0$, $x_1 = \overline{\sigma}_{q-1}^k(e_{q-1})$ *et* $x_2 = \overline{\sigma}_0^{k+1}(e_1)$,
4. $\exists k \geqslant 0, \exists j \in \{0, 1, \ldots, q-2\}$, $x_1 \in F_1(\overline{m})$, $x_2 \in \overline{\sigma}_0^k \circ \overline{\sigma}_{j+1} \circ \overline{\sigma}_j^{-1} \circ \overline{\sigma}_{q-1}^{-k}(\{x_1\})$.

Démonstration. — Cette proposition se montre en utilisant la réécriture de $\overline{m}$ suivante :

$$\overline{m} = e_0e_1 \cdots e_{q-1}\overline{\sigma}(e_1) \cdots \overline{\sigma}(e_{q-1})\overline{\sigma}^2(e_1) \cdots \overline{\sigma}^2(e_{q-1}) \cdots ,$$

cette égalité découlant directement du fait que $\overline{\sigma}(\overline{m}) = \overline{m}$.
Si x_1x_2 est un mot de m, différent de e_0e_1, alors il y a 2 possibilités :

(a) le mot x_1x_2 apparaît dans m à une charnière $\overline{\sigma}^k(e_j)\overline{\sigma}^k(e_{j+1})$ ou à une charnière du type $\overline{\sigma}^k(e_{q-1})\overline{\sigma}^{k+1}(e_1)$ pour un certain entier positif k,

(b) le mot x_1x_2 est à l'intérieur d'un $\overline{\sigma}^k(e_j)$ pour un certain entier positif k et un certain entier $j \in \{0, 1, \ldots, q-2\}$.

(a) Si on trouve x_1x_2 à une charnière $\overline{\sigma}^k(e_j)\overline{\sigma}^k(e_{j+1})$, alors il existe un $k \geqslant 0$ tel que $x_1 = \overline{\sigma}^k_{q-1}(e_j)$ et $x_2 = \overline{\sigma}^k_0(e_{j+1})$. C'est une conséquence de la proposition 3.1 et cela correspond au deuxième cas de figure de la proposition.

De même, si on trouve x_1x_2 à une charnière $\overline{\sigma}^k(e_{q-1})\overline{\sigma}^{k+1}(e_1)$, alors il existe un $k \geqslant 0$ tel que $x_1 = \overline{\sigma}^k_{q-1}(e_{q-1})$ et $x_2 = \overline{\sigma}^{k+1}_0(e_1)$. Cela correspond au troisième cas de figure de la proposition.

(b) Si x_1x_2 apparaît dans m à l'intérieur d'un $\overline{\sigma}^k(e_j)$. La proposition 3.1 permet de dire qu'il existe un entier $0 \leqslant n < q^k - 1$ tel que x_1x_2 est donné par :

$$x_1 = \overline{\sigma}_{n_0} \circ \overline{\sigma}_{n_1} \circ \cdots \circ \overline{\sigma}_{n_l} \circ \overline{\sigma}_0^{k-(l+1)}(e_j)$$

et

$$x_2 = \overline{\sigma}_{(n+1)_0} \circ \overline{\sigma}_{(n+1)_1} \circ \cdots \circ \overline{\sigma}_{(n+1)_l} \circ \overline{\sigma}_0^{k-(l'+1)}(e_j),$$

avec $\overline{n}^q = n_l \cdots n_1 n_0$ et $\overline{n+1}^q = (n+1)_{l'} \cdots (n+1)_1 (n+1)_0$ avec $l \leqslant k$ et $l' = l$ ou $l+1$ selon le cas.

Si $n_0 = j$ pour un certain $j \in \{0, 1, \ldots, q-2\}$, on a $\overline{n+1}^q = n_l \cdots n_1 (j+1)$ et donc $x_2 \in \overline{\sigma}_{j+1} \circ \overline{\sigma}_j^{-1}(\{x_1\})$.

Si $n_{k'} \cdots n_1 n_0 = j(q-1)\cdots(q-1)$ pour un certain $k' \leqslant l-1$ et un certain $j \in \{0, 1, \ldots, q-2\}$, on a alors $\overline{n+1}^q = n_l \cdots n_{k'+1}(j+1)0 \cdots 0$ et donc $x_2 \in \overline{\sigma}_0^{k'} \circ \overline{\sigma}_{j+1} \circ \overline{\sigma}_j^{-1} \circ \overline{\sigma}_{q-1}^{-k'}(\{x_1\})$.

Si $\overline{n}^q = (q-1)\cdots(q-1)$, alors $l \leqslant k-1$ car on se trouve à l'intérieur d'un $\overline{\sigma}^k(e_j)$ et $|n+1|_2 = 10\cdots0$ avec $l' = l+1$, et on a $x_2 \in \overline{\sigma}_0^l \circ \overline{\sigma}_1 \circ \overline{\sigma}_0^{-1} \circ \overline{\sigma}_{q-1}^{-l}(\{x_1\})$.

Pour tout entier n de $\{0, 1, \ldots, q^k - 2\}$, on est dans le quatrième cas de figure de la proposition : pour k fixé, tous les mots x_1x_2 qui apparaissent dans $\overline{\sigma}^k(e_j)$ vérifient $x_2 \in \overline{\sigma}_0^{k'} \circ \overline{\sigma}_{j+1} \circ \overline{\sigma}_j^{-1} \circ \overline{\sigma}_{q-1}^{-k'}(\{x_1\})$ pour un certain $k' \in \{0, 1, \ldots, k-1\}$. □

4. Exemples de majoration de la complexité de mots engendrés par un q-automate dénombrable

Dans ce paragraphe, nous présentons quatre exemples de majoration de la complexité pour des mots engendrés par des q-automates dénombrables. Au terme $\log_q^p(n)$ près, les majorations obtenues sont optimales pour les trois premiers exemples. Des simulations informatiques tendent à montrer que c'est également le cas pour le quatrième.

Notation 4.1. — Pour une substitution $\overline{\sigma}$ de longueur constante q sur un alphabet dénombrable E telle que $\overline{\sigma}(e) = \overline{\sigma}_0(e)\overline{\sigma}_1(e)\cdots\overline{\sigma}_{q-1}(e)$, pour tout $e \in E$, on note

$$\mathcal{Z}_k(e) = \bigcup_{(j_1,\ldots,j_k)\in\{0,1,\ldots,q-1\}^k} \overline{\sigma}_{j_1}^{-1} \circ \cdots \circ \overline{\sigma}_{j_k}^{-1}(\{e\}).$$

D'autre part, pour faciliter la lecture des preuves, on notera, pour une substitution $\overline{\sigma}$ sur un alphabet dénombrable et Π une projection admissible, quel que soit l'entier $k \geqslant 1$, $\sigma^k = \Pi \circ \overline{\sigma}^k$.

4.1. Exemple 1

RÉSULTAT 4.2. — *La substitution de l'ivrogne est définie comme suit :*

$$\begin{array}{rccc} \overline{\sigma}: & \mathbb{Z}\cup\{e_0\} & \to & \mathbb{Z}^2\cup\{e_0 1\} \\ & e & \mapsto & (e-1)(e+1) \\ & e_0 & \mapsto & e_0 1. \end{array}$$

Soit m le mot infini engendré par le 2-automate $(\mathbb{Z}\cup\{e_0\}, e_0, \phi_{\overline{\sigma}}, \{0,1\}, \Pi_0)$ où la projection $\Pi_0 : \mathbb{Z}\cup\{e_0\} \to \{0,1\}$ est définie par $\Pi_0^{-1}(\{1\}) = \{0\}$.

On a la majoration de la complexité de m suivante :

$$\forall n \geqslant 1, \quad p_m(n) \leqslant n(\log_2 n + 9)(\log_2 n + 2).$$

La proposition 3.3 devient, dans ce cas simple :

LEMME 4.3. — *Soit $\overline{m}$ le mot infini point fixe de $\overline{\sigma}$ contenu dans $e_0\mathbb{Z}^\omega$. Les facteurs de $\overline{m}$ de longueur 2 sont donnés par :*

$$F_2(\overline{m}) = \{e_0 1\} \cup \big\{e_1 e_2,\ e_1 \in \mathbb{Z}, e_2 = e_1 - 2p,\ p \geqslant -1 \\ \text{ou } e_2 = -e_1 + 1 \text{ si } e_1 > 0\big\}.$$

Quelques remarques préliminaires : une infinité d'itérés $\overline{\sigma}^k(e)$ se projettent par Π_0 sur 0^{2^k}. En effet, les lettres de $\overline{\sigma}^k(e)$ sont parmi les entiers de l'intervalle $[e-k, e+k]$. Plus précisément :

1. Si e et k sont de même parité, toutes les valeurs entières paires de $[e-k, e+k]$, et uniquement elles, apparaissent dans $\overline{\sigma}^k(e)$.
2. Si e et k sont de parités différentes, toutes les valeurs entières impaires de $[e-k, e+k]$, et uniquement elles, apparaissent dans $\overline{\sigma}^k(e)$.

Ce résultat se montre par récurrence sur k et permet d'affirmer :

$$\sigma^k(e) \neq 0^{2^k} \Leftrightarrow 0 \text{ apparaît dans } \overline{\sigma}^k(e) \Leftrightarrow \exists p \in \{0, 1, \ldots, k\}, e = k - 2p.$$

Si w est un mot de m de longueur 2^k, ce mot est inclus dans un certain $\sigma^k(e_1)\sigma^k(e_2)$. Cependant, il peut exister une infinité de paires (e'_1, e'_2) telles que $\sigma^k(e'_1)\sigma^k(e'_2)$ contiennent w, notamment celles dont les projections par Π_0 des itérés sont les mêmes.

LEMME 4.4. — *On note, pour $k \geqslant 1$:*

$$\mathcal{U}_k = \{e_0, 1\} \cup \Big\{(k - 2q, k - 2p),\ (q, p) \in \mathbb{N}^2 \cap [-1, k] \times [0, k + 1] \setminus \{(-1, k + 1)\},\ p \geqslant q - 1\Big\}.$$

Soit w un mot de m de longueur 2^k. Alors, il existe un couple (e_1, e_2) de $\mathcal{U}_k$ tel que w est un sous-mot de $\sigma^k(e_1)\sigma^k(e_2)$.

Démonstration. — Il existe une paire de $\mathcal{U}_k$ qui convient pour le mot 0^{2^k}, par exemple : $\sigma^k(k)\sigma^k(-k) = 10^{2^{k+1}-2}1$.

Soit w un mot de m de longueur 2^k différent de 0^{2^k}. Par définition de m, il existe au moins une paire (e_1, e_2) pour laquelle w est un sous-mot de $\sigma^k(e_1)\sigma^k(e_2)$.

Lorsque e_1e_2 est différent de e_01, comme w est différent de 0^{2^k}, $\sigma^k(e_1)$ ou $\sigma^k(e_2)$ contient des occurrences de 1 et donc la lettre e_1 ou la lettre e_2 appartient à $[-k, k] \cap \mathbb{N}$ et peut s'écrire $k - 2q$ avec q dans $[0, k] \cap \mathbb{N}$.

Supposons que $e_1 = k - 2q$ avec q dans $[0, k] \cap \mathbb{N}$. D'après le lemme 4.3, on a forcément $e_2 = e_1 - 2p$ avec $p \geqslant -1$ ou $e_2 = -e_1 + 1$ si $e_2 \leqslant 0$. Selon le cas, on obtient des paires de $\mathcal{U}_k$ différentes :

- si $e_2 = e_1 - 2p$ et e_2 est dans $[-k, k] \cap \mathbb{N}$, la paire (e_1, e_2) appartient bien à $\mathcal{U}_k$.
- si $e_2 = -e_1 + 1$ ou $e_2 = e_1 - 2p$ n'est pas dans $[-k, k] \cap \mathbb{N}$, alors $\sigma^k(e_2) = 0^{2^k}$ et donc $\sigma^k(e_2) = \sigma^k(-k - 2)$.

Le lemme 4.3 nous assure que $e_1(-k - 2)$ est bien un mot de $F_2(\overline{m})$ qui vérifie : w est inclus dans $\sigma^k(e_1)\sigma^k(-k - 2)$ et $(e_1, -k - 2)$ appartient à $\mathcal{U}_k$.

Supposons maintenant que $e_2 = k - 2q$ avec q dans $[0, k] \cap \mathbb{N}$. D'après le lemme 4.3, deux cas de figure sont possibles : $e_1 = e_2 + 2p$ avec $p \geqslant -1$ ou $e_1 = -e_2 + 1$, uniquement si $e_2 < 0$. Selon le cas, on obtient des paires de $\mathcal{U}_k$ différentes :

- si $e_1 = e_2 + 2p$ et e_1 est dans $[-k, k] \cap \mathbb{N}$, la paire (e_1, e_2) est dans $\mathcal{U}_k$.
- si $e_1 = -e_2 - 1$ ou $e_1 = e_2 + 2p$ n'est pas dans $[-k, k] \cap \mathbb{N}$, on a $\sigma^k(e_1) = 0^{2^k} = \sigma^k(k + 2)$. Le lemme 4.3 nous assure que $(k + 2)e_2$ est

un mot de $F_2(\overline{m})$ qui vérifie w est inclus dans $\sigma^k(k+2)\sigma^k(e_2)$ et le couple $(k+2, e_2)$ est dans $\mathcal{U}_k$,

On a donc trouvé pour chaque mot w de m, une paire de l'ensemble $\mathcal{U}_k$ qui convient. □

Démonstration du résultat 4.2. — Tout mot de m de longueur 2^k est inclus dans un certain mot $\sigma^k(e_1)\sigma^k(e_2)$ avec (e_1, e_2) dans l'ensemble $\mathcal{U}_k$. On a donc l'inégalité : $p_m(2^k) \leqslant 2^k \cdot \mathrm{Card}(\mathcal{U}_k)$ où 2^k représente le nombre de places différentes où on est susceptible de trouver un mot nouveau de longueur 2^k dans les $\sigma^k(e_1)\sigma^k(e_2)$ et $\mathrm{Card}(\mathcal{U}_k)$ représente le cardinal de $\mathcal{U}_k$. Comme

$$\mathrm{Card}(\mathcal{U}_k) = 1 + (k+1) + (k+2) + \sum_{i=3}^{k+2} i = \frac{k^2+9k+8}{2},$$

on obtient :

$$p_m(2^k) \leqslant 2^k \frac{k^2+9k+8}{2} \quad \text{et} \quad p_m(2^{k+1}) \leqslant 2^{k+1} \frac{k^2+11k+18}{2}.$$

En utilisant le fait que la fonction de complexité est une fonction croissante et que, pour tout entier n de $[2^k, 2^{k+1}[$, $\log_2(n)$ appartient à l'intervalle $[k, k+1[$, on obtient :

$$\forall n \geqslant 1, \quad p_m(n) \leqslant n(\log_2 n + 9)(\log_2 n + 2),$$

c'est-à-dire la majoration annoncée pour $p_m(n)$. □

4.2. Exemple 2

RÉSULTAT 4.5. — *Soit $\overline{\sigma}$ la substitution suivante*

$$\begin{array}{rccc} \overline{\sigma}: & \mathbb{Z} \cup \{e_0\} & \to & \mathbb{Z}^2 \cup \{e_0 2\} \\ & e & \mapsto & (e-1)(e^2) \\ & e_0 & \mapsto & e_0 2. \end{array}$$

Soit m le mot infini engendré par le 2-automate $(\mathbb{Z} \cup \{e_0\}, e_0, \phi_{\overline{\sigma}}, \{0,1\}, \Pi_0)$ où la projection $\Pi_0 : \mathbb{Z} \cup \{e_0\} \to \{0,1\}$ est définie par $\Pi_0^{-1}(\{1\}) = \{0\}$.

La complexité de m admet la majoration suivante :

$$\forall n \geqslant 2, \ p_m(n) \leqslant 2n\big(\log_2 n + \sqrt{\log_2 n + 1} + 3\big)^2.$$

Pour démontrer ce résultat, on introduit les notations suivantes :

– les polynômes $P_0(X) = X - 1$ et $P_1(X) = X^2$,
– l'ensemble de polynômes $\mathcal{F}_k = \big\{P_{i_1} \circ \cdots \circ P_{i_k},\ \forall j \in [1,k],\ i_j \in \{0,1\}\big\}$,
– l'ensemble de nombres entiers $\mathcal{R}_k = \big\{n \in \mathbb{N},\ \exists P \in \mathcal{F}_k,\ P(n) = 0\big\}$.

LEMME 4.6. — *Pour tout $k \geqslant 1$, l'ensemble $\mathcal{R}_k$ est inclus dans l'intervalle $[-\sqrt{k}, k]$.*

En particulier, $\mathrm{Card}(\mathcal{R}_k) \leqslant k + \sqrt{k} + 1$.

Démonstration. — On procède par récurrence forte.

Pour $k = 1$, on a $\mathcal{F}_1 = \{X - 1, X^2\}$ et $\mathcal{R}_1 = \{0, 1\}$, donc l'ensemble $\mathcal{R}_1$ est bien inclus dans l'intervalle $[-1, 1]$.

Supposons maintenant, pour $k \geqslant 2$ et tout $p \leqslant k - 1$, que l'ensemble $\mathcal{R}_p$ soit inclus dans $[-\sqrt{p}, p]$. Soit P un polynôme de $\mathcal{F}_k$. Par définition de l'ensemble $\mathcal{F}_k$, il existe q dans $[1, k]$ et deux q-uplets d'entiers positifs ou nuls $(\alpha_1, \ldots, \alpha_q)$ et $(\beta_1, \ldots, \beta_q)$ tels que :

(1) les couples (α_1, β_1) et (α_q, β_q) sont différents de $(0, 0)$,

(2) pour tout $2 \leqslant i \leqslant q - 1$, α_i et β_i sont non nuls,

(3) $\sum_{i=1}^{q} \alpha_i + \beta_i = k$,

(4) $P = P_0^{\alpha_q} \circ P_1^{\beta_q} \circ \cdots \circ P_0^{\alpha_1} \circ P_1^{\beta_1}$.

On va discuter, selon les q-uplets $(\alpha_1, \ldots, \alpha_q)$ et $(\beta_1, \ldots, \beta_q)$, de l'emplacement des racines entières du polynôme P :

– Si $\alpha_1 = k$, alors $P(X) = X - k$ donc P a une unique racine entière k.

– Si $\beta_1 = k$, alors $P(X) = X^{2^k}$ donc P n'a qu'une racine entière 0.

– Si $\alpha_1 + \beta_1 = k$, avec $\alpha_1 \geqslant 1$ et $\beta_1 \geqslant 1$ (dans ce cas, $q = 1$) alors $P(X) = X^{2^{\beta_1}} - \alpha_1$. Le polynôme P a deux racines réelles $\pm \sqrt[2^{\beta_1}]{\alpha_1}$ qui ne sont pas forcément des valeurs entières, mais qui, dans tous les cas sont contenues dans l'intervalle $[-\sqrt{k}, \sqrt{k}]$.

Pour tous les autres q-uplets $(\alpha_1, \ldots, \alpha_q)$ et $(\beta_1, \ldots, \beta_q)$ vérifiant les conditions ci-dessus, on va pouvoir utiliser l'hypothèse de récurrence : dire que l'entier n est racine du polynôme $P = P_0^{\alpha_q} \circ P_1^{\beta_q} \circ \cdots \circ P_0^{\alpha_1} \circ P_1^{\beta_1}$ revient à dire que $P_0^{\alpha_1} \circ P_1^{\beta_1}(n)$ est racine du polynôme $Q = P_0^{\alpha_q} \circ P_1^{\beta_q} \circ \cdots \circ P_0^{\alpha_2} \circ P_1^{\beta_2}$. Or Q est un polynôme de l'ensemble $\mathcal{F}_{k-(\alpha_1+\beta_1)}$, et donc

$$-\sqrt{k - (\alpha_1 + \beta_1)} \leqslant n^{2^{\beta_1}} - \alpha_1 \leqslant k - (\alpha_1 + \beta_1).$$

Ainsi, pour tous les couples (α_1, β_1) tels que $\alpha_1 + \beta_1 < k$, l'entier n appartient à $[-\sqrt{k}, k]$. On a montré que, quel que soit le polynôme P de $\mathcal{F}_k$, ses racines sont dans l'intervalle $[-\sqrt{k}, k]$, ceci est donc, en particulier, vrai pour les racines entières des éléments de $\mathcal{F}_k$, c'est-à-dire les éléments de $\mathcal{R}_k$. □

Démonstration du résultat 4.5. — On va majorer, dans un premier temps, $p_m(2^k)$ pour un k fixé, puis étendre cette majoration à tout $n > 0$.

Soit $\overline{m}$ le mot infini point fixe de $\overline{\sigma}$ contenu dans $e_0\mathbb{Z}^\omega$.

La majoration de la complexité de m est basée sur l'étude des mots $\sigma^k(e)$, pour e dans $F_1(\overline{m})$, c'est-à-dire pour e dans $\mathbb{Z}$, qui sont différents de 0^{2^k}.

Un mot donné w de m de longueur 2^k provient, par définition, d'un mot $\overline{w}$ de $\overline{m}$ de longueur 2^k. Ce mot $\overline{w}$ est un sous-mot d'un certain $\overline{\sigma}^k(e_1)\overline{\sigma}^k(e_2)$. On va majorer le nombre de $\overline{\sigma}^k(e)$ qui se projettent sur des mots différents de 0^{2^k}.

En conséquence directe du lemme 3.1, on a l'équivalence :

$$\begin{aligned}1 \text{ apparaît dans } \sigma^k(e) \Leftrightarrow 0 \text{ apparaît dans } \overline{\sigma}^k(e)\\ \Leftrightarrow e \in \bigcup_{(j_1,\dots,j_k)\in\{0,1\}^k} P_{j_1}^{-1}\circ\cdots\circ P_{j_k}^{-1}\{0\},\end{aligned}$$

ce que l'on peut traduire par

$$1 \text{ apparaît dans } \sigma^k(e) \iff e \in \mathcal{R}_k.$$

Ainsi, on obtient l'inégalité :

$$\operatorname{Card}\left(\{e\in\mathbb{Z},\ 1 \text{ apparaît dans } \sigma^k(e)\}\right) \leqslant \operatorname{Card}(\mathcal{R}_k).$$

On a donc au plus $k+\sqrt{k}+1$ mots du type $\overline{\sigma}^k(e)$ dont la projection par Π_0 est différente de 0^{2^k}, donc $k+\sqrt{k}+2$ possibilités de projections différentes pour $\overline{\sigma}^k(e_1)$ et $\overline{\sigma}^k(e_2)$.

Comme chaque projection $\sigma^k(e_1)\sigma^k(e_2)$ peut contenir au plus 2^k nouveaux mots différents de longueur 2^k, on a la majoration suivante :

$$\forall k>1,\ p_m(2^k) \leqslant 2^k(k+\sqrt{k}+2)^2.$$

Comme, pour $2^k \leqslant n < 2^{k+1}$, on a $p_m(2^k) \leqslant p_m(n) \leqslant p_m(2^{k+1})$, on obtient :

$$\forall n \geqslant 1, 2^k \leqslant n < 2^{k+1},\ p_m(n) \leqslant 2^{k+1}(k+\sqrt{k+1}+3)^2,$$

et donc, en utilisant le fait que, pour tout entier n contenu dans $[2^k, 2^{k+1}[$, on a $k \leqslant \log_2 n < k+1$:

$$\forall n \geqslant 1,\ p_m(n) \leqslant 4n\left((\log_4 n+1)(\log_4 n+3)(\log_4 n+5)+4\right),$$

ce qui termine la preuve. □

4.3. Exemple 3

RÉSULTAT 4.7. — *On considère la substitution associée à la «marche aléatoire» sur* $\mathbb{Z}^2$ *:*

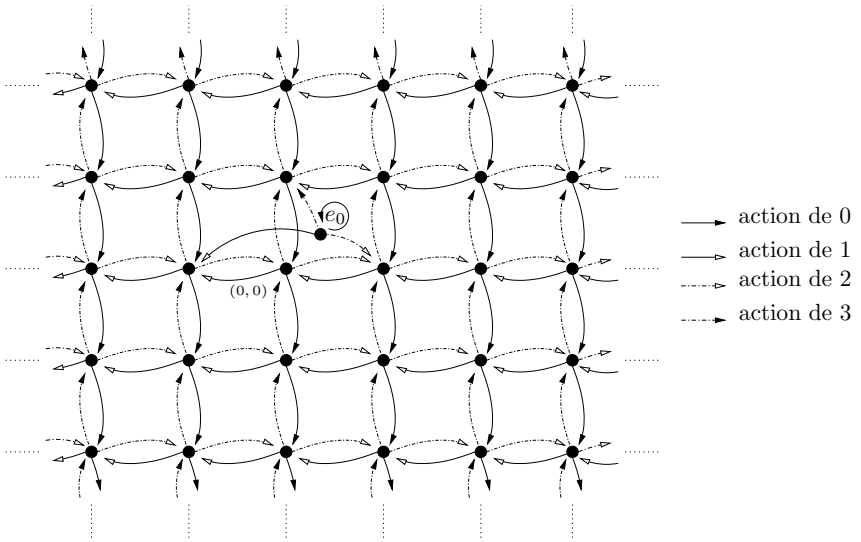
$$\begin{array}{rrcl} \overline{\sigma}: & \mathbb{Z}^2 \cup \{e_0\} & \to & \left(\mathbb{Z}^2 \cup \{e_0\}\right)^4 \\ & e = (a,b) \neq e_0 & \mapsto & (a,b-1)(a-1,b)(a+1,b)(a,b+1) \\ & e_0 & \mapsto & e_0(-1,0)(1,0)(0,1) \end{array}$$

Soit m *le mot infini engendré par le* 4*-automate* $\left(\mathbb{Z}^2 \cup \{e_0\}, e_0, \phi_{\overline{\sigma}}, \{0,1\}, \Pi_0\right)$ *où la projection* $\Pi_0 : \mathbb{Z} \cup \{e_0\} \to \{0,1\}$ *est définie par* $\Pi_0^{-1}(\{1\}) = \{(0,0)\}$.

La complexité du mot m *admet la majoration suivante :*

$$\forall n \geqslant 1, \quad p_m(n) \leqslant 2n \frac{(\log_4 n + 1)}{3} \left(4 \log_2^2 n + 23 \log_4 n + 6\right).$$

Le graphe de l'automate associé à la substitution $\overline{\sigma}$ est donné par :



LEMME 4.8. — *Soit* $\overline{m}$ *le mot infini point fixe de* $\overline{\sigma}$ *contenu dans* $e_0 \left(\mathbb{Z}^2\right)^\omega$. *Le mot* e_1e_2 *est dans* $F_2(\overline{m})$ *si et seulement si un des* 4 *cas suivants se produit :*

(1) $e_1e_2 = e_0(-1,0)$,

(2) $\exists k' \geqslant 0$ *tel que* $e_1 = (-1,k')$ *et* $e_2 = (1,-k')$ *ou* $e_1 = (1,k')$ *et* $e_2 = (0,1-k')$,

(3) $\exists k' \geqslant 0$, $e_1 = (0,k'+1)$ *et* $e_2 = (-1,-k'-1)$,

(4) $\exists k' \geqslant 0$, $e_1 = (a,b) \in F_1(\overline{m})$ *et* $e_2 = (a-1, b-2k'+1)$ *ou* $(a+2, b-2k')$.

Ce lemme est essentiellement une réécriture de la proposition 3.3. L'égalité du point 4. provient du quatrième cas de la proposition 3.3, avec les simplifications suivantes :
$\overline{\sigma}_1 \circ \overline{\sigma}_0^{-1}((a,b)) = \overline{\sigma}_3 \circ \overline{\sigma}_2^{-1}((a,b)) = (a-1, b+1)$ et $\overline{\sigma}_2 \circ \overline{\sigma}_1^{-1}((a,b)) = (a+2, b)$.

Démonstration du résultat 4.7. — Soit $k \geqslant 2$ un entier fixé. On va majorer la complexité au rang 4^k.

Tout mot de m de longueur 4^k est inclus dans un mot de longueur 4^{k+1} du type $\sigma^k(e_1)\sigma^k(e_2)$, on va donc majorer le nombre de mots de deux lettres e_1e_2 tels que $\sigma^k(e_1)\sigma^k(e_2)$ est différent de $0^{4^{k+1}}$. Pour cela, on va distinguer deux cas :

(a) un seul des deux mots $\sigma^k(e_1)$ et $\sigma^k(e_2)$ contient des occurrences de 1,

(b) les deux mots $\sigma^k(e_1)$ et $\sigma^k(e_2)$ contiennent des occurrences de 1.

Dans le cas (b), pour majorer le nombre de mots e_1e_2, on distinguera ensuite six sous-cas, en accord avec le lemme 4.8.

On commence par remarquer que le mot $\overline{\sigma}^k(e)$, pour un élément $e = (a,b)$ de $\mathbb{Z}^2$, contient des occurrences de 0 si et seulement si $|a| + |b| \leqslant k$ et que $(a+b)$ est de même parité que k ; d'autre part, $\overline{\sigma}^k(e_0)$ contient aussi des occurrences de 0 uniquement quand k est pair. Cette «perturbation», due à l'état initial e_0 nécessite une disjonction des cas selon la parité de k. Cependant, comme l'élément e_0 n'intervient que rarement dans la démonstration, la disjonction k pair/k impair sera intégrée aux différents sous-cas uniquement lorsqu'elle sera nécessaire, pour éviter d'alourdir d'avantages la preuve. On retiendra néanmoins que

$$\mathcal{Z}_k((0,0)) = \begin{cases} \{(a,b) \in \mathbb{Z}^2,\ |a|+|b| \leqslant k, a+b \text{ impair}\} & \text{si } k \text{ est impair,} \\ \{(a,b) \in \mathbb{Z}^2,\ |a|+|b| \leqslant k, a+b \text{ pair}\} \cup \{e_0\} & \text{si } k \text{ est pair.} \end{cases}$$

et aussi :

$$\operatorname{Card}(\mathcal{Z}_k((0,0))) = \begin{cases} (k+1)^2 & \text{si } k \text{ est impair,} \\ (k+1)^2 + 1 & \text{si } k \text{ est pair.} \end{cases}$$

On peut donc maintenant majorer le nombre de mots e_1e_2 en distinguant les deux cas annoncés :

(a) Majoration du nombre de mots e_1e_2 tels que seulement un des mots $\sigma^k(e_1)$ ou $\sigma^k(e_2)$ contient des occurrences de 1 :

Lorsque k est impair, comme il existe $(k+1)^2$ éléments différents dans $\mathcal{Z}_k((0,0))$. On peut donc construire de cette manière, au pire, $2(k+1)^2$ mots $\sigma^k(e_1)\sigma^k(e_2)$ différents entre eux et différents de $0^{4^{k+1}}$.

Lorsque k est pair, on peut former $(k+1)^2+1$ mots du type $\sigma^k(e_1)$ contenant des occurrences de 1 et seulement $(k+1)^2$ mots du type $\sigma^k(e_2)$ contenant des occurrences de 1, puisque la lettre e_0 apparaît une seule et unique fois au début du mot $\overline{m}$. On peut donc construire de cette manière, au pire, $2(k+1)^2+1$ mots $\sigma^k(e_1)\sigma^k(e_2)$ différents entre eux et différents de $0^{4^{k+1}}$.

On pose

$$N_0(k) = \begin{cases} 2(k+1)^2 & \text{si } k \text{ est impair,} \\ 2(k+1)^2+1 & \text{si } k \text{ est pair.} \end{cases}$$

(b) Majoration du nombre de mots e_1e_2 tels que les deux mots $\sigma^k(e_1)$ et $\sigma^k(e_2)$ contiennent des occurrences de 1 :

Le lemme 4.8 permet les de classer les mots de deux lettres e_1e_2 de $\overline{m}$ de la manière suivante :

1. $e_1e_2 = e_0(-1,0)$,
2. $\exists k' \geqslant 0$ tel que $e_1 = (-1,k')$ et $e_2 = (1,-k')$,
3. $\exists k' \geqslant 0$ tel que $e_1 = (1,k')$ et $e_2 = (0,1-k')$,
4. $\exists k' \geqslant 0$, $e_1 = (0,k'+1)$ et $e_2 = (-1,-k'-1)$,
5. $\exists k' \geqslant 0$, $e_1 = (a,b) \in F_1(\overline{m})$ et $e_2 = (a-1,b-2k'+1)$,
6. $\exists k' \geqslant 0$, $e_1 = (a,b) \in F_1(\overline{m})$ et $(a+2,b-2k')$.

En explorant ces différents cas de figure, on va majorer le nombre de mots e_1e_2 de $F_2(\overline{m})$ tels que les deux itérés $\sigma^k(e_1)$ et $\sigma^k(e_2)$ contiennent des occurrences de 1 dans chacun des six cas ci-dessus.

1. Si $e_1e_2 = e_0(-1,0)$, les deux itérés $\sigma^k(e_1)$ et $\sigma^k(e_2)$ ne peuvent pas contenir simultanément des occurrences de 1 pour k fixé. En effet, $\sigma^k(e_0)$ contient des occurrences de 1 uniquement lorsque k est pair et $\sigma^k((-1,0))$ contient des occurrences de 1 uniquement lorsque k est impair. Par conséquent, le mot $e_1e_2 = e_0(-1,0)$ ne donne pas d'itéré $\sigma^k(e_1)\sigma^k(e_2)$ différent de $0^{4^{k+1}}$ pour lequel $\sigma^k(e_1)$ et $\sigma^k(e_2)$ contiennent des occurrences de 1.

2. S'il existe $k' \geqslant 0$ tel que $e_1 = (-1,k')$ et $e_2 = (1,-k')$, alors chaque k' de $[0,k] \cap \mathbb{N}$ crée un mot e_1e_2 de $F_2(\overline{m})$, mais les deux itérés $\sigma^k(e_1)$ et $\sigma^k(e_2)$ contiennent des occurrences de 1 uniquement lorsque k' est de parité différente de celle que k. On peut donc former de cette manière, au pire, $N_2(k)$ mots $\sigma^k(e_1)\sigma^k(e_2)$ différents et différents de $0^{4^{k+1}}$ pour lesquels

$\sigma^k(e_1)$ et $\sigma^k(e_2)$ contiennent des occurrences de 1, avec

$$N_2(k) = \begin{cases} (k+1)/2 & \text{si } k \text{ est impair,} \\ k/2 & \text{si } k \text{ est pair.} \end{cases}$$

3. S'il existe $k' \geqslant 0$ tel que $e_1 = (1, k')$ et $e_2 = (0, 1-k')$, alors chaque k' de $[0,k] \cap \mathbb{N}$ crée un mot e_1e_2 de $F_2(\overline{m})$, mais les deux itérés $\sigma^k(e_1)$ et $\sigma^k(e_2)$ contiennent des occurrences de 1 uniquement lorsque k' est de parité différente de celle que k. On peut donc former de cette manière, au pire, $N_3(k)$ mots $\sigma^k(e_1)\sigma^k(e_2)$ différents et différents de $0^{4^{k+1}}$ pour lesquels $\sigma^k(e_1)$ et $\sigma^k(e_2)$ contiennent des occurrences de 1, avec

$$N_3(k) = \begin{cases} (k+1)/2 & \text{si } k \text{ est impair,} \\ k/2 & \text{si } k \text{ est pair.} \end{cases}$$

4. S'il existe $k' \geqslant 0$ tel que $e_1 = (0, k'+1)$ et $e_2 = (-1, -k'-1)$, alors, comme $(k'+1)$ et $(-k'-2)$ sont de parité différentes quel que soit k', les itérés $\sigma^k(e_1)$ et $\sigma^k(e_2)$ ne peuvent pas contenir tous les deux des occurrences de 1. On ne peut donc former de cette manière, aucun mot $\sigma^k(e_1)\sigma^k(e_2)$ différent de $0^{4^{k+1}}$ pour lequel $\sigma^k(e_1)$ et $\sigma^k(e_2)$ contiennent des occurrences de 1.

5. Supposons que l'on puisse écrire le mot e_1e_2 de manière à ce que, pour un certain $k' \geqslant 0$, $e_2 = (a-1, b-2k'+1)$. Puisque e_1 et e_2 doivent être dans $\mathcal{Z}_k((0,0))$ (pour que les itérés $\sigma^k(e_1)$ et $\sigma^k(e_2)$ contiennent des occurrences de 1), on peut donc majorer le nombre de mots e_1e_2 de ce type par le cardinal, noté $N_5(k)$, de l'ensemble :

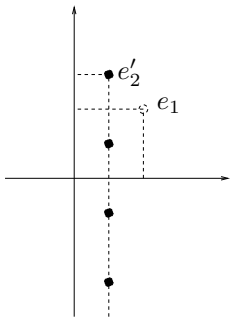
$$\mathcal{U}_5(k) = \Big\{(a,b,k') \in \mathbb{Z}^2 \times \mathbb{N},\ \big((a,b),(a-1,b-2k'+1)\big) \in \mathcal{Z}_k((0,0))^2\Big\}.$$

Pour calculer le cardinal de cet ensemble, on utilise le fait que pour $e_1 = (a,b)$ fixé dans $\mathcal{Z}_k((0,0))$, les éléments du type $e_2 = (a-1, b-2k'+1)$ sont les points entiers de la demi-droite verticale issue du point $e_2' = (a-1, b+1)$:

De cette manière, lorsque $e_1 = (a,b)$ décrit $\mathcal{Z}_k((0,0))$, en dénombrant à chaque fois les points $e_2 = (a-1, b-2k'+1)$ qui sont aussi dans $\mathcal{Z}_k((0,0))$, on obtient la formule suivante pour $N_5(k)$:

$$N_5(k) = \sum_{i=1}^{k} \left((i+1)(k+1-i) + \sum_{j=0}^{i} j \right) + \sum_{j=1}^{k} j = \frac{k(k+4)(k-1)}{3}.$$

Ce calcul s'effectue en regroupant les éléments e_1 de $\mathcal{Z}_k((0,0))$ qui appartiennent à une même diagonale d'équation $y = x + 2i - (k+2)$, pour un certain $i \in \{1, \ldots, k+1\}$. Lorsque $i \leqslant k$ et a est positif ou nul, on a $(i+1)$



éléments du type $e_2 = (a-1, b-2k'+1)$ qui conviennent et lorsque a est négatif, $(i+a+1)$ éléments du type $e_2 = (a-1, b-2k'+1)$ conviennent. Lorsque $i = k+1$, alors a est négatif ou nul et, pour tout $a \in \{-(k), \ldots, 0\}$, on a $k-a$ éléments du type $e_2 = (a-1, b-2k'+1)$ qui conviennent.

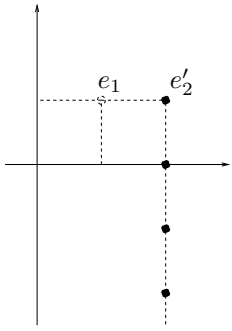
On peut donc former de cette manière, au pire, $\frac{k(k+4)(k-1)}{3}$ mots $\sigma^k(e_1)\sigma^k(e_2)$ différents et différents de $0^{4^{k+1}}$ pour lesquels $\sigma^k(e_1)$ et $\sigma^k(e_2)$ contiennent des occurrences de 1.

6. Il reste à s'intéresser aux couples e_1e_2 qui sont dans $\mathcal{Z}_k((0,0))^2$ et tels que $e_1 = (a,b)$ et $e_2 = (a+2, b-2k')$ pour un certain $k' \geqslant 0$.

On peut majorer le nombre de mots e_1e_2 de ce type par le cardinal, noté $N_6(k)$, de l'ensemble :

$$\mathcal{U}_6(k) = \left\{(a,b,k') \in \mathbb{Z}^2 \times \mathbb{N},\ ((a,b),(a+2,b-2k')) \in \mathcal{Z}_k((0,0))^2\right\}.$$

Pour calculer le cardinal de cet ensemble, on utilise le fait que pour $e_1 = (a,b)$ fixé dans $\mathcal{Z}_k((0,0))$, les éléments du type $e_2 = (a+2, b-2k')$ sont les points entiers de la demi-droite verticale issue du point $e_2' = (a+2, b)$:



De cette manière, lorsque $e_1 = (a, b)$ décrit $\mathcal{Z}_k((0,0))$, en dénombrant à chaque fois les points $e_2 = (a+2, b-2k')$ qui sont aussi dans $\mathcal{Z}_k((0,0))$, on obtient la formule suivante pour $N_6(k)$:

$$N_6(k) = k + \sum_{i=3}^{k+1} \left((i-2) + (i-1)(k+2-i) + \sum_{j=2}^{i-1} j \right) = \frac{k(k+5)(2k-1)}{6}.$$

Ce calcul s'effectue en regroupant les éléments e_1 de $\mathcal{Z}_k((0,0))$ qui appartiennent à une même diagonale d'équation $y = x + 2i - (k+2)$, pour un certain $i \in \{1, \ldots, k+1\}$. Lorsque $i = 1$, aucun élément de type $e_2 = (a+2, b-2k')$ n'appartient à $\mathcal{Z}_k((0,0))$, lorsque $i = 2$ et que e_1 n'appartient pas à la droite d'équation $y = -x + k$, un seul élément de type $e_2 = (a+2, b-2k')$, se situant sur la diagonale d'équation $y = x - k$ appartient à $\mathcal{Z}_k((0,0))$. Lorsque $i \geqslant 3$, on a trois cas de figures : si e_1 appartient à la droite d'équation $y = -x + k$, alors $(i-2)$ éléments du type $e_2 = (a-1, b-2k'+1)$ conviennent, si $a \geqslant -1$, alors $(i-1)$ points $e_2 = (a-1, b-2k'+1)$ conviennent et lorsque $a \leqslant -2$, alors on a $(i+a+1)$ éléments du type $e_2 = (a-1, b-2k'+1)$ qui conviennent.

On peut donc former de cette manière, au pire, $\frac{k(k+5)(2k-1)}{6}$ mots $\sigma^k(e_1)\sigma^k(e_2)$ différents et différents de $0^{4^{k+1}}$ pour lesquels $\sigma^k(e_1)$ et $\sigma^k(e_2)$ contiennent des occurrences de 1.

Il résulte des différentes majorations obtenues en (a) et (b) que :

$$\forall k \geqslant 1,\ p_m(4^k) \leqslant 4^k \left(N_0(k) + N_2(k) + N_3(k) + N_5(k) + N_6(k) \right).$$

Puisque $N_0(k) + N_2(k) + N_3(k) = 2(k+1)(k+2)$ quelque soit la parité de k, il suit :

$$\forall k \geqslant 1,\ p_m(4^k) \leqslant 4^k \Big(2(k+1)(k+2) + \frac{k(k+4)(k-1)}{3} + \frac{k(k+5)(2k-1)}{6} \Big).$$

On obtient donc :

$$p_m(4^k) \leqslant 4^k \frac{k}{6}(4k^2 + 15k - 13)$$

et

$$p_m(4^{k+1}) \leqslant 4^{k+1} \frac{(k+1)}{6}(4k^2 + 23k + 6).$$

En utilisant le fait que p_m est une fonction croissante et que, pour tout entier n de $[4^k, 4^{k+1}[$, $\log_4 n$ appartient à $[k, k+1[$, on obtient :

$$\forall n \geqslant 1,\ \ p_m(n) \leqslant 2n \frac{(\log_4 n + 1)}{3} \left(4 \log_2^2 n + 23 \log_4 n + 6 \right).$$

c'est-à-dire la majoration annoncée pour $p_m(n)$. □

4.4. Exemple 4

RÉSULTAT 4.9. — *Soit $\overline{\sigma}$ la substitution suivante*

$$\begin{array}{rccc} \overline{\sigma}: & \mathbb{N} & \to & \mathbb{N}^2 \\ & e & \mapsto & \left(\lfloor \frac{e}{2} \rfloor\right)(e+1) \end{array}$$

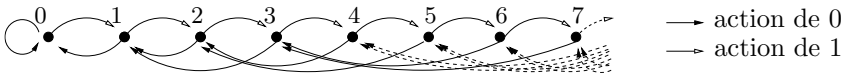
Soit m le mot infini engendré par le 2-automate $(\mathbb{N}, 0, \phi_{\overline{\sigma}}, \{0,1\}, \Pi_0)$ où la projection $\Pi_0 : \mathbb{N} \to \{0,1\}$ est définie par $\Pi_0^{-1}(\{1\}) = \{0\}$.

La complexité de m admet la majoration suivante :

$$\forall n \geqslant 1,\ p_m(n) \leqslant 2n\big(n(8\log_2 n - 1) + 1\big).$$

La fonction $e \mapsto \lfloor e \rfloor$ représente la partie entière de e.

Le graphe de l'automate associé à la substitution est donné par :



Si on pose $\overline{\sigma}_0(e) = \lfloor \frac{e}{2} \rfloor$ et $\overline{\sigma}_1(e) = e+1$, la proposition 3.3 devient :

LEMME 4.10. — *Soit $\overline{m}$ le mot infini point fixe de $\overline{\sigma}$ contenu dans $0\mathbb{N}^\omega$. Le mot e_1e_2 est dans $F_2(\overline{m})$ si et seulement si un des 3 cas suivants se produit :*

1. $e_1e_2 = 01$,
2. $e_1 \in \mathbb{N}^*$ *et* $e_2 = 0$,
3. $\exists k' \geqslant 0, e_1 \in \mathbb{N}$ *et* $e_2 \in \overline{\sigma}_0^{k'} \circ \overline{\sigma}_1 \circ \overline{\sigma}_0^{-1} \circ \overline{\sigma}_1^{-k'}(\{e_1\})$.

Démonstration. — C'est une réécriture des différents cas de la proposition 3.3 : le premier cas du lemme 4.10 correspond au premier cas de la proposition 3.3. Le deuxième cas de la proposition 3.3 disparaît car on a ici une substitution de longueur égale à 2.

Le troisième cas de la proposition 3.3 devient, puisque ici $e_{q-1} = e_1$: il existe $k \geqslant 0$ tel que $e_1e_2 = \overline{\sigma}_1^k(1)\overline{\sigma}_0^{k+1}(1)$. Or, pour tout $k \geqslant 0$, $\overline{\sigma}_1^k(1) = k+1$ et $\overline{\sigma}_0^{k+1}(1) = 0$. Ce qui correspond au deuxième cas du lemme 4.10.

Enfin, le dernier cas du lemme correspond au quatrième cas de la proposition 3.3. □

LEMME 4.11. — *Soit $k \geqslant 1$ fixé.*

L'ensemble $\mathcal{Z}_k(0) = \bigcup_{(j_1,\ldots,j_k)\in\{0,1\}^k} \overline{\sigma}_{j_1}^{-1} \circ \cdots \circ \overline{\sigma}_{j_k}^{-1}(\{0\})$ est exactement $[0, 2^k - 1] \cap \mathbb{N}$.

En particulier quelque soit $p \in [0, 2^{k-1} - 1] \cap \mathbb{N}$, $\sigma^k(2p)$ et $\sigma^k(2p+1)$ ont le même préfixe de longueur 2^{k-1}, donné par $\sigma^{k-1}(p)$ et on a :

$$\forall p \in [2^{k-2}, 2^{k-1} - 1] \cap \mathbb{N},\ \sigma^k(2p) = \sigma^k(2p+1).$$

Démonstration. — Soit $k \geqslant 1$ fixé.

La plus petite lettre de $\overline{\sigma}^k(e)$ est sa première lettre, à savoir $\overline{\sigma}_0^k(e)$. $\overline{\sigma}_0$ étant croissante, $\overline{\sigma}_0^k$ l'est aussi. Comme $\overline{\sigma}_0^k(2^k-1)=0$ et $\overline{\sigma}_0^k(2^k)=1$, Si e est strictement plus grand que 2^k-1, $\overline{\sigma}^k(e)$ ne contient pas d'occurrence de 0, et donc $\mathcal{Z}_k(0)$ est inclus dans $[0,2^k-1]\cap\mathbb{N}$.

La deuxième partie du lemme se montre en utilisant les égalités suivantes :

$$\sigma^k(2p)=\sigma^{k-1}(p)\sigma^{k-1}(2p+1) \text{ et } \sigma^k(2p+1)=\sigma^{k-1}(p)\sigma^{k-1}(2p+2),$$

et en remarquant que pour $p\in[2^{k-2},2^{k-1}-1]\cap\mathbb{N}$, les mots $\sigma^{k-1}(2p+1)$ et $\sigma^{k-1}(2p+2)$ sont tous les deux le mot $0^{2^{k-1}}$. □

Démonstration du résultat 4.9. — Soit $k \geqslant 1$ fixé.

On va majorer le nombre de mots e_1e_2 de $F_2(\overline{m})$ dont les projections $\sigma^k(e_1)\sigma^k(e_2)$ ne sont pas le mot nul $0^{2^{k+1}}$. Pour cela, on va distinguer deux cas :

(a) un seul des deux mots $\sigma^k(e_1)$ et $\sigma^k(e_2)$ contient des occurrences de 1,

(b) les deux mots $\sigma^k(e_1)$ et $\sigma^k(e_2)$ contiennent des occurrences de 1.

Dans le cas (b), pour majorer le nombre de mots e_1e_2, on distinguera ensuite trois sous-cas, en accord avec le lemme 4.10.

(a) Si un seul des mots $\sigma^k(e_1)$ ou $\sigma^k(e_2)$, contient des occurrences de 1, en tenant compte du lemme 4.11, on peut fabriquer au plus $2(2^{k-1}+\frac{2^k-2^{k-1}}{2})=32^{k-1}$ mots $\sigma^k(e_1)\sigma^k(e_2)$ différents. Les mots de m de longueur 2^k contenus dans de tels $\sigma^k(e_1)\sigma^k(e_2)$ sont donc au plus : $2(2^{k-1}2^{k-1}+2^{k-1}\cdot 32^{k-1})$, c'est-à-dire, au plus, $5\cdot 2^{2k-2}$, en dissociant les mots de longueur 2^k qui commencent avant la 2^{k-1}-ième lettre de $\sigma^k(e_1)\sigma^k(e_2)$ et ceux qui commencent après, et en utilisant la deuxième partie du lemme précédent.

(b) Si $\sigma^k(e_1)$ et $\sigma^k(e_2)$ contiennent des occurrences de 1, c'est-à-dire e_1 et e_2 sont dans $[0,2^k-1]\cap\mathbb{N}$. En utilisant le lemme 4.10, on a plusieurs possibilités de mots e_1e_2 :

1. $e_1\in[0,2^k-1]\cap\mathbb{N}$ et $e_2=0$,
2. $\exists k'>0, e_1\in[0,2^k-1]\cap\mathbb{N}$ et $e_2\in\overline{\sigma}_0^{k'}\circ\overline{\sigma}_1\circ\overline{\sigma}_0^{-1}\circ\overline{\sigma}_1^{-k'}(\{e_1\})$.
3. $e_1\in\mathbb{N}$ et $e_2\in\overline{\sigma}_1\circ\overline{\sigma}_0^{-1}(\{e_1\})$.

1. La première situation donne toujours des mots différents de $0^{2^{k+1}}$.

On peut donc fabriquer, en tenant compte du lemme 4.11, au plus, $2^{k-1}+2^{k-2}$ itérés différents $\sigma^k(e_1)\sigma^k(e_2)$ de ce type, ce qui permet, au pire, de fabriquer $2^k(2^{k-1}+2^{k-2})$ mots différents de longueur 2^k de cette manière.

2. Dans le deuxième cas de figure, on a toujours des lettres non nulles dans les 2 itérés $\sigma^k(e_i)$ car la fonction $\overline{\sigma}_0^{k'} \circ \overline{\sigma}_1 \circ \overline{\sigma}_0^{-1} \circ \overline{\sigma}_1^{-k'}$ est décroissante pour $k' > 0$.

Supposons que $e_1 \in [0, 2^k - 1] \cap \mathbb{N}$. Cette situation donne des mots valides uniquement lorsque l'on peut appliquer $\overline{\sigma}_1^{-k'}$ à e_1, c'est-à-dire lorsque $e_1 \in [0, 2^k - 1] \cap \mathbb{N}$ et $k' \leqslant e_1$.

D'autre part, si $e_1 \in [2^\alpha, 2^{\alpha+1}[$ pour $\alpha < k$, on a au plus α possibilités pour e_2, lorsque k' varie. En effet, pour tout $0 < k' \leqslant e_1$, $e_2 = \overline{\sigma}_0^{k'-1}(e_1 - k')$ ou $e_2 = \overline{\sigma}_0^{k'-1}(e_1 - k' + 1)$, donc si $e_1 \in [2^\alpha, 2^{\alpha+1}[$, pour tout $k' \geqslant \alpha + 1$, $e_2 = 0$, ces mots ont été précédemment comptés (1.), par conséquent, on a au plus 2α lettres admissibles $e_2 \neq 0$ à la suite de e_1.

On obtient donc de cette manière, au plus $\sum_{\alpha=0}^{k-1}(2\alpha)2^\alpha = (k-2)2^{k+2} + 2$ mots différents de $0^{2^{k+1}}$ possibles, qui nous permettent de fabriquer au plus $2^k((k-2)2^{k+2} + 2)$ mots de longueur 2^k dans les itérés de ce type.

3. Pour $e_1 \in [0, 2^k - 1] \cap \mathbb{N}$, on a deux antécédents par $\overline{\sigma}_0$: $2e_1$ et $2e_1 + 1$ dans $[0, 2^k - 2] \cap \mathbb{N}$. Si e_2 est aussi dans $[0, 2^k - 1] \cap \mathbb{N}$, il faut donc choisir e_1 dans $[0, 2^{k-1} - 1] \cap \mathbb{N}$ et, dans ce cas, on a deux possibilités pour e_2, lorsque e_1 est fixé dans $[0, 2^{k-1} - 1[\cap \mathbb{N}$ et un seul mot différent de $0^{2^{k+1}}$ pour $e_1 = 2^{k-1} - 1$.

Ainsi, on peut construire de cette manière au plus $2^k(2^k - 1)$ mots de taille 2^k.

On peut donc majorer $p_m(2^k)$ par la somme de tous les mots de longueur 2^k qu'on a pu fabriquer, pour k fixé :

$$p_m(2^k) \leqslant \left(5 \cdot 2^{2k-2} + 2^k(2^{k-1} + 2^{k-2}) + 2^k((k-2)2^{k+2} + 2) + 2^k(2^k - 1)\right).$$

et en simplifiant :

$$\forall k \geqslant 1, p_m(2^k) \leqslant 2^k\left((8k - 9)2^{k-1} + 1\right).$$

Comme la fonction de complexité est une fonction croissante, on obtient :

$$\forall 2^k \leqslant n < 2^{k+1},\ p_m(n) \leqslant 2^{k+1}\left((8k - 1)2^k + 1\right).$$

Et ensuite,

$$\forall n \geqslant 1,\ p_m(n) \leqslant 2n\left(n(8\log_2 n - 1) + 1)\right).$$

En utilisant le fait que, pour tout entier n de $[2^k, 2^{k+1}[$, on a $k \leqslant \log_2 n < k + 1$. □

5. Majoration de la complexité des mots infinis engendrés par un q-automate dénombrable de degré borné

Définition 5.1. — *Une substitution $\overline{\sigma}$ définie par $\overline{\sigma}(e) = \overline{\sigma}_0(e)\overline{\sigma}_1(e)\cdots\overline{\sigma}_{q-1}(e)$ sur un alphabet dénombrable E est* K-uniformément bornée *s'il existe une constante $K \geqslant 1$ telle que :*

$$\forall j \in \{0,1,\ldots,q-1\},\ \forall e \in E,\ \mathrm{Card}\left(\overline{\sigma}_j^{-1}\{e\}\right) \leqslant K. \tag{5.1}$$

Les substitutions du paragraphe 4 données dans les exemples 1 et 2 sont 2-uniformément bornées, la substitution de l'exemple 3 est 4-uniformément bornée, la substitution de l'exemple 4 est 3-uniformément bornée.

La condition (5.1) est équivalente, au niveau du graphe de l'automate associé à $\overline{\sigma}$, par le fait que le nombre de flèches entrantes en chaque état est uniformément borné par K, ainsi, les automates associés aux substitutions de longueur constante K-uniformément bornées sont exactement les automates dont le degré de chaque état (somme du nombre de flèches entrantes et du nombre de flèches sortantes) est borné par $K+q$.

Définition 5.2. — *On dit qu'un mot infini m est* engendré par un q-automate dénombrable de degré borné *s'il est l'image, par une projection admissible du point fixe d'une substitution de longueur constante sur un alphabet dénombrable uniformément bornée.*

Notation 5.3. — Pour Π projection admissible de E sur A, 0n note $|\Pi|$ le minimum de l'ensemble des $\mathrm{Card}(F)$, pour F parcourant l'ensemble est sous-ensembles finis F de E tels que la restriction de Π à $E \smallsetminus F$ est constante.

Nous donnons maintenant une majoration générale de la complexité des mots infinis engendrés par des q-automates dénombrables de degré borné.

Théorème 5.4. — *Soit E un alphabet dénombrable et A un alphabet fini.*

Soient $\overline{\sigma}$ une substitution K-uniformément bornée et Π une projection admissible de E vers A.

Soit m le mot infini engendré par le q-automate $(E, e_0, \phi_{\overline{\sigma}}, A, \Pi)$.

La complexité de m est au plus polynomiale :

$$\forall n \geqslant 2,\ p_m(n) \leqslant qn\Big(|\Pi| qK n^{1+\log_q K} + 1\Big)^2.$$

Démonstration. — Soit F un sous-ensemble fini de E tel que la restriction de Π à $E \smallsetminus F$ est constante égale à a_0.

On va majorer, dans un premier temps, $p_m(q^k)$ pour un k fixé, puis étendre cette majoration à tout $n > 0$.

Soit $\overline{m}$ le mot infini point fixe de $\overline{\sigma}$ contenu dans e_0E^ω. Un mot donné w de m de longueur q^k provient, par définition, de la projection d'un mot $\overline{w}$ de $\overline{m}$ de longueur q^k. Ce mot $\overline{w}$ est un sous-mot d'un certain $\overline{\sigma}^k(e_1)\overline{\sigma}^k(e_2)$. On va majorer le nombre de $\overline{\sigma}^k(e)$ qui se projettent sur des mots différents de $a_0^{q^k}$.

En conséquence directe de la proposition 3.1, on a les équivalences, pour tout $a \in A \setminus \{a_0\}$:

$$a \text{ apparaît dans } \sigma^k(e) \Leftrightarrow \exists s \in F,$$

$$s \text{ apparaît dans } \overline{\sigma}^k(e) \Leftrightarrow e \in \bigcup_{s\in F} \mathcal{Z}_k(s).$$

avec $\mathcal{Z}_k(s) = \bigcup_{(j_1,\dots,j_k)\in\{0,1,\dots,q-1\}^k} \overline{\sigma}_{j_1}^{-1} \circ \cdots \circ \overline{\sigma}_{j_k}^{-1}(\{s\})$.

Par conséquent, on obtient l'inégalité, dans le cas général :

$$\mathrm{Card}\left(\{e \in E,\ \sigma^k(e) \neq a_0^{q^k}\}\right) \leqslant \mathrm{Card}(F)(qK)^k.$$

Il y a ainsi au plus $\mathrm{Card}(F)(qK)^k$ mots du type $\overline{\sigma}^k(e)$ qui se projettent de manière différente de $a_0^{q^k}$, et donc $\mathrm{Card}(F)(qK)^k + 1$ possibilités de projections différentes pour $\overline{\sigma}^k(e_1)$ et $\overline{\sigma}^k(e_2)$.

Comme chaque projection $\sigma^k(e_1)\sigma^k(e_2)$ peut contenir au plus q^k nouveaux mots différents de longueur q^k, on a une première majoration :

$$\forall k > 1,\ p_m(q^k) \leqslant q^k\left(\mathrm{Card}(F)(qK)^k + 1\right)^2.$$

Comme, pour $q^k \leqslant n < q^{k+1}$, on a $p_m(q^k) \leqslant p_m(n) \leqslant p_m(q^{k+1})$, on obtient :

$$\forall n \in [q^k, q^{k+1}[,\ p_m(n) \leqslant q^{k+1}(\mathrm{Card}(F)(qK)^{k+1} + 1)^2.$$

Et ensuite,

$$\forall n \geqslant 1,\ p_m(n) \leqslant qn(\mathrm{Card}(F)qKn^{1+\log_q K} + 1)^2,$$

en utilisant le fait que, pour tout entier n de $[q^k, q^{k+1}[$, on a $k \leqslant \log_q n < k+1$ et donc $(qK)^{k+1} \leqslant qKn^{1+\log_q K}$.

Cette majoration est valable pour tout F sous-ensemble fini de E tel que $\Pi(E \setminus F)$ est constante. On obtient donc

$$\forall n \geqslant 1,\ p_m(n) \leqslant qn(|\Pi|qKn^{1+\log_q K} + 1)^2.$$

Ce qui termine la démonstration. □

6. Ouvertures

Le pivot de la démonstration du théorème 5.4 réside dans le fait que les ensembles d'antécédents par les $\overline{\sigma}_j$ sont de cardinaux finis et uniformément bornés. Il reste à savoir si l'hypothèse «uniformément bornés» est une hypothèse réellement nécessaire ou si d'autres hypothèses, liées à la forme de l'automate suffisent, notamment des conditions assurant la récurrence de l'état 0. Par exemple, que dire de la complexité du mot m engendré par le 2-automate $(\mathbb{Z}, 0, \phi_{\overline{\sigma}}, \{0,1\}, \Pi_0)$, où la substitution $\overline{\sigma}$ est donnée par :

$$\begin{array}{rccl} \overline{\sigma}: & \mathbb{Z} & \to & \mathbb{Z}^* \\ & e \neq 0 & \mapsto & (\lfloor \log_2 e \rfloor)(e+1) \\ & 0 & \mapsto & 01 \end{array}$$

et Π_0 est définie par $\Pi_0^{-1}(\{1\}) = \{0\}$?

$$m = 1101010001010100010101000100010001000\cdots$$

D'autre part, il reste à déterminer si la majoration du théorème 5.4 est une majoration optimale.

BIBLIOGRAPHIE

[1] J.-P. Allouche, « Sur la complexité des suites infinies », *Bull. Belg. Math. Soc. Simon Stevin* **1** (1994), n° 2, p. 133-143.

[2] J.-P. Allouche & J. Shallit, *Automatic sequences. Theory, applications, generalizations*, Cambridge University Press, Cambridge, 2003.

[3] J. Berstel & D. Perrin, *Theory of codes*, Academic Press, 1985.

[4] S. Brlek, « Enumeration of factors in the Thue-Morse word », *Discrete Applied Math.* **24** (1989), p. 83-96.

[5] A. Cobham, « Uniform-tag sequences », *Math. Syst. Th.* **6** (1972), p. 164-192.

[6] A. Ehrenfeucht, K. P. Lee & G. Rozenberg, « Subword complexities of various classes of deterministic developmeltal languages without interactions », *Math. Syst. Theoret. Comput. Science* **63** (1975), p. 59-75.

[7] S. Eilenberg, *Automata, languages and machines*, vol. A, Academic Press, 1974.

[8] S. Ferenczi, « Substitution dynamical systems on infinite alphabets », *to appear in* Ann. Inst. Fourier, 2006.

[9] M. Lothaire, *Combinatorics on words*, Encyclopedia of Mathematics and Its applications, vol. 17, Cambridge University Press, 1983.

[10] ———, *Algebraic combinatorics on words*, Encyclopedia of Mathematics and Its applications, vol. 90, Cambridge University Press, 2002.

[11] C. Mauduit, « Propriétés arithmétiques des substitutions et automates infinis », *to appear in* Ann. Inst. Fourier, 2006.

[12] C. Mauduit & A. Sárközy, « On the arithmetic structure of the integers whose sum of digits is fixed », *Acta Arithmetica* **LXXXI** (1997), n° 2, p. 145-173.

[13] B. Mossé, « Reconnaissabilité des substitutions et complexité des suites automatiques », *Bull. Soc. Math. France* **124** (1996), n° 2, p. 329-346.

[14] J.-J. Pansiot, « Complexité des facteurs des mots infinis engendrés par morphismes itérés », in *Automata, languages and programming (Antwerp, 1984)*, Lecture Notes in Comput. Sci., vol. 172, Springer, Berlin, 1984, p. 380-389.

[15] D. Perrin & J.-E. Pin, *Mots infinis. Technical report 93.40*, Laboratoire Informatique Théorique et Programmation. Institut Blaise Pascal, 1997.

[16] ———, *Infinite words, Automata, Semigroups, Logic and Games*, Pure and Applied Mathematics, vol. 141, Elsevier, 2004.

[17] N. Pytheas Fogg, *Substitutions in dynamics, arithmetics and combinatorics*, Lecture Notes in Mathematics, vol. 1794, Springer-Verlag, Berlin, 2002, Edited by V. Berthé, S. Ferenczi, C. Mauduit and A. Siegel.

[18] M. Queffélec, *Substitution dynamical systems-spectral analysis*, Lecture Notes in Mathematics, vol. 1294, Springer-Verlag, Berlin, 1987.

[19] G. Rozenberg & A. Salomaa (Eds), *Handbook of formal language*, Springer-Verlag, 1997.

Marion LE GONIDEC
IML
Campus de Luminy, case 907
13288 Marseille Cedex 09 (France)
gonidec@iml.univ-mrs.fr