Université de la Méditerranée
Faculté des sciences de Luminy

Thèse

pour obtenir le grade de
Docteur de l'université de la Méditerranée
Spécialité : Mathématiques
Sous la direction de M. Christian MAUDUIT,
Professeur à l'Université de la Méditerranée

soutenue le 6 décembre 2006 par

Marion LE GONIDEC

Sur la complexité des mots q^∞-automatiques

Devant le jury composé de

Mme Valérie BERTHÉ, LIRMM, Montpellier
M. Julien CASSAIGNE, IML, Marseille
M. Didier CAUCAL (rapporteur), IGM, Marne-la-Vallée
M. Sébastien FERENCZI, IML, Marseille
M. Christian MAUDUIT, Université de la Méditerranée
M. Michel MENDÈS FRANCE (rapporteur), Université Bordeaux I

Avant-propos

Je tiens à adresser mes remerciements à toutes les personnes qui m'ont soutenue, aidée et encouragée durant mes années d'étude et de recherche.

J'exprime ma reconnaissance à mon directeur de thèse, Christian Mauduit, pour la liberté qu'il m'a laissée dans l'organisation de mon travail. Ses nombreux conseils et remarques avisés m'ont permis de mener à bien ce travail. D'autre part, son soutien logistique m'a évité de rater certaines dead-lines importantes et de palier à mon « tête-en-l'air-isme administratif » habituel... Je le remercie également de s'être inquiété de ma santé nerveuse et pulmonaire durant la période de rédaction.

Je suis très honorée de l'attention que Michel Mendès France et Didier Caucal ont portée à mes travaux et je les remercie vivement d'avoir accepté de rapporter cette thèse. Les discussions que j'ai pu avoir avec Didier Caucal m'ont permis d'approfondir ma compréhension de nombreuses notions et je lui suis très reconnaissante du temps qu'il m'a accordé.

je suis très heureuse que Valérie Berthé, qui m'a soutenue moralement tout au long de ces premières années de recherche, ait accepté de faire partie de mon jury.

Je tiens également à remercier chaleureusement Julien Cassaigne et Sébastien Ferenczi d'avoir accepter de faire partie de mon jury et pour tous leurs conseils qu'ils m'ont donnés pendant ma thèse.

Durant ces trois années passées à l'Institut de Mathématiques de Luminy, j'ai eu la chance de travailler dans d'excellentes conditions grâce à la très sympathique équipe Dynamique, Arithmétique et Combinatoire qui m'a accueillie. Un grand merci à Aurélia Lozingot qui a su me réconcilier avec l'administratif mais aussi à Catherine Haguenauer et Sophie Realini pour leurs conseils avisés au sein du département de mathématiques.

Je souhaite aussi souligner l'aide précieuse qui m'a été apportée, sans le vouloir, par tous les auteurs qui mettent en libre accès sur internet articles, supports de cours, transparents et livres, tant pour le gain de temps qui cela m'a apporté que pour le message sous-jacent d'encouragement au brassage des connaissances.

J'ai une pensée particulière pour tous les thésard-e-s et post-docs de passage avec qui j'ai partagé, pour quelques semaines ou plus longtemps, ces années de thèse, avec les moments de joie et ceux de déprime profonde qui vont avec : Anne, Matthieu, Julien, Thierry, Nicolas, ainsi que Xavier, Lionel, Etienne, Daniel, Tomi et Annina, Alina, Katalin, Idrissa, Erwan, Samuel... j'en oublie sûrement, mais j'espère qu'ils me pardonneront.

Enfin, je me permets de faire un (gros) paragraphe pour dire Merci à tous ceux qui m'ont sortie régulièrement la tête des mathématiques puisque j'espère qu'ils prendront, à l'occasion, le temps de lire au moins cette page de mon mémoire : Renaud, qui me supporte tous les jours et me rend heureuse, mes parents Martine et Jean-François pour leur soutien constant : ma Maman pour toutes les infimes petites choses que seule une maman sait faire à merveille, du coq aux morilles au séchage de larmes en passant par la correction des fautes d'orthographe, mon Papa, parce qu'il est un modèle de pragmatisme jovial et qu'il accepte qu'une matheuse bricole avec lui, malgré la gène occasionnée par les « et pourquoi ? », Anne-Marie, reine du picnic improvisé, qui m'a herbergée de nombreuses fois durant la rédaction, Guillaume et Max pour les réveils du dimanche avec Hightway to hell au pipeau ou à la guitare, mes deux amies d'enfance Coralie et Flo (et sa p'tite famille), Christine, pour les longues soirées discut' malgré le mal au dos, les repas et les week-ends tranquilous, Anne, bien sûr, qui met une touche de féminité dans ce monde de brutes ! Julien, qui a entrepris la lourde tache de m'initier à la « Pisot Power », Matthieu, avec qui on peut discuter de tout, de l'Amérique du sud aux les mécanismes du cerveau féminin en passant par la botanique et les maths, Thierry, avec qui on peut discuter de tout aussi, mais sous Wiki, du moment qu'on non-dit n'importe quoi et qu'on est oui-ouvert au dialogue, Nicolas, qui est toujours partant pour

une pause-café ou parler billards, s'il y a matière à râler, c'est encore mieux ! Les inconditionnels du Rouveau : Nick et Huguette, Yann et Tina, Erick et Virginie, Greg et Manu (félicitations !), la cousine Frédé, Anne-So, Ber, Paulo, Cruchot, le schwartzwald teufel et le Garabian, le bar du Rivage, la bande du Tarot-Poker du jeudi soir : Blond, φ, Mag, Tom, Toff, Cé, Rasta, Bouba et Sandra, Fa et Nath, tous ceux que je n'ai pas encore cités et qui nous ont aidé pour les travaux de l'appart : Dav' et Rozenn pour le placo et l'équipe de choc composée de mon frangin Guillaume, Eric (M'sieur Vééééérrrron) et Erwan, tous mes « oncles » gad'zarts, mes oncles et tantes du Brusc, Pituit qui ne s'est pas (trop) attaquée à ma bibliothèque, Marvin, le chat noir du parking, qui joue à la perfection le rôle de paillasson vivant de salon, mes potes de Fac, Dju, Karim, Rémi, Mohammed, Boris, Camille, Tony et David, Nath et Célia pour les barbecues très sioux, le bar des As de Saint-Henri, Mr Soulat et Mr Laversin, Gaëlle et Jérémie, pour l'arrosage des plantes quand on est pas là et pour tout le reste quand on est là, Paulette, pour m'avoir accueillie comme une fille et pour le repassage pro, Maxou, pour nous avoir accueillis à Buenos Aires, Jojo la mouette de Porquerolles, ma prof de physique de terminale qui disait que je ferai jamais des études longues, Papi Gogo le Géo trouve-tout du garage, Rémi et Sté parce qu'ils ont une drôle de façon de pas aimer les gens, le chalet de Narreyroux, le général Charix, les petites du placard, le mistral qui sèche mes tomates, le marché des Vans, ses fromages et le p'tit ballon d'rouge du Dardaillon, la balle de base-ball qui m'a maquillée pour un mois, les voitures qui tombent en panne, la SNCF, le 21... Poulpeline, Bradley, Peter Adam Smith Jr, Annette, Gigi, Séraphine, Wolfgang et tous les animaux de la ferme qui m'ont tenu compagnie pendant la rédaction, le Piadon pour le café sur le port du Brusc le dimanche matin, l'archi' pour son cours fabuleux intitulé « la notion d'epsilon dans le bâtiment », les mélèzes de Puy-Saint-Vincent, ...

Pour conclure, je pense que je n'aurai pas pu mener à bien cette thèse sans le soutien moral et logistique de ma famille, au sens large, ni sans l'infinie patience de mon amoureux. Il m'est très difficile de leur exprimer toute ma gratitude et mon affection uniquement avec quelques phrases, surtout quand on sait que je ne les finis une fois sur deux ; je ne m'y risquerais donc pas plus que cela mais j'espère sincèrement qu'ils en ont conscience.

Introduction

Depuis plus d'un demi-siècle, les automates, les langages reconnus par des automates, les mots infinis qu'ils engendrent et les langages associés à ces mots sont apparus en mathématiques et en informatique théorique dans plusieurs domaines : théorie des langages [RS97] , logique [PP04], théorie des codes [BP85], arithmétique et théorie des nombres [AS03a] [PF02], mais aussi en physique théorique, en biologie et neurobiologie.

La recherche effectuée sur les automates et sur tout ce qui s'y rattache est extrêmement dynamique grâce au développement rapide de l'informatique, qui fournit machines, exemples, modèles et problèmes...

D'autre part, ce sont des outils utiles pour la modélisation de systèmes de communication, le codage de systèmes dynamiques ou encore pour la recherche de motifs dans des séquences.

Il semble alors naturel, pour tenter de classer les systèmes que l'on modélise, de se poser la question de la complexité de ces machines et de leurs données de sortie.

Mais comment définir de manière mathématique la complexité ? Cette notion est plus ou moins subjective et étroitement liée aux notions de désordre, hasard et chaos... En donner une définition universelle serait donc hasardeux... voire complexe !

Chaque domaine des mathématiques et de l'informatique théorique possède en fait ses propres définitions de la complexité pour pouvoir classer les objets de son étude. Par exemple, on peut citer la forme de la répartition ou la normalité pour les ensembles de nombres réels ; les propriétés de récurrence, les coefficients de corrélation, la complexité algorithmique et la complexité combinatoire pour les mots finis et infinis ; la dépendance aux conditions initiales, l'ergodicité, l'entropie et les propriété de mélange pour les sytèmes dynamiques, ou encore, pour les langages, leur richesse, le type de grammaire et le type de machine qui l'engendre.

Pour les automates, cette question peut donc prendre plusieurs formes selon le domaine où elle apparait : Quelle est la longueur minimale du programme qui engendre ses données de sortie ? Quelle est la richesse et la nature du langage qu'il engendre ? Quelles peuvent être les corrélations entre deux sorties pour deux entrées proches ou corrélées ? Quels problèmes logiques concernant son graphe sont décidables ? etc...

Parmi le bestiaire des objets associés aux automates qui sont utiles pour répondre à ce type de questions, on trouve ainsi des mots infinis : si on ordonne l'ensemble (supposé dénombrable) des entrées, on peut former un mot infini dont les lettres sont les sorties de l'automate, ordonnées selon le même ordre que les entrées. Ce mot synthétise en quelque sorte l'action de l'automate puisqu'il en liste tous les couples entrée/sortie.

Dans le cas des automates dont l'ensemble des états est fini et l'ensemble des entrées est l'ensemble des représentations propres des entiers dans une base q, pout un certain nombre entier q, ces mots sont connus sous le nom de mots q-automatiques (voir par exemple [AS03a] ou [PF02]). Ils possèdent de nombreuses propriétés combinatoires, statistiques et logiques et dynamiques et ont fait l'objet d'une multitude de travaux notamment parce qu'ils interviennent dans de nombreuses problématiques issues de « milieux mathématiques et informatiques » différents.

Les mots q-automatiques, chaînon manquant entre théorie des langages, théorie des automates, théorie des nombres et dynamique, tiennent une place de choix dans la famille des mots infinis. Le fait qu'ils aient de bonnes propriétés sous tous les aspects incite à envisager d'autre familles de mots infinis, engendrés par des algorithmes proches de celui d'un automate fini ou possédant une propriété proche de celle vérifiée par les mots q-automatiques. C'est dans cette optique là que l'on peut voir les mots substitutifs (voir [Que87], [PF02]) et les mots q-réguliers (voir [AS92, AS03b]). Ces deux classes de mots contiennent les mots q-réguliers et certaines propriétés relatives aux mots q-automatiques s'y généralisent.

Cette étude porte sur une autre famille de mots infinis, les mots q^∞-automatiques, qui sont des mots à valeurs dans un ensemble fini et engendrés par des automates « admissibles » dont l'ensemble des états est dénombrable. Certes cette famille ne contient pas les mots q-automatiques mais les résultats que nous obtenons ici sont en fait valables pour les mots engendrés par des q-automates « admissibles » au plus dénombrables, parmi lesquels ont trouve les q-automates finis même s'ils sont moins précis que ceux qui sont ont déjà été obtenus pour les mots engendrés par des q-automates finis. Nous nous intéressons en particulier à la complexité combinatoire de tels mots.

Plan

Chapitre I

Ce chapitre introduit le vocabulaire et les outils concernant les mots infinis, les langages et les grammaires. Nous nous intéressons de manière plus poussée au langages rationnels et algébriques, notamment aux liens qu'ils entretiennent avec les automates finis et les automates à pile. Nous introduisons ensuite la notion de mot q-automatique, qui lie théorie des langages, théorie des automates et théorie des nombres. Après un survol des propriétés de ces mots, nous abordons, comme deux voies de généralisation de ce concept, la notion de mot régulier et de mot substitutif.

Chapitre II

Nous y détaillons un des principaux outils combinatoires utiles à l'étude des mots infinis : la fonction de complexité. On y introduit également la notion de facteurs spéciaux d'un mot infini. Nous présentons les résultats actuels concernant les différentes questions liées à la fonction de complexité d'un mot : propriétés de ces fonctions, types de fonctions possibles, calcul explicite ou ordre de grandeur possible dans certains cas, et détaillons les résultats concernant les mots automatiques et substitutifs.

Chapitre III

Dans ce chapitre, on introduit une généralisation du concept de mot q-automatique : les mots q^∞-automatiques, engendrés par projections admissibles de mots de sortie de q-automates dénombrables déterministes « admissibles » (q-ADDA). Nous présentons différentes classes de tels automates et explorons les propriétés combinatoires et dynamiques des mots de sortie de ces automates.

Chapitre IV

Ce chapitre présente une première étude de la complexité de mots q^∞-automatiques : les mots engendrés par les marches aléatoires isotropes sur des réseaux de $\mathbb{R}^d$, pour lesquelles on montre que la complexité vérifie $p(n) = O(n \log_{2d}^{d+1} n)$, et plus généralement, la complexité d'un mot engendré par une marche aléatoire sur un réseau de $\mathbb{R}^d$, dont les déplacements possibles sont donnés par une famille contenant q éléments ($q \geq d+1$), vérifie $p(n) = O(n \log_{2d}^{q+1} n)$. De plus, dans le cas de la marche aléatoire isotrope sur Z, on montre que la fonction de complexité est exactement équivalente à $n \log_2^2 n$.

Chapitre V

Nous exposons dans ce chapitre deux théorèmes généraux pemettant, sous des hypothèses de nature très différentes, de majorer de la fonction de complexité des mots q^∞-automatiques :

Pour les mots q^∞-automatiques engendrés par les q-ADDA de degré borné, nous obtenons que la fonction de complexité de tels mots vérifie $p(n) = O(n^{3+2\log_q K})$, où K représente l'entier qui borne le degré de tout état de l'automate. Dans ce cadre, sous l'hypothèse supplémentaire que les fonctions de transition de l'automate indexées par des entrées différentes commutent nous obtenons $p(n) = O(n^{1+2\log_q K}(\log_q)^{2q})$.

Pour les mots q^∞-automatiques engendrés par les q-ADDA monotones, nous obtenons que la fonction de complexité de tels mots vérifie $p(n) = O(n^3)$.

Nous verrons aussi comment la méthode développée pour démontrer ce théorème général concernant les mots q^∞-automatiques engendrés par les q-ADDA de degré borné peut-être améliorée sur des exemples.

Chapitre VI

Nous étudions la complexité du mot infini de Dyck sur deux types de parenthèses. On montre que la fonction de complexité de ce mot est du même ordre de croissance que $n\sqrt{n}$.

Chapitre VII

Ce chapitre regroupe un certain nombre de remarques concernant les extensions possibles des résultats du chapitre V aux mots q^∞-automatiques engendrés par des q automates de degré infini non monotones et les cas de générescence possibles. Nous terminons par plusieurs questions ouvertes.

Table des matières

Chapitre I

Bienvenue dans le monde des mots

Les mots et les langages sont des objets qui apparaissent en mathématiques et en informatique théorique dans de nombreux domaines : théorie des langages [RS97] , logique [PP04], théorie des codes [BP85], arithmétique et théorie des nombres [AS03a], étude des systèmes dynamiques [PF02]...

Bien que le travail effectué dans cette thèse soit plus axé autour de la combinatoire et la complexité de certains mots infinis, ce chapitre tente de rassembler le vocabulaire et les outils usuels liées aux langages formels ainsi qu'aux mots infinis.

I.1 Terminologie des mots finis et infinis

Les notations utilisées ici sont principalement issues de [PF02].

Pour $a < b$ deux entiers relatifs, on note $[\![a, b]\!]$ l'ensemble $\{a, a+1, \dots b\}$.

Alphabet :

Un ensemble fini ou dénombrable est appelé *alphabet*. Ses éléments sont appelé *lettres*.

Mots finis :

Pour tout entier n positif, on note les éléments de A^n sous forme concaténée : un élément m de A^n se note $m = m_0 m_1 \dots m_{n-1}$. On appelle ces éléments des *mots de longueur* n et la longueur d'un mot m donné est notée $|m|$. Le nombre d'*occurences* de la lettre a dans le mot m, c'est-à-dire le nombre de fois où apparait la lettre a dans la lecture de w, est noté $|m|_a$ et défini par la formule $|m|_a = Card\left(\{k \in [\![0, |m|-1]\!],\ m_k = a\}\right)$. Par exemple, 10110011 est un mot de longueur 8 sur l'alphabet $\{0, 1\}$ et $|10110011|_1 = 5$, *ab* et *pituit* sont des mots de longueurs respectives 2 et 6 sur l'alphabet latin, $\uparrow\downarrow\rightarrow\uparrow\uparrow\leftarrow\downarrow\downarrow\uparrow$ est un mot de longueur 9 sur l'alphabet $\{\downarrow, \uparrow, \rightarrow, \leftarrow\}$.

On note l'ensemble des mots finis $A^* = \cup_{n\geq 0} A^n$ (L'opérateur $*$ est appelé étoile de Kleene). On définit, sur A^*, l'opération de concaténation : le mot $v = mw$, concaténé de m et de w, est défini par $|v| = |m| + |w|$ et $v_i = m_i$ si $0 \leq i \leq |m| - 1$ et $v_i = w_{i-|m|}$ pour $|m| \leq i \leq |v| - 1$. L'ensemble A^* muni de l'opération de concaténation est un monoïde, c'est-à-dire que l'opérateur de concaténation est une loi de composition interne dans A^*, associative qui admet un élément neutre : le mot vide, noté ε par convention.

Mots infinis et bi-infinis :

On note $A^{\mathbb{N}}$ l'ensemble des suites indéxées par $\mathbb{N}$ à valeurs dans A. Ces suites seront notées sous forme concaténée et appelées *mots infinis* : $m \in A^{\mathbb{N}}$ s'écrit $m = m_0 m_1 \dots m_n \dots$ avec, pour tout n, $m_n \in A$.

On note $A^{\mathbb{Z}}$ l'ensemble des suites indéxées par $\mathbb{Z}$ à valeurs dans A. Ces suites seront notées également sous forme concaténée et appelées indifféremment *suites* ou *mots bi-infinis* : $m \in A^{\mathbb{Z}}$ s'écrit $m = \dots m_{-2} m_{-1} \centerdot m_0 m_1 \dots m_n \dots$ avec, pour tout n, $m_n \in A$.

La notation A^ω désignera d'une manière générale l'ensemble $A^{\mathbb{N}}$ ou l'ensemble $A^{\mathbb{Z}}$.

Facteurs d'un mot :

Un mot fini w de longueur n qui apparait dans un mot m fini, infini ou bi-infini, c'est-à-dire qui vérifie pour un certain entier k, $w = m_k m_{k+1} \dots m_{k+n-1}$, est appellé *facteur* ou *sous-mot de m de longueur n*. On dit alors que l'entier k est un *rang d'apparition de w dans m*.

Pour un mot $m = m_0 m_1 \dots m_{n-1}$ sur l'alphabet A et pour tout $i \in [\![1, n]\!]$, on appelle *préfixe de longueur i de m* (resp. *suffixe de longueur i de m*) le mot $Pref_i(m) = m_0 \dots m_{i-1}$ (*resp.* $Suff_i(m) =$

$m_{n-i} \dots m_{n-1}$).

Pour tout entier $n \geq 0$, on note $F_n(m)$ l'*ensemble des facteurs de longueur n de m* et $L(m) = \cup_{n\in\mathbb{N}} F_n(m)$ le *langage du mot m*. Lorsque $n > |m|$, $F_n(m) = \emptyset$ et $F_0(m) = \{\varepsilon\}$.

On appelle *factorisation* d'un mot m toute écriture de m sous la forme explicite d'une concaténation de facteurs.

Ainsi, par exemple, si $u = ab$ et $v = ba$ sur l'alphabet $\{a, b\}$ sont des facteurs du mot $m = abababa$, et m se factorise en $m = avvvv = av^4$ ou en $uuuua = u^4a$.

Ordre sur A^*

On peut définir sur A^* plusieurs ordres :

L'*ordre préfixiel* est un ordre partiel défini par $w < m$ si et seulement si w est un préfixe strict de m.

L'*ordre lexicographique* est un ordre total qui étend l'ordre préfixiel. Il est nécessaire, pour le définir que l'alphabet A soit totalement ordonné de façon arbitraire (s'il est ordonné naturellement, c'est encore mieux !). On ordonne donc $A = \{a_1 \prec a_2 \prec \dots\}$, et on pose :

$$\forall w, w' \in (A^*)^2,\ w \prec w' \Leftrightarrow w < w' \text{ ou } w = uav_w, w' = ubv_{w'} \text{ avec } (a, b) \in A^2, a \prec b.$$

L'*ordre hiérarchique* est un ordre total pour lequel les mots sont d'abord classés par leur longueur puis ordonnés lexicographiquement ; il est aussi nécessaire pour le définir, d'ordonner arbitrairement A. C'est, par exemple, l'ordre que l'on associe naturellement à l'ensemble des représentations des nombres entiers en base q, pour un entier $q \geq 2$.

Récurrence

Un facteur w d'un mot infini m est *récurrent* s'il apparait une infinité de fois comme facteur de m ; autrement dit, si pour tout entier n tel que $m_n m_{n+1} \dots m_{n+|w|-1} = w$, il existe un entier $k > n$ pour lequel $m_k m_{k+1} \dots m_{k+|w|-1} = w$. Un facteur w d'un mot infini m est *récurrent à lacunes bornées* ou *uniformément récurrent* le facteur w apparait une infinité de fois comme facteur de m et pour lequel les plages entre deux occurrences consécutives sont de longueurs bornées uniformément, c'est-à-dire que w est récurrent et qu'il existe un entier K_w tel que, pour tout entier n tel que $m_n m_{n+1} \dots m_{n+|w|-1} = w$, c'est-à-dire qu'il existe un entier $n < k < n + K_w$ pour lequel $m_k m_{k+1} \dots m_{k+|w|-1} = w$.

Un mot infini est dit *récurrent* (resp. *uniformément récurrent*) si tous ses facteurs sont récurrents (resp. uniformément récurrents).

Les mots uniformément récurrents les plus simples sont les *mots périodiques*, c'est-à-dire les mots m qui vérifient : il existe un entier P tel que, pour tout $n \in \mathbb{N}$ ou $\mathbb{Z}$ (selon que m est infini ou bi-infini), $m_{n+P} = m_n$. Si $v = m_0 m_1 \dots m_{P-1}$, on note alors $m = v^\omega$. Un mot infini w sur l'alphabet A est dit *ultimement périodique* s'il peut s'écrire $w = xv^\omega$, où x et v sont des éléments de A^*.

Système dynamique associé à un mot infini

L'ensemble A^ω est naturellement muni de la topologie produit $\mathcal{T}$ associée à la topologie discrète sur A. Cette topologie est définie par la distance suivante :

$$\forall m \neq m' \in A^\omega, d(m, m') = 2^{-\inf\{|n|,\ m_n \neq m'_n\}}.$$

Cette topologie est appelée *topologie des cylindres*. Lorsque $A^\omega = A^{\mathbb{N}}$, elle est engendrée par l'ensemble des ouverts-fermés du type $[w] = \{m \in A^{\mathbb{N}}, m_0 m_1 \dots m_{|w|-1} = w\}$, pour tout mot w de A^*, appelés *cylindres*. Lorsque $A^\omega = A^{\mathbb{Z}}$, elle est engendrée par l'ensemble des ouverts-fermés du type $[u \centerdot w] = \{m \in A^{\mathbb{Z}}, m_{-|u|} \dots m_{-2} m_{-1} \centerdot m_0 m_1 \dots m_{|w|-1} = u \centerdot w\}$, pour tout mot $u \centerdot w$ de A^*, aussi appelés *cylindres*.

Il est intéressant de souligner ici la différence de structure de A^ω dans le cas où l'alphabet A est fini et celui où l'alphabet A est dénombrable : lorsque A est un alphabet fini, $(A^\omega, \mathcal{T})$ est un espace compact de Cantor et lorsque A est dénombrable, $(A^\omega, \mathcal{T})$ est un espace polonais, c'est-à-dire un espace métrisable à base dénombrable dont la topologie (ici, celle des cylindres) est définie par une distance qui en fait un espace complet.

Sur $(A^{\mathbb{N}}, \mathcal{T})$, on définit l'*opérateur de décalage* ou *shift unilatéral* S, qui consiste à « couper »la première lettre d'un mot :

$$\begin{array}{rccc} S : & A^{\mathbb{N}} & \longrightarrow & A^{\mathbb{N}} \\ & m = m_0 m_1 \dots & \mapsto & S(m) = m_1 m_2 \dots \end{array}$$

On note $Orb(m) = \{S^n(m), n \in \mathbb{N}\}$ l'orbite de m dans $A^{\mathbb{N}}$ sous l'action du shift.

Sur $(A^{\mathbb{Z}}, \mathcal{T})$, on peut aussi définir l'opérateur de décalage de la manière suivante :

$$\begin{array}{rccc} S: & A^{\mathbb{Z}} & \longrightarrow & A^{\mathbb{Z}} \\ & m = \ldots m_{-2}m_{-1} \centerdot m_0 m_1 \ldots & \mapsto & S(m) = \ldots m_{-2}m_{-1}m_0 \centerdot m_1 m_2 \ldots \end{array}$$

On parle alors de *shift bitatéral* et on note $Orb(m) = \{S^n(m), n \in \mathbb{Z}\}$ l'orbite de m dans $A^{\mathbb{Z}}$ sous l'action du shift.

Le système dynamique symbolique associé au mot infini ou bi-infini m est alors $(\overline{Orb(m)}, S)$. Sous de bonnes conditions sur le mot infini m, en particulier la récurrence, on peut s'intéresser aux propriétés dynamiques du système $(\overline{Orb(m)}, S)$: minimalité, existence (et unicité) de mesure invariante. Ces propriétés peuvent fournir des renseignements précieux, notamment si le mot m est obtenu par codage d'un système dynamique géométrique (flot géodésique sur une surface, codage d'une rotation...). L'étude spectrale ainsi que les mesures de corrélations peuvent également permettre d'obtenir des propriétés statistiques sur la structure du mot. Nous invitons le lecteur à consulter [Que87] ou [PF02] pour une introduction à toutes ces notions.

I.2 Morphismes de monoïde et substitutions

Définition I.2.1. Soient A et A' deux alphabets.

Un *morphisme* ou *codage* $\psi : A^* \to A'^*$ est une application de A^* vers A'^* telle que $\psi(wm) = \psi(w)\psi(m)$ pour tout mots w et m de A^*. L'application ψ est entièrement définie par l'image des éléments de A. Lorsque l'alphabet de sortie est l'alphabet de depart A, on parlera de *substitution sur l'alphabet* A.

On dit que le morphisme $\psi : A^* \to A'^*$ est :
- *alphabétique* si $\forall a \in A,\ |\psi(a)| \leq 1$,
- *strictement alphabétique* si $\forall a \in A,\ |\psi(a)| = 1$,
- *continu ou non effaçant* si $\forall a \in A,\ |\psi(a)| \geq 1$,
- *uniforme* ou *de longueur constante* s'il existe un nombre entier $q \geq 1$ tel que $\forall a \in A,\ |\psi(a)| = q$.

On peut décrire une substitution σ de la manière suivante :
σ est d'abord définie sur A :

$$\begin{array}{rccc} \sigma: & A & \to & A^* \\ & a & \mapsto & \sigma_0(a)\sigma_1(a)\ldots\sigma_{|\sigma(a)|-1}(a). \end{array}$$

où, pour tout $i \in \{0, 1, \ldots, |\sigma(a)| - 1\}$, σ_i est une application de A dans A, et l'application σ est ensuite étendue au monoïde A^* et à A^ω par concaténation.

Remarque I.2.2. Mise en garde : La définition des substitutions faite ici n'est pas usuelle et ne doit pas être confondue avec celle ayant cours en général dans la littérature concernant les langages, où elle sont définies comme des homomorphismes entre les ensembles des parties de deux monoïdes. Elle provient du fait que l'application d'une substitution à un mot consiste à substituer à chaque lettre son mot-image.

Exemples I.2.3. La *substitution de Morse* σ sur l'alphabet $\{a, b\}$ est définie par $\sigma(a) = ab$ et $\sigma(b) = ba$, ce que l'on peut aussi noter sous la forme suivante $a \to ab$, $b \to ba$.

La *substitution de Fibonacci* ϕ sur l'aphabet $\{a, b\}$ est définie par $\phi(a) = ab$ et $\phi(b) = a$.

La *substitution de l'ivrogne* $\overline{\sigma}$ sur l'alphabet $\mathbb{Z}$, qui apparait dans [Fer06] et [Mau06], est définie pour tout entier relatif n par $\overline{\sigma}(n) = (n-1)(n+1)$. C'est « l'équivalent substitutif » de la marche aléatoire sur $\mathbb{Z}$.

On associe naturellement à une substitution σ sur l'alphabet A un graphe $\mathcal{G}(\sigma)$. L'ensemble des sommets de $\mathcal{G}(\sigma)$ est A et l'ensemble des arcs orientés et étiquetés de $\mathcal{G}(\sigma)$ est donné par l'ensemble $\left\{(a, a', i) \in A^2 \times \{0, 1, \ldots, |\sigma(a)|\}, \sigma_i(a) = a'\right\}$. Les arcs étiquetés sont représentés sur le graphe par des flèches de a vers a' indexées par i.

Les graphes représentés dans les figures I.1, I.2 et I.3 sont les graphes associés aux trois exemples que nous venons de citer.

Lorsque l'alphabet A est fini, on peut classer les substitutions en fonction de la croissance des mots images, par itération de la substitution (voir [Mau88]).

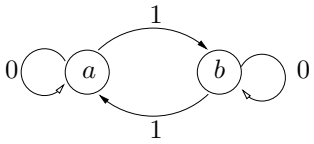


FIG. I.1 – Graphe associé à la substitution de Morse.

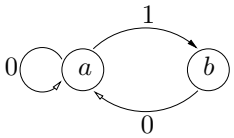


FIG. I.2 – Graphe associé à la substitution de Fibonacci.

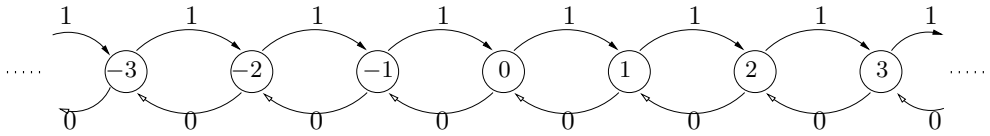


FIG. I.3 – Graphe associé à la substitution de l'ivrogne.

Proposition et définition I.2.4. Soit A un alphabet fini.

Pour une substitution donnée $\sigma : A^* \to A^*$ et a une lettre de A, la suite des longueurs $|\sigma^n(a)|$ est soit bornée, soit croissante comme une fonction $n^{\alpha_a}\beta_a^n$ pour un certain entier $\alpha_a \geq 0$ et un certain nombre $\beta_a \geq 1$. La fonction $n^{\alpha_a}\beta_a^n$ est appellé *ordre de croissance de a*.

On dit que la substitution $\sigma : A^* \to A^*$ est :

- *quasi-uniforme* si tous les éléments a de A ont un même ordre de croissance de la forme β^n,
- *polynomialement divergente* si tous les éléments a de A ont un ordre de croissance de la forme $n^{\alpha_a}\beta^n$, avec $\beta > 1$ et les α_a non nuls,
- *exponentiellement divergente* s'il existe deux lettres a et b d'ordres respectifs $n^{\alpha_a}\beta_a^n$ et $n^{\alpha_b}\beta_b^n$ avec $1 < \beta_a < \beta_b$ et quelque soit la lettre l de A, $\beta_l > 1$.

On dit qu'une substitution est *croissante* si toute lettre de A a un ordre de croissance non borné. Une substitution est croissante si et seulement si elle est soit quasi-uniforme, soit polynomialement ou exponentiellement divergente (c'est la définition utilisée par J.-J. Pansiot [Pan85]).

Si une substitution est de longueur constante q, alors, pour un entier n fixé, toutes les longueurs $|\sigma^n(a)|$ sont égales à q^n. On parle alors de *substitution uniforme*.

Voici quelques exemples de substitutions sur l'alphabet $\{a, b\}$, pour lesquelles sont donnés les types et les ordres de croissance des itérés :

Substitution	$a \mapsto ab$ $b \mapsto ba$	$a \mapsto aaab$ $b \mapsto bb$	$a \mapsto aba$ $b \mapsto bb$	$a \mapsto aab$ $b \mapsto b$
type	Uniforme	Exp. divergente	Pol. divergente	non croissante
ordre de a	2^n	3^n	$n2^n$	2^n
ordre de b	2^n	2^n	2^n	1

Remarque I.2.5. La classification des substitutions sur un alphabet dénombrable, par l'ordre de croissance des itérés, reste une question ouverte : quelles familles de fonctions croissantes peuvent-elles être les ordres de croissance des itérés d'une substitution sur un alphabet dénombrable ? Si la famille des longueurs des images est bornée, quels types d'ordre de croissance pouvont-nous obtenir ? De quelle forme sont les substitutions dont les ordre sont de type $n^{\alpha_a}\beta_a^n$ pour un certain entier $\alpha_a \geq 0$ et un certain nombre $\beta_a \geq 1$?... Toutes ces questions attendent encore leur réponse.

Comme une substitution peut s'étendre en une application de $A^{\mathbb{N}}$ dans lui-même, il est naturel de se poser la question de l'existence de points fixes. Lorsqu'il en existe, ces mots infinis sont particulièrement représentatifs de l'action de la substitution :

Définition I.2.6. Soit σ une substitution sur un alphabet A.

On dit que la substitution admet *un point fixe* s'il existe un mot infini m qui vérifie $\sigma(m) = m$.

Les substitutions n'admettent pas toutes des points fixes, mais il existe des conditions suffisantes sur une substitution pour qu'elle admette un point fixe.

Une façon plus générale de garantir l'existence d'au moins un point fixe pour une substitution est de poser les conditions suivantes :

Proposition I.2.7. *Soit σ est une substitution sur un alphabet A. S'il existe un élément a_0 de A tel que $\sigma(a_0) \in a_0A^*$ et $\lim_{n\to+\infty} |\sigma^n(a_0)| = +\infty$ (on retrouve aussi cette propriété sous l'expression : σ est* prolongeable à a_0*), alors σ admet au moins un point fixe m, et ce point fixe commence par la lettre a_0 et on note $m = \lim_{n\to\infty} \sigma^n(a_0)$.*

En particulier, tous les mots itérés $\sigma^n(a_0)$ sont des préfixes de m.

■ **Démonstration de la proposition I.2.7** Soit $(m^{(n)})_{n\in\mathbb{N}}$, la suite de mots infinis définie par $m^{(n)} = \sigma^n(a_0)a^\omega$ pout un élément $a \in A$ fixé.

On a, dans $\mathcal{A}^{\mathbb{N}}$ et pour tout couple d'entier (n, p), $d(m^{(n)}, m^{(n+p)}) = 2^{-|\sigma^n(a_0)|}$. En effet, puisque $\sigma(a_0)$ commence par a_0, alors il en est de même de $\sigma^p(a_0)$ pour tout entier p. Ainsi, $\sigma^{n+p}(a_0)$ commence par $\sigma^n(a_0)$ et $Pref_{|\sigma^n(a_0)|}(m^{(n)} = Pref_{|\sigma^n(a_0)|}(m^{(n+p)}$.

La suite $(m^{(n)})_{n\in\mathbb{N}}$ est donc une suite de Cauchy, puisque $\lim_{n\to+\infty} |\sigma^n(a_0)| = +\infty$. Ainsi, que A soit fini ou dénombrable $A^{\mathbb{N}}$ est complet et donc il existe un mot infini m de $A^{\mathbb{N}}$ vers lequel converge la suite $(m^{(n)})_{n\in\mathbb{N}}$ (pour la topologie des cylindres).

D'autre part, on a $d(\sigma(m^{(n)}), m^{(n+1)}) \leq 2^{-|\sigma^n(a_0)|}$ donc le mot limite m vérifie $d(\sigma(m), m) = 0$, et ainsi, il existe un mot infini m ■

Exemple I.2.8. La substitution de Morse admet deux points fixes, l'un commençant par la lettre a et l'autre par la lettre b :

$$m = \sigma(m) = abbabaabbaababbabaababbaabbabaabbaababbaabbabaababbabaabbaababba\ldots$$

$$m' = \sigma(m') = baababbaabbabaababbabaabbaababbaabbabaabbaababbaabbabaababbabaabbaababbaabbabaab\ldots$$

Cette substitution et ses points fixes possèdent de nombreuses propriétés. A titre d'exemples, nous en mentionnerons quelques unes tout au long de cette étude. Nous invitons le lecteur à consulter [PF02] pour plus de détails.

Exemple I.2.9. La substitution de Fibonacci rentre dans les critères de la proposititon ci-dessus ; l'unique point fixe de cette substitution est le mot infini f, dit mot de Fibonacci :

$$f = \phi(f) = \lim_{n\to\infty} \phi^n(a) = abaababaabaababaababaabaababaabaababaababaabaababaabaababaab\ldots$$

Définition I.2.10. On dit qu'une substitution sur un alphabet A est *primitive* ou *minimale* si elle vérifie : $\exists k \geq 1, \forall a, b \in A,\ |\sigma^k(b)|_a \neq 0$.

Cette propriété n'est possible que lorsque l'alphabet A est fini. Lorsque la substitution est définie sur un alphabet fini ou dénombrable, on peut tout de même définir une propriété plus faible :

On dit qu'une substitution sur un alphabet A est *faiblement primitive* si elle vérifie : $\forall a, b \in A, \exists k \geq 1,\ |\sigma^k(b)|_a \neq 0$.

La notion de substitution primitive est liée, par la *matrice d'incidence de la substitution* : M_σ, définie par :

$$M_\sigma = \left(|\sigma^k(b)|_a\right)_{(a,b)\in A^2},$$

à la minimalité du système dynamique associé au point fixe de la substitution.

Cette matrice d'incidence, sorte d'abélianisée de la substitution, peut posséder des propriétés ayant une influence directe sur la substitution. Dans le cas d'un alphabet fini, c'est grâce à elle que l'on peut calculer les ordres de croissance des itérés et si cette matrice vérifie certaines propriétés, alors on peut interpreter géométriquement le système dynamique associé aux points fixes de la substitution (Voir par exemple [AI01] ou [PF02] pour une introduction, et l'article précurseur [Rau82]).

I.3 Langages formels et grammaires

Dans ce paragraphe, A désigne un alphabet fini.

I.3.1 Langages et opérations sur les langages

Définition I.3.1. On appelle *langage sur l'alphabet* A tout sous-ensemble de A^* ; un langage L est donc une collection de mots finis, dont les éléments sont appelés *mots admissibles* de L.

Par exemple, l'ensemble des suites de lettres finies de l'alphabet latin ayant une définition attitrée dans le dictionnaire français forme un langage.

L'ensemble des mots sur l'alphabet $\{a, b\}$ qui contiennent autant d'occurrences de a que d'occurrences de b dans leur écriture forment un langage.

L'ensemble des représentations propres en base 2 des entiers pairs strictement positifs est un langage sur l'alphabet $\{0, 1\}$ (c'est le langage formé des mots de $\{0, 1\}^*$ qui commencent par un 1 et se terminent par un 0).

Définitions I.3.2. Opérations sur les langages.

On définit sur l'ensemble des langages sur l'alphabet A, les opérations booléennes usuelles : pour deux langages L et L', on définit

- l'union de deux langages, qui est usuellement notée sous forme additive, par un "+", $L \cup L' = L + L' = \{w,\ w \in L \text{ ou } w \in L'\}$,
- l'intersection $L' \bigcap L = \{w,\ w \in L, w \in L'\}$,
- le complément $L \setminus L' = \{w,\ w \in L, w \notin L'\}$,
- le produit de deux langages, induit par la concaténation, $LL' = \{ww',\ w \in L, w' \in L'\}$.

Par extension, on définit L^n pour tout entier n par la relation $L^0 = \{\varepsilon\}$ et $L^n = LL^{n-1}$. On définit également l'*étoile* d'un langage $L^* = \sum_{n\geq 0} L^n$ et l'*étoile propre* d'un langage $L^+ = \sum_{n>0} L^n$.

Définition I.3.3. Un langage L sur l'alphabet A est dit *prolongeable à droite* (resp. *prolongeable à gauche*) si pour tout mot w de L, il existe une lettre $a \in A$ telle que wa est un mot de L (*resp.* aw est un mot de L). On parle de *langage bi-prolongeable* lorsqu'il est à la fois prolongeable à gauche et à droite.

Un langage L sur l'alphabet A est dit *factoriel* si, pour tout mot w de L, tous les facteurs de w sont aussi des mots de L.

Lorsque L est le langage des facteurs d'un mot infini, L est un langage factoriel, de plus, si m est un mot récurrent, alors L est bi-prolongeable.

Remarque I.3.4. Lorsqu'un langage L n'est pas factoriel, il se peut que l'on puisse tout de même factoriser ses mots à l'aide d'un nombre fini de mots de L. Pour C, partie finie de L, on appelle *factorisation* d'un mot w de L toute chaîne $w_1,\ w_2, \ldots, w_n$ de mots de C telle que $w = w_1 w_2 \ldots w_n$. Si tout mot de L admet une factorisation en mots de C, on dit que le monoïde engendré par C est un *langage codé* ou *code*. Un code C est dit *préfixe* (resp. *suffixe*) si aucun mot de C n'est pas le préfixe (*resp.* suffixe) d'un autre mot de C. On parle de *code bipréfixe* lorsqu'il est à la fois préfixe et suffixe. Nous renvoyons à [BP85] pour une étude approfondie de la théorie des codes.

I.3.2 Un peu de grammaire...

Lorsqu'on parle de langages, la notion de grammaire apparait naturellement. Une *grammaire*, c'est l'ensemble des règles qui permettent d'engendrer un langage. Avant d'en donner une définition formelle, donnons deux exemples de construction de grammaires, à partir d'un langage. L'idée est de systématiser la formation des mots d'un langage, à l'aide d'un nombre fini de symboles et de transformations.

Exemple I.3.5. On considère le langage sur l'alphabet $\{a, b\}$ défini par $L = a^+b^+$. Les mots du langage L peuvent être construits de la manière suivante : lorsqu'on écrit un mot $w = a^n b^m$ de L, il contient au moins un a, il faut spécifier combien : n , ensuite, w contient au moins un b à la suite des n occurrences de a, il faut spécifier combien : m. Cette construction directe des mots de L peut se traduire de la manière décrite ci-dessous :

On part d'un symbole S fixé, qui marque le début du processus de fabrication d'un mot de L. On applique une ou plusieurs fois la transformation $S \to aS$, puis on applique la transformation $S \to aTb$,

où T est un autre symbole, on peut alors utiliser la transformation $T \to Tb$ une ou plusieurs fois et pour finir, on applique la transformation $T \to \varepsilon$. On a par exemple :

$$S \to aS \to aaS \to aaaS \to aaaaTb \to aaaaTbb \to aaaaTbbb \to aaaabbb,$$

Le mot a^4b^3 est un mot de L et les transformations successives pour l'engendrer sont données par : $S \to aS$, $S \to aS$, $S \to aS$, $S \to aTb$, $T \to Tb$, $T \to Tb$, $T \to \varepsilon$.

La grammaire utilisée ici pour engendrer L est la donnée du quadruplet : $(\{a, b\}, \{S, T\}, P, S)$, où $P = \{S \to aS, S \to aTb, T \to Tb, T \to \varepsilon\}$ pour signifier que l'alphabet support du langage est $\{a, b\}$, les symboles « accessoires »prennent valeurs dans l'ensemble $\{S, T\}$, les transformations autorisées sont les éléments de $\{S \to aSb, S \to aTb, T \to Tb, T \to \varepsilon\}$ et le point de départ de la formation des éléments de L est S.

Exemple I.3.6. On considère le langage sur l'alphabet $\{a, b\}$ défini par $L = \{a^n b^n, n \geq 0\}$. Les mots du langage L s'obtiennent exactement de la manière suivante : ε est un mot de L et si w est un mot de L, alors awb est aussi un mot de L. Cette relation permet d'engendrer récursivement tous les mots de L. On peut traduire ce mécanisme de création des mots de L de la façon suivante :

On part d'un symbole S fixé, qui marque le début du processus de fabrication d'un mot de L. On applique une ou plusieurs fois la transformation $S \to aSb$ puis on applique la transformation $S \to \varepsilon$:

$$S \to aSb \to aaSbb \to aaaSbbb \to aaaaSbbbb \to aaaabbbb.$$

Le mot engendré à partir de S et des transformations successives $S \to aSb$, $S \to aSb$, $S \to aSb$, $S \to aSb$, $S \to \varepsilon$ est le mot a^4b^4 qui appartient à L. La grammaire utilisée ici pour engendrer le langage L est la donnée du quadruplet : $(\{a, b\}, \{S\}, \{S \to aSb, S \to \varepsilon\}, S)$, pour signifier que l'alphabet support du langage est $\{a, b\}$, les symboles « accessoires »prennent valeurs dans l'ensemble $\{S\}$, les transformations autorisées sont les éléments de $\{S \to aSb, S \to \varepsilon\}$ et le point de départ de la formation des éléments de L est S.

Donnons maintenant la définition formelle d'une grammaire :

Définition I.3.7. Une *grammaire* est un quadruplet $G = (A, X, P, S)$ où A et X sont deux alphabets disjoints, S est un élément de X, appelé *axiome* et P est une partie de $(A+X)^*X(A+X)^* \times (A+X)^*$, dont les éléments (M, W) sont appelés *règles de la grammaire* et notés $M \to W$.

L'alphabet A est appelé *alphabet terminal* et l'alphabet X est appelé *alphabet non-terminal*

Remarques I.3.8. Si V est un mot sur l'alphabet $A+X$ et que M est un facteur de V, l'application de la règle de la grammaire $M \to W$ consiste à substituer le facteur W au facteur M dans l'écriture de V : le mot $V = mMw$ devient mWw par application de la règle $M \to W$. Dans la littérature concernant les langages formels, ces transformations sont appellées substitutions. Nous n'emploierons jamais ce terme pour éviter les confusions avec les endomorphismes définis au paragraphe I.2.

Si $M \to W$ et $M \to W'$ sont deux règles de la grammaire, on les note de la manière condensée $M \to W + W'$.

Hiérarchie de Chomsky I.3.9. Selon les restrictions supplémentaires que l'on peut imposer sur la nature des mots V et W dans la définitions des règles de la grammaire $G = (A, X, P, S)$, on obtient une classification des grammaires, dont les principes ont été donnés par Chomsky [Cho56] [Cho59] :

type 0 Aucune restriction n'est imposée à la grammaire, on parle de *grammaire non restreinte*.

type 1 On impose que toute règle de la grammaire est de la forme : $uTv \to uMw$ avec $u, v, w \in (A+X)^*$, $T \in X$ et $M \in (A + X)^+$. Le fait que M ne puisse pas être ε est essentiel. Ces grammaires sont dites *contextuelles* ou *sensible au contexte* (context-sensitive en anglais).

type 2 On impose que toute règle de la grammaire est de la forme : $T \to M$ avec $T \in X$ et $M \in (A+X)^*$. Ces grammaires sont dites *algébriques* (context-free en anglais).

type 3 On impose que toute règle de la grammaire est de la forme : $T \to M$ avec $T \in X$ et $M \in AX + X + \varepsilon$ (ou $T \in X$ et $M \in XA + X + \varepsilon$). Ces grammaires sont dites *régulières* ou *rationnelles*.

Le qualificatif « contextuel » associé à une grammaire de type 1 vient directement de la forme des règles. En effet, elles sont du type $uTv \to uMw$, avec $u, v \in (A+X)^*$ et $T \in X$ et $M \in (A+X)^+$ donc les règles applicables à l'élément non-terminal T dépendent du contexte où T se trouve : selon les mots u et v qui encadrent S, les règles applicables peuvent être différentes. Le qualificatif « hors-contexte » que l'on associe à une grammaire de type 2 vient, de même, de la forme des règles : elles sont du type $T \to M$ avec $S \in X$ et $M \in (A+X)^*$, donc les règles applicables à T ne dépendent pas des mots qui encadrent T, c'est-à-dire qu'elles ne dépendent pas du contexte dans lequel T se trouve.

I.3.3 Langage associé à une grammaire

Définition I.3.10. Soit $G = (A, X, P, S)$ une grammaire sur l'alphabet A. On dit qu'un mot w de A^* est engendré par G s'il existe une suite de règles de P qui permet de passer de S à w. La suite de règles de S vers w est appelée *dérivation de la grammaire de S vers w*.

Le *langage $L(G)$ sur A^* engendré par la grammaire G* est l'ensemble des mots de A^* qui sont engendrés par G.

On donne à un langage le qualificatif de la grammaire de plus grand type qui l'engendre. Les langages de type 0 sont qualifiés de *récursivement énumérables*.

Si on note $\mathcal{L}_i$ l'ensemble des langages engendrés par des grammaires de type i, on a les inclusions strictes suivantes :

$$\mathcal{L}_3 \subset \mathcal{L}_2 \subset \mathcal{L}_1 \subset \mathcal{L}_0,$$

Un langage est dit *propre* lorsqu'il ne contient pas le mot vide.

Les grammaires de type 3 sont dites « régulières » car les langages associés à de telles grammaires peuvent être décrit par des expressions régulières. Nous y reviendrons au paragraphe I.4.1.

La grammaire $(\{a,b\}, \{S\}, \{S \to aSb, S \to \varepsilon\}, S)$, de l'exemple I.3.6, qui engendre $L = \{a^n b^n, n \geq 0\}$, est de type 2. La grammaire $(\{a,b\}, \{S,T\}, \{S \to aS, S \to aT, T \to Tb, T \to \varepsilon\}, S)$, qui engendre $L = a^+b^+$, est de type 3. Voici un exemple de grammaire de type 1, ainsi qu'un exemple de grammaire de type 0 :

Exemple I.3.11. Le langage sur $\{a,b,c\}$, décrit par $L = \{a^n b^n c^n,\ n > 0\}$ est engendré par la grammaire de type 1 $G = (\{a,b,c\}, X, P, S)$ avec $X = \{S,T\}$ et $P = \{S \to abc, S \to aSTc, cT \to Tc, bT \to bb\}$. C'est un langage sensible au contexte. Voici une dérivation qui engendre le mot $a^4b^4c^4$:

$$S \to aSTc \to aaSTcTc \to aaaSTcTcTc \to aaaabcTcTcTc \to aaaabTccTcTc \to aaaabbcTcTc \dots$$
$$\dots \to aaaabbTccTc \to aaaabbTcTcc \to aaaabbbcTccc \to aaaabbbbTcccc \to aaaabbbbcccc.$$

Exemple I.3.12. Le langage sur $\{a\}$ décrit par $L = \{a^{2^n},\ n > 0\}$ est engendré par la grammaire non restreinte suivante : $G = (\{a\}, X, P, S)$ avec $X = \{S, A, B, C, D, E\}$ et

$$P = \{S \to ACaB, Ca \to aaC, CB \to DB, CB \to E, aD \to Da, AD \to AC, aE \to Ea, AE \to \varepsilon\}.$$

Voici la dérivation qui engendre le mot a^8 :

$$S \to ACaB \to Aa^2CB \to Aa^2DB \to AaDaB \to ADa^2B \to ACa^2B \to Aa^2CaB \to Aa^4CB \dots$$
$$\dots \to Aa^4DB \to Aa^3DaB \to Aa^2Da^2B \to AaDa^3B \to ADa^4B \to ACa^4B \to Aa^2Ca^3B \dots$$
$$\dots \to Aa^4Ca^2B \to Aa^6CaB \to Aa^8CB \to Aa^8E \to Aa^7Ea \to Aa^6Ea^2 \to Aa^5Ea^3 \to Aa^4Ea^4 \dots$$
$$\dots \to Aa^3Ea^5 \to Aa^2Ea^6 \to Aa^7Ea^1 \to AEa^8 \to a^8.$$

En fait, A et B marquent toujours le début et la fin du mot. La lettre C agit comme un curseur, allant de gauche à droite en doublant tous les a qu'il rencontre. Lorsqu'il rencontre B, il y a alors deux possibilités :

- C se change en D et se déplace de droite à gauche sans changer le nombre de a, jusqu'à rencontrer A, alors D redevient C et le processus décrit ci-dessus peut recommencer, doublant à chaque fois le nombre de a,
- CB devient E, E se déplace de droite à gauche jusqu'à rencontrer A. La seule règle applicable est alors $AE \to \varepsilon$, donc AE disparait et on obtient un mot du type a^{2^n}.

Remarque I.3.13. Il n'est pas exclus que plusieurs grammaires engendrent le même langage, ce que sous entend la phrase « on donne à un langage le qualificatif de la grammaire de plus petit type qui

l'engendre ». Par exemple, le langage $L = a^+b^+$, présenté à l'exemple I.3.5 est également engendré par la grammaire de type 2 suivante : $(\{a, b\}, \{S\}, \{S \to aS, S \to Sb, S \to ab\}, S)$.

Dans le cas où deux grammaires G_1 et G_2 engendrent le même langage, on dira que les grammaires G_1 et G_2 sont *équivalentes*. Il existe des moyens algorithmiques de réduire les grammaires pour obtenir des grammaires équivalentes dites de *forme normale*, plus compactes. Par exemple, dans le cas des grammaires hors-contexte, il existe une forme normale fréquemment utilisée pour simplifier les preuves de résultats concernant les langages algébriques :

Théorème I.3.14. *Toute grammaire hors-contexte $G = (A, X, P, S)$ propre, c'est-à-dire dont aucune des règles n'a pour second membre le mot vide ou un élément de X, est équivalente à une grammaire G' dont toutes les règles sont de la forme $T \to aW$ avec $a \in A$ et $W \in X^*$.*

On dit que la grammaire G' est en forme normale de Greibach *de G.*

Remarque I.3.15. Il est aussi possible, pour une grammaire donnée, qu'il y ait plusieurs *dérivations* de S vers un même mot du langage, c'est-à-dire, au moins deux chaines de transformations distinctes de P, de S vers un mot du langage ; on parle alors d'*ambiguïté* de la grammaire.

De nombreuses questions de décidabilité ont été abordées concernant les grammaires : peut-on savoir avec certitude si un mot donné est engendré par une grammaire fixée ? Peut-on en modifiant la grammaire, enlever les ambiguités sans changer le langage engendré ? Nous invitons le lecteur à consulter [Eil74], [HU79] ou [RS97] pour une introduction générale à la théorie des langages, à la réduction de grammaires ainsi que pour les questions de décidabilité, d'ambiguïté et de logique.

I.3.4 Langages et séries formelles

Les qualificatifs *rationnel* et *algébrique* (introduit par N. Chomsky et M.-P. Schutzenberger [CS66]) donnés respectivement aux grammaires hors-contextes et aux grammaires régulières proviennent du fait que les séries formelles associées aux langages engendrés par ces grammaires sont respectivement algébriques et rationnelles.

Ce lien entre langages et séries formelles, qui passe, bien entendu, par l'intermédiaire des grammaires, est ici présenté de manière simpliste. Pour une étude plus poussée, nous invitons le lecteur à consulter [ABB97] :

On introduit la notion de série formelle à coefficients dans un demi-anneau $(K, +, \cdot)$, c'est-à-dire, de manière informelle, à coefficients dans un anneau duquel on aurai retiré les inverses pour les deux lois ($(K, +)$ et $(K, \cdot)$ sont deux monoïdes de la multiplication est distributive par rapport à l'addition) et qui vérifie $k \cdot 0 = 0 \cdot k = 0$.

Une série formelle S, sur l'alphabet A, à coefficients dans K est une application de A^* vers K. On note S_w l'image de $w \in A^*$, et S est notée

$$S = \sum_{w \in A^*} S_w w.$$

On note $K \ll A \gg$ l'ensemble des séries formelles sur l'alphabet A, à coefficients dans K. Cet ensemble a une structure de demi-anneau. le demi-anneau $K \ll A \gg$ peut être muni de la topologie produit associée à la topologie discrète sur K. Cette topololologie pouvant être définie à l'aide d'une distance qui le rend complet, on peut définir, comme pour les séries entières, une notion de sommabilité, qui permet, entre autre, de définir l'opérateur $*$ sur $K \ll A \gg$ pour les séries formelles propres (dont le terme constant est nul), par $S^* = \sum_{n \geq 0} S^n$.

En particulier, si U et T sont deux séries formelle, avec T propre, alors l'équation $S = TS + U$ (*resp.* $S = ST + U$) admet une unique solution : $S = T^*U$ (*resp.* $S = UT^*$).

Pour E une partie de $K \ll A \gg$, on appelle clôture rationnelle de E la plus petite partie de $K \ll A \gg$ contenant E et stable par addition, multiplication, multplication par un élément de K et opérateur $*$ (Cette opération remplace en quelque sorte l'inversion). L'ensemble $K \ll A \gg$ est quant à lui rationnellement clos. N. Chomsky et M.-P. Schützenberger [CS66] ont aussi défini dans ce cadre la clôture algébrique de $K \ll A \gg$.

Définition I.3.16. Système d'équations associé à une grammaire $G = (A, X, P, S)$.

On associe à chaque élément x de $X + A + \{\varepsilon\}$ un langage L_x par une application $\Lambda : X \to \mathcal{P}((A+X)^*)$ telle que :

– $\Lambda(\varepsilon) = \varepsilon$,
– $\Lambda(a) = a$, pour tout élément a de A,
– $\Lambda(T) = L_T$, pour tout élément T de X,
– $\Lambda(MW) = \Lambda(M)\Lambda(W)$, pour tous éléments M et W de $(A + X)^*$,
– $\Lambda(M + W) = \Lambda(M) + \Lambda(W)$, pour tous éléments M et W de $(A + X)^*$.

Le système d'équations Σ_G associé à G est la donnée des égalités

$$(\Sigma_G) : \qquad \forall (M \to W) \in P, \quad \Lambda(M) = \Lambda(W).$$

Si Σ_G admet une solution $L = (L_T,\ T \in X)$, le langage $L_S \cap A^*$ est alors le langage engendré par la grammaire $G = (A, X, P, S)$.

Par exemple, on associe à la grammaire hors-contexte $G = (A, X, P, S)$, définie par $A = \{a, b\}$, $X = \{S, T\}$, $P = \{S \to ST + \varepsilon, T \to aTb\}$, qui engendre le langage des mots bien parenthésés (avec $a = ($ et $b =)$), le système d'équations suivant :

$$\begin{cases} L_S = L_S L_T + \varepsilon, \\ L_T = aL_T b. \end{cases}$$

Lorsque la grammaire est régulière, les égalités $\Lambda(M) = \Lambda(W)$ de Σ_G, sont linéaires en les variables L_T, à cause de la forme des règles d'une grammaire régulière. De même, lorsque la grammaire est hors-contexte et sous forme normale de Greibach, ces égalités sont polynomiales en les variables $\Lambda(L_T)$ pour $T \in X$.

Ainsi, puisque les composantes des solutions d'un système linéaire (*resp.* polynomial) sont dans la clôture rationnelle (*resp.* algébrique) de l'ensemble des coefficients du système, on parle alors, par analogie, de langage rationnel pour tout langage régulier et de langage algébrique pour tout langage hors-contexte.

Par la suite, nous allons nous intéresser plus particulièrement aux langages réguliers et aux langages hors-contextes, et developper les différents aspects qui les relient aux séries formelles et les automates. En ce qui concerne les machines associées aux les langages sensibles au contexte (machines linéairement bornées), nous invitons le lecteur à consulter [MS01] [MR05] ou [CM06].

I.4 Langages rationnels et automates finis

De nombreux aspects des langages rationnels ont fait l'objet de travaux car, d'une manière générale, ces langages possèdent de bonnes propriétés : ils sont reconnus par des automates finis, leur séries génératrices sont rationnelles, leurs grammaires sont régulières et ils ont de bons critères de décidabilité et de logique... Le seul désavantage, qui découle de cela, est une question de vocabulaire : on les retrouve dans la littérature sous les termes de *langages réguliers, langages rationnels* ou encore *langages reconnaissables* (ce qui en fait des langages légèrement schizophrènes !). Ce sont pourtant les mêmes objets. Ce paragraphe expose en partie les liens entre langages rationnels, expressions régulières et automates finis.

Dans ce paragraphe, A désigne un alphabet fini.

I.4.1 Langages rationnels et expressions régulières

Proposition I.4.1. *La classe $\mathcal{L}_3$ des langages rationnels sur un alphabet A est la plus petite classe de langages contenant les langages finis et stables par addition, produit et opérateur $*$ (passage au sous-monoïde engendré) et l'ensemble des langages rationnels est stable par :*
– *passage au complémentaire (dans E^*),*
– *intersection,*
– *union,*
– *produit et opérateur $*$,*
– *homomorphisme et substitution,*
– *quotient à droite : si L_1 et L_2 sont deux langages sur A, le quotient à droite de L_1 par L_2 est définit par $L_1/L_2 = \{m \in L_1,\ \exists w \in L_2, mw \in L_1\}$.*

On peut ainsi caractériser les langages rationnels de la manière suivante :

Proposition I.4.2. *Un langage est engendré par une grammaire régulière, s'il peut être défini par une* expression régulière *r, c'est-à-dire un élément de la clôture rationnelle de $A + \{\varepsilon\}$, pour les opérations d'addition, de produit de concaténation et opérateur étoile $*$.*

La famille $\mathcal{R}$ des expressions régulières est, en fait, la plus petite famille de phrases mathématiques qui vérifie $A + \{\varepsilon\}$ et $\emptyset$ sont inclus dans $\mathcal{R}$ et si r et r' sont deux expressions régulières, $r + r'$, rr' et r^* sont aussi régulières.

La définition faite ici doit s'entendre comme suit : une expression régulière est un mot sensé sur l'alphabet $A + \{\varepsilon, \emptyset, (,), +, \cdot, *\}$, où $\cdot$ représente le produit de concaténation et où l'étoile de Kleene $*$ est distributive par rapport à l'addition.

Exemple I.4.3. L'expression régulière $r = (a+ba)^*(b+\varepsilon)$ définit le langage $L(r)$ des mots sur l'alphabet $\{a, b\}$ qui ne contiennent pas deux b consécutifs. Son complementaire $E^* \setminus L(r)$ est aussi un langage rationnel : le langage L des mots sur l'alphabet $\{a, b\}$ qui contiennent deux b consécutifs peut être définit par l'expression régulière $r' = (a + b)^*bb(a + b)^*$.

Cette caractérisation des langages rationnels par les expressions régulières est en fait une conséquence du lien entre langages rationnels et séries formelles rationnelles, que l'on peut condenser par la proposition ci-dessous :

Proposition I.4.4. *Un langage L sur l'alphabet A est rationnel si et seulement s'il existe une série rationnelle $\mathcal{S} = \sum_{w \in A^*} \mathcal{S}_w w$ de $\mathbb{N} \ll A \gg$, anneau des séries entières indexées sur A^*, dont L est le support (*i.e. *$\mathcal{S}_w \neq 0$ équivaut à $w \in L$).*

D'autre part, la série caractéristique d'un langage rationnel L : $\mathcal{S}_L = \sum_{w \in L} w$ est une série rationnelle de $\mathbb{N} \ll A \gg$.

Cette proposition est une conséquence d'un théorème de Schützenberger [Sch61, Sch62b, Sch62a] qui donne l'équivalence entre les notions de reconnaissabilité et rationnalité pour les séries formelles à valeurs dans un demi-anneau quelconque. Nous renvoyons le lecteur à [BR84] pour une démonstration.

I.4.2 Automates finis

Les automates finis sont des « machines mathématiques » très simples ; ce sont des accepteurs. On leur donne une chaîne de caractères —un mot sur un alphabet donné—, chaque lettre lue change la position de l'automate (un nombre fini de positions étant possibles), après lecture successive des lettres de ce mot, l'automate se fige et renvoie une information du type « oui/non »ou plus généralement une information pouvant prendre un nombre fini de valeurs, selon sa position finale. Plus formellemnt :

Définitions I.4.5. Un *automate fini* $\mathcal{A}$ est un quintuplet du type $\mathcal{A} = (E, A, \Phi, e_0, F)$, où :
- E est un ensemble fini, appelé *ensemble des états*,
- A est un alphabet fini,
- Φ est une partie de $E \times A \times E$ dont les éléments sont appelés *transitions*,
- I est un sous-ensemble de E, appelé *ensemble des états initiaux*,
- F est un sous-ensemble de E, appelé *ensemble des états finaux.*

Si $Card(I) = 1$ et si pour tout couple (e, a) de $E \times A$, l'ensemble $\Phi \cap \{e\} \times \{a\} \times E$ est au plus un singleton, l'automate est dit *déterministe*. L'ensemble des transitions Φ peut être vu comme une fonction définie sur une partie de $E \times A$ et le triplet $(e, a, e') \in \Phi$ est noté $e \cdot a = e'$. Si, de plus, pour tout couple (e, a) de $E \times A$, il existe un élément e' de E tel que $(e, a, e') \in \Phi$, l'automate est dit *complet*.

On note AFD un automate fini déterministe et complet, et AFN un automate non-déterministe.

Lorsqu'on a un AFD $\mathcal{A} = (E, A, \phi, e_0, F)$, on étend la définition Φ au monoïde A^* de la manière suivante : on définit, pour tout $e \in E$, $e \cdot \varepsilon = e$ et pour tout mot v de A^* et toute lettre a de A, on pose $e \cdot (va) = (e \cdot v) \cdot a$.

On appelle *trajectoire* ou *exectution d'un mot d'entrée $m = m_0 m_1 \dots m_n \in A^*$ dans l'automate*, le $(n + 1)$-uplet $(e_0, e_1, \dots, e_{n+1})$ tel que pour tout $i \in [\![0, n]\!]$, $e_{i+1} = e_i \cdot m_i$, de sorte que $e_{i+1} = e_0 \cdot (m_0 m_1 \dots m_i)$.

Le *langage accepté par l'automate* $\mathcal{A}$, noté $L(\mathcal{A})$ est défini par :

$$L(\mathcal{A}) = \{w \in A^*,\ e_0 \cdot w \in F\}.$$

C'est l'ensemble des mots dont les trajectoires dans l'automate finissent sur un élément de F. Un langage est *reconnaissable* s'il est accepté par un automate fini.

On représente un automate par son *graphe des transitions*, graphe dont les sommets sont indexés par E où les éléments $e \cdot a = s$ de Φ sont symbolisés par des flèches de e vers s étiquetées par a.

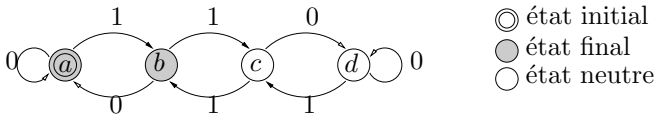


FIG. I.4 – AFD complet associé au langage de $\{0,1\}^*$ des mots contiennant un nombre pair de motifs 11

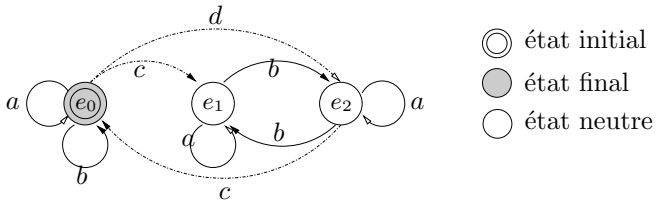


FIG. I.5 – AFD non complet associé à la transformation de Chacon.

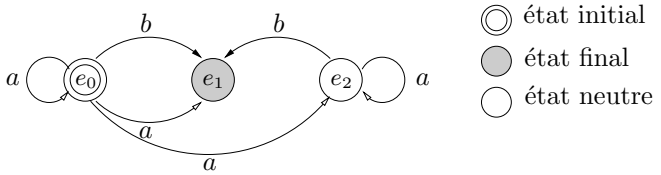


FIG. I.6 – Un exemple d'AFN.

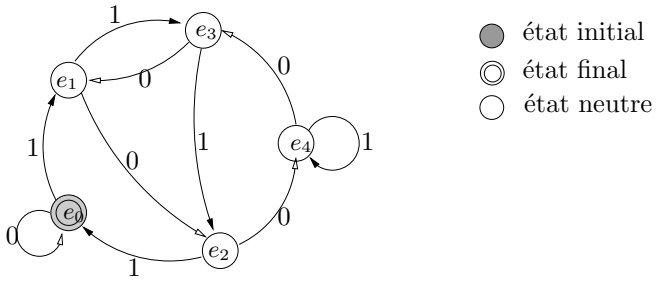


FIG. I.7 – AFD reconnaissant l'ensemble des écriture propres en base 2 des entiers divisibles par 5.

Remarque I.4.6. En ordonnant arbitrairement l'alphabet $A = \{a_0, a_1, \ldots, a_{q-1}\}$, on peut représenter la fonction de transition ϕ d'un AFD complet sous la forme d'une substitution σ, définie sur E, de la manière suivante : $\sigma(e) = (e \cdot a_0)(e \cdot a_1)\ldots(e \cdot a_{q-1})$.

On peut remplacer la donnée de F dans la définition de l'automate $\mathcal{A} = (E, A, \Phi, e_0, F)$ par la donnée, plus générale, d'un autre alphabet A et une fonction $\Pi : E \to A'$, appelée *fonction de sortie*. La fonction de sortie Π peut, par exemple, être une *projection lettre-à-lettre* de E vers A' ou une *projection lettre-à-mot* (ou codage) si A' est un langage fini sur un autre alphabet A.

Définition I.4.7. Un *automate fini déterministe avec fonction de sortie* (AFDS) est la donnée du sextuplet $\mathcal{A} = (E, A, \phi, e_0, A', \Pi)$ où

– E est un ensemble fini, appelé *ensemble des états*,

– A est un alphabet,
– ϕ est une fonction de $E \times A$ vers E dont les éléments sont appelés *transitions*. Un triplet $(e, a, s) \in \Phi$ est noté $e \cdot a = s$;
– e_0 est un élément de E, appelé *état initial*,
– E' est un alphabet,
– Π est un codage de E vers A'^*, appelé *fonction de sortie*.

Un AFDS $\mathcal{A}$ fonctionne alors, informellement, de la manière suivante : on "rentre" un mot w dans l'automate, l'automate se déplace d'états en états, en suivant les instructions de la fonction de transition, au fur et à mesure que l'automate lit les lettres successives de w. Lorsque toutes les lettres de w ont été lues, l'automate s'arrête, sur un état e, et renvoie la sortie $\Pi(e)$.

La représentation d'un AFD (E, A, ϕ, e_0, F) peut être alors vue comme un cas particulier d'AFDS du type $(E, A, \phi, e_0, A', \Pi)$. En effet, le modèle de l'automate (E, A, ϕ, e_0, F) est avant tout une machine qui accepte ou rejette des mots w de A^*, selon que $e_0 \cdot w$ appartient ou n'appartient pas à l'ensemble des états terminaux F. On peut donc interpréter ce modèle comme une fonction de A^* dans $\{0, 1\}$ qui renvoit 1 si w est accepté et 0 sinon ; et cela revient, en fait, à remplacer la donnée de F par la donnée de la fonction Π_F de E vers $\{0, 1\}$, définie par $\Pi_F^{-1}(\{1\}) = F$: on a $\Pi_F(e_0 \cdot w) = 1$ si et seulement si w est accepté par l'automate.

Lorsque le monoïde A^* des entrées de l'AFDS $(E, A, \phi, e_0, A', \Pi)$ est totalement ordonné, on peut définir un mot infini associé à cet automate.

Définition I.4.8. Soit $\mathcal{A} = (E, A, \phi, e_0, A', \Pi)$ un AFDS.

Si le monoïde A^* est ordonné, $A^* = \{w_0 < w_1 < w_2 \ldots\}$, alors, on appelle *mot de sortie de l'automate* $\mathcal{A}$ le mot infini $\mu(\mathcal{A})$ défini par :

$$\forall n \in \mathbb{N},\ \mu(\mathcal{A})_n = \Pi(e_0 \cdot w_n).$$

Nous reviendrons sur la nature des ces mots dans le paragraphe I.6.

Une façon d'étendre le concept des automates finis est de considérer des machines (E, A, Φ, e_0, F) où Φ est une partie de $E \times \mathcal{R} \times E$, où $\mathcal{R}$ désigne l'ensembles des expressions régulières sur A ; on parle alors d'*automates finis généralisés* (AFG). Par exemple, la figure I.8 présente un automate généralisé qui reconnait le langage défini par l'expression régulière $1 + 01^*$.

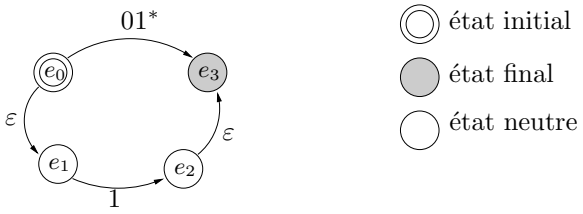


FIG. I.8 – Un exemple d'AFG qui reconnait le langage $1 + 01^*$.

Les AFN et les AFD sont des cas particuliers d'AFG : ce sont les AFG pour lesquels toutes les transitions sont étiquetées par des éléments de A mais, outre le fait que les AFD sont plus précis, ces notions sont en fait presque équivalentes car on peut construire à partir d'un AFG, un AFD qui reconnait le même langage. C'est cette construction qui est à la base de la démonstration du résultat qui relie les automates finis et les langages rationnels : le théorème de Kleene.

Théorème I.4.9. *(Kleene 1956)*
Un langage est rationnel si et seulement s'il est reconnaissable par un AFD.

■ **Démonstration du théorème I.4.9** Soit L un langage rationnel et r une expression régulière telle que $L = L(r)$. Ce langage L est accepté par l'AFG $G_0 = \{E_0 = \{e_0, e_1\}, A, \Phi_0, e_0, F_0 = \{e_1\}\}$, avec $\Phi_0 = \{(e_0, r, e_1)\}$. L'arc du graphe de G_0 est étiqueté par l'expresssion régulière r. On transforme le graphe de l'automate de manière réccursive : tant que toutes les expressions qui étiquettent les arcs ne

sont pas dans $A+\{\varepsilon\}+\{\emptyset\}$, on transforme les arcs et les états du graphe, en utilisant les transformations précisées dans la figue I.9.

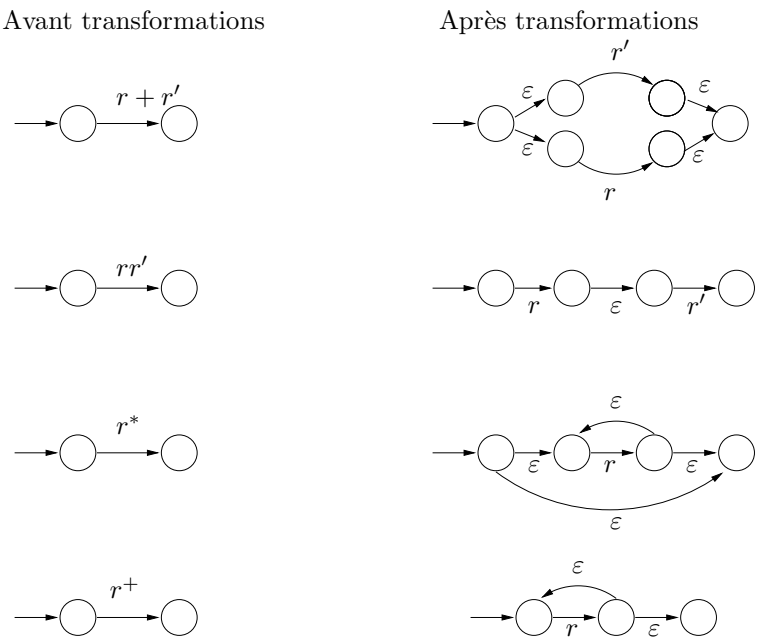


FIG. I.9 – Transformations permettant de passer d'un AFG à un AFN.

L'itération de ces opérations aboutit à un AFN $G_1 = \{E_1, A + \{\varepsilon\} + \{\emptyset\}, \Phi_1, e'_0, F_1\}$ reconnaissant L, à partir de l'AFN G_0 qui reconnait L.

Soit $G_2 = \{E_2, A, \Phi_2, e'_0, F_2\}$ définit de la manière suivante :

- $E_2 = E_1$,
- $(e, a, e') \in \Phi_2$ si et seulement si il existe une chaine d'éléments de Φ_1 :
  $(e, \alpha_1, e_1)(e_1, \alpha_2, e_2)\dots(e_{n-1}, \alpha_n, e_n)(e_n, \alpha_{n+1}, e')$ telle que $a \in \alpha_1\alpha_2\dots\alpha_{n+1}$,
- F_2 est construit à partir de F_1 en y ajoutant l'état e'_0 s'il existe une chaine d'éléments de Φ_1 :
  $(e'_0, \alpha_1, e_1)(e_1, \alpha_2, e_2)\dots(e'_{n-1}, \alpha_n, e_n)(e_n, \alpha_{n+1}, e')$, où un des α_i est ε et $e' \in F_1$.

L'AFN ainsi fabriqué $G_2\{E_2, A, \Phi_2, e'_0, F_2\}$, reconnait toujours L.

Il ne reste plus qu'à construire un automate G_3 déterministe à partir de G_2. Pour cela, on construit l'automate G_3 dont l'ensemble les états est indexé par les parties de E_2 : $E_3 = \mathcal{P}(E_2)$, la fonction de transition ϕ est donnée, pour tout $S \in E_3$ par $\phi(S, a) = \cup_{s\in S}\{e,\ (s, a, e) \in \Phi_2\}$, l'état initial est $\{e'_0\}$ et l'ensemble des états finaux est $F_3 = \{S \in \mathcal{P}(E_2),\ S \cap F_2 \neq \emptyset\}$. (On montre par récurrence le fait que G_2 et G_3 reconnaissent le même langage L).

Pour ce qui est de la réciproque, on considère un AFD $G = \{E, A, \Phi, e_0, F\}$ qui reconnait L et on construit à partir de G un AFG du type $G_0 = \{E_0 = \{e_0, e_1\}, A, \Phi_0, e_0, F_0 = \{e_1\}\}$, avec $\Phi_0 = \{(e_0, r, e_1)\}$. Cette construction est basée sur les principes décrits ci-dessous :

On traite G comme un AFG. On rajoute un état e'_0 qui sera notre nouvel état initial ainsi que la transition (e'_0, ε, e_0) dans Φ. On rajoute également un état f, qui sera notre seul nouvel état final, et, pour tout état $e' \in F$, on rajoute la transition (e', ε, f) dans Φ. Ensuite, on élimine successivement tous les états qui ne sont pas e'_0 ou f de la manière suivante :

Si $(e, r_1, e')(e', r_2, e'')$ sont des éléments de Φ, alors ces deux éléments sont éliminés de Φ et remplacés par (e, r_1r_2, e'') dans Φ.

Si (e, r_1, e') et (e, r_2, e') sont des éléments de Φ, alors ces deux éléments sont éliminés de Φ et remplacés par $(e, r_1 + r_2, e')$ dans Φ.

Si (e, r, e) est un élément de Φ, alors cet élément est éliminé de Φ et remplacé par (e, r^*, e) dans Φ.

La figure I.10 présente le passage de l'AFD qui reconnait un langage à l'expression régulière qui le définit sur l'exemple du langage L des mots sur l'alphabet $\{a, b\}$ qui contiennent un nombre pair de b. ■

Pour clore ce paragraphe, nous donnons une propriété des langages rationnels, connue sous le nom de *lemme de pompage* (*pumping lemma* en anglais) qui est un outil très utile pour prouver la non-rationnalité

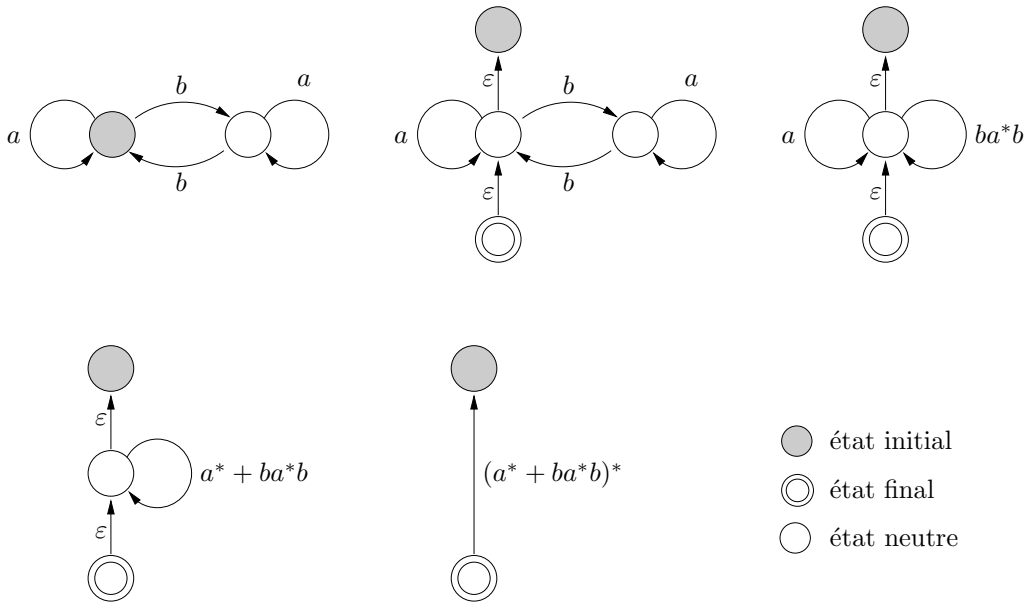


FIG. I.10 – Passage de l'AFD à l'AFN sur un exemple.

d'un langage.

Lemme I.4.10. *Soit L un langage rationnel sur l'alphabet A. Il existe un entier strictement positif n tel que, pour tout mot m de L vérifiant $|m| \geq n$, il existe une décomposition de $m = uvw$ où u, v et w sont des mots de L tels que :*

$$|uw| \leq n,\ |v| \geq 1 \text{ et, pour tout } i \in \mathbb{N},\ uv^i w \in L.$$

■ **Démonstration du lemme I.4.10** L'entier n peut être choisi égal au nombre d'états de l'automate qui engendre L ; en effet, la trajectoire que suit un mot m de longueur supérieure à n lorsqu'il est entré dans l'automate passe alors par au moins $(n+1)$ états, donc passe forcément deux fois par un même état (c'est le principe des tiroirs), ainsi, la trajectoire de m : $(e_0, e_1, \ldots, e_{|m|+1})$ fait une « boucle » dans l'automate : il existe $i < j$ tels que $e_i = e_j$. Ce sont les étiquettes des arcs utilisés pour parcourir cette boucle qui donne le mot v.

Le mot u est donné par le préfixe de m, donné par les étiquettes successives des arcs utilisés pour le morceau de la trajectoire de m qui va de e_0 à e_i et w est le suffixe de m donné par les étiquettes successives des arcs utilisés pour le morceau de la trajectoire de m qui va de e_j à $e_{|m|+1}$. ■

I.5 Langages algébriques et automates à pile

Dans ce paragraphe, A désigne un alphabet fini.

De la même manière que pour les langages rationnels, les langages algébriques sont les langages reconnus par un certains type de machines : les automates à pile.

La correspondance langages algébriques/automates à pile est cependant moins satisfaisante que la correspondance langages rationnels/automates finis. En effet, les automates à pile reconnaissent les langages algébriques mais sont des machines non-déterministes et il n'y a pas, comme pour les automates finis, équivalence entre les notions de reconnaissance par automate à pile non-déterministe et de reconnaissance par automate à pile déterministe. De plus, la famille des langages reconnus par des automates à pile déterministes n'est pas stable par intersection, ni par union.

Néanmoins, les automates à pile déterministes reconnaissent des langages algébriques parmi lesquels figurent la majeure partie des langages de programmation.

Les automates à pile sont des modèles mathématiques que l'on pourrait apparenter à des automates finis non déterministes ayant une mémoire. Cette mémoire (écrite avec un autre alphabet), pouvant

potentiellement être infinie, est stockée dans une « pile » qui se vide ou se remplit au fil de la lecture du mot d'entrée, en suivant le principe du « dernier entré–premier sorti » :

Lorsqu'on lit une lettre d'entrée, couplée avec la lettre du dessus de pile : on retire cette lettre de la pile et on pose sur la pile un autre mot (qui peut être vide) qui dépend de la lettre d'entrée et de celle du haut de pile.

Nous donnons, pour deux exemples de langages algébriques, la construction plus visuelle que formelle d'un automate à pile qui les reconnait.

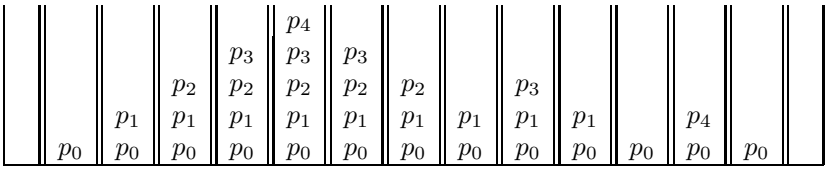
Exemple I.5.1. Le langage de Dyck sur $n+1$ symboles ($n \in \mathbb{N}$) est un exemple important de langage algébrique : c'est le langage des mots bien parenthésés sur $n+1$ types de parenthèses. Plus formellement, c'est le langage sur l'alphabet $A = \{p_0, p_1, p_2, \dots p_n\} + \{\overline{p}_0, \overline{p}_1, \overline{p}_2, \dots \overline{p}_n\}$ engendré par la grammaire $G = \{A, \{S, T\}, P, S\}$ où les éléments de P sont : $S \to ST + \varepsilon$ et, pour tout i, $T \to p_i T \overline{p}_i$.

Les parenthèses ouvrantes sont les p_i et les parenthèses fermantes correspondantes sont les $\overline{p}_i$.

On considère le dispositif suivant, qui permet de conserver en mémoire les parenthèses ouvertes qui n'ont pas encore été fermées :

Soit m un mot sur l'alphabet A, qui l'on lit de gauche à droite. Si je lis une lettre p_i, alors, je la mets sur la pile ; si je lis une lettre $\overline{p}_i$, alors, puisque la lettre de pile est forcément p_i (la dernière parenthèse ouverte est la première que l'on ferme), on la retire de la pile.

Par exemple, voici l'évolution de la pile lorsque l'on applique cet algorithme au mot de Dyck $m = p_0p_1p_2p_3p_4\overline{p}_4\overline{p}_3\overline{p}_2p_3\overline{p}_3\overline{p}_1p_4\overline{p}_4\overline{p}_0$:



Lorsqu'on part de la pile vide, ces mouvements de pile ne sont possibles que lorsque le mot lu est bien un mot du langage de Dyck. D'autre part, la pile est vide à la fin de la lecture d'un mot si et seulement si le mot appartient au langage de Dyck.

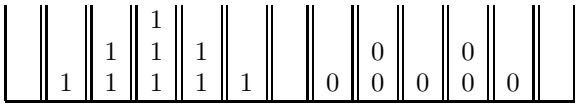
Exemple I.5.2. Considèrons l'ensemble des représentations propres des entiers en base 2 qui contiennent autant d'occurrences de 0 que d'occurrences de 1. Ce langage est engendré par la grammaire algébrique suivante : $G = (\{0,1\}\{S,T\}, P, S)$ où $P = \{S \to 1T, T \to 0T1 + 1T0 + 0\}$.

Pour reconnaître les mots de ce langage, on met en place le dispositif suivant :

On lit un mot sur l'alphabet $\{0,1\}$: on part d'une pile vide. La première lettre lue doit être un 1, puisqu'on ne retient que les écritures en base 2 qui sont propres. On met alors ce premier 1 dans la pile. On applique ensuite l'algorithme suivant :

- si la pile n'est pas vide et que le symbole du dessus de pile est un 1, la lecture d'un 1 rajoute 1 au dessus de la pile et la lecture de 0 enlève le 1 du haut de pile,
- si la pile est vide, on rajoute le symbole lu à la pile,
- si la pile n'est pas vide et que le symbole du dessus de pile est un 0, la lecture d'un 0 rajoute 0 au dessus de la pile et la lecture de 1 enlève le 0 du haut de pile.

Par exemple, voici l'évolution de la pile lorsque l'on applique au mot $m = 111000001011$ cet algorithme :



Donnons maintenant une définition formelle d'un automate à pile.

Définition I.5.3. Un *automate à pile* (AP) (*pushdown automata* en anglais) est la donnée d'un septuplet $\mathcal{A} = (E, A, P, \phi, e_0X_0, F)$ où :

- E est un alphabet fini, appelé *ensemble des états*,
- A est un alphabet fini, appelé *ensemble des entrées*,
- P est un alphabet fini, appelé *ensemble des symboles de pile*,
- ϕ est une application de $E \times (A + \{\varepsilon\}) \times (P + \{\varepsilon\})$ vers l'ensemble des parties finies de EP^* est une application appelée *fonction de transition*,

– e_0X_0 est un élément de EP, appelé *configuration interne initiale*
– F est un sous-ensemble de EP^*, appelé *ensemble des configurations internes finales.*

Un AP est *déterministe* lorsqu'il vérifie les conditions suivantes :
– quelque soit le triplet (e, a, X) de $E \times (A + \{\varepsilon\}) \times P + \{\varepsilon\}$, $\phi(e, a, X)$ est au plus un singleton, c'est-à-dire que ϕ est une fonction d'une partie de $E \times (A + \{\varepsilon\}) \times (P + \{\varepsilon\})$ vers $E \times P^*$,
– Si $\phi(e, \varepsilon, X) \neq \emptyset$, alors, pour tout $a \in A$, $\phi(e, a, X) = \emptyset$.

On parle d'*automate à pile en temps réel* lorsque $\phi(e, \varepsilon, X) = \emptyset$, pour tout couple $(e, X) \in E \times P$. Si de plus, E est réduit à un singleton, alors on parle d'*automate à pile simple.*

Pour pouvoir décrire correctement l'automate à pile $\mathcal{A} = (E, A, P, \phi, e_0X_0, F)$, on a besoin d'introduire la notion de configuration. Une *configuration* est un triplet d'éléments (e, m, W) de $E \times A^* \times P^*$ où m est le mot d'entrée qui reste à lire et W est le mot stocké dans la pile. Pour $a \in A^*$ et $X \in P$, on note $(e, am, XW) \to (e', m, X'VW)$ si $\phi(e, a, X)$ contient $(e', X'V)$ (ici, on a fait apparaitre en plus la première lettre du mot stocké dans la pile, qui représente le symbole du haut de pile) ; la flèche $\to$ symbolise le mouvement de l'automate lors de la lecture de la première lettre du mot am : lorsque l'automate est sur l'état e et que la lettre du dessus de pile est X, la lecture de a fait passer l'automate sur l'état e', enlève X du dessus de la pile et le remplace par le mot $X'V$, de sorte que la nouvelle lettre du dessus de pile est X'.

S'il existe un chemin de l'automate permettant de passer de la configuration (e, m, XW) à la configuration $(e', m', X'W')$ en n mouvements de l'automate, on note cela $(e, m, XW) \xrightarrow{n} (e', m', X'W')$ et on définit par extension la flèche $\xrightarrow{*}$: $(e, m, XW) \xrightarrow{*} (e', m', X'W')$ équivaut à dire qu'il existe un entier n pour lequel $(e, m, XW) \xrightarrow{n} (e', m', X'W')$.

Une *configuration interne* est un mot eW de EP^* (le couple $(e, W) \in E \times P^*$ est donné sous forme concaténée), où e est un état de l'AP et W est un mot qui représente le haut de pile ou la pile entière (la première lettre de W est le symbole du dessus de pile). On introduit, pour tout mot m de A^*, la relation $\overset{m}{\Longrightarrow}$ entre les configurations internes, définie par :

$$eW \overset{m}{\Longrightarrow} e'W' \text{ si et seulement si } (e, m, W) \xrightarrow{*} (e', \varepsilon, W').$$

La relation $eW \overset{m}{\Longrightarrow} e'W'$ signifie que lorsqu'on est dans l'état e et que le haut de pile est le mot W, alors la lecture du mot m par l'automate change l'état e en l'état e' et sur le dessus de la pile, W est remplacé par W' on dit alors que $e'W'$ est une *configuration interne accessible depuis* eW. Lorsque $e'W'$ est accessible depuis eW et eW est accessible depuis $e'W'$, on dit que les couples eW et $e'W'$ sont des *configurations co-accessibles* ou *inter-accessibles.*

Puisqu'on a la relation $\overset{m}{\Longrightarrow}\overset{m'}{\Longrightarrow} = \overset{mm'}{\Longrightarrow}$ pour tous mots m et m' de A^*, tous les mouvements possibles entre les configurations internes de l'automate, c'est-à-dire la donnée d'ensembles des transitions Φ, sont donnés par l'ensemble des flèches

$$\mathcal{M} = \left\{ eX \overset{a}{\Longrightarrow} e'X'V,\ (e, e', a, X, X', V) \in E^2 \times A + \{\varepsilon\} \times P^2 \times P^* \right\}.$$

Remarque I.5.4. On peut aussi définir les configurations en choisissant de décrire la pile avec une récriture suffixe plutôt que préfixe : la pile dont la lettre du dessus est X s'écrirait alors WX et l'action de l'automate ferait passer la pile de WX à WVX'. Cettre écriture semblerait plus naturelle : on écrit le mot de pile de gauche à droite et l'action de l'automate gomme la dernière lettre et rajoute un mot à la fin. C'est cependant la réécriture préfixe qui a été retenue dans la majorité de la littérature qui concerne les automates à pile car elle est finalement plus facile à manipuler.

Définition I.5.5. On dit qu'un mot m sur l'alphabet A est reconnu par l'AP $\mathcal{A} = (E, A, P, \phi, e_0X_0, F)$ s'il existe une configuration interne finale eW telle que $e_0X_0 \overset{m}{\Longrightarrow} eW$.

Le langage $L(\mathcal{A})$ sur l'alphabet A reconnu par l'AP est l'ensemble des mots reconnus.

Il y a plusieurs manières d'engendrer des langages à l'aide d'automates à pile, selon la nature des configurations finales (nature de K) que l'on choisit ; voici les différents modes d'acceptation usuels pour un AP :
– si $F = F'P^*$ avec $F' \subset E$, on parle d'*acceptation par états finaux*,
– si $F = E\varepsilon$, on parle d'*acceptation par pile vide*,
– si $F = F'\varepsilon$ avec $F \subset E$, on parle d'*acceptation par états finaux et pile vide*,
– si $F = EXP^*$, pour un certaine lettre X de P, on parle d'*acceptation par lettre de sommet de pile.*

Remarques I.5.6. Il est possible que les langages engendrés par un même AP mais par les différents modes d'acceptation soient différents (voir exemple I.5.10), mais il existe des « passerelles »entre ces differents modes d'acceptation : si un langage est reconnu par un AP par états d'acceptation, il existe un autre AP qui reconnait ce même langage par pile vide. De même, si un langage est reconnu par un AP par pile vide et états d'acceptation, il existe un autre AP qui reconnait ce même langage par lettre de sommet de pile... voir [Aut87]. Ces résultats permettent en fait de pouvoir parler de *langage reconnu par un AP* sans préciser le mode d'acceptation.

D'autre part, de la même manière que pour les automates finis, il se peut que plusieurs AP reconnaissent le même langage et il existe des moyens de modifier les AP sans changer le langage engendré, pour obtenir des AP de forme sympathique sans, par exemple, de mouvements entre configurations internes indéxés par ε. Voir par exemple [Aut87] [ABB97] ou [Sti05] pour différentes approches et pour les questions d'équivalence entre deux AP et d'accessibilité entre configurations.

Théorème I.5.7. *La famille des langages de A^* reconnus par les AP est exactement la famille des langages algébriques sur l'alphabet A.*

La famille des langages de A^* reconnus par les AP simples est exactement la famille des langages algébriques propres sur l'alphabet A.

Nous invitons le lecteur à consulter [Aut87] ou [ABB97] pour la démonstration de ce théorème. Pour montrer qu'il existe un AP qui reconnait un langage algébrique L fixé, l'idée de la démonstration passe par la grammaire sous forme normale de Greibach $G = (A, P, \mathcal{P}, S)$ qui engendre L, lorsque L est un langage propre (voir le théorème I.3.14). L'AP $\mathcal{A} = (\{e\}, A, P, \phi, (e, \varepsilon), K)$ qui reconnait L est construit avec les équivalences entre les règles de la grammaire et les mouvements de configurations internes suivantes :

$$\forall (T, a, W) \in P \times A \times P^*,\ eT \overset{a}{\Longrightarrow} eW \in \mathcal{M} \Longleftrightarrow T \to aW \in \mathcal{P}.$$

Lorsque L n'est pas propre, il suffit de rajouter un symbole de pile Y à l'automate simple qui reconnait $L \setminus \{\varepsilon\}$ et de rajouter les deux transitions $eX \overset{\varepsilon}{\Longrightarrow} eY$ et $eY \overset{\varepsilon}{\Longrightarrow} e\varepsilon$.

Exemple I.5.8. L'automate à pile simple décrivant le langage de Dyck sur $n+1$ types de parenthèses dans l'exemple I.5.1 est donné par $\mathcal{A}_n = (E, A, P, \phi, (e, \varepsilon), F)$ où $E = \{e\}$, $A = \{p_1, p_2, \ldots, p_n\} + \{\overline{p}_1, \overline{p}_2, \ldots, \overline{p}_n\}$, $P = \{p_1, p_2, \ldots, p_n\}$ et les transitions de ϕ sont données par :

$$\begin{aligned} &e\varepsilon \overset{p_i}{\Longrightarrow} ep_i, \\ &ep_j \overset{p_i}{\Longrightarrow} ep_ip_j, \\ &ep_i \overset{\overline{p}_i}{\Longrightarrow} e\varepsilon, \end{aligned}$$

et la configuration interne finale est l'élément $e\varepsilon$. Le langage de Dyck sur n type de parenthèses est reconnu par l'AP simple $\mathcal{A}_{n+1}$ par pile vide.

Ici, puisque l'automate est simple, on aurait pu l'oublier dans l'écriture des transitions. Le graphe $\mathcal{G}(\mathcal{A})_{n+1}$ est donné à la figure I.11.

Exemple I.5.9. L'automate à pile simple décrit dans l'exemple I.5.2, qui reconnait le langage L sur $\{0,1\}$ des représentations propres des entiers en base 2 qui contiennent autant d'occurrences de 0 que d'occurrences de 1 est donné par $\mathcal{A}_r = (E, A, P, \phi, (e, \varepsilon), F)$ où $E = \{e_0, e_1\}$, $A = \{0,1\}$, $P = \{0,1\}$ et les transitions de ϕ sont données par :

$$\begin{aligned} &e_0\varepsilon \overset{1}{\Longrightarrow} e_1 1, && \\ &e_1\varepsilon \overset{1}{\Longrightarrow} e_1 1, && e_1\varepsilon \overset{0}{\Longrightarrow} e_1 0, \\ &e_1 1 \overset{1}{\Longrightarrow} e_1 11, && e_1 1 \overset{0}{\Longrightarrow} e_1\varepsilon, \\ &e_1 0 \overset{1}{\Longrightarrow} e_1\varepsilon, && e_1 0 \overset{0}{\Longrightarrow} e_1 00, \end{aligned}$$

et la configuration interne finale est l'élément $e_1\varepsilon$. Le langage L est reconnu par l'AP simple $\mathcal{A}_r$ par pile vide et état finaux. Le graphe $\mathcal{G}(\mathcal{A})_n$ est donné à la figure I.12.

Exemple I.5.10. Voici un autre exemple d'automate à pile : $\mathcal{A} = (E, A, P, \phi, e_0X, K)$, défini par $A = \{a, b\}$, $E = \{e_0, e_1, e_2, e_3\}$, $P = \{X, Y\}$ et où ϕ est donnée par les transitions suivantes :

$$\begin{aligned}
&e_0X \overset{a}{\Longrightarrow} e_1, XY + e_2, XYY,\\
&e_1X \overset{a}{\Longrightarrow} e_1XY,\\
&e_2X \overset{a}{\Longrightarrow} e_2XYY + e_1XY,\\
&e_1X \overset{\varepsilon}{\Longrightarrow} e_3\varepsilon,\\
&e_2X \overset{\varepsilon}{\Longrightarrow} e_3\varepsilon,\\
&e_3X \overset{b}{\Longrightarrow} e_3\varepsilon.
\end{aligned}$$

Cet automate fonctionne ainsi : lorsqu'on est dans la configuration interne initiale e_0X, la lecture de a revient soit à rajouter un symbole Y sous le haut de pile X et se déplacer en e_1 soit à rajouter deux symboles Y sous le haut de pile X et se déplacer en e_2.

Lorsqu'on est dans la configuration e_1X,la lecture de a revient à glisser un symbole Y sous le X du haut de pile, sans changer d'état et la lecture du mot vide revient à supprimer la lettre du haut de la pile X et migrer vers e_3.

Lorsqu'on est dans la configuration e_2X,la lecture de a revient à glisser deux symboles Y sous le X du haut de pile, sans changer d'état ou à rajouter un symbole Y sous le haut de pile X et se déplacer en e_1 et la lecture du mot vide revient à supprimer la lettre de haut de pile X et migrer vers e_3.

Lorsqu'on est dans la configuration e_3X, la seule chose possible est de lire un b, qui supprime la lettre Y du haut de la pile et ne change pas l'état.

Tous ces mouvements possibles se lisent sur le graphe $\mathcal{G}(\mathcal{A})$ est donné à la figure I.13.

Selon la nature de F, l'automate $\mathcal{A}$ ne reconnait pas le même langage :
- si $F = E\varepsilon$, $L(\mathcal{A}) = \{a^nb^p,\ 0 < n < p < 2n\}$,
- si $F = e_3P^*$, $L(\mathcal{A}) = \{a^nb^p,\ 0 < n, 0 < p < 2n\}$,
- si $F = e_2\varepsilon$, $L(\mathcal{A}) = \emptyset$,
- si $F = EP^*p$, $L(\mathcal{A}) = \{a^nb^p,\ 0 < n, 0 \leq p < 2n\}$.

La reconnaissance des langages algébriques par les automates à piles permet d'obtenir les propriétés de stabilité suivantes :

Proposition I.5.11. *La famille des langages algébriques sur un alphabet A fixé est fermée par union, produit, opérateur $*$ et homomorphisme.*

Elle est aussi fermée par intersection avec un langage rationnel.

On perd, par rapport aux langages rationnels, la stabilité par passage au complémentaire ainsi que par intersection.

Par exemple, $L_1 = \{a^nb^nc^p,\ n, p > 0\}$ et $L_2 = \{a^nb^pc^p,\ n, p > 0\}$ sont des langages algébriques mais leur intersection $L_1 \cap L_2 = \{a^nb^nc^n,\ n, p > 0\}$ n'est pas algébrique.

La démonstration du fait que $\{a^nb^nc^n,\ n, p > 0\}$ n'est pas un langage algébrique se montre par l'absurde et repose sur le *lemme d'Ogden*, qui est l'équivalent du lemme de pompage qui a été énoncé pour les langages rationnels. Ce lemme assure, en substance, que si un mot engendré par une grammaire algébrique $G = (A, X, P, S)$ est assez long, alors, sa construction peut se décomposer en trois phases : on applique un cerain nombre de fois des règles de type $S \to uTv$, ensuite un cerain nombre de fois des règles de type $T \to wTm$ et enfin un cerain nombre de fois des règles de type $T \to s$. Ce lemme fait intervenir la notion d'arbres de dérivation d'une grammaire (arbres par lesquels ont visualise la formation des mots d'un langage), liée à la question d'ambigüité des grammaires algébriques. Nous renvoyons à [Aut87] pour une introduction à ces notions.

On peut associer à un AP $\mathcal{A} = (E, A, P, \phi, e_0X_0, F)$, un graphe $\mathcal{G}(\mathcal{A})$, appelé simplement *graphe de l'automate à pile*, dont les états sont indexés par E et les flèches sont indexées par les mouvements simples entre les configurations internes : une flèche du graphe $\mathcal{G}(\mathcal{A})$ allant de e vers e' est indexée par $a, X/X'V$ si $eX \overset{a}{\Longrightarrow} e'X'V$ appartient à $\mathcal{M}$. Voir les figures I.11 I.12 et I.13 pour des exemples.

La représentation de l'automate par son graphe est une représentation interne de l'action l'AP. Il en décrit les mécanismes et il est suffisant pour définir un AP, mais il ne montre pas l'AP fonctionner. Un certain nombre de questions, comme, par exemple, « quelles sont toutes les configurations internes

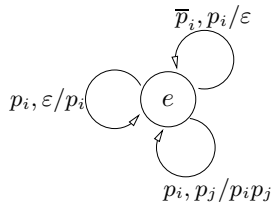


FIG. I.11 – Graphe de l'automate à pile de l'exemple I.5.8

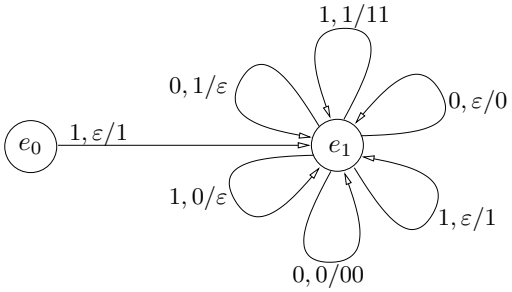


FIG. I.12 – Graphe de l'automate à pile de l'exemple I.5.2

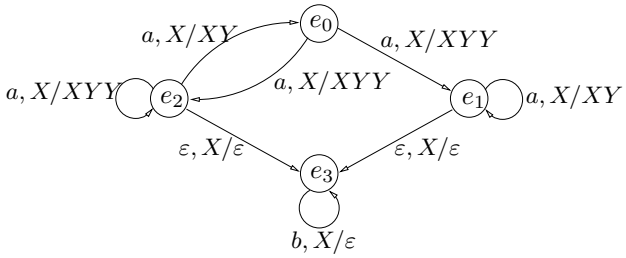


FIG. I.13 – Graphe de l'automate à pile de l'exemple I.5.10

accessibles de l'AP ? » ou « comment et en combien de temps peut-on passer de l'une à l'autre ? » n'ont pas de réponse évidente à la simple vue du graphe de l'AP.

On introduit donc un autre graphe, infini, associé à l'AP et appelé graphe des transitions de l'AP, donne une représentation externe de l'automate et permet de visualiser son mécanisme : il décrit l'ensemble des mouvements possibles entre configurations internes de l'automate.

Définition I.5.12. Soit $\mathcal{A} = (E, A, P, \phi, e_0 X_0, K)$ un AP, on appelle *graphe des transitions de* $\mathcal{A}$ le graphe infini $\mathcal{G}_T(\mathcal{A})$ dont l'ensemble des sommets est $E \times P^*$ et tel qu'il existe un arc indexé par $a \in A + \{\varepsilon\}$ de eW vers $e'W'$ si et seulement si $(e, XW) \overset{a}{\Longrightarrow} (e', W')$.

Exemple I.5.13. Soit $\mathcal{A} = (\{p, q, r\}, \{a, b\}, \{X\}, \phi, pX, F)$ un automate à pile dont les transitions sont données par

$$pX \overset{a}{\Longrightarrow} pXX \quad pX \overset{b}{\Longrightarrow} r\varepsilon + q\varepsilon$$
$$rX \overset{b}{\Longrightarrow} r\varepsilon \qquad qX \overset{b}{\Longrightarrow} q\varepsilon$$

Le graphe de l'AP est représenté à la figure I.14 et le graphe des transitions associé à $\mathcal{A}$ est le graphe représenté à la figure I.15.

Exemple I.5.14. Le graphe des transitions associé à l'AP qui reconnait le mots du langage de Dyck sur deux types de parenthèses, présenté à l'exemple I.5.8 a pour support l'arbre binaire. voir figure I.16 (l'unique état de l'automate est oublié dans l'écriture des états du graphe).

La géométrie des graphes de transition des AP a été caractérisée par D. Muller et P. Schupp [MS85]. Pour pouvoir énoncer cette caractérisation, nous avons besoin d'introduire des notions générales concer-

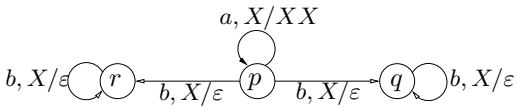


FIG. I.14 – Graphe de l'automate à pile de l'exemple I.5.13

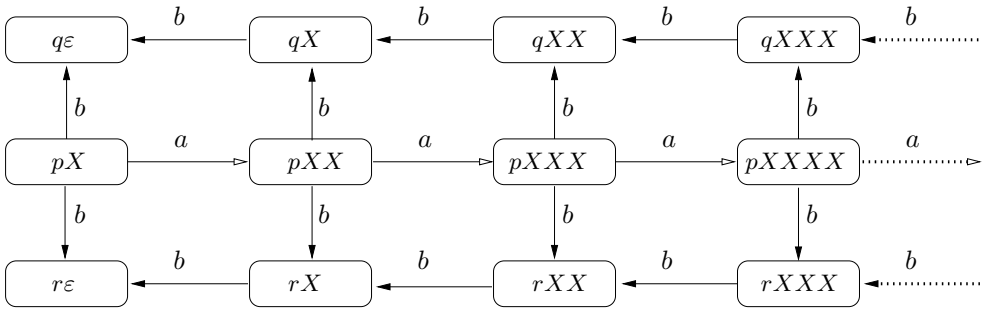


FIG. I.15 – Graphe des transitions de l'automate à pile de l'exemple I.5.13

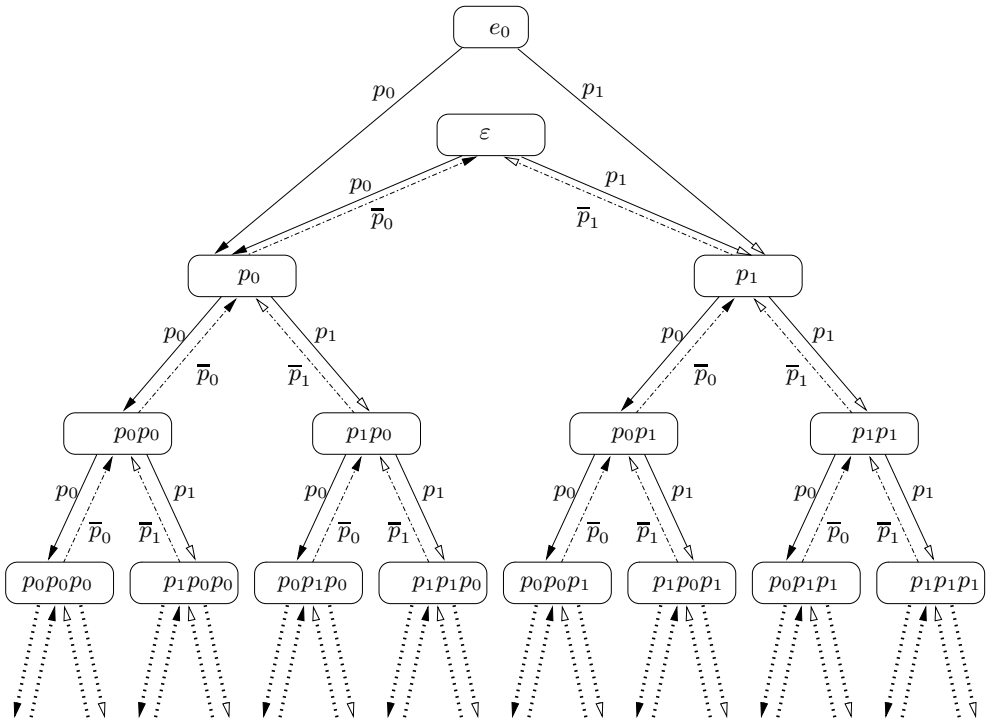


FIG. I.16 – Graphe des transitions de l'automate à pile associé au langage de Dyck à deux types de parenthèses.

nant les graphes dénombrables. Pour plus de détails relatifs aux graphes infinis en général, nous renvoyons à [Tho02], [Cau03] et [CK02] où sont présentées et ordonnées différentes classes de graphes infinis, en fonction de leur propriétés constructives et leurs propriétés logiques.

Définitions I.5.15. Soit $\mathcal{G} = (E, \phi, e_0)$ un graphe dont l'ensemble des sommets est E, l'ensemble des flèches est ϕ et e_0 est l'origine (ou racine) de $\mathcal{G}$.

Le degré de $\mathcal{G}$ est *borné* par un consante c (resp. *fini*) si, pour tout sommet du graphe, le nombre de flèches entrantes et sortantes est borné par c (*resp.* fini).

On définit la famille des sous-graphes de $\mathcal{G}$: $(\mathcal{G}_n)_{n\in\mathbb{N}}$, qui forme la *décomposition par distance à* e_0

de G. $\mathcal{G}_n$ est le graphe dont l'ensemble des sommets E_n est formé des éléments de E sont à distance au moins n de e_0 (par les flèches de ϕ et leurs inverses) et dont les flèches sont données par les éléments de ϕ qui relient deux sommets de E_n. Les éléments de E_n qui sont à distance n de e_0 dans $\mathcal{G}$ forment la *frontière de* $\mathcal{G}_n$.

Théorème I.5.16. *Soit $\mathcal{G} = (E, \phi, e_0)$ un graphe et $(\mathcal{G}_n)_{n\in\mathbb{N}}$ sa décomposition par distance à e_0.*

Le graphe $\mathcal{G}$ est le graphe des transitions d'un automate à pile si et seulement si la famille des composantes connexes des $\mathcal{G}_n$ (par les flèches de ϕ et leurs inverses), pour tout $n \in \mathbb{N}$ est finie, à isomorphismes près, en respectant les frontières. On dit que $\mathcal{G}$ admet une décomposition finie par distance.

Remarques I.5.17. Dans l'article [MS85], les composantes connexes des $\mathcal{G}_n$ sont appelés *ends*, c'est-à-dire, littéralement, les « bouts » du graphe.

Dans le cas du graphe des transitions $\mathcal{G}_T$, décrit sur la figure I.16, dont le support est l'arbre binaire complet, lorsqu'on gomme les sommets de la racine jusqu'à la n-ième ligne, on obtient 2^n graphes qui sont tous isomorphes à $\mathcal{G}_T$ lui-même.

La caractérisation de D. Muller et P. Schupp est liée à la manière dont on peut construire le graphe des transitions d'AP à l'aide de grammaires de graphes déterministes ou par répétitions de motifs. Voir [Cau92], dans lequel D. Caucal donne une preuve alternative à celle de [MS85].

I.6 Mots automatiques et extensions

I.6.1 Mots automatiques : lien entre langages formels et numération

Les automates qui vont nous intéresser dans ce paragraphe et dans la suite, sont les automates dont l'alphabet des entrées est $[\![0, q-1]\!]$, c'est-à-dire de la forme $\mathcal{A} = (E, [\![0, q-1]\!], \phi, e_0, A, \Pi)$. On parle alors de *$q$-automate fini déterministe* (q-AFD).

Lorsqu'on ce situe dans ce cadre, on peut créer une fonction f de $\mathbb{N}$ vers A qui associe à tout entier n la valeur de sortie de l'automate $\mathcal{A} = (E, [\![0, q-1]\!], \phi, e_0, A, \Pi)$ dans lequel on a rentré l'écriture propre en base q de n : on écrit l'entier n en base q, on fait rentrer la représentation de n en base q dans l'automate (en commençant par le chiffre le plus représentatif), l'automate finit par s'arrêter sur un état e, et on définit $f(n) = \Pi(e)$. Cette fonction f peut être représentée comme un mot infini m avec, pour tout $n \geq 0$, $m_n = f(n)$. Ce type de mots infinis a été introduit en 1972 par A. Cobham [Cob72] sous le nom de « uniform-tag sequences » et en 1974 par S. Eilenberg sous le nom de « q-recognizable sequences" »[Eil74].

Notation I.6.1. Un q-AFDS complet $\mathcal{A} = (E, [\![0, q-1]\!], \phi, e_0, A, \Pi)$ sera noté $\mathcal{A} = (E, \phi, e_0, A, \Pi)$, la donnée de l'ensemble qui agit sur l'automate n'étant plus nécessaire, puisqu'elle est implicitement donnée par l'appellation q-AFD.

Définition I.6.2. Soit $\mathcal{A} = (E, \phi, e_0, A, \Pi)$ un q-AFDS. On appelle *mot de sortie de l'automate* $\mathcal{A}$, le mot infini $f(\mathcal{A})$ défini par : pour tout $n \geq 0$, $\mu(\mathcal{A})_n = e_0 \cdot \overline{n}^q$, où $\overline{n}^q = n_l \dots n_1 n_0$ est la représentation en base q de $n = \sum_{i=0}^{l} n_i q^i$.

On dit qu'un mot infini m à valeurs dans un alphabet E est *q-automatique* s'il existe un q-AFDS $\mathcal{A} = (E, \phi, e_0, A, \Pi)$ tel que $e_0 \cdot 0 = e_0$ et pour tout entier n, $m_n = \Pi(\mu(\mathcal{A})_n)$.

Autrement dit, un mot q-automatique est une projection lettre-à-lettre d'un mot de sortie d'un q-AFD.

Un sous-ensemble d'entiers R est *q-automatique* si son mot infini caractérisque m, défini par :

$$\forall n \in \mathbb{N},\ m_n = \begin{cases} 1 & \text{si } n \in R, \\ 0 & \text{sinon.} \end{cases}$$

est un mot q-automatique.

Remarque I.6.3. La convention imposée $e_0 \cdot 0 = e_0$ permet en fait de pouvoir rentrer dans l'automate des écritures en base q qui ne sont pas propres, c'est-à-dire complétées à gauche par un nombre quelconque de zéros, sans changer l'état de sortie. Cela parait être une hypothèse forte mais on peut toujours s'y ramener en modifiant l'automate $\mathcal{A} = (E, \phi, e_0, A, \Pi)$ qui engendre notre mot : On rajoute à E un état e'_0, qui sera le nouvel état initial, on étend ϕ et Π à $E + \{e'_0\} \times [\![0, q-1]\!]$ en posant $e'_0 \cdot 0 = e'_0$, pour tout i de $[\![1, q-1]\!]$, $e'_0 \cdot i = e_0 \cdot i$ et $\Pi(e'_0) = \Pi(e_0)$. Le nouvel automate $\mathcal{A}'$ engendre le même mot q-automatique que $\mathcal{A}$ et vérifie notre hypothèse.

Comme conséquence de ces définitions, on a la proprosition suivante :

Proposition I.6.4. *Soit R un sous-ensemble de $\mathbb{N}$ q-automatique si et seulement si la famille $\overline{R}^q$ des représentations en base q des éléments de R forme un langage rationnel de $[\![0, q-1]\!]^*$.*

Exemples I.6.5. Le mot de Thue-Morse m, engendré par le 2-AFDS $\mathcal{A} = \{\{a, b\}, \phi, a, \{a, b\}, Id_{\{a,b\}}\}$ dont le graphe est présenté à la figure I.1, est 2-automatique.

$$m = abbabaabbaababbabaababbaabbabaabbaababbaabbabaababba\ldots$$

On remarquera au passage, que ce mot est identique au point fixe engendré par la substitution de Morse présentée à l'exemple I.2.8...

L'ensemble des nombres entiers $E_a = \{n \in \mathbb{N},\ a \cdot \overline{n}^2 = a\}$ est l'ensemble des entiers dont la somme des chiffres de l'écriture propre en base 2 est paire ; cet ensemble est 2-automatique. De même, l'ensemble des nombres entiers $E_b = \{n \in \mathbb{N},\ a \cdot \overline{n}^2 = b\}$ est l'ensemble des entiers dont la somme des chiffres de l'écriture propre en base 2 est impaire ; il est également 2-automatique.

Le mot de Rudin-Shapiro, engendré par le 2-AFDS $\mathcal{A}_{RS} = \{\{a, b, c, d\}, \phi, a, \{-1, 1\}, \Pi\}$ dont le graphe est présenté à la figure I.4 et où Π est définie par $\Pi^{-1}(\{1\}) = \{a, b\}$, est 2-automatique.

$$r = 111(-1)11(-1)1111(-1)(-1)(-1)1(-1)111(-1)11(-1)1(-1)(-1)(-1)111\ldots$$

La valeur de $\Pi(a \cdot \overline{n}^2)$ donne la parité du nombre de motifs 11 rencontrés lors de la lecture de $\overline{n}^2$ (avec chevauchements).

Le mot de Baum-Sweet β, défini sur l'alphabet $\{0, 1\}$ par $\beta_n = 1$ si et seulement si la représentation binaire propre de n ne contient aucun bloc de 0 de longueur impaire, est engendré par l'AFDS suivant : $\mathcal{A}_{BS} = \{\{a, b, c, d\}, \phi, a, \{0, 1\}, \Pi\}$, représenté à la figure I.17 où Π est définie par $\Pi^{-1}(\{1\}) = \{a, b\}$. Ce mot est donc 2-automatique.

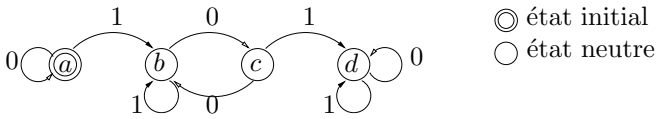


FIG. I.17 – AFD associé au mot de Baum-Sweet

Un mot périodique m à valeurs dans un alphabet A est q-automatique pour tout entier $q \geq 2$. En effet, il existe un entier $T \geq 1$ tel que, pour tout $n \geq 0$, $m_{n+T} = m_n$, le mot m est alors reconnu par le q-ADFS $\mathcal{A}_m = (E, \phi, e_0, A, \Pi)$, où $E = \{e_0, e_1, \ldots, e_{T-1}\}$, pour tout $(i, j) \in [\![0, T-1]\!] \times [\![0, q-1]\!]$, $e_i \dot{j} = e_{qi+j \mod T}$ et $\Pi : E \to A$ est définie, pour tout $i \in [\![0, T-1]\!]$ par $\Pi(e_i) = m_i$.

Il existe un lien très fort entre les points fixes de substitutions de longueur constante, découvert par A. Cobham [Cob72] :

Théorème I.6.6. *(Cobham 1972)*
Le mot infini m est q-automatique si et seulement s'il est l'image par un codage d'un point fixe de substitution de longueur constante supérieure ou égale à 2.

Ce théorème découle du fait qu'il existe une correspondance entre les q-AFD dont l'ensemble des états est E et les substitutions de longueur constante q sur l'alphabet E. Cette correspondance est décrite par Cobham dans [Cob72] et se généralise dans le cas où l'alphabet E est dénombrable [Mau06] (voir aussi le chapitre III).

En effet, si σ est une substitution de longueur constante q sur un alphabet E qui admet un point fixe $\overline{m}$ contenu dans $e_0 E^{\mathbb{N}}$, le graphe $\mathcal{G}_\sigma$ associé à σ est exactement le graphe associé au q-AFD $\mathcal{A}_\sigma = (E, e_0, \phi_\sigma, E, Id_E)$, où :

- l'ensemble des états est E,
- l'état initial est e_0,
- la fonction de transition ϕ_σ est donnée par : $\forall i \in \{0, 1\ldots, q-1\}, \forall e \in E,\ \phi_\sigma(e, i) = \sigma_i(e)$.

$\overline{m}$ est alors exactement la suite de sortie de l'automate $\mathcal{A}_\sigma$:

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

On retrouvera d'ailleurs ce résultat comme corollaire de la proposition III.5.1.

Ainsi, si le mot m est un mot q-automatique, engendré par le q-AFDS $\mathcal{A} = (E, \phi, e_0, A, \Pi)$, alors m est la projection par Π du point fixe de la substitution σ de longueur constante (ou morphisme uniforme) définie sur E par $\sigma(e) = (e \cdot 0)(e \cdot 1) \dots (e \cdot (q-1))$.

L'étude de ces mots peut donc être envisagée sous les deux éclairages différents : comme image par un morphisme d'un point fixe de substitution de longueur constante q ou comme mot de sortie d'un q-AFDS.

La reconnaissance des mots q-automatiques par les q-AFDS permet d'obtenir, à partir d'un mot q-automatique m, d'autres mots q-automatiques, par modification de l'automate qui reconnait m ; ainsi, on a les propriétés suivantes :

Proposition I.6.7. *Soit m et w deux mots q-automatiques à valeurs dans des alphabets finis A et A'. les mots décrits ci-dessous sont tous q-automatiques :*
- *les mots m' tels que $m'_n = m_n$ sauf pour un nombre fini d'entiers n,*
- *les mots $h(m)$, où h est un codage de E vers A''^*, pour un alphabet A'' fini quelconque,*
- *les mots v^f, où $f : A \times A' \to A''$ et $v_n = f(m_n, w_n)$.*

De plus, pour tout $a \in A$, le support de a, $N_a = \{n \in \mathbb{N},\ m_n = a\}$, est un ensemble q-automatique.

Lorsque q est un nombre premier, le lien entre mots infinis q-automatiques et automates peut aussi être interprété à l'aide de séries sur un corps fini, c'est ce qui a été montré dans [CKMFR80](voir également [Chr79]) :

Théorème I.6.8. *(G. Christol, T.Kamae, M. Mendès France et G. Rauzy 1980)*

Soit q un nombre premier et A un alphabet fini.

Un mot infini m de $\mathcal{A}^{\mathbb{N}}$ est q-automatique si et seulement s'il existe un corps fini $\mathcal{K}$ de caractéristique q et une injection ι de A dans $\mathcal{K}$ tels que la série $F_m(X) = \sum_{n\in\mathbb{N}} \iota(m_n)X^n$ est une série algébrique sur $\mathcal{K}(X)$, c'est-à-dire si et seulement s'il existe un entier k est $(k+1)$ polynômes $a_i(X)$ de $\mathcal{K}[X]$ tels que :

$$\sum_{i=0}^{k} a_i(X)(F_m(X))^i = 0.$$

On peut également, du point de vue de la théorie des nombres, citer le théorème de B. Adamczewski et Y. Bugeaud [AB06] qui permet de montrer qu'un nombre irrationnel dont le developpement en base q est un mot q-automatique est transcendant.

Un autre moyen de caractérisation des mots q-automatiques est la forme de leurs q-noyaux :

Définition I.6.9. q-noyau d'un mot infini q-automatique :

Le q-noyau $K_q(m)$ d'un mot infini m quelconque est l'ensembles des mot infinis extraits suivants :

$$K_q(m) = \{m^{(a,b)} \in A^{\mathbb{N}},\ m_n^{(a,b)} = m_{q^a n+b},\ a \geq 0, 0 \leq b < q^a\}.$$

Théorème I.6.10. *[Eil74]*

Le mot m est q-automatique si et seulement si son q-noyau $K_q(m)$ est fini.

Voir [AS03a] pour un démonstration. La démonstration utilise le fait qu'il y a équivalence entre être q-automatique est être le mot de sortie d'un AFDS dont les entrées $\overline{n}^q$ sont lues « à l'envers », c'est-à-dire à partir du chiffre le moins significatif vers le plus significatif de la représentation en base q de n.

Cette caractérisation des mots automatiques est un outil utile pour montrer élégamment certaines propriétés de stabilité de la famille des mots automatiques et de la famille des sous-ensembles automatiques :

Proposition I.6.11. *Soit m un mot q-automatique sur l'alphabet A. Pour tout couple d'entiers positifs a et b, le* mot extrait *$m^{(a,b)}$, défini par $m_n^{(a,b)} = m_{an+b}$ pour tout $n \geq 0$, est q-automatique.*

Proposition I.6.12. *La classe des ensembles q-automatiques est stable par passage au complémentaire, union, intersection, multiplication par un entier positif, par l'addition d'ensembles : $R+S = \{r+s,\ r \in R, s \in S\}$ et par l'opération $J_{a,b}$, définie par $J_{a,b}(R) = \{r,\ ar+b \in R\}$, pour tout couple $(a,b) \in \mathbb{N} \times \mathbb{Z}$.*

D'autre part, la propriété de q-automaticité étant liée aux propriétés arithmétiques des représentations propres en base q, cela lui confère une certaine rigidité ; ainsi, on a la propriété suivante :

Proposition I.6.13. *le mot m est q-automatique si et seulement si il est q^k-automatique pour un certain entier k.*

De même, un ensemble d'entiers est q-automatique si et seulement si il est q^k-automatique pour un certain entier k.

Cette dépendance à la base dans laquelle on prend les représentations propres des nombres entiers a été affinée par A. Cobham [Cob69] :

Théorème I.6.14. *(Cobham 1969) Soient deux nombres entiers q et p multiplicativement indépendants, c'est-à-dire pour lesquels il n'existe pas de couple d'entiers positifs (a,b) pour lequel $q^a = p^b$.*

Un ensemble d'entiers q- et p-automatique est une réunion finie de progressions arithmétiques, c'est-à-dire que son mot caractéristique est ultimement périodique.

De manière simpliste, ce théorème affirme qu'un mot qui serait automatique dans 2 bases entières indépendantes ne peut forcement pas être « très compliqué ».

Des généralisations de ce théorème existent, notamment pour d'autres systèmes de numération et pour certaines substitutions de longueur non constante, citons par exemple [Fav94], [Dur02], [Bès00] et [HS03], et [Bel06] pour les mots q-réguliers (voir le paragraphe I.6.2.1) ainsi que dans le cas multidimensionnel (théorème de Cobham-Semënov) [BHMV94a, BHMV94b].

La démonstration initiale est assez difficle [Cob69], on peut en trouver une présentation claire dans [AS03a] et [Per90] avec l'errata [RW06]. Le résultat charnière permettant la démonstration est le fait qu'un ensemble d'entier X q- et p-automatique est syndétique, c'est-à-dire que si on ordonne X de manière croissante $X = \{x_1 < x_2 < \ldots\}$, alors il existe une constante c qui vérifie $x_{n+1} - x_n < c$ (Nous renvoyons à [RW06] pour la démonstration de ce résultat).

I.6.2 Extensions de la notion de mot automatique

Comme nous venons de le voir, les mots automatiques sont des objets fascinants car on peut les voir sous de multiples facettes, et que, sous tous ces angles d'approche, il ont des propriétés sympathiques. La question de généralisation de ce concept vient alors naturellement :

Quelles autres familles de mots vérifient les mêmes type de propriétés de stabilité que la famille des mots automatiques ?

Dans quelles limites peut-on étendre un théorème vérifié par les mots automatiques ?

C'est dans cette optique que l'on peut envisager les différentes extensions du théorème de Cobham (énoncé au théorème I.6.14) dont nous avons parlé.

Nous présentons ici les notions de *mot substitutif* et de *mot q-régulier* qui englobent, chacunes à leur manière, la notion de mot automatique.

I.6.2.1 Mots q-réguliers

La notion de mot q-régulier, introduite par J.-P. Allouche et J. Shallit [AS92] [AS03b], est en fait basée sur le théorème d'Eilenberg (énnoncé au théorème I.6.10), qui caractérise les mots q-automatiques à l'aide de leur q-noyau. L'idée de leur travaux est de considérer des mots infinis à valeurs dans un anneau (donc à valeurs dans un alphabet infini), ayant un q-noyau finiment engendré.

On rappelle que le q-noyau $K_q(m)$ d'un mot infini m quelconque à valeurs dans un alphabet A est l'ensembles des mots extraits suivants :

$$K_q(m) = \{m^{(a,b)} \in A^*,\ m_n^{(a,b)} = m_{q^a n+b},\ a \geq 0, 0 \leq b \leq q^a\}.$$

Définition I.6.15. Soit R' un anneau noetherien et R un anneau contenant R'. On dit qu'un mot infini m à valeurs dans R est (R',q)-*régulier* si le R'-module engendré par son q-noyau est un R'-sous-module finiment engendré de $R^{\mathbb{N}}$.

Autrement dit, il existe un nombre fini de mots infinis $w^{(1)}$, $w^{(2)}$, ..., $w^{(r)}$ à valeurs dans R et telles que tous les mots de $K_q(m)$ sont des R'-combinaisons linéaires des $w^{(i)}$.

Remarques I.6.16. Lorsque l'anneau R' est fini, alors l'ensemble des mots (R', q)-réguliers contient les mots q-automatiques à valeurs dans tout anneau R contenant R'. Lorsqu'il n'y a pas de confusion possible, on parle simplement de mots q-réguliers.

Cette notion est introduite dans [AS92] sous la forme de *suites q-régulière*. C'est uniquement par souci d'homogénéité que nous l'avons transposée en terme de mots infinis, mais ces deux objets sont tout à fait identiques, seule diffère la présentation.

Exemple I.6.17. Le mot infini lg tel que $lg_0 = 1$ et pour tout entier $n > 0$, $lg_n = |\overline{n}^2|$ est $(\mathbb{N}, 2)$-régulier. En effet, on peut montrer que $lg_{2n+1} = lg_n + 1$, $lg_{4n} = 2lg_{2n} - lg_n$ et $lg_{4n+2} = lg_n + 2$. Ainsi, le 2-noyau de m est engendré par les suites $(lg_n)_{n\in\mathbb{N}}$, $(lg_{2n})_{n\in\mathbb{N}}$ et la suite constante égale à 1. Nous renvoyons à [AS92] pour d'autres exemples.

Théorème I.6.18. *(Allouche-Shallit 1992)*
les propositions suivantes sont équivalentes :

1. *le mot m est (R', q)-régulier,*
2. *le R'-module engendré par $K_q(m)$ est généré par un nombre fini de mots $w^{(i)}$ définis par $w_n^{(i)} = m_{q^{a_i}n+b_i}$ pour tout $n \geq 0$, pour un certain entier a_i et un entier $0 \leq b_i < q^{a_i}$,*
3. *il existe un entier r et r mots $w^{(1)}$, $w^{(2)}$, ..., $w^{(r)}$ à valeurs dans R tels que les q mots $m^{(q,b)}$ pour tout entier $b \in [\![0, q-1]\!]$ sont des R'-combinaisons linéaires des $w^{(i)}$ et de m.*
4. *il existe un entier r et r mots $w^{(1)}$, $w^{(2)}$, ..., $w^{(r)}$ à valeurs dans R et q matrices M_0, M_1, ..., M_{q-1} de $\mathcal{M}_{r+1}(R')$ telles que, si on pose*

$$V_n = \begin{pmatrix} m_n \\ w_n^{(1)} \\ \vdots \\ w_n^{(r)} \end{pmatrix}$$

alors, pour tout entier $a \in [\![0, q-1]\!]$, on a $V_{qn+a} = M_a V_n$.

La généralisation des mots q-automatiques par les mots q-réguliers est une « bonne » généralisation, dans le sens du théorème suivant :

Théorème I.6.19. *(Allouche-Shallit 1992)*
Un mot infini q-régulier qui prend uniquement un nombre fini de valeurs est q-automatique.

La famille des mots q réguliers possède aussi des propriétés de stabilité intéressantes :

Proposition I.6.20. *Si m et w sont deux mots q-réguliers, alors, les mots suivants sont q-réguliers :*

- *les mots obtenu par addition, multiplication ou multiplication par un élément de R' terme à terme des lettres de m et de w,*
- *pour tout couple de nombres entiers positifs (a, b) avec $a \neq 0$, le mot extrait $m^{(a,b)}$ défini par $m_n^{(a,b)} = m_{an+b}$ pour tout $n \geq 0$,*
- *pour tout nombre rationnel positif r, les mots extraits $m^{(r)}$ et $w^{(r)}$ définis par $m_n^{(r)} = m_{\lfloor nr \rfloor}$ et $w_n^{(r)} = m_{\lceil nr \rceil}$ pour tout $n \geq 0$,*
- *le mot $m \star w$, défini par $(m \star w)_n = \sum_{i+j=n} m_i w_j$,*

Il y a aussi, comme pour les mots automatiques, équivalence entre q-régularité et q^k-régularité.

Comme nous l'avons mentionné au paragraphe I.6, le théorème de Cobham admet une généralisation, donnée par J. Bell [Bel06], pour les mots q-réguliers, dans le cas commutatif :

Théorème I.6.21. *(Bell 2006)*
Soit R un anneau commutatif et soient q et p, deux entiers positifs multiplicativement indépendants.
Si m est un mot (R, q)- et (R, p)-régulier, à valeurs dans un R-module, alors il satisfait une récurrence linéaire sur R, c'est-à-dire, qu'il existe un entier k et k constantes c_1, c_2,..., c_k dans R, pour lesquelles on a :

$$\forall n \geq k,\ m_n = \sum_{i=1}^{k} c_i m_{n-i}.$$

I.6.2.2 Mots substitutifs

Le théorème de Cobham (énoncé au théorème I.6.6), qui lie les mots automatiques aux substitutions de longueur constante, amène naturellement à considérer l'extension de la notion de mot automatique, par les mots engendrés par des substitutions de longueur non constante :

Définition I.6.22. Un mot infini m est un *mot substitutif* s'il existe deux alphabets E et A, une substitution σ sur l'alphabet E admettant un point fixe $\overline{m}$ infini et un codage Π de E vers A^* tels que pour tout $n \geq 0, m_n = \Pi(\overline{m}_n)$; ce que l'on note $m = \Pi(\overline{m})$.

Dans la littérature, on retrouve aussi la notion de mot substitutif sous le terme de *mot morphique*.

Exemples I.6.23. Les mots automatiques sont tous des mots substitutifs.

Le mot de Fibonacci, présenté à l'exemple I.2.9 est un mot substitutif.

La substitution définie sur $\{a, b, c\}^*$ par $a \mapsto ab$, $b \mapsto ac$ et $c \mapsto a$, appellée substitution de Tribonacci, admet un unique point fixe $t = abacabaabacababacaba \ldots$. Le mot t est substitutif. Cette substitution est au centre de la relation de semi-conjugaison entre deux systèmes dynamiques : un échange d'intervalles et une translation sur le tore [Arn88].

Proposition I.6.24. *Soit m un mot infini sur l'alphabet A.*

Les propriétés suivantes sont équivalentes :

1. *m est un mot substitutif,*
2. *m est un mot substitutif, image par un morphisme d'un point fixe d'une substitution non effaçante,*
3. *il existe un mot automatique w sur l'alphabet $A+\{e_0\}$ ($e_0 \notin A$), non ultimement égal au mot infini constant $e_0^{\mathbb{N}}$, tel que m est obtenu à partir de w en effaçant toutes les lettres e_0 de w : $m_n = w_{N(n)}$ où $N(n)$ est défini par : $N(0) = \inf\{k \in \mathbb{N},\ w_k \neq e_0\}$ et $N(n+1) = \inf\{k > N(n),\ w_k \neq e_0\}$.*
4. *Il existe $Card(A)$ sous-ensembles automatiques de $\mathbb{N}$, notés S_a pour $a \in A$, tels que $T = \cup_{a \in A} S_a = \{t_1 < t_2 < \cdots < t_n < \ldots\}$ est un sous ensemble infini de $\mathbb{N}$ et pour tout $a \in A$, $m_n = a$ si et seulement si $t_n \in S_a$.*

La famille des mots substitutifs est aussi stable pour plusieurs opérations :

Proposition I.6.25. *Si m est un mot substitutif sur l'alphabet A, alors, les mots décrits ci-dessus sont tous substitutifs :*

- *le mot décalé $S(m)$ défini par $S(m)_n = m_{n+1}$,*
- *le mot wm, pour tout mot fini w de A^*,*
- *le mot M défini sur l'alphabet $A \times A$ par $M_n = (m_n, m_{n+1})$,*
- *s'il n'est pas fini, le mot obtenu en effaçant de m toutes les lettres d'un sous-alphabet $B \subset A$: $m(A \setminus B) = m_{N(n)}$ où $N(n)$ est défini par : $N(0) = \inf\{k \in \mathbb{N},\ w_k \notin B\}$ et $N(n+1) = \inf\{k > N(n),\ w_k \notin B\}$,*
- *s'il n'est pas fini, le mot $h(m)$, image de m par un morphisme h.*

L'atout principal des mots substitutifs, des substitutions et des systèmes dynamiques qui leurs sont associés vient du fait qu'ils ont naturellement des interactions avec de nombreux domaines : combinatoire, géométrie des pavages du plan, algèbre linéaire, théorie ergodique et théorie spectrale, informatique théorique, approximation diophantienne, transcendance... Voir [PF02] ou [Que87] pour une présentation (non exhaustive) de ces interactions.

Le chapitre III présente une autre forme de généralisation des mots infinis automatiques. Les mots automatiques étant les mots de sortie des automates finis, on peut poser la question suivante : à quoi ressemblent les mots de sortie des automates infinis ou, au moins, les automates dénombrables ? Ou, de manière équivalente, à quoi ressemblent les projections des points fixes de substitutions de longueur constante sur un alphabet dénombrable ?

Ces questions viennent, forcément, avec d'autres : ont-ils des propriétés similaires ? Dans quelles mesures peut-on transposer les résultats connus sur les mots automatiques ?...

Nous aborderons certains aspects combinatoires, en particulier la complexité, de ces mots ainsi que certaines propriétés des systèmes dynamiques associés aux points fixes de substitutions sur un alphabet dénombrable.

Chapitre II

Survol sur la complexité

S'il est une obsession en mathématiques, c'est bien de distinguer ce qui est compliqué de ce qui ne l'est pas, ou en tous cas, de ce qui l'est moins : du nombre de pages nécessaire à une démonstration à la longueur d'un programme informatique, des mesures de probabilité d'un évènement aux nombres de lignes pour écrire une formule... Avec ou sans fondement rationnels, ces quantités permettent de mesurer, de manière plus ou moins rigoureuse et utile, la complexité d'un objet.

Par exemple, on peut « quantifier » la simplicité d'une figure géométrique par plusieurs critères : nombre de cotés, régularité, nature des angles et des longueurs des cotés, connexité, dimension, genre....

De la même manière, on peut « quantifier » la simplicité du comportement d'un système dynamique ou d'un mot infini selon plusieurs critères : existence de mesures invariantes, ergodicité, entropie, propriétés de mélange...

En ce qui concerne les mots infinis, de nombreux outils issus des domaines où ils interviennent peuvent nous renseigner sur leur « complexité » : propriétés de récurrence, fréquences des lettres, mots évités, type de système dynamique associé, propriétes statistiques et coefficients de corrélation mais aussi la complexité de Kolmogorov, introduite dans [KU87] ou complexité algorithmique, qui mesure la taille du plus petit programme nécessaire à engendrer ce mot (nous renvoyons à [LV97] pour une introduction) et la complexité combinatoire, qui compte le nombre de mots de taille fixe dans le langage des facteurs d'un mot. C'est cette dernière complexité qui fait l'objet de ce chapitre.

Dans ce chapitre, A désigne un alphabet fini.

II.1 Mesurer le caractère aléatoire d'un mot infini...

Définition II.1.1. Pour un mot m de A^ω à valeurs dans un alphabet fini A, on appelle *fonction de complexité* du mot m, notée p_m, la fonction suivante :

$$\begin{array}{rccl} p_m : & \mathbb{N} & \to & \mathbb{N} \\ & n & \mapsto & Card\left(F_n(m)\right). \end{array}$$

Puisque le langage des facteurs d'un mot infini est un langage factoriel, on a $F_{n+n'}(m) \subset F_n(m)F_{n'}(m)$ pour tout couple d'entiers n et n', et donc la fonction de complexité est sous-multiplicative :

$$p_m(n+n') \leq p_m(n)p_m(n').$$

On a, de plus,

$$\forall m \in A^*, \forall n \in \mathbb{N}, 1 \leq p_m(n) \leq \left(Card(A)\right)^n, \tag{II.1}$$

La fonction extrémale $p_m(n) = 1$ (*resp.* $p_m(n) = (Card(A))^n$) est obtenue pour les mots constants (*resp.* pour presque tous les mots infinis aléatoires).

En particulier, la suite $\left(\frac{\ln p_m(n)}{n}\right)_{n\geq 1}$ converge et on peut montrer que l'entropie topologique du système dynamique (Ω_m, S) associé à m peut être exprimée à l'aide de la complexité par la formule suivante :

$$h_{top}(\Omega_m, S) = \lim_{n\to\infty} \frac{\ln p_m(n)}{n},$$

puisque la fermeture Ω_m de l'orbite de m sous l'action du décalage S admet une famille naturelle de recouvrements formés des cylindres $[w]$, pour tout facteur w de m de taille fixée (voir, par exemple [Kůr03]).

On trouve également dans la littérature une autre définition de l'entropie topologique d'un mot infini, proportionnelle à $h_{top}(\Omega_m, S)$ (voir par exemple [Que87]) définie par : $h(m) = \lim_{n\to\infty} \frac{\ln p_m(n)}{n \ln(Card(A))}$, de sorte qu'on a toujours $0 \leq h(m) \leq 1$, avec l'idée que plus le langage d'une suite est riche, plus elle est « compliquée » et son entropie topologique se rapproche de 1. On distingue deux grandes classes de mots : les *mots infinis de faible complexité*, qui sont les mots dont la fonction de complexité est sous-exponentielle (bornée, linéaire, polynomiale, etc...) pour lesquels l'entropie topologique est nulle et les *mots infinis de forte complexité*, dont la complexité croît de manière exponentielle, pour lesquels l'entropie topologique est strictement positive.

Parmi les mots de faible complexité, on compte les mots ultimement périodiques, qui sont caractérisés par une fonction de complexité bornée [HM38], les mots sturmiens, qui sont caractérisés par une fonction de complexité égale à $n+1$ mais aussi les mots automatiques et les mots substitutifs.

Remarque II.1.2. De la même manière que pour les langages des facteurs d'un mot infini, lorsque A est un alphabet fini, on peut définir la complexité d'un langage quelconque L par la fonction :

$$\begin{array}{rccl} p_L : & \mathbb{N} & \to \mathbb{N} \\ & n & \mapsto Card\left(L \cap A^n\right). \end{array}$$

La fonction de complexité d'un langage factoriel est croissante.

II.2 Facteurs spéciaux

Certains facteurs d'un mot infini m sont particulièrement précieux lorsqu'il s'agit de calculer, de trouver un équivalent ou de donner l'ordre de croissance de la fonction de complexité, notamment pour les mots infinis engendrés par une substitution ou un automate ; ce sont les facteurs spéciaux :

Définition II.2.1. Les *facteurs spéciaux à gauche* (resp. *spéciaux à droite, bispéciaux*) d'un mot m infini ou bi-infini à valeurs dans l'alphabet A sont les sous-mots de m qui apparaissent dans m prolongés à gauche (resp. à droite, à gauche et à droite) de plusieurs manières.

Pour un mot infini m, fixé et n un entier, on note $\mathcal{S}d_m(n)$ l'ensemble des facteurs spéciaux à droite de longueur n du mot m, de sorte qu'on a :

$$\mathcal{S}d_m(n) = \{w \in F_n(m),\ \exists a_1 \neq a_2 \in A, wa_1 \in L(m) \text{ et } wa_2 \in L(m)\}.$$

De la même manière, on note $\mathcal{S}g_m(n)$ l'ensemble des facteurs spéciaux à gauche de longueur n de m :

$$\mathcal{S}g_m(n) = \{w \in F_m(n),\ \exists a_1 \neq a_2 \in A, a_1w \in L(m) \text{ et } a_2w \in L(m)\}.$$

L'ensemble des facteurs bispéciaux de longueur n du mot m, noté $\mathcal{B}_m(n)$ est donné par :

$$\mathcal{B}_m(n) = \mathcal{S}g_m(n) \cap \mathcal{S}d_m(n) = \{w \in F_m(n),\ \exists a_1 \neq a_2 \in A, \exists a_3 \neq a_4 \in A, (a_1w, a_2w, wa_3, wa_4) \in L(m)^4\}.$$

Remarques II.2.2. L'utilisation des facteurs spéciaux à gauche peut s'avérer délicate lorsque l'on travaille avec des mots qui ne sont pas bi-infinis et non récurrents. En effet, tous les facteurs d'un mot bi-infini sont prolongeables à gauche mais ce n'est pas forçément le cas pour tous les facteurs d'un mot infini. Par exemple, le mot ab qui apparaît comme facteur du mot infini $m = ab^\omega$ n'admet pas de prolongement à gauche dans m. Ce genre de problèmes n'a pas lieu lorsque le mot est récurrent, puisque tous les facteurs, y compris les préfixes de m puisqu'ils réapparaissent plus loin dans l'écriture de m, admettent un prolongement à gauche.

Les résultats faisant intervenir les facteurs spéciaux à gauche seront donc donnés pour les mots bi-infinis et les mots infinis récurrents et pour calculer la fonction de complexité d'un mot infini non récurrent, on utilisera l'astuce suivante : On complète le mot m en un mot bi-infini $m' = \ldots xxxx\centerdot m$ où x est une lettre additionné à l'alphabet A et on calcule la complexité du mot m', vu comme élément de $(\{x\} + A)^{\mathbb{Z}}$. La complexité de m est alors donnée par la formule :

$$p_m(n) = p_{m'}(n) - n.$$

Par ailleurs, de même que pour la complexité, on peut définir, pour un langage factoriel prolongeable, les *mots spéciaux à gauche* (resp. *spéciaux à droite*, *bispéciaux*) , qui sont les mots du langage L qui se prolongent par au moins deux lettres distinctes à gauche (*resp.* à droite, à gauche et à droite) en des mots admissibles. Les mots bispéciaux jouent alors le même rôle fondamental dans le calcul de la complexité d'un langage que pour celui que nous avons décrit pour les mots infinis. Par conséquent, les résultats suivants sont valables aussi pour les langages factoriels prolongeables.

Graphiquement, le rôle que jouent les facteurs spéciaux peut se lire sur l'*arbre des facteurs à droite* (resp. *à gauche*) du langage d'un mot infini m sur l'alphabet A, est un arbre infini, dont les sommets sont indexés par $L(m) \cup \{\varepsilon\}$, de racine ε et de sorte qu'il y ait une arête entre v et w si et seulement si w s'écrit va (resp. av) pour une certaine lettre a de A. Comme un sommet w est indexé par le mot indiqué par l'écriture successive des arêtes qui l'on lit sur les arêtes du chemin qui va de la racine à w, on nomme conventionnelement les sommets de l'arbre par la lettre qui indexe la (seule) flèche entrante en ce sommet, et on représente le graphe avec la racine à gauche.

Par exemple, l'arbre des facteurs à droite du mot de Fibonacci f est représenté à la figure II.1.

$$f = abaababaabaababaababaabaababaabaababaababaabaababaababaabaababaab\dots$$

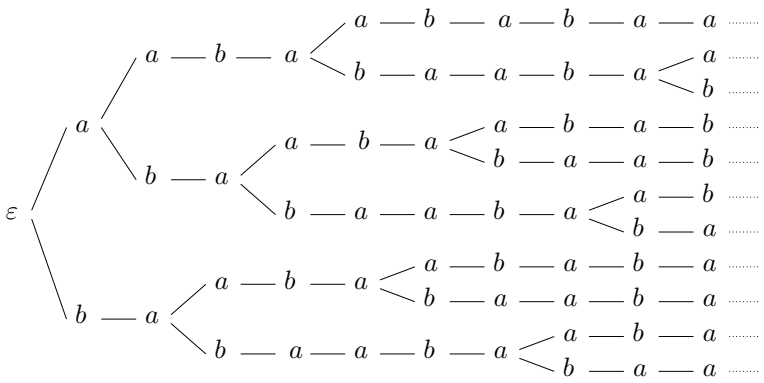


FIG. II.1 – Arbre à droite des facteurs du mot de Fibonacci.

On peut alors, à partir de l'arbre des facteurs à droite (*resp.* à gauche), voir ce que représentent la fonction de complexité p_m et les facteurs spéciaux de m : la complexité $p_m(n)$ est représentée par le nombre de sommets à distance n de la racine, et les facteurs spéciaux à droite (*resp.* à gauche) sont les points de branchement, c'est-à-dire les sommets dont au moins deux arêtes partent, ainsi, s'il y a exactement k branchements dont partent deux branches à distance n de ε, alors $p_m(n+1) - p_m(n) = k$. Dans le cas du mot de Fibonacci (et comme pour tous les *mots sturmiens*), il y a exactement un branchement à chaque distance, c'est-à-dire un unique facteur spécial et $p_m(n) = n + 1$.

En fait, dans le cas d'un alphabet de plus de deux lettres, tous les facteurs spéciaux ne jouent pas le même rôle, puisque le rôle d'un facteur spécial est en quelque sorte déterminé par le nombre de branches sortantes de son sommet correspondant dans l'arbre des facteurs à droite. Précisons cela :

Proposition II.2.3. *Pour tout facteur w d'un mot m, on appelle* ordre de multiplicité à droite (*resp.* à gauche), *noté $M_d(w)$ (*resp. *$M_g(w)$), la quantité suivante :*

$$M_d(w) = Card\,(\{e,\ we \in L(m)\}) - 1 \quad \Big(\ resp.\ M_g(w) = Card\,(\{e,\ ew \in L(m)\}) - 1\Big).$$

Lorsque m est infini et récurrent ou bi-infini, on a, pour tout entier $n \geq 1$, une formule qui lie les facteurs spéciaux à la complexité de m :

$$s(n) = p_m(n+1) - p_m(n) = \sum_{w \in \mathcal{S}d_m(n)} M_d(w) = \sum_{w \in \mathcal{S}g_m(n)} M_g(w).$$

■ **Démonstration de la proposition II.2.3** En effet, supposons que l'on connaisse la quantité $p_m(n)$ pour un mot infini m et un entier n fixé. Pour tout facteur w de $F_m(n)$, la quantité $M_d(u) + 1$ compte

le nombre de mots de $F_m(n+1)$ dont le préfixe de longueur n est w. Ainsi,

$$p_m(n+1) = \sum_{w \in F_m(n)} M_g(w) + 1,$$

et donc

$$p_m(n+1) = p_m(n) + \sum_{w \in F_m(n)} M_g(w).$$

Comme, pour tout élément de $F_m(n) \setminus \mathcal{S}d_m(n)$, la quantité $M_g(w)$ est nulle, on obtient la formule annoncée. ∎

Lorsque l'alphabet A est un alphabet binaire, tous les facteurs spéciaux à gauche (*resp.* à droite) sont d'ordre 2, de sorte que $s(n) = p_m(n+1) - p_m(n) = Card\,(\mathcal{S}g_m(n)) = Card\,(\mathcal{S}d_m(n))$.

La quantité $s(n) = p_m(n+1) - p_m(n)$ représente, quelque soit la taille de l'alphabet, le nombre de facteurs spéciaux (comptés avec ordre de multiplicité). La manière dont elle varie avec n est étroitement liée aux facteurs bispéciaux. D'ailleurs, on peut représenter graphiquement les facteurs d'un mot m de sorte à ce qu'apparaissent naturellement ses facteurs bispéciaux.

On construit une famille de graphes $(\mathcal{G}_n(m))_{n\in\mathbb{N}}$, appelés *graphes de Rauzy de* m. Les sommets de $\mathcal{G}_n(m)$, appelé *graphe de Rauzy de* m *à l'ordre* n, sont étiquetés par les facteurs de longueur n de m et les arcs sont indexées par les mots de longueur $n+1$ et de sorte qu'il y ait une arc de v vers w s'il existe un facteur u de m de longueur $n+1$ tel que $Pref_n(u) = v$ et $Suff_n(u) = w$.

La figure II.2 présente les premiers graphes de Rauzy du mots de Fibonacci.

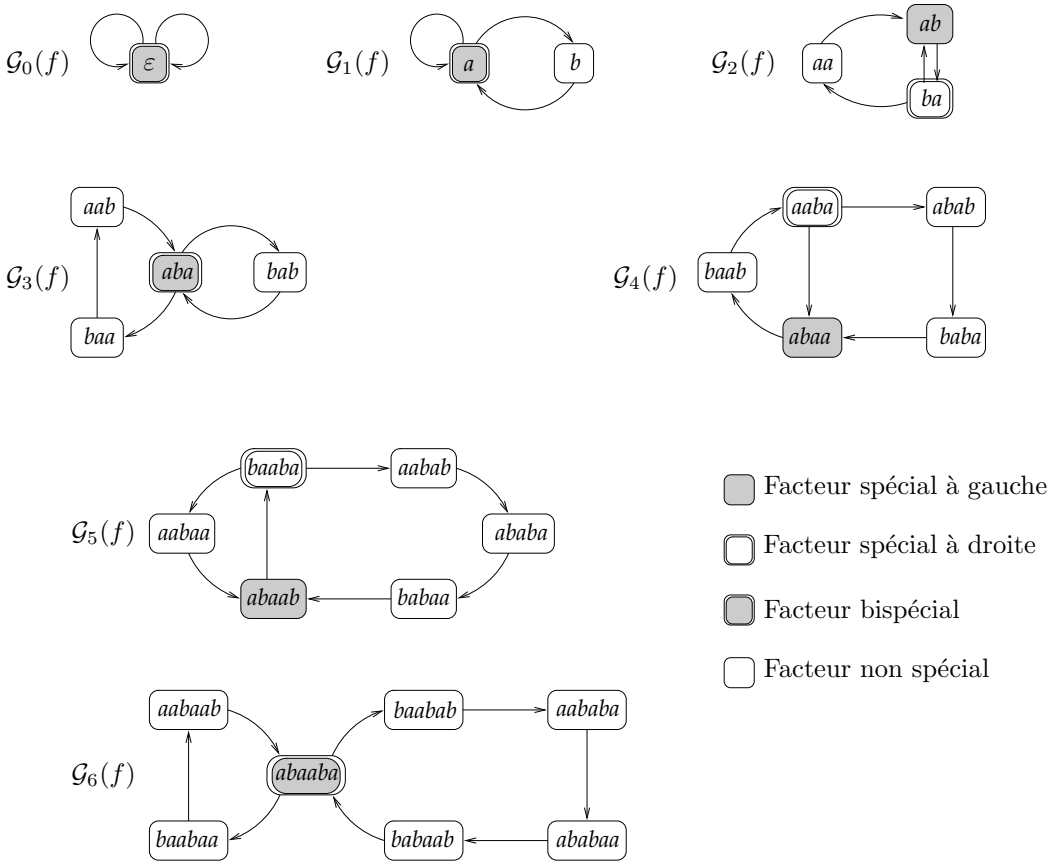


FIG. II.2 – Graphes de Rauzy d'ordre 0 à 6 du mot de Fibonacci.

Sur le graphe $\mathcal{G}_n(m)$, les facteurs de longueur n spéciaux à gauche sont les étiquettes des sommets sur lesquels arrivent au moins deux flèches et les facteurs de longueur n spéciaux à droite sont les étiquettes des sommets desquels partent au moins deux flèches. Lorsque les facteurs spéciaux sont peu nombreux,

on peut, dans certains cas, décrire complètement la structure des graphes de Rauzy (voir par exemple [AR91] et [Rot94]).
Tous les facteurs bispéciaux ne jouent pas le même rôle vis-à-vis de la fonction $n \mapsto s(n) = p_m(n+1) - p_m(n)$; on distingue trois classes de mots bispéciaux.

Définition II.2.4. Un w un facteur bispécial d'un mot m, selon le cardinal de $L(m) \cap AwA$, on qualifie le facteur w de :

bispécial strict si le nombre de prolongement à gauche et à droite de w en un mot de $L(m) \cap AwA$ est maximal, c'est-à-dire égal à $(M_d(w)+1)(M_g(w)+1)$,

bispécial faible si le nombre de prolongement à gauche et à droite de w en un mot de $L(m) \cap AwA$ est minimal, c'est-à-dire égal à $\max(M_d(w), M_g(w)) + 1$,

bispécial ordinaire si le nombre de prolongement à gauche et à droite de w en un mot de $L(m) \cap AwA$ est strictement compris entre $\max(M_d(w), M_g(w)) + 1$ et $(M_d(w)+1)(M_g(w)+1)$.

On peut représenter graphiquement les facteurs spéciaux dans un arbre, afin que les différentes classes de facteurs bispéciaux apparaissent naturellement.

Si un mot est spécial à droite, tous ses suffixes sont encore des facteurs spéciaux à droite mais on peut se poser la question de savoir si on peut le completer à gauche en un facteur de m qui soit aussi spécial à droite. De même, si un facteur est spécial à gauche, alors, tous ses préfixes sont des facteurs spéciaux à gauche mais peut-on le prolonger à droite en un facteur de m qui soit aussi spécial à gauche ?

Cela se lit sur l'*arbre des facteurs spéciaux à gauche* (resp. *à droite*), dont les sommets sont indexés par l'ensemble des facteurs spéciaux à gauche (*resp.* à droite), de racine ε et de sorte qu'il y ait un arc de v vers w si $Pref_{|v|-1}(v) = w$ (*resp.* $Suff_{|v|-1}(v) = w$).

Les facteurs spéciaux à gauche (*resp.* à droite) peuvent ainsi être rangés en familles : deux facteurs spéciaux sont dans la même famille s'ils appartiennent à la même branche de l'arbre des facteurs spéciaux à gauche (*resp.* à droite).

L'arbre des facteurs spéciaux à droite du mot m de Thue-Morse (voir I.2.8) est représenté à la figure II.3.

$$m = \sigma(m) = abbabaabbaababbabaababbaabbabaabbaababbaabbabaababbabaabbaababba\ldots$$

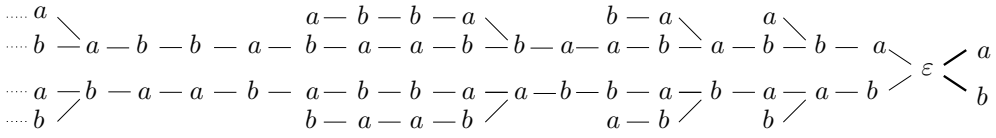


FIG. II.3 – Arbre des facteurs spéciaux à droite du mot de Thue-Morse.

Les feuilles de l'arbre sont en fait les bispéciaux faibles et les bispéciaux ordinaires et stricts sont les points de branchements de l'arbre.

Par exemple, le mot ab, est un bispécial strict : les quatres mots que l'ont peut former $aaba$, $aabb$, $baba$ et $babb$ en le prolongeant des deux cotés sont des facteurs du mot de Thue-Morse. Par contre, aba est un bispécial faible car ses seuls prolongements en facteurs du mot de Thue-Morse possibles sont $aabab$ et $babaa$. Le mot $abba$ est quant à lui bispécial ordinaire car ses seuls prolongements possibles sont $aabbab$, $babbaa$ et $babbab$.

Dans certains cas, la structure de l'arbre des facteurs spéciaux à gauche, combiné avec celle des graphes de Rauzy, permettent aussi, de donner des renseignements sur le système dynamique symbolique associé au mot infini, notamment sur le nombre de mesures ergodiques [Mon05].

Proposition II.2.5. *On définit l'*ordre de multiplicité bilatère *d'un mot w de m par :*

$$M(w) = Card\,(L(m) \cap AwA) - (M_d(w) + M_g(w) + 1)\,.$$

Lorsque m est un mot bi-infini ou infini et récurrent, alors, pour tout entier $n \geq 1$:

$$s(n+1) - s(n) = \sum_{w \in B_m(n)} M(u).$$

Les formules qui lient la complexité aux nombres de facteurs spéciaux et bispéciaux sont très utiles lorsqu'on cherche la complexité d'un mot infini, et encore plus lorsqu'on considère des mots infinis engendrés par des algorithmes simpes et/ou de faible complexité, pour lesquels les facteurs bispéciaux sont peu nombreux ([Cas96]), mais, conservation de la difficulté oblige, exhiber les facteurs spéciaux d'un mot fixé n'est pas forçément aisé. Il existe toutefois des méthodes pour les mots engendrés par des substitutions, basées sur l'étude des mots bispéciaux et sur des principes de synchronisation (qui permettent, à partir d'un facteur arbitraire d'un mot infini, de savoir où exactement le situer dans le mot infini et de savoir comment il a été fabriqué). Nous reviendrons sur cela dans le paragraphe II.4.

II.3 Quelques résultats généraux

Plusieurs angles d'approche pour l'étude de la fonction de complexité des mots infinis ont été abordés (voir par exemple [All94] ou [Fer99] pour un survol), notamment dans l'optique de répondre aux questions suivantes :

- Peut-on calculer exactement la complexité d'un mot infini donné ?
- S'il est difficile de calculer exactement la complexité, est-il au moins possible d'en donner une fonction équivalente ou un ordre de croîssance ?
- Pour une famille de mots infinis donnée, quels types de complexité sont possibles ?
- Etant donnée une fonction f de $\mathbb{N}$ dans $\mathbb{N}$ croissante et sous-multiplicative, peut-on trouver un mot infini de complexité f ?
- Quelles particularités ont en commun tous les mots infinis ayant la même fonction de complexité ?

Avant de répondre partiellement à ces réponses, nous donnons quelques résultats généraux sur la complexité.

Proposition II.3.1. *Soit m un mot infini sur un alphabet A de complexité p_m.*

1. *S'il existe un entier n tel que $p_m(n) \leq n$ ou $p_m(n+1) = p_m(n)$, alors p_m est bornée et donc m est ultimement périodique.*
2. *S'il existe un entier n tel que $p_m(n) < (Card(A))^n$, alors il existe un nombre réel $c < Card(A)$ tel que $p_m(n) < c^n$ pour tout n.*
3. *S'il existe une constante $\alpha > 0$ telle que, pour tout entier n, $p_m(n) \leq \alpha n$, alors, il existe une constante $C > 0$ telle que $s_m(n) = p_m(n+1) - p_m(n) \leq C\alpha^3$.*
4. *Si $s_m(n) = p_m(n+1) - p_m(n)$ est une suite bornée, alors l'ensemble des entiers n tels que $s_m(n) < s_m(n+1)$ est de densité nulle dans $\mathbb{N}$. De plus, si la suite $(s_m(n))_{n\in\mathbb{N}}$ est ultimement périodique, alors elle est ultimement constante.*

Ces résultats sont principalement issus de [Ale96], [Cas96] et [Cas97].

D'une manière générale, la détermination d'une formule exacte pour la fonction de complexité d'un mot infini est un problème délicat qui a fait l'objet de nombreux travaux. L'outil le plus utile pour de tels calculs reste l'utilisation des facteurs spéciaux.

Par exemple, dans le cas du mot m de Thue-Morse (engendré par la substitution $\sigma(a) = ab$ et $\sigma(b) = ba$), on a la formule suivante exprimant la complexité ([Brl89]) :

$p_m(1) = 2$, $p_m(2) = 4$ et pour tout $n \geq 3$, si $n = 2^r + q + 1$ avec $r \geq 0$ et $q \in [1, 2^r]$ alors :

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

La preuve de ce résultat repose sur le fait que tous les facteurs bispéciaux à droite de m sont en fait les suivants :

- les bispéciaux stricts sont ε, $\sigma^k(ab)$ ou $\sigma^k(ba)$ pour tout $k \geq 0$,
- les bispéciaux ordinaires sont a et b,
- les bispéciaux simples sont $\sigma^k(aba)$ ou $\sigma^k(bab)$ pour tout $k \geq 0$.

Pour d'autre exemples, pour renvoyons à [All92] pour la complexité des mots infini codant des pliages de papiers, égale à $4n$, ainsi que pour la complexité du mot de Rudin-Shapiro, égale à $8n - 8$, à [Fer95] pour la complexité des poins fixes de la substitution de Chacon, égale à $2n + 1$.

Un autre moyen de calculer la complexité d'une suite donnée est de passer par les séries génératrices de la complexité, et celles des facteurs spéciaux et des facteurs bispéciaux, en exploitant les relations qui les lient ([Cas97]). Si les calculs n'aboutissent pas à une expression explicite de la complexité, ils peuvent néanmoins permettre d'obtenir, dans certains cas, au moins un équivalent ou l'ordre de grandeur de la complexité.

Pour certaines familles de mots infinis, les ordres de grandeurs possibles de la complexité sont entièrement déterminés. C'est le cas des mots substitutifs et des mots automatiques. Nous y reviendrons au paragraphe II.4 : s'il n'est pas ultimement périodique, un mot substitutif a pour complexité une fonction de l'ordre de n, $n \log\log n$, $n \log n$ ou n^2.

D'une manière générale, les mots de basse complexité, en particulier ceux dont la fonction de complexité est linéaire ou sous linaire, ont été largement étudiés, aussi bien du point de vue de leurs propriétés combinatoires arithmétiques que dynamiques (voir par exemple [HM40], [AR91], [All92],[FM97],[AB06] [Abe01], [AB06]).

Parmi ces mots, les mots infinis non ultimement périodiques et de complexité minimale, c'est-à-dire tels que $\forall n \in N,\ p_m(n) = n + 1$, sont appelées *mots sturmiens* (et sont forçément des mots infinis binaires). Ces mots ont fait l'objet de nombreux travaux (pour un survol, voir par exemple le chapitre 2 de [Lot02], le chapitre 6 de [PF02] ou [Ber01]). Il trouvent une représentation dynamique dans le codage des lignes brisées de meilleure approximation des droites irrationnelles du plan ou encore dans le codage de rotations irrationnelles du cercle [HM40].

La question de savoir quelles fonctions peut-on réaliser en tant que fonction de complexité ou en tant qu'ordre de grandeur de la complexité de mots infinis vient alors naturellement.

Dans cette optique, J. Cassaigne [Cas97] a caractérisé toutes les fonctions linéaires qui sont exactement les fonction de complexité d'un mot.

Théorème II.3.2. *Soit $(\alpha, \beta) \in \mathbb{N} \times \mathbb{Z}$. Il existe un mot infini dont la fonction de complexité vaut $\alpha n + \beta$ pour tout $n \geq 1$ si et seulement si $\alpha + \beta > 1$ et $2\alpha + \beta \leq (\alpha + \beta)^2$.*

En ce qui concerne les ordres de grandeurs possibles, des réponses partielles ont été données par J. Goyon [Goy97] et par J. Cassaigne [Cas96].

Théorème II.3.3. *[Goy97] Pour toute fonction f croissante de $[1, +\infty[$ dans $[1, +\infty[$ vérifiant :*

$$\exists k > 1, \exists H > 0,\ \forall i \geq 0, f(k^{i+1}) \leq H f(k^i), \tag{II.2}$$

il existe un mot infini $m^{(f)}$ dont la fonction de complexité est du même ordre que $nf(n)$.

En particulier, quelque soit le k-uplet de réels $(\alpha_1, \alpha_2, \ldots, \alpha_k)$, avec $\alpha_1 > 1$, il existe un mot infini dont la fonction de complexité est du même ordre que $n^{\alpha_1} (\log n)^{\alpha_2} (\log\log n)^{\alpha_3} \ldots (\log\log \ldots \log n)^{\alpha_k}$.

L'idée est en fait de montrer que toutes les fonctions qui vérifient (II.2) peuvent être réalisée comme fonctions de complexité d'un mot de Toeplitz (c'est-à-dire un mot m qui vérifie $\forall n \in \mathbb{N}, \exists p \geq 1,\ \forall k \in \mathbb{N}, m_{n+kp} = m_n$) : on introduit une famille de mots de Toeplitz, paramétrée par des suites d'entiers strictement plus grands que 2 et, pour une fonction f vérifiant (II.2), on construit un mot de la famille (construction de la suite d'entiers qui le paramètre) dont la complexité est de l'ordre de f. Ses travaux s'inscrivent dans le prolongement de ceux de J. Cassaigne et J. Karhumaki [CK97].

De manière plus fine, J. Cassaigne, dans [Cas96], a donné une condition suffisante pour qu'une fonction puisse être asymptotiquement équivalente à la fonction de complexité d'un mot infini binaire.

Théorème II.3.4. *Si ϕ est une fonction de $\mathbb{R}$ dans $\mathbb{R}$ deux fois dérivable telle que :*

1. *il existe $x_0 \in \mathbb{R}$ tel que, pour tout $x \geq x_0$, $0 \leq \phi''(x) \geq 1$,*
2. *$x \log_2 x = o(\phi(x))$ ou $\lim_{x \to \infty} \phi'(x) = +\infty$ alors il existe un mot infini binaire m dont la complexité vérifie*

$$p_m(n) \sim \phi(n).$$

La démonstration de théorème est constructive et donne des exemples de mots de complexité f, pour toute fonction f comprise entre $n \log_2 n$ et n^2 et à croissance suffisamment régulière.

On peut également construire des mots infinis m de complexité intermédiaire [Cas02] pour lesquels l'ordre de croissance de $\log_2 p_m(n)$ est fixé, c'est-à-dire des mots d'entropie nulle, mais pour lesquels la suite $\left(\frac{\log_2 p_m(n)}{\log_2 n}\right)_{n\in\mathbb{N}}$ n'est pas bornée (la complexité n'est, dans ce cas, ni polynomiale ni exponentielle, par exemple $p_m(n) = 2^{\sqrt{n}}$).

Théorème II.3.5. *Si ϕ est une fonction de $\mathbb{R}$ dans $\mathbb{R}$ différentiable, sauf peut-être en 0, telle que :*

1. $\log_2 x = o(\phi(x))$,
2. ϕ' *est décroissante et* $\phi'(x)) = o(x^{-a})$ *pour un certain réel positif* a,

alors il existe un mot infini binaire m récurrent dont la complexité vérifie :

$$\log_2 p_m(n) \sim \phi(n).$$

L'idée de ces deux derniers résultats repose sur la construction d'un mot infini obtenu par morphisme à partir du point fixe d'une substitution sur un alphabet dénombrable. Nous aurons l'occasion de revenir sur ce type de mots dans la chapitre III.

Remarque II.3.6. Toutes les fonctions de complexité ne sont pas forçément aussi régulières que celles dont nous avons parlé jusqu'à présent... En effet, il existe un mot m infini de Toeplitz dont la complexité vérifie [Goy97] :

$$\forall \alpha > 1,\ \limsup \frac{p_m(n)}{n^\alpha} = +\infty \text{ et } \liminf \frac{p_m(n)}{n} \leq 6,$$

ou encore, il existe un mot m infini de Toeplitz d'entropie nulle et tel que :

$$\forall \alpha > 1,\ \limsup \frac{p_m(n)}{n^\alpha} = +\infty.$$

Dans une autre optique, dans [MM06], C. Mauduit et C. G. Moreira donnent un encadrement du nombre de facteurs apparaissant dans la famille des mots infini dont la complexité est bornée une fonction, dans le but d'estimer les dimensions de Hausdorff généralisées de l'ensemble des réels de $[0, 1]$ dont le développement en base entière est de complexité bornée par une fonction donnée.

II.4 Complexité des mots subsitutifs et automatiques

La famille des mots automatiques et substitutifs est une large famille de mots de faible complexité. L'étude des fonctions de complexité de ces mots a été motivée par l'implication de cette famille de mots dans de nombreux domaines.

L'ordre de grandeur de la fonction de complexité des mots automatiques a été déterminé par A. Cobham [Cob72].

Théorème II.4.1. *(Cobham 1972)*
Si m est un mot q-automatique, engendré par un q-automate fini à p états, alors :

$$\forall n > 0,\ p_m(n) \leq qp^2 n.$$

Ainsi, si, m n'est pas ultimement périodique, alors p_m est de l'ordre de n.

■ Démonstration du théorème II.4.1
Soient σ la substitution associée à l'automate $\mathcal{A} = (E, [\![0, q-1]\!], \phi, A, \Pi)$ un q-automate à p états qui engendre m et $\overline{m}$ le point fixe de σ tel que $m = \Pi(\overline{m})$.

Soient un entier n fixé et k un entier tel que $q^k \leq n < q^{k+1}$. Un mot w de longueur q^k provient de la projection par Π d'un certain mot $\overline{w}$ de $\overline{m}$ de longueur q^k. Ce relevé $\overline{w}$ est sous-mot d'un facteur de m du type $\sigma^k(e_1)\sigma^k(e_2)$ avec $(e_1, e_2) \in E^2$. Or il existe p^2 couples (e_1, e_2) dans E^2 et chacun des mots $(e_1, e_2) \in E^2$ a, au plus, q^k sous-facteurs de longueur q^k différents, qui donnent au maximum q^k mots projetés par Π, et donc q^k mots de m différents. Ainsi, $p_m(q^k) \leq q^k p^2$. La fonction de complexité d'un mot infini étant croissante, on a $p_m(n) \leq q^{k+1}p^2$ et ainsi, $p_m(n) \leq qp^2 n$. ■

Le premier résultat sur la complexité des suites substitutives, dû à A. Ehrenfeucht, K. P. Lee et G. Rozenberg [ELR75] indique que la complexité d'un mot m engendré par une substitution sur un alphabet fini vérifie $p_m(n) = O(n^2)$. Les ordres de croissance possibles de leurs fonctions de complexité ont ensuite été entièrement déterminés par J.-J. Pansiot [Pan85].

Théorème II.4.2. *(Pansiot 1984)*

Pour tout mot infini m engendré par une substitution σ sur un alphabet fini A, il existe une fonction $f_\sigma(n)$ égale à 1, n, $n\log n$, $n\log\log n$ ou n^2 telle que :

$$\exists (c,C) \in \left(\mathbb{R}_+^*\right)^2,\ cf_\sigma(n) \leq p_m(n) \leq Cf_\sigma(n).$$

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

La fonction f_σ est appelée ordre de croissance de p_m.

Ce théorème souligne en fait les cinq classes de substitutions sur un alphabet A, caractérisées par le type des ordres de croissance des itérés $\sigma^n(a)$, pour a dans A. On peut donc préciser le théorème ci-dessus.

Théorème II.4.3. *Si m est un point fixe non ultimement périodique d'une substitution σ sur l'alphabet A et que toutes les lettres de A apparaissent dans m, alors*

- *si σ est croissante, alors $f_\sigma(n) = n$, $n\log\log n$, $n\log n$ selon que σ est quasi-uniforme, polynomialement divergente ou exponentiellement divergente,*
- *si σ est n'est pas croissante, alors $f_\sigma(n) = n^2$ ou s'il existe une substitution τ sur un alphabet A' qui admet un point fixe m' et un morphisme non effaçant $h : A'^* \to A^*$ tels que $m = h(m')$, alors $f_\sigma = f_\tau$.*

Remarque II.4.4. Il est intéressant de souligner que l'hypothèse faite sur l'apparition de toutes les lettres de A dans m est essentielle. Si la substitution σ admet deux points fixes m et w et s'il existe un sous-alphabet propre A' de A tel que $\sigma(A') \subset A'^*$ et $m \in A'^*$, il se peut que m et w n'aient pas le même ordre de complexité, puisque l'ordre de m dépend de la manière dont croissent les itérés $\sigma^n(a)$ pour $a \in A'$ et l'ordre de w dépend de la manière dont croissent les itérés $\sigma^n(a)$ pour $a \in A$.

La démonstration du théorème de Pansiot utilise les facteurs spéciaux (qu'il appelle facteurs biprolongeables) pour minorer la fonction de complexité. Plus exactement, il utilise la formule $p_m(n+1) - p_m(n) \geq Card(\mathcal{S}_m d(n))$, qu'il minore au cas par cas. En ce qui concerne les majorations, l'idée est de « désubstituer »pour comprendre d'où vient un mot, c'est-à-dire quelles lettres et combien d'itérations sont nécessaires pour le fabriquer.

Dans le cas des morphismes croissants, les majorations sont obtenues en utilisant le même principe de désubsitution utilisé par Cobham : pour un mot w de m de longueur n fixé , on construit une suite de mots w_i de longueur dé croissante, telle que w est contenu dans $\sigma^i(w_i)$. Pour tous les mots de longueur n, la suite $(|w_i|)_{i\geq 1}$ est décroissante et $|w_i| \geq 3$ à partir d'un certain rang k_w ; on trouve ensuite une majoration du nombre de k_w possibles, pour tous les mots de longueur n, selon la nature de la croissance des itérés.

Dans le cas des morphismes non-croissants, la disjonction des cas dépend des longueurs possibles des facteurs contenant uniquement des lettres d'ordre de croissance borné. En effet, l'application d'un morphisme non effacant à un point fixe de substitution ne change pas les ordres de croissance des itérés, donc s'il existe une substitution τ sur un alphabet A' qui admet un point fixe m' et un morphisme non effaçant $h : A'^* \to A^*$ tels que $m = h(m')$, alors $f_\sigma = f_\tau$. D'autre part, lorsque qu'il existe pas de telle substitution τ, on montre que m contient des facteurs arbitrairement longs dont les lettres sont toutes à croissance bornée (on parle alors de *substitution poussive*). La majoration est donnée par le théorème 2 de [ELR75] et la minoration du nombre de facteurs spéciaux utilise le fait que le mot m contient forcement une infinité d'occurences d'une lettre dont l'ordre de croisssance n'est pas borné. Voici un tableau regroupant des exemples de substitutions, leurs types ainsi que l'ordre de grandeur de la complexité de leur point fixe contenu dans $a\{a,b\}^{\mathbb{N}}$:

Substitution	$a \mapsto aba$ $b \mapsto bab$	$a \mapsto ab$ $b \mapsto ba$	$a \mapsto aaab$ $b \mapsto bb$	$a \mapsto aba$ $b \mapsto bb$	$a \mapsto aab$ $b \mapsto b$
type	Uniforme	Uniforme	Exp. divergente	Pol. divergente	Poussive
ordre de a	3^n	2^n	3^n	$n2^n$	2^n
ordre de b	3^n	2^n	2^n	2^n	1
ordre de croissance de la complexité	1	n	$n\log n$	$n\log\log n$	n^2

Les majorations de $p_m(n)$ effectuées par J.-J. Pansiot utilisent le même principe de base utilisé par Cobham dans le cas des mots q-automatiques : on majore le nombre de manières possible de creer des mots m de longueur « un peu plus grande« que n du type $\sigma^k(e_1)\sigma^k(e_2)\dots\sigma^k(e_p)$ (pour k le plus grand possible en fonction des e_i) qui sont nécessaires pour fabriquer un mot de longueur au moins n.

Dans le prolongement de cette idée, on peut se demander, pour un facteur w d'un mot infini substitutif m, si, du premier coup d'oeil, on sait exactement combien de lettres, lesquelles et combien d'itérations sont nécessaires pour le fabriquer. C'est ce qu'on appelle les *principes de synchronisation* d'un mot substitutif.

Définition II.4.5. Soit m un point fixe d'une substitution σ sur l'alphabet A. On appelle *k-découpage* d'un facteur w de m, toute écriture de w de la forme $w = s\sigma^k(u_1)\sigma^k(u_2)\dots\sigma^k(u_n)p$, où n est un entier, $u = u_0u_1u_2\dots u_n$ est un facteur de m, s est un suffixe strict de u_0 et p est un préfixe strict de u_{n+1}.

On dit alors que *w provient du mot u.*

On dit qu'un facteur w de m est *déterminé à l'ordre k* s'il apparait dans m avec toujours le même k-découpage. On dit que la substitution σ est *déterminée à l'ordre k* si pour tout facteur w d'un de ses points fixes, de longueur supérieure à un entier C_k, est déterminé à l'ordre k.

Lorsqu'une substitution est déterminée à l'ordre k, on peut « désubstituer »tous les facteurs de m suffisamment longs sauf éventuellement sur leurs préfixes et leurs suffixes de longueur inférieure à $\max_{a\in A}|\sigma^k(a)| - 1$.

Exemples II.4.6. Soit σ la substitution définie sur $\{a, b, c\}$ par $\sigma(a) = abc$, $\sigma(b) = ca$ et $\sigma(c) = bb$. La substitution σ admet un point fixe $w = abccabbbbabccacacacaabccabbbbabcbbabcbbabcbb\dots$. On symbolise les k-découpages de w par des traits verticaux entre deux itérés du type $\sigma^k(a)$:

$$\begin{array}{ll} w = |abc|ca|bb|bb|abc|ca|ca|ca|abc|ca|bb|bb|abc|bb|abc|bb|abc|bb|\dots & \text{1-découpage de } w \\ w = |abccabb|bbabc|caca|caca|abccabb|bbabc|bbabc|bbabc|bb\dots & \text{2-découpage de } w \\ w = |abccabbbbabccaca|cacaabccabb|bbabcbbabc|bbabcbb\dots & \text{3-découpage de } w \end{array}$$

Les k-découpages des points fixes d'une substitution de longueur constante q sont symbolisés par des barres verticales toutes le q^k lettres. Par exemple, le mot de Morse présenté à l'exemple I.2.8, donné par $m = abbabaabbaababbabaabbaabbabaabbaababbaabbabaababba\dots$, admet les k-découpages suivants :

$$\begin{array}{ll} m = ab|ba|ba|ab|ba|ab|ab|ba|ba|ab|ab|ba|ab|ba|ba|ab|ba|ab|ab|ba|ab|ba|\dots & \text{1-découpage de } m \\ m = abba|baab|baab|abba|baab|abba|abba|baab|baab|abba|abba|\dots & \text{2-découpage de } m \\ m = abbabaab|baababba|baababba|abbabaab|baababba|abba\dots & \text{3-découpage de } m \end{array}$$

Le mot aba n'a pas le même 1-découpage lorsqu'il apparait au rang 0 : $ab|a$ et lorsqu'il apparait au rang 3 : $a|ba$, mais si on observe les mots de 4 lettres du mot de Morse : $abba$, $bbab$, $babb$, $baab$, $aaba$ et $abaa$, leurs 1-découpages sont indépendants du rang où ils apparaissent. En effet, puisque aa et bb ne sont pas des images de lettres, il y a donc une barre de 1-découpage entre deux occurrences consécutives de a ou de b. Ainsi, on a les découpages suivants :

$ab|ba$, $b|ba|b$, $b|ab|b$, $ba|ab$, $a|ab|a$ et $a|ba|a$. La substitution de Morse est donc déterminée à l'ordre 1.

La propriété de détermination permet, dans le cas d'une substitution de longueur constante, d'obtenir une relation linéaire vérifiée de la fonction de complexité de ses points fixes, qui sont des mots automatiques.

Théorème II.4.7. *(Mossé 1996)*

Soient m un point fixe d'une substitution primitive σ de longueur constante q, déterminée à l'ordre 1 et C_1 l'entier attaché à cette propriété.

Pour tout entier n, on note x_n le quotient et r_n le reste de la division euclidienne de n par q : $n = qx_n + r_n$ avec $0 \le r_n < q$. On a alors :

1. *Pour $n \ge C_1$, la valeur de la fonction de complexité $p_m(n)$ est donnée par les relations suivantes :*

$$p_m(n) = \begin{cases} p_m(x_n) + (q-1)p_m(x_n+1) & \text{si } r_n = 0, \\ (q - r_n + 1)p_m(x_n+1) + (r_n - 1)p_m(x_n+2) & \text{sinon.} \end{cases}$$

2. *Pour $n \geq C_1$, le mot s, défini, pour tout $n \geq 0$, par $s_n = p_m(n+1) - p_m(n)$ est q-automatique et on a :*

$$s_n = p_m(n+1) - p_m(n) = \begin{cases} p_m(x_n+1) - p_m(x_n) & \text{si } r_n = 0, \\ p_m(x_n+2) - p_m(x_n+1) & \text{sinon.} \end{cases}$$

En particulier, s_n, qui donne le nombre de facteurs spéciaux de m de longueur n compté avec multiplicités prend un nombre fini de valeurs.

Nous renvoyons à [Mos96] pour la démonstration de ce théorème.

La propriété de détermination à l'ordre K est une propriété très forte. Il existe une notion plus faible, appelée reconnaissabilité.

Définition II.4.8. Soit σ une substitution sur un alphabet A fini qui admet un point fixe m.

On note $E_1 = \{|\sigma(m_0 m_1 \dots m_n)|,\ n \in \mathbb{N}\}$.

On dit que σ est *reconnaissable* s'il existe un entier C tel que si $m_{i-C} \dots m_{i+C} = m_{j-C} \dots m_{j+C}$ pour un certain $i \in E_1$, alors j appartient aussi à E_1.

Remarque II.4.9. Cette notion apparait sous le terme de *reconnaissabilité bilatère* dans [Mos92] et [Mos96]. Cette propriété dit, en substance, que lorsqu'on extrait un mot w suffisamment long d'un point fixe d'une substitution reconnaissable, son 1-découpage ne dépend pas de l'endroit d'où a été extrait w (sauf eventuellement sur son préfixe de longueur C et son suffixe de longueur C) puisqu'il existe alors une écriture $w = s\sigma(u_1)\sigma(u_2)\dots\sigma(u_n)p$ où n est un entier, et $u_1u_2\dots u_n$ est un facteur de m, s est un suffixe strict de u_0 et p est un préfixe strict de u_{n+1}.

Attention : il se peut qu'il n'y ait pas unicité du mot $u_1u_2\dots u_n$, si par exemple, deux lettres de A ont la même image par σ ou à cause d'« effets de bord » : par exemple, il peut exister une lettre l telle que $\sigma(u_n)$ soit un préfixe de $\sigma(l)$, $\sigma(l)$ soit préfixe de $\sigma(u_n)p$ et que $u_1u_2\dots u_{n-1}l$ soit un mot de m ; ainsi, w admet également le 1-découpage $w = s\sigma(u_1)\sigma(u_2)\dots\sigma(l)p'$.

Il y a unicité de la décomposition du mot w uniquement au cas où, pour tout couple de lettres (a, b) de A, $\sigma(a)$ et $\sigma(b)$ commencent et finissent par deux lettres distinctes (dans ce cas, une substitution reconnaissable est détemininée) ;

d'autre part, lorsque σ est injective, les problèmes d'unicité ne peuvent se produire qu'au niveau des couples (s, u_1) et (u_n, p).

Il est nécessaire d'imposer plusieurs conditions sur la substitution σ qui engendre m pour avoir l'unicité des mots dont proviennent les facteurs suffisamment longs de m (nous renvoyons à [Mos92] et [Mos96]).

Dans le cas des substitutions primitives de longueur constantes, ces differentes notions de synchronisation permettent de montrer le théorème suivant.

Théorème II.4.10. *(Mossé 1996)*

Si m est un point fixe d'une substitution primitive de longueur constante q, alors, les mots p et s définis, pour tout entier n, par $p_n = p_m(n)$ et $s_n = p_m(n+1) - p_m(n)$ ont les propriétés suivantes :

- *p est un mot q-régulier,*
- *s est un mot q-automatique.*

Ainsi, ce théorème affirme que pour les mots q-automatiques engendrés par des substitutions primitives de longueur constante, la complexité p_m vérifie une relation du type : il existe un entier r et r mots infinis $w^{(1)}$, $w^{(2)}$, ... et $w^{(r)}$ à valeurs dans $\mathbb{N}$ (voir I.6.2.1) tels que :

$$\forall i \in [\![0, q-1]\!], \exists (a_{i,1}, a_{i,2}, \dots, a_{i,r}) \in \mathbb{N}^r, \quad \forall n \in \mathbb{N},\ p_m(nq+i) = \sum_{j=1}^{r} a_{i,j} w_n^{(j)},$$

et que, d'autre part, le mot formé par les quantités successives de facteurs spéciaux est reconnu par un automate fini ; en particulier, le nombre de facteurs spéciaux ne prend qu'un nombre fini de valeurs.

Dans les chapitres IV et V, nous présentons des résultats sur les fonctions de complexité de mots infinis dits q^∞-automatiques, que nous allons introduire au chapitre suivant. Le chapitre V présente un résultat analogue au théorème de Cobham (théorème II.4.1).

Dans ces deux chapitres, nous effectuons également des minorations de fonctions de complexité de tels mots en faisant appel aux facteurs spéciaux et à des principes de synchronisation similaires à ceux présentés ci-dessus.

Chapitre III

Mots q^∞-automatiques

Ce chapitre présente une classe de mots, proche de celle des mots automatiques : les mots q^∞-automatiques. De le même manière que les mots automatiques sont reconnus par les q-AFDS, les mots q^∞-automatiques sont engendrés par des q-automates dénombrables avec fonction de sortie « admissible ». Ils peuvent aussi être vus comme projections de points fixes de substitutions de longueur constante sur un alphabet dénombrable. Nous définissons ici ces mots et explorons les propriétés combinatoires et dynamiques des points fixes de substitutions de longueur constante sur un alphabet dénombrable. Les propriétés combinatoires de ces objets permettrons d'obtenir des résultats sur la complexité des mots q^∞-automatiques, qui seront exposés aux chapitres suivants.

Dans tout ce chapitre, q désigne un nombre entier supérieur ou égal à 2.

III.1 q-automates dénombrables

Définitions III.1.1. Un *q-automate dénombrable déterministe* (q-ADD) est la donnée d'un quadruplet $\mathcal{A} = (E, \phi, e_0, F)$, où :
- E est un ensemble dénombrable, appelé *ensemble des états*,
- ϕ est une fonction d'une partie de $E \times [\![0, q-1]\!]$ dans E. Un triplet (e, i, e') tel que $\phi(e, i) = e'$ est appelé *transition* et noté $e \cdot i = e'$,
- e_0 est un élément de E, appelé *état initial*,
- F est un sous-ensemble fini de E, appelé *ensemble des états finaux*.

Si pour tout couple (e, i) de $E \times [\![0, q-1]\!]$, $e \cdot i$ est défini, l'automate est dit *complet*.

Un q-ADD est représenté par son *graphe des transitions*, graphe dont les sommets sont indexés par E où les transitions $e \cdot a = s$ sont symbolisées par des flèches de e vers s indexées par a.

Exemple III.1.2. L'automate $\mathcal{A} = (\mathbb{Z}, \phi, \{0\}, \{0\})$, où $\phi(n, 0) = n - 1$ et $\phi(n, 1) = n + 1$ pour tout $n \in \mathbb{Z}$ est un 2-ADD complet, dont le graphe des transitions est représenté à la figure III.1.

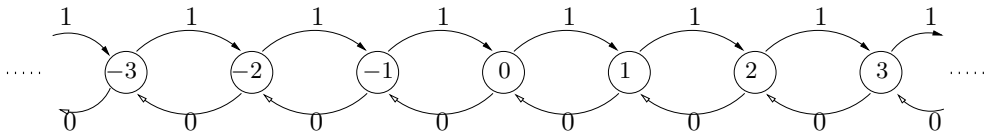


FIG. III.1 – 2-automate dénombrable déterministe complet

Exemple III.1.3. L'automate représenté à la figure III.2 est un 4-ADD non complet. C'est essentiellement le graphe des transitions de l'automate à pile associé au langage de Dyck sur lequel on aurait modifié les étiquettes des arcs, avec les correspondances $\overline{p}_1 \mapsto 0$, $\overline{p}_0 \mapsto 1$, $p_0 \mapsto 2$ et $p_1 \mapsto 3$.

Remarque III.1.4. L'hypothèse de complétude sera toujours faite dans la suite. En effet, il est possible de modifier un q-ADD, sans changer le langage engendré, de sorte que l'automate soit complet. il suffit pour cela de rajouter un état « poubelle » x, sur lequel on fait arriver toutes les transitions (arcs du graphe) qui ne sont pas définies : si $e \cdot i$ n'est pas définie, on pose $e \cdot i = x$.

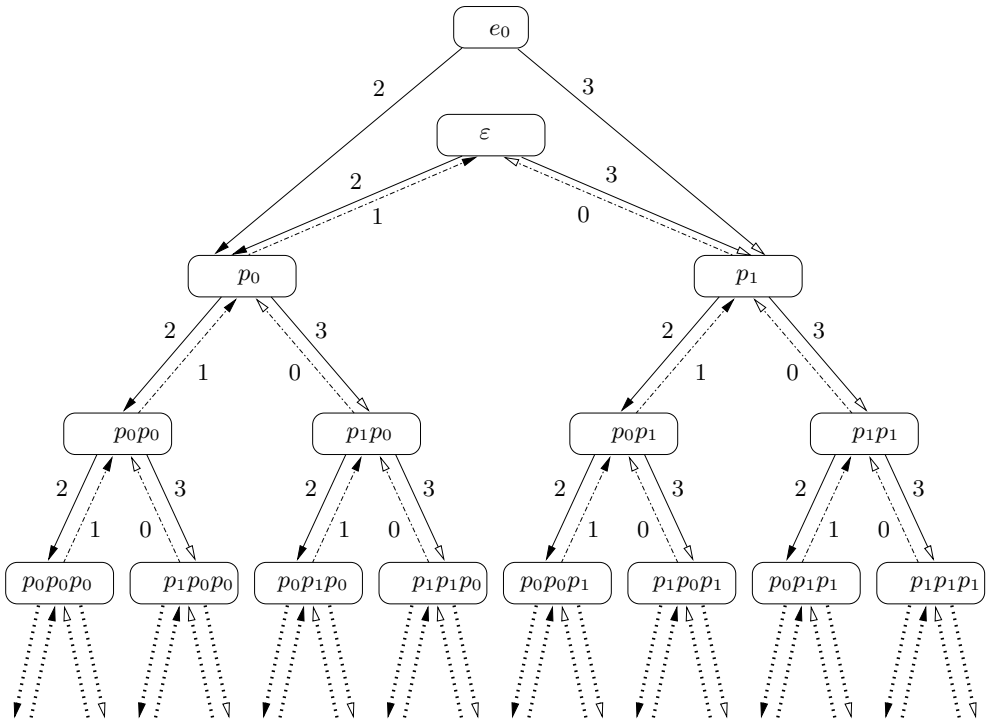


FIG. III.2 – 4-automate dénombrable déterministe non complet

Définitions III.1.5. Soit $\mathcal{A} = (E, \phi, e_0, F)$ un q-ADD. On étend la définition des transitions de $\mathcal{A}$ au monoïde $[\![0, q-1]\!]^*$ de la manière suivante :

On définit, pour tout $e \in E$, $e \cdot \varepsilon = e$ et pour tout mot v de A^* et toute lettre a de A, on pose $e \cdot (va) = (e \cdot v) \cdot a$.

On appelle *trajectoire d'un mot* $m = m_0 m_1 \dots m_n$ de A^* *dans l'automate*, le $(n+1)$-uplet d'élements $(e_0, e_1, \dots, e_{n+1})$ de E, tel que, pour tout $i \in [\![0, n]\!]$, $e_{i+1} = e_i \cdot m_i$, de sorte que $e_{i+1} = e_0 \cdot m_0 m_1 \dots m_i$.

Si $\mathcal{A}$ est un q-ADD complet, on appelle *mot de sortie de* $\mathcal{A}$, le mot infini $\mu(\mathcal{A})$ de $E^{\mathbb{N}}$ défini pour tout entier $n \geq 0$ par $\mu(\mathcal{A})_n = e_0 \cdot \overline{n}^q$, où $\overline{n}^q = n_l \dots n_1 n_0$ est la représentation en base q de $n = \sum_{i=0}^{l} n_i q^i$.

Le *langage accepté par l'automate* $\mathcal{A}$, noté $L(\mathcal{A})$ est défini par :

$$L(\mathcal{A}) = \{w \in A^*,\ e_0 \cdot w \in F\}.$$

C'est l'ensemble des mots dont les trajectoires dans l'automate finissent sur un élément de F.

De la même manière que pour les q-automates finis, on peut étendre le concept de q-automate dénombrable par celui de q-automate dénombrable avec fonction de sortie.

Il est évident que l'introduction d'une fonction de sortie multiplie énormément les possibilités de description des ADD. Nous nous restreindrons donc à un certain type de fonction de sortie, appelées projections admissibles :

Définition III.1.6. On dit qu'une projection lettre-à-lettre Π d'un alphabet dénombrable E vers un alphabet fini A est une *projection admissible* s'il existe un sous-ensemble fini F_Π de E tel que la restriction de Π à $E \setminus F_\Pi$ est constante.

Définition III.1.7. Un *q-automate dénombrable déterministe et admissible* (q-ADDA) est la donnée du plet $\mathcal{A} = (E, \phi, e_0, A, \Pi)$ où
- E est un ensemble dénombrable, appelé *ensemble des états*,
- ϕ est une fonction d'une partie de $E \times [\![0, q-1]\!]$ vers E. Un triplet (e, i, e') tel que $\phi(e, i) = e'$ est appelé *transition* et noté $e \cdot i = e'$,
- e_0 est un élément de E, appelé *état initial*,
- A est un alphabet fini,
- Π est une projection admissible de E vers A.

III.2 Correspondance entre q-ADD et substitutions sur un alphabet dénombrable

Il existe une correspondance entre les q-ADD complets dont l'ensemble des états est E et les substitutions de longueur constante q sur l'alphabet E (voir [Mau06]). Cette correspondance est analogue à celle décrite par Cobham [Cob72] 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 I.2, 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 $\{(e, i, e') \in E \times \{0, 1, \ldots, q-1\} \times E, \overline{\sigma}_i(e) = e'\}$. 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 $\mathcal{A}_{\overline{\sigma}} = (E, \phi_{\overline{\sigma}}, e_0, E, Id_E)$, où :
- l'ensemble des états est E,
- la fonction de transition $\phi_{\overline{\sigma}}$ est donnée par :
  $\forall i \in \{0, 1 \ldots, q-1\}, \forall e \in E,\ \phi_{\overline{\sigma}}(e, i) = \overline{\sigma}_i(e)$.
- l'état initial est e_0.

De plus, si $\overline{\sigma}$ admet un point fixe infini $\overline{m}$ contenu dans $e_0 E^{\mathbb{N}}$, 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 III.5.1.

Notation III.2.1. Soient E un alphabet dénombrable et A un alphabet fini.

Si $\overline{\sigma}$ est une substitution donnée sur l'alphabet E et Π est une projection admissible de E vers A, on note donc $(E, \phi_{\overline{\sigma}}, e_0, A, \Pi)$ le q-ADDA associé à $\overline{\sigma}$ et Π.

La correspondance que nous venons de décrire permet alors d'énoncer la proposition suivante :

Proposition III.2.2. *Soient E un alphabet dénombrable et A un alphabet fini.*

Si $\overline{\sigma}$ est une substitution de longueur constante q sur l'alphabet E qui admet un point fixe infini dans $e_0 E^{\mathbb{N}}$ et Π est une projection admissible de E vers A alors l'automate $(E, e_0, \phi_{\overline{\sigma}}, A, \Pi)$, associé à $\overline{\sigma}$ est un q-ADDA complet qui vérifie $e_0 \cdot 0 = e_0$.

Si (E, e_0, ϕ, A, Π) est un q-ADD complet, tel que $e_0 \cdot 0 = e_0$, on peut construire une substitution $\overline{\sigma}$ de longueur q sur l'alphabet E en posant, pour tout $e \in E$ et tout $i \in [\![0, q-1]\!]$, $\overline{\sigma}_i(e) = e \cdot i$. et la substitution $\overline{\sigma}$ a un point fixe infini dans $e_0 E^{\mathbb{N}}$.

Exemple III.2.3. La figure III.3 est le graphe de 2-ADD complet tel que $e_0 \cdot 0 = e_0$, obtenu à partir de l'ADD présenté à l'exemple III.1.2.

La substitution associée à ce 2-ADDA est

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

Exemple III.2.4. La modification de l'4-ADD présenté à l'exemple III.1.3 consiste à rajouter deux états e_0 et x ainsi que les transisions suivantes : $\forall i \in [\![0, 3]\!],\ x \cdot i = x$, $e_0 \cdot 0 = e_0$, $e_0 \cdot 1 = x$, $e_0 \cdot 2 = p_0$, $e_0 \cdot 3 = p_1$ et si $e \cdot i$ n'est pas définie, on pose $e \cdot i = x$. On retrouvera cet automate à la figure VI.1.

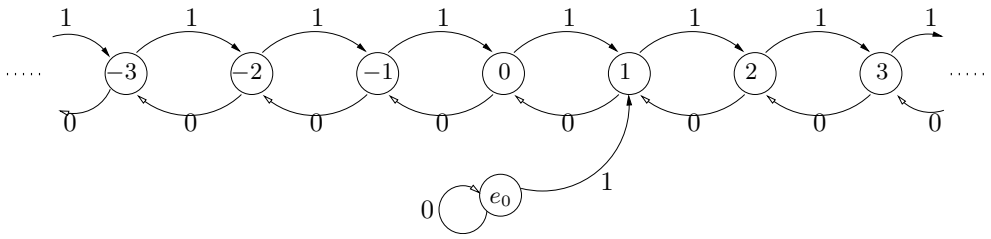


FIG. III.3 – ADD complet obtenu à partir de l'ADD représenté à la figure III.1.

La substitution que l'on associe alors à ce 4-ADD modifié est donnée par :

$$\begin{array}{rccc} \overline{\delta}: & E & \to & (\{p_0,p_1\}^* \cup \{e_0,x\})^4 \\ & \varepsilon & \mapsto & xx(p_0)(p_1) \\ & e=p_0v & \mapsto & x(v)(p_0e)(p_1e) \\ & e=p_1v & \mapsto & (v)x(p_0e)(p_1e) \\ & x & \mapsto & xxxx \\ & e_0 & \mapsto & e_0x(p_0)(p_1) \end{array}$$

III.3 Mots q^∞-automatiques

L'idée sous-jacente à l'introduction des q-ADD et des q-ADDA est de construire, de la même manière que les mots automatiques sont engendrés par des automates finis, une classe de mots engendrés par les q-automates dénombrables admissibles.

Définition III.3.1. Soit A un alphabet fini.

On dit qu'un mot infini m à valeurs dans A est q^∞-*automatique* s'il existe un q-ADDA complet $\mathcal{A}=(E,\phi,e_0,A,\Pi)$ vérifiant $e_0\cdot 0=e_0$, tel que, pour tout n, $m_n=\Pi(e_0\cdot n)$.

Autrement dit, les mots q^∞-automatiques sont les mots de sortie des q-ADDA.

Un sous-ensemble d'entiers R est q^∞-*automatique* si son *mot infini caractérisque* m, défini par :

$$\forall n\in\mathbb{N},\ m_n=\begin{cases} 1 & \text{si } n\in R,\\ 0 & \text{sinon.}\end{cases}$$

est un mot q^∞-automatique.

Remarque III.3.2. la convention $e_0\cdot 0=e_0$ imposée à l'automate permet de pouvoir rentrer dans l'automate des écritures d'entier en base q qui ne sont pas propres, c'est-à-dire complétées à gauche par un nombre quelconque de zéros, sans changer l'état de sortie. Cela parait être une hypothèse forte mais on peut toujours s'y ramener en modifiant l'automate $\mathcal{A}=(E,\phi,e_0,E',\Pi)$ qui engendre notre mot : On rajoute à E un état e_0', qui sera le nouvel état initial, on étend ϕ et Π à $E\cup\{e_0'\}\times[\![0,q-1]\!]$ en posant $e_0'\cdot 0=e_0'$, pour tout i de $[\![1,q-1]\!]$, $e_0'\cdot i=e_0\cdot i$ et $\Pi(e_0')=\Pi(e_0)$. Le nouvel automate $\mathcal{A}'$ engendre le même mot q^∞-automatique que $\mathcal{A}$ et vérifie notre hypothèse.

Exemple III.3.3. Le mot de sortie de l'automate représenté à la figure III.3 est le mot infini $\overline{m}$ suivant :

$$\begin{array}{rl} \overline{m}= & e_0102(-1)113(-2)0020224(-3)(-1)(-1)1(-1)113(-1)1131335(-4)(-2)(-2)0 \\ & (-2)002(-2)0020224(-4)(-2)(-2)0(-2)002(-2)002022402242446\ldots \end{array}$$

et l'image par Π, projection admissible définie sur $\{0,1\}$ par $\Pi^{-1}(\{1\})=\{0\}$, est le mot 2^∞-automatique suivant :

$$m=0010000001101000000000000000000000010110011010000110100010000000\ldots$$

Ce mot est le mot indicateur de l'ensemble des entiers qui contiennent autant d'occurrences de 0 que d'occurrences de 1 dans leur écriture binaire propre. Nous retrouverons cet exemple au paragraphe IV.3.

Exemple III.3.4. Le mot de sortie de l'automate de l'exemple III.2.4 est le mot infini $\overline{\delta}$ suivant :

$$\overline{d} = e_0x(p_0)(p_1)(x)^5\varepsilon(p_0^2)(p_1p_0)\varepsilon x(p_0p_1)(p_1^2)(x)^{22}(p_0)(p_1)x(p_0)(p_0^3)(p_1p_0^2)(p_0)x(p_0p_1p_0)(p_1^2p_0)(x)^2\dots$$
$$\dots(p_0)(p_1)(x)^5(p_1)(p_0^2p_1)(p_1p_0p_1)(x)^{72}\varepsilon x(p_0^2)(p_1p_0)x\varepsilon(p_0p_1)(p_1^2)(x)^{16}\varepsilon x(p_0^2)(p_1p_0)(x)^4(p_0^2)x(p_0^4)\dots$$
$$\dots(p_1p_0^3)x(p_0^2)(p_0p_1p_0^2)(p_1^2p_0^2)(x)^4\varepsilon x(p_0^2)(p_1p_0)(p_1p_0)x(p_0^2p_1p_0)x(p_1p_0)(p_0p_1^2p_0)(p_1^3p_0)x^{24}\varepsilon x(p_0^2)\dots$$
$$\dots(p_1p_0)x\varepsilon(p_0p_1)(p_1^2)x\varepsilon(p_0p_1)(p_1^2)(x)^4\dots$$

et l'image par Π, projection admissible définie sur $\{0,1\}$ par $\Pi^{-1}(\{1\}) = \{0\}$, est le mot 4^∞-automatique suivant :

$$d = 10^710^410^{118}10^410^310^{18}10^{19}10^{34}10^410^310\dots$$

Ce mot est le mot indicateur de l'ensemble des entiers dont l'image de l'écriture propre en base 4 par le morphisme de monoïdes $\mathcal{D} : [\![0,3]\!]^* \to \{\overline{p}_0, \overline{p}_1, p_0, p_1\}^*$ défini par $0 \mapsto \overline{p}_0$, $1 \mapsto \overline{p}_1$, $2 \mapsto p_0$, et $3 \mapsto p_1$ est un mot de Dyck sur les deux types de parenthèses p_0 et p_1. Nous reviendrons sur cet exemple au paragraphe VI.

L'intérêt majeur des deux hypothèses (complétude et $e_0 \cdot 0 = e_0$) que nous imposons aux q-ADDA que nous considérons dans la suite, réside dans le fait que les substitutions associées à ces automates sont de longueur constante q (grâce à la complétude) et admettent au moins un point fixe (grâce à $e_0 \cdot 0 = e_0$).

Ainsi, l'étude des mots q^∞-automatiques peut être envisagée sous les deux éclairages différents : comme projection admissible d'un point fixe de substitution de longueur constante q ou comme un mot de sortie d' un q-ADDA complet. Les deux visions (substitutives et automatiques) des mots q^∞-automatiques se complètent. Dans la suite, c'est le coté substitutif qui sera privilégié pour les démonstrations. Cependant, la vision automatique de ces objets restera omniprésente, essentiellement car il amène une représentation visuelle importante à la compréhension.

Les mots indicateurs de certains langages algébriques de $[\![0, q-1]\!]^*$ sont des mots q^∞-automatiques. Plus précisément :

Proposition III.3.5. *Soit S un sous-ensemble de $\mathbb{N}$.*

Si la famille $\overline{S}^q$ des représentations propres en base q des éléments de S forme un langage algébrique de $[\![0, q-1]\!]^*$, engendré par admissibilité par pile vide (ou par pile vide et états d'acceptation) par un automate à pile déterministe et temps réel, alors l'ensemble S est q^∞-automatique.

Cette proposition illustre le fait que la famille des q-ADD contient les graphes des transitions des automates à pile déterministes en temps réel dont l'alphabet des entrées est $[\![0, q-1]\!]$:

■ **Démonstration de la proposition III.3.5** Soit $\mathcal{A}_\mathcal{P} = (E, A, P, \phi, e_0W_0, K)$ l'AP déterministe et en temps réel qui engendre, par vidage de pile, le langage $\overline{S}^q$ des représentations en base q des éléments de S et soit $\mathcal{G}_T(\mathcal{A}_P)$ son graphe des transitions. K est du type $F\varepsilon$, pour $F \subset E$, l'ensemble des sommets de $\mathcal{G}_T(\mathcal{A}_P)$ est $E \times P^*$ et il existe un arc de $\mathcal{G}_T(\mathcal{A}_P)$, étiqueté par $a \in A$ de eW vers $e'W$ si et seulement si $(e, XW) \stackrel{a}{\Longrightarrow} (e', X'VW)$.

Le q-ADDA qui reconnait le mot infini indicateur de S est donné par : $\mathcal{A} = (E \times P^*, \phi, e_0, \{0,1\}, \Pi)$, où les transitions sont définies par $eW \cdot a = e'X'VW$ si et seulement si $(e, XW) \stackrel{a}{\Longrightarrow} (e', X'VW)$, et la projection admissible est donnée par $\Pi : E \times P^* \to \{0,1\}$ vérifie : $\Pi^{-1}(\{1\}) = F\varepsilon$.

Si e_0W_0 ne vérifie pas $e_0W_0 \cdot 0 = e_0W_0$, on rajoute un état $e_0' \notin EP^*$, on définit les transitions $e_0' \cdot 0 = e_0'$ et pour tout $a \in [\![1, q-1]\!]$, on définit $e_0' \cdot a = e_0W_0 \cdot a$ et $\Pi(e_0') = \Pi(e_0W_0)$.

Si l'automate obtenu n'est pas complet, ou rajoute un état $x \notin EP^*$ et toutes les transitions manquantes, c'est-à-dire que s'il existe des couples (eW, a) de $EP^* \times [\![0, q-1]\!]$ pour lesquels $\phi(eW, a)$ n'est pas défini, on les définit par $eW \cdot a = x$, et, pour tout $a \in [\![0, q-1]\!]$, on définit $x \cdot a = x$ et $\Pi(x) = 0$. L'automate dénombrable est bien un q-ADDA qui renvoie 1 pour les éléments de $\overline{S}^q$ et 0 sinon. Le mot infini indicateur de S est donc q^∞-automatique est donc S est q^∞-automatique. ■

Remarque III.3.6. Tous les mots q^∞-automatiques sur $\{0,1\}$ ne sont pas les mots caractéristiques d'ensembles algébriques engendrés par des AP déterministes en temps réels car la géométrie des q-ADD que nous considérons n'est pas aussi rigide que celle imposée aux graphes des transitions des AP (Voir I.5).

Proposition III.3.7. *Soit m un mot q^∞-automatique à valeurs dans un alphabet A.*

Le mot infini w tel que $m_n = w_n$ sauf pour un nombre fini d'entiers n est q^∞-automatique.

Si m' est un autre mot q^∞-automatique à valeurs dans un alphabet A' et $f : A \times A' \to A''$ est une fonction, le mot w tel que $w_n = f(m_n, m'_n)$ est aussi q^∞-automatique.

Proposition III.3.8. *La famille des sous-ensembles d'entiers q^∞-automatiques est stable par intersection et par intersection avec un sous-ensemble q-automatique. Elle n'est cependant pas stable ni par passage au complémentaire ni par union.*

On montre la stabilité par intersection et par intersection avec un sous-ensemble q-automatique en construisant les q-ADDA correspondants. Ainsi, par exemple, l'ensemble des entiers $\mathcal{E}$ dont les représentations propres en base 4 contiennent autant d'occurrences de 0 que de 1, un nombre pair d'occurrences de 3 et un nombre quelconque de 2 est 4^∞-automatique.

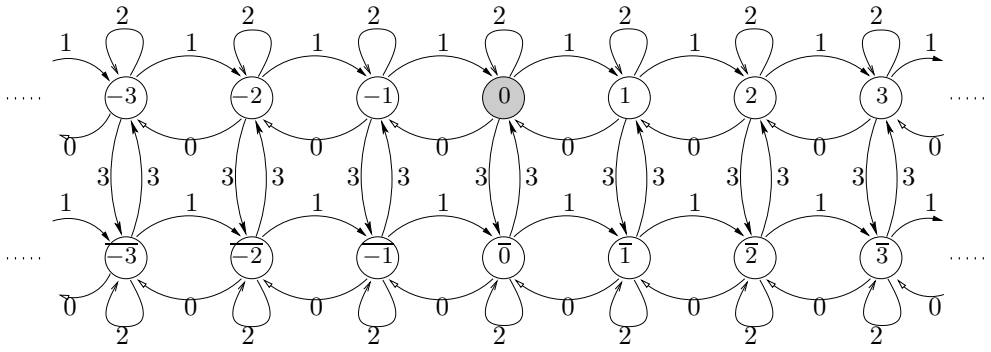


FIG. III.4 – 4-ADDA reconnaissant les représentations en base 4 des entiers de $\mathcal{E}$.

III.4 Classes de q-ADDA remarquables

Notations III.4.1. Soient $\overline{\sigma}$ une substitution sur l'alphabet dénombrable E, admettant un point fixe dans $e_0E^{\mathbb{N}}$ et Π une projection admissible de E dans un alphabet fini A.

On note F_Π le sous ensemble fini de E de cardinal minimal, en dehors duquel Π est constante ($\Pi(E \setminus F_\Pi) = \{a_0\}$).

On note $|\Pi|$ le cardinal de F_Π.

Soit $\mathcal{A} = (E, e_0, \phi_{\overline{\sigma}}, A, \Pi)$ le q-ADDA associé à $\overline{\sigma}$ et Π.

On note, pour tout état $e \in E$ et tout entier $k \geq 0$, $\mathcal{Z}_k(e)$ l'ensemble des éléments de E, accessibles dans le graphe de $\mathcal{A}$, en utilisant uniquement les arcs inverses des transitions, à partir de l'élément e :

$$\mathcal{Z}_k(e) = \{e' \in E,\ \exists w \in [\![0, q-1]\!]^k, e' \cdot w = e\} = \bigcup_{(i_1,\ldots,i_k)\in[\![0,q-1]\!]^k} \overline{\sigma}_{i_1}^{-1} \circ \overline{\sigma}_{i_2}^{-1} \circ \cdots \circ \overline{\sigma}_{i_k}^{-1} (\{e\}).$$

On note E_Π l'ensemble des éléments de E, accessibles dans le graphe de $\mathcal{A}$, en utilisant uniquement les arcs inverses des transitions, à partir d'un élément de F_Π :

$$E_\Pi = \{e \in E,\ \exists w \in [\![0, q-1]\!]^*, \exists f \in F_\Pi, e \cdot w = f\} = \bigcup_{k\geq 0} \bigcup_{f\in F_\Pi} \mathcal{Z}_k(f).$$

Définition III.4.2. Soit $\mathcal{A} = (E, e_0, \phi, A, \Pi)$ un q-ADDA.

Soit e un élément de E. On appelle *degré entrant de e* le cardinal de l'ensemble des transitions de $\mathcal{A}$ qui terminent sur e, c'est-à-dire le cardinal de l'ensemble $\{(e', i) \in E \times [\![0, q-1]\!],\ \phi(e', i) = e\}$.

On dit que $\mathcal{A}$ est *de degré fini pour* Π ou simplement *de degré fini* si le degré entrant de tout les éléments de E_Π est fini.

On dit que $\mathcal{A}$ est *de degré borné par K pour* Π ou simplement *de degré borné par K* si le degré entrant de tous les éléments de E_Π est borné par K.

Si $\mathcal{A}$ n'est pas de degré fini, on dit qu'il est *de degré infini*.

Une substitution $\overline{\sigma}$ de longueur constante sur E définie par $\overline{\sigma}(x) = \overline{\sigma}_0(x)\overline{\sigma}_1(x)\ldots\overline{\sigma}_{q-1}(x)$ est *K-uniformement bornée pour* Π ou simplement *K-uniformement bornée*, pour une projection admissible Π, s'il existe une constante $K \geq 1$ telle que, pour $j = 0, 1, \ldots, q-1$:

$$\forall x \in E_\Pi,\ Card\left(\overline{\sigma}_j^{-1}\{x\}\right) \leq K. \tag{III.1}$$

Remarques III.4.3. La condition (III.1) dépend uniquement de la géométrie du graphe non orienté qui porte le graphe du q-ADDA. Intuitivement, cette condition purement géométrique impose le fait que la structure des q-automates dénombrables considérés est localement la même que celle des q-automates finis ayant au plus K états. De plus, le fait que la condition (III.1) ne porte que sur les éléments de E_Π permet que certains q-ADDA non complets puissent être aussi considérés comme de degré borné malgré l'ajout d'un état « poubelle » qui peut être de degré infini.

Les graphes des transitions d'automates à pile déterministes et en temps réel vérifient la condition (III.1) lorsque Π est une projection constante sauf sur un nombre fini de configurations internes (voir paragraphe I.5).

Exemples III.4.4. Le 2-ADDA et le 4-ADDA des exemples III.3.3 et III.3.4 sont respectivement de degré borné par 1 et de degré borné par 2. Par contre le 2-ADDA associé à la substitution

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

pour un certain entier $K \geq 2$ et à la projection admissible Π qui vérifie $\Pi^{-1}(\{1\}) = \{0\}$ est de degré fini.

Le 2-ADDA associé à la substitution

$$\begin{array}{rccl} \overline{\theta}: & \mathbb{N} & \to & \mathbb{N}^* \\ & e & \mapsto & 0(e+1) \end{array}$$

et à la projection admissible Π qui vérifie $\Pi^{-1}(\{1\}) = \{1\}$ est de degré infini.

Définitions III.4.5. Soit $\overline{\sigma}$ une substitution sur un alphabet dénombrable E. On suppose qu'elle est définie par $\overline{\sigma}(e) = \overline{\sigma}_0(e)\overline{\sigma}_1(e)\dots\overline{\sigma}_{q-1}(e)$.

La substitution $\overline{\sigma}$ est *monotone* s'il existe un ordre total sur E_Π pour lequel toutes les applications coordonnées $\overline{\sigma}_i$ sont monotones.

De manière équivalente, un q-ADDA $\mathcal{A} = (E, e_0, \phi, A, \Pi)$ est *monotone* s'il existe un ordre total sur E_Π pour lequel les q applications $\phi(., i)$ sont monotones.

Remarque III.4.6. La condition de monotonie impose une certaine « compatibilité » entre les sous-graphes orientés $\mathcal{G}_i$ dont l'ensemble des états est E_Π et les ensembles des transitions sont $\{e \cdot i = e', (e, e') \in E_\Pi\}$, puisqu'elle impose de pouvoir ordonner les éléments de E de sorte que les applications $\phi(., i)$ respectent toutes cet ordre, indépendamment les unes des autres.

Exemples III.4.7. Tous les q-ADDA présentés au III.4.4 sont monotones : Le 2-ADDA de l'ivrogne est monotone pour l'ordre naturel sur $\mathbb{Z}$, le 4-ADDA de Dyck sur deux types de parenthèses l'est pour l'ordre hiérarchique sur $\{p_0, p_1\}^*$ et les deux autres le sont pour l'ordre naturel sur $\mathbb{N}$.

Remarques III.4.8. La classe des q-ADDA de degré borné et la classe des q-ADDA monotones sont d'intersection non vide et aucune des deux n'est contenue dans l'autre. Cela est assez logique étant donné que les objets sur lesquels on impose des conditions sont de nature différente. Par exemple, le 3-ADDA associé à la substitution

$$\begin{array}{rccl} \overline{\varphi}: & \mathbb{N} \cup \{-1\} & \to & \mathbb{Z}^3 \\ & e & \mapsto & \left(\lfloor \frac{e}{2} \rfloor\right)((-1)^e)(e+1) \end{array}$$

et à la projection admissible Π qui vérifie $\Pi^{-1}(\{1\}) = \{0\}$ n'est pas monotone et de degré infini.

Le 3-ADDA associé à la substitution

$$\begin{array}{rccl} \overline{\psi}: & \mathbb{N} \cup \{-1\} & \to & \mathbb{Z}^3 \\ & e & \mapsto & (e-1)(e+(-1)^e)(e+1) \end{array}$$

et à la projection admissible Π qui vérifie $\Pi^{-1}(\{1\}) = \{0\}$ n'est pas monotone et de degré borné par 2.

Nous verrons dans les chapitres suivants que l'on peut obtenir des majorations de l'ordre de grandeur de la complexité des mots q^∞-automatiques engendrés par les q-ADDA de degré borné ainsi que par les q-ADDA monotones.

III.5 Combinatoire des mots de sortie des q-ADD

Les propriétés combinatoires d'un mot q^∞-automatique sont étroitement liées à la combinatoire du mot de sortie du q-ADD sous-jacent au q-ADDA qui l'engendre, c'est-à-dire aux points fixes de la substitution de longueur constante associée à ce q-ADDA. On se propose alors d'explorer certaines des propriétés combinatoires des points fixes de substitutions de longueur constante sur un alphabet dénombrable.

Voici donc le cadre de ce paragraphe :

Soient E un alphabet dénombrable et q fonctions $\overline{\sigma}_i : E \to E$ avec $i = 0, \dots, q-1$ et soit $\overline{\sigma}$ la substitution définie, comme au paragraphe I.2, sur E par $\overline{\sigma}(e) = \overline{\sigma}_0(e)\overline{\sigma}_1(e)\dots\overline{\sigma}_{q-1}(e)$. On suppose que $\overline{\sigma}_0(e_0) = e_0$, ce qui garantit l'existence d'un point fixe dans $e_0E^{\mathbb{N}}$.

Proposition III.5.1. *Pour un élément e donné de $F_1(\overline{m})$ et un entier $k \geq 1$, le k-ième itéré par $\overline{\sigma}$ de la lettre e, $\overline{\sigma}^k(e) = u_0u_1\dots u_{q^k-1}$, vérifie :*

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

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

■ **Démonstration de la proposition III.5.1** Cette proposition se démontre par récurrence sur k.

Avant de commencer cette récurrence, on remarquera que, $\overline{\sigma}$ étant une substitution de longueur constante, pour tout e appartenant à E et tout entier $k \geq 0$, le mot $\overline{\sigma}^k(e)$ est de longueur q^k.

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

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

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 $w = \overline{\sigma}^k\left(\overline{\sigma}_0(e)\overline{\sigma}_1(e)\dots\overline{\sigma}_{q-1}(e)\right) = \overline{\sigma}^k\left(\overline{\sigma}_0(e)\right)\overline{\sigma}^k\left(\overline{\sigma}_1(e)\right)\dots\overline{\sigma}^k\left(\overline{\sigma}_{q-1}(e)\right)$.

Soit $n \in \{0, 1, \dots, q^{k+1} - 1\}$ et $\overline{n}^q = n_l \dots n_1 n_0$ 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

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

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

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. ■

Une des principales utilités de cette proposition, est de pouvoir « valider » la correspondance entre point fixe de substitution de longueur q et mot de sortie du q-ADD que nous avons exposé au paragraphe III.2 :

Corollaire III.5.2. *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 du q-ADD $\mathcal{A} = (E, e_0, \phi_{\overline{\sigma}}, E, Id_E)$, défini au paragraphe III.2.*

On a l'égalité $\overline{m} = \mu(\mathcal{A})$.

■ **Démonstration du corollaire III.5.2**

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

Soit n un entier et soit k un entier tel que $n \leq q^k - 1$. Soit $\overline{n}^q = n_l \dots n_1 n_0$ 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 III.5.1, 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. ■

La proposition suivante explicite les mots de deux lettres qui apparaissent dans $\overline{m}$. Connaitre ces facteurs se révèlera essentiel pour pouvoir étudier la complexité des mots q^∞-automatiques.

Proposition III.5.3. *Soit $(e_1, \dots, e_{q-1}) \in E^{q-1}$ tel que $\overline{\sigma}(e_0) = e_0e_1\dots e_{q-1}$.*

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. $\exists k \geq 0, \exists i \in \{1,\dots,q-2\}$, $x_1 = \overline{\sigma}^k_{q-1}(e_i)$ *et* $x_2 = \overline{\sigma}^k_0(e_{i+1})$,
3. $\exists k \geq 0$, $x_1 = \overline{\sigma}^k_{q-1}(e_{q-1})$ *et* $x_2 = \overline{\sigma}^{k+1}_0(e_1)$,
4. $\exists k \geq 0, \exists j \in \{0,1,\dots,q-2\}$, $x_1 \in F_1(\overline{m})$ *et* $x_2 \in \overline{\sigma}^k_0 \circ \overline{\sigma}_{j+1} \circ \overline{\sigma}_j^{-1} \circ \overline{\sigma}^{-k}_{q-1}(\{x_1\})$.

■ **Démonstration de la proposition III.5.3** Cette proposition peut se montrer en utilisant le q-ADD associé à la substitution. Nous en donnons ici la version substitutive de la démonstration, en utilisant la réécriture de $\overline{m}$ suivante :

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

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 $\overline{\sigma}^k(e_{q-1})\overline{\sigma}^{k+1}(e_1)$ 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 $j \in \{0,1,\dots,q-2\}$.

(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 $k \geq 0$ 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 III.5.1 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 $k \geq 0$ 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 III.5.1 permet de dire qu'il existe un entier $0 \leq n < q^k - 1$ tel que x_1x_2 est donné par :

$$x_1 = \overline{\sigma}_{n_0} \circ \overline{\sigma}_{n_1} \circ \dots \circ \overline{\sigma}_{n_l} \circ \overline{\sigma}_0^{k-(l+1)}(e_j) \quad et \quad x_2 = \overline{\sigma}_{(n+1)_0} \circ \overline{\sigma}_{(n+1)_1} \circ \dots \circ \overline{\sigma}_{(n+1)_l} \circ \overline{\sigma}_0^{k-(l'+1)}(e_j),$$

avec $\overline{n}^q = n_l \dots n_1n_0$ et $\overline{n+1}^q = (n+1)_{l'} \dots (n+1)_1(n+1)_0$ avec $l \leq k$ et $l' = l$ ou $l+1$ selon le cas.

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

Si $n_{k'} \dots n_1n_0 = j(q-1)\dots(q-1)$ pour un certain $k' \leq l-1$ et un certain $j \in \{0,1,\dots,q-2\}$, on a alors $\overline{n+1}^q = n_l \dots n_{k'+1}(j+1)0\dots0$ et donc $x_2 \in \overline{\sigma}^{k'}_0 \circ \overline{\sigma}_{j+1} \circ \overline{\sigma}_j^{-1} \circ \overline{\sigma}^{-k'}_{q-1}(\{x_1\})$.

Si $\overline{n}^q = (q-1)\dots(q-1)$, alors $l \leq k-1$ car on se trouve à l'intérieur d'un $\overline{\sigma}^k(e_j)$ et $|n+1|_2 = 10\dots0$ avec $l' = l+1$, et on a $x_2 \in \overline{\sigma}^l_0 \circ \overline{\sigma}_1 \circ \overline{\sigma}_0^{-1} \circ \overline{\sigma}^{-l}_{q-1}(\{x_1\})$.

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

Remarque III.5.4. Comme nous le verrons aux chapitres suivants, l'utilisation de cette proposition est utile dans la détermination d'une majoration fine de la complexité d'un mot q^∞- automatique, notamment lorsque le q-ADDA qui l'engendre est de degré borné.

III.6 Un peu de dynamique des mots de sortie des q-ADD

Soit $\overline{\sigma}$ une substitution sur un alphabet dénombrable E, ayant un point fixe $\overline{m}$ dans $e_0E^{\mathbb{N}}$. On associe à m le système dynamique symbolique $(\Omega_{\overline{m}}, S)$, où $\Omega_{\overline{m}} = \overline{Orb_S(\overline{m})}$ est la fermeture de l'orbite de $\overline{m}$ sous l'action du décalage S ($S : m_0m_1m_2\dots \mapsto m_1m_2m_3\dots$).

On rappelle que la *matrice de la substitution* $\overline{\sigma}$, est la matrice $M_{\overline{\sigma}} = (M_{e,e'})_{(e,e')\in E^2}$ où $M_{(e,e')}$ est le nombre d'occurence de e' dans $\overline{\sigma}(e)$.

Remarque III.6.1. Lorsque le point fixe de la substitution est créé artificiellement, par l'ajout d'un élément e_0 dans E tel que $\overline{\sigma}_0(e_0) = e_0$ (cela se traduit par le fait que e_0 apparait dans m une unique fois) on pose, dans ce cas, $\Omega_{\overline{m}} = \overline{Orb_S(S(\overline{m}))}$ et $M_{\overline{\sigma}} = (M_{e,e'})_{(e,e')\in(E\setminus\{e_0\})^2}$.

De plus, on choisira toujours des substitutions ayant une propriété semblable (mais plus faible) à la condition de minimalité dans le cas où E est fini : on suppose que pour tout couple de lettres (e, e'), il existe un couple d'entiers (k, k') tels que e' apparait dans $\overline{\sigma}^k(e)$ et e apparait dans $\overline{\sigma}^{k'}(e')$; on appellera ces substitutitions des *substitutions faiblement primitives* ou *irreductibles*.

L'espace Ω_m est un espace fermé de $E^{\mathbb{N}}$ muni de la topologie des cylindres sur E ($E^{\mathbb{N}}$ n'est pas compact pour cette topologie) et invariant par le décalage S. Cette topologie sur $E^{\mathbb{N}}$ peut être aussi définie par la distance suivante :

$$\forall m \neq w \in E^{\mathbb{N}}, d(m, m') = 2^{-\inf\{n \in \mathbb{N} | m_n \neq m'_n\}},$$

L'opérateur de décalage S est continu sur $\Omega_{\overline{m}}$ pour cette topologie.

Proposition III.6.2. *Si $\overline{m} \in e_0 E^{\mathbb{N}}$ est le point fixe de $\overline{\sigma}$, alors, on a l'équivalence des 3 affirmations suivantes :*

1. *$w \in \Omega_{\overline{m}}$,*
2. *il existe une suite $(k_n)_{n\geq 0}$ telle que $w_0 w_1 \dots w_n = \overline{m}_{k_n} \overline{m}_{k_n+1} \dots \overline{m}_{k_n+n}$,*
3. *$F_n(w) \subset F_n(\overline{m})$.*

Cette proposition se montre exactement comme dans le cas où l'alphabet est fini (Voir [PF02]).

Définitions III.6.3. Soit $M = (M_{(e,e')})_{(e,e')\in E^2}$ une matrice dénombrable à coefficients dans $\mathbb{N}$. On note $M^{(n)}_{(e,e')}$ le terme général de la matrice M^n.

- La matrice M est *irréductible* si, pour tout couple $(e, e') \in E^2$, il existe un entier n tel que $M^{(n)}_{(e,e')} > 0$.
- La matrice M est *de période p* si, pour tout $e \in E$, $p = \mathrm{Pgcd}\left\{n, M^{(n)}_{(e,e)} > 0\right\}$. Si $p = 1$, on dit que M est *apériodique*.

Remarque III.6.4. L'irréductibilité de la matrice $M_{\overline{\sigma}}$ se traduit au niveau du graphe associé à $\overline{\sigma}$ par la propriété de forte connexité ainsi que par la faible minimalité de la substitution. On perd cette propriété lorsqu'on crée artificiellement un point fixe par rajout d'un état.

L'existence d'un point fixe pour $\overline{\sigma}$ assure l'apériodicité de la matrice $M_{\overline{\sigma}}$. Lorsqu'il n'y a pas d'autres points fixes que le point fixe créé artificiellement, comme c'est le cas, par exemple, pour la substitution de l'ivrogne et si la matrice de la substitution $\overline{\sigma}$ est périodique de période p ($p = 2$ dans le cas de la substitution de l'ivrogne) ; on considèrera alors le système dynamique associé au point fixe artificiel $\overline{m}$ contenu dans $e_0 E'^N$ de la substitution $\overline{\sigma}^p$, où E' désigne le sous-alphabet de $E \setminus \{e_0\}$ dans lequel $\overline{m}$ prend ses valeurs (la réduction de l'alphabet permet d'obtenir l'irréductibilité en plus de l'apériodicité) ; voir [Fer06].

Dans le cas de la substitution de l'ivrogne, on considère donc la substitution $\overline{\sigma}^2$ sur l'alphabet $E' = 2\mathbb{Z}$, définie par $\overline{\sigma}^2(e) = (e-2)ee(e+2)$. Cette substitution admet un point fixe artificel

$$\overline{m} = e_0 002(-2)002(-2)0020224(-4)(-2)(-2)0(-2)002(-2)0020224\dots$$

Le point fixe ayant été créé artificiellement, on considère alors $\Omega_{\overline{m}} = \overline{Orb_S(S(\overline{m}))}$ et le système dynamique qu'on associe à la substitution est $(\Omega_{\overline{m}}, S)$.

Il existe, pour les matrices dénombrables à coefficients dans $\mathbb{N}$, une extension de la théorie de Perron-Frobenius [Kit98]. C'est principalement cette extension qui nous fournira les résultats de ce paragraphe. On pourra aussi consulter [Fer06] pour certaines démonstrations.

Proposition III.6.5. *Un matrice dénombrable, à coefficients dans $\mathbb{N}$, irréductible et apériodique admet une valeur propre de Perron-Frobenius λ définie par*

$$\lambda = \lim_{n\to\infty} \left(M^{(n)}_{(e,e')}\right)^{\frac{1}{n}}.$$

De plus, on peut classifier les matrices irréductibles et apériodiques grâce à cette valeur propre :

- *M est* transiente *si $\sum_{n>0} M^{(n)}_{(e,e')} \lambda^{-n} < +\infty$,*

– M *est* récurrente *sinon.*

Les matrices récurrentes se divisent en 2 catégories : Pour M récurrente, on introduit, pour tout couple (e,e'), la suite $\left(l_{(e,e')}(n)\right)_{n>0}$, définie par $l_{(e,e')}(1) = M_{(e,e')}$ et $l_{(e,e')}(n+1) = \sum_{s\neq e} l_{(e,s)}(n)M_{(s,e')}$, on dit que :

– M *est* récurente nulle *si* $\sum_{n>0} n l_{(e,e')}(n)\lambda^{-n} < +\infty$,

– M *est* récurrente positive *sinon.*

On parlera de substitution transiente, récurrente nulle ou récurrente positive *selon la nature de la matrice $M_{\overline{\sigma}}$.*

Les termes utilisés ici proviennent de la théorie des chaînes de Markov.

Théorème III.6.6. *(Ferenczi 2006)*

Soit $\overline{\sigma}$ une substitution sur un alphabet dénombrable E.

La substitution $\overline{\sigma}$ admet un point fixe $\overline{m}$ et son graphe associé est fortement connexe si et seulement si le système dynamique $(\Omega_{\overline{m}}, S)$ associé à $\overline{\sigma}$ est topologiquement mélangeant.

Si $\overline{\sigma}$ est récurrente positive : *le système dynamique $(\Omega_{\overline{m}}, S)$ admet une mesure de probabilité invariante naturelle. Si de plus, $\overline{\sigma}$ est de longueur constante, alors cette mesure est ergodique.*

Si $\overline{\sigma}$ est récurrente nulle : *le système dynamique associé $(\Omega_{\overline{m}}, S)$ admet une mesure infinie invariante naturelle.*

Nous renvoyons à [Kit98] et à [Fer06] pour la démonstration de l'ergodicité dans le cas de substitutions récurrentes positives de longueur constante (cette démonstration utilise les tours de Rokhlin).

Lorsque la substitution est transiente, la question de l'existence de mesure invariante reste ouverte.

D'autre part, ces résultats sont aussi vrais pour tout *sous-décalage de type fini* de $E^{\mathbb{N}}$, c'est-à-dire pour toute partie $\mathcal{S}$ de $E^{\mathbb{N}}$ fermée par l'action de décalage définie à l'aide d'une matrice $M = (M_{(e,e')})_{(e,e')\in E^2}$ à coefficients dans $\{0,1\}$, par $\mathcal{S} = \{m \in E^{\mathbb{N}},\ \forall e \in E, M_{(e,e+1)} = 1\}$.

Remarques III.6.7. La question de l'entropie des systèmes dynamiques associés à des mots infinis à valeurs dans un alphabet dénombrable n'est pas aussi simple que dans le cas fini (voir par exemple [Wal82]), notamment car le passage fini/dénombrable fait perdre la propriété de compacité de l'ensemble $E^{\mathbb{N}}$.

En effet, Lorsque E est dénombrable, et donc $E^{\mathbb{N}}$ n'est pas compact, il existe plusieurs manières de définir l'entropie d'un système dynamique associé à un mot infini $\overline{m}$ à valeurs dans E : soit on la définit comme borne supérieure de la famille des entropies de sous-systèmes dynamiques compacts associés à $\overline{m}$, soit on compactifie $\Omega_{\overline{m}}$ en $\Omega'_{\overline{m}}$ et on considère l'entropie du compactifié $(\Omega'_{\overline{m}}, S)$ comme l'entropie du système considéré $(\Omega_{\overline{m}}, S)$.

Ces deux constructions de l'entropie topologique sont en fait équivalentes [Gur69][Gur70] lorsque l'on fixe une métrique compatible avec la topologie. Cependant, elles ne sont pas indépendantes de la métrique choisie, et l'entropie change lorsque l'on change de métrique (voir[Bow73] et [HK95]), et ce, même si deux métriques sont équivalentes. Les mesures pour lesquelles l'entropie est maximale, c'est-à-dire égale à l'entropie du système, sont appelées *mesures maximales.* Nous renvoyons le lecteur à [Kit98] ou [Wal82] pour plus de détails concernant ce sujet.

D'autres notions d'entropie ont été introduites, par exemple, dans [FF97], D. et U. Fiebig définissent l'entropie de transition et l'entropie périodique pour des sous-décalages localement compacts.

Il existe aussi une définition de l'entropie en rapport direct avec la « complexité combinatoire » du mot infini : l'entropie introduite par I. Salama [Sal84] [Sal88] ; elle est définie par

$$H^*(\Omega_{\overline{m}}) = \sup_{e\in E} \limsup_n \frac{\log p_{\overline{m}}(n,e)}{n},$$

où $p_{\overline{m}}(n,e)$ représente le nombre de facteurs de longueur n de $\overline{m}$ dont la première lettre est e. Cette entropie renseigne sur la manière dont croît, avec n, le nombre de cylindres de la famille $\{[w],\ w \in E^n\}$.

Chapitre IV

Complexité des marches aléatoires

IV.1 Contexte

Soient deux entiers d et q supérieurs ou égaux à 1 et tels que $d+1 \leq q$, et une famille $\mathcal{G} = \{e_i, i \in [\![1, q]\!]\}$ de vecteurs de $\mathbb{R}^d$ qui génère un réseau $\mathcal{R}$ de $\mathbb{R}^d$ de rang d et un élément v de $\mathcal{R}$ fixé.

On considère la substitution $\overline{\varsigma}$ de longueur q associée à la marche aléatoire par les vecteurs de $\mathcal{G}$ sur $\mathcal{R}$ par

$$\begin{array}{rccc} \overline{\varsigma}: & \mathcal{R} \cup \{e_0\} & \to & (\mathcal{R} \cup \{e_0\})^q \\ & a = (a_1, \ldots, a_d) & \mapsto & (a + e_1) \ldots (a + e_q) \\ & e_0 & \mapsto & e_0(e_2) \ldots (e_q) \end{array}$$

On s'intéresse dans ce chapitre, à la famille des mots q^∞-automatiques engendrés par les q-ADDA du type $\mathcal{A}_{(d,\mathcal{G},v)} = (\mathcal{R} \cup \{e_0\}, e_0, \phi_{\overline{\varsigma}}, \{0,1\}, \Pi_v)$ où la projection $\Pi_v : \mathcal{R} \cup \{e_0\} \to \{0,1\}$ est définie par $\Pi^{-1}(\{1\}) = \{v\}$.

Par exemple, lorsque la dimension d est égale à 1, que $\mathcal{G} = \{-1, 1\}$ et que $v = (0, \ldots, 0)$, la substitution $\overline{\sigma}$, aussi appelée *substitution de l'ivrogne*, est donnée par

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

l'unique point fixe $\overline{m}$ de $\overline{\sigma}$ est le mot infini suivant :

$$\begin{array}{rl} \overline{m} = & e_0102(-1)113(-2)0020224(-3)(-1)(-1)1(-1)113(-1)1131335(-4)(-2)(-2)0 \\ & (-2)002(-2)0020224(-4)(-2)(-2)0(-2)002(-2)002022402242446\ldots \end{array}$$

Le mot 2^∞-automatique engendré par le 2-ADDA $\mathcal{A}_{(1,\{1\},0)}$, représenté à la figure IV.1, associé à la marche aléatoire en dimension 1, appelé *mot de l'ivrogne* est donc donné par $m = \Phi(\overline{m})$:

$$m = 00100000011010000000000000000000000010110011010000110100010000000\ldots$$

On trouvera dans [FM05] et [Mau06] l'étude de certaines propriétés arithmétiques et statistiques de l'ensemble d'entiers dont m est le mot indicateur.

La figure IV.2 représente le graphe du 4-ADDA qui génère le mot associé à la marche aléatoire sur $\mathbb{Z}^2$, pour la famille génératrice formée des vecteurs $\{(0,1), (0,-1), (1,0)(-1,0)\}$.

On peut aussi considérer des mots associés à des marches moins régulières, comme par exemple, dans $\mathbb{C} \approxeq \mathbb{R}^2$, sur le réseau triangulaire $\mathcal{R}$ engendré par la famille $\mathcal{G} = \{1, j, j^2\}$, où $j = e^{\frac{2i\pi}{3}}$. La substitution $\overline{\tau}$ qu'on lui associe est donnée par

$$\begin{array}{rccc} \overline{\tau}: & \mathcal{R} \cup \{e_0\} & \to & \mathcal{R}^2 \cup \{e_0(j)(j^2)\} \\ & e & \mapsto & (e+1)(e+j)(e+j^2) \\ & e_0 & \mapsto & e_0(j)(j^2) \end{array}$$

l'unique point fixe $\overline{m}$ de $\overline{\tau}$ est le mot infini suivant :

$$\begin{array}{rl} \overline{m} = & e_0(j)(j^2)(j+1)(2j)(j+j^2)(j^2+1)(j^2+j)(2j^2)(j+2)(2j+1)(0)(2j+1)(3j)(2j+j^2)(0)\ldots \\ & (2j+j^2)(j+2j^2)(j^2+2)(0)(1+2j^2)(0)(j^2+2j)(j+2j^2)(2j^2+1)(2j^2+j)(3j^2)(j+3)\ldots \end{array}$$

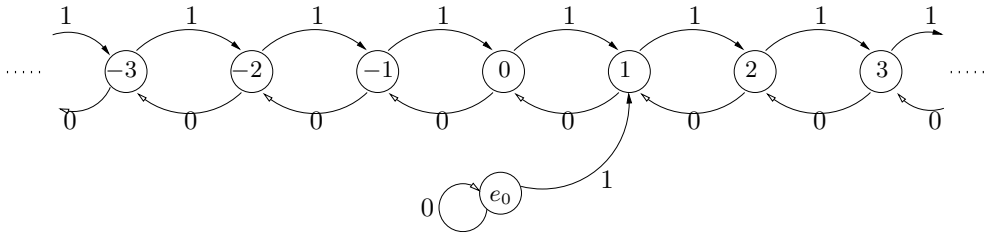


FIG. IV.1 – Graphe du 2-ADDA qui engendre le mot de l'ivrogne.

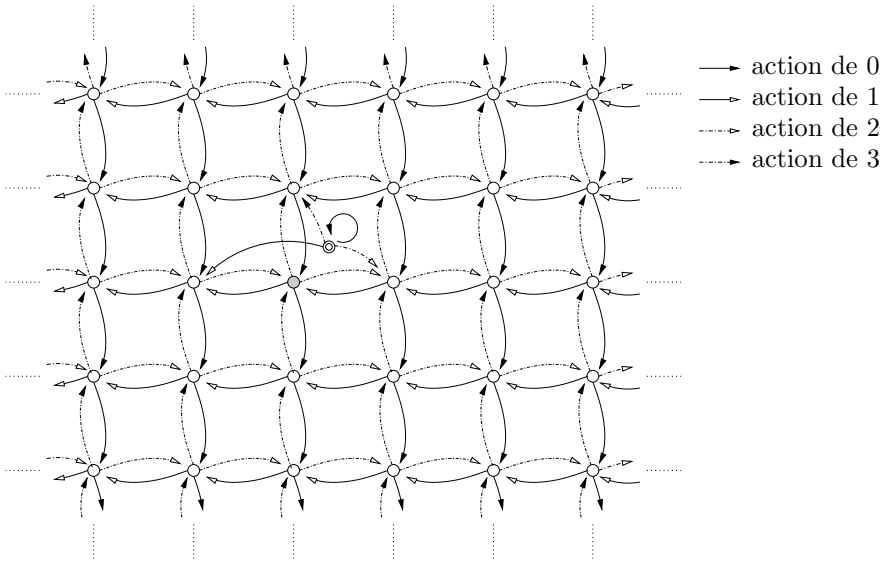


FIG. IV.2 – Graphe du 4-ADDA qui engendre le mot de la marche aléatoire en dimension 2.

Le mot 3^∞-automatique engendré par le 2-ADDA $\mathcal{A} = (\mathcal{R} \cup \{e_0\}, e_0, \phi_{\overline{\tau}}, \{0,1\}, \Pi_0)$, représenté à la figure IV.3, est le mot indicateur de l'ensemble des entiers dont l'écriture en base 3 contient autant de 0 que de 1 et de 2. Cet ensemble apparait dans [Mau06].

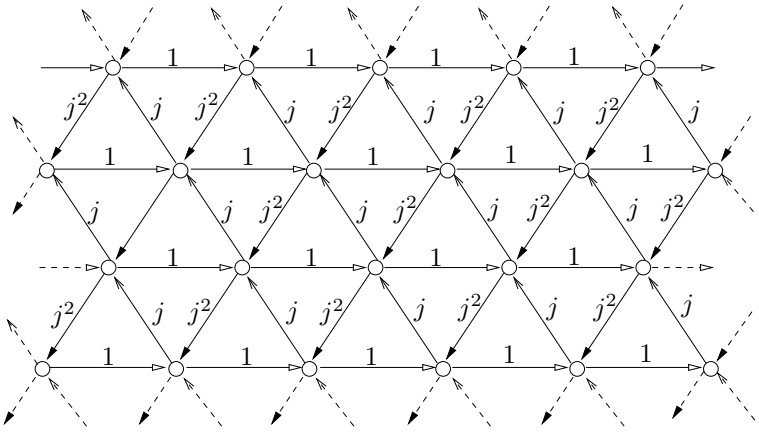


FIG. IV.3 – Graphe du 3-ADDA associé à la marche aléatoire sur le réseau triangulaire.

Nous donnons dans ce chapitre une majoration de la complexité des mots engendrés par les q-ADDA du type $\mathcal{A}_{(d,\mathcal{G},v)}$ en dimension quelconque ainsi qu'un équivalent de la complexité de m en dimension 1.

IV.2 Majoration de la complexité en dimension quelconque

Théorème IV.2.1. *Soient d un entier supérieur ou égal à 1, $\mathcal{G} = \{e_i, i \in [\![1, q]\!]\}$ une partie génératrice d'un réseau $\mathcal{R}$ de $\mathbb{R}^d$ de rang d et un élément v de $\mathcal{R}$ fixé.*

Si m est le mot q^∞-automatique engendré par le q-ADDA $\mathcal{A}_{(d,\mathcal{G},v)}$, on a

$$p_m(n) = O(n \log_q^{q+1} n).$$

Nous allons démontrer ce théorème dans un cas simple, puis donner les indications qui permettent de généraliser la preuve pour tous ces mots, car la preuve générale ne diffère que très peu de la preuve dans le cas simple suivant et introduit une complexité importante au niveau des notations.

On va donc considérer, dans un premier temps, uniquement les mots engendrés par les $2d$-ADDA du type $\mathcal{A}_{(d,\mathcal{G},v)} = (\mathcal{R} \cup \{e_0\}, e_0, \phi_{\overline{\tau}}, \{0,1\}, \Pi_v)$, lorsque la famille $\mathcal{G}$ est formée d'une base du réseau et de ses inverses ; on parle alors de *marche aléatoire isotrope* :

Soient d un entier supérieur ou égal à 1, $\mathcal{B} = \{e_i, i \in [\![1, d]\!]\}$ une base d'un réseau $\mathcal{R}$ de $\mathbb{R}^d$ de rang d et un élément v de $\mathcal{R}$ fixé.

Un élément $x = \sum_{i=1}^{d} x_i e_i$ de $\mathcal{R}$, où les x_i sont dans Z, est noté $x = (x_1, x_2, \ldots, x_d)$.

On considère la substitution $\overline{\sigma}$ de longueur $2d$ associée à la marche aléatoire isotrope sur $\mathcal{R}$ par :

$$\begin{array}{rrcl} \overline{\sigma} : & \mathcal{R} \cup \{e_0\} & \to & (\mathcal{R} \cup \{e_0\})^{2d} \\ & x = (x_1, \ldots, x_d) & \mapsto & (x - e_d) \ldots (x - e_i) \ldots (x - e_1)(x + e_1) \ldots (x + e_i) \ldots (x + e_d) \\ & e_0 & \mapsto & e_0(-e_{d-1}) \ldots (-e_i) \ldots (-e_1)(e_1) \ldots (e_i) \ldots (e_d) \end{array}$$

On considère la famille des mots $(2d)^\infty$-automatiques engendrés par la famille des $2d$-ADDA du type $\mathcal{A}_{(d,\mathcal{B},v)} = (\mathcal{R} \cup \{e_0\}, e_0, \phi_{\overline{\sigma}}, \{0,1\}, \Pi_v)$ où la projection $\Pi_v : \mathcal{R} \cup \{e_0\} \to \{0,1\}$ est définie par $\Pi^{-1}(\{1\}) = v$.

Ces mots étant définis comme les images par les projections admissibles Π_v de l'unique point fixe $\overline{m}$ de $\overline{\sigma}$ contenu dans $e_0\mathcal{R}^{\mathbb{N}}$.

Notation IV.2.2. Si $\overline{\sigma}$ est une substitution sur un alphabet dénombrable E et Π est une projection admissible de E vers un alphabet fini A, on note, pour tout entier $k \geq 0$, $\sigma^k = \Pi \circ \overline{\sigma}^k$.

Théorème IV.2.3. *Soient d un entier supérieur ou égal à 1, $\mathcal{B} = \{e_i, i \in [\![1, d]\!]\}$ une base de $\mathcal{R}$ et un élément v de $\mathcal{R}$ fixé.*

Si m est le mot $(2d)^\infty$-automatique engendré par le $2d$-ADDA $\mathcal{A}_{(d,\mathcal{B},v)}$, on a

$$p_m(n) = O(n(\log_{2d} n)^{d+1}).$$

Remarque IV.2.4. On notera au passage que dans le cadre des marches aléatoires isotropes, le théorème IV.2.1 est amélioré puisqu'on aurait attendu un résultat du type $p_m(n) = O(n(\log_{2d} n)^{2d+1})$. On a donc gagné un facteur $(\log_{2d} n)^d$. Cela est dû au fait que les relations de dépendance entre les vecteurs de $\mathcal{B}$ et leurs inverses sont connues. Dans le cadre général, où on considère une famille génératrice quelconque, ces relations de dépendances ne peuvent pas être exploitées. En revanche on peut, sur des cas particuliers, s'en servir pour faire baisser la fonction qui majore l'ordre de la complexité.

La démonstration de ce théorème consiste à montrer qu'il existe un polynôme P_d, de degré $(d+1)$ tel que :

$$\forall n \geq 1, \quad p_m(n) \leq 2dnP_d(\log_{2d} n).$$

La preuve se construit à partir de l'observation suivante :

Puisque, pout tout entier k $\overline{\sigma}^k(\overline{m}) = \overline{m}$, on peut écrire m sous la forme $m = \sigma^k(\overline{m}_0)\sigma^k(\overline{m}_1)\sigma^k(\overline{m}_2)\ldots$, cela implique que tout facteur de longueur $(2d)^k$ est contenu dans un mot de longueur $(2d)^{k+1}$ du type $\sigma^k(xy)$.

La première étape consiste donc à déterminer quels sont tous les mots de deux lettres de $\overline{m}$. Cela fait l'objet du lemme IV.2.5.

On construit ensuite un polynôme P_d tel que, pour tout $k \geq 1$, la valeur $P_d(k-1)$ majore le nombre de mots de deux lettres xy du point fixe $\overline{m}$ de $\overline{\sigma}$ contenu dans $e_0(\mathcal{R})^{\mathbb{N}}$, pour lesquels les images par Π de $\overline{\sigma}^k(xy)$ contiennent des occurrences de 1.

On peut alors en déduire une majoration du nombre de facteurs de m de longueur $(2d)^k$ par $(2d)^k P_d(k-1)$. La croissance de la fonction de complexité permet alors d'obtenir la majoration pour tout n.

Lemme IV.2.5. *Soit $\overline{m}$ le mot infini point fixe de $\overline{\sigma}$ contenu dans $e_0(\mathcal{R})^{\mathbb{N}}$. Le mot xy est dans $F_2(\overline{m})$ si et seulement si un des 4 cas suivants se produit :*

1. $xy = e_0(-e_{d-1})$,
2. $\exists k' \geq 0$ *et* $\exists i \in [\![2, d]\!]$ *tels que* $x = -e_i + k'e_d$ *et* $y = -e_{i-1} - k'e_d$,
3. $\exists k' \geq 0$ *tel que* $x = e_d + k'e_d$ *et* $y = -e_{d-1} - k'e_d$,
4. $\exists k' \geq 0$ *et* $\exists i \in [\![1, d-1]\!]$ *tels que* $x = e_i + k'e_d$ *et* $y = e_{i+1} - k'e_d$,
5. $\exists k' \geq 0$, $x = (k'+1)e_d$ *et* $y = e_1 - (k'+1)e_d$,
6. $\exists k' \geq 0$ *et* $\exists i \in [\![2, d]\!]$ *tels que* $x \in F_1(\overline{m})$ *et* $y = x + e_i - e_{i-1} - 2k'e_d$,
7. $\exists k' \geq 0$ *tel que* $x \in F_1(\overline{m})$ *et* $y = x - 2k'e_d$,
8. $\exists k' \geq 0$ *et* $\exists i \in [\![1, d-1]\!]$ *tels que* $x \in F_1(\overline{m})$ *et* $y = x + e_{i+1} - e_i - 2k'e_d$.

Ce lemme est essentiellement une réécriture de la proposition III.5.3. Le point III.5.3 2. correspond ici aux points 2, 3 et 4, le point III.5.3 3. correspond ici au point 5 et le point III.5.3 4. se scinde ici en les points 6, 7 et 8.

▪ Démonstration du théorème IV.2.3

Soit $k \geq 2$ un entier fixé. On pose $k = 2q + p$, avec $p \in \{0, 1\}$. On va majorer la complexité au rang $(2d)^k$.

Tout mot de m de longueur $(2d)^k$ est inclus dans un mot de longueur $(2d)^{k+1}$ du type $\sigma^k(x)\sigma^k(y)$. On donc va majorer le nombre de mots de deux lettres xy tels que $\sigma^k(x)\sigma^k(y)$ est différent de $0^{4^{k+1}}$. Pour cela, on va distinguer deux cas :

(a) un seul des deux mots $\sigma^k(x)$ et $\sigma^k(y)$ contient des occurences de 1,

(b) les deux mots $\sigma^k(x)$ et $\sigma^k(y)$ contiennent des occurences de 1.

Dans le cas (b), pour majorer le nombre de mots xy, on distinguera ensuite six sous-cas, en accord avec le lemme IV.2.5.

On commence par remarquer que le mot $\overline{\sigma}^k(x)$, pour un élément $x = (x_1, \ldots, x_n)$ de $\mathcal{R}$, contient des occurrences de 0 si et seulement s'il s'écrit $x = v + \sum_{i=1}^{d} a_i e_i$, où les coefficients a_i sont des éléments de $\mathbb{Z}$ vérifiant $\sum_{i=1}^{d} |a_i| \leq k$ et tels que $\sum_{i=1}^{d} a_i$ soit de même parité que k. On a donc

$$\mathcal{Z}_k(v) = \left\{ v + \sum_{i=1}^{d} a_i e_i \in \mathcal{R},\ \sum_{i=1}^{d} |a_i| \leq k, \sum_{i=1}^{d} a_i = k \mod 2 \right\}$$

et si on note $N_d(k) = Card(\mathcal{Z}_k(v))$, alors, la famille $N_d(k)$ est indépendante de l'élément v choisi dans $\mathbb{Z}$ et est donnée par les formules suivantes :

$$\forall d \in \mathbb{N}^*,\ N_d(0) = 1 \text{ et } N_d(1) = 2d,$$

$$\forall k \in \mathbb{N},\ N_1(k) = k + 1,$$

$$\forall d \in \mathbb{N}^*, \forall k \in \mathbb{N},\ N_d(k) = N_{d-1}(k) + 2\sum_{i=0}^{k-1} N_{d-1}(i).$$

On a, par exemple, $N_2(k) = (k+1)^2$ pour tout k, et plus, généralement, on montre par récurrence que N_d est un polynôme de degré d en la variable k, équivalent à $\frac{2^{d-1}}{d!}k^d$, lorsque k tend vers l'infini.

On peut donc maintenant majorer le nombre de mots xy en distinguant les deux cas annoncés :
(a) Majoration du nombre de mots xy tels que seulement un des mots $\sigma^k(x)$ ou $\sigma^k(y)$ contient des occurrences de 1 :

Comme il existe $N_d(k)$ éléments différents dans $\mathcal{Z}_k(v)$, on peut donc construire de cette manière, au pire, $2N_d(k)$ mots $\sigma^k(x)\sigma^k(y)$ différents entre eux et différents de $0^{(2d)^{k+1}}$.

(b) Majoration du nombre de mots xy tels que les deux mots $\sigma^k(x)$ et $\sigma^k(y)$ contiennent des occurrences de 1 :

Le lemme IV.2.5 permet de classer les mots de deux lettres xy de $\overline{m}$ de la manière suivante :

1. $xy = e_0(-e_{d-1})$,
2. $\exists k' \geq 0$ et $\exists i \in [\![2, d]\!]$ tels que $x = -e_i + k'e_d$ et $y = -e_{i-1} - k'e_d$,
3. $\exists k' \geq 0$ tel que $x = e_d + k'e_d$ et $y = -e_{d-1} - k'e_d$,
4. $\exists k' \geq 0$ et $\exists i \in [\![1, d-1]\!]$ tels que $x = e_i + k'e_d$ et $y = e_{i+1} - k'e_d$,
5. $\exists k' \geq 0$, $x = (k'+1)e_d$ et $y = e_1 - (k'+1)e_d$,
6. $\exists k' \geq 0$ et $\exists i \in [\![2, d]\!]$ tels que $x \in F_1(\overline{m})$ et $y = x + e_i - e_{i-1} - 2k'e_d$,
7. $\exists k' \geq 0$ tel que $x \in F_1(\overline{m})$ et $y = x - 2k'e_d$,
8. $\exists k' \geq 0$ et $\exists i \in [\![1, d-1]\!]$ tels que $x \in F_1(\overline{m})$ et $y = x + e_{i+1} - e_i - 2k'e_d$.

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

1. Si $xy = e_0(-e_{d-1})$, les deux itérés $\sigma^k(x)$ et $\sigma^k(y)$ ne peuvent pas contenir simultanément des occurences de 1 pour k fixé. En effet, $\sigma^k(e_0)$ contient des occurences de 1 uniquement si $-\sum_{i=1}^d v_i$ et k sont de même parité et $\sigma^k((-e_1))$ contient des occurences de 1 uniquement lorsque $-1-\sum_{i=1}^d v_i$ et k sont de même parité. Par conséquent, le mot $xy = e_0(-e_1)$ ne donne pas d'itéré $\sigma^k(x)\sigma^k(y)$ différent de $0^{4^{k+1}}$ pour lequel $\sigma^k(x)$ et $\sigma^k(y)$ contiennent des occurences de 1.

2. S'il existe $k' \geq 0$ et $i \in [\![2, d]\!]$ tels que $x = -e_i + k'e_d$ et $y = -e_{i-1} - k'e_d$, les deux itérés $\sigma^k(x)$ et $\sigma^k(y)$ contiennent des occurences de 1 uniquement lorsque

$$|-1-v_i| + |k'-v_d| + \sum_{\substack{j=1\\ j\neq i,d}}^{d} |v_j| \leq k \quad \text{et} \quad |1+v_{i-1}| + |k'+v_d| + \sum_{\substack{j=1\\ j\neq i-1,d}}^{d} |v_j| \leq k,$$

et que $1+k'$ et $k-\sum_{i=1}^d v_i$ sont de même parité. On peut donc former de cette manière, au pire, $(d-1)\frac{k+1}{2}$ mots itérés $\sigma^k(x)\sigma^k(y)$ différents et différents de $0^{4^{k+1}}$ pour lesquels $\sigma^k(x)$ et $\sigma^k(y)$ contiennent des occurences de 1 (lorsque k est suffisamment grand pour qu'il existe de tels couples).

3. S'il existe $k' \geq 0$ tel que $x = e_d + k'e_d$ et $y = -e_{d-1} - k'e_d$, les deux itérés $\sigma^k(x)$ et $\sigma^k(y)$ contiennent des occurences de 1 uniquement si

$$+|(k'+1)-v_d| + \sum_{\substack{j=1\\ j\neq d}}^{d} |v_j| \leq k \quad \text{et} \quad |1+v_d| + |k'+v_d| + \sum_{\substack{j=1\\ j\neq d-1,d}}^{d} |v_j| \leq k,$$

et que $1+k'$ et $k-\sum_{i=1}^d v_i$ sont de même parité.

On peut donc former de cette manière, au pire, $\frac{k+1}{2}$ mots itérés $\sigma^k(x)\sigma^k(y)$ différents et différents de $0^{4^{k+1}}$ pour lesquels $\sigma^k(x)$ et $\sigma^k(y)$ contiennent des occurences de 1 (lorsque k est suffisamment grand pour qu'il existe de tels couples).

4. S'il existe $k' \geq 0$ et $i \in [\![1, d-1]\!]$ tels que $x = e_i + k'e_d$ et $y = e_{i+1} - k'e_d$, les deux itérés $\sigma^k(x)$ et $\sigma^k(y)$ contiennent des occurences de 1 uniquement si

$$|1-v_i| + |k'-v_d| + \sum_{\substack{j=1\\ j\neq i,d}}^{d} |v_j| \leq k \quad \text{et} \quad |1-v_{i-1}| + |k'+v_d| + \sum_{\substack{j=1\\ j\neq i-1,d}}^{d} |v_j| \leq k,$$

et que $1+k'$ et $k-\sum_{i=1}^d v_i$ sont de même parité. On peut donc former de cette manière, au pire, $(d-1)\frac{k+1}{2}$ mots itérés $\sigma^k(x)\sigma^k(y)$ différents et différents de $0^{4^{k+1}}$ pour lesquels $\sigma^k(x)$ et $\sigma^k(y)$ contiennent des occurences de 1 (lorsque k est suffisamment grand pour qu'il existe de tels couples).

5. S'il existe $k' \geq 0$ tel que $x = (k'+1)e_d$ et $y = e_1 - (k'+1)e_d$, les deux itérés $\sigma^k(x)$ et $\sigma^k(y)$ ne peuvent pas contenir simultanément des occurences de 1 pour k fixé. En effet, $\sigma^k((k'+1)e_d)$ contient des

occurences de 1 uniquement lorsque $(k'+1)-\sum_{i=1}^{d} v_i$ et k sont de même parité et $\sigma^k((e_1-(k'+1)e_d)$ contient des occurences de 1 uniquement lorsque $k'-\sum_{i=1}^{d} v_i$ et k sont de même parité. Par conséquent, les mots xy de ce type ne donne pas d'itéré $\sigma^k(x)\sigma^k(y)$ différent de $0^{4^{k+1}}$ pour lequel $\sigma^k(x)$ et $\sigma^k(y)$ contiennent des occurences de 1.

6. Supposons que l'on puisse écrire le mot xy de manière à ce que, pour un certain $k' \geq 0$ et un certain $i \in [\![2,d]\!]$, $y = x + e_i - e_{i-1} - 2k'e_d$. On note $x = v + x'$ et $y = v + y'$. Puisque x et y doivent être dans $\mathcal{Z}_k(v)$ pour que les itérés $\sigma^k(x)$ et $\sigma^k(y)$ contiennent des occurences de 1, alors x' et y' doivent être dans $\mathcal{Z}_k((0,\dots,0))$ pour que les itérés $\sigma^k(x)$ et $\sigma^k(y)$ contiennent des occurences de 1. On va donc majorer le nombre de mots xy de ce type par un majoration du cardinal de l'ensemble :

$$\mathcal{U}_i(k) = \left\{(x',k') \in \mathcal{R}\times\mathbb{N},\ (x', x'+e_i-e_{i-1}-2k'e_d)) \in \mathcal{Z}_k((0,\dots,0))^2\right\}.$$

Pour x' fixé dans $\mathcal{Z}_k((0,\dots,0)$, les éléments du type $y' = x'+e_i-e_{i-1}-2k'e_d$ sont des points entiers de la demi-droite de pente $-e_d$ issue de $x+e_i-e_{i-1}$. Or toutes les droites parallèles l'axe e_d contiennent au plus $N_1(k)$ éléments de $\mathcal{Z}_k((0,\dots,0)$, donc pour chaque point x' de $\mathcal{Z}_k((0,\dots,0)$, on peut majorer le nombre d'éléments $y' = x' + e_i - e_{i-1} - 2k'e_d$ par $N_1(k)$ et donc

$$Card(\mathcal{U}_i(k)) \leq (k+1)N_d(k).$$

On peut donc former de cette manière, moins de $(d-1)(k+1)N_d(k)$ mots $\sigma^k(x)\sigma^k(y)$ différents et différents de $0^{4^{k+1}}$ pour lesquels $\sigma^k(x)$ et $\sigma^k(y)$ contiennent des occurences de 1.

7. Supposons que l'on puisse écrire le mot xy de manière à ce que, pour un certain $k' \geq 0$, $y = x-2k'e_d$. On note $x = v + x'$ et $y = v + y'$. Puisque x et y doivent être dans $\mathcal{Z}_k(v)$ pour que les itérés $\sigma^k(x)$ et $\sigma^k(y)$ contiennent des occurences de 1, alors x' et y' doivent être dans $\mathcal{Z}_k((0,\dots,0))$ pour que les itérés $\sigma^k(x)$ et $\sigma^k(y)$ contiennent des occurences de 1. On va donc majorer le nombre de mots xy de ce type par un majoration du cardinal de l'ensemble :

$$\mathcal{V}(k) = \left\{(x',k') \in \mathcal{R}\times\mathbb{N},\ (x', x'-2k'e_d)) \in \mathcal{Z}_k((0,\dots,0))^2\right\}.$$

Pour x' fixé dans $\mathcal{Z}_k((0,\dots,0))$, les éléments du type $y' = x' - 2k'e_d$ sont des points entiers de la demi-droite de pente $-e_d$ issue de x. Or toutes les droites parallèles l'axe e_d contiennent au plus $N_1(k)$ éléments de $\mathcal{Z}_k((0,\dots,0)$, donc, pour chaque point x' de $\mathcal{Z}_k((0,\dots,0)$, on peut majorer le nombre d'éléments $y' = x' - 2k'e_d$ par $N_1(k)$ et donc

$$Card(\mathcal{V}(k)) \leq (k+1)N_d(k).$$

On peut donc former de cette manière, moins de $(k+1)N_d(k)$ mots $\sigma^k(x)\sigma^k(y)$ différents et différents de $0^{4^{k+1}}$ pour lesquels $\sigma^k(x)$ et $\sigma^k(y)$ contiennent des occurences de 1.

8. Il reste à s'intéresser aux couples xy qui sont dans $\mathcal{Z}_k(v)^2$ et tels que $x \in \mathcal{R}$ et $y = x+e_{i+1}-e_i-2k'e_d$ pour un certain $k' \geq 0$ et un certain $i \in [\![1,d-1]\!]$. On note $x = v + x'$ et $y = v + y'$. Puisque x et y doivent être dans $\mathcal{Z}_k(v)$ pour que les itérés $\sigma^k(x)$ et $\sigma^k(y)$ contiennent des occurences de 1, alors x' et y' doivent être dans $\mathcal{Z}_k((0,\dots,0))$ pour que les itérés $\sigma^k(x)$ et $\sigma^k(y)$ contiennent des occurences de 1. On va donc majorer le nombre de mots xy de ce type par un majoration du cardinal de l'ensemble :

$$\mathcal{W}_i(k) = \left\{(x',k') \in \mathcal{R}\times\mathbb{N},\ (x', x'+e_{i+1}-e_i-2k'e_d)) \in \mathcal{Z}_k((0,0)^2\right\}.$$

Pour x' fixé dans $\mathcal{Z}_k((0,\dots,0)$, les éléments du type $y' = x' - 2k'e_d$ sont des points entiers de la demi-droite de pente $-e_d$ issue de $x+e_{i+1}-e_i$. Or toutes les droites parallèles l'axe e_d contiennent au plus $N_1(k)$ éléments de $\mathcal{Z}_k((0,\dots,0)$, donc, pour chaque point x' de $\mathcal{Z}_k((0,\dots,0)$, on peut majorer le nombre d'éléments $y' = x' + e_{i+1} - e_i - 2k'e_d$ par $N_1(k)$, donc

$$Card(\mathcal{W}_i(k)) \leq (k+1)N_d(k).$$

On peut donc former de cette manière, moins de $(d-1)(k+1)N_d(k)$ mots $\sigma^k(x)\sigma^k(y)$ différents et différents de $0^{4^{k+1}}$ pour lesquels $\sigma^k(x)$ et $\sigma^k(y)$ contiennent des occurences de 1.
Tout mot de m de longueur $(2d)^k$ etant inclus dans un mot de longueur $(2d)^{k+1}$ du type $\sigma^k(x)\sigma^k(y)$, il résulte des différentes majorations obtenues en (a) et (b) que :

$$\forall k \geq 1,\ p_m((2d)^k) \leq (2d)^k\left(2N_d(k) + \frac{(2d-1)(k+1)}{2}\left(2N_d(k)+1\right)\right).$$

Soit le polynôme P_d de degré $(d+1)$ défini par $P_d(X) = 4dN_d(X+1)+d(2d-1)(X+2)\,(2N_d(X+1)+1)$. On a alors $p_m((2d)^{k+1}) \leq (2d)^k P_d(k)$. En utilisant le fait que p_m est une fonction croissante et que, pour tout entier n de $[\![(2d)^k, (2d)^{k+1}[\![$, $\log_{2d} n$ appartient à $[\![k, k+1[\![$, on obtient :

$$\forall n \geq 1, \quad p_m(n) \leq nP_d(\log_{2d} n).$$

Donc on a construit un polynôme P_d de degré $(d+1)$ tel que $nP_d(\log_{2d} n)$ majore $p_m(n)$. ∎

Le théorème IV.2.1 se montre de la même manière que le théoreme IV.2.3. Dans ce cadre là, le lemme IV.2.5 devient :

Le mot xy est dans $F_2(\overline{m})$ si et seulement si un des 4 cas suivants se produit :

1. $xy = e_0(e_2)$,
2. $\exists k' \geq 0$ et $\exists i \in [\![1, q-1]\!]$ tels que $x = e_i + k'e_d$ et $y = e_{i+1} + k'e_1$,
3. $\exists k' \geq 0$, $x = (k'+1)e_q$ et $y = e_2 - (k'+1)e_q$,
4. $\exists k' \geq 0$ et $\exists i \in [\![1, q-1]\!]$ tels que $x \in F_1(\overline{m})$ et $y = x + e_{i+1} - e_i - k'(e_q - e_1)$.

La difficulté supplémentaire qu'implique cette généralisation est au niveau du calcul du cardinal de $\mathcal{Z}_k(v)$ puisque on a :

$$\mathcal{Z}_k(v) = \left\{ x = a + v \in \mathcal{R},\ \inf_{\sum_1^q a_i e_i = a} \left(\sum_{i=1}^{d} |a_i| \right) \leq k, \sum_{i=1}^{d} a_i = k \mod 2 \right\}.$$

Ce que l'on peut traduire par l'égalité suivante :

$$\mathcal{Z}_k(v) = \bigcup_{i=1}^{q} \bigcup_{\{e_{j_1}, \ldots, e_{j_i}\} \subset \mathcal{G}} \left\{ x = v + \sum_{j=1}^{i} a_j e_{i_j} \in \mathcal{R},\ \left(\sum_{j=1}^{i} |a_j| \right) \leq k, \sum_{j=1}^{i} a_j = k \mod 2 \right\}.$$

Ainsi, on a la majoration suivante :

$$Card(\mathcal{Z}_k(v) \leq \sum_{i=1}^{q} \binom{q}{i} Card \left(\left\{ (a_1, \ldots, a_i) \in \mathbb{Z}^i,\ \sum_{j=1}^{i} |a_j| \leq k \right\} \right).$$

Or, on a déjà vu que, pour tout entier i, $Card\left(\left\{(a_1, \ldots, a_i) \in Z^i,\ \sum_{j=1}^{i} |a_j| \leq k\right\}\right)$ est un polynôme de degré i en la variable k (cela se montre par récurrence sur i), donc $Card(\mathcal{Z}_k(v))$ est inférieur à un polynôme de degré q en la variable k. En utilisant le même raisonnement que celui de la preuve de IV.2.3, on obtient alors : $p_m(n) = O(n \log_q^{q+1} n)$.

Remarque IV.2.6. Lorsqu'on effectue la majoration de $Card(\mathcal{Z}_k(v))$ sur un exemple précis, on peut alors exploiter les relations de dépendance entre les éléments de $\mathcal{G}$ afin d'obtenir une majoration moins barbare. C'est d'ailleurs ce qui se produit dans la preuve du théorème IV.2.3, et qui permet d'obtenir une meilleure majoration de la complexité.

IV.3 Cas de la marche aléatoire isotrope sur $\mathbb{Z}$

Soit un élément $v \in \mathbb{Z}$ et $\overline{\sigma}$ la substitution définie sur $\mathbb{Z} \cup \{e_0\}$ par :

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

On considère ici, le mot infini m engendré par le 2-ADDA $\mathcal{A}_v = (\mathbb{Z} \cup \{e_0\}, \phi_{\overline{\sigma}}, e_0, \{0,1\}, \Pi)$ où la projection $\Pi_0 : \mathbb{Z} \cup \{e_0\} \to \{0,1\}$ est définie par $\Pi^{-1}(\{1\}) = \{v\}$ dont le graphe est représenté à la figure IV.1.

Théorème IV.3.1. *La fonction de complexité de m est équivalente à $n \log_2^2 n$.*

Pour montrer ce théorème, on cherche en fait une majoration puis une minoration de la fonction p_m qui soient toutes les deux des fonctions équivalentes à $n\log_2^2 n$. La majoration de la complexité que nous devons effectuer doit être plus précise que celle que nous obtenons grâce au théorème IV.2.3, puisqu'il fournit une majoration de $p_m(n)$ par $n(\log_2 n+2)(2\log_2 n+9)$. On va montrer les deux résultats suivants :

Proposition IV.3.2. *La fonction de complexité de m admet la majoration suivante :*

$$\forall n \geq 1, \quad p_m(n) \leq n(\log_2 n + 9)(\log_2 n + 2).$$

Proposition IV.3.3. *La fonction de complexité de m admet la minoration suivante :*

$$\forall n \in \mathbb{N}, n \geq 3, \ p_m(n) \geq n\left((\log_2 n)^2 - 2\right) + 2.$$

Nous démontrons ces deux propositions l'une après l'autre. La majoration de la proposition IV.3.2 découle d'une majoration fine du nombre de facteurs de longueur 2 qui apparaissent dans le mot de l'ivrogne.

La minoration de la proposition IV.3.3 repose sur le comptage de facteurs spéciaux. Pour cela, on étudiera en détails la structure de certains itérés $\overline{\sigma}^k(e)$.

La proposition III.5.3 devient, dans ce cas simple :

Lemme IV.3.4. *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 :

$$F_2(\overline{m}) = \{e_0 1\} \cup \{x_1x_2, \ x_1 \in \mathbb{Z}, x_2 = x_1 - 2p, \ p \geq -1 \text{ ou } x_2 = -x_1 + 1 \text{ si } e_1 > 0\}.$$

Remarques IV.3.5. Quelques remarques préliminaires :

Soit un élément $x = v + a$ de $\mathbb{Z}$. Une infinité d'itérés $\overline{\sigma}^k(x)$ se projettent par Π_v sur 0^{2^k}. En effet, les lettres de $\overline{\sigma}^k(x)$ sont dans $[\![x-k, x+k[\![$. Plus précisemment :

1. Si a et k sont de même parité, toutes les valeurs entières, de même parité que celle de v, de $[\![x-k, x+k]\!]$, et uniquement elles, apparaîssent dans $\overline{\sigma}^k(x)$.
2. Si a et k sont de parité différentes, toutes les valeurs entières de parité différente de celle de v de $[\![x-k, x+k]\!]$, et uniquement elles, apparaîssent dans $\overline{\sigma}^k(x)$.

Ce résultat se montre par récurrence sur k et permet d'affirmer que $\sigma^k(x) \neq 0^{2^k}$ si et seulement si v apparaît dans $\overline{\sigma}^k(x)$ et donc si et seulement s'il existe un entier $p \in [\![0, k]\!]$ tel que $x = v + k - 2p$.

Si w est un mot de m de longueur 2^k, ce mot est inclus dans un certain $\sigma^k(x_1)\sigma^k(x_2)$. Cependant, il peut exister une infinité de paires (x_1', x_2') telles que $\sigma^k(x_1')\sigma^k(x_2')$ contiennent w, notamment celles dont les projections par Π_v des itérés sont les mêmes. Mais en réalité, on peut trouver un ensemble fini de mots x_1x_2 tels que tout facteur de longueur 2^k soit facteur d'un des $\sigma^k(x_1)\sigma^k(x_2)$, comme nous l'avons fait au paragraphe précédent. plus précisemment, on a, dans ce cas :

Lemme IV.3.6.
On note, pour $k \geq 1$:

$$\mathcal{U}_k = \{e_0, 1\} \cup \Big\{(v + k - 2q, v + k - 2p), \ (q, p) \in [\![-1, k]\!] \times [\![0, k+1]\!] \backslash \{(-1, k+1)\}, \ p \geq q - 1\Big\}.$$

Pour tout mot w de m de longueur 2^k, il existe un couple (x_1, x_2) de $\mathcal{U}_k$ tel que w est un sous-mot de $\sigma^k(x_1)\sigma^k(x_2)$.

■ Démonstration du lemme IV.3.6

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

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) de $\mathbb{Z}^2$ pour laquelle w est un sous-mot de $\sigma^k(e_1)\sigma^k(e_2)$.

Comme w est différent de 0^{2^k}, $\sigma^k(e_1)$ ou $\sigma^k(e_2)$ contient des occurences de 1 et donc la lettre e_1 ou la lettre e_2 appartient à $[\![v-k, v+k]\!]$ et peut s'écrire $v + k - 2q$ avec q dans $[\![0, k]\!]$.

Supposons que $e_1 = v+k-2q$ avec q dans $[\![0, k]\!]$. D'après le lemme IV.3.4, on a forcement $e_2 = e_1 - 2p$ avec $p \geq -1$ ou $e_2 = -e_1 + 1$ si $e_2 \leq 0$. Selon le cas, on obtient des paires de $\mathcal{U}_k$ différentes :

– si $e_2 = e_1 - 2p$ et e_2 est dans $[\![v-k, v+k]\!]$, 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 $[\![v-k, v+k]\!]$, alors $\sigma^k(e_2) = 0^{2^k}$ et donc $\sigma^k(e_2) = \sigma^k(v-k-2)$.
Le lemme IV.3.4 nous assure que $e_1(v-k-2)$ est bien un mot de $F_2(\overline{m})$ qui vérifie : w est inclus dans $\sigma^k(e_1)\sigma^k(v-k-2)$ et $(e_1, v-k-2)$ appartient à $\mathcal{U}_k$.

Supposons maintenant que $e_2 = v+k-2q$ avec q dans $[\![0,k]\!]$. D'après le lemme IV.3.4, deux cas de figure sont possibles : $e_1 = e_2 + 2p$ avec $p \geq -1$ 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 $[\![v-k, v+k]\!]$, 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 $[\![v-k, v+k]\!]$, on a $\sigma^k(e_1) = 0^{2^k} = \sigma^k(v+k+2)$. Le lemme IV.3.4 nous assure que $(v+k+2)e_2$ 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 $(v+k+2, e_2)$ 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 de la proposition IV.3.2

D'après le lemme IV.3.6, tout mot de m de longueur 2^k est inclus dans un certain mot $\sigma^k(x_1)\sigma^k(x_2)$ avec (x_1, x_2) dans l'ensemble $\mathcal{U}_k$. Cela nous donne alors l'inégalité : $p_m(2^k) \leq 2^k \cdot Card(\mathcal{U}_k)$ où 2^k représente le nombre de places différentes où on est susceptible de trouver un mot nouveau de longueur 2^k dans les $\sigma^k(x_1)\sigma^k(x_2)$ et $Card(\mathcal{U}_k)$ représente le cardinal de $\mathcal{U}_k$. Comme

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

on obtient :

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

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 :

$$\forall n \geq 1, \quad p_m(n) \leq n(\log_2 n + 9)(\log_2 n + 2),$$

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

La minoration de la complexité de m nécessite d'explorer d'avantage la structure des $\sigma^k(v+k-2p)$ pour k fixé et $p \in [\![0,k]\!]$, c'est à dire pour les mots $\sigma^k(x)$ qui sont différents du mot 0^{2^k}. En effet, cela va nous permettre de localiser des facteurs spéciaux de longueur fixée et grâce à ces facteurs, nous pourrons minorer la complexité de m.

Tout d'abord, il y a un moyen très simple de calculer les mots $\sigma^k(v+a)$ pour $a \in [\![-k, k]\!]$. On peut construire une sorte de « triangle de Pascal » (Voir figure IV.4) en utilisant la relation substitutive $\sigma^k(v+a) = \sigma^{k-1}(v+a-1)\sigma^{k-1}(v+a+1)$, sur lequel le p-ième élément de la $k^{\text{ième}}$ ligne est $\sigma^k(v+2p-k)$.

Proposition IV.3.7. *Propriétés des $\sigma^k(v+k-2p)$ pour $p \in [\![0,k]\!]$:*

Pour $k > 0$ et p dans $[\![0,k]\!]$, on note $x = v+k-2p$. On a les propriétés suivantes :

1. *Pour tout entier n, $0 \leq n \leq 2^k - 1$, $(\sigma^k(v-(k-2p)))_n = (\sigma^k(v+k-2p))_{2^k-n-1}$. C'est-à-dire que les mots $\sigma^k(v-(k-2p))$ et $\sigma^k(v+k-2p)$ sont image miroir l'un de l'autre. En particulier, $\sigma^{2k}(v)$ est un palindrome.*
2. *Le nombre d'occurences de 1 dans $\sigma^k(x)$ est exactement $\binom{k}{p}$.*
3. *Si on note $R_k(x,q)$ la place de la $q^{ième}$ occurence de 1 de $\sigma^k(x)$, on a la formule suivante :*

$$\forall p \in [\![0,k]\!],\ \forall q \in [\![1, \tbinom{k}{p}]\!], \quad R_k(x,q) = \begin{cases} R_{k-1}(x-1,q) & \text{si } q \leq \binom{k-1}{p}, \\ 2^{k-1} + R_{k-1}(x+1, q - \binom{k-1}{p}) & \text{sinon.} \end{cases}$$

En particulier, on peut montrer, grâce à cette formule, que :

– $R_k(x,1) = 2^p$,
– $R_k(x,2) = 2^p + 2^{p-1}$,

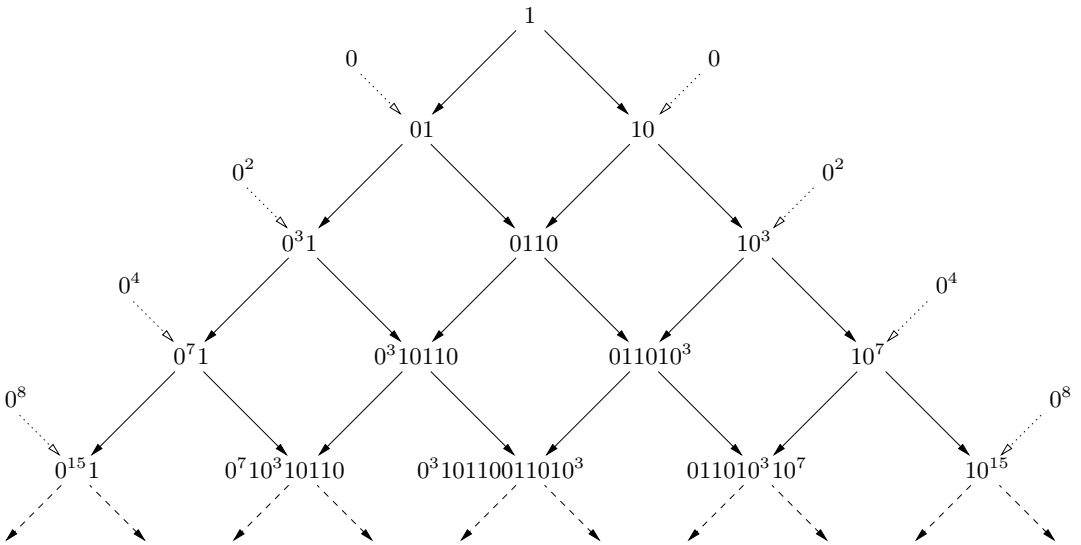


FIG. IV.4 – Processus de fabrication des itérés $\overline{\sigma}^k(v+a) \neq 0^{2^k}$.

- $R_k(x, \binom{k}{p}) = 2^k - 2^{k-p} + 1$,
- $R_k(x, \binom{k}{p} - 1) = 2^k - 2^{k-p} - 2^{k-p-1} + 1$.

■ **Démonstration de la proposition IV.3.7**

1. La première chose à remarquer, c'est que $\overline{\sigma}^k(v) = \overline{u}_0 \dots \overline{u}_{2^k-1}$ vérifie :

$$\forall n \in [\![0, 2^k - 1]\!], \forall b \in [\![-k, k]\!],\ \overline{u}_n = v + b \Longleftrightarrow \overline{u}_{2^k-1-n} = v - b.$$

Ce résultat se montre par récurrence sur $k \geq 1$.

En notant $\sigma^k(v) = u_0 \dots u_{2^k-1}$, on obtient directement :

$$\forall n \in [\![0, 2^k - 1]\!],\ u_n = 1 \iff u_{2^k-1-n} = 1,$$

ce qui montre que $\sigma^k(v)$ est un palindrome de longueur 2^k, non nul lorsque k est de même parité que v.

Pour montrer que, pour tout $x = v + k - 2p$, $\sigma(v - (k-2p))$ est l'image miroir de $\sigma^k(x)$, on note

$$\overline{\sigma}^k(v + k - 2p) = \overline{y}_0 \dots \overline{y}_{2^k-1}, \quad \sigma^k(v + k - 2p) = y_0 \dots y_{2^k-1},$$

$$\overline{\sigma}^k(v - (k-2p)) = \overline{w}_0 \dots \overline{w}_{2^k-1}, \quad \sigma^k(v - (k-2p)) = w_0 \dots w_{2^k-1}.$$

Toutes les applications coordonnées de $\overline{\sigma}$ étant linéaires, on a : $\forall n \in [\![0, 2^k - 1]\!],\ \overline{y}_n = \overline{u}_n + (k-2p)$ et $\overline{w}_n = \overline{u}_n - (k-2p)$. Ainsi, on a les équivalences suivantes, pour n dans $[\![0, 2^k - 1]\!]$:

$$y_n = 1 \iff \overline{y}_n = v \iff \overline{u}_n = v - (k-2p),$$

et d'autre part,

$$w_{2^k-1-n} = 1 \iff \overline{w}_{2^k-1-n} = v \iff \overline{u}_{2^k-1-n} = v + (k-2p).$$

On en déduit donc, à partir de ces équivalences et du résultat précédent, que $\sigma^k(v-(k-2p))$ est l'image miroir de $\sigma^k(v+k-2p)$.

2. Pour $x = v + k - 2p$, un élément fixé de $[\![v-k, v+k]\!]$, on note $N_k(k-2p)$ le nombre d'occurence de 1 dans $\sigma^k(x)$. l'ensemble $\{N_k(k-2p), k \geq 1, p \in \mathbb{N}\}$ vérifie :

- $\forall p \notin [\![0, k]\!],\ N_k(k-2p) = 0$.
- comme $\sigma^k(v+k) = 10^{2^k-1}$ et $\sigma^k(v-k) = 0^{2^k-1}1$, on obtient :

$$\forall k \geq 1,\ N_k(k) = N_k(-k) = 1.$$

- La relation $\sigma^k(v+k-2p) = \sigma^{k-1}(v+k-1-2p)\sigma^{k-1}(v+k-1-2(p-1))$ donne

$$N_k(k-2p) = N_{k-1}(k-1-2p) + N^{k-1}(k-1-2(p-1)).$$

La fonction $(k,p) \mapsto N_k(k-2p)$ vérifie donc les mêmes propriétés fonctionnelles que la fonction $(k,p) \mapsto \binom{k}{p}$ donc :

$$\forall k \geq 1,\ \forall p \in [\![0,k]\!],\ N_k(k-2p) = \binom{k}{p}.$$

Notons que le triangle de construction présenté au début du paragraphe permet de voir directement cette propriété.

3. De la relation $\sigma^k(v+k-2p) = \sigma^{k-1}(v+k-1-2p)\sigma^{k-1}(v+k-1-2(p-1))$ et du résultat $N_k(k-2p) = \binom{k}{p}$ découle :

$$\forall q \in [\![1, \tbinom{k-1}{p}]\!],\ R_k(v+k-2p,q) = R_{k-1}(v+k-1-2p,q).$$

car les premiers 1 de $\sigma^k(x)$ sont les 1 de $\sigma^{k-1}(x-1)$ et

$$\forall q \in [\![\ \tbinom{k-1}{p}+1, \tbinom{k}{p}]\!],\ R_k(v+k-2p,q) = 2^{k-1} + R_{k-1}(v+k-1-2(p-1),q).$$

car les derniers 1 de $\sigma^k(x)$ sont ceux de $\sigma^{k-1}(x+1)$.

les valeurs exactes des $R_k(x,q)$ pour $x = v+k-2p$ avec p dans $[\![0,k]\!]$ se calculent par récurrence sur k, de proche en proche pour p à k fixé. ∎

On rappelle que les facteurs spéciaux à droite d'un mot m de $E^{\mathbb{N}}$ sont les sous-mots de m qui apparaîssent dans m prolongés à droite de plusieurs manières différentes (voir chapitre II).

Lemme IV.3.8. *Soit (a,q) un élément de $\mathbb{Z}\times\mathbb{N}$. Le mot $(v+a)(v+a-2q)$ est un mot de $\overline{m}$ qui se prolonge à gauche dans m manière unique par $(v+a-2)$, quelque soit p. C'est à dire :*

$$\forall (a,q) \in \mathbb{Z}\times\mathbb{N}, \forall n \in \mathbb{N},\ \ \overline{m}_n\overline{m}_{n+1} = (v+a)(v+a-2q) \Longrightarrow \overline{m}_{n-1} = (v+a-2).$$

∎ **Démonstration du lemme IV.3.8**

Pour tout couple (a,q) de $\mathbb{Z}\times\mathbb{N}$, le mot $(v+a)(v+a-2q)$ est un mot de $\overline{m}$, d'après le lemme IV.3.4. D'autre part, comme aucun élément de $\mathbb{Z}$ n'admet pour image par $\overline{\sigma}$ un mot de ce type, le mot $(v+a)(v+a-2q)$ apparait donc à une jonction du type $\overline{\sigma}(x_1)\overline{\sigma}(x_2)$, avec, forcement $x_1 = v+a-1$ et $x_2 = v+a-2q+1$. ∎

∎ **Démonstration de la proposition IV.3.3** Soit k un entier fixé, et soit n un entier tel que $2^k \leq n \leq 2^{k+1}-1$.

L'idée de la preuve consiste à exhiber un certain nombre de facteurs spéciaux à droite de longueur n du mot m, pour ainsi minorer la quantité $p_m(n+1)-p_m(n)$, quelque soit n. En sommant ces inégalités, on obtient ensuite une minoration de la fonction p_m.

L'idée que nous allons exploiter pour exhiber ces mots spéciaux est que tout mot w de m de longueur n est contenu dans un itéré $\sigma^k(x_0)\sigma^k(x_1)\sigma^k(x_2)$ pour un certain mot $x_0x_1x_2$ de m.

Pour tout couple d'entiers (p,q) avec $0 \leq p \leq q \leq k-1$ et $q \in [\![0,k-1]\!]$, on appelle $w^{(p,q)}$ le sous-mot de $\sigma^k(v+k-2(p+1))\sigma^k(v+k-2p)\sigma^k(v+k-2q)$ de longueur n extrait de telle sorte que la dernière lettre de $w^{(p,q)}$ soit la $R_k(v+(k-2q),1)-1$-ième lettre de $\sigma^k(v+k-2q)$ (avec les notations de la proposition IV.3.7), c'est-à-dire la (2^q-1)-ième lettre de $\sigma^k(v+k-2q)$ (voir figure IV.5). Autrement dit, si on pose $\sigma^k(v+k-2(p+1))\sigma^k(v+k-2p)\sigma^k(v+k-2q) = y_0y_1\ldots y_{32^k-1}$, on a

$$w^{(p,q)} = y_{2^{k+1}+2^q-n-1}y_{2^{k+1}+2^q-n}\cdots y_{2^{k+1}+2^q-1}.$$

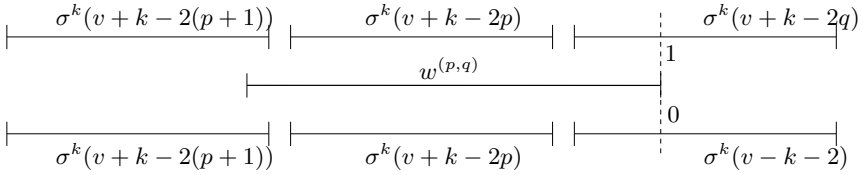


FIG. IV.5 – Construction du facteur spécial $w^{(p,q)}$.

Les mots $w^{(p,q)}$ sont tous des mots spéciaux à droite. En effet, comme $(v+k-2(p+1))(v+k-2p)(v+k-2q)$ et $(v+k-2(p+1))(v+k-2p)(v+k-2(q+1))$ apparaissent dans la suite, le mot $w^{(p,q)}$ peut se prolonger à droite par 1 si on le voit comme sous-mot de $\sigma^k(v+k-2(p+1))\sigma^k(v+k-2p)\sigma^k(v+k-2q)$ ou par 0 si on le considère comme sous-mot de $\sigma^k(v+k-2(p+1))\sigma^k(v+k-2p)\sigma^k(v+k-2(q+1))$, puisque la première occurence de 1 dans $\sigma^k(v+k-2(q+1))$ n'apparaît qu'au rang 2^{q+1}.

Cependant, les mots spéciaux ainsi formés ne sont pas forçément tous différents ; mais si on suppose que p appartient à $[\![1, k-1]\!]$ (ce qui assure que $\sigma^k(v+k-2p)$ contient au moins deux occurences de 1) et que $w^{(p,q)}$ contient les deux dernières occurences de 1 de $\sigma^k(v+k-2p)$, tous les $w^{(p,q)}$ formés sont différents.

En effet, dans ce cas, les deux dernières occurences de 1 de $w^{(p,q)}$ sont les deux dernières de $\sigma^k(v+k-2p)$. La longueur du bloc d'occurences de 0 entre les deux dernières occurences de 1 de $w^{(p,q)}$ détermine de manière unique p, ce nombre devant être égal à $R_k(v+k-2p, \binom{k}{p}) - R_k(v+k-2p, \binom{k}{p}-1) - 1$, c'est-à-dire égal à 2^{k-p-1}.

L'entier q est ensuite entièrement déterminé par le nombre d'occurences de 0 suivant la dernière occurence de 1 dans $w^{(p,q)}$, qui doit être égal à $R_k(v+k-2q,1)-1+2^k-R_k(v+k-2p, \binom{k}{p})$, c'est-à-dire égal à $2^{k-p}+2^q-1$.

D'après la propriété IV.3.7 3., $w^{(p,q)}$ contient les deux dernières occurences de 1 de $\sigma^k(v+k-2p)$ si et seulement si :

$$2^k + R_k(v+k-2q,1) - R_k(v+k-2p, \tbinom{k}{p}-1) \leq n,$$

Comme n appartient à $[\![2^k, 2^{k+1}-1]\!]$, quite à « perdre » quelques facteurs spéciaux, on est sûr d'obtenir des mots spéciaux différents lorsque :

$$2^k + R_k(v+k-2q,1) - R_k(v+k-2p, \tbinom{k}{p}-1) \leq 2^k,$$

c'est-à-dire lorsque $(p,q) \in \mathcal{S}$, avec

$$\mathcal{S} = \{(1,q), q \in [\![1,k-3]\!]\} \cup \{(p,q), p \in [\![2,k-1]\!], q \in [\![p,k-1]\!]\}.$$

On a donc au moins autant de facteurs spéciaux qui contiennent au moins deux occurences de 1 que de couples dans $\mathcal{S}$, c'est-à-dire $k-3+\sum_{i=2}^{k-1} i - 1$.

De plus, pour tout $q \leq 1$, le mot $w^{(-2,q)} = 0^n$ aussi est spécial.

Il existe d'autre mots spéciaux de longueur 2^k qui ne contiennent qu'un seul 1. Pour q dans $[\![1,k-2]\!]$, les mots $(v+k+2)(v+k+4)(v+k-2q)$ et $(v+k+2)(v+k+4)(v+k-2(q+1))$ sont des mots de m. On note $u^{(q)}$ le mot extrait de $\sigma^k(v+k+2)\sigma^k(v+k+4)\sigma^k(v+k-2q)$ de sorte que la dernière lettre de $u^{(q)}$ soit la $(R_k(v+k-2q,2)-1)$-ième lettre de $\sigma^k(v+k-2q)$, donc la $(2^q+2^{q-1}-1)$ième lettre de $\sigma^k(v+k-2q)$. C'est-à-dire (voir figure IV.6), si on note $\sigma^k(v+k+2)\sigma^k(v+k+4)\sigma^k(v+k-2q) = u_0u_1\dots u_{32^k-1}$, alors

$$u^{(q)} = u_{2^{k+1}+2^q+2^{q-1}-n-1}u_{2^{k+1}+2^q+2^{q-1}-n}\dots u_{2^{k+1}+2^q+2^{q-1}-1}.$$

On a donc $u^{(q)} = 0^{n-2^{q-1}}10^{2^{q-1}-1}$. Mais le mot $u^{(q)}$ peut aussi être extrait de $\sigma^k(v+k+2)\sigma^k(v+k+4)\sigma^k(v+k-2(q+1))$ de sorte que la dernière lettre de $u^{(q)}$ soit la $(R_k(v+k-2q,1)+2^{q-1}-1)$-ième lettre de $\sigma^k(v+k-2(q+1))$, donc la $(2^{q+1}+2^{q-1}-1)$-ième lettre de $\sigma^k(v+k-2(q+1))$. Autrement dit, si on note $\sigma^k(v+k+2)\sigma^k(v+k+4)\sigma^k(v+k-2(q+1)) = x_0x_1\dots x_{32^k-1}$,

$$u^{(q)} = x_{2^{k+1}+2^{q+1}+2^{q-1}-n-1}u_{2^{k+1}+2^{q+1}+2^{q-1}-n}\dots u_{2^{k+1}+2^{q+1}+2^{q-1}-1}.$$

Les mots $u^{(q)}$ sont tous des mots spéciaux à droite. en effet, comme $(v+k+2)(v+k+4)(v+k-2q)$ et $(v+k+2)(v+k+4)(v+k-2(q+1))$ apparaîssent dans la suite, le mot $u^{(q)}$ peut se prolonger a droite par 1 si on le voit comme sous-mot de $\sigma^k(v+k+2)\sigma^k(v+k+4)\sigma^k(v+k-2q)$ ou par 0 si on le considère comme sous-mot de $\sigma^k(v+k+2)\sigma^k(v+k+4)\sigma^k(v+k-2(q+1))$, puisque la deuxième occurence de 1 dans $\sigma^k(v+k-2(q+1))$ n'apparaît qu'au rang $2^{q+1}+2^q-1$.

La restriction $q \leq k-2$ que l'on s'est imposé est nécessaire pour que $u^{(q)}0$ puisse être entièrement inclus dans $\sigma^k(v+k+2)\sigma^k(v+k+4)\sigma^k(v+k-2(q+1))$, mais le mot $u^{(k-1)} = 0^{n-2^{k-2}}10^{2^{k-2}-1}$ est lui aussi spécial. En utilisant la construction que nous venons de faire, le dernier 1 de $u^{(k-1)}$ se trouve être l'unique 1 de $\sigma^k(v-k) = 0^{2^k-1}1$ mais aussi la dernière lettre de ce mot. Cependant, le mot $(v+k+2)(v-k)(v-k+2)$ est un facteur de m et $\sigma^k(v+k+2)\sigma^k(v-k)\sigma^k(v-k+2)$ contient $u^{(k-1)}0$

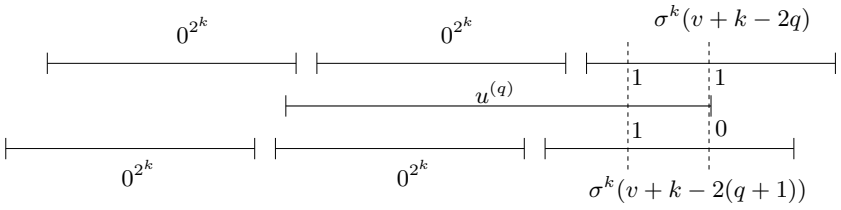


FIG. IV.6 – Construction du facteur spécial $u^{(p)}$.

car la longueur du bloc d'occurences de 0 entre les deux premières occurrences de 1 de ce facteur est 2^{k-1}.

Ainsi, on a exhibé de cette manière $k-1$ nouveaux facteurs spéciaux de longueur n différents les uns des autres et différents de ceux obtenus plus haut.

On a donc au minimum $2k-3+\sum_{i=1}^{k-2} i = (\sum_{i=1}^{k} i) - 2$ facteurs spéciaux de longueur n, d'où

$$\forall k \in \mathbb{N}^*, \forall n \in [\![2^k, 2^{k+1}-1]\!],\ p_m(n+1) - p_m(n) \geq \frac{k(k+1)}{2} - 2.$$

En sommant ces inégalités, on obtient :

$$\forall k \in \mathbb{N}^*, \forall n \in [\![2^k, 2^{k+1}-1]\!], n \geq 3, p_m(n) - p_m(2) \geq (n-2^k)\left(\frac{k(k+1)}{2} - 2\right) + \sum_{i=0}^{k-1} 2^i \left(\frac{i(i+1)}{2} - 2\right).$$

Comme $p_m(2) = 4$, on obtient :

$$\forall k \in \mathbb{N}^*, \forall n \in [\![2^k, 2^{k+1}-1]\!], n \geq 3,\ p_m(n) \geq (n-2^k)\left(\frac{k(k+1)}{2} - 2\right) + (k+1)(k+2)2^{k-1} - 2^{k+1} - 2 + 4,$$

il suit

$$\forall k \in \mathbb{N}^*, \forall n \in [\![2^k, 2^{k+1}-1]\!], n \geq 3,\ p_m(n) - p_m(2) \geq n\left(\frac{k(k+1)}{2} - 2\right) + (k+1)2^k + 2.$$

En minorant k par $\log_2 n - 1$ et 2^k par $\frac{n}{2}$, on obtient alors :

$$\forall n \in \mathbb{N}, n \geq 3,\ p_m(n) \geq n\left(\frac{(\log_2 n + 1)\log_2 n}{2} - 2\right) + \frac{n \log_2 n}{2} + 2,$$

et donc

$$\forall n \in \mathbb{N}, n \geq 3,\ p_m(n) \geq n\left((\log_2 n)^2 - 2\right) + 2,$$

ce qui termine la démonstration. ■

■ **Démonstration du théorème IV.3.1** Les résultats IV.3.2 et IV.3.3 permettent l'encadrement suivant de la complexité :

$$\forall n \in \mathbb{N}, n \geq 3,\ n\left((\log_2 n)^2 - 2\right) + 2 \leq p_m(n) \leq n(\log_2 n + 9)(\log_2 n + 2).$$

les deux fonctions qui encadrent la complexité du mot m sont toutes les deux équivalentes à $n \log_2^2 n$. On en déduit donc que la fonction $p_m(n)$ est elle aussi équivalente à $n \log_2^2 n$. ■

Remarque IV.3.9. On pourrait extrapoler ce résultat en conjecturant que la complexité du mot $(2d)^\infty$-automatique associé à la marche aléatoire régulière dans $\mathcal{R}$ est de l'ordre de grandeur de $n \log_{2d}^{d+1} n$. La difficulté à surmonter pour prouver ce résultat est de connaitre suffisamment la structure des itérés $\sigma^k(e)$ pour pouvoir, avec certitude, minorer le nombre de facteurs spéciaux.

Chapitre V

Complexité des mots q^∞-automatiques

V.1 Deux théorèmes généraux

Nous donnons, dans ce paragraphe, deux théorèmes généraux de majoration de la complexité pour des classes de mots q^∞-automatiques :

L'un concerne les mots engendrés par des q-ADDA auxquels on impose une condition sur la structure des graphes non orienté qui les portent, à savoir les q-ADDA de degré borné. L'autre concerne les mots engendrés par les q-ADDA auxquels on impose une condition sur l'ensemble des arcs orienté de leur graphe, à savoir les q-ADDA monotones (voir paragraphe III.4 pour les définitions).

Dans les deux cas, on obtient une majoration polynomiale de la fonction de complexité de ces mots.

V.1.1 Un théorème pour les mots q^∞-automatiques engendrés par des q-ADDA de degré borné

Nous donnons ici une majoration polynomiale de la complexité des mots infinis q^∞-automatiques, engendrés par les substitutions K-uniformement bornées ou, de manière équivalente, par les q-ADDA de degré borné (voir paragraphe III.4).

Théorème V.1.1. *Soient E un alphabet dénombrable, A un alphabet fini, $\overline{\sigma}$ une substitution de longueur constante q sur E et Π une projection admissible de E vers A.*

Soit m le mot infini q^∞-automatique, engendré par le q-ADDA $\mathcal{A} = (E, \phi_{\overline{\sigma}}, e_0, A, \Pi)$.

Si $\mathcal{A}$ est de degré borné par K pour Π, alors la complexité de m vérifie :

$$p_m(n) = O(n^{3+2\log_q K}).$$

Si de plus, les fonctions $\overline{\sigma}_i$ commutent deux à deux, alors

$$p_m(n) = O(n^{1+2\log_q K}(log_q n)^{2q}).$$

Remarque V.1.2. Par exemple, la deuxième partie du théorème est intéressante pour la famille de mots infinis $\{m^{(\alpha)}, \alpha \in \mathbb{Q}, \alpha \geq 0\}$ des suites indicatrices des sous-ensembles E_α de $\mathbb{N}$ formés des entiers qui vérifie : la proportion de 1 dans l'écriture binaire de n est égal à α fois la proportion de 0, c'est-à-dire, donnés par $E_\alpha = \{n, \sum_{i=0}^{|\overline{n}^2|} n_i = \frac{\alpha}{1+\alpha}|\overline{n}^2|\}$.

Le mot $m^{(\alpha)}$ peut être vu comme projection par Π_0 du point fixe de la substitution suivante :

$$\begin{array}{rccc} \overline{\sigma}: & \mathbb{Z} \cup \{\varepsilon\} & \rightarrow & \mathbb{Z}^2 \cup \{\varepsilon a_\alpha\} \\ & x & \mapsto & (x - b_\alpha)(x + a_\alpha) \\ & \varepsilon & \mapsto & \varepsilon a_\alpha. \end{array}$$

où a_α et b_α sont tels que $\alpha = \frac{a_\alpha}{b_\alpha}$.

Pour tout $\alpha \in \mathbb{Q}$, ce théorème donne donc une meilleure majoration de la fonction de complexité $(p_{m^{(\alpha)}}(n) = O(n \log^2 n))$, que celle obtenue au théorème IV.2.1.

■ **Démonstration du théorème V.1.1** Pour démontrer ce théorème, on montre en fait les affirmations suivantes :

Si $\mathcal{A}$ est de degré borné par K pour Π, alors la complexité de m, on a :

$$\forall n \geq 2,\ p_m(n) \leq qn(|\Pi| qK n^{1+\log_q K} + 1)^2,$$

et si de plus, les fonctions $\overline{\sigma}_i$ commutent deux à deux, la fonction de complexité admet la majoration suivante :

$$p_m(n) \leq qKn \left(\frac{K|\Pi| n^{\log_q K}}{(q-1)!} \prod_{i=1}^{q} \left(\log_q(n) + i\right) + 1 \right)^2.$$

L'idée est de majorer, dans un premier temps, $p_m(q^k)$ pour un k fixé, puis étendre cette majoration à tout $n > 0$.

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 $\overline{\sigma}^k(x_1)\overline{\sigma}^k(x_2)$. On va majorer le nombre de $\overline{\sigma}^k(x)$ qui se projètent sur des mots différents de $a_0^{q^k}$.

En conséquence directe de la proposition III.5.1, on a l'équivalence, pour tout $a \in A \setminus \{a_0\}$:

$$a \subset \sigma^k(x) \Rightarrow \exists s \in F_\Pi, s \subset \overline{\sigma}^k(x) \Rightarrow x \in \bigcup_{s \in F_\Pi} \mathcal{Z}_k(s).$$

Et donc, puisque $\mathcal{Z}_k(s)$ est invlus dans E_Π, on obtient l'inégalité, dans le cas général :

$$\left| \left\{ x \in E,\ \sigma^k(x) \neq a_0^{q^k} \right\} \right| \leq |\Pi| (qK)^k.$$

Il y a ainsi au plus $|\Pi| (qK)^k$ mots du type $\overline{\sigma}^k(x)$ qui se projètent de manière différente de $a_0^{q^k}$, et donc $|\Pi| (qK)^k + 1$ possibilités de projections différentes pour $\overline{\sigma}^k(x_1)$ et $\overline{\sigma}^k(x_2)$.

Comme chaque projection $\sigma^k(x_1)\sigma^k(x_2)$ peut contenir au plus q^k nouveaux mots différents de longueur q^k, on a une première la majoration :

$$\forall k > 1,\ p_m(q^k) \leq q^k |\Pi| ((qK)^k + 1)^2.$$

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

$$\forall n \in [\![q^k, q^{k+1}[\![,\ p_m(n). \leq q^{k+1} |\Pi| (qK)^{k+1} + 1)^2.$$

Et ensuite,

$$\forall n \geq 1,\ p_m(n) \leq qn(q |\Pi| K n^{1+\log_q K} + 1)^2.$$

En utilisant le fait que, pour n dans $[\![q^k, q^{k+1}[\![$, on a $k \leq \log_q n$ et donc $(qK)^{k+1} \leq qKn^{1+\log_q K}$.

Pour montrer la deuxième partie du théorème, il suffit de remarquer que lorsque les $\overline{\sigma}_i$ commutent, on a l'inégalité suivante :

$$\left| \left\{ x \in E, \sigma^k(x) \neq a_0^{q^k} \right\} \right| \leq |\Pi| \binom{k+q-1}{q-1} K^q.$$

On utilise ensuite le même raisonnement. ■

V.1.2 Un théorème pour les mots q^∞-automatiques engendrés par les q-ADDA monotones

Lorsque le q-ADDA qui engendre un mot q^∞-automatique n'est pas de degré borné, le schéma de preuve que nous avons mis en place dans le paragraphe précédent ne tient plus. Cependant, lorsque le q-ADDA qui engendre un mot q^∞-automatique est monotone, il existe un autre moyen de garantir une complexité au plus polynomiale :

Théorème V.1.3. *Soient E un alphabet dénombrable, A un alphabet fini, $\overline{\sigma}$ une substitution de longueur constante q sur E et Π une projection admissible de E vers A.*

Soit m le mot infini q^∞-automatique, engendré par le q-ADDA $\mathcal{A} = (E, e_0, \phi_{\overline{\sigma}}, A, \Pi)$.

Si $\mathcal{A}$ est un q-ADDA monotone, alors la complexité de m vérifie :

$$p_m(n) = O(n^3).$$

Ce théorème permet, par exemple, de majorer la complexité du mot 2^∞-automatique engendré par l'automate de degré fini (et non borné) associé à la substitution

$$\begin{array}{rccc} \overline{\theta}: & \mathbb{N} & \to & \mathbb{N}^* \\ & e \neq 0 & \mapsto & (\lfloor \log_K e \rfloor)(e+1) \\ & 0 & \mapsto & 01 \end{array}$$

pour un certain entier $K \geq 2$ et à la projection admissible Π qui vérifie $\Pi^{-1}(\{1\}) = \{0\}$. Ce que nous ne pouvons pas faire avec le théorème V.1.1.

D'autre part, il donne une meilleure majoration « brute » de l'ordre de grandeur de la fonction de complexité des mots q^∞-automatiques qui sont engendrés par des q-ADDA qui sont à la fois de degré borné et monotones.

■ **Démonstration du théorème V.1.3** On va montrer qu'il existe une constante C_Π telle que :

$$\forall n \geq 2,\ p_m(n) \leq C_\Pi^2 q^3 n^3.$$

On ordonne F_Π de manière croîssante : $F_\Pi = \{f_1, f_2, \ldots, f_\Pi\}$ et on suppose que $\Pi(E \setminus F_\Pi) = \{a_0\}$.

On dira que f_i et f_{i+1} sont *consécutifs dans* E s'il n'existe pas d'élément x de E, différent de f_i et f_{i+1} qui vérifie $f_i \leq x \leq f_{i+1}$.

On introduit la famille d'entiers $(c_1, \ldots, c_{|\pi|})$ définie par $c_{|\Pi|} = 1$ et

$$\forall i \in [\![1, |\Pi| - 1]\!],\ c_i = \begin{cases} 0 & \text{si } f_i \text{ et } f_{i+1} \text{ sont consécutifs,} \\ 1 & \text{sinon.} \end{cases}$$

On pose $C_\Pi = |\Pi| + \sum_{i=1}^{|\Pi|} c_i$.

Soit k un entier fixé. On va majorer le nombre de mots $\sigma^k(x)$ différents.

Toutes les applications du type $\overline{\sigma}_{i_k} \circ \overline{\sigma}_{i_{k-1}} \circ \cdots \circ \overline{\sigma}_{i_1}$ sont des fonctions monotones (puisque $\overline{\sigma}$ est monotone). La projection par Π de cette application est donc une fonction palier de x. Dans le pire des cas, $\sigma_{i_k} \circ \sigma_{i_{k-1}} \circ \cdots \circ \sigma_{i_1}$ commence par être constante, égale à a_0, passe, par tous les paliers $\Pi(f_i)$ de manière monotone en i, entrecoupés par des paliers a_0 lorsque f_i et f_{i+1} ne sont pas consécutifs et termine constante égale à a_0.

La fonction $x \mapsto \sigma_{i_k} \circ \sigma_{i_{k-1}} \circ \cdots \circ \sigma_{i_1}(x)$ fait donc, au plus, C_Π sauts. Ceci est valable pour les q^k applications coordonnées de σ^k. En admettant que les sauts de toutes les applications coordonnées ont lieu pour les éléments de E différents, on obtient, grâce à la proposition III.5.1, au pire $C_\Pi q^k$ sauts différents. et donc $C_\Pi q^k$ mots $\sigma^k(x)$ différents.

Tout mot w de m de longueur q^k étant contenu dans un mot de longueur q^{k+1} du type $\sigma^k(x_1)\sigma^k(x_2)$, on majore le nombre de mots w par q^k fois le nombre de mots $\sigma^k(x_1)\sigma^k(x_2)$ possibles. Il vient :

$$p_m(q^k) \leq C_\Pi^2 q^{3k}.$$

en utilisant le fait que p_m est croîssante et que, pour n dans $[\![q^k, q^{k+1}[\![$, on a $k \leq \log_q n$, on obtient :

$$\forall n \geq 2, p_m(n) \leq C_\Pi^2 q^3 n^3.$$

ce qui termine la preuve. ■

Lorsqu'en pratique, on utilise les deux théorèmes V.1.1 et V.1.3 sur un mot q^∞-automatique engendré par un q-ADDA de degré borné et monotone, il apparait que le théorème V.1.1 donne une moins bonne majoration que le second. Cependant, nous allons voir que la méthode sur laquelle est basée le théorème V.1.1 peut être raffinée lorsque le q-ADD est donné explicitement et donner ainsi une borne de l'ordre de la complexité plus intéressante.

Le théorème V.1.3 reste tout de même intéressant car c'est le seul outil pour majorer la fonction de complexité des mots q^∞-automatiques engendrés par des automates monotones de degré non borné.

V.2 Exemples d'étude de complexité de mots q^∞-automatiques

La majoration de l'ordre de grandeur des fonctions de complexité de cette classe de mots q^∞-automatiques est une majoration finalement assez barbare, mais c'est le prix à payer pour la généralité. La méthode utilisée peut donner, sur des exemples fixés, de bien meilleures majorations, notamment en exploitant les deux remarques suivantes :

1. Dans des cas particuliers, on peut obtenir une meilleure majoration du cardinal de $\bigcup_{s\in F_\Pi} \mathcal{Z}_k(s)$ et donc une meilleure majoration de l'ordre de grandeur de la fonction de complexité.
2. Dans certains cas, la proposition III.5.3 peut aussi être utilisée pour faire baisser le nombre de mots de deux lettres de m et permettre également d'obtenir une meilleure majoration de l'ordre de grandeur de la fonction de complexité ; c'est ce qui a été fait pour les marches aléatoires.

Nous donnons ici un exemple de mot 2-automatique pour lequel on peut utiliser la première remarque et une famille de mots 2-automatiques pour laquelle ont peut utiliser la deuxième remarque fait l'objet de la section suivante.

L'exemple du mot infini de Dyck sur deux types de parenthèses, présenté au chapitre VI, permet de valider en quelque sorte ces deux manières d'améliorer le théorème V.1.1 car la majoration obtenue par ces méthodes donne en fait exactement l'ordre de croissance de la fonction de complexité.

V.2.1 Un premier exemple

Soit $\overline{\tau}$ la substitution suivante

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

On considère le mot 2^∞-automatique engendré par le 2-ADDA $(\mathbb{Z}\cup\{e_0\}, \phi_{\overline{\tau}}, e_0, \{0,1\}, \Pi_0)$ où la projection $\Pi_0 : \mathbb{Z}\cup\{e_0\} \to \{0,1\}$ est définie par $\Pi_0^{-1}(\{1\}) = \{0\}$.

Le 2-ADDA qui engendre m est de degré borné par 2 mais n'est pas monotone. Le résultat V.1.1 donne donc une première majoration de de la complexité de m : $p_m(n) = O(n^5)$. Cette majoration peut être largement améliorée, en adaptant la preuve du théorème V.1.1, pour obtenir le résultat suivant :

Proposition V.2.1. *La fonction de complexité de m vérifie :*

$$p_m(n) = O(n\log_2^2 n).$$

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 $\mathcal{F}_k = \left\{P_{i_1}\circ\cdots\circ P_{i_k},\ \ \forall j\in[\![1,k]\!], i_j\in\{0,1\}\right\}$,
- l'ensemble de nombres entiers $\mathcal{R}_k = \left\{n\in\mathbb{N},\ \ \exists P\in\mathcal{F}_k, P(n)=0\right\}$.

Lemme V.2.2. *Pour tout $k\geq 1$, l'ensemble $\mathcal{R}_k$ est inclus dans l'intervalle $[-\sqrt{k}, k]$.*
En particulier, $Card(\mathcal{R}_k) \leq k+\sqrt{k}+1$.

■ **Démonstration du lemme V.2.2**

On procède par récurrence forte.

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

Supposons maintenant, pour $k\geq 2$ et tout $p\leq k-1$, 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 $2\leq i\leq q-1$, 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 $(\alpha_1,\dots,\alpha_q)$ et $(\beta_1,\dots,\beta_q)$, 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 $\alpha_1 \geq 1$ et $\beta_1 \geq 1$ (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 forcement des valeurs entières, mais qui, dans tout les cas sont contenues dans l'intervalle $[-\sqrt{k}, \sqrt{k}]$.

Pour tous les autres q-uplets $(\alpha_1,\dots,\alpha_q)$ et $(\beta_1,\dots,\beta_q)$ 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

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

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 de la proposition V.2.1** Pour montrer ce résultat, on va en fait montrer que pour tout $n \geq 2$,

$$p_m(n) \leq 2n\big(\log_2 n + \sqrt{\log_2 n + 1} + 3\big)^2.$$

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 $\overline{\tau}$ contenu dans $e_0\mathbb{Z}^\omega$.

La majoration de la complexité de m est basée sur l'étude des mots $\tau^k(e)$, 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 $\overline{\tau}^k(e_1)\overline{\tau}^k(e_2)$. On va majorer le nombre de $\overline{\tau}^k(e)$ qui se projettent sur des mots différents de 0^{2^k}.

En conséquence directe de la proposition III.5.1, le lettre 1 apparait dans $\tau^k(e)$ si et seulement si 0 apparait dans $\overline{\tau}^k(e)$ et donc, si et seulement si $e \in \bigcup_{(j_1,\dots,j_k)\in\{0,1\}^k} P_{j_1}^{-1} \circ \cdots \circ P_{j_k}^{-1}\{0\}$. On peut traduire cela par : la lettre 1 apparait dans $\tau^k(e)$ si et seulement si $e \in \mathcal{R}_k$.

Ainsi, on obtient l'inégalité :

$$Card\big(\{e \in \mathbb{Z},\ 1 \text{ apparaît dans } \tau^k(e)\}\big) \leq Card(\mathcal{R}_k).$$

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

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

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

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

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

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

$$\forall n \geq 1,\ p_m(n) \leq 2n(\log_2 n + \sqrt{\log_2 n + 1} + 3)^2,$$

ce qui termine la preuve. ■

V.2.2 Sur une famille de mots 2^∞-automatiques engendrés par des q-ADDA de degré borné

On introduit, pour un nombre entier $K \geq 2$, $\overline{\kappa}$ la substitution $\overline{\kappa}$ suivante :

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

ainsi que la projection admissible $\Pi_0 : \mathbb{N} \to \{0,1\}$, définie par $\Pi_0^{-1}(\{1\}) = \{0\}$.

Le 2-ADDA $(\mathbb{N}, \phi_{\overline{\kappa}}, 0, \{0,1\}, \Pi_0)$ est de degré borné par K et est monotone.

On peut donc utiliser les théorèmes V.1.1 et V.1.3 pour obtenir des majorations de la fonction de complexité du mot 2^∞-automatique m engendré par le 2-ADDA $(\mathbb{N}, \phi_{\overline{\kappa}}, 0, \{0,1\}, \Pi_0)$. La meilleure majoration est celle obtenue grâce au théorème V.1.3 : $p_m(n) = O(n^3)$. Ce résultat peut cependant être amélioré :

Proposition V.2.3.
La complexité du mot 2^∞-automatique m engendré par le 2-ADDA $(\mathbb{N}, \phi_{\overline{\kappa}}, 0, \{0,1\}, \Pi_0)$ vérifie :

$$p_m(n) = O(n^{1+\log_2 K} \log_2 n)$$

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

Le graphe de l'automate associé à la substitution $\overline{\kappa}$ pour $K = 2$ est représenté à la figure V.1

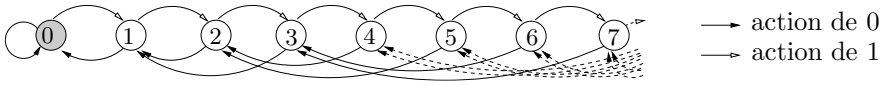


FIG. V.1 – 2-ADD associé à la substitution $\overline{\kappa}$ pour $K = 2$.

Si on pose $\overline{\kappa}_0(e) = \lfloor \frac{e}{K} \rfloor$ et $\overline{\kappa}_1(e) = e + 1$, la proposition III.5.3 devient :

Lemme V.2.4. *Soit $\overline{m}$ le mot infini point fixe de $\overline{\kappa}$ 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. $\exists k' \geq 0, e_1 \in \mathbb{N}$ *et* $e_2 \in \overline{\kappa}_0^{k'} \circ \overline{\kappa}_1 \circ \overline{\kappa}_0^{-1} \circ \overline{\kappa}_1^{-k'}(\{e_1\})$.

■ **Démonstration du lemme V.2.4**

C'est une réécriture des différents cas de la proposition III.5.3 : le premier cas du lemme V.2.4 correspond au premier cas de la proposition III.5.3. Le deuxième cas de la proposition III.5.3 disparait car on a ici une substitution de longueur égale à 2.

Le troisième cas de la proposition III.5.3 devient, puisque ici $e_{q-1} = e_1$: il existe $k \geq 0$ tel que $e_1e_2 = \overline{\kappa}_1^k(1)\overline{\kappa}_0^{k+1}(1)$. Or, pour tout $k \geq 0$, $\overline{\kappa}_1^k(1) = k+1$ et $\overline{\kappa}_0^{k+1}(1) = 0$. Ce qui correspond au deuxième cas du lemme V.2.4.

Enfin, le dernier cas du lemme correspond au quatrième cas de la proposition III.5.3. ■

Lemme V.2.5. *Soit $k \geq 1$ fixé.*

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

En particulier, quelque soit $p \in [\![0, K^{k-1} - 1]\!]$, $\kappa^k(Kp)$ et $\kappa^k(Kp+i)$, pour tout $i \in [\![0, K-1]\!]$ ont le même préfixe de longueur 2^{k-1}, donné par $\kappa^{k-1}(p)$ et on a :

$$\forall p \in [\![K^{k-2}, K^{k-1} - 1]\!], \forall i \in [\![0, K-1]\!],\ \kappa^k(Kp) = \kappa^k(Kp+i).$$

■ **Démonstration du lemme V.2.5**

Soit $k \geq 1$ fixé.

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

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

$$\kappa^k(Kp) = \kappa^{k-1}(p)\kappa^{k-1}(Kp+1) \text{ et } \kappa^k(Kp+i) = \kappa^{k-1}(p)\kappa^{k-1}(Kp+i+1),$$

et en remarquant que pour $p \in [\![K^{k-2}, K^{k-1}-1]\!]$, les mots $\kappa^{k-1}(Kp+1)$ et $\kappa^{k-1}(Kp+i+1)$, sont tous les deux le mot $0^{2^{k-1}}$, pour tout $i \in [\![0, K-1]\!]$. ■

■ **Démonstration de la proposition V.2.3** Pour montrer ce résultat, on va en fait pouver que p_m admet la majoration suivante :

$$\forall n \geq 1,\ p_m(n) \leq \frac{n^{1+\log_2 K}}{K}(2K^2(K+1))(\log_2 n + 6K^3 + 9K^2 - 5).$$

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

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

(b) les deux mots $\kappa^k(e_1)$ et $\kappa^k(e_2)$ contiennent des occurences 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 V.2.4.

(a) Si un seul des mots $\kappa^k(e_1)$ ou $\kappa^k(e_2)$, contient des occurences de 1, en tenant compte du lemme V.2.5, on peut distinguer $K^{k-1}+1$ mots $\kappa^k(e_1)$ qui ne sont pas le mot nul et qui sont différents. On peut, de plus, les classer en K^{k-2} classes de K itérés ayant le même préfixe de longueur 2^k, en regroupant les itérés $\overline{\kappa}^k(Kp+i)$, pour tout $i \in [\![0, K-1]\!]$ et pour un p fixé dans $[\![0, K^{k-1}-1]\!]$ et, plus une classe formée de l'élément $\overline{\kappa}^k(K^{k-1})$.

On peut alors former $2^{k-1}(K^{k-1}+1) + 2^{k-1}(K^{k-1})$ mots, a priori différents, si $\kappa^k(e_1)$ est le seul itéré à contenir des occurrences de 1, en dissociant les mots de longueur 2^k qui commencent avant la 2^{k-1}-ième lettre de $\kappa^k(e_1)\kappa^k(e_2)$ et ceux qui commencent après, puisque les itérés $\overline{\kappa}^k(K^{k-1}+i)$ finissent tous par le mot $0^{2^{k-1}}$, les mots formés à partir du 2^{k-1}-ième rang de ces itérés sont le mot nul.

D'autre part, on peut former $2^{k-1}(\frac{K^{k-1}}{K}+1) + 2^{k-1}(K^{k-1}+1)$ mots, a priori différents, si $\kappa^k(e_2)$ est le seul itéré à contenir des occurrences de 1, en dissociant les mots de longueur 2^k qui commencent avant la 2^{k-1}-ième lettre de $\kappa^k(e_1)\kappa^k(e_2)$ et ceux qui commencent après, et en utilisant la deuxième partie du lemme précédent.

(b) Si $\kappa^k(e_1)$ et $\kappa^k(e_2)$ contiennent des occurences de 1, c'est-à-dire e_1 et e_2 sont dans $[\![0, K^k-1]\!]$. En utilisant le lemme V.2.4, on a plusieurs possibilités de mots e_1e_2 :

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

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

On peut donc fabriquer, en tenant compte du lemme V.2.5, au plus, $K^{k-1} + K^{k-2}$ itérés différents $\kappa^k(e_1)\kappa^k(e_2)$ de ce type, ce qui permet, au pire, de fabriquer $2^k(K^{k-1}+K^{k-2})$ 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 $\kappa^k(e_i)$ car la fonction $\overline{\kappa}_0^{k'} \circ \overline{\kappa}_1 \circ \overline{\kappa}_0^{-1} \circ \overline{\kappa}_1^{-k'}$ est décroissante pout $k' > 0$.

Supposons que $e_1 \in [\![0, K^k-1]\!]$. Cette situation donne des mots valides uniquement lorsque l'on peut appliquer $\overline{\kappa}_1^{-k'}$ à e_1, c'est-à-dire lorsque $e_1 \in [\![0, K^k-1]\!]$ et $k' \leq e_1$.

D'autre part, si $e_1 \in [\![K^a, K^{a+1}[\![$ pour un $a < k$ fixé, on a, au plus, $2a$ possibilités pour $e_2 \neq 0$, lorsque k' varie. En effet, pour tout $0 < k' \leq e_1$, e_2 peut prendre les valeurs $\overline{\kappa}_0^{k'}(K(e_1-k')+i+1)$ pour tout $i \in [\![0, K-1]\!]$. Comme $k' > 0$, la composition d'au moins une fois $\overline{\kappa}_0$ assure qu'on a au maximum deux valeurs possibles pour e_2 : $\overline{\kappa}_0^{k'-1}(e_1-k')$ ou $\overline{\kappa}_0^{k'-1}(e_1-k'+1)$, selon que $i < K-1$ ou $i = K-1$. D'autre part, si $e_1 \in [\![K^a, K^{a+1}[\![$, pour tout $k' \geq a+1$, $e_2 = 0$ et ces mots ont déjà étés comptés (1.), par conséquent, on a au plus $2a$ lettres admissibles $e_2 \neq 0$ à la suite de e_1.

On obtient donc de cette manière, au plus $\sum_{a=0}^{k-1}(2a)K^a = \frac{(k-1)K^{k+1}+(k+2)K^k+K-1}{(K-1)^2}$ mots différents de $0^{2^{k+1}}$ possibles, qui nous permettent de fabriquer au plus $2^k\frac{(k-1)K^{k+1}+(k+2)K^k+K-1}{(K-1)^2}$ mots de longueur

2^k dans les itérés de ce type.
3. Pour $e_1 \in [\![0, K^k - 1]\!]$, on a K image par $\overline{\kappa}_1 \circ \overline{\kappa}_0^{-1}$: $Ke_1 + i$, pour $i \in [\![1, K]\!]$. Si e_2 est aussi dans $[\![0, K^k - 1]\!]$, il faut donc choisir e_1 dans $[\![0, K^{k-1} - 1]\!]$ et, dans ce cas, on a K possibilités pour e_2, lorsque e_1 est fixé dans $[\![0, 2^{k-1} - 1$ 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 $2^k(K^k + 1)$ 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, :

$$p_m(2^k) \leq 2^{k-1} K^{k-2} \left((2K^2(K+1))k + 4K^3 + 7K^2 - 5\right).$$

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

$$\forall 2^k \leq n < 2^{k+1},\ p_m(n) \leq 2^k K^{k-1} \left((2K^2(K+1))k + 6K^3 + 9K^2 - 5\right).$$

Et ensuite,

$$\forall n \geq 1,\ p_m(n) \leq \frac{n^{1+\log_2 K}}{K} \left((2K^2(K+1))(\log_2 n + 6K^3 + 9K^2 - 5\right).$$

En utilisant le fait que, pourtout entier n de $[\![2^k, 2^{k+1}[\![$, on a $k \leq \log_2 n < k + 1$. ∎

V.3 Commentaires

A la vue des exemples que nous venons de traiter, on voit que, si l'on étudie la fonction de complexité d'un mot q^∞-automatique défini par une substitution $\overline{\sigma}$, les défauts d'injectivité des fonctions $\overline{\sigma}_{i+1} \circ \overline{\sigma}_i^{-1}$ et surtout ceux de $\overline{\sigma}_q^{-1}$ ainsi que les défauts de commutativité des couples de fonctions $\overline{\sigma}_i$ et $\overline{\sigma}_j$ sont en corrélation directes avec la majoration de l'ordre de la fonction de complexité d'un mot q^∞-automatique.

En effet, en reprenant les notations du paragraphe III.4), les défauts de commutativité des fonctions $\overline{\sigma}_i$ et $\overline{\sigma}_j$, pour tout couple $(i, j) \in [\![0, q-1]\!]$ jouent un rôle essentiel pour la majoration du cardinal de $\cup_{s \in F_\Pi} \mathcal{Z}_k(s)$, et sur des exemples de q-ADDA de degré borné par K, on peut alors obtenir une meilleure majoration que $(qK)^k$. L'ordre de croissance de $Card(\cup_{s \in F_\Pi} \mathcal{Z}_k(s))$ est lié à la manière dont grandit le graphe lorsqu'on s'éloigne de F_Π par les arcs inverses : lorsqu'il grandit polynomialement, alors on obtiendra une complexité d'ordre inférieur à une fonction du type $n \log_q^a n$. Par contre lorsqu'il croît exponentiellement, on obtient alors une complexité d'ordre inférieur à une fonction du type n^a, avec $a > 1$. Ce résultat peut ensuite être amélioré en utilisant la proposition III.5.3, qui restreint le nombre de facteurs de deux lettres à ceux qui apparaissent effectivement dans m. C'est alors essentiellement le quatrième point de cette proposition, dans lequel intervient la famille de fonctions $\overline{\sigma}_0^k \circ \overline{\sigma}_{i+1} \circ \overline{\sigma}_i^{-1} \circ \overline{\sigma}_q^{-k}$, qui peut permettre d'obtenir une complexité d'ordre inférieur à une fonction du type $n^{1+b} \log_q^c n$. Sur ce point, les défauts d'injectivité de $\overline{\sigma}_q^{-k}$ jouent donc un rôle important.

La question qui vient assez naturellement est celle de savoir s'il existe un mot q^∞-automatique de complexité de même ordre de croissance que la borne $n^{3+2\log_q K}$, pour q et K fixés. Cette question est assez délicate. En effet, la manière efficace pour pouvoir obtenir une minoration de l'ordre de croissance de la complexité est l'utilisation des facteurs spéciaux. Mais minorer le nombre de tels facteurs est plus fastidieux que dans le cas des mots q-automatiques. En effet, il ne suffit pas de trouver les facteurs spéciaux du mot de sortie de l'automate pour lequel il existe deux prolongements différents ; encore faut-il que les projections des facteurs et des lettres qui les prolongent soient elles aussi différentes. La composition par la projection admissible nivèle une grande partie de l'information. Ce comptage nécessite donc une connaissance approfondie de la struture des itérés des états de l'automate par la substitution qui lui est associée (voir les sections IV.3 et VI).

C'est dans cette optique que se situe le chapitre suivant : nous y présentons une étude de la complexité du mot infini d de Dyck sur deux types de parenthèses. Nous allons montrer que la fonction de complexité de d est du même ordre de croissance que la fonction $n\sqrt{n}$.

Chapitre VI

Complexité du mot infini de Dyck sur deux types de parenthèses

Ce chapitre présente un exemple d'étude de la complexité d'un mot 4^∞-automatique engendré par un 4-ADDA de degré borné et monotone. Nous allons montrer que l'ordre de croissance de la fonction de complexité du mot infini de Dyck sur deux types de parenthèses est $n\sqrt{n}$.

VI.1 Cadre

Soit $P = \{p_1, p_0\}$ et $E = P^* \cup \{e_0, x\}$.
On définit, sur $\{p_1, p_0\}^* \cup \{e_0, x\}$, la substitution suivante :

$$\begin{array}{rccc} \overline{\delta}: & E & \to & (P^* \cup \{e_0, x\})^4 \\ & \varepsilon & \mapsto & xx(p_1)(p_0) \\ & e = p_1 v & \mapsto & x(v)(p_1 e)(p_0 e) \\ & e = p_0 v & \mapsto & (v)x(p_1 e)(p_0 e) \\ & x & \mapsto & xxxx \\ & e_0 & \mapsto & e_0 x(p_1)(p_0) \end{array}$$

Soit m le mot infini engendré par le 4-ADDA $\left(E, e_0, \phi_{\overline{\delta}}, \{0,1\}, \Pi\right)$ où la projection $\Pi : E \to \{0,1\}$ est définie par $\Pi^{-1}(\{1\}) = \{\varepsilon, e_0\}$, dont le graphe est représenté à la figure VI.1. Ce 4-ADDA est de degré borné par 2 et monotone pour l'ordre hiérarchique sur P^*. L'unique point fixe $\overline{d}$ de $\overline{\delta}$ est le mot infini suivant :

$$\begin{array}{l} \overline{d} = e_0x(p_1)(p_0)(x)^5\varepsilon(p_1^2)(p_0p_1)\varepsilon x(p_1p_0)(p_0^2)(x)^{22}(p_1)(p_0)x(p_1)(p_1^3)(p_0p_1^2)(p_1)x(p_1p_0p_1)(p_0^2p_1)(x)^2\dots \\ \dots(p_1)(p_0)(x)^5(p_0)(p_1^2p_0)(p_0p_1p_0)(x)^{72}\varepsilon x(p_1^2)(p_0p_1)x\varepsilon(p_1p_0)(p_0^2)(x)^{16}\varepsilon x(p_1^2)(p_0p_1)(x)^4(p_1^2)x(p_1^4)\dots \\ \dots(p_0p_1^3)x(p_1^2)(p_1p_0p_1^2)(p_0^2p_1^2)(x)^4\varepsilon x(p_1^2)(p_0p_1)(p_0p_1)x(p_1^2p_0p_1)x(p_0p_1)(p_1p_0^2p_1)(p_0^3p_1)x^{24}\varepsilon x(p_1^2)\dots \\ \dots(p_0p_1)x\varepsilon(p_1p_0)(p_0^2)x\varepsilon(p_1p_0)(p_0^2)(x)^4\dots \end{array}$$

Le mot infini d est donc donné par $m = \Phi(\overline{d})$:

$$d = 10^7 10^4 10^{118} 10^4 10^3 10^{18} 10^{19} 10^{34} 10^4 10^3 10\dots$$

On appelle d le *mot infini de Dyck sur deux types de parenthèses*. En effet, si on considère le morphisme de monoïdes $\mathcal{D}$ défini par :

$$\begin{array}{rccc} \mathcal{D}: & [\![0,3]\!]^* & \to & \{\overline{p}_1, \overline{p}_0, p_1, p_0\}^* \\ & 0 & \mapsto & \overline{p}_1 \\ & 1 & \mapsto & \overline{p}_0 \\ & 2 & \mapsto & p_0 \\ & 3 & \mapsto & p_1 \end{array}$$

alors, $d_n = 1$ si et seulement si $\mathcal{D}(\overline{n}^4)$ est un mot bien parenthésé sur les deux types de parenthèses p_1 et p_0, c'est-à-dire que $\mathcal{D}(\overline{n}^4)$ appartient au langage de Dyck sur deux types de parenthèses $\mathfrak{D}_2$, qui est un langage algébrique, et en particulier, d est le mot indicateur de l'ensemble $D = \{n \in \mathbb{N},\ \mathcal{D}(\overline{n}^4) \in \mathfrak{D}_2\}$.

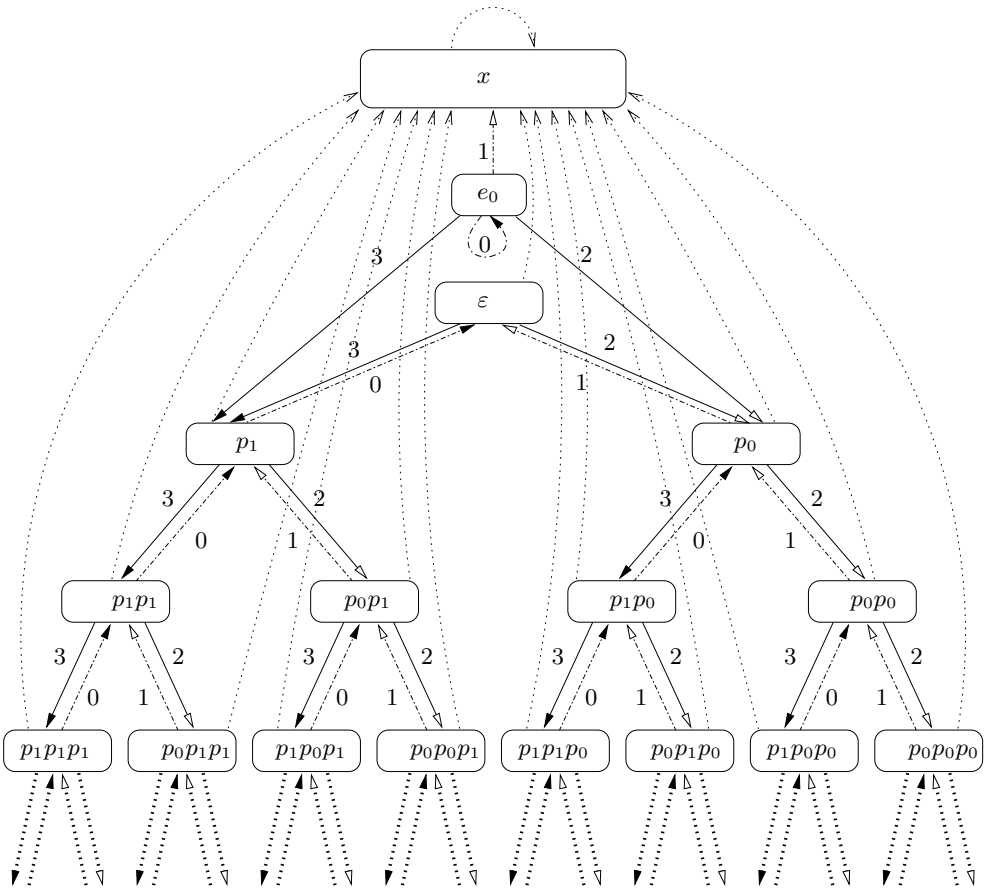


FIG. VI.1 – Graphe du 4-ADD qui engendre le mot infini de Dyck sur deux types de parenthèses.

Le résultat V.1.1 donne $p_d(n) = O(n^5)$, le résultat V.1.3 donne $p_d(n) = O(n^3)$. On va voir qu'en utilisant la méthode utilisée pour montrer le résultat V.1.1, bien qu'il donne, a priori, la plus mauvaise des deux majorations, on peut obtenir un résultat plus fin : $p_m(n) = O(n\sqrt{n})$ et que cette majoration est optimale, dans le sens où la fonction $n\sqrt{n}$ est exactement l'ordre de croissance de la fonction de complexité de d.

Théorème VI.1.1. *L'ordre de croissance de la fonction de complexité de d est* $n\sqrt{n}$.

Pour montrer ce théorème, on va construire deux fonctions de n, équivalentes à une constante fois $n\sqrt{n}$ et qui encadrent p_d. Plus précisemment, on montre les deux résultats suivants :

Proposition VI.1.2.

$$\forall n \geq 1, \quad p_d(n) \leq \frac{20}{9} n(\sqrt{n} - 6\log_4 n + 16),$$

Proposition VI.1.3. *Il existe une constante* C *telle que, pour tout* $n \geq 256$,

$$p_d(n) \geq \frac{8}{42\sqrt{2}} n\sqrt{n} + \frac{n}{9} + C.$$

Les preuves de ces résultats sont l'objet des deux paragraphes suivants.

VI.2 Majoration de la complexité

Pour montrer le résultat VI.1.2, on utilise le schéma de preuve du théorème V.1.1 :

Lemme VI.2.1. *Une infinité d'itérés $\overline{\delta}^k(e)$ se projettent par Π sur 0^{4^k}. Plus précisemment :*

$\overline{\delta}^k(e)$ contient des occurrences de ε si et seulement si $e \in \cup_{i=0^k} P^i$ et $|e|$ est de même parité que k. Pour $k = 2q + p$, avec $p \in \{0, 1\}$ et $q \in \mathbb{N}^*$, il y a donc $\frac{1+p}{3}(2^{k+1} - 1)$ éléments de E tels que $\delta(e)$ est différent de 0^{4^k}.

Ce lemme se montre facilement par récurrence, mais il est facilement "lisible" sur le graphe du q-ADDA $\mathcal{A}$.

Lemme VI.2.2. *Soit $\overline{d}$ le mot infini point fixe de $\overline{\delta}$ contenu dans $e_0E^{\mathbb{N}}$. x_1x_2 est un mot de deux lettres de $\overline{d}$ si et seulement si*

1. $x_1x_2 = (e_0)x$,
2. $\exists e \in P^*$, $x_1x_2 = (e)x$ *ou* $x(e)$,
3. $\exists e \in P^*$, $x_1x_2 = (p_1e)(p_0e)$, *ou* $(e)(p_1^2e)$,
4. $\exists e \in P^*$ *et* $k \geq 1$, $x_1x_2 = (p_0^k p_1 p_0^{k-1} e)(e)$.

■ **Démonstration du lemme VI.2.2** On reprend les différents cas de la proposition III.5.3 en posant $x(p_1)(p_0) = e_1e_2e_3$:

Le mot x_1x_2 est dans $F_2(\overline{d})$ si et seulement si un des 4 cas suivants se produit :

i. $x_1x_2 = e_0x$,

ii. $\exists k > 0, \exists i \in \{1, 2\}$, $x_1 = \overline{\delta}_3^k(e_i)$ et $x_2 = \overline{\delta}_0^k(e_{i+1})$,

iii. $\exists k > 0$, $x_1 = \overline{\delta}_3^k(e_3)$ et $x_2 = \overline{\delta}_0^{k+1}(e_1)$,

iv. $\exists k \geq 0, \exists j \in \{0, 1, 2\}$, $x_1 \in E \setminus \{e_0\}$, $x_2 \in \overline{\delta}_0^k \circ \overline{\delta}_{j+1} \circ \overline{\delta}_j^{-1} \circ \overline{\delta}_3^{-k}(\{x_1\})$.

Les différents mots de longueur 2 de $\overline{d}$, anoncés dans le lemme VI.2.2 apparaîtront au fur et à mesure.

ii. Les mots de $F_2(\overline{d})$ obtenus par cette construction sont $x_1x_2 = xx$, quelque soit $k \geq 0$ lorsque $i = 1$ et $x_1x_2 = (p_0^k p_1)x$, pour tout $k \geq 0$ lorsque $i = 2$.

iii. Les mots de $F_2(\overline{d})$ obtenus par cette construction sont les mots $x_1x_2 = (p_0^k)x$, pour tout $k \geq 1$.

iv. Les mots de $F_2(\overline{d})$ que l'ont peut obtenir par la construction (iv.) paraissent nombreux. Nous allons les scinder en deux modes de construction :

(a) $\exists j \in \{0, 1, 2\}$, $x_1 \in E$ $x_2 \in \overline{\delta}_{j+1} \circ \overline{\delta}_j^{-1} \circ (\{x_1\})$.

(b) $\exists a \in \mathbb{N}^*, \exists j \in \{0, 1, 2\}$, $x_1 \in P^*$ et $x_2 \in \overline{\delta}_0^a \circ \overline{\delta}_{j+1} \circ \overline{\delta}_j^{-1} \circ \overline{\delta}_3^{-a}(\{x_1\})$.

Avant de décrire les mots engendrés par ces deux modes de constructions, il est utile de savoir que $F_1(\overline{d}) = E$ et d'expliciter les différentes valeurs de $\overline{\delta}_{j+1} \circ \overline{\delta}_j^{-1}(\{e\})$ pour $e \in E \setminus \{e_0\}$:

$$\overline{\delta}_1 \circ \overline{\delta}_0^{-1}(\{e\}) = \begin{cases} \{e',\ e' \in P^*\} \cup \{x\} & \text{si } e = x, \\ \{x\} & \text{sinon.} \end{cases}$$

$$\overline{\delta}_2 \circ \overline{\delta}_1^{-1}(\{e\}) = \begin{cases} \{p_1p_0e',\ e' \in P^*\} \cup \{p_1\} \cup \{x\} & \text{si } e = x, \\ \{p_1^2e\} & \text{sinon.} \end{cases}$$

$$\overline{\delta}_3 \circ \overline{\delta}_2^{-1}(\{e\}) = \begin{cases} \{p_0v\} & \text{si } e = p_1v, \\ \{x\} & \text{sinon.} \end{cases}$$

(a). L'expression de $\overline{\delta}_1 \circ \overline{\delta}_0^{-1}$ permet d'obtenir tous les mots de $F_2(\overline{d})$ du type $x_1x_2 = (e)x$ et $x_1x_2 = x(e)$ pour tout élément e de $E \setminus \{e_0, x\}$.

L'expression de $\overline{\delta}_2 \circ \overline{\delta}_1^{-1}$ permet d'obtenir tous les mots de $F_2(\overline{d})$ du type $x_1x_2 = (e)(p_1^2e)$ pour tout élément e de P^*.

L'expression de $\overline{\delta}_3 \circ \overline{\delta}_2^{-1}$ permet d'obtenir tous les mots de $F_2(\overline{d})$ du type $x_1x_2 = (p_1e)(p_0e)$ pour tout élément e de P^*.

(b). Soit a un entier fixé. on cherche l'ensemble des images par les fonctions $\overline{\delta}_0^a \circ \overline{\delta}_{j+1} \circ \overline{\delta}_j^{-1} \circ \overline{\delta}_3^{-a}$ d'un élément e de P^* (tous les mots de deux lettres ayant pour première lettre x on déja été déterminés). Pour pouvoir appliquer $\overline{\delta}_3^{-a}$ il faut impérativement que e se factorise en $e = p_0^a v$ pour un élément $v \in P^*$. Nous discernons des cas différents selon la fonction $\overline{\delta}_0^a \circ \overline{\delta}_{j+1} \circ \overline{\delta}_j^{-1}$ que nous appliquons à $\overline{\delta}_3^{-a}(e) = v$:

Comme

$$\overline{\delta}_0^a \circ \overline{\delta}_1 \circ \overline{\delta}_0^{-1} \circ \overline{\delta}_3^{-a}(\{p_0^a v\}) = \{x\} \text{ et } \overline{\delta}_0^a \circ \overline{\delta}_2 \circ \overline{\delta}_1^{-1} \circ \overline{\delta}_3^{-a}(\{p_0^a v\}) = \{x\},$$

Ces constructions ne peuvent former des mots de $F_2(m)$ dont la deuxième lettre est x.

Si on veut pouvoir appliquer $\overline{\delta}_0^a \overline{\delta}_3 \circ \overline{\delta}_2^{-1}$ à v sans obtenir la lettre x, il faut que $Pref_a(v) = p_1 p_0^{a-1}$. Si on pose $e = p_0^a p_1 p_0^{a-1} u$, alors,

$$\overline{\delta}_0^a \circ \overline{\delta}_3 \circ \overline{\delta}_2^{-1} \circ \overline{\delta}_3^{-a}(\{p_0^a p_1 p_0^{a-1} u\}) = \{u\}$$

On forme donc de cette manière, les mots de $F_2(m)$ du type $x_1x_2 = (p_0^a p_1 p_0^{a-1} e)(e)$ pour $e \in P^*$ et un certain entier $a \geq 1$. ■

■ **Démonstration de la proposition VI.1.2** On va montrer que p_m admet la majoration suivante :

$$\forall n \geq 1, \quad p_m(n) \leq 4n(10\sqrt{n} - 2\log_4 n - 3).$$

Soit $k = 2g + p$ un entier fixé, pour $p \in \{0, 1\}$ et $g \in \mathbb{N}$.

Tout mot de m de longueur 4^k est un facteur d'un certain mot $\delta^k(x_1)\delta^k(x_2)$ avec $x_1x_2 \in F_2(m)$, à un rang compris entre 1 et 4^k.

On va donc, dans un premier temps, compter le nombre de mots $x_1x_2 \in F_2(\overline{d})$ dont la projection pr Π des itérés $\overline{\delta}^k(x_1)$ et $\overline{\delta}^k(x_2)$ ne donne pas deux mots égaux à 0^{4^k}. Le comptage du nombre de ces mots s'effectue en utilisant les lemmes VI.2.1 et VI.2.2. Nous allons les distinguer en deux classes de mots :

1. les mots x_1x_2 dont l'itéré d'une seule des deux lettres x_i donne un mot $\delta^k(x_i)$ non nul par la projection Π, ce qui est typiquement le cas des mots du type $(e)x$ ou $x(e)$ lorsque $e \in P^*$ est ε apparait dans $\overline{\delta}^k(e)$,
2. les mots x_1x_2 pour lequels les deux itérés $\delta^k(x_1)$ et $\delta^k(x_2)$ donnent des mots non nuls par la projection Π, qui peuvent être de plusieurs type :
   (a) $(p_1e)(p_0e)$ pour un certain $e \in P^*$,
   (b) $(e)(p_1^2e)$ pour un certain $e \in P^*$,
   (c) $(p_0^a p_1 p_0^{a-1} e)(e)$ pour un certain $e \in P^*$ et un certain $a \in \mathbb{N}$.

1. Etant donné qu'on a $\frac{1+p}{3}(2^{k+1} - 1)$ éléments e de E dont les itéres $\overline{\delta}^k(e)$ donnent des projections par Π différentes de 0^{4^k}, on peut construire, en supposant qu'a priori elles sont différentes entre elles, $2\frac{1+p}{3}(2^{k+1} - 1)$ mots $\delta^k(x_1)\delta^k(x_2)$ différents de 0^{4^k} et tels que seulement $\delta^k(x_1)$ ou $\delta^k(x_2)$ contienne des occurrences de 1.

2. On va maintenant compter les mots x_1x_2 pour lequels les deux itérés $\delta^k(x_1)$ et $\delta^k(x_2)$ donnent des mots non nuls par la projection Π.

(a) Les mots du type $x_1x_2 = (p_1e)(p_0e)$, pour $e \in P^*$ ne donnent des mots souhaités uniquement lorsque la longueur de e est plus petite que k et est de parité différente de k (voir lemme VI.2.1). Ainsi, on obtient donc $\frac{(2-p)}{3}(2^k - 1)$ mots de deux lettres pour lequels les deux itérés $\delta^k(x_1)$ et $\delta^k(x_2)$ donnent des mots non nuls par la projection Π,

(b) Les mots du type $x_1x_2 = (e)(p_1^2e)$, pour $e \in P^*$ ne donnent des mots souhaités uniquement lorsque la longueur de e strictement plus petite que k et est de même parité que k (voir lemme VI.2.1).

On obtient alors, de cette manière, $\frac{1+p}{3}(2^k - 1)$ mots de deux lettres pour lequels les deux itérés $\delta^k(x_1)$ et $\delta^k(x_2)$ donnent des mots non nuls par la projection Π.

(c) Il ne nous reste plus qu'à compter les mots du type $x_1x_2 = (p_0^a p_1 p_0^{a-1} e)(e)$ pour lequels les deux itérés $\delta^k(x_1)$ et $\delta^k(x_2)$ donnent des mots non nuls par la projection Π :

Soit $e \in P^*$ tel que $|e| = k - 2b$, pour un $b \in [\![1, g]\!]$ fixé. Tous les mots $(p_0^a p_1 p_0^{a-1} e)(e)$, pour $1 \leq a \leq b$, et uniquement eux, ont la propriété souhaitée. Ainsi, lorsqu'on fait varier e dans P^{k-2b} puis b dans $[\![1, g]\!]$, on obtient $N = \sum_{b=1}^{g} b2^{k-2b}$ mots du type $x_1x_2 = (p_0^a p_1 p_0^{a-1} e)(e)$ pour lequels les deux itérés $\delta^k(x_1)$ et $\delta^k(x_2)$ donnent des mots non nuls par la projection Π. Tous calculs faits, on trouve

$$N = \begin{cases} \frac{4}{9}(2^k - \frac{3}{2}k - 1) \text{ si } k \text{ est pair} \\ \frac{1}{9}(2^k - 3k + 1) \text{ si } k \text{ est impair} \end{cases}$$

On retiendra juste que pour tout k, $N \leq \frac{4}{9}(2^k - \frac{3}{2}k + 1)$.

On a donc majoré l'ensemble des mots de deux lettres qui donnent des itérés par δ^k différents de 0^{4^k}. Il suit :

$$p_d(2^k) \leq 4^k \left(\frac{2}{3}(2^{k+1}-1) + \frac{4}{3}(2^k-1) + \frac{4}{9}(2^k - \frac{3}{2}k + 1) \right)$$

et ainsi

$$p_d(2^k) \leq \frac{5}{9}4^k(2^k - 6k + 22) \quad \text{et} \quad p_d(2^{k+1}) \leq \frac{5}{9}4^{k+1}(2^{k+1} - 6k + 16).$$

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 :

$$\forall n \geq 1, \quad p_d(n) \leq \frac{20}{9} n(\sqrt{n} - 6\log_4 n + 16),$$

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

VI.3 Minoration de la complexité

La minoration de la complexité de d nécessite d'explorer d'avantage la structure des $\delta^k(w)$ qui se projettent vers des mots différents de 0^{4^k}, pour un entier k fixé. En effet, cela va nous permettre de localiser des facteurs spéciaux de longueur donnée et grâce à ces facteurs, nous pourrons minorer la complexité de d.

Notations VI.3.1. Etant donné que nous allons un peu jongler entre les entiers et leurs représentations en base 4, nous introduisons la notation suivante :

Si w est un mot de $[\![0,3]\!]^*$, on note $[w]_4$ l'entier pour lequel w est une représentation en base 4 est w (même si la représentation n'est pas propre) et $\overline{\delta}_w$ l'application définie par :

$$\overline{\delta}_w = \overline{\delta}_{w_{|w|-1}} \circ \cdots \circ \overline{\delta}_{w_1} \circ \overline{\delta}_{w_0}.$$

Si e est un mot de P^*, On note $\tilde{e}$ le mot de $\{0,1\}^*$ obtenu à partir de e par application du morphisme de monoïdes défini par $p_i \mapsto i$ et $\hat{e}$ le mot de $\{2,3\}^*$ obtenu à partir de e par application du morphisme de monoïdes défini par $p_i \mapsto 3-i$ et de l'image miroir, c'est-à-dire $\hat{e}_i = 3 - \tilde{e}_{|e|-i-1}$.

Remarques VI.3.2. Pour $e \in P^*$, le mot $\hat{e}$ est le mot de longueur minimale tel que $\overline{\delta}\hat{e}(\varepsilon) = e$. Il prend ses valeurs dans $\{2,3\}^*$.

D'autre part, le mot $\tilde{e}$ est le facteur de $\{0,1\}^*$ qui vérifie $\mathcal{D}(\hat{e})\mathcal{D}(\tilde{e})$ est le mot de Dyck de longueur minimale dont le préfixe est $\hat{e}$.

Par exemple, si $e = p_0p_0p_1p_0p_1$, alors on a $\tilde{e} = 00101$ et $\hat{e} = 23233$ et le mot $\mathcal{D}(\hat{e})\mathcal{D}(\tilde{e}) = p_1p_0p_1p_0p_0\overline{p}_0\overline{p}_0\overline{p}_1\overline{p}_0\overline{p}_1$ est un mot de Dyck.

On rappelle que $d_n = 1$ si et seulement si $\mathcal{D}(\overline{n}^4)$ appartient au langage de Dyck $\mathfrak{D}_2$ sur deux types de parenthèses p_0 et p_1, et lorsque $n \neq 0$, on a :

$$d_n = 1 \iff \overline{\delta}_{\overline{n}^4}(\varepsilon) = \varepsilon.$$

Lemme VI.3.3. *Soient e un mot de P^* et w un mot de $[\![0,3]\!]^*$.*

Le mot $\mathcal{D}(\hat{e}w)$ est un mot de $\mathfrak{D}_2$ si et seulement s'il existe $(|e|+1)$ mots $w^{(0)}$, $w^{(1)}$, …, $w^{(|e|)}$ de $[\![0,3]\!]^*$ tels que, pour tout $i \in [\![0,|e|]\!]$, $\mathcal{D}(w^{(i)}) \in \mathfrak{D}_2$ et

$$w = w^{(0)}\tilde{e}_0 w^{(1)}\tilde{e}_1 \dots w^{(|e|-1)}\tilde{e}_{|e|-1}w^{(|e|)}.$$

Proposition VI.3.4. *Soient $k \geq 3$ et $e \in P^{k-2b}$ pour un certain $b \in [\![0, \lfloor \frac{k}{2} \rfloor]\!]$.*

Si $|e| = k$ ($b = 0$), alors le mot $\delta^k(w)$ contient une unique occurrence de 1 au rang $[\tilde{e}]_4$.

Si $|e| < k$ ($b \neq 0$) , le mot $\delta^k(e)$ contient au moins deux occurrences de 1. De plus, le nombre d'occurence de 1 est uniquement dépendant de b.

Plus précisemment, pour tout n de $[\![0, 4^k - 1]\!]$, $\delta^k(e)_n = 1$ si et seulement s'il existe $(k-2b+1)$ mots $w^{(0)}$, $w^{(1)}$, …, $w^{(k-2b)}$ de $[\![0,3]\!]^*$ tels que $|w^{(0)}w^{(1)}w^{(k-2b)}| = 2b$ et vérifiant :

$$\forall i \in [\![0, k-2b]\!], \ \mathcal{D}(w^{(i)}) \in \mathfrak{D}_2 \quad et \quad n = [w^{(0)}\tilde{e}_0 w^{(1)}\tilde{e}_1 \dots w^{(k-2b-1)}\tilde{e}_{k-2b-1}w^{(k-2b)}]_4.$$

■ **Démonstration de la proposition VI.3.4** Soient $b \in [\![0, \lfloor \frac{k}{2} \rfloor]\!]$ et $e \in P^{k-2b}$. On commence par remarquer que $\overline{\delta}_{\tilde{e}}(e) = \varepsilon$. C'est d'ailleurs l'unique façon de composer $k-2b$ fois les δ_i pour obtenir ε (c'est l'unique chemin de longueur $k-2b$ qui relie e à ε).

Soit un mot e de P^{k-2b} pour un certain entier $b \in [\![1, \lfloor \frac{k}{2} \rfloor]\!]$.

Soit $n \in [\![0, 4^k - 1]\!]$ et $w \in [\![0,3]\!]^*$ défini par $w = 0^{|k-|\overline{n}^q|}\overline{n}^q$. On a, d'après la proposition III.5.1 et la remarque VI.3.2 :

$$\delta^k(e)_n = 1 \Longleftrightarrow \overline{\delta}_w(e) = \varepsilon \Longleftrightarrow \overline{\delta}_w\left(\overline{\delta}_{\hat{e}}(\varepsilon)\right) = \varepsilon \Longleftrightarrow \overline{\delta}_{\hat{e}w}(\varepsilon) = \varepsilon.$$

Ainsi, $\overline{\delta}_w(e) = \varepsilon$ si et seulement si $\mathcal{D}(\hat{e}w)$ est un mot de $\mathfrak{D}_2$.

Le lemme VI.3.3 donne alors la forme générale de w : il existe un $(k-2b+1)$-uplet de mots $w^{(0)}$, $w^{(1)}$, $\ldots$, $w^{(k-2b)}$ de $[\![0,3]\!]^*$ tels que $|w^{(0)}w^{(1)}w^{(k-2b)}| = 2b$ et vérifiant pour tout i de $[\![0, k-2b]\!]$, $\mathcal{D}(w^{(i)}) \in \mathfrak{D}_2$, et $w = w^{(0)}\tilde{e}_0 w^{(1)}\tilde{e}_1 \ldots w^{(k-2b-1)}\tilde{e}_{k-2b-1}w^{(k-2b)}$.

Ainsi, $\delta^k(e)_n = 1$ si et seulement si n est un entier du type

$$n = [w^{(0)}\tilde{e}_0 w^{(1)}\tilde{e}_1 \ldots w^{(k-2b-1)}\tilde{e}_{k-2b-1}w^{(k-2b)}]_4.$$

En fait, cela se voit très bien sur le graphe de l'automate car le chemin de longueur k de e vers ε contruit à partir du chemin indexé par $\tilde{e}$ en intercalant les boucles indexées par les $w^{(j)}$ entre les $\tilde{e}_i$ est le chemin indéxé par $w^{(0)}\tilde{e}_0 w^{(1)}\tilde{e}_1 \ldots w^{(k-2b-1)}\tilde{e}_{k-2b-1}w^{(k-2b)}$.

L'ensemble, des $(k-2b+1)$-uplets de mots $(w^{(0)}, w^{(1)}, \ldots, w^{(k-2b)})$ de $([\![0,3]\!]^*)^{k-2b+1}$ tels que $|w^{(0)}w^{(1)}w^{(k-2b)}| = 2b$ et vérifiant pour tout i de $[\![0, k-2b]\!]$, $\mathcal{D}(w^{(i)}) \in \mathfrak{D}_2$ indexe donc l'ensemble des rangs d'occurence de 1 dans les $\delta^k(e)$ pour tout élément e de P^{k-2b} et cela, de manière indépendante du mot e choisi. Ainsi, tous les itérés par δ^k des éléments de P^* de même longueur ont le même nombre d'occurrences de 1. ■

Nous allons nous intéresser plus en détails aux ensembles des $(k-2b+1)$-uplets de mots de $[\![0,3]\!]^*$, notés $W = (w^{(0)}, w^{(1)}, \ldots, w^{(k-2b)})$, tels que $|w^{(0)}w^{(1)}w^{(k-2b)}| = 2b$ et vérifiant pour tout i de $[\![0, k-2b]\!]$, $\mathcal{D}(w^{(i)}) \in \mathfrak{D}_2$, car on va voir (lemme VI.3.7) que l'on peut indexer l'ensemble des occurrences de 1 dans les itérés grâce à ces ensembles.

Définition VI.3.5. Soient un entier $k \geq 2$ et un entier $b \in [\![1, \lfloor \frac{k}{2} \rfloor]\!]$.

On note $\mathcal{U}_k(b)$ l'ensemble des $(k-2b+1)$-uplets de mots de $[\![0,3]\!]^*$, notés $W = (w^{(0)}, w^{(1)}, \ldots, w^{(k-2b)})$, tels que $|w^{(0)}w^{(1)}w^{(k-2b)}| = 2b$ et vérifiant pour tout i de $[\![0, k-2b]\!]$, $\mathcal{D}(w^{(i)}) \in \mathfrak{D}_2$.

On ordonne l'ensemble $\mathcal{U}_k(b)$ à partir de l'ordre hiérarchique $(<)$ sur $[\![0,3]\!]^*$:

$W = (w^{(0)}, w^{(1)}, \ldots, w^{(k-2b)}) < V = (v^{(0)}, v^{(1)}, \ldots, v^{(k-2b)})$ si et seulement si :

$$\exists i_0 \in [\![0, k-2b]\!],\ \forall i < i_0, w^{(i)} = v^{(i)} \text{ et } w^{(i_0)} < v^{(i_0)}.$$

Si on note $\mathcal{U}_k(b) = \{W_1 < W_2 < \cdots < W_{Card(\mathcal{U}_k(b))}\}$, on définit, pour tout $i \in [\![0, k-2b]\!]$, l'entier $N_i \leq Card(\mathcal{U}_k(b))$, dépendant de k et b, tel que $W_{N_i} = (\underbrace{\varepsilon, \ldots, \varepsilon}_{i}, 3^b0^b, \varepsilon, \ldots, \varepsilon)$.

Remarque VI.3.6. Pour l'ordre que nous avons définit sur $\mathcal{U}_k(b)$, pour $k \geq 2$ et $b \in [\![1, \lfloor \frac{k}{2} \rfloor]\!]$, le plus petit élément W_1 est $(\varepsilon, \ldots, \varepsilon, (21)^b)$, le plus grand élément $W_{Card(\mathcal{U}_k(b))}$ est $W_{N_0} = (3^b0^b, \varepsilon, \ldots, \varepsilon)$ et le successeur de $W_{N_i} = (\underbrace{\varepsilon, \ldots, \varepsilon}_{i}, 3^b0^b, \varepsilon, \ldots, \varepsilon)$ dans $\mathcal{U}_k(b)$ est $W_{N_i+1} = (\underbrace{\varepsilon, \ldots, \varepsilon}_{i-1}, 21, \varepsilon, \ldots, \varepsilon, (21)^{b-1})$.

D'autre part, les entiers N_i dépendent uniquement de k et de b et, lorsque k est fixé, N_i croît avec b. Ils nous seront utiles par la suite pour déterminer les facteurs spéciaux.

D'autre part, nous tenons à souligner le fait que la valeur exacte de $Card(\mathcal{U}_k(b))$ ne sera pas nécéssaire par la suite. Cependant, on peut montrer que le nombre de mots bien parenthésés avec deux types de parenthèses de longueur $2n$ est $2^n\mathcal{C}_n$ où $\mathcal{C}_n = \frac{1}{n+1}\binom{2n}{n}$ est le n-ième nombre de Catalan. Et en particulier, on a :

$$N_{k-2b-1} = 2^b\mathcal{C}_b,$$

$$\forall i \in [\![0, k-2b]\!],\ N_i - N_{i+1} = Card(\mathcal{U}_{k-i}(b)),$$

$$N_0 = Card(\mathcal{U}_k(b)) = 2^b \sum_{i_0+\cdots+i_{k-2b}=b} \prod_{j=0}^{k-2b} \mathcal{C}_{i_j}.$$

Lemme VI.3.7. *Soient un entier $k \geq 2$ et un entier $b \in [\![1, \lfloor \frac{k}{2} \rfloor]\!]$.*

Soint un couple d'éléments $W = (w^{(0)}, w^{(1)}, \ldots, w^{(k-2b)})$ et $V = (v^{(0)}, v^{(1)}, \ldots, v^{(k-2b)})$ de $\mathcal{U}_k(b)$ et un élément e de P^{k-2b} :

On a $W < V$ si et seulement si :

$$[w^{(0)}\tilde{e}_0 w^{(1)}\tilde{e}_1 \ldots w^{(k-2b-1)}\tilde{e}_{k-2b-1}w^{(k-2b)}]_4 < [v^{(0)}\tilde{e}_0 v^{(1)}\tilde{e}_1 \ldots v^{(k-2b-1)}\tilde{e}_{k-2b-1}v^{(k-2b)}]_4.$$

Ainsi, si $W_n = (w^{(0)}, w^{(1)}, \ldots, w^{(k-2b)})$, le rang d'apparition du n-ième 1 de $\delta^k(e)$ est donné par l'entier $[w^{(0)}\tilde{e}_0 w^{(1)}\tilde{e}_1 \ldots w^{(k-2b-1)}\tilde{e}_{k-2b-1}w^{(k-2b)}]_4$.

■ **Démonstration du lemme VI.3.7** Ce lemme découle du fait que le mot $\tilde{e}$ est un élément de $\{0,1\}^*$ et que les mots non nuls de $[\![0,3]\!]^*$ dont les images par $\mathcal{D}$ sont des mots de Dyck commencent par 2 ou 3.

Soit un élément e de P^{k-2b} et deux éléments $(w^{(0)}, w^{(1)}, \ldots, w^{(k-2b)})$ et $(v^{(0)}, v^{(1)}, \ldots, v^{(k-2b)})$ de $\mathcal{U}_k(d)$. On pose

$$w = w^{(0)}\tilde{e}_0 w^{(1)}\tilde{e}_1 \ldots w^{(k-2b-1)}\tilde{e}_{k-2b-1}w^{(k-2b)} \text{ et } v = v^{(0)}\tilde{e}_0 v^{(1)}\tilde{e}_1 \ldots v^{(k-2b-1)}\tilde{e}_{k-2b-1}v^{(k-2b)}.$$

Supposons que $[w]_4 < [v]_4$, il existe donc un entier minimal $n_0 \in [\![0, k-1]\!]$ tel que pour tout $n < n_0$, $w_n = v_n$ et $w_{n_0} < v_{n_0}$. On peut, sans perte de généralités supposer, que $n_0 \leq \max(|w^{(0)}|, |v^{(0)}|)$.

D'autre part, on ne peut pas avoir $n_0 > |v^{(0)}|$. En effet, si $n_0 > |v^{(0)}|$, alors $v^{(0)}\tilde{e}$ est un préfixe strict de $w^{(0)}$ puisque les images par $\mathcal{D}$ des mots $w^{(0)}$ et $v^{(0)}$ sont des mots de Dyck et ils sont donc tous les deux de longueur paire. Mais cela implique également que les préfixes stricts de longueur impaire de $w^{(0)}$ contiennent strictement plus de lettres 2 que de lettres 1 et strictement plus de lettres 3 que de lettres 0, et $v^{(0)}\tilde{e}$ ne vérife pas ces conditions puisque $\tilde{e}$ appartient à $\{0,1\}$.

Ainsi, $n_0 \leq |v^{(0)}|$, et donc $w^{(0)} < v^{(0)}$ pour l'ordre hiérarchique sur $[\![0,3]\!]^*$, et ainsi, si $[w]_4 < [v]_4$, alors

$$(w^{(0)}, w^{(1)}, \ldots, w^{(k-2b)}) \leq (v^{(0)}, v^{(1)}, \ldots, v^{(k-2b)}).$$

La réciproque se montre aisément et la deuxième partie du lemme découle alors de la proposition VI.3.4 et ce que nous venons de démontrer. ■

Lemme VI.3.8. *Pour tout élément e de P^*, le mot $(x)(x)(e)$ est un mot de $\overline{\delta}$.*

En effet, pour tout $e \in P^*$, le mot $((x)(p_1 e)$ est un mot de $\overline{d}$. Lorsqu'on applique la substitution $\overline{\delta}$ à ce mot de deux lettres, on obtient le mot $(x)^4(e)x(p_0 p_1 e)(p_1^2 e)$, qui est aussi un mot de $\overline{d}$. On peut alors en extraire le mot $(x)(x)(e)$, qui est donc également un mot de $\overline{d}$.

Nous sommes à présents prêts pour isoler des facteurs spéciaux à droite de d, afin de minorer la fonction de complexité de d. Soit k un entier fixé, et soit n un entier tel que $2^k \leq n \leq 2^{k+1} - 1$.

L'idée de la preuve consiste à exhiber un certain nombre de facteurs spéciaux à droite de longueur n du mot d, pour ainsi minorer la quantité $p_d(n+1) - p_d(n)$, quelque soit n. En sommant ces inégalités, on obtient ensuite une minoration de la fonction p_d.

Pour exhiber ces mots spéciaux, nous allons exploiter le fait que tout mot w de d de longueur $4^k \leq n < 4^{k+1}$ est contenu dans un itéré $\delta^k(x_0)\delta^k(x_1)\delta^k(x_2)$ pour un certain mot $x_0x_1x_2$ de d. Les couples de facteurs du type $(x)(x)(p_1 e)$ et $(x)(x)(p_0 e)$ nous fournirons les facteurs spéciaux à droite. Le travail précédent concernant les ensembles $\mathcal{U}_k(b)$ permettra d'assurer que les images obtenues par la projection Π de ces facteurs sont différrentes.

■ **Démonstration de la proposition VI.1.3**

On fixe l'entier $k \geq 3$ et un entier n vérifiant $4^k \leq n < 4^{k+1}$. Avant de construire une famille de facteurs spéciaux de longueur n, on commence par remarquer que le mot 0^n est un facteur spécial à droite de d. En effet, 0^n est un mot que l'on peut extraire de $\delta^k((x)(x))$, donc c'est un mot de d que l'on peut prolonger à droite par 0, mais on peut aussi l'extraire de $\delta^k((x)(x)(e))$ pour tout e de P^k, puisque le préfixe de longueur 4^{k+1} de $\delta^k((x)(x)(e))$ est le mot $0^{4^{k+1}}$. Comme $\delta^k(e)$ contient une seule occurence de 1, qui apparaît au rang $[\tilde{e}]$, le mot extrait de $\delta^k((x)(x)(e))$ dont la première lettre est la $(24^k + [\tilde{e}]_4 - 1)$-ième lettre de $\delta^k((x)(x)(e))$, est le mot 0^n est il s'y prolonge à gauche par 1. Ainsi, 0^n0 et 0^n1 sont des mots de d donc 0^n est un facteur spécial à droite de d.

Construisons maintenant d'autres facteurs spéciaux. Pour cela, on fixe un entier $b \in [\![1, \lfloor \frac{k}{2} \rfloor]\!]$ et un élément e de P^{k-2b-1}.

Grâce au lemme VI.3.7, les rangs d'apparition des occurences de 1 dans $\delta^k(p_0e)$ sont donnés par l'ensemble ordonné :

$$\mathcal{R}_k(0e) = \{[w^{(0)}0w^{(1)}\tilde{e}_1 \dots w^{(k-2b-1)}\tilde{e}_{k-2b-1}w^{(k-2b)}]_4, (w^{(0)}, w^{(1)}, \dots, w^{(k-2b)}) \in \mathcal{U}_k(b)\},$$

et les rangs d'apparition des occurences de 1 dans $\delta^k(p_0e)$ sont donnés par l'ensemble ordonné :

$$\mathcal{R}_k(1e) = \{[w^{(0)}1w^{(1)}\tilde{e}_1 \dots w^{(k-2b-1)}\tilde{e}_{k-2b-1}w^{(k-2b)}]_4, (w^{(0)}, w^{(1)}, \dots, w^{(k-2b)}) \in \mathcal{U}_k(b)\}.$$

Ainsi, si $r \leq N_1$ (voir la définition VI.3.5) et $(\varepsilon, w^{(1)}, \dots, w^{(k-2b)})$ est le r-ième élément de $\mathcal{U}_k(b)$, les rangs des r-ièmes apparitions de 1 dans $\delta^k(0e)$ et $\delta^k(1e)$ sont donnés respectivement par les entiers $[0w^{(1)}\tilde{e}_1 \dots w^{(k-2b-1)}\tilde{e}_{k-2b-1}w^{(k-2b)}]_4$ et $[1w^{(1)}\tilde{e}_1 \dots w^{(k-2b-1)}\tilde{e}_{k-2b-1}w^{(k-2b)}]_4$.

En particulier, les rangs des N_1-ièmes apparitions de 1 dans $\delta^k(0e)$ et $\delta^k(1e)$ sont donnés respectivement par $[03^b0^b\tilde{e}]_4$ et par $[13^b0^b\tilde{e}]_4$.

Les rangs d'apparitions des N_1 premieres occurences de 1 dans $\delta^k(1e)$ se déduisent donc des rangs d'apparitions de N_1 premieres occurences de 1 dans $\delta^k(0e)$ en leur rajoutant 4^{k-1}, car

$$[1w^{(1)}\tilde{e}_1 \dots w^{(k-2b-1)}\tilde{e}_{k-2b-1}w^{(k-2b)}]_4 - [0w^{(1)}\tilde{e}_1 \dots w^{(k-2b-1)}\tilde{e}_{k-2b-1}w^{(k-2b)}]_4 = 4^{k-1}.$$

Par contre, les rangs des (N_1+1)-ièmes apparitions de 1 dans $\delta^k(0e)$ et $\delta^k(1e)$ sont donnés respectivement par $[210\tilde{e}(21)^{b-1}]_4$ et par $[211\tilde{e}(21)^{b-1}]_4$. On a, en particulier :

$$[211\tilde{e}(21)^{b-1}]_4 - [210\tilde{e}(21)^{b-1}]_4 = 4^{k-3}.$$

Donc le rang de la (N_1+1)-ième apparition de 1 dans $\delta^k(1e)$ s'obtient en rajoutant 4^{k-3} à la valeur du rang de la (N_1+1)-ième apparition de 1 dans $\delta^k(0e)$.

On va fabriquer un mot spécial que l'on peut extraire de $\delta^k(0e)$ et de $\delta^k(1e)$ en exploitant cette propriété :

Pour tout $e \in P^{k-2b-1}$, on définit le mot $u^{(e)}$ extrait de $\delta^k(x)\delta^k(x)\delta^k(p_1e)$ de sorte que la dernière lettre de $u^{(e)}$ soit la $([211\tilde{e}(21)^{b-1}]_4 - 1)$-ième lettre de $\delta^k(p_1e)$, c'est-à-dire (voir figure VI.2) que si on pose $\delta^k(x)\delta^k(x)\delta^k(p_1e) = v_0v_1 \dots v_{34^k-1}$, alors

$$u^{(e)} = v_{24^k+[211\tilde{e}(21)^{b-1}]_4-1-n} \cdots v_{24^k+[211\tilde{e}(21)^{b-1}]_4-2}$$

Les mots $u^{(e)}$ sont tous des mots spéciaux à droite. En effet, comme $(x)(x)(p_0e)$ et $(x)(x)(p_1e)$ ap-

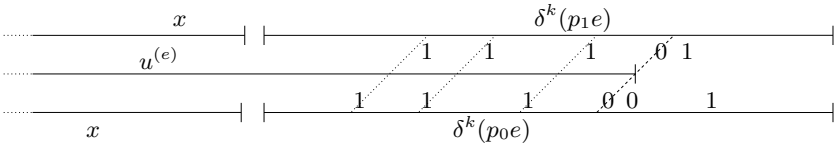


FIG. VI.2 – Construction du facteur spécial $u^{(\tilde{e})}$.

paraîssent dans la suite, le mot $u^{(e)}$ peut se prolonger à droite par 1 si on le voit comme sous-mot de $\delta^k(x)\delta^k(x)\delta^k(p_1e)$ ou, d'après ce que l'on vient de faire, par 0 si on le considère comme sous-mot de $\delta^k(x)\delta^k(x)\delta^k(p_0e)$.

Cependant, pour pouvoir minorer la fonction $p_d(n+1)-p-d(n)$, il faut être sûr que les mots spéciaux ainsi formés sont différents lorsque l'on fait varier $e \in P^{k-2b-1}$.

Il faut donc que le nombre et la répartition des occurrences de 1 dans $u^{(e)}$ caractérise entièrement l'élément e. C'est effectivement le cas. En effet, le nombre d'occurence de 1 dans $u^{(e)}$ est N_i. L'entier N_i étant injectivement lié à b lorsque k est fixé, il impose b et il détermine la longueur de e : $k-2b-1$. D'autre part, une fois b déterminé, la longueur du bloc de 0 qui termine $u^{(e)}$ est donnée par $[211\tilde{e}(21)^{b-1}]_4 - [13^b0^b\tilde{e}]$, cela permet donc de déterminer entièrement $[\tilde{e}]_4$. On en déduit donc e comme l'image par le morphisme de $\{0,1\}^*$ vers $\{p_0,p_1\}^*$ défini par $i \mapsto p_i$ de l'élément $0^a\overline{[\tilde{e}]_4}^4$, où $a = k-2b-1-|\overline{[\tilde{e}]_4}^4|$ (le mot $\overline{[\tilde{e}]_4}^4$ est complété à gauche par suffisamment de zéros pour qu'il soit de longueur $k-2b-1$).

Ainsi, on a exhibé un ensemble de facteurs spéciaux à droite de longueur n, en bijection avec l'ensemble des éléments e de $\cup_{b\in[\![1,\lfloor\frac{k}{2}\rfloor]\!]}P^{k-2b-1}$. On a donc au minimum $\frac{1}{3}(2^k-1)$ facteurs spéciaux de longueur n, d'où

$$\forall k\geq 3, \forall n\in[\![4^k,4^{k+1}-1]\!],\ p_d(n+1)-p_d(n)\geq\frac{1}{3}(2^k-1).$$

En sommant ces inégalités, on obtient :

$$\forall k\geq 4, \forall n\in[\![4^k,4^{k+1}-1]\!], p_d(n)-p_d(16)\geq\frac{1}{3}(n-4^k)(2^k-1)+\frac{1}{3}\sum_{i=3}^{k-1}4^i(2^i-1).$$

et il suit, pour tout $k\geq 4$ et tout $n\in[\![2^k,2^{k+1}-1]\!]$:

$$p_d(n)\geq p_d(16)+\frac{1}{3}\left(n(2^k-1)-4^k(2^k-1)+\frac{1}{7}(2^k4^k-512)-\frac{1}{3}(4^k-8)\right).$$

En minorant 2^k par $\frac{\sqrt{n}}{\sqrt{2}}$ et 4^k par $\frac{n}{2}$, on obtient, pour tout $n\geq 256$,

$$p_d(n)\geq\frac{8}{42\sqrt{2}}n\sqrt{n}+\frac{n}{9}+p_d(16)-\frac{1480}{63}.$$

Ainsi, il existe une constante C telle que, pour tout $n\geq 256$,

$$p_d(n)\geq\frac{8}{42\sqrt{2}}n\sqrt{n}+\frac{n}{9}+C.$$

le résultat VI.1.3 est donc démontré. ■

■ **Démonstration du théorème VI.1.1** Les résultats VI.1.2 et VI.1.3 permettent l'encadrement suivant de la complexité :

$$\forall n\in\mathbb{N}, n\geq 256,\ \frac{8}{42\sqrt{2}}n\sqrt{n}+\frac{n}{9}+C\leq p_d(n)\leq 4n(10\sqrt{n}-2\log_4 n-3).$$

les deux fonctions qui encadrent la complexité du mot m sont toutes les deux équivalentes en l'infini à une constante fois $n\sqrt{n}$. On en déduit donc que la fonction $p_d(n)$ est donc de l'ordre de croissance de $n\sqrt{n}$. ■

Chapitre VII

Commentaires et perspectives

VII.1 À propos des mots q^∞-automatiques engendrés par des q-ADDA de degré non borné

Lorsqu'on cherche à majorer la fonction de complexité d'un mot q^∞-automatique m engendré par un q-ADDA de degré non borné et non monotone, il est parfois possible d'adapter la méthode de la démonstration du théorème V.1.1 afin d'obtenir une majoration de l'ordre de la fonction de complexité de m. D'autre part, certains mots obtenus grâce à des q-ADDA de degré infini peuvent présenter des dégénérescence. Nous explorons ces problèmes dans ce paragraphe.

VII.1.1 Majoration de la complexité d'un mot q^∞ engendré par un q-ADDA de degré infini non monotone

Soit $\overline{\varphi}$ la substitution suivante

$$\begin{array}{rccc} \overline{\varphi}: & \mathbb{N}\cup\{-1\} & \to & \mathbb{Z}^3 \\ & e & \mapsto & \left(\lfloor\frac{e}{2}\rfloor\right)\left((-1)^e\right)(e+1) \end{array}$$

La fonction $e \mapsto \lfloor e \rfloor$ représente la partie entière de e, pour $e \geq 0$ et avec la convention $\lfloor\frac{-1}{2}\rfloor = 0$.

Proposition VII.1.1. *Soit le m mot le mot engendré par le 2-ADDA $\mathcal{A} = (\mathbb{N}\cup\{-1\}, \phi_{\overline{\varphi}}, 0, \{0,1\}, \Pi_0)$, où Π_0 est la projection admissible qui vérifie $\Pi_0^{-1}(\{1\}) = \{0\}$.*

La fonction de complexité de m vérifie :

$$p_m(n) = O(n^{1+2\log_3 2}).$$

Lemme VII.1.2.
Le mot e_1e_2 est dans $F_2(\overline{m})$ si et seulement si un des 3 cas suivants se produit :

1. $e_1e_2 \in \mathbb{N}^2$,
2. $e_1 = -1$ *et* $e_2 \in 2\mathbb{N}$,
3. $e_1 \in \mathbb{N}$ *et* $e_2 = -1$.

■ **Démonstration du lemme VII.1.2**

On rappelle que $\overline{m} = 0\overline{\varphi}(0)\overline{\varphi}^2(0)\ldots\overline{\varphi}^k(0)\ldots$. Quleque soit k, le mot $\overline{\varphi}^k(0)$ se termine par $\left((-1)^k\right)k$ (conséquence de la proposition III.5.1). Donc les mots $(-1)(2p)$ et $1(2p+1)$ sont des mots du language de $\overline{m}$. La lettre (-1) n'apparaissant qu'au milieu d'un $\overline{\varphi}(e)$, le mot $(-1)(2p)$ se prolonge à gauche en $p(-1)(2p)$, et cela, quelque soit p. On a donc décrit tous les mots de $\overline{m}$ de 2 lettres qui contiennent la lettre (-1).

On va maintenant voir que tous les mots de 2 lettres choisies dans $\mathbb{N}$ sont des mots de $\overline{m}$. Comme $(-1)(2p)$ apparait dans $\overline{m}$, $\overline{\varphi}(-1)\overline{\varphi}(2p)$ est aussi un mot de $\overline{m}$. Le mot de 2 lettres formant la jonction de $\overline{\varphi}(-1)$ et $\overline{\varphi}(2p)$ est le mot $0p$. Donc, quelque soit $p \in \mathbb{N}$, $0p$ est un mot de $\overline{m}$.

Soit $e_1e_2 \in \mathbb{N}^2$. Le mot $0\left(e_2 2^{e_1}\right)$ apparaît dans $\overline{m}$ d'après ce que l'on vient de faire et, par conséquent, tous les itérés par $\overline{\varphi}$ de ce mot sont aussi des mots de $\overline{m}$. Le mot de 2 lettres qui se trouve à la jonction de $\overline{\varphi}^{e_1}(0)\overline{\varphi}^{e_1}\left(e_2 2^{e_1}\right)$ est exactement e_1e_2. ■

Lemme VII.1.3. *Soit $k \geq 2$ fixé.*

Quelque soit e, $\varphi^k(2e)$ et $\varphi^k(2e+1)$ ont le même préfixe de longueur $\frac{3^k-1}{2}$.

De plus,

$$\forall e, e' > 2^k + 1, \quad e = e' \mod 2^k \Longrightarrow \varphi^k(e) = \varphi^k(e').$$

Par conséquent, quel que soit le mot w de longueur 3^k, il est contenu dans un mot de longueur 3^{k+1} du type $\overline{\varphi}^k(e_1)\overline{\varphi}^k(e_2)$, où le couple (e_1, e_2) prend ses valeurs dans l'ensemble suivant :

$$\mathcal{U}_k = [\![0, 2^{k+1}+1]\!]^2 \bigcup [\![0, 2^{k+1}+1]\!] \times \{-1\} \bigcup \{-1\} \times [\![0, 2^{k+1}]\!] \cap 2\mathbb{N}$$

■ **Démonstration du lemme VII.1.3**

La première partie du lemme vient du fait que, pour $k \geq 2$, $\overline{\varphi}^k(2e)$ et $\overline{\varphi}^k(2e+1)$ ont le préfixe suivant en commun :

$$\overline{\varphi}^{k-1}(e)\overline{\varphi}^{k-2}(0)\overline{\varphi}^{k-2}(0)\ldots\overline{\varphi}(0)0.$$

et ce préfixe est suivit d'un 1 dans $\overline{\varphi}^k(2e)$ et d'un -1 dans $\overline{\varphi}^k(2e+1)$. Cela se montre facilement par récurrence sur k :

En effet, $\overline{\varphi}^2(2e) = \overline{\varphi}(e)\overline{\varphi}(1)\overline{\varphi}(2e+1) = \overline{\varphi}(e)012\overline{\varphi}(2e+1)$ et $\overline{\varphi}^2(2e+1) = \overline{\varphi}(e)\overline{\varphi}(-1)\overline{\varphi}(2e+1) = \overline{\varphi}(e)0(-1)0\overline{\varphi}(2e+2)$ et le préfixe commun $\overline{\varphi}(e)0$ est suivit d'un 1 dans $\overline{\varphi}^2(2e)$ et d'un -1 dans $\overline{\varphi}^2(2e+1)$. Si, pour un certain $k \geq 2$, $\overline{\varphi}^k(2e)$ et $\overline{\varphi}^k(2e+1)$ ont le préfixe $\overline{\varphi}^{k-1}(e)\overline{\varphi}^{k-2}(0)\overline{\varphi}^{k-2}(0)\ldots\overline{\varphi}(0)0$ en commun et ce préfixe est suivit d'un 1 dans $\overline{\varphi}^k(2e)$ et d'un -1 dans $\overline{\varphi}^k(2e+1)$, il suffit d'appliquer $\overline{\varphi}$ à $\overline{\varphi}^k(2e)$ et à $\overline{\varphi}^k(2e+1)$ pour obtenir la propriété pour $k+1$.

L'application de la projection Π_0 préservant l'égalité des préfixes, on obtient :

$$\forall e, e' > 2^k + 1, \quad e = e' \mod 2^k \Longrightarrow \varphi^k(e) = \varphi^k(e').$$

La deuxième partie du lemme découle du fait que les 0 de $\varphi^k(e)$ pour $e \geq 2^k$ proviennent exclusivement des 1 et -1 qui apparaissent en itérant $\overline{\varphi}$: soient $e = 2^k p + \alpha$ et $e' = 2^k q + \alpha$, avec $(p, q, \alpha) \in \mathbb{N}^* \times \mathbb{N}^* \times [\![0, 2^k - 1]\!]$, deux entiers suprérieurs ou égaux à $2^k + 2$. On a, quel que soit l'entier $n \leq k$, $\overline{\varphi}_0^n(e) = \overline{\varphi}_0^n(e') \mod 2$ et $\overline{\varphi}_2^n(e) = \overline{\varphi}_2^n(e') \mod 2$, de plus, tous ces entiers sont supérieurs ou égaux à 2. Les occurences de 0 apparaîssent donc uniquement aux places où intervient la fonction $\overline{\varphi}_1$ (voir proposition III.5.1) et on a, quelque soit $n \leq k-1$,

$$\overline{\varphi}_1 \circ \overline{\varphi}_0^n(e) = \overline{\varphi}_1 \circ \overline{\varphi}_0^n(e') \text{ et } \overline{\varphi}_1 \circ \overline{\varphi}_2^n(e) = \overline{\varphi}_1 \circ \overline{\varphi}_2^n(e').$$

Ainsi, quel que soit l'entier $n \leq k-1$, l'apparition d'un 1 dans $\overline{\varphi}^{n+1}(e)$ coïncide avec l'apparition d'un 1 dans $\overline{\varphi}^n(e')$ et l'apparition d'un -1 dans $\overline{\varphi}^n(e)$ coïncide avec l'apparition d'un -1 dans $\overline{\varphi}^n(e')$. Les itérés $\overline{\varphi}^{n+1}(e)$ et $\overline{\varphi}^{n+1}(e')$, auront donc des occurences de 0 au mêmes places. On obtient donc, pour tout $n \leq k$, $\varphi^n(e) = \varphi^n(e')$.

Montrons maintenant la dernière affirmation du lemme. Soit k un entier plus grand ou égal à 1. Un mot w de longueur 3^k est contenu dans un itéré du type $\varphi^k(e_1)\varphi^k(e_2)$ où e_1e_2 est un mot du point fixe $\overline{m}$. D'après le lemme VII.1.2, le couple (e_1, e_2) appartient à l'ensemble $\mathcal{U} = \mathbb{N}^2 \cup \mathbb{N} \times \{-1\} \cup \{-1\} \times 2\mathbb{N}$.

On distingue deux possibilité pour le couple (e_1, e_2) : soit il appartient à $\mathcal{U} \cap [\![-1, 2^k+1]\!]^2$, et donc (e_1, e_2) est un couple de $\mathcal{U}_k$, soit au moins un des deux e_i est plus grand que $2^k + 1$. Dans la deuxième éventualité, si $e_i \leq 2^{k+1} + 1$, on pose $e_i' = e_i$, sinon, il existe, pour chacun des $e_i > 2^{k+1} + 1$, un entier $2^k + 1 < e_i' \leq 2^{k+1} + 1$ tel que $e_i = e_i' \mod 2^k$. et donc, d'après le point précédent du lemme, $\varphi^k(e_i) = \varphi^k(e_i')$. Ainsi, le mot w est aussi contenu dans l'itéré $\varphi^k(e_1')\varphi^k(e_2')$, qui est un élément de $\mathcal{U}_k$. Ainsi, pour tout mot w de longueur 3^k, il est existe un couple (e_1, e_2) de $\mathcal{U}_k$ tel que w soit un facteur de $\varphi^k(e_1)\varphi^k(e_2)$. ■

■ **Démonstration de la proposition VII.1.1**

Pour démontrer la majoration de la complexité de m annocée au résultat VII.1.1, on va majorer, pour $k \geq 1$, le nombre de mots w de m de longueur 3^k, et ainsi, montrer que

$$\forall n \geq 1, \quad p_m(n) \leq 3n(8n^{2\log_3 2} + 14n^{\log_3 2} + 7).$$

Tout mot de m de longueur 3^k est inclus dans un certain mot $\varphi^k(e_1)\varphi^k(e_2)$ avec (e_1, e_2) dans l'ensemble $\mathcal{U}_k$. On a donc l'inégalité : $p_m(2^k) \leq 3^k \cdot Card(\mathcal{U}_k)$ où 3^k représente le nombre de places différentes où on est susceptible de trouver un mot nouveau de longueur 3^k dans les $\varphi^k(e_1)\varphi^k(e_2)$. Comme

$$Card(\mathcal{U}_k) = (2^{k+1}+2)^2 + (2^{k+1}+2) + (2^k+1) = 2^{2k+2} + 72^k + 7,$$

on obtient :

$$p_m(2^k) \leq 3^k(2^{2k+2} + 72^k + 7) \quad \text{et} \quad p_m(2^{k+1}) \leq 3^{k+1}(2^{2k+3} + 142^k + 7).$$

En utilisant le fait que la fonction de complexité est une fonction croissante et que, pour tout entier n de $[\![3^k, 3^{k+1}[\![$, $2^k \leq n^{\log_3 2}$, on obtient :

$$\forall n \geq 1, \quad p_m(n) \leq 3n(8n^{2\log_3 2} + 14n^{\log_3 2} + 7),$$

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

Il semble donc qu'il soit possible de majorer la complexité d'une classe de mots engendrés par des q-ADDA de degré infini. Les hypothèses qui semblent être essentielles pour pouvoir utiliser ce raisonnement sont les deux suivantes :

1. les états de degré infini sont en nombre fini et tous les autres états sont de degré borné par une constante K,
2. tout les états à partir desquels on peut accéder à un état de degré infini sont à distance uniformément bornée des l'ensemble des états de degré infini,c'est-à-dire qu'il existe un entier B pour lequel on a la propriété suivante :
   S'il existe un chemin du graphe orienté de e vers un état de degré infini e', alors, il existe un chemin du graphe orienté, de longueur inférieure à B de e vers un état e'' de degré infini.

Ces deux conditions permettent de partitionner les éléments de $\cup_{s\in F_\Pi} \mathcal{Z}_k(s)$, de sorte que tous les éléments qui sont à distance inférieure à k de F_Π uniquement par des chemins qui passent par les états de degré infini soient regroupés en un nombre fini et indépendant de k de classes selon le nombre et les index des chemins qui les relient aux éléments de degré infini.

VII.1.2 Cas de dégénérescence

Parmi les q-ADD de degré infini, certains engendrent des mots q^∞-automatiques qui sont des mots non-constants périodiques.

Par exemple, le 2-ADD dont le graphe est représenté à la figure VII.1, avec 0 pour état initial et $\Pi : \mathbb{N} \to \{0,1\}$ définie par $\Pi^{-1}(\{1\}) = \{0\}$ pour projection admissible, engendre le mot périodique $m = 10101010101010101010\ldots..$

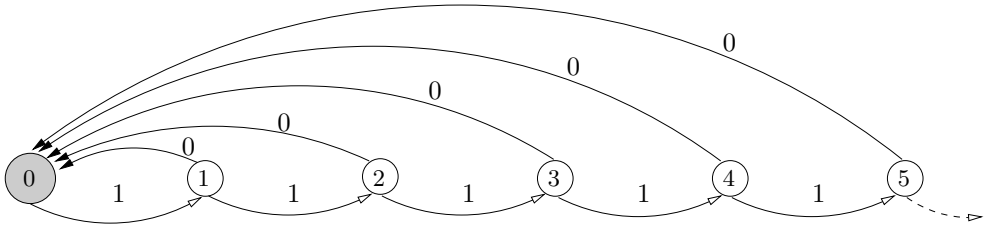


FIG. VII.1 – 2-ADD qui engendre un mot périodique.

Le 2-ADD qui engendre m est associé à la substitution

$$\begin{array}{rccc} \overline{\rho}: & \mathbb{N} & \to & \mathbb{N} \\ & n & \mapsto & 0(n+1) \end{array}$$

En fait, cela vient du fait que l'on peut partitionner l'ensemble des états du 2-ADD en deux parties $E_0 = \{0\} = \Pi^{-1}(\{1\})$ et $E_1 = \mathbb{N}^* = \Pi^{-1}(\{1\})$, et pour lesquelles on a

$$\forall i \in \{0,1\}, \quad \exists j \in \{0,1\}, \ \overline{\rho}(E_i) \subset E_j.$$

Cela traduit en fait la finitude du nombre de comportements des éléments de E sous l'action de la substitution. D'une manière générale, on a le résultat suivant :

Proposition VII.1.4. *Soit* Π *une projection admissible de* E *vers* A *et* $\overline{\varphi}$ *est une substitution de longueur constante sur l'alphabet* E *qui vérifie :*

Il existe un entier $k \geq 1$ *et une partition finie* $(P_i)_{i\in I}$ *de* E *dont les singletons* $\{s\}$*, pour tout* $s \in F_\Pi$*, font partie, tels que :*

$$\forall i \in I, \quad \exists j \in I,\ \overline{\varphi}^k(P_i) \subset P_j.$$

Alors, le mot q^∞*-automatique engendré par le* q^k*-ADDA associé à* $\overline{\varphi}^k$ *et* Π *est un mot* q*-automatique.*

Ce résultat n'est cependant pas très facile à utiliser car exhiber une partition qui correspond aux critères souhaités n'est pas forçément aussi facile que dans l'exemple que nous avons fourni et pour lequel l'entier $k = 1$ et la partition $(\{0\}, \mathbb{N}^*)$ de $\mathbb{N}$ conviennent.

VII.2 Perspectives et questions ouvertes

Les q-ADDA et les mots q^∞-automatiques offrent un large domaine de recherche sur lequel de nombreuses questions restent en suspend.

Tout d'abord, du point de vue de la théorie des langages formels, on a vu que certains langages algébriques sur $[\![0, q-1]\!]$ sont engendrés par des q-ADDA de degré borné. Cependant, la spécificité de la forme des graphe de transitions des automates à pile permettent d'affirmer que des langages non algébriques peuvent être engendrés par les q-ADD. Ainsi, on peut poser les questions suivantes : Quels sont tous les langages de $[\![0, q-1]\!]^*$ engendrés par les q-ADDA de degré borné ou monotones ? Les grammaires associées à ces langages ont-elles des propriétés spécifiques ou peut-on engendrer tous les langages récursivement énumerables ?

Nous avons montré que les mots q^∞-automatiques sont de faible complexité lorsque les automates ou, de manière équivalente, les substitutions qui les engendrent sont de degré borné ou monotones mais nous n'avons pas encore de résultats concernant les mots q^∞-automatiques engendrés par les q-ADDA de degré fini (non borné) et non monotones. Cette dernière classe de mots reste à étudier.

On peut également se demander si le théorème V.1.1 est optimal (voir V.3) et s'il est possible de déterminer quel types de fonctions peuvent être la fonction de complexité d'un mot q^∞-automatique engendré par un q-ADDA de degré borné.

L'idée d'une généralisation de ce théorème, dans la même optique que J.-J. Pansiot [Pan85] à des mots engendrés par des substitutions de longueur non constantes sur un alphabet dénombrable se heurte à de nombreux problèmes, le premier étant de cerner la nature de la croissance des longueurs des itérés des lettres par la substitution. On doit pouvoir généraliser ce théorème aux substitutions quasi-uniformes (pour lesquelles les ordres de croissance des longueurs des itérés des lettres par la substitution sont tous les même) mais il serait indispensable de pouvoir caractériser simplement ces substitutions pour que cela soit utilisable. La question reste ouverte pour les autres substitutions.

D'un tout autre point de vue, il serait intéressant de trouver une caractérisation des q-ADD dont l'ensemble des états est un module R, pour que le mot de sortie du q-ADD $\overline{m}$ soit un mot régulier. Cela permettrait de lier q^∞-automatiques et mots q-réguliers et de fournir des renseignements sur la nature des ensembles $\{n \in \mathbb{N},\ \overline{m}_n = a\}$ pour $a \in R$.

Bibliographie

[AB06] B. Adamczewski and Y. Bugeaud. On the complexity of algebraic numbers I. expansions in integer bases. to appear in *Annals of Math.*, 2006.

[ABB97] J.-M. Autebert, J. Berstel, and L. Boasson. Context-free languages and pushdown automata. In *Handbook of formal languages, Vol. 1*, pages 111–174. Springer, 1997.

[Abe01] A. Aberkane. Exemples de suites de complexité inférieure à 2n. *Bull. Belg. Math. Soc. Simon Stevin*, 8 : 161–180, 2001.

[AI01] P. Arnoux and S. Ito. Pisot substitutions and Rauzy fractals. *Bull. Belg. Math. Soc. Simon Stevin*, 8(2) : 181–207, 2001. Journées Montoises (Marne-la-Vallée, 200).

[Ale96] P. Alessandri. Codages de rotations et basses complexités, *Ph. D. thesis. Université Aix-Marseille 2*, 1996.

[All92] J.-P. Allouche. The number of factors in a paperfolding sequence. *Bull. Austr. Math. Soc.*, 26 : 23–32, 1992.

[All94] J.-P. Allouche. Sur la complexité des suites infinies. *Bull. Belg. Math. Soc. Simon Stevin*, 1(2) : 133–143, 1994.

[AR91] P. Arnoux and G. Rauzy. Représentation géométrique de suites de complexité 2n+1. *Bull. Soc. Math. France*, 119 : 199–215, 1991.

[Arn88] P. Arnoux. Un exemple de semi-conjugaison entre un échange d'intervalles et une translation sur le tore. *Bull. Soc. Math. France*, 116 : 489–500, 1988.

[AS92] J.-P. Allouche and J. Shallit. The ring of k-regural sequences. *Theoret. Comp. Sci.*, 98 : 163–197, 1992.

[AS03a] J.-P. Allouche and J. Shallit. *Automatic sequences. Theory, applications, generalizations.* Cambridge University Press, 2003.

[AS03b] J.-P. Allouche and J. Shallit. The ring of k-regural sequences 2. *Theoret. Comp. Sci.*, 307 : 3–29, 2003.

[Aut87] J.-M. Autebert. *Langages algébriques.* Etudes et recherce en informatique. Masson, 1987.

[Bel06] J. Bell. A generalisation of Cobham's theorem for regular sequences. *Séminaire Lotharingien de combinatoire*, 54A :Article B54Ap, 2006.

[Ber01] V. Berthé. Sequences of low complexity : automatic and Sturmian sequences. In *Symbolic dynamics and its applications*, volume 279 of *Lecture Notes Series*, page 3. F. Blanchard, A. Mass and A. Nogueira Eds., Cambridge University Press, 2001.

[BHMV94a] V. Bruyère, G. Hansel, C. Michaux, and R. Villemaire. Logic and p-recognizable sets of integers. *Bull. Belg. Math. Soc. Simon Stevin*, 1(2) : 191–238, 1994. Journées Montoises (Mons, 1992).

[BHMV94b] V. Bruyère, G. Hansel, C. Michaux, and R. Villemaire. Logic and p-recognizable sets of integers (corrections). *Bull. Belg. Math. Soc. Simon Stevin*, 1(4) :577, 1994.

[Bow73] R. Bowen. Topological entropy for non-compact sets. *Trans. Amer. Math. Soc.*, 184 : 125–136, 1973.

[BP85] J. Berstel and D. Perrin. *Theory of codes.* Academic Press, 1985.

[BR84] J. Berstel and C. Reutenauer. *Les séries rationnelles et leurs langages.* Etudes et recherce en informatique. Masson, 1984.

[Brl89] S. Brlek. Enumeration of factors in the Thue-Morse word. *Discrete Applied Math.*, 24 : 83–96, 1989.

[Bès00] A. Bès. A extention of the Cobham-Semënov theorem. *J. Symb. Logic*, 65(1) : 201–211, 2000.

[Cas96] J. Cassaigne. Special factors of sequences with linear subword complexity. In *Developments in language theory (Magdeburg, 1995)*, pages 25–34. World Sci. Publishing, 1996.

[Cas97] J. Cassaigne. Complexité et facteurs spéciaux. *Bull. Belg. Math. Soc. Simon Stevin*, 4(1) : 67–88, 1997.

[Cas02] J. Cassaigne. Constructing infinite words of intermediate complexity. In *Developments in language theory (Kyoto 2002)*, volume 2450 of *Lecture Notes in Comp. Sci.*, pages 173–184. Springer, 2002.

[Cau92] D. Caucal. On the regular structure of prefix rewriting. *Theor. Comp. Sci.*, 106 : 61–86, 1992.

[Cau03] D. Caucal. On infinite transition graphs having a decidable monadic theory. *Theor. Comp. Sci.*, 290 : 79–115, 2003.

[Cho56] N. Chomsky. Three models for the description of language. *IRE transactions on information theory*, 2(3) : 113–124, 1956.

[Cho59] N. Chomsky. On certain formal properties of grammars. *Information and control*, 2(2) : 137–167, 1959.

[Chr79] Gilles Christol. Ensembles presque périodiques k-reconnaissables. *Theoret. Comput. Sci.*, 9(1) : 141–145, 1979.

[CK97] J. Cassaigne and J. Karhumaki. Toeplitz words, generalized periodicity and periodically iterated morphisms. *Eur. J. Comb.*, 18(5) : 497–510, 1997.

[CK02] D. Caucal and T. Knapik. A chomsky-like hierarchy of infinite graphs. In *Proc. of the 27th MFCS, Warsaw*, volume 2420 of *Lecture Notes in Comput. Sci.*, pages 177–187. Springer, 2002.

[CKMFR80] G. Christol, T. Kamae, M. Mendès France, and G. Rauzy. Suites algébriques, automates et substitutions. *Bull. Math. Soc. France*, 108(4) : 401–419, 1980.

[CM06] A. Carayol and A. Meyer. Context-sensitive languages, rational graphs and determinism. to appear in *Logical Methods in Comp. Sci.*, 6, 2006.

[Cob69] A. Cobham. On the base-dependence of sets of numbers recognizable by finite automata. *Math. Syst. Theory*, 3 : 186–192, 1969.

[Cob72] A. Cobham. Uniform-tag sequences. *Math. Syst. Theory*, 6 : 164–192, 1972.

[CS66] N. Chomsky and M.-P. Schürtzenberger. The algebraic theory of context-free languages. In *Computer Programming and formal systems*, pages 195–242. North-Holland, 1966. P. Braffort and D. Hirshberg Eds.

[Dur02] F. Durand. A theorem of Cobham for non primitive substitutions. *Acta Arith.*, 104 : 225–241, 2002.

[Eil74] S. Eilenberg. *Automata, languages and machines*, volume Vol. A. Academic Press, 1974.

[ELR75] A. Ehrenfeucht, K. P. Lee, and G Rozenberg. Subword complexities of various classes of deterministic developmental languages without interactions. *Theor. Comp. Sci.*, 1 : 59–75, 1975.

[Fav94] S. Favre. Une généralisation du théorème de Cobham. *Acta Arith.*, 67 : 197–208, 1994.

[Fer95] S. Ferenczi. Transformations de Chacon : combinatoire, structure générale, lien avec les systèmes de complexité $2n + 1$. *Bull. Soc. Math. France*, 123(2) : 271–292, 1995.

[Fer99] S. Ferenczi. Complexity of sequences and dynamical systems. *Discrete Math.*, 206(1-3) : 145–154, 1999.

[Fer06] S. Ferenczi. Substitution dynamical systems on infinite alphabets. to appear in *Ann. Inst. Fourier*, 56, 2006.

[FF97] D. Fiebig and U. Fiebig. Entropy and finite generators for locally compact subshifts. *Ergodic Th. and Dyn. Syst.*, 17 : 349–368, 1997.

[FM97] S. Ferenczi and C. Mauduit. Transcendence of numbers with a low complexity expansion. *J. Number Theory*, 67 : 146–161, 1997.

[FM05] E. Fouvry and C. Mauduit. Sur les entiers dont la somme des chiffres est moyenne. *J. Number Theory*, 114 : 135–152, 2005.

[Goy97] J. Goyon. Construction de suites dont la complexité est donnée. Preprint, *Institut des Mathématiques de Luminy*, 1997.

[Gur69] B. M. Gurevic. Topological entropy of enumerable markov chains. *Soviet Math. Dokl.*, 10(4) : 911–915, 1969.

[Gur70] B. M. Gurevic. Shift entropy and markov measures in the path space of an enumerable graph. *Soviet Math. Dokl.*, 11(3) : 744–747, 1970.

[HK95] M. Handel and B. Kitchens. Metrics and entropy for non-compact spaces. *Israel J. Math.*, 91 : 253–271, 1995.

[HM38] G. A. Hedlund and M. Morse. Symbolics dynamics. *Amer. J. Math.*, 60 : 815–866, 1938.

[HM40] G. A. Hedlund and M. Morse. Symbolics dynamics II. Sturmian trajectories. *Amer. J. Math.*, 62 : 1–42, 1940.

[HS03] G. Hansel and T. Safer. Vers un théorème de Cobham pour les entiers de Gauss. *Bull. Belg. Math. Soc. Simon Stevin*, 10 : 723–735, 2003.

[HU79] J. E. Hopcroft and J. D. Ullman. *Introduction to automata theory, languages and computation.* Addison-Wesley, 1979.

[Kit98] B. Kitchens. *Symbolic dynamics. One-sided, two-sided and countable Markov shifts.* Universitext. Springer, 1998.

[KU87] A. N. Kolmogorov and V. A. Uspensky. Algorithms and randomness. *SIAM J. Theory of probability and Its applications*, 32 : 389–412, 1987.

[Kůr03] P. Kůrka. *Topological ans symbolic dynamics*, volume 11 of *Cours spécialisés.* SMF, 2003.

[LG06] M. Le Gonidec. Sur la complexité de mots infinis engendrés par des q-automates dénombrables. to appear in *Ann. Inst. Fourier*, 56, 2006.

[Lot02] M. Lothaire. *Algebraic ombinatorics on words*, volume 90 of *Encyclopedia of Mathematics and Its applications.* Cambridge University Press, 2002.

[LV97] M. Li and P. Vitanyi. *An introduction to Kolmogorov complexity and its applications.* Springer, 1997.

[Mau88] C. Mauduit. Sur l'ensemble normal des substitutions de longueur quelconque. *J. Number Theory*, 29(3) : 235–250, 1988.

[Mau06] C. Mauduit. Propriétés arithmétiques des substitutions et automates infinis. to appear in *Annales de l'Institut Fourier*, 56, 2006.

[MM06] C. Mauduit and C. G. Moreira. Generalized fractal dimensions and complexity of infinite sequences. *Preprint*, 2006.

[Mon05] T. Monteil. Estimating the number of ergodic measures of a minimal subshift knowing the geometry of its Rauzy graphs. *Preprint*, 2005.

[Mos92] B. Mossé. Puissances de mots et reconnaissabilité des points fixes de substitutions. *Theor. Comp. Sci.*, 99 : 327–334, 1992.

[Mos96] B. Mossé. Reconnaissabilité des substitutions et complexité des suites automatiques. *Bull. Soc. Math. France*, 124(2) : 329–346, 1996.

[MR05] C. Morvan and C. Rispal. Families of automata characterizing context-sensitive languages. *Acta Informatica*, 41(4-5) : 293–314, 2005.

[MS85] D. Muller and P. Schupp. The theory onf ends, pushdown automata ans second-order logic. *Theor. Comp. Sci.*, 37 : 51–75, 1985.

[MS01] C. Morvan and C. Stirling. Rational graphs trace context-sensitive languages. In *Proc. of the 26th MFCS*, volume 2136 of *Lecture Notes in Comput. Sci.*, pages 548–559. Springer, 2001.

[Pan85] J.-J. Pansiot. Complexité des facteurs des mots infinis engendrés par morphismes itérés. *Lecture Notes in Comp. Sci.*, 172 : 380–389, 1985. Automata, languages and programming (Antwerp, 1984).

[Per90] D. Perrin. Finite automata. In *Handbook of Theoretical Computer Science, Vol. B*, pages 3–57. J. Van Leeuwen Ed., Elsevier, 1990.

[PF02] N. Pytheas Fogg. *Substitutions in dynamics, arithmetics and combinatorics*. Lecture Notes in Mathematics. Springer, 2002. Edited by V. Berthé, S. Ferenczi, C. Mauduit and A. Siegel.

[PP04] D. Perrin and J.-E. Pin. *Infinite words, Automata, Semigroups, Logic and Games*, volume 141 of *Pure and Applied Mathematics*. Elsevier, 2004.

[Que87] M. Queffélec. *Substitution dynamical systems—spectral analysis*. Lecture Notes in Mathematics. Springer, 1987.

[Rau82] G. Rauzy. Nombres algébriques et substitutions. *Bull. Soc. Math. France*, 110 : 147–178, 1982.

[Rot94] G. Rote. Sequences with subword complexity $2n$. *J. Number Theory*, 46 : 196–213, 1994.

[RS97] G. Rozenberg and A. Salomaa. Eds. *Handbook of formal language*. Springer, 1997.

[RW06] M. Rigo and L. Waxweiler. A note on syndeticity, recognizable sets and Cobham's theorem, to appear in. *Bull. of EATCS*, 2006.

[Sal84] I. Salama. Topological entropy and classification of countable chains. Ph. D. thesis, *University of north Carolina*, 1984.

[Sal88] I. Salama. Topological entropy and recurrence of countable chains. *Pacific J. Math.*, 134(2) : 325–341, 1988. Errata, **140**(2) : 397.

[Sch61] M.-P. Schützenberger. On a definition of a family of automata. *Information and control*, 4 : 245–270, 1961.

[Sch62a] M.-P. Schützenberger. Certain elementary families of automata. In *Proc. Symposium on Math. Th. of Automata*, pages 139–153. Polytechnic Insitute of Brooklin, 1962.

[Sch62b] M.-P. Schützenberger. On a theorem of R. Jungen. *Proc. Amer. Math. Soc.*, 13 :885–889, 1962.

[Sti05] C. Stirling. Language theory and infinite graphs. *Marktoberdorf summerschool*, 2005. Notes for *Logical aspects of secure computer systems*.

[Tho02] W. Thomas. An short introduction to infinite automata. In *Proceedings of the International Conference of Developments in Language Theory*, volume 2295 of *Lecture Notes in Comput. Sci.*, pages 130–144. Springer, 2002.

[Wal82] P. Walters. *An introduction to ergodic theory*. Springer, 1982.

RÉSUMÉ : L'objet de cette thèse est l'étude de mots infinis, à valeurs dans un alphabet fini, engendrés par des q-automates dont l'ensemble des états est dénombrable (q est un nombre entier supérieur ou égal à 2). Ces mots, appelés mots q^∞-automatiques, sont les images par morphismes « admissibles » de points fixes de substitutions de longueur constante sur un alphabet dénombrable et peuvent être perçus comme une généralisation des mots automatiques.

Nous montrons que la fonction de complexité p du mot $(2d)^\infty$-automatique engendré par l'automate associé à la marche aléatoire régulière dans $\mathbb{Z}^d$ vérifie $p(n) = O(n \log_{2d}^{d+1} n)$. Lorsque $d = 1$, on montre que la fonction $n \log_2^2 n$ est un équivalent de la fonction de complexité. Nous verrons également que l'on peut généraliser cette majoration de l'ordre de grandeur de la fonction de complexité aux mots q^∞-automatiques engendrés par les automates associé à marches aléatoires sur des réseaux de $\mathbb{R}^d$ par une relation du type $p(n) = O(n \log_q^{q+1} n)$.

Dans le cadre plus général des mots q^∞-automatiques engendrés par des automates de degré borné, nous montrons que la fonction de complexité est au plus polynomiale. Nous verrons, sur des exemples, comment améliorer ce résultat pour obtenir une majoration plus fine de l'ordre de grandeur de la fonction de complexité. Nous donnons également un résultat de majoration polynomiale de l'ordre de grandeur de la fonction de complexité pour les mots q^∞-automatiques engendrés par des automates monotones.

On présentera pour finir un exemple de mot q^∞-automatique dont la complexité a pour ordre de grandeur $n\sqrt{n}$.

ABSTRACT : In this thesis, we study a class of infinite words on a finite alphabet, generated by q-automata with countable states set, for an integer $q \geq 2$. Those words, called q^∞-automatic words are images by some « admissibles » morphisms of fixed points of substitutions of constant length q over countable alphabets. In this sense, thoses words can be viewed as a generalisation of automatic words.

We show that the complexity function p of the $(2d)^\infty$-automatic word generated by the automaton related to regular random walk on $\mathbb{Z}^d$ satisfies $p(n) = O(n \log_{2d}^{d+1} n)$. Moreover, in the one-dimension case, the function $n \log_2^2 n$ is an equivalent of the complexity function. This result can be generalized to q^∞-automatic words generated by automata related to more general random walks on $\mathbb{Z}^d$ by the relation $p(n) = O(n \log_q^{q+1} n)$.

In the more general case of q^∞-automatic words generated by bounded degree automata, we prouve that the complexity function is at most a polynomial function. We show how this result can be improved on examples, to obtain a thiner majoration of the complexity function order of growth. Moreover, on one exemple related to the Dyck language over two types of parenthesis, we obtain the exact complexity function order of growth.

We also give a polynomial majoration of the complexity function for q^∞-automatic words generated by monotone automata.

Mots clés : mots infinis, automates, automates dénombrables, substitutions, combinatoire des mots, complexité.