UNIVERSITÉ DU QUÉBEC À MONTRÉAL

MOTS ÉQUILIBRÉS ET MOTS LISSES

THÈSE

PRÉSENTÉE

COMME EXIGENCE PARTIELLE

DU DOCTORAT EN MATHÉMATIQUE

PAR

GENEVIÈVE PAQUIN

AOÛT 2008

REMERCIEMENTS

Merci à mon directeur Christophe Reutenauer qui, sans me connaître, a accepté de me diriger. En me donnant une grande liberté, en étant d'une patience remarquable et en étant toujours présent au bon moment, il m'a permis de faire mes études selon mon rythme et mes envies et ainsi, de m'épanouir pleinement durant mon doctorat.

Merci à mon co-directeur Srecko Brlek de m'avoir initiée à ce domaine. Son dynamisme et son sens de l'initiative m'ont permis de bien m'intégrer dans la communauté. Lors des moments plus difficiles, son écoute et sa confiance en moi ont permis de bien me motiver.

Merci aux membres de mon jury : Jean Berstel, Valérie Berthé et Gilbert Labelle. Merci d'avoir accepté de prendre le temps de lire ma thèse. C'est un honneur pour moi d'avoir un jury d'une telle qualité.

Merci à Pierre Leroux. Sans être directement impliqué dans mon cheminement mathématique, il a cru en moi dès le début et est venu vers moi afin de m'encourager à poursuivre après la maîtrise. Sans lui, je n'aurais probablement jamais décidé de relever ce grand défi et je regrette de ne jamais avoir pris le temps de le remercier avant son départ.

Merci à mes collaborateurs, Laurent Vuillon et Damien Jamet, avec qui j'ai eu le plaisir de travailler. Elles sont les premières personnes extérieures à l'UQAM à m'avoir considérée comme une chercheure et cela a été fort bénéfique pour la continuité de ma thèse.

Merci à Nicolas pour son soutien, son écoute et sa sensibilité qui ont permis d'égayer les nombreuses années passées au 5218. Comme un rayon de soleil, sa présence a compensé l'absence de fenêtre dans notre bureau.

Merci à mes collègues et amis uqamiens aux côtés de qui j'évolue depuis 8 ans, plus particulièrement mes frères et ma soeur mathématiques qui ont été une grande source de motivation et d'encouragement.

Merci à ma famille, à mes amis et à mon chéri. Je vous aime.

Merci au Conseil de recherches en sciences naturelles et en génie du Canada qui m'a soutenue financièrement durant les 4 dernières années.

TABLE DES MATIÈRES

LISTE DES TABLEAUX

LISTE DES FIGURES

RÉSUMÉ

Dans ce travail, nous nous intéressons à divers problèmes de la combinatoire des mots, portant principalement sur deux familles : les mots équilibrés et les mots lisses infinis. Il est facile de vérifier que l'intersection entre ces deux familles de mots est vide ; c'est pourquoi nous les traiterons de façon indépendante.

Dans la première partie, nous étudions les mots équilibrés et des familles de mots dérivées de ces derniers. Les mots infinis sturmiens, aussi appelés des suites sturmiennes, sont étudiés depuis plus de cent ans et sont caractérisés de plusieurs façons : pour un alphabet à deux lettres, ce sont exactement les suites équilibrées non ultimement périodiques et les suites de complexité minimale, c'est-à-dire les suites ayant seulement $(n+1)$ facteurs de longueur n. Une autre propriété caractéristique des suites sturmiennes est qu'elles décrivent une droite discrète. Les suites épisturmiennes ont récemment été introduites comme étant l'une des généralisations sur plus de 2 lettres des suites sturmiennes ; un mot de Christoffel est la version finie d'une suite sturmienne. Dans un premier temps, nous introduisons donc une généralisation des mots de Christoffel sur un alphabet à plus de 2 lettres. Pour ce faire, nous utilisons la propriété qu'un mot de Christoffel est l'image d'une lettre par morphisme sturmien. Nous appelons les mots ainsi obtenus des *mots épichristoffels*. Il est intéressant de remarquer que ces mots ne sont généralement pas équilibrés, tout comme les suites épisturmiennes. Nous montrons comment obtenir des mots épichristoffels, comment les reconnaître et nous montrons que certaines propriétés des mots de Christoffel se généralisent bien aux mots épichristoffels.

Dans un deuxième temps, nous nous intéressons aux mots équilibrés, en lien avec la conjecture de Fraenkel. Cette conjecture énonce que pour un alphabet à k lettres, avec $k \geq 3$, il n'existe qu'un unique mot infini équilibré, à permutation des lettres et à décalage près, ayant des fréquences de lettres toutes différentes. Par exemple, pour l'alphabet $\{1,2,3\}$, ce mot est $(1213121)^\omega$. Nous montrons que la conjecture est vraie si elle est restreinte à la famille des suites épisturmiennes et du coup, nous caractérisons les suites épisturmiennes équilibrées.

Nous approchons ensuite la conjecture de Fraenkel en travaillant sur la superposition de mots de Christoffel. Nous traduisons les travaux de R. Simpson et de R. Morikawa sur les suites de Beatty en terme de mots de Christoffel et nous fournissons les détails passés sous silence dans leurs preuves. Nous obtenons ainsi une condition nécessaire et suffisante pour que deux mots de Christoffel se superposent. Comme un mot équilibré à k lettres peut être vu comme la superposition de k mots équilibrés sur 2 lettres, cette condition nous permet de nous approcher de la conjecture de Fraenkel, sans toutefois la prouver. Nous prouvons toutefois une formule donnant le nombre de superpositions pour deux mots de Christoffel superposables et nous montrons de nouvelles propriétés concernant les mots de Christoffel.

La deuxième partie de ce travail porte sur l'étude des mots lisses infinis, principalement les mots lisses extrémaux, c'est-à-dire le plus petit et le plus grand selon l'ordre lexicographique. Nous caractérisons ces mots sur des alphabets à deux lettres de même parité. Pour ce faire, nous décrivons d'abord des algorithmes qui permettent de les construire en temps linéaire selon le nombre d'opérations. Nous montrons ensuite des propriétés de fermeture et de récurrence pour les mots lisses en général sur un alphabet de même parité, nous fournissons une formule explicite pour la fréquence des lettres dans les mots extrémaux et nous décrivons la factorisation de Lyndon pour une sous-classe des mots extrémaux. Ces résultats sont forts intéressants puisque les propriétés démontrées ne sont pour la plupart que des conjectures pour les mots lisses sur l'alphabet $\{1, 2\}$. Par ailleurs, nous montrons que les mots lisses maximaux sur des alphabets contenant deux lettres paires et certains mots de Kolakoski généralisés coïncident. Du coup, nous prouvons plusieurs propriétés concernant la factorisation de Lyndon, les fréquences, la fermeture de l'ensemble des facteurs sous l'image miroir et la récurrence pour les mots de Kolakoski généralisés. Finalement, nous étudions les mots lisses infinis en lien avec les surfaces discrètes. Nous montrons que les seuls pavages lisses du quart de plan décrivant un morceau de surface discrète sont engendrés par des mots de Kolakoski généralisés.

Mots clés : Combinatoire des mots, mot de Christoffel, suite sturmienne, mot épichristoffel, suite épisturmienne, conjecture de Fraenkel, suite équilibrée, mot lisse, mot de Kolakoski, mot de Lyndon, mot extrémal, surface discrète.

INTRODUCTION

Même si elle apparaît dans la littérature depuis près de cent ans, la combinatoire des mots est un domaine de l'informatique mathématique récent. En effet, malgré que certains articles étudiaient déjà les mots au début du 20ième siècle (Thue, 1906; Thue, 1910; Thue, 1912; Thue, 1914), les premiers recueils publiés sur ce sujet sous le pseudonyme Lothaire (Lothaire, 1983; Lothaire, 2002; Lothaire, 2005) n'apparaissent qu'à partir de la fin du 20ième siècle. Dans plusieurs domaines des mathématiques tels que la théorie des nombres, la dynamique symbolique, la géométrie discrète et l'infographie, il est possible de représenter certains problèmes en terme de mots, c'est-à-dire une suite finie ou infinie de symboles. La combinatoire des mots est donc un outil permettant de modéliser des problèmes apparaissant dans des domaines très variés.

Dans ce travail, nous étudions deux ensembles de mots : les mots équilibrés et les familles reliées, et les mots lisses. Rappelons que les suites sturmiennes sont exactement les suites équilibrées sur un alphabet à deux lettres et que ces dernières décrivent les droites discrètes (Morse et Hedlund, 1938; Coven et Hedlund, 1973; Lothaire, 2002). Les mots de Christoffel sont la version finie des suites sturmiennes. Ces derniers ont été aussi fortement étudiés (Christoffel, 1875; Borel et Laubie, 1993; Berstel et de Luca, 1997; Borel et Reutenauer, 2005; Borel et Reutenauer, 2006; Berthé, de Luca et Reutenauer, 2007) et sont maintenant bien caractérisés. Dans ce travail, nous donnons une généralisation des mots de Christoffel sur un alphabet à plus de deux lettres et nous montrons quelles sont les propriétés qui se généralisent. Par la suite, nous nous intéressons à la conjecture de Fraenkel (Fraenkel, 1973). Ce dernier prétend que pour un alphabet à k lettres, avec $k \geq 3$, il n'existe qu'une seule suite équilibrée, à permutation des lettres et conjugaison près, ayant des fréquences de lettres toutes différentes. Nous abordons cette conjecture d'abord en circonscrivant le problème aux suites épisturmiennes introduites dans (Droubay, Justin et Pirillo, 2001). Ensuite, nous approchons cette conjecture en

utilisant la superposition de mots équilibrés, comme les mots satisfaisant la conjecture sont obtenus par superposition de mots équilibrés. Nous nous intéressons ensuite aux mots lisses infinis introduits dans (Brlek et al., 2006), c'est-à-dire les mots qui restent sur le même alphabet sous l'application de l'opérateur de codage par blocs un nombre infini de fois. Plus particulièrement, nous voulons caractériser le plus petit et le plus grand mots selon l'ordre lexicographique afin de donner des bornes à l'ensemble des mots lisses infinis, pour un alphabet fixé. Pour terminer, nous étudions le lien entre les surfaces discrètes (voir (Jamet, 2004)), c'est-à-dire la généralisation des droites discrètes en 3 dimensions, et les mots lisses. Cette idée provient du fait que les surfaces discrètes peuvent être codées par un pavage sur l'alphabet $\{1, 2, 3\}$ et qu'il est possible d'associer un pavage du plan à un mot lisse pour un alphabet fixé. Nous voulons savoir à quoi correspond l'intersection de ces deux objets.

Ce travail est présenté comme suit. Dans le Chapitre I, nous rappelons les définitions et notations de bases en combinatoire des mots, d'abord pour les mots finis, puis pour les mots infinis et bi-infinis. Ces notations sont celles qui seront utilisées tout au long de ce travail. Par ailleurs, notons qu'en cas de besoin, le lecteur peut se référer en tout temps à l'index afin de retrouver facilement les définitions nécessaires.

Dans le Chapitre II, après avoir rappelé les différentes définitions et caractérisations des suites sturmiennes, nous donnons les principaux résultats les concernant. Nous énonçons ensuite la définition d'un mot de Christoffel, nous donnons sa représentation géométrique et sa construction à l'aide des graphes de Cayley ainsi que quelques unes de ses propriétés qui nous seront utiles pour la suite. Comme les suites sturmiennes sont caractérisées de plusieurs façons, il y a une multitude de généralisations possibles, tout dépendant de la propriété utilisée. La généralisation qui nous intéresse ici est la famille des suites épisturmiennes. Après avoir rappelé les définitions et certaines propriétés des suites épisturmiennes, nous introduisons finalement la généralisation des mots de Christoffel sur plus de deux lettres : les mots épichristoffels. Par analogie avec les mots de Christoffel, les mots épichristoffels sont définis comme étant les mots finis obtenus par morphisme épisturmien sur une lettre et qui sont plus petits que tous ses suffixes ; ils ne sont pas nécessairement équilibrés. Dans un premier temps, nous montrons que tout comme les mots de Christoffel, les mots épichristoffels sont primitifs et s'écrivent de façon

unique comme le produit de deux palindromes. Par la suite, nous donnons un algorithme rapide qui permet de dire si un k-tuplet décrit les fréquences d'un mot épichristoffel ou pas. Si tel est le cas, nous montrons que ce mot épichristoffel est unique. Nous montrons ensuite que quelques résultats sur les mots de Christoffel se généralisent bien aux mots épichristoffels. Finalement, nous prouvons une condition nécessaire et suffisante qui permet de dire si un mot est dans la classe de conjugaison d'un mot épichristoffel ou pas.

Le Chapitre III concerne les suites épisturmiennes et la conjecture de Fraenkel. Comme les suites sturmiennes sont équilibrées et que ce n'est généralement pas le cas pas pour les suites épisturmiennes, il est naturel de vouloir caractériser quelles sont les suites qui sont à la fois épisturmiennes et équilibrées. Par ailleurs, il y a plus de 30 ans, A. S. Fraenkel a conjecturé que pour un $k > 2$ fixé, il existe une seule façon de couvrir les entiers $\mathbb{Z}$ avec k suites de la forme $\lfloor an + b \rfloor$ ayant des fréquences différentes. En combinatoire des mots, cette conjecture peut se reformuler ainsi : il existe une seule suite équilibrée sur un alphabet à k lettres ayant des fréquences de lettres différentes, à permutation de lettres et à conjugaison près. Nous montrons qu'il existe exactement trois différentes suites épisturmiennes équilibrées, à permutations des lettres et conjugaison près, décrites par les suites directrices suivantes :

a) $\Delta(s) = 1^n 23 \dots (k-1)k^\omega$, avec $n \geq 1$;

b) $\Delta(s) = 12 \dots (k-1)1k(k+1) \dots (k+\ell-1)(k+\ell)^\omega$, avec $\ell \geq 1$;

c) $\Delta(s) = 123 \dots k1^\omega$,

où $k \geq 3$. Parmi elles, une seule a des fréquences de lettres toutes différentes. Ainsi, en plus de fournir une belle caractérisation des suites épisturmiennes équilibrées, ce résultat fournit une preuve partielle de la conjecture de Fraenkel pour les suites épisturmiennes. Les résultats de ce chapitre font l'objet de la publication (Paquin et Vuillon, 2007).

Toujours motivé par la conjecture de Fraenkel, nous étudions dans le Chapitre IV la superposition de deux mots de Christoffel. Des travaux précédents ont montré que les suites satisfaisant à la conjecture de Fraenkel sont périodiques. Ainsi, puisque tout mot fini u tel que u^ω satisfait à la conjecture de Fraenkel peut être obtenu par la superposition de k mots équilibrés sur deux lettres, nous abordons la conjecture de Fraenkel en cherchant sous quelle condition deux mots de Christoffel se superposent. Nous utilisons les résultats de (Simpson, 2004) dans lequel l'au-

teur donne une condition nécessaire et suffisante pour que deux suites de Beatty se superposent. Nous traduisons tout en terme de mots de Christoffel. Nous obtenons donc une condition nécessaire et suffisante afin que deux mots de Christoffel se superposent et de la preuve de ce résultat découlent de nouvelles propriétés des mots de Christoffel. En plus de fournir les détails passés sous silence dans les preuves de (Simpson, 2004), nous allons plus loin en donnant le nombre de superpositions possibles pour deux mots de Christoffel. Nous terminons ce chapitre en utilisant les mots de Christoffel afin de prouver que le nombre d'éléments du sous-monoïde engendré par a et b inférieurs à $(a-1)(b-1)$ est $(a-1)(b-1)/2$.

La deuxième partie de ce travail débute au Chapitre V : nous abandonnons les mots équilibrés et travaillons plutôt sur les mots lisses infinis. Après avoir rappelé ce qu'est le mot de Kolakoski, nous introduisons une famille de mots dérivée de ce dernier, les mots lisses infinis. Nous définissons d'abord la fonction de codage par blocs que nous notons Δ. Cette fonction est ensuite utilisée pour définir la famille des mots lisses infinis. Nous rappelons ensuite les différentes propriétés de Δ, nous montrons de quelle façon les mots lisses infinis sont en bijection avec l'ensemble des mots infinis, puis nous définissons les facteurs, préfixes et suffixes lisses. Ensuite, après avoir introduit les graphes de De Bruijn et sa version réduite, nous entrons dans le sujet principal : les mots lisses infinis extrémaux. Il s'agit du plus petit et du plus grand mots lisses infinis, selon l'ordre lexicographique. Nous nous intéressons d'abord aux algorithmes qui les engendrent et nous montrons qu'il est possible de construire un préfixe de longueur n d'un mot lisse extrémal en $\mathcal{O}(n^2)$ opérations. Nous étudions ensuite les propriétés structurelles des mots lisses extrémaux, c'est-à-dire leurs dérivées successives et leurs factorisations de Lyndon. Selon les dérivées successives obtenues, aucune régularité ne semble apparaître dans les mots lisses extrémaux. Les factorisations de Lyndon suggèrent que les mots lisses extrémaux admettent un suffixe lisse minimal et nous montrons que le mot lisse minimal n'est pas un mot de Lyndon infini. Les résultats de ce chapitre font l'objet de la publication (Brlek, Melançon et Paquin, 2007).

Dans le Chapitre VI, nous nous intéressons encore aux mots lisses extrémaux, mais en considérant maintenant les mots lisses sur des alphabets autres que l'alphabet $\{1,2\}$. Plus particulièrement, nous étudions les mots lisses extrémaux sur des alphabets à deux lettres de même

parité. Nous commençons par généraliser les définitions reliées aux mots lisses pour un alphabet ordonné quelconque à deux lettres. Nous étudions ensuite le mot lisse infini minimal sur l'alphabet $\{1, 3\}$. Nous montrons qu'il existe un algorithme linéaire pour le construire et que le mot infini de Fibonnaci et le mot lisse minimal sur $\{1, 3\}$ sont dans le même orbitre sous l'opérateur Δ. Une question naturelle est de savoir si ces propriétés se généralisent pour tout alphabet à deux lettres impaires. Nous montrons ensuite que les mots lisses sur un alphabet impair contiennent une infinité de palindromes préfixes, qu'ils sont récurrents, que l'ensemble des facteurs est fermé sous l'image miroir et que les mots lisses extrémaux s'obtiennent par un algorithme linéaire. De plus, nous montrons que le mot lisse minimal est un mot de Lyndon infini si et seulement si l'alphabet s'écrit comme $\{a < b\}$, avec $a \neq 1$ et a, b impaires. Nous montrons aussi que pour un alphabet de la forme $\{1, b\}$ avec b impair, la fréquence de la lettre b du mot lisse minimal est donnée par $\frac{1}{\sqrt{2b-1}+1}$. Ensuite, l'étude des mots lisses extrémaux sur un alphabet à deux lettres paires nous indique que ces derniers peuvent aussi être obtenus par un algorithme linéaire. Par ailleurs, nous montrons que pour cet alphabet, la fréquence des lettres dans les mots lisses extrémaux est $1/2$, que tout mot lisse est récurrent et que les ensembles des facteurs des mots lisses extrémaux ne sont pas fermés sous l'image miroir, ni sous la complémentation. Nous montrons aussi que les mots maximaux correspondent exactement aux mots de Kolakoski généralisés. Cela implique que toutes les propriétés prouvées pour les mots lisses extrémaux pour un alphabet pair s'appliquent aux mots de Kolakoski généralisés sur ces alphabets. Pour finir, nous prouvons que le mot lisse minimal est un mot de Lyndon infini. Les résultats de ce chapitre font l'objet de la publication (Brlek, Jamet et Paquin, 2008).

Nous terminons ce travail par le Chapitre VII dans lequel nous étudions le lien entre les surfaces discrètes et les mots lisses. Les surfaces discrètes sont la généralisation des suites sturmiennes en 3 dimensions : elles approximent des surfaces. Il est possible d'associer un pavage de plan à une surface discrète et ce pavage s'écrit sur l'alphabet $\{1, 2, 3\}$. Comme à tout mot lisse, il est aussi possible d'associer un pavage du quart de plan, l'objectif du chapitre est de caractériser les mots lisses qui ont un pavage de plan décrivant un morceau de surface discrète. Pour ce faire, nous rappelons les notions nécessaires concernant les surfaces discrètes. Nous énonçons un résultat de (Jamet, 2004) qui fournit une condition nécessaire et suffisante sur un pavage

du plan afin de déterminer s'il décrit ou pas une surface discrète. Pour finir, nous étudions tous les cas possibles et montrons par élimination que les seuls mots lisses ayant un pavage de plan décrivant un morceau de surface discrète sont les mots $K_{(3,1)}, 1K_{(3,1)}$ et $2K_{(3,1)}$, où $K_{(3,1)}$ est le mot de Kolakoski généralisé sur l'alphabet $\{1, 3\}$ et débutant par la lettre 3. Les résultats de ce chapitre font l'objet de la publication (Jamet et Paquin, 2005).

Chapitre I

GÉNÉRALITÉS SUR LES MOTS

L'objectif de ce chapitre est de rappeler les concepts et définitions de base en combinatoire des mots qui seront utilisés tout au long de ce travail. Nous fixons aussi les notations qui apparaîtront dans les chapitres ultérieurs. Les définitions et notations seront fortement basées sur celles utilisées dans les Lothaire (Lothaire, 1983; Lothaire, 2002). Nous définissons les trois types de mots que nous considérerons : les mots finis, infinis et bi-infinis. Finalement, nous introduisons les concepts utiles qui s'y rattachent.

Dans ce qui suit, $\mathbb{N}$, $\mathbb{Z}$ et $\mathbb{R}$ désignent respectivement l'ensemble des naturels (entiers positifs ou nuls), des entiers et des réels. La cardinalité d'un ensemble E sera notée Card(E). Un *alphabet* $\mathcal{A}$ est un ensemble fini de symboles qui sont appelés des *lettres*. On écrit $\mathcal{A} = \{a_0, a_1, \ldots, a_{n-1}\}$ pour désigner un alphabet à n lettres et $\mathcal{A} = \{a_0 < a_1 < \ldots < a_{n-1}\}$ s'il est *ordonné*. Un *semi-groupe* est un ensemble muni d'une opération binaire associative. L'ensemble des mots sur un alphabet $\mathcal{A}$ avec l'opération de concaténation est un semi-groupe.

Un *morphisme de semi-groupe* s d'un semi-groupe S dans un semi-groupe T est une fonction $s : S \to T$ telle que $s(uv) = s(u)s(v)$, pour tout $u, v \in S$. Un *monoïde* M est un semi-groupe ayant un élément neutre, c'est-à-dire un élément e tel que $me = em = m$ pour tout $m \in M$. Un *morphisme de monoïde* s d'un monoïde M dans un monoïde N est un morphisme de semi-groupe $s : M \to N$ tel que $s(e_M) = e_N$.

1.1 Mots finis

Un *mot fini* w sur l'alphabet $\mathcal{A}$ est une suite finie de lettres $w = w[0]w[1]\cdots w[n-1]$, où $w[i-1] \in \mathcal{A}$ est la i-ième lettre de w. On dit alors que w est de *longueur* n et on note $\lg(w) = n$. L'ensemble des mots finis de longueur n sur l'alphabet $\mathcal{A}$ est désigné par $\mathcal{A}^n$ et $\mathcal{A}^* = \cup_{n\in\mathbb{N}}\mathcal{A}^n$ est l'ensemble de tous les mots finis sur l'alphabet $\mathcal{A}$. Le *mot vide* est noté ε et sa longueur est 0. On note $\mathcal{A}^+ = A^* - \{\varepsilon\}$ l'ensemble des mots finis non vides sur l'alphabet $\mathcal{A}$. L'ensemble des mots finis de longueur $\leq n \in \mathbb{N}$ sur l'alphabet $\mathcal{A}$ est noté $\mathcal{A}^{\leq n}$.

L'ensemble $\mathcal{A}^*$ est un monoïde. En effet, la concaténation des mots est associative et le mot vide est bien un élément neutre pour la concaténation. L'ensemble $\mathcal{A}^+$ est appelé le *semi-groupe libre* sur l'alphabet $\mathcal{A}$, alors que l'ensemble $\mathcal{A}^*$ est appelé le *monoïde libre*.

Un mot u est appelé un *facteur* (resp. un *préfixe*, resp. un *suffixe*) du mot w s'il existe des mots x, y tels que $w = xuy$ (resp. $w = uy$, resp. $w = xu$). On dit que le facteur (resp. préfixe, resp. suffixe) est *propre* si $xy \neq \varepsilon$ (resp. $y \neq \varepsilon$, resp. $x \neq \varepsilon$). On note Pref(w) (resp. Suff(w)) l'ensemble des préfixes (resp. suffixes) propres de w. L'ensemble des facteurs du mot w est noté $F(w)$ et $F_n(w)$ désigne l'ensemble des facteurs de longueur n du mot w. On désigne le facteur de w commençant à la i-ième lettre et terminant à la j-ième par $w[i, j]$.

Exemple 1.1.1 Soit le mot $w = abbaaabbaa \in \{a, b\}^*$. On a $\lg(w) = 10$. Les facteurs de longueur 3 de w sont donnés par l'ensemble $F_3(w) = \{aaa, abb, baa, bba\}$. On a $abba \in$ Pref(w), mais $abbaaabbaa \notin$ Pref(w) comme il n'est pas un préfixe propre de w. Par ailleurs, $w[2, 4] = baa$ et $w[0, 5] = abbaaa$.

On écrit Alph(w) pour désigner l'ensemble des lettres ayant au moins une occurrence dans le mot w. Le nombre d'occurrences d'une lettre $a \in \mathcal{A}$ dans le mot $w \in \mathcal{A}^*$ est noté $|w|_a$ et de façon similaire, pour un facteur f de w, $|w|_f$ désigne le nombre d'occurrences du facteur f dans le mot w. La *fréquence d'une lettre* a dans un mot w est définie par $f_a(w) = |w|_a/\lg(w)$.

Exemple 1.1.2 Fixons $\mathcal{A} = \{1, 2, 3, 4\}$ et considérons le mot $w = 131431 \in \mathcal{A}^*$. Alors Alph$(w) = \{1, 3, 4\}$ et $f_1(w) = 3/6 = 1/2$, $f_2(w) = 0/6 = 0$, $f_3(w) = 2/6 = 1/3$ et $f_4(w) = 1/6$.

L'*image miroir* d'un mot $w = w[0]w[1]\cdots w[n-1]$ est le mot noté $\widetilde{w}$ et défini par $\widetilde{w} = w[n-1]w[n-2]\cdots w[1]w[0]$. Un *palindrome* est un mot w tel que $\widetilde{w} = w$. L'ensemble des palindromes sur l'alphabet $\mathcal{A}$ est noté pal$(\mathcal{A}^*)$. Si lg(w) est paire (resp. impaire), alors w est un palindrome si et seulement s'il s'écrit comme $w = x\widetilde{x}$ (resp. $w = xa\widetilde{x}$, où a est une lettre) pour un certain mot x. Soient u, v deux palindromes. Alors u est un *facteur central* de v si $v = wu\tilde{w}$ pour $w \in \mathcal{A}^*$. La *fermeture palindromique* de $w \in \mathcal{A}^*$ est le plus court palindrome $u = w^{(+)}$ ayant w comme préfixe. Si $w = xp$, où p est le plus long suffixe palindrome de w, alors $w^{(+)} = xp\widetilde{x}$.

Exemple 1.1.3 Soient les mots $u = 1234123121$ et $v = 1243421$. Alors $\widetilde{u} = 1213214321 \neq u$ et $\widetilde{v} = 1243421 = v$. Ainsi, u n'est pas un palindrome et v en est un de longueur impaire et s'écrit comme $v = x3\widetilde{x}$, avec $x = 124$. D'autre part, $u^{(+)} = 1234123 \cdot 121 \cdot 3214321$ et $v^{(+)} = v$, comme v est un palindrome.

Soit $w \in \mathcal{A}^*$, où Card$(\mathcal{A}) = 2$. Le *complément* du mot $w = w[0]w[1]w[2]\cdots w[n-1]$ est noté $\overline{w}$ et est défini par $\overline{w} = \overline{w}[0]\overline{w}[1]\overline{w}[2]\cdots\overline{w}[n-1]$, où $\overline{w}[i]$ est la lettre complémentaire de $w[i]$ dans l'alphabet $\mathcal{A}$.

Exemple 1.1.4 Le complément de $w = abbaaabbaa$ est $\overline{w} = baabbbaabb$.

On dit que $p \in \mathbb{N}$ est une *période* d'un mot $w = w[0]w[1]\cdots[n-1]$ si $w[i] = w[i+p]$, pour $0 \leq i < n-p$. Si $p = 0$, on dit que la période est *triviale* .

Exemple 1.1.5 Soit le mot $w = abbaaabbaa \in \{a,b\}^*$. w admet les périodes 5 et 9, puisque $w = abbaa \cdot abbaa$ et $w = abbaaabba \cdot a$.

On note w^n, avec $n \in \mathbb{N}$ et $w \in \mathcal{A}^*$, le mot w répété n fois. On dit alors que w^n est la *n-ième puissance* de w. Un mot $w \in \mathcal{A}^+$ est dit *primitif* si $w = u^n$ pour un mot $u \in \mathcal{A}^+$ implique que $n = 1$. D'autre part, pour $w \in \mathcal{A}^m$ et pour $q = n + p/m$ avec $n, p, m \in \mathbb{N}$, $p < m$ et $m \neq 0$, w^q désigne le mot $w^n w[0, p-1]$.

Exemple 1.1.6 Soit $w = 123112$. Alors $w^3 = www = 123112 \cdot 123112 \cdot 123112$. D'autre part, $u = 112112$ n'est pas primitif, puisqu'il s'écrit comme $u = (112)^2$.

Exemple 1.1.7 Soit $w = 1123$. Alors $w^{10/4} = 1123112311$.

Deux mots $u, v \in \mathcal{A}^+$ sont dits *conjugués* s'il existe $x, y \in A^*$ tels que $u = xy$ et $v = yx$. La conjugaison est une relation d'équivalence. En effet, on peut facilement vérifier la réflexivité, la symétrie et la transitivité de la conjugaison. La *classe de conjugaison* du mot w est notée $[w]$. Si w est primitif, alors il a $\lg(w)$ conjugués.

Exemple 1.1.8 a) Le mot $w = aababc$ est un mot primitif. Alors
$[w] = \{aababc, ababca, babcaa, abcaab, bcaaba, caabab\}$ et Card$([w]) = 6 = \lg(w)$.

b) Le mot $u = aabaab = (aab)^2$ est un mot non primitif. Alors
$[u] = \{aabaab, abaaba, baabaa\}$ et Card$([u]) = 3$.

c) Le mot $v = 12112$ est un mot primitif. Alors
$[v] = \{12112, 21121, 11212, 12121, 21211\}$ et Card$([v]) = 5$.

L'*ordre lexicographique* pour deux mots $u, v \in \mathcal{A}^*$, aussi appelé l'*ordre alphabétique*, où $\mathcal{A}$ est un alphabet totalement ordonné, est défini comme suit. On écrit $u < v$ si u est un préfixe propre de v ou s'il existe des factorisations $u = xau'$ et $v = xbv'$ telles que $a < b \in \mathcal{A}$. On écrit $u \leq v$ si $u < v$ ou si $u = v$. Cela correspond à l'ordre dans un dictionnaire. Si l'alphabet est numérique, on suppose que l'ordre sur les lettres est l'ordre des naturels.

Un mot $\ell \in \mathcal{A}^*$ est un *mot de Lyndon* s'il est primitif et s'il est minimal dans sa classe de conjugaison, selon l'ordre lexicographique. Un mot de longueur 1 est nécessairement un mot de Lyndon. De plus, tout mot non vide est nécessairement un conjugué d'une puissance d'un mot de Lyndon. L'ensemble des mots de Lyndon est noté $\mathcal{L}$.

Exemple 1.1.9 a) Soit le mot $v = 12112$ de l'Exemple 1.1.8 c). On a $v \notin \mathcal{L}$, puisqu'il n'est pas minimal dans sa classe de conjugaison. En effet, $11212 \in [v]$ et $11212 < 12112$. Donc, $11212 \in \mathcal{L}$, puisqu'il est primitif et qu'il est le plus petit de sa classe de conjugaison selon l'ordre lexicographique.

b) Soit le mot $abab$. On peut écrire $abab = (ab)^2$. Donc $abab \notin \mathcal{L}$.

c) Soit le mot 23121. On a que 12123 et 23121 sont conjugués et 12123 est le mot minimal de sa classe de conjugaison. Donc $23121 \notin \mathcal{L}$, mais son conjugué $12123 \in \mathcal{L}$.

Rappelons le théorème suivant concernant la factorisation de Lyndon.

Théorème 1.1.10 *(Lothaire, 1983) Tout mot fini non vide w s'écrit de façon unique comme un produit non croissant de mots de Lyndon :*

$$w \quad = \quad \ell_0\ell_1\cdots\ell_n = \bigodot_{i=0}^{n} \ell_i, \text{ où } \quad \ell_i \in \mathcal{L} \text{ et } \quad \ell_0 \geq \ell_1 \geq \cdots \geq \ell_n. \tag{1.1}$$

Exemple 1.1.11 La factorisation de Lyndon du mot $u = 121221122112$ est

$$u = 12122 \cdot 1122 \cdot 112.$$

On a bien $12122 \geq 1122 \geq 112$ et $12122, 1122, 112 \in \mathcal{L}$.

Un mot $w \in \mathcal{A}^*$ est dit *équilibré* si pour tous facteurs u, v de même longueur de w et pour toute lettre $a \in \mathcal{A}$, $\big||u|_a - |v|_a\big| \leq 1$. Pour un alphabet à deux lettres, il suffit de vérifier cette condition pour une seule des deux lettres, puisque par complémentarité, $|u|_a - |v|_a = k$ si et seulement si $|v|_b - |u|_b = k$.

Exemple 1.1.12

a) Soit le mot $w = xyxyyxx$. Ce mot n'est pas équilibré. En effet, $xx, yy \in F_2(w)$, et $\big||xx|_x - |yy|_x\big| = |2-0| = 2 > 1$. Remarquons qu'on a aussi $\big||xx|_y - |yy|_y\big| = |0-2| = 2 > 1$.

b) Soit le mot $u = ababb$. Ce mot est équilibré. En effet,

$$F(u) = \{a, b, ab, ba, bb, aba, abb, bab, abab, babb, ababb\}$$

et on peut facilement vérifier que pour deux facteurs $f, f' \in F(u)$ de même longueur, $\big||f|_a - |f'|_a\big| \leq 1$.

c) Soit le mot $v = 1213$. On a

$$F(v) = \{1, 2, 3, 12, 13, 21, 121, 213, 1213\}.$$

On peut vérifier que pour tous facteurs f, f' tels que $\lg(f) = \lg(f')$, $\big||f|_1 - |f'|_1\big| \leq 1$, $\big||f|_2 - |f'|_2\big| \leq 1$ et $\big||f|_3 - |f'|_3\big| \leq 1$. Ainsi, v est équilibré.

d) Soit le mot $t = 1123$. On a $F(t) = \{1, 2, 3, 11, 12, 23, 112, 123, 1123\}$. Pour tous facteurs f, f' tels que $\lg(f) = \lg(f')$, $\big||f|_2 - |f'|_2\big| \leq 1$, $\big||f|_3 - |f'|_3\big| \leq 1$, mais $\big||11|_1 - |23|_1\big| = |2-0| > 1$. Donc t n'est pas équilibré.

Définition 1.1.13 On note γ le *conjugueur* défini de la manière suivante : pour une lettre y et un mot $w \in \mathcal{A}^*$, $\gamma(yw) = wy$.

1.2 Mots infinis

Un *mot infini* à droite s sur l'alphabet $\mathcal{A}$ est une suite infinie de lettres $s = s[0]s[1]s[2]\cdots$, où $s[i] \in \mathcal{A}$ pour $i \geq 0$. On désigne par $\mathcal{A}^{\mathbb{N}}$ l'ensemble des mots infinis à droite. Cet ensemble est aussi parfois désigné par $\mathcal{A}^{\omega}$. C'est l'ensemble des suites de symboles de $\mathcal{A}$ indicées par les naturels. Nous appelerons *suites* ou *mots infinis* les mots infinis à droite.

On définit $\mathcal{A}^{\infty} = \mathcal{A}^* \cup \mathcal{A}^{\omega}$, l'ensemble des mots finis ou infinis sur l'alphabet $\mathcal{A}$.

La concaténation uv est bien définie pour $u \in \mathcal{A}^*$ et $v \in \mathcal{A}^{\omega}$. Le résultat est une suite. Si $u \in \mathcal{A}^*$, alors u^{ω} désigne la suite $uuu\cdots$.

Un mot fini u est un *facteur* d'une suite s s'il existe $p \in \mathcal{A}^*$, $s' \in \mathcal{A}^{\omega}$ tel que $s = pus'$. L'ensemble des facteurs de s est noté $F(s)$ et l'ensemble de ses facteurs de longueur n est noté $F_n(s)$. Une suite $t \in \mathcal{A}^{\omega}$ est un *suffixe* de la suite $s \in \mathcal{A}^{\omega}$ s'il existe un mot $p \in \mathcal{A}^*$ tel que $s = pt$. Si $p \neq \varepsilon$, on dit que le suffixe est *propre*.

Si $w = pus \in \mathcal{A}^{\infty}$, avec $p, u \in \mathcal{A}^*$ et $s \in \mathcal{A}^{\infty}$, alors $p^{-1}w$ désigne le mot us. De façon similaire, ws^{-1} est le mot pu.

Exemple 1.2.1 Soient $w = 11232121242$, $u = 11231(213)^{\omega}$, $s = 21242$ et $p = 1123$. Alors $p^{-1}w = 2121242$, $p^{-1}u = 1(213)^{\omega}$ et $ws^{-1} = 112321$.

L'*ordre lexicographique* pour les suites $s, t \in \mathcal{A}^{\omega}$ est défini comme suit. On dit que $s < t$ s'il existe des factorisations $s = uas'$ et $t = ubt'$, avec $a < b \in \mathcal{A}$, $u \in \mathcal{A}^*$ et $s', t' \in \mathcal{A}^{\omega}$.

Siromoney et al. (Siromoney et al., 1994) ont généralisé le Théorème 1.1.10 aux mots infinis. L'ensemble $\mathcal{L}_{\infty}$ des *mots de Lyndon infinis* consiste en les mots qui sont plus petits que tous leurs suffixes. La définition de factorisation de Lyndon se généralise aux mots infinis de la façon suivante.

Théorème 1.2.2 (Siromoney et al., 1994) *Tout mot infini w s'écrit de façon unique comme un produit, fini ou infini, non croissant de mots de Lyndon de l'une des deux formes suivantes :*

i) *soit il existe une suite infinie* $(\ell_k)_{k\geq 0}$ *d'éléments de* $\mathcal{L}$ *tel que*

$$w = \ell_0\ell_1\ell_2\cdots \text{ et pour tout } k,\ \ell_k \geq \ell_{k+1}.$$

ii) *soit il existe une suite finie* $\ell_0, \ldots, \ell_m\ (m \geq 0)$ *d'éléments de* $\mathcal{L}$ *et* $\ell_{m+1} \in \mathcal{L}_\infty$ *tel que*

$$w = \ell_0\ell_1\cdots\ell_m\ell_{m+1} \text{ et } \ell_0 \geq \ldots \geq \ell_m > \ell_{m+1}.$$

Exemple 1.2.3 a) La factorisation de Lyndon de la suite $s = 123121133424121^\omega$ est

$$s = 123 \cdot 12 \cdot 113342412 \cdot 1 \cdot 1 \cdot 1 \cdots$$

Cette factorisation est du type i).

b) La factorisation de Lyndon de la suite $t = 1231211334241(12)^\omega$ est

$$t = 123 \cdot 12 \cdot 1133424 \cdot 1(12)^\omega.$$

Cette factorisation est du type ii).

La fonction de *décalage* σ pour les suites est définie par $\sigma : \mathcal{A}^{\mathbb{N}} \to \mathcal{A}^{\mathbb{N}}$ telle que $\sigma(s) = t$, où $t[n] = s[n+1]$, pour $n \in \mathbb{N}$. On a alors $\sigma(s[0]s[1]s[2]\cdots) = s[1]s[2]s[3]\cdots$.

Exemple 1.2.4 Soit $s = (123)^\omega$. Alors $\sigma(s) = 23(123)^\omega = (231)^\omega$ et $\sigma^2(s) = \sigma(\sigma(s)) = \sigma((231)^\omega) = (312)^\omega$.

L'opérateur σ est l'analogue de l'opérateur γ sur les mots infinis. En effet : $\sigma(u^\omega) = (\gamma w)^\omega$.

Soit $w \in \mathcal{A}^\infty$. La *fonction de complexité* du mot w est la fonction qui associe pour chaque longueur $n \in \mathbb{N}$, le nombre $P(w,n)$ de facteurs de longueur n dans le mot w. Ainsi, $P(w,n) =$ Card$(F_n(w))$. On a toujours $P(w,0) = 1$ et $P(w,1) =$ Card(Alph(w)). S'il n'y a pas d'ambiguïté, on écrit simplement $P(n)$.

Exemple 1.2.5 Soit $w = bbabaa$. Alors

$$F(w) = \{a, b, aa, ab, ba, bb, aba, baa, bba, abaa, bbab, babaa, bbaba, bbabaa\}$$

et donc $P(w,1) = 2$, $P(w,2) = 4$, $P(w,3) = 3$, $P(w,4) = 2$, $P(w,5) = 2$ et $P(w,6) = 1$.

Si w est une suite, tout facteur peut être prolongé vers la droite et donc, la propriété $P(w,n) \leq P(w,n+1)$ est toujours satisfaite. Un facteur f de w est dit *spécial à droite* (resp. *spécial à gauche*) dans w s'il existe $a, b \in \mathcal{A}$, $a \neq b$, tels que $fa, fb \in F(w)$ (resp. $af, bf \in F(w)$).

Exemple 1.2.6 Dans le mot $abcabacbc$, le facteur ab est spécial à droite, puisque abc et aba sont aussi facteurs. D'autre part, le facteur bc est spécial à gauche, puisque abc et cbc sont aussi facteurs.

Soit $s \in \mathcal{A}^\omega$. On note $\mathrm{Ult}(s)$ l'ensemble des lettres qui apparaissent infiniment souvent dans s. On dit qu'une suite s est *récurrente* si $|s|_f$ est infini pour tout facteur $f \in F(s)$. On dit que s est *purement périodique* (resp. *ultimement périodique*) si s s'écrit comme $s = u^\omega$ (resp. $s = pu^\omega$, $p \in \mathcal{A}^*$), avec $u \in \mathcal{A}^*$.

Exemple 1.2.7 On a $\mathrm{Ult}(1213121412(142)^\omega)) = \{1,2,4\}$. Cette suite n'est pas récurrente, puisque le facteur 1213 n'apparaît qu'une seule fois, mais elle est ultimement périodique.

La *fréquence* (voir (Pytheas-Fogg, 2002)) de la lettre a dans la suite s est définie par

$$f_a(s) = \lim_{n\to\infty} \frac{|s[0,n-1]|_a}{n}$$

lorsque cette limite existe.

Exemple 1.2.8 Soit $s = 1213121412(142)^\omega$. On a $f_1(s) = f_2(s) = f_4(s) = 1/3$ et $f_3(s) = 0$.

Deux mots $u, v \in \mathcal{A}^\infty$ sont dits *équivalents* si $F(u) = F(v)$: s'ils ont le même ensemble de facteurs.

Soit $w \in \mathcal{A}^\infty$. Un facteur f de w de la forme $f = \alpha^k$, avec $\alpha \in \mathcal{A}$ et $k \geq 1$ maximal, est appelé *bloc* de α de w et sa longueur est k.

Exemple 1.2.9 Soit $w = \underline{1}23\underline{1}3\underline{111}3\underline{1}22\underline{1}3\underline{11}323$. Ce mot contient contient successivement des blocs de 1 de longueur respectivement $1, 1, 3, 1, 1$ et 2.

1.3 Mots bi-infinis

L'ensemble des *mots bi-infinis*, noté $\mathcal{A}^{\mathbb{Z}}$, est défini comme étant l'ensemble des fonctions $\mathbb{Z} \to \mathcal{A}$.

Si $u \in \mathcal{A}^*$, alors ${}^{\omega}u^{\omega}$ désigne le mot bi-infini $s = \cdots u \bullet uu \cdots$. Le point $\bullet$ est placé juste avant $s[0]$ et représente l'*origine* du mot s. On aura alors que $s[i] = u[j]$, où $j \equiv i \mod \lg(u)$.

Pour $s \in \mathcal{A}^{\mathbb{Z}}$, on définit la *fonction de décalage* σ par $\sigma : \mathcal{A}^{\mathbb{Z}} \to \mathcal{A}^{\mathbb{Z}}$ telle que $\sigma(s) = t$, où $t[n] = s[n+1]$, pour $n \in \mathbb{Z}$. Contrairement au décalage d'un mot infini à droite, le décalage d'un mot bi-infini donne une bijection dans $\mathcal{A}^{\mathbb{Z}}$. Un décalage σ^k correspond à un déplacement de l'origine de k positions vers la droite.

Soit G, l'ensemble des entiers n tels que $\sigma^n(s) = s$, avec $s \in \mathcal{A}^{\mathbb{Z}}$. Alors G forme un groupe d'élément neutre 0 et dans lequel l'inverse de n est $-n$. Les éléments de G sont appelés des *périodes*. On définit la *plus petite période non triviale* du mot bi-infini $s \in \mathcal{A}^{\mathbb{Z}}$ comme étant le plus grand commun diviseur p des entiers $n \geq 1$ de G. Ce p correspond exactement au générateur du groupe G. Si Card$(G) > 1$, alors $p \geq 1$. Sinon, $G = \{0\}$ et on dit que la période est *infinie*.

Exemple 1.3.1 Soit le mot bi-infini $s = {}^{\omega}(12334)^{\omega}$ de périodes $5, 10, 15, \ldots = 5\mathbb{N}$ et de plus petite période non triviale 5. On a $\sigma(s) = {}^{\omega}(23341)^{\omega}$ et $\sigma^3(s) = {}^{\omega}(34123)^{\omega}$.

1.4 Autres définitions et notations

Notation 1.4.1 On écrit $\alpha \perp \beta$ si pgcd$(\alpha, \beta) = 1$. Sinon, on écrit $\alpha \not\perp \beta$.

Notation 1.4.2 À moins de mentionner le contraire, $[a, b]$ désigne l'ensemble qui contient tous les entiers de a à b inclusivement, où $a, b \in \mathbb{Z}$. On appelle cet ensemble un *intervalle entier*.

Définition 1.4.3 La *projection* $\Pi_{\alpha}(s)$ de $s \in \mathcal{A}^{\infty}$ par rapport à la lettre α est définie par $\Pi_{\alpha}(s)[i] = \alpha$ si $s[i] = \alpha$, et $\Pi_{\alpha}(s)[i] = x$ sinon.

Exemple 1.4.4 Soit $w = 123124321$. Alors $\Pi_1(w) = 1xx1xxxx1$ et $\Pi_2(w) = x2xx2xx2x$.

Chapitre II

MOTS ÉPICHRISTOFFELS

Les suites sturmiennes, aussi appelées les mots de Sturm, apparaissent dans la littérature dès le 18ième siècle dans les travaux précurseurs de l'astronome Bernoulli (Bernoulli, 1772) : ces suites caractériseraient les phases de la lune. Selon Markov (Markov, 1882), il considère la suite $\{F(\alpha n+\rho)\}_{n\in\mathbb{Z}}$, où $F(x)$ désigne l'entier le plus proche de $x \in \mathbb{R}$, et tente de déterminer pour quelles valeurs de $\alpha \in \mathbb{R}$ la suite $\{F(\alpha(n+1)+\rho) - F(\alpha n+\rho)\}_{n\in\mathbb{Z}}$ est périodique. Il conjecture que c'est le cas si et seulement si α est rationnel.

On retrouve ensuite les suites sturmiennes au 19ième siècle dans les travaux de Christoffel (Christoffel, 1875), puis dans ceux de Markov (Markov, 1882). Ce dernier prouve la conjecture de Bernoulli. La première étude en profondeur de ces suites demeure toutefois celle de Morse et Hedlund. Dans leurs travaux (Morse et Hedlund, 1938; Morse et Hedlund, 1940; Hedlund, 1944), ils montrent comment les suites sturmiennes s'obtiennent en considérant les zéros des solutions d'équations différentielles homogènes linéaires de degré 2

$$y'' + h(x)y = 0,$$

où $h(x)$ est continu et de période 1. C'est d'ailleurs dans leurs travaux que le nom *suites sturmiennes* apparaît pour la première fois dans la littérature.

À la fin du 20ième siècle et plus récemment, une foule de chercheurs se sont intéressés à ces suites. Par exemple (Coven et Hedlund, 1973; Coven, 1974; Storlarsky, 1976; Brown, 1993; Ziccardi, 1995; de Luca, 1997a; Bender, Patashnik et Rumsey, 1994; Berstel, 2002). Trois livres récents soulignent cet intérêt (Lothaire, 2002; Pytheas-Fogg, 2002; Allouche et Shal-

lit, 2003). Dans cette vaste littérature, on retrouve plusieurs caractérisations des suites sturmiennes. Entre autres, elles sont les suites infinies sur deux lettres de complexité minimale, elles sont les suites équilibrées sur deux lettres, elles codent des droites discrètes, elles codent l'orbite d'un point de l'intervalle réel $[0, 1[$ sous l'action d'une rotation d'angle irrationnel suivant une partition de $[0, 1[$ en deux intervalles de longueurs respectives α et $1 - \alpha$. Ces différentes caractérisations proviennent du fait que les suites sturmiennes apparaissent dans une multitude de domaines tels que la théorie des nombres (Morikawa, 1985b; Simpson, 1991; Tijdeman, 2000b; Tijdeman, 2000a; Barát et Varjú, 2003; Simpson, 2004; Graham et O'Bryant, 2005), la géométrie discrète, la cristallographie (Bombieri et Taylor, 1986), la reconnaissance de formes, la dynamique symbolique (Morse et Hedlund, 1938; Morse et Hedlund, 1940; Hedlund, 1944; Queffélec, 1987), l'infographie (Bresenham, 1991). Selon le domaine, on retrouve différentes appelations pour les suites sturmiennes ou pour une sous classe : suites de Beatty, mots de coupures, suites caractéristiques, mots de Christoffel, mots de billard.

Depuis la fin des années 1990, plusieurs chercheurs ont généralisé les suites sturmiennes à des suites à plus de deux lettres, en utilisant l'une de ses multiples caractérisations. Une généralisation naturelle sur trois lettres et plus est la famille des suites *épisturmiennes*. Cette dernière utilise la propriété de fermeture palindromique des suites sturmiennes. La première construction de ces mots apparaît dans (de Luca, 1997b). Cette classe apparaît aussi dans (Rauzy, 1985; Arnoux et Rauzy, 1991) et a été étudiée plus récemment par (Justin et Vuillon, 2000; Risley et Zamboni, 2000; Droubay, Justin et Pirillo, 2001; Justin et Pirillo, 2002; Justin et Pirillo, 2004; Justin, 2005; Glen, 2008; Glen, 2007). Certains auteurs (Cassaigne, Ferenczi et Zamboni, 2000; Berstel, 2002; Vuillon, 2003) se sont intéressés à une généralisation de la notion d'équilibre sur des suites sur un alphabet à plus de trois lettres et il a été prouvé qu'il existe toujours une suite d'Arnoux-Rauzy qui ne soit pas équilibrée, selon la notion d'équilibre généralisée. Une autre généralisation des suites sturmiennes consiste en les mots de billard en trois dimensions (Arnoux et al., 1994; Borel et Reutenauer, 2005). Remarquons que pour un alphabet à deux lettres, ces trois classes coïncident avec la famille des suites sturmiennes. Par contre, dès que l'alphabet comporte plus de deux lettres, elles ne coïncident plus.

La version finitaire des suites sturmiennes, les mots de Christoffel, s'avère aussi un objet fort

étudié (Christoffel, 1875; Lothaire, 2002; Borel et Reutenauer, 2006; Berthé, de Luca et Reutenauer, 2007; Kassel et Reutenauer, 2007). Malgré que les facteurs finis de suites épisturmiennes aient été étudiés entre autres par (Glen, Justin et Pirillo, 2008), personne n'a encore introduit une généralisation des mots de Christoffel. Dans ce chapitre, nous introduisons une généralisation des mots de Christoffel que nous appelons les mots *épichristoffels* en utilisant les morphismes épisturmiens. Dans un premier temps, nous rappelons ce que sont les suites sturmiennes. Ensuite, nous introduisons la version finie des suites sturmiennes : les mots de Christoffel. Puis, nous nous intéressons à la généralisation naturelle des suites sturmiennes : les suites épisturmiennes. Nous présentons ensuite la nouvelle classe de mots : les *mots épichristoffels*. Malgré que ces mots ne soient généralement pas équilibrés, nous montrons de quelle façon certaines propriétés des mots de Christoffel se généralisent pour un alphabet à plus de deux lettres.

2.1 Suites sturmiennes

Dans cette section, nous n'introduisons que les propriétés des suites sturmiennes qui nous intéressent pour la suite. Pour plus de propriétés, nous référons le lecteur à la section du (Lothaire, 2002) qui est consacrée à l'étude de ces suites.

L'une des définitions classiques des suites sturmiennes est celle de Morse et Hedlund.

Définition 2.1.1 (Morse et Hedlund, 1940) Soit α et ρ deux nombres réels, avec $0 \leq \alpha < 1$ irrationnel, et posons, pour $n \geq 0$,

$$s[n] = \begin{cases} a \text{ si } \lfloor \alpha(n+1) + \rho \rfloor = \lfloor \alpha n + \rho \rfloor, \\ b \text{ sinon,} \end{cases}$$

$$s'[n] = \begin{cases} a \text{ si } \lceil \alpha(n+1) + \rho \rceil = \lceil \alpha n + \rho \rceil, \\ b \text{ sinon.} \end{cases}$$

Alors les deux suites

$$s_{\alpha,\rho} = s[0]s[1]s[2]\cdots \qquad \text{et} \qquad s'_{\alpha,\rho} = s'[0]s'[1]s'[2]\cdots$$

sont *sturmiennes* et réciproquement, tout mot sturmien est de la forme $s_{\alpha,\rho}$ ou $s'_{\alpha,\rho}$ pour un nombre irrationnel α et un réel ρ. On appelle *la pente* le réel α et *l'intercept* le réel ρ.

Définition 2.1.2 Soit une suite sturmienne s. Si l'intercept est 0, alors on dit que s est *standard* ou *caractéristique*.

Définition 2.1.3 (Lothaire, 2002) La suite sturmienne $s_{\alpha,\rho}$ (resp. $s'_{\alpha,\rho}$) est appelée le *mot mécanique inférieur* (resp. *mot mécanique supérieur*) de pente α et d'intercept ρ.

Cette définition provient de la représentation géométrique suivante. Considérons une droite d'équation $y = \alpha x + \rho$ (voir la Figure 2.1). Les points à coordonnées entières tout juste sous la droite sont les points $P_n = (n, \lfloor \alpha n + \rho \rfloor)$. Deux points consécutifs P_n et P_{n+1} sont reliés par une ligne droite horizontale si $s_{\alpha,\rho}(n) = a$ et diagonale si $s_{\alpha,\rho}(n) = b$. On a la même chose pour $P_n = (n, \lceil \alpha n + \rho \rceil)$ pour les points situés juste au dessus de la droite. Le mot associé à cette droite est $baabaabaabaabaaa\cdots$.

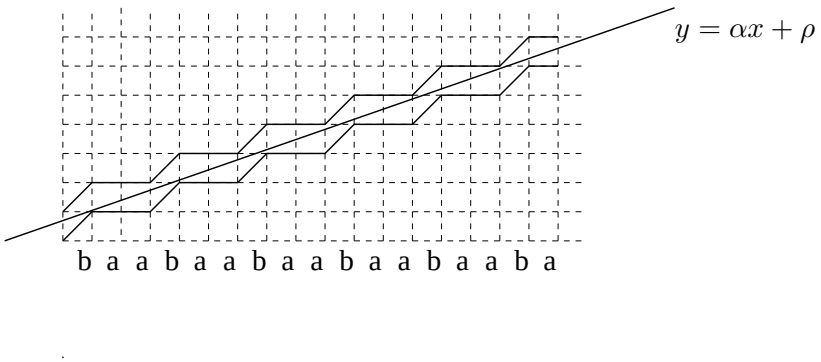


Figure 2.1 Exemple de mots mécaniques supérieurs et inférieurs.

Notation 2.1.4 Dans ce travail, nous ne considérons que les mots de pente irrationnelle, appelé des mots mécaniques *irrationnels*. Il existe aussi des mots mécaniques *rationnels* : les mots ayant une pente rationnelle.

Proposition 2.1.5 *(Morse et Hedlund, 1938) Une suite s est sturmienne si et seulement si elle est apériodique et équilibrée.*

De plus, on a la Proposition 2.1.6.

Proposition 2.1.6 *(Lothaire, 2002) Un mot fini w est facteur d'une suite sturmienne si et seulement s'il est équilibré.*

Les suites sturmiennes sont caractérisées par leur complexité minimale. En effet, on a le résultat suivant.

Proposition 2.1.7 *(Coven et Hedlund, 1973) Une suite s est sturmienne si et seulement si $P(n) = n + 1$, pour tout $n \in \mathbb{N}$.*

Théorème 2.1.8 *(Lothaire, 2002) Soit s une suite. Les énoncés suivants sont équivalents :*
i) *s est sturmienne ;*
ii) *s est équilibrée et apériodique ;*
iii) *s est un mot mécanique irrationnel.*

Une *suite sturmienne standard* correspond au mot mécanique décrit par une droite débutant à un point entier.

Définition 2.1.9 (Lothaire, 2002) Un *morphisme* f est *sturmien* si $f(s)$ est une suite sturmienne pour toute suite sturmienne s.

Exemple 2.1.10 Le morphisme identité et le morphisme qui échange les deux lettres de l'alphabet sont des morphismes sturmiens.

Proposition 2.1.11 *(Séébold, 1991) Les morphismes ψ et $\overline{\psi}$ définis par*

$$\begin{aligned} \psi(a) = ab \quad & \psi(b) = a \\ \overline{\psi}(a) = ba \quad & \overline{\psi}(b) = a \\ E(a) = b \quad & E(b) = a \end{aligned}$$

sont sturmiens.

Proposition 2.1.12 *(Lothaire, 2002) L'ensemble des morphismes sturmiens est le monoïde engendré par les morphismes $\psi, \overline{\psi}, E$ sous la composition.*

2.2 Mots de Christoffel

Informellement, un mot $w \in \{a, b\}^*$ est un *mot de Christoffel* s'il peut être obtenu en discrétisant un segment de droite dans le plan, comme dans la Figure 2.2, en prenant cette fois des sauts horizontaux et verticaux.

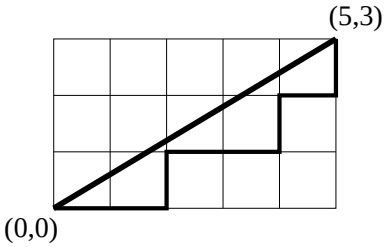


Figure 2.2 Le mot de Christoffel $aabaabab$ de pente $3/5$.

Tout mot fini w sur un alphabet ordonné à deux lettres $\mathcal{A} = \{a, b\}$ définit un chemin continu dans le plan, partant de l'origine $(0, 0)$ et allant jusqu'à un point $(p, q) \in \mathbb{N}^2$. On part du point $(0, 0)$ et en lisant le mot w, on fait un pas horizontal vers la droite pour chaque lettre a, et on fait un pas vertical vers le haut pour chaque lettre b. Ainsi, si $|w|_a = p$ et $|w|_b = q$, alors le chemin se termine au point (p, q).

Définition 2.2.1 On dit qu'un chemin décrit un *mot de Christoffel* (inférieur) s'il satisfait les 3 conditions suivantes :

i) il est situé sous le segment de droite reliant $(0, 0)$ et (p, q), où $p, q \in \mathbb{N}$, et il joint $(0, 0)$ à (p, q) ;

ii) aucun point de $\mathbb{N}^2$ ne se situe entre le chemin et le segment ;

iii) $p \perp q$.

On dit alors que le mot de Christoffel w est de *pente* q/p ; cela correspond exactement à la pente du segment de droite représentant le mot de Christoffel. De plus, on a toujours que $\lg(w) = p + q$. Par définition, une lettre est bien un mot de Christoffel. On dit qu'un mot de Christoffel est *propre* s'il n'est pas réduit à une seule lettre.

Il existe aussi des *mots de Christoffel supérieurs* qui sont situés au dessus du segment de droite, mais nous ne nous y intéressons pas dans ce travail. Nous ne considérons que les *mots de Christoffel inférieurs*, c'est-à-dire ceux situés sous le segment de droite.

Cette définition d'un mot de Christoffel est celle du point de vue de la géométrie discrète introduite dans (Borel et Laubie, 1993). En dynamique symbolique, les mots de Christoffel sont définis à l'aide des échanges d'intervalles (Morse et Hedlund, 1940) de la façon qui suit.

Définition 2.2.2 (Morse et Hedlund, 1940) Soit l'alphabet $\mathcal{A} = \{a < x\}$, soient $\alpha, \beta \in \mathbb{N}$ tels que $\alpha \perp \beta$ et posons $n = \alpha + \beta$. Le *mot de Christoffel* $u \in \mathcal{A}^*$ avec α occurrences de a et β occurrences de x est défini par $u = u[0]u[1]\cdots u[n-1]$, où

$$u[i] = \begin{cases} a & \text{si } i\beta \mod n < \alpha \mod n \\ x & \text{si } i\beta \mod n \geq \alpha \mod n \end{cases}$$

pour $0 \leq i < n$, où $i\beta \mod n$ désigne le reste de la division euclidienne de $i\beta$ par n.

Comme modulo n on a

$$i\beta < \alpha \iff i\beta < n - \beta \iff i\beta < i\beta + \beta = (i+1)\beta,$$

on obtient la définition équivalente suivante.

Définition 2.2.3 Soit l'alphabet $\mathcal{A} = \{a < x\}$, soient $\alpha, \beta \in \mathbb{N}$ tels que $\alpha \perp \beta$ et posons $n = \alpha + \beta$. Le *mot de Christoffel* $u \in \mathcal{A}^*$ avec α occurrences de a et β occurrences de x est défini par $u = u[0]u[1]\cdots u[n-1]$, où

$$u[i] = \begin{cases} a & \text{si } (i+1)\beta \mod n > i\beta \mod n \\ x & \text{si } (i+1)\beta \mod n \leq i\beta \mod n \end{cases}$$

pour $0 \leq i < n$, où $i\beta \mod n$ désigne le reste de la division euclidienne de $i\beta$ par n. Ce mot est de pente β/α.

Remarquons que la Définition 2.2.3 peut se généraliser aux puissances de mots de Christoffel en supprimant la condition de primalité entre α et β.

Proposition 2.2.4 *Un mot de Christoffel est toujours primitif.*

Preuve Soit un mot de Christoffel $w \in \{a, x\}^*$ tel que $|w|_a = p$ et $|w|_x = q$. Supposons w non primitif. Il existe donc u tel que $w = u^n$, avec $n \geq 2$. Mais alors, $p = np'$ et $q = nq'$, avec $n \geq 2$. Donc pgcd$(p, q) \geq n \geq 2$. D'où une contradiction, comme par définition d'un mot de Christoffel, $p \perp q$. ■

Définition 2.2.5 Soit $C(n, \alpha)$ un mot de longueur n sur $\{a < x\}^*$ ayant α occurrences de la lettre a et posons $r =$ pgcd(n, α).

i) Si $r = 1$, alors $C(n, \alpha)$ désigne le mot de Christoffel de pente $\dfrac{n-\alpha}{\alpha}$.

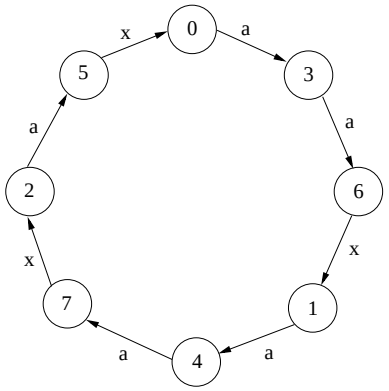
ii) Si $r > 1$, alors $C(n, \alpha) = C(r\frac{n}{r}, r\frac{\alpha}{r}) = \left(C\left(\frac{n}{r}, \frac{\alpha}{r}\right)\right)^r$ et représente la r-ième puissance du mot de Christoffel $C\left(\frac{n}{r}, \frac{\alpha}{r}\right)$.

Considérons le graphe orienté ayant l'ensemble de sommets $\{0, 1, 2, \ldots, \alpha + \beta - 1\}$ et comportant une flèche du sommet i au sommet j si $i + \beta \equiv j \mod n$ étiquetée a si $i < j$, et x si $j < i$.

i) Si $\alpha \perp \beta$, ce graphe est appelé le *graphe de Cayley* du mot de Christoffel u sur l'alphabet $\{a < x\}$ de pente $\dfrac{\beta}{\alpha}$.

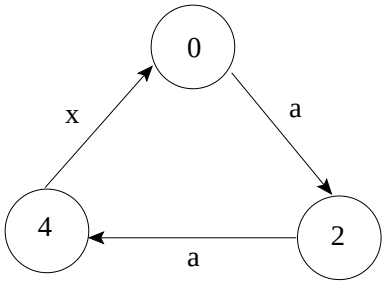
ii) Si pgcd$(\alpha, \beta) = r > 1$, alors le graphe ainsi obtenu est isomorphe au graphe de Cayley du mot de Christoffel $C\left(\dfrac{\alpha+\beta}{r}, \dfrac{\alpha}{r}\right)$. Ce graphe parcouru r fois est le graphe de Cayley de $C(\alpha + \beta, \alpha)$.

Exemple 2.2.6 Le graphe de Cayley associé au mot de Christoffel sur $\{a < x\}$ de pente $3/5$ est

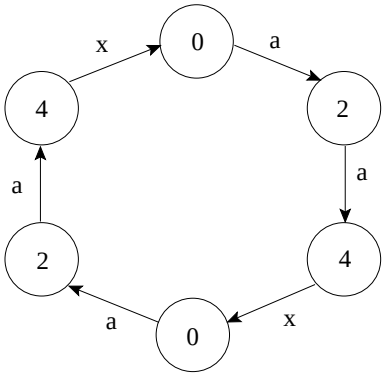


et $u = aaxaaxax$, $|u|_a = 5, |u|_x = 3$.

Exemple 2.2.7 Le graphe de Cayley associé au mot $C(6,4) = aaxaax$ sur $\{a < x\}$ est le suivant,



parcouru deux fois. Son graphe de Cayley peut aussi être représenté par



Soit $C(n,\alpha)$ un mot de Christoffel et posons $\alpha + \beta = n$. Le graphe de Cayley de $C(n,\alpha)$ est le graphe du sous-groupe engendré par $(n-\alpha)$ dans $\mathbb{Z}/n\mathbb{Z}$. Ce sous-groupe est exactement $\mathbb{Z}/n\mathbb{Z}$, puisque $\alpha \perp n$ comme un mot de Christoffel est primitif, donc $(n-\alpha) \perp n$. De plus, on peut vérifier que le graphe de Cayley du mot de Christoffel $C(n,\alpha)$ a bien α occurrences de a et β occurrences de x. En effet, comme $\alpha + \beta = n$, on a $(i-1)\beta \mod n = j$ si et seulement si soit $i\beta \mod n = j + \beta$, soit $i\beta \mod n = j - \alpha$. Si $i\beta \mod n = j + \beta$, alors $C(n,\alpha)[i] = a$, comme $i\beta \mod n > (i-1)\beta \mod n$. Sinon, $i\beta \mod n = j - \alpha$, et donc $i\beta \mod n < (i-1)\beta \mod n$ implique que $C(n,\alpha)[i] = x$. Mais $i\beta \mod n = j + \beta$ pour α valeurs de j et $i\beta \mod n = j - \alpha$ pour β valeurs de j. Ainsi, le mot obtenu par le graphe de Cayley a bien α occurrences de a et β occurrences de x.

Lemme 2.2.8 *L'image miroir d'un mot de Christoffel (resp. une puissance d'un mot de Christoffel)* $C(n,\alpha) \in \{a < x\}^*$*, noté* $\widetilde{C}(n,\alpha)$*, est aussi un mot de Christoffel (resp. une puissance d'un mot de Christoffel) sur le même alphabet, mais pour lequel l'ordre des lettres est inversé. Plus précisément,* $\widetilde{C}(n,\alpha) = C(n, n-\alpha) \in \{x < a\}^*$.

Preuve Considérons le graphe de Cayley du mot $C(n, n-\alpha) \in \{x < a\}^*$. Par définition, ce graphe a une flèche du sommet i au sommet j si $i+\alpha \equiv j \mod n$ et cette flèche est étiquetée x si $i < j$ et a sinon. D'autre part, remarquons que le graphe de Cayley de $\widetilde{C}(n,\alpha)$ est le graphe de Cayley de $C(n,\alpha)$ pour lequel le sens des flèches est inversé. Si l'on change le sens des flèches dans le graphe de Cayley de $C(n,\alpha)$, alors dans le graphe ainsi obtenu, il y a une flèche du sommet i au sommet j si $i+\alpha \equiv j \mod n$ et elle sera étiquetée x si $i < j$ et a sinon. ■

Lemme 2.2.9 *(Berstel et de Luca, 1997) Un mot* w *est un mot de Christoffel si et seulement si* w *est un mot de Lyndon équilibré.*

Lemme 2.2.10 *Dans la classe de conjugaison d'un mot de Christoffel, le mot de Lyndon, c'est-à-dire le plus petit mot selon l'ordre lexicographique, est le mot de Christoffel.*

Le Théorème suivant apparaît dans (de Luca et de Luca, 2006).

Théorème 2.2.11 *Soit* $w \in \mathcal{A}^*$ *un mot primitif tel que tout* $w' \in [w]$ *est facteur d'une suite sturmienne (pas nécessairement la même suite sturmienne). Alors,* w *est conjugué à un mot de Christoffel.*

Preuve Supposons que tout conjugué de w est facteur de suites sturmiennes. Alors, par la Proposition 2.1.6, w et ses conjugués sont tous équilibrés. Comme w est primitif, il existe un mot de Lyndon w' dans la classe de conjugaison de w. Plus particulièrement, ce mot de Lyndon w' est équilibré et donc, par la Proposition 2.2.9, est un mot de Christoffel. Ainsi, w est dans la classe de conjugaison d'un mot de Christoffel. ■

Sans jamais n'avoir été énoncée explicitement, la proposition suivante découle des travaux de Séébold, entre autres (Séébold, 1996; Séébold, 1998), et de Richomme, Kassel et Reutenauer (Richomme, 2007; Kassel et Reutenauer, 2007).

Proposition 2.2.12 *Les mots de Christoffel et leurs conjugués sont exactement les mots obtenus par l'application d'un morphisme sturmien sur une lettre.*

Définition 2.2.13 (de Luca et de Luca, 2006) Soit le mot $w \in \mathcal{A}^*$ et soit p, la plus petite période non triviale de w. On appelle *racine fractionnaire* de w le préfixe z_w de w de longueur p.

La notion de racine fractionnaire est bien définie : si la plus petite période est $\lg(w)$, on a $z_w = w$.

Théorème 2.2.14 *(de Luca et de Luca, 2006) Soit w un mot non vide. Les conditions suivantes sont équivalentes :*

i) *w est facteur d'une suite sturmienne ;*

ii) *la racine fractionnaire z_w de w est conjugué à un mot de Christoffel.*

2.3 Suites épisturmiennes

L'ensemble des suites épisturmiennes a d'abord été obtenu par une construction de de Luca (de Luca, 1997a), puis a été étudié par Rauzy (Rauzy, 1985) et par Arnoux-Rauzy (Arnoux et Rauzy, 1991). Plus récemment, on retrouve entre autres les travaux (Justin et Vuillon, 2000; Droubay, Justin et Pirillo, 2001; Justin, 2005).

Rappelons la définition de suite épisturmienne standard introduite par Droubay, Justin et Pirillo.

Définition 2.3.1 (Droubay, Justin et Pirillo, 2001) Une suite s est dite *épisturmienne standard* si elle satisfait à l'une des conditions équivalentes qui suit.

i) Pour tout préfixe u de s, $u^{(+)}$ est aussi préfixe de s.

ii) Toute première occurrence d'un palindrome dans s est un facteur central d'un palindrome préfixe de s.

iii) Il existe une suite $u_0 = \varepsilon, u_1, u_2, \ldots$ de palindromes et une suite $\Delta(s) = x[0]x[1]\cdots$, avec $x[i] \in \mathcal{A}$, telle que tout u_n défini par $u_{n+1} = (u_n x[n])^{(+)}$, $n \geq 0$, avec $u_0 = \varepsilon$, est préfixe de s.

Pour la preuve de l'équivalence des conditions de la Définition 2.3.1, nous référons le lecteur à (Droubay, Justin et Pirillo, 2001).

Définition 2.3.2 (Droubay, Justin et Pirillo, 2001) Une suite t est *épisturmienne* si $F(t) = F(s)$ pour une suite épisturmienne standard s.

Notation 2.3.3 (Justin, 2005) Soit $w = w[0]w[1]\cdots w[n-1]$, $w[i] \in \mathcal{A}$, et $u_0 = \varepsilon$, ..., $u_{n+1} = (u_n w[n])^{(+)}$, les préfixes palindromiques de u_{n+1}. Alors Pal(w) désigne le mot u_{n+1}.

Dans la Définition 2.3.1, le mot $\Delta(s)$ est appelé la *suite directrice* de la suite épisturmienne standard s. On écrit alors $s = \text{Pal}(\Delta(s))$.

L'exemple suivant illustre comment la suite directrice nous permet de construire facilement des suites épisturmiennes standards.

Exemple 2.3.4 Sur l'alphabet $\mathcal{A} = \{1, 2, 3\}$, la *suite de Tribonacci* T est épisturmienne standard et est déterminée par la suite directrice $\Delta(T) = (123)^\omega$. En effet, on trouve
$u_0 = \varepsilon$,
$u_1 = \underline{1}$,
$u_2 = (1\underline{2})^{(+)} = 1\underline{2}1$,
$u_3 = (121\underline{3})^{(+)} = 121\underline{3}121$,
et finalement, $T = \underline{12}\underline{3}121\underline{1}2131212\underline{1}31211213121\cdots = \text{Pal}((123)^\omega)$.

Dans l'exemple précédent comme dans tous les exemples à venir, les lettres de la suite directrice sont soulignées dans la suite épisturmienne standard correspondante pour des fins de clarté.

Rappelons de (Justin, 2005) une propriété utile de l'opérateur Pal.

Lemme 2.3.5 *(Justin, 2005) Soit $x \in \mathcal{A}$, $w \in \mathcal{A}^*$. Si $|w|_x = 0$, alors* Pal$(wx) =$ Pal$(w)x$Pal(w). *Sinon, on écrit $w = w_1 x w_2$ avec $|w_2|_x = 0$. Le plus long préfixe palindromique de* Pal(w) *qui est suivi de x dans* Pal(w) *est donc* Pal(w_1). *D'où* Pal$(wx) =$ Pal(w)Pal$(w_1)^{-1}$Pal(w).

Dans le lemme précédent, rappelons du Chapitre 1 que si $w = pus \in \mathcal{A}^\infty$, avec $p, u \in \mathcal{A}^*$ et $s \in \mathcal{A}^\infty$, alors $p^{-1}w$ désigne le mot us.

Exemple 2.3.6 Soit $w = \text{Pal}(123) = 1213121$. Alors,

$$\begin{aligned}
\text{Pal}(123 \cdot 4) &= \text{Pal}(123) \cdot 4 \cdot \text{Pal}(123) \\
&= 121312141213121, \\
\text{Pal}(123 \cdot 2) &= \text{Pal}(123) \cdot \text{Pal}^{-1}(w_1) \cdot \text{Pal}(123) \\
&= 1213121(1)^{-1}1213121 \\
&= 1213121213121 \quad \text{et} \\
\text{Pal}(123 \cdot 3) &= \text{Pal}(123) \cdot \text{Pal}^{-1}(w_1) \cdot \text{Pal}(123) \\
&= 1213121(121)^{-1}1213121 \\
&= 12131213121.
\end{aligned}$$

Lemme 2.3.7 *(Droubay, Justin et Pirillo, 2001) Une suite épisturmienne non périodique s sur un alphabet à k lettres est de complexité $P(n) = (h-1)n + q$, avec $q \in \mathbb{N} - \{0\}$ et $h = \text{Card}(\text{Ult}(\Delta(s)))$.*

Définition 2.3.8 (Arnoux et Rauzy, 1991) Une suite s est une *suite d'Arnoux-Rauzy* si elle est uniformément récurrente et qu'elle admet exactement un seul facteur spécial à droite et un seul facteur spécial à gauche pour chaque longueur.

Proposition 2.3.9 *(Justin et Pirillo, 2002) Sur un alphabet à k lettres, toute suite d'Arnoux-Rauzy est de complexité $(k-1)n+1$, mais toutes les suites de complexité $(k-1)n+1$ ne sont pas nécessairement des suites d'Arnoux-Rauzy.*

Définition 2.3.10 Une suite épisturmienne standard $s \in \mathcal{A}^\omega$ ou n'importe quelle suite épisturmienne équivalente est dite $\mathcal{B}$-*stricte* si $\text{Ult}(s) = \text{Alph}(s) = \mathcal{B} \subseteq \mathcal{A}$: toute lettre dans $\mathcal{B} = \text{Alph}(s)$ apparaît infiniment souvent dans la suite directrice $\Delta(s)$.

Plus particulièrement, les suites épisturmiennes $\mathcal{B}$-strictes correspondent aux suites d'Arnoux-Rauzy (Arnoux et Rauzy, 1991).

Définition 2.3.11 Un morphisme f est *épisturmien* si $f(s)$ est une suite épisturmienne pour toute suite épisturmienne s.

Définition 2.3.12 Soit $\mathcal{A}$, un alphabet à k lettres. Pour toute lettre $a, b \in \mathcal{A}$, on définit les endormorphimes de $\mathcal{A}^*$ suivant :
i) $\psi_a(a) = \overline{\psi}_a(a) = a$;
ii) $\psi_a(x) = ax$, si $x \in \mathcal{A} - \{a\}$;
iii) $\overline{\psi}_a(x) = xa$, si $x \in \mathcal{A} - \{a\}$;
iv) $\theta_{ab}(a) = b$, $\theta_{ab}(b) = a$, $\theta_{ab}(x) = x$, $x \in \mathcal{A} - \{a, b\}$.

Rappelons la définition suivante de Justin et Pirillo.

Définition 2.3.13 (Justin et Pirillo, 2002) L'ensemble $\mathscr{E}$ des *morphismes épisturmiens* est le monoïde engendré par les morphismes $\psi_a, \overline{\psi}_a, \theta_{ab}$ sous la composition. L'ensemble $\mathscr{S}$ des morphismes épisturmiens standards est le sous-monoïde engendré par ψ_a et θ_{ab}.

Théorème 2.3.14 *(Justin et Pirillo, 2002) La suite s est épisturmienne standard si et seulement s'il existe une suite épisturmienne standard t et une lettre a tels que $s = \psi_a(t)$. De plus, t est unique et sa suite directrice satisfait $\Delta(s) = a\Delta(t)$.*

2.4 Mots épichristoffels

Dans cette section, nous généralisons les mots de Christoffel à un alphabet à k lettres et nous appelons cette généralisation des *mots épichristoffels*. Nous utiliserons la propriété énoncée dans le Lemme 2.2.12.

Définition 2.4.1 Un mot fini $w \in \mathcal{A}^*$ est un *mot épichristoffel* s'il est obtenu par morphismes épisturmiens et qu'il est le plus petit de sa classe de conjugaison, selon l'ordre lexicographique.

Définition 2.4.2 Un mot fini $w \in \mathcal{A}^*$ est dans une *classe épichristoffelle* s'il est conjugué à un mot épichristoffel.

On peut facilement se convaincre que tout mot appartenant à une classe épichristoffelle peut toujours s'obtenir par morphismes épisturmiens sur une lettre.

Exemple 2.4.3 Soit $\mathcal{A} = \{a, b, c\}$. Alors le mot $\psi_a(\psi_b(\psi_c(a))) = \psi_a(\psi_b(ca)) = \psi_a(bcba) = abacaba$ est dans une classe épichristoffelle, mais n'est pas un mot épichristoffel, comme son conjugué $aabacab < abacaba$. Le mot $aabacab$ est un mot épichristoffel.

Exemple 2.4.4 Soit $\mathcal{A} = \{1, 2, 3, 4, 5\}$. Alors le mot $u = 1213121412131215$ est dans une classe épichristoffelle. En effet, on peut vérifier que 12131412131215 est obtenu par le morphisme $\psi_{1234}(5)$. Par ailleurs, comme u est plus petit que tous ses conjugués, il est un mot épichristoffel.

Dans (Justin et Pirillo, 2002), les auteurs ont utilisé des mots épichristoffels sans mentionner qu'il s'agit d'une généralisation des mots de Christoffel. Dans leur article, les conjugués de mots épichristoffels sont notés h_n. Ils ont prouvé deux propriétés qui montrent clairement que les h_n sont une généralisation des mots de Christoffel. La proposition suivante rappelle ces propriétés.

Proposition 2.4.5 ((Justin et Pirillo, 2002), prop. 2.8, prop. 2.12) *Tout mot fini dans une classe épichristoffelle est primitif et s'écrit de façon unique comme un produit de deux palindromes.*

2.4.1 k-tuplets épichristoffels

Rappelons que pour un couple (p, q) donné, il existe un mot de Christoffel avec des fréquences de lettres p et q si et seulement si $p \perp q$. De plus, il est possible de le construire facilement en utilisant les graphes de Cayley. Dans cette sous-section, on donne un algorithme qui détermine si pour un k-tuplet $p = (p_0, p_1, \ldots, p_{k-1})$, il existe ou non un mot épichristoffel w sur l'alphabet $\mathcal{A} = \{a_0, a_1, \ldots, a_{k-1}\}$ tel que $p = (|w|_{a_0}, |w|_{a_1}, \ldots, |w|_{a_{k-1}})$ représente les fréquences des lettres.

Définition 2.4.6 Soit $p = (p_0, p_1, \ldots, p_{k-1})$ un k-tuplet d'entiers non négatifs. Alors, l'opérateur $T : \mathbb{N}^k \to \mathbb{Z}^k$ est défini sur un k-tuplet p comme

$$T(p) = T(p_0, p_1, \ldots, p_{k-1}) = (p_0, p_1, \ldots, p_{i-1}, \left(p_i - \sum_{j=0, j\neq i}^{k-1} p_j \right), p_{i+1}, \ldots, p_{k-1}),$$

où $p_i \geq p_j$, $\forall\, j \neq i$.

Proposition 2.4.7 *Soit p un k-tuplet. Il existe un mot appartenant à une classe épichristoffelle ayant les fréquences de lettres p si et seulement si l'itération de l'opérateur T sur p retourne un k-tuplet p' avec $p'_j = 0$ pour tout j sauf pour $j = m$ où $p'_m = 1$.*

La proposition précédente et la définition de l'opérateur T sont inspirées de l'algorithme calculant le plus grand diviseur commun de 3 entiers décrit à la section 3 de (Castelli, Mignosi et Restivo, 1999) utilisé pour la généralisation du théorème de Fine et Wilf à 3 périodes et des tuplets décrits dans (Justin, 2000).

Avant de faire la preuve de la Proposition 2.4.7, certains lemmes sont nécessaires.

Lemme 2.4.8 *Soit $w = \psi_{a_0}(u)$, avec $w, u \in \mathcal{A}^*$, ψ un morphisme épisturmien, $a_0 \in \mathcal{A}$ et $\mathcal{A} = \{a_0, a_1, \ldots, a_{k-1}\}$. Alors*

i) $|w|_{a_0} = \sum_{i=0}^{k-1} |u|_{a_i} = \lg(u)$;

ii) $|w|_{a_0} = |u|_{a_0} + \sum_{i=1}^{k-1} |w|_{a_i}$.

Preuve La première égalité découle de la définition de ψ_{a_0}. Pour chaque lettre $\alpha \neq a_0$, $\psi_{a_0}(\alpha) = a_0\alpha$ et $\psi_{a_0}(a_0) = a_0$: ψ_{a_0} ajoute autant de a_0 que le nombre d'occurrences des autres lettres dans le mot. La deuxième égalité découle de la première, comme $|w|_{a_i} = |u|_{a_i}$ pour $i > 0$. ■

Lemme 2.4.9 *Soit $w \in \mathcal{A}^*$ un mot dans une classe épichristoffelle. Alors il existe $u \in \mathcal{A}^*$, $\lg(u) > 1$ et un morphisme épisturmien ψ_{a_0}, $a_0 \in \mathcal{A}$, tels que $w = \psi_{a_0}(u)$ si et seulement si $|w|_{a_0} > |w|_{a_i}$ pour tout $a_i \in \mathcal{A}$, $i \neq 0$.*

Preuve

($\Longrightarrow$) Par contradiction. Supposons que w est dans une classe épichristoffelle, $a_0 \in \mathcal{A}$, $w = \psi_{a_0}(u)$ et $|w|_{a_0}$ non maximum. Alors il existe au moins une lettre $a_i \in \mathcal{A}$ telle que $|w|_{a_i} \geq |w|_{a_0}$. Sans perte de généralité, supposons que $i = 1$. Par le Lemme 2.4.8, $|w|_{a_0} = \sum_{i=0}^{k-1} |u|_{a_i} = |u|_{a_0} + |w|_{a_1} + \sum_{i=2}^{k-1} |u|_{a_i}$ et cela implique que $|w|_{a_0} - |w|_{a_1} = |u|_{a_0} + \sum_{i=2}^{k-1} |u|_{a_i} \leq 0$, qui est impossible, comme $|w|_\alpha \geq 0$ pour tout mot w et toute lettre $\alpha \in \mathcal{A}$. Alors, si $|w|_{a_0}$ n'est

pas maximum, $|u|_{a_i} = 0$ pour tout $i \neq 1$ et alors $|w|_{a_1} = |w|_{a_0}$. Ainsi, on aurait $u = {a_1}^n$ et $w = \psi_{a_0}({a_1}^n)$. La seule possibilité est que $n = 1$, comme un mot dans une classe épichristoffelle est primitif. Alors $\lg(u) = 1$: contradiction. Ainsi, si $w = \psi_{a_0}(u)$, avec $\lg(u) > 1$, $|w|_{a_0}$ est maximum.

($\Longleftarrow$) Supposons maintenant que $|w|_{a_0} > |w|_{a_i}$ pour tout $a_i \in \mathcal{A}$, $i \neq 0$. Comme w est un mot épichristoffel, il existe un morphisme épisturmien ψ_{a_i} et un mot $u \in \mathcal{A}^*$ dans une classe épichristoffelle tels que $\psi_{a_i}(u) = w$. Supposons que $i \neq 0$. En utilisant le Lemme 2.4.8, $|w|_{a_i} = |w|_{a_0} + |u|_{a_i} + \sum_{1 \leq j \leq k-1, j \neq i} |w|_{a_j}$. Comme $|w|_{a_0} > |w|_{a_i}$, cela impliquerait que $|u|_{a_i} + \sum_{1 \leq j \leq k-1, j \neq i} |w|_{a_j} < 0$, ce qui est impossible. Donc $i = 0$.

■

Preuve (de la Proposition 2.4.7) Découle des Lemmes 2.4.8 et 2.4.9. ■

Exemple 2.4.10 Le 5-tuplet $p = (3, 4, 5, 6, 7)$ ne décrit pas les fréquences d'un mot appartenant à une classe épichristoffelle. En effet, $T(p) = T(3, 4, 5, 6, 7) = (3, 4, 5, 6, 7 - 18) = (3, 4, 5, 6, -11)$. Par contre, le 6-tuplet $q = (1, 1, 2, 4, 8, 16)$ décrit les fréquences des lettres d'un mot appartenant à une classe épichristoffelle :

$$\begin{aligned} T(1,1,2,4,8,16) &= (1,1,2,4,8,0) \\ T^2(q) &= T(1,1,2,4,8,0) = (1,1,2,4,0,0) \\ T^3(q) &= T(1,1,2,4,0,0) = (1,1,2,0,0,0) \\ T^4(q) &= T(1,1,2,0,0,0) = (1,1,0,0,0,0) \\ T^5(q) &= T(1,1,0,0,0,0) = (1,0,0,0,0,0). \end{aligned}$$

De cet algorithme découle une construction d'un mot appartenant à une classe épichristoffelle et ayant les fréquences de lettres p. Soit p le k-tuplet initial correspondant aux fréquences des lettres du mot appartenant à une classe épichristoffelle. Avec l'opérateur T décrit précédemment, on obtient une suite finie de k-tuplets $p^{(0)}$, $p^{(1)}$, $p^{(2)} \ldots$, $p^{(r-1)}, p^{(r)}$. Notons

$$p^{(s)} \xrightarrow{i} p^{(s+1)}$$

la relation $T(p^{(s)}) = p^{(s+1)}$, où $p_i^{(s)}$ est un entier maximal de $p^{(s)}$. On applique T jusqu'à ce que $p_i^{(r)} = 0$ pour tout i sauf pour une valeur i_{r-1} pour laquelle $p_{i_{r-1}}^{(r)} = 1$.

Si la suite de k-tuplets est

$$p^{(0)} \xrightarrow{i_0} p^{(1)} \xrightarrow{i_1} p^{(2)} \xrightarrow{i_2} \cdots \xrightarrow{i_{r-2}} p^{(r-1)} \xrightarrow{i_{r-1}} p^{(r)},$$

alors un mot appartenant à une classe épichristoffelle et ayant les fréquences p est

$$\psi_{a_{i_0}}(\psi_{a_{i_1}}(\ldots(\psi_{a_{i_{r-1}}}(\alpha))\ldots)),$$

où α est la lettre pour laquelle $p_{i_{r-1}}^{(r)} = 1$.

Exemple 2.4.11 Pour le triplet $(5, 10, 16)$ représentant les fréquences des lettres a, b et c respectivement, on obtient la suite

$$(5,10,16) \xrightarrow{c} (5,10,1) \xrightarrow{b} (5,4,1) \xrightarrow{a} (0,4,1) \xrightarrow{b} (0,3,1) \xrightarrow{b} (0,2,1) \xrightarrow{b} (0,1,1) \xrightarrow{b} (0,0,1).$$

En appliquant l'algorithme, on trouve le mot

$$\begin{aligned}
\psi_{cbabbbb}(c) &= \psi_{cbabbb}(bc) \\
&= \psi_{cbabb}(\psi_b(bc)) \\
&= \psi_{cbab}(\psi_b(bbc)) \\
&= \psi_{cba}(\psi_b(bbbc)) \\
&= \psi_{cb}(\psi_a(bbbbc)) \\
&= \psi_c(\psi_b(abababababac)) \\
&= \psi_c(babbabbabbabbabc) \\
&= cbcacbcbcacbcbcacbcbcacbcbcacbc.
\end{aligned}$$

Ce mot est dans une classe épichristoffelle et le mot épichristoffel de cette classe est le conjugué $acbcbcacbcbcacbcbcacbcbcacbc \cdot cbc$.

Remarquons que dans l'exemple précédent, on aurait pu décider de choisir la transition $(0,1,1) \xrightarrow{c} (0,1,0)$ et le mot obtenu aurait été un conjugué de $\psi_{cbabbbb}(c)$.

2.4.2 Propriétés des mots épichristoffels

Dans cette sous-section, nous montrons que certaines propriétés des mots de Christoffel se généralisent bien aux mots épichristoffels, mais que ce n'est toutefois pas toujours le cas.

Remarquons d'abord que l'équivalence du Théorème 2.2.14 ne se généralise pas aux mots épichristoffels. En effet, considérons la suite épisturmienne

$$s = aabaacaabaacaabaabaa \cdot caabaacaabaaa \cdots$$

Alors le mot $w = caabaacaabaaa$ est facteur de s, mais sa racine fractionnaire $z_w = w$ n'est pas dans une classe épichristoffelle, car la Proposition 2.4.7 nous assure qu'il n'existe aucun mot épichristoffel ayant les fréquences $(2, 2, 9)$. En effet, $T(2, 2, 9) = (2, 2, 5)$, $T^2(2, 2, 9) = T(2, 2, 5) = (2, 2, 1)$, $T^3(2, 2, 9) = T(2, 2, 1) = (2, -1, 1)$.

Par contre, l'implication suivante est vraie pour les mots épichristoffels.

Théorème 2.4.12 *Soit w un mot non vide tel que sa racine fractionnaire est dans une classe épichristoffelle. Alors w est facteur d'une suite épisturmienne.*

Preuve Posons $w = z_w^k$, avec $k \geq 1 \in \mathbb{Q}$, la racine fractionnaire de w. Supposons que z_w est dans une classe épichristoffelle. Il existe donc $x \in \mathcal{A}^*$ et $a \in \mathcal{A}$ tels que $\psi_x(a) = z_w$. Alors, w est facteur de $z_w^{\lceil k \rceil} = \psi_x^{\lceil k \rceil}(a) = \psi_x(a^{\lceil k \rceil})$. Il suffit donc de prendre une suite épisturmienne ayant $a^{\lceil k \rceil}$ comme facteur et d'appliquer le morphisme ψ_x. On trouve alors que $\psi_x(a^{\lceil k \rceil})$ est facteur d'une suite épisturmienne, et donc que w aussi. ■

Proposition 2.4.13 *Soit $w \in \mathcal{A}^*$ un mot épichristoffel. Alors l'ensemble des facteurs de longueur $\leq \lg(w)$ de la classe épichristoffelle $[w]$ est fermé sous l'image miroir.*

Preuve Remarquons d'abord que l'ensemble des facteurs de longueur $\leq \lg(w)$ d'une classe épichristoffelle est le même que celui de w^2. Comme tout mot w épichristoffel est produit de deux palindromes (Proposition 2.4.5), posons $w = p_1p_2$, avec p_1, p_2 palindromes. Alors $w^2 = p_1p_2p_1p_2$ et donc $\widetilde{w} = \widetilde{p_1p_2} = p_2p_1$ est facteur de w^2. Ainsi, l'image miroir de tout

facteur de w est aussi un facteur de w^2 et par conséquent, est dans la classe épichristoffelle de w. ■

Remarque 2.4.14 La fermeture palindromique à droite d'un mot w dans une classe épichristoffelle est souvent préfixe de w^2, mais ce n'est pas toujours le cas. Il suffit de prendre le mot $w = abcbab$ pour lequel $w^{(+)} = abcbab \cdot cba$.

Lemme 2.4.15 *Un mot épichristoffel ne s'écrit pas toujours comme produit de deux mots épichristoffels.*

Preuve Il suffit de considérer le mot épichristoffel $aabacab$. Les seules décompositions en mots de classe épichristoffelle sont $a \cdot abacab$ et $aab \cdot acab$, mais $abacab$ et $acab$ ne sont pas les plus petits mots de leur classe de conjugaison respective. ■

Lemme 2.4.16 *Tout mot épichristoffel s'écrit de façon non unique comme un produit de deux mots appartenant à des classes épichristoffelles.*

Preuve Pour la non unicité, il suffit de considérer l'exemple du mot $aabacab$ donné dans la preuve du Lemme 2.4.15. Pour l'existence, comme tout mot faisant partie d'une classe épichristoffelle s'écrit comme $\psi_w(a)$, avec $a \in \mathcal{A}$, $w \in \mathcal{A}^n$ et $w[n-1] \neq a$, il suffit de prendre les mots $\psi_{w[0]w[1]\cdots w[n-2]}(w[n-1])$ et $\psi_{w[0]w[1]\cdots w[n-2]}(a)$, puisque

$$\begin{aligned}\psi_w(a) &= \psi_{w[0]w[1]\cdots w[n-1]}(a)\\ &= \psi_{w[0]w[1]\cdots w[n-2]}(w[n-1]a)\\ &= \psi_{w[0]w[1]\cdots w[n-2]}(w[n-1]) \cdot \psi_{w[0]w[1]\cdots w[n-2]}(a).\end{aligned}$$

■

Lemme 2.4.17 *(Justin, 2005) Pour tout $w \in \mathcal{A}^*$, $x \in \mathcal{A}^*$,*

$$\text{Pal}(xw) = \psi_x(\text{Pal}(w))\text{Pal}(x).$$

Proposition 2.4.18 *Soit w un mot appartenant à une classe épichristoffelle. Soit $\psi_{x_0}, \psi_{x_1}, \ldots, \psi_{x_n}$, avec $x_i \in \mathcal{A}$, une suite de morphismes à appliquer sur une lettre $a \in \mathcal{A}$ fixée pour obtenir w. Alors*

$$w = \psi_{x_n x_{n-1} \cdots x_0}(a) \in \text{Pref}(\text{Pal}(x_n x_{n-1} \cdots x_0 a)).$$

Preuve Procédons par récurrence. Vérifions pour $n = 0$. Il y a 2 cas à considérer. Soit $w = \psi_a(a) = a$, soit $w = \psi_b(a) = ba$, avec $b \neq a$. On a respectivement $\text{Pal}(aa) = aa$ et $\text{Pal}(ba) = bab$. Dans les 2 cas, $w \in \text{Pref}(\text{Pal}(x_0 a))$. Supposons maintenant que $w = \psi_{x_n \cdots x_0}(a)$ est préfixe de $\text{Pal}(x_n \cdots x_0 a)$, pour tout $k \leq n$. Vérifions pour $n+1$. D'une part, on a $\psi_{x_{n+1} x_n \cdots x_0}(a) = \psi_{x_{n+1}}(w)$. D'autre part, par le Lemme 2.4.17, on a $\text{Pal}(x_{n+1} x_n \cdots x_0 a) = \psi_{x_{n+1}}(\text{Pal}(x_n \cdots x_0 a)) x_{n+1}$. Par hypothèse de récurrence, $\text{Pal}(x_n \cdots x_0 a)$ s'écrit comme ws, pour un $s \in \mathcal{A}^*$. Donc, $\text{Pal}(x_{n+1} x_n \cdots x_0 a) = \psi_{x_{n+1}}(ws) x_{n+1} = \psi_{x_{n+1}}(w) \psi_{x_{n+1}}(s) x_{n+1}$. D'où la conclusion. ∎

Exemple 2.4.19 Soit $w = \psi_{abaa}(a) = aba$. Alors $\text{Pal}(abaaa) = (aba)^3$ et $w \in \text{Pref}(\text{Pal}(abaaa))$.

2.4.3 Critère pour être dans une classe épichristoffelle

Le but de cette section est de prouver le théorème suivant, qui est une généralisation du Théorème 2.2.11.

Théorème 2.4.20 *Soit w un mot fini tel que tous ses conjugués sont facteurs d'une même suite épisturmienne z. Alors w est dans une classe épichristoffelle.*

Remarquons que pour les suites épisturmiennes, une condition supplémentaire est nécessaire : les conjugués doivent être facteurs de la même suite épisturmienne. Par exemple, tous les conjugués du mot abc sont facteurs de suites épisturmiennes, mais abc n'est pas un mot épichristoffel.

Avant de prouver ce théorème, quelques résultats sont préalables.

Lemme 2.4.21 *Soient $u, v \in \mathcal{A}^*$ et un morphisme ψ. Si u et v sont conjugués, alors $\psi(u)$ et $\psi(v)$ le sont aussi.*

Proposition 2.4.22 *Soient u et v des mots finis appartenant à des classes épichristoffelles. Si $|u|_\alpha = |v|_\alpha$ pour tout $\alpha \in \mathcal{A}$, alors u et v sont dans la même classe épichristoffelle. Autrement dit, un k-tuplet de fréquences de lettres détermine une unique classe épichristoffelle.*

Preuve Soit $\alpha_0 \in \mathcal{A}$ tel que $|u|_{\alpha_0}$ est maximum. Alors, en utilisant le Lemme 2.4.9, il existe $s_0, t_0 \in \mathcal{A}^*$ dans des classes épichristoffelles tels que $u = \psi_{\alpha_0}(s_0)$ et $v = \psi_{\alpha_0}(t_0)$. De plus, pour toute lettre $\beta \in \mathcal{A}$, si $|\psi_{\alpha_0}(s_0)|_\beta = |\psi_{\alpha_0}(t_0)|_\beta$, alors $|s_0|_\beta = |t_0|_\beta$, comme le nombre de lettres d'un mot obtenu par un morphisme sur un mot w dépend seulement du nombre de chaque lettre de w et non pas de leur position (voir Lemme 2.4.8). Par le même argument, on trouve une suite de morphismes $\psi_{\alpha_1}, \psi_{\alpha_2}, \ldots, \psi_{\alpha_n}$ et deux suites de mots $s_0, s_1, \ldots, s_{n-1}, s_n$ et $t_0, t_1, \ldots, t_{n-1}, t_n$ telles que $s_i = \psi_{\alpha_{i+1}}(s_{i+1})$, $t_i = \psi_{\alpha_{i+1}}(t_{i+1})$, $|s_i|_\beta = |t_i|_\beta$ pour toute lettre $\beta \in \mathcal{A}$ et pour tout $0 \leq i < n$ et $s_n = a_1 a_2$ et $t_n = a_2 a_1$, $a_1, a_2 \in \mathcal{A}$, $a_1 \neq a_2$. On conclut en utilisant le Lemme 2.4.21 n fois : $u = \psi_{\alpha_0\alpha_1\cdots\alpha_n}(a_1 a_2)$ est conjugué à $v = \psi_{\alpha_0\alpha_1\cdots\alpha_n}(a_2 a_1)$. ■

Lemme 2.4.23 *La longueur des blocs d'un mot fini w est exactement la même que dans sa fermeture palindromique $w^{(+)}$.*

Preuve Soit $w = w_1 w_2$, avec w_2 le plus long suffixe palindromique de w. Alors, par définition, $w^{(+)} = w_1 w_2 \widetilde{w_1}$. La seule façon de construire un bloc d'une longueur différente de ceux de w est lors de la concaténation de w_2 avec $\widetilde{w_1}$. Comme w_2 est un palindrome, on peut écrire $w_2 = a$ ou $w_2 = aua$, avec $a \in \mathcal{A}$ et $u \in \mathcal{A}^*$ palindrome. Il y a deux cas à considérer :

i) $w_1 = va$, $v \in \mathcal{A}^*$: $w^{(+)}$ s'écrit donc comme $vaw_2a\widetilde{v}$. Le bloc de la lettre a créé par la concaténation de w_2 et $\widetilde{w_1}$ a la même longueur que celui de la concaténation de w_1 et w_2, comme w_2 est un palindrome. Aucun bloc d'une nouvelle longueur n'est créé. Notons que cet argument fonctionne comme $w_2 \notin a^+$. En effet, si $w_2 \in a^+$ et $w_1 = va$, alors w_2 n'est pas le plus long suffixe palindromique de w.

ii) $w_1 = vb$, $a \neq b$ et $v \in \mathcal{A}^*$: la concaténation de w_2 avec $\widetilde{w_1}$ ne crée pas de bloc d'une nouvelle longueur, comme la dernière lettre de w_2 est différente de la première de w_1.

■

Exemple 2.4.24 Dans le mot $w = abacabaab$, les blocs de a sont de longueur 1 et 2, les blocs des lettres b et c sont de longueur 1, tout comme pour sa fermeture palindromique

$$w^{(+)} = abacabaab \cdot acaba.$$

Lemme 2.4.25 *Dans une suite épisturmienne standard, les blocs des lettres sont de longueur* 1, *sauf exactement ceux d'une lettre, dont les longueurs sont* k *et* $(k+1)$, *pour* $k \geq 1$ *fixé.*

Preuve Soit $\Delta(s) = a^k z$ la suite directrice d'une suite épisturmienne s, avec k maximal, $a \in \mathcal{A}$ et $z \in \mathcal{A}^\omega$. Alors la suite s a le préfixe Pal$(a^k) = a^k$. En utilisant le Lemme 2.3.5, il en découle que pour toute première occurrence d'une lettre $b \neq a \in \mathcal{A}$ dans la suite directrice, Pal$(a^k vb) =$ Pal$(a^k v) \cdot b \cdot$ Pal$(a^k v)$, avec $v \in \mathcal{A}^*$, qui se réécrit comme $a^k t a^k \cdot b \cdot a^k t a^k$ (resp. $a^k b a^k$), en supposant que Pal$(a^k v) = a^k t a^k$ (resp. $v = \varepsilon$). Ainsi, la longueur des blocs de la lettre b est 1, comme le b est précédé et suivi de la lettre a.

Si ce n'est pas la première occurrence de la lettre b dans la suite directrice, alors par définition, Pal$(a^k vb) =$ (Pal$(a^k v) \cdot b)^{(+)}$. Comme Pal$(a^k v) \cdot b$ a la lettre a comme avant-dernière lettre et comme la fermeture palindromique ne change pas la longueur des blocs (voir la Proposition 2.4.23), l'ajout de la lettre b à la suite directrice ne crée pas une nouvelle longueur de bloc. Si la lettre suivante de la suite directrice était un a, alors on aurait Pal$(a^k va) =$ (Pal$(a^k v) \cdot a)^{(+)}$. La concaténation de Pal$(a^k v)$ avec a crée un bloc de a de longueur $(k+1)$, puisque Pal$(a^k v)$ a le suffixe a^k. Comme la fermeture palindromique préserve la longueur des blocs, les longueurs des blocs de la lettre a sont k et $(k+1)$. Remarquons qu'il est impossible d'avoir des blocs de la lettre a de longueur $(k+2)$ puisque les blocs de longueur $(k+1)$ ne sont jamais suffixes d'un préfixe palindromique. ■

Remarque 2.4.26 Une suite épisturmienne peut avoir des blocs d'une seule longueur pour toutes ses lettres. C'est le cas si la suite directrice s'écrit comme $a^k z$, avec $z \in \{\mathcal{A} - \{a\}\}^*$. Alors, ces suites épisturmiennes ont des blocs de longueurs k pour la lettre a et de longueur 1 pour toutes les autres lettres.

Remarque 2.4.27 Pour une suite épisturmienne standard $w = \psi_a(t)$, toutes les lettres différentes de a sont précédées et suivies de la lettre a. Si s est une suite épisturmienne non standard, cette remarque est aussi vraie, sauf possiblement pour la première lettre de la suite, si elle est différente de a.

Lemme 2.4.28 *Dans une suite épisturmienne non standard, les blocs des lettres sont de longueur* 1*, sauf exactement ceux d'une lettre* α*, dont les longueurs sont* ℓ*,* k*,* $(k+1)$*, pour* $k \geq 1$ *fixé, et où* $\ell < k$ *est la longueur du bloc de* α *préfixe du mot, s'il y a lieu.*

Preuve Comme pour toute suite épisturmienne, il existe une suite épisturmienne standard ayant exactement le même ensemble de facteurs, il n'y a que le préfixe du mot qui peut créer un bloc de lettres de longueur plus petite, d'où la longueur ℓ. ∎

Exemple 2.4.29 Soit la suite épisturmienne standard $s = 1121131121141121131121112113\cdots$ et la suite épisturmienne non standard $t = 121131121141121131121112113\cdots$. Ces 2 suites ont le même langage : $F(s) = F(t)$. La suite s a des blocs de 1 de longueur 2 et 3, mais la suite t a des blocs de 1 de longueur $1, 2$ et 3.

Lemme 2.4.30 *Soit une suite épisturmienne standard* $z = \psi_{a_0}(t)$ *et* $w = a_0 y a_1$ *un facteur de* z*, avec* $a_0, a_1 \in \mathcal{A}$ *et* $y \in \mathcal{A}^*$*. Alors, il existe un facteur* u *de* t *tel que* $\psi_{a_0}(u) = w$*.*

Preuve Si $z = \psi_{a_0}(t)$, $t = t[0]t[1]t[2]\cdots$ et Card$(\mathcal{A}) = k$, alors par définition de ψ, $z = \psi_{a_0}(t[0])\psi_{a_0}(t[1])\cdots \in \{a_0, a_0a_1, a_0a_2, \ldots, a_0a_{k-1}\}^*$. Comme tout facteur $w = a_0 y a_1$ de $z = \psi_{a_0}(t)$ s'écrit comme un $w \in \{a_0, a_0a_1, a_0a_2, \ldots, a_0a_{k-1}\}^*$, on peut construire un mot u en faisant correspondre à a_0a_i la lettre a_i pour $i \neq 0$ et à a_0 la lettre a_0. Ainsi, w est l'image du mot u par le morphisme ψ_{a_0}. ∎

Proposition 2.4.31 *Soit* $z = \psi_a(t)$*, où* t *et* z *sont des suites épisturmiennes standards. Soit* w *un facteur de* z *non puissance d'une lettre tel que* $\lg(w) > 1$ *et tous ses conjugués sont aussi facteurs de* z*. Alors, il existe* u *facteur de* t *tel que* $w = \psi_a(u)$ *ou* $w = \overline{\psi}_a(u)$*.*

Preuve Puisque $z = \psi_a(t)$, les blocs de toute lettre $\alpha \in \{\mathcal{A} - \{a\}\}$ sont de longueur 1. Posons k et $(k+1)$ les longueurs des blocs de a. Soit $\beta, \gamma \in \mathcal{A}$, avec $\beta, \gamma \neq a$ et $y \in \mathcal{A}^*$. Il y a 4 cas à considérer.

i) $w = \beta y \gamma$: si w est facteur de z, alors son conjugué $y\gamma\beta$ doit aussi être facteur de z. C'est impossible, puisque toute occurrence de la lettre β doit être précédé de la lettre a. On exclut donc ce cas.

ii) $w = ay\beta$: si w est facteur de $z = \psi_a(t)$, alors par le Lemme 2.4.30, il existe u facteur de t tel que $\psi_a(u) = w$.

iii) $w = \beta ya$: de façon symétrique au cas ii), si $w = \beta ya$ est facteur de $z = \psi_a(t)$, alors il existe u facteur de t tel que $\overline{\psi_a}(u) = w$.

iv) $w = aya$: réécrivons $w = a^m y' a^n$, où m, $n \geq 1$ et m et n sont maximums. Le facteur y' est non vide, puisque w est supposé non puissance d'une lettre. Supposons que wa n'est pas facteur de z. On a donc que $w\beta = a^m y' a^n \beta$ est facteur de z, pour $\beta \in \mathcal{A}$, $\beta \neq a$ fixé. Comme les blocs de a sont de longueur k et $(k+1)$ seulement, on a que $n = k$ ou $n = k+1$. D'autre part, comme $w = a^m y' a^n$ est facteur de z, son conjugué $y' a^{m+n}$ est aussi facteur de z. Donc $m + n \leq k + 1$. Mais comme $m \neq 0$, la seule possibilité est que $n = k$ et $m = 1$. Ainsi, $w = ay'a^k$. Son conjugué $y'a^{k+1}$ est aussi facteur de z et comme y' ne commence pas par a, il doit être précédé de a : $ay'a^{k+1} = ay'a^k a = wa$ est facteur de z. Donc wa est facteur de z et on conclut en utilisant le Lemme 2.4.30.

■

Corollaire 2.4.32 *Soit $z = \psi_a(t)$, où t et z sont des suites épisturmiennes standards. Soit w un facteur de z non puissance d'une lettre tel que* $\lg(w) > 1$ *et tous ses conjugués sont aussi facteurs de z. Alors, pour tout conjugué w' de w, il existe u' facteur de t tel que $w' = \psi_a(u')$ ou $w' = \overline{\psi_a}(u')$.*

Preuve Il suffit d'utiliser la Proposition 2.4.31 pour tous les conjugués de w. ■

Proposition 2.4.33 *Soient w et w', deux mots finis conjugués tels que $w = \phi(u)$ et $w' = \phi'(u')$, où $\phi, \phi' \in \{\psi_a, \overline{\psi}_a\}$, pour $u, u' \in \mathcal{A}^*$. Alors u et u' sont aussi conjugués.*

Preuve Sans perte de généralité, on peut supposer que $\phi = \phi' = \psi_a$, comme $\psi_a(w) = a\overline{\psi}_a(w)a^{-1}$ et donc, $\psi_a(w)$ est conjugué à $\overline{\psi}_a(w)$ pour tout mot w. Ainsi, on peut écrire $w = a^{n_0}v[0]a^{n_1}v[1]\cdots a^{n_k}v[k]$, où $v[i] \neq a$, $n_0 > 0$ et $n_i > 0$. Comme $w = \psi_a(u)$, on a $u = a^{n_0-1}v[0]a^{n_1-1}v[1]\cdots a^{n_k-1}v[k]$. Comme w et w' sont conjugués, on peut écrire $w' = a^{\alpha}v[i]a^{n_{i+1}}v[i+1]\cdots a^{n_{i-1}}v[i-1]a^{\beta}$, où $\alpha+\beta = n_i$ et $\alpha \geq 1$. De la même façon, $u' = a^{\alpha-1}v[i]a^{n_{i+1}-1}v[i+1]\cdots a^{n_{i-1}-1}v[i-1]a^{\beta}$, qui est conjugué à $u'' = a^{\alpha+\beta-1}v[i]a^{n_{i+1}-1}v[i+1]\cdots a^{n_{i-1}-1}v[i-1]$. En comparant u et u'', on conclut que u est conjugué à u'. ∎

Avant de faire la preuve du Théorème 2.4.20, prouvons le théorème suivant qui en est un cas particulier.

Théorème 2.4.34 *Soit w un mot fini primitif tel que tous ses conjugués sont facteurs d'une même suite épisturmienne standard z. Alors, w est dans une classe épichristoffelle.*

Preuve Par récurrence sur le nombre de morphismes. On utilise le Corollaire 2.4.32 et la Proposition 2.4.33 : on trouve une suite de morphismes épisturmiens $\{\phi_0, \phi_1, .., \phi_k\} \in \{\psi_a, \overline{\psi}_a\}^{k+1}$ et une suite de mots $\{w, w_1, w_2, ..., w_k\}$ tel que $\lg(w) \geq \lg(w_1) \geq \lg(w_2) \geq ... \geq \lg(w_k) = 1$, $w = \phi_0(\phi_1(...(\phi_k(w_k))...))$ et $w_i = \phi_i(\ldots(\phi_k(w_k)))$. Ainsi, w est bien l'image d'une lettre par morphismes épisturmiens, et donc, w est dans une classe épichristoffelle. Remarquons que l'utilisation du Corollaire 2.4.32 nécessite l'hypothèse que les conjugués soient facteurs de la même suite épisturmienne. ∎

Rappelons que la condition que tous les conjugués soient facteurs de la même suite épisturmienne est nécessaire. Par exemple, il suffit de prendre le mot abc pour lequel tous les conjugués sont facteurs de suites épisturmiennes, mais qui n'est pas un mot épichristoffel.

Preuve (Preuve du Théorème 2.4.20) Supposons que w est un mot fini tel que tous ses conjugués sont facteurs d'une même suite épisturmienne s. Si s n'est pas standard, par la Définition 2.3.2, on sait qu'il existe une suite épisturmienne standard s' telle que $F(s) = F(s')$. On peut donc considérer cette suite s' et on conclut en utilisant le Théorème 2.4.34. ∎

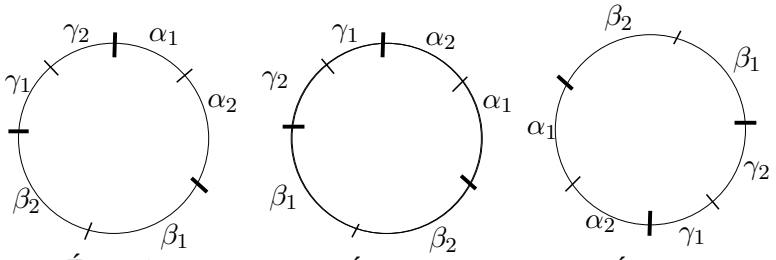
Théorème 2.4.35 *Soit w un mot fini primitif. Alors w est tel que tous ses conjugués sont facteurs d'une même suite épisturmienne z si et seulement si w est dans une classe épichristoffelle.*

Preuve Le sens direct découle du Théorème 2.4.20 et la réciproque est obtenue par définition de suites épisturmiennes et de mots épichristoffels. ■

2.5 Problèmes ouverts

Inspirée d'une construction décrite dans (Arnoux et Rauzy, 1991), nous avons trouvé une façon analogue aux graphes de Cayley permettant de construire géométriquement un mot épichristoffel à partir des fréquences des lettres pour une sous-classe des mots épichristoffels sur un alphabet à trois lettres. Cette construction est la version discrète de la construction suivante sur un cercle de longueur $\alpha + \beta + \gamma$.

1. Diviser le cercle en trois intervalles ayant des longueurs respectives de α, β et γ.
2. Couper chacun de ces intervalles en deux parties égales, puis permuter ces deux parties sur chacun des intervalles en préservant l'orientation. Les intervalles $\alpha_1, \alpha_2, \beta_1, \beta_2, \gamma_1$ et γ_2 sont obtenus.
3. Faire une rotation d'un demi-tour.



Exemple 2.5.1 Soient les fréquences $(2, 3, 5)$. Comme $2 + 3 + 5 = 10$, on prend l'intervalle entier $[1, 10]$, que l'on subdivise en trois sous-intervalles, de longueurs 2, 3 et 5. Pour chacun des sous-intervalles, on le coupe en deux et on échange les deux moitiés. Si la longueur est impaire, la valeur du milieu reste inchangée. On obtient ainsi :

1	2	3	4	5	6	7	8	9	10
2	1	5	4	3	9	10	8	6	7

Ensuite, on coupe le nouvel intervalle en deux et on échange les deux moitiés. On obtient :

$$\begin{array}{cccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 2 & 1 & 5 & 4 & 3 & 9 & 10 & 8 & 6 & 7 \\ 9 & 10 & 8 & 6 & 7 & 2 & 1 & 5 & 4 & 3 \end{array}$$

En ne considérant que la première et la troisième lignes, on obtient la permutation en un seul cycle $(1, 9, 4, 6, 2, 10, 3, 8, 5, 7)$. En associant aux valeurs 1 et 2 la lettre a, aux valeurs $3, 4, 5$ la lettre b et aux valeurs 6 à 10 la lettre c, on obtient le mot $acbcacbcbc$ qui a les fréquences désirées et qui s'obtient par morphisme épisturmien. En effet, $acbcacbcbc = \overline{\psi}_{cb}\psi_{aa}(b)$.

Le problème est le suivant : pour quelle sous-classe des mots épichristoffels cette construction permet-elle d'obtenir un conjugué d'un mot épichristoffel ?

Nous avons effectué cette construction pour plusieurs triplets de fréquences. Pour les triplets $(1, 1, 2), (1, 2, 4), (2, 4, 7), (1, 1, 4), (1, 1, 6), (1, 4, 6), (1, 2, 3), (1, 2, 6), (2, 3, 6), (1, 3, 6),$ $(2, 3, 5), (2, 5, 7), (1, 4, 5), (1, 6, 7)$, cette construction fonctionne très bien. Par contre, pour certains autres triplets, on rencontre un problème : soit le mot obtenu n'est pas dans une classe épichristoffelle, comme c'est la cas pour le triplet $(1, 2, 7)$, soit la permutation consiste en plus d'un cycle, comme c'est le cas pour les triplets $(1, 4, 7), (1, 3, 4), (3, 4, 7), (1, 5, 6)$.

Il serait bien de caractériser cette sous-classe des mots épichristoffels. Peut-être est-elle d'un intérêt sous estimé. D'autre part, comme les mots épichristoffels sont définis sur des alphabets à 3 lettres et plus, il serait pertinent de trouver une construction analogue pour un mot épichristoffel sur k lettres, pour $k \geq 3$.

Un autre problème intéressant relié aux mots épichristoffels serait de caractériser les permutations obtenues par la bijection de Gessel-Reutenauer (Gessel et Reutenauer, 1993) sur les mots épichristoffels. Rappelons que pour un mot fini w de longueur n, la bijection de Gessel-Reutenauer associe à un mot de longueur n une permutation P des éléments de 1 à n de la façon suivante :

$$\begin{aligned} P \quad &: \quad \mathcal{A}^* \to S_n \\ &\quad\; w \mapsto \sigma, \end{aligned}$$

où $\sigma^i(1) = j$ si le conjugué de w commençant à la $(i+1)$-ième lettre est le j-ième plus petit, pour $1 \leq i \leq n-1$ et $\sigma^n(1) = 1$. Comme un mot épichristoffel est primitif, il sera toujours possible d'ordonner les conjugués d'un mot épichristoffel.

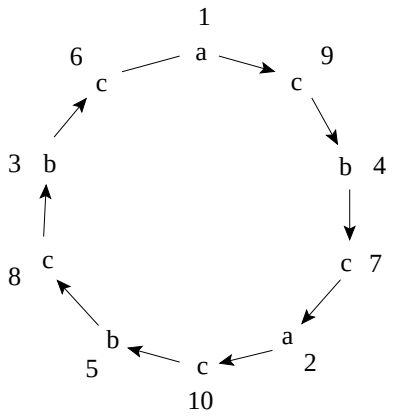


Figure 2.3 Illustration de l'Exemple 2.5.2

Exemple 2.5.2 Soit le mot épichristoffel $w = acbcacbcbc$. En ordonnant les conjugués de w selon l'ordre lexicographique, on obtient

$$\begin{aligned}
w &= acbcacbcbc \\
\gamma^4(w) &= acbcbcacbc \\
\gamma^8(w) &= bcacbcacbc \\
\gamma^2(w) &= bcacbcbcac \\
\gamma^6(w) &= bcbcacbcac \\
\gamma^9(w) &= cacbcacbcb \\
\gamma^3(w) &= cacbcbcacb \\
\gamma^7(w) &= cbcacbcacb \\
\gamma^1(w) &= cbcacbcbca \\
\gamma^5(w) &= cbcbcacbca.
\end{aligned}$$

La permutation obtenue est donc $(1, 9, 4, 7, 2, 10, 5, 8, 3, 6)$. Cette permutation se réécrit sous la

forme

$$\begin{array}{cccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 9 & 10 & 6 & 7 & 8 & 1 & 2 & 3 & 4 & 5 \end{array}$$

Dans le cas de mot de Christoffel, la bijection de Gessel-Reutenauer donne la même permutation que celle qui engendre le graphe de Cayley. Dans le cas de mots épichristoffels, il serait intéressant de trouver à quoi correspondent les permutations ainsi obtenues. Notons que la régularité observée dans l'exemple précédent, c'est-à-dire le fait qu'on retrouve 3 intervalles d'entiers ayant des longueurs respectivement 2, 3 et 5, n'est pas présente pour tout mot épichristoffel, comme l'illustre l'exemple qui suit.

Exemple 2.5.3 Soit le mot épichristoffel $w = aabacababacab$. La permutation obtenue par la bijection Gessel-Reutenauer est

$$\begin{array}{cccccccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 \\ 5 & 8 & 9 & 10 & 11 & 12 & 13 & 1 & 4 & 6 & 7 & 2 & 3 \end{array}$$

Remarque 2.5.4 Les mots épichristoffels ne sont généralement pas équilibrés, au sens traditionnel d'équilibre. En effet, il suffit de considérer le mot $w = aabacababacab$ de l'exemple précédent et les facteurs bab et aca.

Le fait que les mots épichristoffels ne soient pas tous équilibrés est cohérent avec les résultats du chapitre suivant. En effet, à la Section 2.5, nous verrons que les suites épisturmiennes ne sont pas équilibrées en général. Il pourrait donc être fort intéressant d'éventuellement caractériser les mots épichristoffels équilibrés.

Chapitre III

SUITES ÉPISTURMIENNES ET CONJECTURE DE FRAENKEL

Les suites sturmiennes sont bien connues comme étant les suites non périodiques équilibrées sur un alphabet à deux lettres. Elles sont aussi caractérisées par leur complexité : elles ont exactement $(n+1)$ facteurs de longueur n. Une généralisation naturelle des suites sturmiennes est l'ensemble des suites épisturmiennes. Ces suites ne sont pas nécessairement équilibrées sur un alphabet à k lettres, ni nécessairement apériodiques. Dans ce chapitre, nous caractérisons les suites épisturmiennes équilibrées, périodiques ou non, et prouvons la conjecture de Fraenkel pour la classe des suites épisturmiennes. Nous montrons que les suites épisturmiennes équilibrées sont toutes ultimement périodiques et sont divisées en 3 classes.

Dans un premier temps, nous rappelons la définition d'une suite de Beatty. Ensuite, nous introduisons la conjecture de Fraenkel telle qu'initialement introduite, en terme de suites de Beatty, puis en terme de mots infinis. Puis, après avoir caractérisé les suites épisturmiennes équilibrées, nous utilisons ce résultat afin de prouver la conjecture de Fraenkel pour les suites épisturmiennes.

3.1 Suites de Beatty

Comme mentionné au début du Chapitre II, les suites de Beatty correspondent aux suites sturmiennes étudiées du point de vue de la théorie des nombres. La forme des suites de Beatty apparaît dans la littérature pour la première fois dans (Beatty, 1926) et ce n'est qu'une trentaine d'années plus tard que le nom apparaît (Connell, 1959; Connell, 1960).

Définition 3.1.1 Une *suite de Beatty* est une suite de la forme

$$S(\alpha, \beta) = \{\lfloor \alpha n + \beta \rfloor : n \in \mathbb{Z}\},$$

avec $\alpha, \beta \in \mathbb{R}$.

Définition 3.1.2 Une *suite de Beatty rationnelle* est une suite de Beatty $S(\alpha, \beta)$ telle que $\alpha \in \mathbb{Q} - \{0\}$. On la note $S(p/q, \beta)$, avec $p, q \in \mathbb{Z}$ et $q \neq 0$.

Exemple 3.1.3 Pour $p = 2, q = 3$ et $\beta = 0$, on obtient la suite de Beatty

$$S\left(\frac{2}{3}, 0\right) = \left\{\left\lfloor \frac{2n}{3} \right\rfloor : n \in \mathbb{Z}\right\}.$$

Le tableau suivant donne quelques unes des valeurs de la suite.

n	1	2	3	4	5	6	7	8
$\left\lfloor \frac{2n}{3} \right\rfloor$	0	1	2	2	3	4	4	5

Ainsi, $S\left(\frac{2}{3}, 0\right) = \ldots, 0, 1, 2, 2, 3, 4, 4, 5, \ldots$

Proposition 3.1.4 *Soit* $S\left(\frac{p}{q}, \beta\right) = \left\{\left\lfloor \frac{pn}{q} + \beta \right\rfloor : n \in \mathbb{Z}\right\}$, *une suite de Beatty rationnelle.*

i) *Si* $p = q$, *alors la suite correspond exactement à* $\mathbb{Z}$.

ii) *Si* $p < q$, *alors la suite correspond à* $\mathbb{Z}$ *et comporte des répétitions.*

iii) *Si* $p > q$, *alors la suite est de la forme*

$$\left\{\lfloor \beta \rfloor, \left\lfloor \frac{p}{q} + \beta \right\rfloor, \left\lfloor \frac{2p}{q} + \beta \right\rfloor, \ldots, \left\lfloor \frac{(q-1)p}{q} + \beta \right\rfloor\right\} + p\mathbb{Z}.$$

Preuve Découle de la définition de $S(\alpha, \beta)$. ∎

Les suites de Beatty rationnelles qui nous intéressent sont celles pour lesquelles $\alpha > 1$, c'est-à-dire $p > q$.

Exemple 3.1.5 Pour $p = 5, q = 2$ et $\beta = -2$, on obtient la suite de Beatty

$$S\left(\frac{5}{2}, -2\right) = \left\{\left\lfloor \frac{5n}{2} - 2 \right\rfloor : n \in \mathbb{Z}\right\}$$

et on a

n	1	2	3	4	5	6	7	8
$\left\lfloor \frac{5n}{2} - 2 \right\rfloor$	0	3	5	8	10	13	15	18

Ainsi, $S\left(\frac{5}{2}, -2\right) = \ldots, 0, 3, 5, 8, 10, 13, 15, 18, \ldots = \{0, 3\} + 5\mathbb{Z}$.

Définition 3.1.6 Un *système couvrant de suites de Beatty disjointes* est un ensemble de k suites de Beatty $\left\{ S\left(\frac{p_i}{q_i}, b_i\right) : i = 1, \ldots, k \right\}$ tel que

i) $\bigcup_{1 \leq i < j \leq k} \left(S\left(\frac{p_i}{q_i}, b_i\right) \bigcap S\left(\frac{p_j}{q_j}, b_j\right) \right) = \emptyset$;

ii) $\bigcup_{i=1}^{k} S\left(\frac{p_i}{q_i}, b_i\right) = \mathbb{Z}$.

Définition 3.1.7 Un *système couvrant éventuel de suites de Beatty disjointes* est un ensemble de k suites de Beatty $\left\{ S\left(\frac{p_i}{q_i}, b_i\right) : i = 1, \ldots, k \right\}$ tel que

i) $\bigcup_{1 \leq i < j \leq k} \left(S\left(\frac{p_i}{q_i}, b_i\right) \bigcap S\left(\frac{p_j}{q_j}, b_j\right) \right) = \emptyset$;

ii) $\{i\}_{i \geq \ell} \subseteq \bigcup_{i=1}^{k} S\left(\frac{p_i}{q_i}, b_i\right)$, pour un $\ell \in \mathbb{Z}$ fixé.

Autrement dit, l'ensemble doit couvrir les entiers à partir d'une certaine valeur ℓ.

Exemple 3.1.8 Soient les suites de Beatty $S_1\left(\frac{7}{4}, 0\right)$, $S_2\left(\frac{7}{2}, 6\right)$ et $S_3\left(\frac{7}{1}, 4\right)$. On trouve que $S_1 = \{0, 1, 3, 5\} + 7\mathbb{Z}$, $S_2 = \{2, 6\} + 7\mathbb{Z}$ et $S_3 = \{4\} + 7\mathbb{Z}$. L'union de ces suites forme un système couvrant de suites de Beatty disjointes, puisque

i) $S_1 \cap S_2 = S_1 \cap S_3 = S_2 \cap S_3 = \emptyset$;

ii) $S_1 \cup S_2 \cup S_3 = \{0, 1, 2, 3, 4, 5, 6\} + 7\mathbb{Z} = \mathbb{Z}$.

3.2 Conjecture de Fraenkel

Il est naturel de se demander quels ensembles de suites de Beatty forment un système qui couvre les entiers.

D'abord, en 1926, Beatty a étudié le cas où $\beta = 0$. Il énonce le résultat suivant :

Proposition 3.2.1 *(Beatty, 1926) $S(\alpha_1, 0)$ et $S(\alpha_2, 0)$ forment un système couvrant de suites de Beatty disjointes si et seulement si α_1 est irrationnel et $\alpha_1^{-1} + \alpha_2^{-1} = 1$.*

Des preuves indépendantes de ce résultat ont été données un peu plus tard dans (Bang, 1957; Skolem, 1957), puis dans (Fraenkel, Levitt et Shimshoni, 1972). Le cas d'un ensemble de 2 suites de Beatty avec les α_i rationnels et les β quelconques a été prouvé dans (Niven, 1963). En 1927, l'auteur de (Uspensky, 1927) montre qu'il n'existe aucun système couvrant éventuel de suites de Beatty pour 3 suites de Beatty ou plus de la forme $S(\alpha_i, 0)$. Pour ce faire, il utilise une application du théorème d'approximation de Kronecker. Une preuve plus simple de ce résultat est ensuite donnée dans (Graham, 1963).

Des travaux de Skolem découle le résultat suivant.

Proposition 3.2.2 *(Skolem, 1957) Si α_1 est irrationnel, alors $S(\alpha_1, \beta_1)$ et $S(\alpha_2, \beta_2)$ forment un système couvrant éventuel si et seulement si $\frac{1}{\alpha_1} + \frac{1}{\alpha_2} = 1$ et $\frac{\beta_1}{\alpha_1} + \frac{\beta_2}{\alpha_2} \in \mathbb{Z}$.*

Fraenkel a ensuite prouvé (Fraenkel, 1969) le même résultat, mais pour α_1 rationnel. Il a ainsi donné une condition nécessaire et suffisante pour que deux suites de Beatty forment un système couvrant. Dans un autre article (Fraenkel, 1973), l'auteur énonce la conjecture suivante.

Conjecture 3.2.3 *Si $\{S(\alpha_i, \beta_i) : i = 1, \ldots, k\}$ forme un système couvrant éventuel, avec $k \geq 3$ et $\alpha_i \neq \alpha_j$ pour tout $i \neq j$, alors il existe deux α_i rationnels.*

Fraenkel a prouvé cette conjecture pour $k = 3, 4$ et pour quelques cas particuliers pour $k > 4$. Erdos et Graham (Erdos et Graham, 1980) attribuent la conjecture suivante à Fraenkel.

Conjecture 3.2.4 (Conjecture de Fraenkel) *Si les suites de Beatty rationnelles*

$$\left\{ S\left(\frac{p_i}{q_i}, b_i\right) : i = 1, \ldots, k \right\}$$

forment un système couvrant éventuel, avec $k \geq 3$ et $p_1/q_1 < p_2/q_2 < \cdots < p_k/q_k$, alors pour $1 \leq i \leq k$,

$$p_i = 2^k - 1, \qquad q_i = 2^{k-i}.$$

Entre autre, Graham (Graham, 1973) a utilisé le théorème d'approximation de Kronecker et les travaux de Skolem sur les suites de Beatty afin de montrer que si tous les α d'un système couvrant éventuel de plus de 2 suites de Beatty sont différents, alors tous les α sont rationnels. De plus, si α est rationnel, alors la suite des différences entre deux éléments consécutifs de $S(\alpha, \beta)$ est une suite périodique. Ainsi, aucune différence n'est nécessaire entre les systèmes couvrants éventuels et les systèmes couvrants lorsque l'on étudie les systèmes couvrants de suites de Beatty avec des α tous différents, comme c'est le cas pour la Conjecture 3.2.4. C'est pourquoi qu'à partir de maintenant, nous nous intéressons à la conjecture de Fraenkel que pour des systèmes couvrants.

Depuis l'énoncé de la conjecture de Fraenkel, plusieurs autres chercheurs se sont intéressés aux suites de Beatty. Il sufffit de penser aux travaux suivants : (Storlarsky, 1976; Morikawa, 1982; Morikawa, 1983; Morikawa, 1984; Morikawa, 1985a; Morikawa, 1985b; Morikawa, 1985c; Morikawa, 1988; Morikawa, 1989; Simpson, 1991; Morikawa, 1992; Morikawa, 1993; Morikawa, 1995; Tijdeman, 1996a; Tijdeman, 1996b; Tijdeman, 1998; Tijdeman, 2000a; Tijdeman, 2000b; Simpson, 2004).

Le travail (Tijdeman, 2000a) offre un excellent survol des derniers travaux sur ce sujet.

Proposition 3.2.5 *Si l'ensemble des k suites de Beatty $\left\{ S\left(\frac{p_i}{q_i}, \beta_i\right) : i = 1, \ldots, k \right\}$ forme un système couvrant, avec $k \geq 3$, alors on peut lui associer une unique suite bi-infinie s sur l'alphabet à k lettres $\mathcal{A} = \{1, 2, \ldots, k\}$, à permutation des lettres et à décalage près, de la façon suivante : $s[i] = j$, si $i \in S\left(\frac{p_j}{q_j}, \beta_j\right)$.*

Exemple 3.2.6 Avec les suites de Beatty de l'Exemple 3.1.8, on trouve $s[-2] = s[5] = 1, s[-1] = s[6] = 2, s[0] = 1, s[1] = 1, s[2] = 2, s[3] = 1, s[4] = 3, s[5] = 1, s[6] = 2, s[7] = s[0] = 1, s[8] = s[1] = 1, \ldots$ et donc,

$$s = \cdots 12 \cdot 112131211 \cdots = {}^{\omega}(1121312)^{\omega} \in {}^{\omega}\{1, 2, 3\}^{\omega}.$$

Dans une suite d'entiers, un *intervalle* est un ensemble fini d'entiers consécutifs. Un intervalle de n entiers est appelé un *intervalle de longueur* n. Une suite d'entiers S est dite *équilibrée* si

le nombre de termes de S dans n'importe quels deux intervalles d'entiers de même longueur diffère d'au plus 1.

Exemple 3.2.7 Soit la suite d'entiers $S = 1, 3, 4, 5, 7, 9, 11, 13, 14, 16, \ldots$. Considérons les intervalles de longueur 3 : $[3, 5]$ et $[8, 10]$. Dans le premier, il y a 3 entiers de S et dans le deuxième, il n'y en a qu'un. Ainsi, S n'est pas équilibrée.

Proposition 3.2.8 *(Tijdeman, 2000a) Soit $S = \{\lfloor \alpha n + \beta \rfloor\}_{n \in \mathbb{Z}}$, une suite de Beatty périodique, avec $\alpha, \beta \in \mathbb{R}$. S est équilibrée si et seulement s'il existe des nombres rationnels α', β' avec $\alpha' > 1$ tels que $S = \{\lfloor \alpha' n + \beta' \rfloor\}_{n \in \mathbb{Z}}$.*

Lemme 3.2.9 *(Tijdeman, 2000a) Soit la suite de Beatty $S(\alpha, \beta)$, avec α rationnel. Alors, on peut supposer que β est un nombre rationnel de même dénominateur que α.*

Proposition 3.2.10 *Les suites de Beatty rationnelles sont équilibrées.*

Preuve Par le Lemme 3.2.9, une suite $S(\alpha, \beta)$, avec $\alpha = p/q$ rationnel, peut toujours s'écrire comme

$$\{\lfloor n\alpha + \beta \rfloor\}_{n \in \mathbb{Z}} = \{\lfloor np/q + \beta \}_{n \in \mathbb{Z}} = \{\lfloor np/q + r/q \rfloor\}_{n \in \mathbb{Z}}.$$

On conclut en utilisant la Proposition 3.2.8. ■

Corollaire 3.2.11 *Les suites de Beatty satisfaisant à la conjecture de Fraenkel sont équilibrées.*

Preuve Rappelons que la conjecture de Fraenkel suppose les suites de Beatty rationnelles. On conclut en utilisant la Proposition 3.2.10. ■

Dans (Altman, Gaujal et Hordijk, 1998), les auteurs montrent que tout mot infini sur 3 lettres ou plus avec des fréquences de lettres toutes différentes est périodique. Ils prouvent aussi la conjecture de Fraenkel pour 4 suites de Beatty. Puis dans (Tijdeman, 2000a), l'auteur prouve la conjecture pour 5 suites de Beatty.

En utilisant la Proposition 3.2.10 et la bijection décrite dans la Proposition 3.2.5, on peut réécrire la conjecture de Fraenkel de la façon qui suit.

Conjecture 3.2.12 (Conjecture de Fraenkel en termes de combinatoire des mots) *Toute suite bi-infinie équilibrée sur un alphabet à k lettres, où $k \geq 3$, et ayant des fréquences de lettres toutes différentes est de la forme ${}^{\omega}\mathrm{Fr}_k^{\omega}$, où*

$$\mathrm{Fr}_k = \mathrm{Fr}_{k-1} \cdot k \cdot \mathrm{Fr}_{k-1},$$

avec $\mathrm{Fr}_2 = 121$.

Notation 3.2.13 Les mots Fr_k sont aussi connus dans la littérature sous le nom de *mots de Zimin.*

Lemme 3.2.14 $\lg(\mathrm{Fr}_k) = 2^k - 1$.

Preuve Procédons par récurrence. Pour $k = 2$, on a $\mathrm{Fr}_2 = 121$ et donc, $\lg(\mathrm{Fr}_2) = 2^2 - 1 = 3$. Supposons que $\lg(\mathrm{Fr}_k) = 2^k - 1$ pour tout $k \leq n$. Montrons que c'est alors vrai pour $(n+1)$. Comme $\mathrm{Fr}_{n+1} = \mathrm{Fr}_n n \mathrm{Fr}_n$, on a

$$\lg(\mathrm{Fr}_{n+1}) = \lg(\mathrm{Fr}_n n \mathrm{Fr}_n) = 2\lg(\mathrm{Fr}_n) + 1 = 2(2^n - 1) + 1 = 2^{n+1} - 1.$$

■

Lemme 3.2.15 *La fréquence de la lettre i dans* Fr_k *est* $2^{k-i}/(2^k - 1)$.

Preuve Montrons par récurrence que le nombre d'occurrences de la lettre i dans Fr_k est 2^{k-i}. Pour $k = 2$, on a $|121|_1 = 2^{2-1} = 2$ et $|121|_2 = 2^{2-2} = 1$. Supposons que $|\mathrm{Fr}_k|_i = 2^{k-i}$ pour tout $k \leq n$ et montrons que c'est alors vrai pour $(n+1)$. Comme $\mathrm{Fr}_{n+1} = \mathrm{Fr}_n(n+1)\mathrm{Fr}_n$, on a $|\mathrm{Fr}_{n+1}|_i = |\mathrm{Fr}_n(n+1)\mathrm{Fr}_n|_i = 2|\mathrm{Fr}_n|_i = 2(2^{n-i}) = 2^{n+1-i}$ si $i \neq (n+1)$, et $|\mathrm{Fr}_{n+1}|_i = |\mathrm{Fr}_n(n+1)\mathrm{Fr}_n|_i = 1 = 2^{(n+1)-(n+1)}$ si $i = (n+1)$. On conclut en calculant la fréquence :

$$\frac{|\mathrm{Fr}_k|_i}{\lg(\mathrm{Fr}_k)} = \frac{2^{k-i}}{2^k - 1}.$$

■

Définition 3.2.16 On appelle *mots de Fraenkel* les mots de la forme Fr_k^{ω}.

Exemple 3.2.17 Ainsi, la seule suite équilibrée ayant des fréquences de lettres différentes sur 3 lettres serait

$$^{\omega}(1213121)^{\omega}.$$

Sur 4 lettres, ce serait

$$^{\omega}(1213121412131 21)^{\omega}.$$

Lemme 3.2.18 $\text{Pal}(12\cdots k^{\omega}) = (\text{Pal}(12\cdots(k-1))k)^{\omega}$, *où* Pal *est l'opérateur introduit au Chapitre II.*

Preuve Il suffit d'utiliser le Lemme 2.3.5 un nombre infini de fois. ■

Lemme 3.2.19 $\text{Fr}_k = \text{Pal}(12\cdots k)$ *et* $\text{Pal}(12\cdots k^{\omega}) = (\text{Fr}_{k-1}k)^{\omega}$.

Preuve Pour la première égalité, on procède par récurrence sur k. Pour $k = 3$, on a $\text{Fr}_3 = 1213121$ et $\text{Pal}(123) = 1213121$. Supposons que l'égalité soit vraie pour tout $n \leq k$ et vérifions pour $(k+1)$. Comme $\text{Fr}_{k+1} = \text{Fr}_k(k+1)\text{Fr}_k$, en utilisant l'hypothèse de récurrence, on trouve $\text{Fr}_{k+1} = \text{Pal}(12\cdots k)(k+1)\text{Pal}(12\cdots k) = \text{Pal}(12\cdots k(k+1))$. Pour la deuxième égalité, par le Lemme 3.2.18, on a

$$\text{Pal}(12\cdots k^{\omega}) = (\text{Pal}(12\cdots k-1)k)^{\omega} = (\text{Fr}_{k-1}k)^{\omega}.$$

■

Lemme 3.2.20 *(Justin, 2005) Soit* $w \in \mathcal{A}^*$, $a \in \mathcal{A}$. *Alors* $\text{Pal}(wa) = \psi_w(a)\text{Pal}(w)$.

Proposition 3.2.21 *Les mots* Fr_k *appartiennent à des classes épichristoffelles. Plus précisément,*

$$\text{Fr}_k = \psi_{12\cdots k}(1).$$

Preuve Procédons par récurrence. Pour $k = 3$, on a $\text{Fr}_3 = 1213121$ et $\psi_{123}(1) = 1213121$. Supposons que l'énoncé est vrai pour tout $k \leq n$. Soit $\text{Fr}_{n+1} = \text{Fr}_n(n+1)\text{Fr}_n$. En utilisant les

Lemmes 3.2.19 et 3.2.20, on a $\text{Fr}_{n+1} = \text{Pal}(12\cdots(n+1)) = \psi_{12\cdots n}(n+1)\text{Pal}(1\cdots n)$. Par hypothèse de récurrence, on a

$$\psi_{12\cdots n}(n+1)\text{Pal}(1\cdots n) = \psi_{12\cdots n}(n+1)\psi_{12\cdots n}(1) = \psi_{12\cdots n}((n+1)1) = \psi_{12\cdots(n+1)}(1).$$

■

3.3 Caractérisation des suites épisturmiennes équilibrées

Définition 3.3.1 Un mot fini $w \in \mathcal{A}^*$ est *linéaire* si toute lettre de l'alphabet apparaît au plus une fois dans w.

Exemple 3.3.2 Le mot $w = 123415$ n'est pas linéaire puisque $|w|_1 > 1$, alors que le mot $u = 526134$ l'est.

Lemme 3.3.3 *Soit* $1 \in \mathcal{A}$ *et* $w \in \mathcal{A}^*$. *Si* $1w$ *est un mot linéaire, alors* $\text{Pal}(1w1^k) = (\text{Pal}(1w))^{k+1}$.

Preuve On procède par récurrence sur k et on utilise le Lemme 2.3.5. Pour $k = 0$, c'est trivial. Pour $k = 1$, on a la factorisation $1w = \varepsilon \cdot 1 \cdot w$ et on trouve $\text{Pal}(1w1) = \text{Pal}(1w) \cdot (\text{Pal}(\varepsilon))^{-1}\text{Pal}(1w) = (\text{Pal}(1w))^2$. Supposons maintenant que pour tout $k \leq n$, on ait $\text{Pal}(1w1^k) = (\text{Pal}(1w))^{k+1}$. Alors

$$\begin{aligned}
\text{Pal}(1w1^{n+1}) &= \text{Pal}(1w1^n)(\text{Pal}(1w1^{n-1})^{-1}\text{Pal}(1w1^n) \\
&= (\text{Pal}(1w))^{n+1}((\text{Pal}(1w))^n)^{-1}(\text{Pal}(1w))^{n+1} \\
&= (\text{Pal}(1w))^{n+1-n+n+1} \\
&= (\text{Pal}(1w))^{n+2}.
\end{aligned}$$

■

Définition 3.3.4 Soit $w \in \mathcal{A}^\infty$ un mot non linéaire et écrivons $w = pw'$, où p est le plus long préfixe linéaire de w. Alors $w'[0]$ est appelé la *première lettre répétée*.

Exemple 3.3.5 Soit $u = 1432$, $v = 1212312$ et $w = 12321$. Seul u est linéaire, la première lettre répétée de v est 1 et celle de w est 2.

Exemple 3.3.6 Soit $\mathrm{Fr}_3 = 1213121$. Ce mot est décrit par la suite directrice $\Delta(\mathrm{Fr}_3) = 123$. On a aussi $\mathrm{Fr}_4 = 121312141213121$ et $\Delta(\mathrm{Fr}_4) = 1234$. De manière générale, on peut vérifier que $\Delta(\mathrm{Fr}_k) = 123\cdots k$.

Remarquant dans l'Exemple 3.3.6 que les mots de Fraenkel semblent tous être déterminés par une suite directrice, nous nous intéressons aux suites épisturmiennes équilibrées. Dans un premier temps, nous caractérisons les suites épisturmiennes standards équilibrées, puis en utilisant cette caractérisation et la propriété que pour toute suite épisturmienne non standard, il existe une suite épisturmienne standard ayant le même langage, nous caractérisons de façon générale l'ensemble des suites épisturmiennes équilibrées.

Dans ce chapitre, afin de montrer qu'une suite n'est pas équilibrée, nous allons toujours donner deux facteurs f et f' de même longueur tels que $\big||f|_{f[0]} - |f'|_{f[0]}\big| \geq 2$: le déséquilibre sera sur la première lettre de f, soit la lettre $f[0]$.

Commençons par quelques exemples.

Exemple 3.3.7

a) Soit s une suite épisturmienne standard définie par la suite directrice $\Delta(s) = 1232\cdots$. Alors

$$s = \mathrm{Pal}(1232\cdots) = \underline{12}1\underline{3}121\underline{2}13121\cdots$$

qui contient les facteurs 212 et 131. Ainsi, s n'est pas équilibrée par rapport à la lettre 2.

b) Soit t une suite épisturmienne standard définie par la suite directrice $\Delta(t) = 12131\cdots$. Alors

$$t = \mathrm{Pal}(12131\cdots) = \underline{12}1\underline{1}21\underline{3}12112\underline{1}213121121\cdots$$

qui contient les facteurs 11211 et 21312. Ainsi, t n'est pas équilibrée par rapport à la lettre 1.

c) Soit u une suite épisturmienne standard définie par la suite directrice $\Delta(u) = 12341^\omega$. Alors

$$u = \mathrm{Pal}(12341^\omega) = (\underline{12}1\underline{3}121\underline{4}1213121)^\omega$$

qui est une suite équilibrée.

La Proposition 3.3.9 décrit le cas des suites épisturmiennes standards équilibrées ayant des suites directrices avec comme première lettre répétée une lettre différente de la première lettre

du mot. La Proposition 3.3.11 étudie celles ayant des suites directrices de la forme $\Delta(s) = 11z$, avec $z \in \mathcal{A}^\omega$. Avant de présenter ces propositions, le lemme suivant est préalable.

Lemme 3.3.8 *Soit $\Delta(s) = x\alpha^\ell y$ la suite directrice d'une suite épisturmienne standard équilibrée s, avec $\alpha \in \mathcal{A}$, $x \in \mathcal{A}^+$, $y \in \mathcal{A}^\omega$, $\ell \geq 2$ et $y[0] \neq \alpha$. Si $x\alpha$ est linéaire, alors*

$$\text{Alph}(x) \cap \text{Alph}(y) = \emptyset.$$

Preuve Supposons qu'il existe $\beta \in \text{Alph}(x) \cap \text{Alph}(y)$ tel que $\Delta(s) = x'\beta x''\alpha^\ell y'\beta y''$, avec $x'\beta x''y'$ linéaire. Posons $p = \text{Pal}(x'\beta x'')$. Il y a 3 cas possibles.

a) Si $x' \neq \varepsilon$, alors

$$s = p(\underline{\alpha}p)^\ell \cdots (p\alpha)^\ell p\underline{\beta}p[0] \cdots ,$$

qui contient les facteurs $\alpha p\alpha$ et $p\beta p[0]$, $p[0] \neq \alpha$, $p[0] \neq \beta$. Ainsi, s n'est pas équilibrée.

b) Si $x' = \varepsilon$ et $x'' \neq \varepsilon$, alors

$$s = (p(\underline{\alpha}p)^\ell \cdots (p\alpha)^\ell p)^2 \cdots ,$$

qui contient $\alpha p\alpha$ et $pp[0]p[1]$, $p[0] = \beta \neq \alpha$, $p[1] = x''[0] \neq \alpha$. Alors s n'est pas équilibrée.

c) Si $x' = \varepsilon$ et $x'' = \varepsilon$, alors comme s est sur un alphabet contenant au moins 3 lettres, il existe $\gamma \in \{\mathcal{A} - \{\alpha, \beta\}\}$ tel qu'à sa première occurrence,

$$s = \underline{\beta}(\underline{\alpha}\beta)^\ell \cdots \beta\underline{\gamma}\beta \cdots ,$$

qui contient $\alpha\beta\alpha$ et $\beta\gamma\beta$. Ainsi, s n'est pas équilibrée.

■

Proposition 3.3.9 *Soit $\Delta(s)$ la suite directrice d'une suite épisturmienne standard équilibrée s sur un alphabet à k lettres $\mathcal{A} = \{1, 2, \ldots, k\}$, $k \geq 3$. Supposons que k soit la première lettre répétée de s. Si $k \neq s[0]$, alors la suite directrice s'écrit comme $\Delta(s) = 12\cdots(k-1)k^\omega$, à permutation de lettres près.*

Preuve Soit $\Delta(s) = xkykz$ la suite directrice d'une suite épisturmienne standard équilibrée s, avec $x \in \mathcal{A}^+$, $y \in \mathcal{A}^*$, $z \in \mathcal{A}^\omega$ et xky un mot linéaire. Posons $p = \text{Pal}(x)$ et supposons

$y \neq \varepsilon$. Alors

$$s = p\underline{k}p\underline{y[0]}pkp \cdots pkpy[0]pkp\underline{k}p \cdots$$

qui contient les facteurs kpk et $py[0]p[0]$, $p[0] \neq k$ et $y[0] \neq k$ (comme xky est linéaire). Alors s n'est pas équilibrée par rapport à la lettre k. Ainsi, $y = \varepsilon$. Conséquemment, on a $\Delta(s) = xk^2z$. Supposons $z \neq k^\omega$. On réécrit $\Delta(s) = xk^\ell z'$, avec $z'[0] \neq k$ et $\ell \geq 2$. Alors, par le Lemme 3.3.8, comme xk est linéaire, $z'[0] \notin \text{Alph}(x)$. Ainsi, $xz'[0]$ est linéaire et donc

$$s = p(\underline{k}p)^\ell \underline{z'[0]}p[0] \cdots .$$

Donc s contient les facteurs kpk et $pz'[0]p[0]$, avec $z'[0] \neq k$ et $p[0] = x[0] \neq k$. ■

Lemme 3.3.10 *Soit $\Delta(s) = 1^\ell w$, avec $\ell \geq 1$ et $w \in \mathcal{A}^\omega$. Alors toute lettre $\alpha \neq 1$ dans s est séparée par un facteur contenant 1^ℓ.*

Preuve Considérons le préfixe $1^\ell\alpha$ de $\Delta(s)$ et supposons $\alpha \neq 1$. Alors $s = 1^\ell\alpha 1^\ell \cdots$. La lettre α est bien précédée et suivie de 1^ℓ. Considérons maintenant le préfixe $1^\ell y\alpha$ de la suite directrice, avec $y \in \mathcal{A}^+$. Alors $(\text{Pal}(1^\ell y)\alpha)^{(+)}$ est préfixe de s. Mais $\text{Pal}(1^\ell y)$ commence et termine par 1^ℓ. Il s'écrit donc comme $1^\ell z 1^\ell$. Ainsi, $(\text{Pal}(1^\ell y)\alpha)^{(+)} = 1^\ell z 1^\ell \alpha 1^\ell \cdots$. D'où la conclusion. ■

Proposition 3.3.11 *Soit $\Delta(s)$ la suite directrice d'une suite épisturmienne standard équilibrée s sur un alphabet à k lettres $\mathcal{A} = \{1, 2, \ldots, k\}$, $k \geq 3$. Si $\Delta(s) = 1^\ell z$, avec $z \in \mathcal{A}^\omega$, $z[0] \neq 1$ et $\ell \geq 2$, alors $\Delta(s) = 1^\ell 23 \cdots (k-1)k^\omega$, à permutation de lettres près.*

Preuve Soit $\Delta(s) = 1^\ell z$, la suite directrice d'une suite épisturmienne standard équilibrée s, avec $z[0] \neq 1$, $\ell \geq 2$. Supposons $|z|_1 > 0$. Alors $\Delta(s) = 1^\ell z' 1 z''$, avec $z' \neq \varepsilon$ et $|z'|_1 = 0$. Comme s est sur un alphabet à au moins 3 lettres, il existe au moins une lettre α dans z' ou z'' différente de $z'[0]$ et 1. À sa première occurrence dans s, α est précédée et suivie par 1^ℓ (voir Lemme 3.3.10). On a

$$s = \underline{1^\ell}\underline{z'}[0]1^\ell \cdots 1^\ell z'[0]1^\ell \underline{1}z'[0] \cdots$$

qui contient les facteurs $z'[0]1^{\ell+1}z'[0]$ et $1^\ell\alpha 1^2$. Ainsi, $|z|_1 = 0$. Par ailleurs, comme l'alphabet est fini, il existe au moins une lettre différente de 1 qui apparaît au moins deux fois dans z. Considérons γ, la première lettre répétée de z. Alors $\Delta(s) = 1^\ell u\gamma v\gamma w$, avec $u\gamma v$ linéaire. Supposons $v \neq \varepsilon$ et posons $p = \text{Pal}(1^\ell u)$. Alors

$$s = p\underline{\gamma}p\underline{v[0]}p\gamma p\cdots p\gamma p\underline{\gamma}\cdots$$

qui contient les facteurs $\gamma p\gamma$, $pv[0]p[0]$, $v[0] \neq \gamma$ et $p[0] = 1$. Il en découle que $v = \varepsilon$. Considérons maintenant $\Delta(s) = 1^\ell u\gamma^2 w$, que nous réécrivons comme $\Delta(s) = 1^\ell u\gamma^m w'$, avec u linéaire, $m \geq 2$ et $w'[0] \neq \gamma$. Alors

$$s = p(\underline{\gamma}p)^m\underline{w'[0]}p[0]\cdots$$

qui contient les facteurs $\gamma p\gamma$ et $pw'[0]p[0]$. Ainsi, $w' = \gamma^\omega$ et on conclut. ■

Les Propositions 3.3.13 et 3.3.16 étudient les suites épisturmiennes standards équilibrées ayant des suites directrices de la forme $\Delta(s) = 1y1z$, avec $y \neq \varepsilon$ et $1y$ linéaire. Mais avant, introduisons le résultat suivant.

Lemme 3.3.12 *Soit $\Delta(s) = 1y1z$ la suite directrice d'une suite épisturmienne standard équilibrée s, avec $1y \in \mathcal{A}^*$ un mot linéaire, $y \neq \varepsilon$, $z \in \mathcal{A}^\omega$. Alors,* $\text{Alph}(y) \cap \text{Alph}(z) = \emptyset$.

Preuve Supposons qu'il existe $\alpha \in \text{Alph}(y) \cap \text{Alph}(z)$. Alors on peut écrire $\Delta(s) = 1y'\alpha y''1z'\alpha z''$, avec $y'\alpha y''1z'$ un mot linéaire. Posons $p = \text{Pal}(1y')$. Il y a deux cas à considérer.

a) Si $y' \neq \varepsilon$, comme $\text{Pal}(1y'\alpha y''1) = (\text{Pal}(1y'\alpha y''))^2$, on a

$$s = (p\underline{\alpha}p\cdots p\alpha p)^2\cdots(p\alpha p\cdots p\alpha p)^2\underline{\alpha}\cdots$$

qui contient les facteurs $\alpha p\alpha$ et $pp[0]p[1] = p1y'[0]$, $1 \neq \alpha$, $y'[0] \neq \alpha$.

b) Si $y' = \varepsilon$, alors comme s est sur un alphabet à au moins 3 lettres, il existe une lettre $\beta \in \{\mathcal{A} - \{1, \alpha\}\}$ dans s telle qu'à sa première occurrence, elle est précédée et suivie par 1 (voir Lemme 3.3.10). On a alors

$$s = (\underline{1\alpha}1\cdots 1\alpha 1)^2\cdots(1\alpha 1\cdots 1\alpha 1)^2\underline{\alpha}\cdots$$

qui contient les facteurs $\alpha 1\alpha$ et $1\beta 1$.

D'où la conclusion. ■

Proposition 3.3.13 *Soit $\Delta(s) = 1y1z$ la suite directrice d'une suite épisturmienne standard équilibrée s sur un alphabet à au moins 3 lettres. Supposons que $|1y1z|_1 = 2$ et que $y \neq \varepsilon$ est un mot linéaire. Alors $\Delta(s) = 12\cdots(k-1)1k\cdots(k+\ell-1)(k+\ell)^\omega$, à permutation de lettres près, avec $k \geq 3$.*

Preuve Considérons $\Delta(s) = 1y1z$ avec $|1y1z|_1 = 2$, y linéaire et supposons que $\alpha \in \mathcal{A}$ est la première lettre répétée différente de 1, donc la première lettre répétée de yz. Alors, il y a deux cas à considérer.

a) $\Delta(s) = 1y'\alpha y''1z'\alpha z''$, avec $|y'y''1z'|_\alpha = 0$. Impossible par le Lemme 3.3.12.

b) $\Delta(s) = 1y'1z'\alpha z''\alpha z'''$, avec $|y'1z'z''|_\alpha = |z'\alpha z''\alpha z''|_1 = 0$. Supposons $z'' \neq \varepsilon$ et posons $p = \text{Pal}(1y'1z')$. Alors

$$s = p\underline{\alpha}p\underline{z''[0]}p\alpha p\cdots p\alpha p z''[0]p\alpha p\underline{\alpha}\cdots$$

qui a les facteurs $\alpha p\alpha$ et $pz''[0]p[0]$. On conclut que $z'' = \varepsilon$.

La seule possibilité est donc que $\Delta(s) = 1y'1z'\alpha^2 z'''$, que nous réécrivons comme $\Delta(s) = 1y'1z'\alpha^\ell z''''$, avec $\ell \geq 2$ et $z''''[0] \neq \alpha$. Posons $p = \text{Pal}(1y'1z')$. Alors,

$$s = p(\underline{\alpha}p)^\ell \underline{z''''[0]}p[0]\cdots$$

qui a les facteurs $\alpha p\alpha$ et $pz''''[0]p[0]$, avec $z''''[0] \neq \alpha$, $p[0] = 1 \neq \alpha$. Il en découle que $z'''' = \alpha^\omega$, donc que $z''' = \alpha^\omega$. Finalement, $\Delta(s) = 1y'1z'\alpha^\omega$ avec $y'1z'\alpha$ linéaire. ■

Exemple 3.3.14 Soit s une suite épisturmienne standard ayant comme suite directrice $\Delta(s) = 12321\cdots$. Alors,

$$s = \underline{12}1\underline{3}121\underline{2}13121\underline{1}213121213121\cdots$$

qui contient les facteurs 212 et 131. Ainsi, s n'est pas équilibrée par rapport à la lettre 2.

Exemple 3.3.15 Soit t une suite épisturmienne standard ayant comme suite directrice $\Delta(t) = 12312\cdots$. Alors,

$$t = \underline{12}1\underline{3}1211213121\underline{2}13121\cdots$$

qui contient les facteurs 212 et 131. Ainsi, t n'est pas équilibrée par rapport à la lettre 2.

Proposition 3.3.16 *Soit $\Delta(s) = 1y1z$ la suite directrice d'une suite épisturmienne standard équilibrée s sur un alphabet à k lettres, $k \geq 3$, telle que $1y$ est linéaire, $y \neq \varepsilon$ et $z \in \mathcal{A}^\omega$.*
i) *Si $z[0] \neq 1$, alors $|z|_1 = 0$ et donc, $\Delta(s)$ est de la forme donnée dans la Proposition 3.3.13.*
ii) *Si $z[0] = 1$, alors $z = 1^\omega$ et donc, $\Delta(s) = 123\cdots k(1)^\omega$.*

Preuve Supposons qu'il y ait une troisième occurrence de 1 dans la suite directrice : $|z|_1 \geq 1$. Alors $\Delta(s) = 1y1z'1z''$, avec $|z'|_1 = 0$. Supposons que $z' \neq \varepsilon$. Posons $p = \text{Pal}(1y)$. Comme $\text{Pal}(1y1) = (\text{Pal}(1y))^2 = p^2$, on a

$$s = p^2\underline{z'[0]}p^2\cdots p^2z'[0]p^2\underline{1}\cdots$$

qui contient les facteurs $1p1$ (dans p^21) et $1^{-1}pz'[0]p[0]p[1]$, où $z'[0] \neq p[0] = 1$ et $z'[0] \neq p[1] = y[0]$ (assuré par le Lemme 3.3.12). Alors, $z' = \varepsilon$ et $\Delta(s) = 1y11z''$. Réécrivons $\Delta(s) = 1y1^\ell z'''$, avec $\ell \geq 2$ et $z'''[0] \neq 1$. Comme $\text{Pal}(1y1^\ell) = (\text{Pal}(1y))^{\ell+1}$ si $1y$ est linéaire, on a

$$s = p^{\ell+1}\cdots p^{\ell+1}\underline{z'''[0]}1y[0]\cdots$$

qui contient les facteurs $1p1$ (dans p^3) et $1^{-1}pz'''[0]1y[0]$, avec $y[0] \neq 1$. On aurait donc que s n'est pas équilibrée. On conclut que $z = 1^\omega$. Donc $\Delta(s) = 1y1^\omega$, avec w linéaire. ∎

On a maintenant considéré toutes les possibilités de suites directrices pour une suite épisturmienne standard équilibrée. En effet, la Proposition 3.3.9 décrit les suites directrices ayant comme première lettre répétée une lettre différente de la première. La Proposition 3.3.11 donne la forme d'une suite directrice débutant par un bloc de lettres. Finalement, les Propositions 3.3.13 et 3.3.16 décrivent les suites directrices de la forme $1y1z$ ayant ou pas d'autres occurrences de la lettre 1.

Le Théorème 3.3.17 résume ces résultats.

Théorème 3.3.17 *Toute suite épisturmienne standard équilibrée s sur un alphabet à trois lettres et plus a une suite directrice, à permutation de lettres près, dans l'une ou l'autre des trois familles de suites suivantes :*

i) $\Delta(s) = 1^n \left(\prod_{i=2}^{k-1} i\right) (k)^\omega = 1^n 23\cdots(k-1)(k)^\omega$, *avec* $n \geq 1$ *;*

ii) $\Delta(s) = \left(\prod_{i=1}^{k-1} i\right) 1 \left(\prod_{i=k}^{k+\ell-1} i\right) (k+\ell)^\omega = 12\cdots(k-1)1k\cdots(k+\ell-1)(k+\ell)^\omega$, *avec* $\ell \geq 1$ *;*

iii) $\Delta(s) = \left(\prod_{i=1}^{k} i\right) (1)^\omega = 123\cdots k(1)^\omega$,

avec $k \geq 3$.

Preuve La Proposition 3.3.9 implique a) pour $n = 1$, la Proposition 3.3.11 implique a) pour $n \geq 2$, alors que b) (resp. c)) découle de la Proposition 3.3.13 (resp. Proposition 3.3.16). ∎

Nous devons maintenant vérifier que ces suites directrices décrivent bien des suites épisturmiennes qui sont équilibrées. Pour ce faire, certains résultats de (Hubert, 2000) concernant les suites équilibrées sont nécessaires.

3.4 Quelques notions de suites équilibrées

Définition 3.4.1 (Hubert, 2000) Soit s une suite sur un alphabet fini $\mathcal{A}$. On dit que s est *fortement équilibrée* si pour tout nombre entier n, pour tout couple f et f' de $F_n(s) \cup F_{n-1}(s)$ et pour toute lettre a de l'alphabet $\mathcal{A}$, on a

$$\big||f|_a - |f'|_a\big| \leq 1.$$

Lemme 3.4.2 *Toute suite fortement équilibrée est aussi équilibrée.*

Preuve Il suffit de remarquer que $F_n(s) \subset (F_n(s) \cup F_{n-1}(s))$. ∎

Exemple 3.4.3 La suite $s = 1211212112112121121\cdots$ est équilibrée, mais n'est pas fortement équilibrée. En effet, pour les facteurs $f = 11211, f' = 2121 \in F_5(s) \cup F_4(s)$, on trouve que

$$\big||f|_1 - |f'|_1\big| = |4-2| = 2 > 1.$$

Proposition 3.4.4 *((Hubert, 2000), Proposition 2.2) Soient $\mathcal{A}$ un alphabet fini, v une suite sturmienne sur l'alphabet $\{a, b\}$ et $\{\mathcal{A}_1, \mathcal{A}_2\}$ une partition de $\mathcal{A}$. Soit la suite u obtenue en remplaçant les a du mot v par les lettres successives d'une suite fortement équilibrée $z \in \mathcal{A}_1^\omega$ et les lettres b du mot v par les lettres successives d'une suite fortement équilibrée $z' \in \mathcal{A}_2^\omega$. Alors u est une suite équilibrée sur l'alphabet $\mathcal{A}$.*

Les Exemples 3.4.9 et 3.4.10 illustrent la Proposition 3.4.4.

Définition 3.4.5 (Hubert, 2000) Soit m une suite sur un alphabet fini $\mathcal{A}$. On dit que m est une *suite à lacunes constantes* si, pour toute lettre $\alpha \in \mathcal{A}$, les lacunes entre deux apparitions consécutives de α ont une longueur constante.

Exemple 3.4.6 La suite $(23242325)^\omega$ est à lacunes constantes. En effet, la lacune entre deux occurrences de 2 est 2, celle entre deux occurrences de 3 est 4 et celle entre deux occurrences de 4 ou de 5 est 8.

Proposition 3.4.7 *((Hubert, 2000), Proposition 3.1) Soit z une suite sur un alphabet fini $\mathcal{A}$. Alors z est fortement équilibrée si et seulement si z est à lacunes constantes. En particulier, une suite fortement équilibrée est périodique.*

Exemple 3.4.8 Comme la suite $(23242325)^\omega$ de l'Exemple 3.4.6 est à lacune constante, elle est aussi fortement équilibrée.

Exemple 3.4.9 Soit $\mathcal{A} = \{1, 2, 3, 4\}$ et $\{\mathcal{A}_1, \mathcal{A}_2\}$ une partition de $\mathcal{A}$: $\mathcal{A}_1 = \{1, 2\}$ et $\mathcal{A}_2 = \{3, 4\}$. Soit la suite sturmienne $s = abaababaabaababaaba\cdots$ et soient les suites fortement équilibrée $(12)^\omega$ et $(34)^\omega$. En remplaçant dans s les a par les lettres de la suite $(12)^\omega$ et les b par les lettres de la suite $(34)^\omega$, par la Proposition 3.4.4, on obtient la suite équilibrée

$$1321423124123141231\cdots$$

Exemple 3.4.10 Considérons la suite sturmienne périodique $(a^n b)^\omega$. En utilisant la Proposition 3.4.4, on sait que si l'on remplace la suite $aaa\cdots$ par 1^ω, une suite fortement équilibrée, et si

l'on remplace la suite $bbb\cdots$ par $(23242325)^\omega$, aussi fortement équilibrée, alors on obtient la suite

$$(1^n21^n31^n21^n41^n21^n31^n21^n5)^\omega$$

qui est une suite équilibrée sur l'alphabet $\{1,2,3,4,5\}$.

Proposition 3.4.11 *Les suites de la forme* $\text{Pal}(123\cdots k^\omega)$ *sont à lacunes constantes.*

Preuve D'abord, par le Lemme 3.2.19, on sait que

$$\text{Pal}(123\cdots k^\omega) = (\text{Pal}(123\cdots(k-1))k)^\omega.$$

Procédons par récurrence pour montrer que $(\text{Pal}(123\cdots(k-1))k)^\omega$ est à lacune constante. Pour $k=4$, on a $(\text{Pal}(123)4)^\omega = (12131214)^\omega$ qui est bien à lacune constante. Supposons donc que cette suite soit à lacune constante pour tout $n \leq k$. Pour $(k+1)$, on a

$$(\text{Pal}(12\cdots k)(k+1))^\omega = (\text{Pal}(12\cdots(k-1))k\text{Pal}(12\cdots(k-1))(k+1))^\omega\,.$$

Par l'hypothèse de récurrence, on sait que les lacunes sont constantes pour les lettres $\leq (k-1)$ et comme k et $(k+1)$ n'apparaissent qu'une seule fois chacune dans la période, on conclut. ■

3.5 Retour à la caractérisation

Proposition 3.5.1 *Les suites directrices données au Théorème 3.3.17 sont les suites directrices de suites épisturmiennes standards qui sont équilibrées.*

Preuve Considérons chacun des 3 types de suites directrices du Théorème 3.3.17.

i) En construisant la suite épisturmienne standard correspondante à la suite directrice i) du Théorème 3.3.17, on obtient une suite qui s'écrit comme

$$s = (\underline{1}^n\underline{2}1^n\underline{3}1^n21^n\underline{4}1^n21^n31^n21^n\cdots1^n21^n\underline{(k-1)}1^n21^n\cdots1^n21^n31^n21^n41^n21^n31^n21^n\underline{k})^\omega.$$

Considérons maintenant la suite obtenue en faisant la projection Π_1 par rapport à la lettre 1. On obtient alors la suite sturmienne $\Pi_1(s) = (1^nx)^\omega$. En remplaçant la lettre x par la suite fortement équilibrée $\text{Pal}(23\cdots k^\omega) = (\underline{23}2\underline{4}232\cdots232\underline{k})^\omega$ (voir Proposition 3.4.11), par la Proposition 3.4.4, on conclut que la suite s est équilibrée .

ii) La suite définie par $\Delta(s) = 12\cdots(k-1)1k\cdots(k+\ell-1)(k+\ell)^\omega$ est donnée par

$$s = \Big(u^2\underline{k}u^2\underline{(k+1)}u^2ku^2\cdots u^2ku^2\underline{(k+\ell-1)}u^2ku^2\cdots u^2(k+1)u^2ku^2\underline{(k+\ell)}\Big)^\omega,$$

où $u = \text{Pal}(12\cdots(k-1))$. Comme $u = \text{Fr}_{k-1}$, s s'écrit comme $(\text{Fr}_{k-1}\text{Fr}_{k-1}A)^\omega$, où A est périodiquement remplacé par les lettres de la suite

$$(\text{Pal}(k(k+1)\cdots(k+\ell-1))(k+\ell))^\omega.$$

Considérons maintenant la suite sturmienne $(a^{2(2^{k-1}-1)}b)^\omega$, qui consiste en la suite épisturmienne s pour laquelle on a remplacé les lettres 1 à $k-1$ par a et les lettres k à $k+\ell$ par b. En effet, par le Lemme 3.2.14, on sait que $\text{lg}(\text{Fr}_{k-1}) = 2^{k-1}-1$. Remplaçons les b par la suite périodique à lacune constante (voir Proposition 3.4.11), donc fortement équilibrée, $\text{Pal}(k(k+1)\cdots(k+\ell-1)(k+\ell)^\omega)$. La suite ainsi obtenue est équilibrée (voir Proposition 3.4.4). Ainsi, s est équilibrée par rapport aux lettres k à $k+\ell$. Afin de montrer que s est équilibrée, il suffit de montrer que s est équilibrée par rapport à la lettre α, pour $1 \leq \alpha \leq (k-1)$. Considérons $s' = (\text{Fr}_{k-1}\text{Fr}_{k-1}b)^\omega = \text{Fr}_{k-1}(\text{Fr}_{k-1}b\text{Fr}_{k-1})^\omega$, avec $b \notin \{1,2,\ldots,k-1\}$ fixé. On a alors que s' est de la forme $\text{Fr}_{k-1}\text{Fr}_k^\omega$. On sait que la suite Fr_k^ω est une suite périodique équilibrée. De plus, comme $\text{Fr}_k^\omega = (\text{Fr}_{k-1}k\text{Fr}_{k-1})^\omega = \text{Fr}_{k-1}k\text{Fr}_{k-1}\text{Fr}_k^\omega$, le préfixe Fr_{k-1} de s' ne peut faire en sorte que s' ne soit pas équilibrée, sinon il en serait de même pour Fr_k^ω. Ainsi, puisque s' est équilibrée, s est équilibrée par rapport aux lettres $1,2,\ldots,(k-1)$ et on conclut que s est une suite épisturmienne équilibrée.

iii) Les suites décrites par les suites directrices données en $c)$ sont de la forme Fr_k^ω et sont connues pour être équilibrées.

■

Théorème 3.5.2 *Une suite épisturmienne standard s sur un alphabet à trois lettres et plus est équilibrée si et seulement si sa suite directrice (à permutation de lettres près) est dans l'une ou l'autre des trois familles de suites suivantes :*

i) $\Delta(s) = 1^n\left(\prod_{i=2}^{k-1} i\right)(k)^\omega = 1^n23\cdots(k-1)(k)^\omega$, *avec* $n \geq 1$ *;*

ii) $\Delta(s) = \left(\prod_{i=1}^{k-1} i\right)1\left(\prod_{i=k}^{k+\ell-1} i\right)(k+\ell)^\omega = 12\cdots(k-1)1k\cdots(k+\ell-1)(k+\ell)^\omega$, *avec* $\ell \geq 1$ *;*

iii) $\Delta(s) = \left(\prod_{i=1}^{k} i\right)(1)^{\omega} = 123\cdots k(1)^{\omega}$,
avec $k \geq 3$.

Preuve Découle du Théorème 3.3.17 et de la Proposition 3.5.1. ■

Rappelons le résultat suivant.

Théorème 3.5.3 ((Droubay, Justin et Pirillo, 2001), Théorème 3) *Une suite épisturmienne standard s est ultimement périodique si et seulement si sa suite directrice $\Delta(s)$ est de la forme $w\alpha^{\omega}$, $w \in \mathcal{A}^*$, $\alpha \in \mathcal{A}$.*

Corollaire 3.5.4 *Une suite épisturmienne standard ne peut pas être à la fois périodique et $\mathcal{A}$-stricte.*

Preuve Découle du Théorème 3.5.3 et de la définition d'une suite $\mathcal{A}$-stricte. ■

Corollaire 3.5.5 *Toute suite épisturmienne standard équilibrée sur un alphabet à trois lettres ou plus est ultimement périodique.*

Preuve Découle directement du Théorème 3.5.2 et du Théorème 3.5.3. ■

Corollaire 3.5.6 *Aucune suite de Arnoux-Rauzy (suites épisturmiennes $\mathcal{A}$-strictes) n'est équilibrée.*

Preuve Découle des Théorèmes 3.5.2 et 3.5.3 et du Corollaire 3.5.4 : une suite épisturmienne $\mathcal{A}$-stricte ne peut être ultimement périodique. ■

Remarque 3.5.7 Une suite $s \in \mathcal{A}^{\omega}$ est *c-équilibrée* si pour tous facteurs u, v de s de même longueur,

$$\big||u|_a - |v|_a\big| \leq c$$

pour tout $a \in \mathcal{A}$. Dans (Cassaigne, Ferenczi et Zamboni, 2000), les auteurs ont prouvé qu'il est possible de construire une suite d'Arnoux-Rauzy qui n'est pas c-équilibrée pour tout $c \in \mathbb{N}$.

Corollaire 3.5.8 *Toute suite épisturmienne standard équilibrée s sur un alphabet à 3 lettres et plus appartient à l'une ou l'autre des familles suivantes, à permutation de lettres près :*
i) $s = (p(k-1)pk)^\omega$, *avec* $p = \mathrm{Pal}(1^n 2 \cdots (k-2))$ *;*
ii) $s = ((\mathrm{Fr}_{k-1})^2 t)^\omega$, *où* t *est successivement remplacé par les lettres de la suite*

$$[\mathrm{Pal}(k(k+1)\cdots(k+\ell-1))(k+\ell)]^\omega;$$

iii) $s = [\mathrm{Pal}(123\cdots k)]^\omega = (\mathrm{Fr}_k)^\omega$.

Preuve On obtient ce résultat directement en calculant $\mathrm{Pal}(\Delta(s))$ en prenant les suites directrices du Théorème 3.5.2. ∎

3.6 Retour à la conjecture de Fraenkel

Proposition 3.6.1 *Toute suite épisturmienne standard équilibrée s sur un alphabet à 3 lettres et plus et ayant des fréquences de lettres différentes pour toutes les lettres peut être écrite, à permutation de lettres près, sous la forme*

$$s = [\mathrm{Pal}(123\cdots k)]^\omega .$$

Preuve Dans le Corollaire 3.5.8 i), les lettres k et $(k-1)$ apparaissent une fois par période. Ainsi, les fréquences de k et $(k-1)$ dans s sont les mêmes. Dans ii), on a $s = (\mathrm{Fr}_{k-1}\mathrm{Fr}_{k-1}t)^\omega$, avec t périodiquement remplacée par $[\mathrm{Pal}(k(k+1)\cdots(k+\ell-1))(k+\ell)]^\omega$. Comme les lettres $(k+\ell)$ et $(k+\ell-1)$ n'apparaissent qu'une seule fois chacune dans $\mathrm{Pal}(k(k+1)\cdots(k+\ell-1))(k+\ell)$, la fréquence de ces deux lettres est la même dans une période. Dans iii), la période de s est $\lg(\mathrm{Pal}(123\ldots k)) = 2^k - 1$ (voir Lemme 3.2.14) et la fréquence de la lettre i est $2^{k-i}/(2^k-1)$ (voir Lemme 3.2.15). Ainsi, les fréquences de deux lettres différentes sont différentes. ∎

Comme pour toute suite épisturmienne t, il existe une suite épisturmienne standard s ayant le même langage, le résultat de la Proposition 3.6.1 peut se généraliser à n'importe quelle suite épisturmienne équilibrée. On obtient alors le résultat général suivant, qui est une preuve de la conjecture de Fraenkel réduite à la classe des suites épisturmiennes.

Théorème 3.6.2 *Soit s une suite épisturmienne équilibrée sur un alphabet fini $\mathcal{A} = \{1, 2, \ldots, k\}$, avec $k \geq 3$ et $f_i(s) \neq f_j(s)$ pour tout $i \neq j$: toutes les lettres ont des fréquences différentes. Alors, à permutation de lettres près et à décalage près, $s = [\text{Pal}(123\cdots k)]^\omega$.*

Exemple 3.6.3 Pour respectivement $k = 3, 4, 5$, on obtient

$$\begin{aligned} s &= (1213121)^\omega, \\ t &= (121312141213121)^\omega, \\ u &= (12131214121312151213121412131213121)^\omega. \end{aligned}$$

De plus, les suites

$$\begin{aligned} s' &= 3121(1213121)^\omega, \\ t' &= 213121(121312141213121)^\omega, \\ u' &= 12141213121512131214121312151213121412131213121(12131214121312151213121412131213121)^\omega \end{aligned}$$

sont aussi des suites épisturmiennes (non standards) équilibrées avec des fréquences de lettres différentes.

La conjecture de Fraenkel considère les suites bi-infinies. Étant donné que les suites satisfaisant la conjecture de Fraenkel sont toutes périodiques, il suffit de considérer les suites infinies et non pas bi-infinies.

Dans ce chapitre, nous avons réussi à caractériser les suites épisturmiennes équilibrées. Il serait intéressant de généraliser ce résultat en considérant la notion générale d'équilibre énoncée à la Remarque 3.5.7. Par ailleurs, remarquons que notre résultat fournit un nouvel angle d'attaque pour la conjecture de Fraenkel : il suffit de prouver que toute suite satisfaisant à la conjecture de Fraenkel est une suite épisturmienne.

Chapitre IV

SUPERPOSITION DE 2 MOTS DE CHRISTOFFEL

Rappelons que la conjecture de Fraenkel a d'abord été étudiée du point de vue de la théorie des nombres, en termes des suites de Beatty. Un cas particulier de la conjecture est de savoir sous quelles conditions deux suites de Beatty sont disjointes. Dans les années 1980-1990, les suites de Beatty ont fait l'objet d'une dizaine d'articles de Morikawa, entre autres (Morikawa, 1985a). Dans ce dernier article, l'auteur prouve le théorème suivant, qui fournit une condition nécessaire et suffisante pour que deux suites de Beatty soient disjointes.

Théorème. (Morikawa, 1985a) *Soient p_1, p_2, q_1, q_2 des entiers et posons $p = \text{pgcd}(p_1, p_2)$, $q = \text{pgcd}(q_1, q_2)$, $u_1 = q_1/q$, $u_2 = q_2/q$. Il existe β_1 et β_2 tels que les suites de Beatty $S_1 = \{\lfloor p_1 n/q_1 + \beta_1 \rfloor : n \in Z\}$ et $S_2 = \{\lfloor p_2 n/q_2 + \beta_2 \rfloor : n \in Z\}$ sont disjointes si et seulement s'il existe des entiers positifs x et y tels que*

$$xu_1 + yu_2 = p - 2u_1u_2(q-1).$$

Exemple 4.0.4 Soient $p_1 = p_2 = 10$, $q_1 = 3$ et $q_2 = 4$. Alors $p = 10$, $q = 1$, $u_1 = 3$, $u_2 = 4$. Il existe β_1 et β_2 tels que $S_1 = \{\lfloor 10n/3 + \beta_1 \rfloor : n \in \mathbb{Z}\}$ et $S_2 = \{\lfloor 10n/4 + \beta_2 \rfloor : n \in \mathbb{Z}\}$ sont disjointes si et seulement s'ils existent $x, y > 0$ tels que $3x + 4y = 10$. Il suffit de prendre $x = 2$ et $y = 1$ et on a l'égalité. En effet, avec $\beta_1 = 0$ et $\beta_2 = 3$, on a les suites de Beatty disjointes $S_1 = \{1, 4, 7\} + 10\mathbb{Z}$ et $S_2 = \{0, 3, 5, 8\} + 10\mathbb{Z}$.

20 ans plus tard, dans (Simpson, 2004), l'auteur donne une preuve différente de ce résultat et prouve de nouveaux résultats intermédiaires. Cet article s'avère contenir des propriétés in-

téressantes concernant les mots de Christoffel. En effet, en traduisant les résultats de Simpson en terme de mots de Christoffel, certaines caractéristiques de ces mots apparaissent naturellement. Dans ce chapitre, nous introduisons d'abord quelques définitions et résultats préliminaires. Après avoir traduit les principaux résultats de Simpson, nous allons plus loin en donnant le nombre de superpositions de deux mots de Christoffel superposables. Nous donnons aussi quelques détails passés sous silence dans certaines preuves de Simpson. Entre autres, la preuve du Théorème 4.5.1 présentée ici est la même que celle donnée dans (Simpson, 2004), mais pour laquelle plusieurs détails ont été précisés. Nous terminons le chapitre en parlant d'un problème relié au problème de la monnaie, dans lequel les mots de Christoffel apparaissent géométriquement.

La motivation principale de ce chapitre est de caractériser un cas particulier de la conjecture de Fraenkel. En effet, la conjecture en terme de mots énonce que pour un alphabet à k lettres, $k \geq 3$, il existe une unique suite équilibrée et ayant des fréquences de lettres toutes différentes, à permutation des lettres et à conjugaison près. Dans ce chapitre, nous montrons comment une suite satisfaisant les hypothèses de la conjecture de Fraenkel peut être obtenue par superposition de k puissances de mots de Christoffel. L'idée de ce chapitre est d'étudier un cas particulier : la superposition de deux mots (ou puissances) de mot de Christoffel.

Définition 4.0.5 Un mot $w \in \mathcal{A}^*$ est *circulairement équilibré* si le mot $w^2 = ww$ est équilibré.

Exemple 4.0.6 Le mot $u = 112121$ est équilibré, mais n'est pas circulairement équilibré. En effet, $111, 212 \in F(uu)$, mais $\big||111|_1 - |212|_1\big| > 1$.

Exemple 4.0.7 On peut facilement vérifier que le mot 112112 est équilibré et donc que $v = 112$ est circulairement équilibré.

Lemme 4.0.8 *Si $w \in \mathcal{A}^*$ est circulairement équilibré, alors pour tout $a \in \mathcal{A}$, la projection $\Pi_a(w)$ est aussi circulairement équilibrée.*

Preuve Comme w est circulairement équilibré, pour tous facteurs u, v de w^2 tels que $\lg(u) = \lg(v)$ et pour toute lettre $a \in \mathcal{A}$, on a $\big||u|_a - |v|_a\big| \leq 1$. Par ailleurs, on a $|\Pi_a(u)|_a = |u|_a$ et

$|\Pi_a(v)|_a = |v|_a$. Ainsi, pour tout facteur $\Pi_a(u)$ et $\Pi_a(v)$ de même longueur de $\Pi_a(w^2)$, on a $\big||\Pi_a(u)|_a - |\Pi_a(v)|_a\big| = \big||u|_a - |v|_a\big| \leq 1$. Comme $\Pi_a(w) \in \{a,x\}^*$, par complémentarité, on a aussi que $\big||\Pi_a(u)|_x - |\Pi_a(v)|_x\big| \leq 1$. Donc $\Pi_a(w)$ est circulairement équilibré. ■

Définition 4.0.9 Soient $u \in {}^{\omega}\{a,x\}^{\omega}$ et $v \in {}^{\omega}\{b,x\}^{\omega}$, deux mots bi-infinis. Soient A l'ensemble des positions des a dans u et B l'ensemble des positions des b dans v. On dit que u et v sont *superposables* si $A \cap B = \emptyset$. Autrement dit, les ensembles A et B forment une partition des entiers.

Exemple 4.0.10 Soient $u = {}^{\omega}(aaxaxx)^{\omega}$ et $v = {}^{\omega}(xxbxxx)^{\omega}$. Alors $A = \{0,1,3\} + 6\mathbb{Z}$ et $B = \{2\} + 6\mathbb{Z}$. Donc u et v sont superposables.

Définition 4.0.11 Soient u, v des mots finis. Soit A l'ensemble des positions des a dans $u \in \{a,x\}^*$ et soit B l'ensemble des positions des b dans $v \in \{b,x\}^*$. Supposons que $\lg(u) = n$ et $\lg(v) = m$. On dit que u et v sont *superposables* si et seulement s'il existe $k \in \mathbb{Z}$ tel que ${}^{\omega}u^{\omega}$ et $\sigma^k({}^{\omega}v^{\omega})$ (σ est la fonction de décalage introduite dans la Section 0.3) soient superposables, c'est-à-dire si

$$(A + n\mathbb{Z}) \cap (B - k + m\mathbb{Z}) = \emptyset.$$

Si $k = 0$, on dit alors que les mots u et v se *superposent exactement*.

Remarque 4.0.12 Dans la Définition 4.0.11, la condition $k \in \mathbb{Z}$ peut être remplacée par $k \in [0, \min\{m,n\} - 1]$. En effet, on peut facilement vérifier que s'il existe un décalage k hors de cet intervalle permettant la superposition, alors il existe $k' \in [0, \min\{m,n\} - 1]$ pour lequel il en est de même.

Exemple 4.0.13 Soient les mots $u = axxxax$ et $v = xbx$. Considérons ${}^{\omega}u^{\omega} = {}^{\omega}(axxxax)^{\omega}$ et $\sigma^{-1}({}^{\omega}v^{\omega}) = {}^{\omega}(xxb)^{\omega}$. Les positions des a dans ${}^{\omega}u^{\omega}$ sont données par l'ensemble $\{0,4\}+6\mathbb{Z}$ et les positions des b dans $\sigma^{-1}({}^{\omega}v^{\omega})$ sont données par l'ensemble $\{2\} + 3\mathbb{Z} = \{2,5\} + 6\mathbb{Z}$. Comme ces deux ensembles sont disjoints, u et v sont superposables et le décalage nécessaire est $k = -1$.

Exemple 4.0.14 Soient les mots $u = axax$ et $v = bxb$. Ces deux mots ne sont pas superposables. En effet, les positions des a dans ${}^{\omega}u^{\omega}$ sont données par l'ensemble $A = \{0, 2\} + 4\mathbb{Z}$ et les positions des b dans ${}^{\omega}v^{\omega}$ sont données par l'ensemble $B = \{0, 2\} + 3\mathbb{Z}$. Comme $A = \{0, 2\} + 4\mathbb{Z} = \{0, 2, 4, 6, 8, 10\} + 12\mathbb{Z}$ et que $B = \{0, 2\} + 3\mathbb{Z} = \{0, 2, 3, 5, 6, 8, 9, 11\} + 12\mathbb{Z}$, on conclut par inspection qu'il n'existe aucun k tel que $A \cap (B - k) = \emptyset$. Donc u et v ne sont pas superposables.

Lemme 4.0.15 *Soient u, v, A et B tels que décrits dans la Définition 4.0.11. Les mots $u \in \{a, x\}^*$ et $v \in \{b, x\}^*$ sont superposables et sont tels que*

$$(A + n\mathbb{Z}) \cap (B - k + m\mathbb{Z}) = \emptyset$$

si et seulement si u et $\gamma^k(v)$ se superposent exactement.

Preuve Rappelons que A est l'ensemble des positions des a dans u et que B est l'ensemble des positions des b dans v. Par définition, les mots u et v sont superposables si et seulement s'il existe $k \in \mathbb{Z}$ tel que $(A+n\mathbb{Z}) \cap (B-k+m\mathbb{Z}) = \emptyset$. Cette condition est satisfaite si et seulement si les positions des a dans ${}^{\omega}u^{\omega}$ et des b dans $\sigma^k({}^{\omega}v^{\omega})$ forment des ensembles disjoints. Il suffit donc de montrer que les positions des b dans ${}^{\omega}(\gamma^k(v))^{\omega}$ sont les mêmes que dans $\sigma^k({}^{\omega}v^{\omega})$. Ces deux mots ont la période m et donc, les positions des b seront les même si et seulement si ${}^{\omega}(\gamma^k(v))^{\omega} = \sigma^k({}^{\omega}v^{\omega}) \iff \gamma^k(v) = \sigma^k({}^{\omega}v^{\omega})[0, m-1]$. Par la définition du conjugueur γ, cette dernière condition est satisfaite. ■

Corollaire 4.0.16 *Un mot fini $w \in \{1, 2, \ldots, k\}^n$ circulairement équilibré et ayant des fréquences de lettres toutes différentes peut être obtenu par la superposition de k mots circulairement équilibrés $w_1 \in \{1, x\}^n, w_2 \in \{2, x\}^n, \ldots, w_k \in \{k, x\}^n$ tels que $|w_i|_i \neq |w_j|_j$ pour tout $i \neq j$.*

Preuve Découle directement du Lemme 4.0.8. Il suffit d'appliquer la projection $\Pi_a(w)$ pour toute lettre $a \in \mathcal{A}$.

Le Corollaire 4.0.16 motive l'étude de ce chapitre : tout mot satisfaisant aux conditions de la conjecture de Fraenkel peut être obtenu par la superposition de mots circulairement équilibrés,

autrement dit, par la superposition de conjugués de puissances de mots de Christoffel. Un mot satisfaisant aux conditions de la conjecture de Fraenkel pour un alphabet à k lettres est donc obtenu par la superposition de k mots circulairement équilibrés.

Exemple 4.0.17 a) Le mot 1213121 satisfait les conditions du Corollaire 4.0.16. Il peut être obtenu en superposant le mots circulairement équilibrés $1x1x1x1$, $x2xxx2x$ et $xxx3xxx$.

b) Le mot 312112131214121 est le résultat de la superposition des mots $3xxxxxx3xxxxxxx$, $x1x11x1x1x1x1x1$, $xx2xx2xxx2xxx2x$ et $xxxxxxxxxxx4xxx$, qui sont tous circulairement équilibrés et qui ont respectivement les fréquences de lettres $2/15$, $8/15$, $4/15$ et $1/15$.

Dans ce chapitre, nous nous intéressons d'abord à la superposition de deux mots circulairement équilibrés. Nous ne considérons que des mots primitifs et nous donnons une condition nécessaire et suffisante pour que deux mots primitifs u, v soient tels que u^ω et v^ω se superposent. De plus, puisqu'un mot de Christoffel est un mot circulairement équilibré, primitif et minimal selon l'ordre lexicographique dans sa classe de conjugaison, si u est primitif et circulairement équilibré, il existe un conjugué de u qui soit un mot de Christoffel. Ainsi, nous considérons les mots de Christoffel correspondant w_1 et w_2 et nous donnons un critère pour que w_1^ω et w_2^ω se superposent, à conjugaison près de w_2. Pour ce faire, nous utilisons les résultats de (Simpson, 2004) qui sont une extension du travail de (Morikawa, 1985a).

4.1 Quelques résultats de base

Le Lemme 4.1.1 est une traduction du Théorème 3 de (Simpson, 2004) en terme de mots de Christoffel. Ce résultat apparaît aussi dans (Berthé, de Luca et Reutenauer, 2007) sous une forme équivalente, mais en utilisant la dualité des mots de Christoffel.

Lemme 4.1.1 *(Simpson, 2004; Berthé, de Luca et Reutenauer, 2007) Soit le mot de Christoffel $C(n,\alpha) \in \{a < x\}^*$ et soit $\overline{\alpha}$ tel que $\alpha\overline{\alpha} \equiv -1 \mod n$. Alors les positions modulo n des a dans $C(n,\alpha)$ sont données par l'ensemble $\{0, \overline{\alpha}, 2\overline{\alpha}, \ldots, (\alpha-1)\overline{\alpha}\}$.*

Preuve Posons $i \equiv j\overline{\alpha} \mod n$ et $\alpha + \beta = n$. Comme $\alpha \perp n$, alors son opposé inverse $\overline{\alpha}$ est aussi tel que $\overline{\alpha} \perp n$. Donc $\overline{\alpha}$ est un générateur du groupe cyclique $\mathbb{Z}/n\mathbb{Z}$ et alors

$$\{j\overline{\alpha}\}_{0 \leq j < n} = \{i\}_{0 \leq i < n}.$$

Soit $i \in [0, n-1]$. Nous pouvons donc écrire $i = j\overline{\alpha}$, avec $j \in [0, n-1]$. Ainsi,

$$i\beta \equiv j\overline{\alpha}\beta \equiv j\overline{\alpha}(n-\alpha) \equiv -j\overline{\alpha}\alpha \equiv j \mod n.$$

Par la Définition 2.2.3 d'un mot de Christoffel, $C(n,\alpha)[i] = a$ si et seulement si $(i+1)\beta \mod n > i\beta \mod n$. Mais cette condition est satisfaite si et seulement si $i\beta \mod n \in [0, \alpha - 1]$. Ainsi, $i\beta \mod n \in [0, \alpha-1]$ si et seulement si $j \in [0, \alpha-1]$, ce qui prouve le lemme. ■

Corollaire 4.1.2 *Soit $C(nq, \alpha q) = (C(n,\alpha))^q$ avec $n \perp \alpha$, satisfaisant aux mêmes hypothèses que dans le Lemme 4.1.1. Alors les positions modulo nq des a dans $C(nq, \alpha q)$ sont données par*

$$\bigcup_{i=0}^{q-1} \{0, \overline{\alpha}, 2\overline{\alpha}, \ldots, (\alpha-1)\overline{\alpha}\} + in.$$

Preuve Découle directement du Lemme 4.1.1. ■

4.2 Premiers résultats concernant la superposition de deux mots de Christoffel

Théorème 4.2.1 (Th. du reste chinois de (Simpson, 2004)) *Soient $C(n,1) \in \{a < x\}^*$ et $C(m,1) \in \{b < x\}^*$, deux mots de Christoffel. Alors $C(n,1)$ et $C(m,1)$ se superposent si et seulement si $n \not\perp m$.*

Preuve Par la Définition 4.0.11, les mots $C(n,1)$ et $C(m,1)$ sont superposables si et seulement si les mots $(C(n,1))^m = C(nm,m)$ et $(C(m,1))^n = C(nm,n)$ le sont. Les positions des a dans $C(nm,m)$ sont $A = \{0, n, 2n, \ldots, (m-1)\, n\}$ et celles des b dans $C(nm,n)$ sont $B = \{0, m, 2m, \ldots, (n-1)m\}$.

($\Longrightarrow$) Supposons que $n \perp m$ et que les mots $C(nm,m)$ et $C(nm,n)$ se superposent avec un décalage k. Ainsi, les positions des b dans $C(nm,n)$ suite au décalage k sont données par l'ensemble $B' = \{-k, m-k, \ldots, (n-1)m - k\}$. Comme $n \perp m$, par le Théorème de Bezout,

il existe $0 \leq \ell < n$ tel que $\ell m - k \equiv 0 \mod n$. Il y a donc conflit à la ℓ-ième position de b, puisqu'à cette position, il y a un a dans $C(nm, m)$: contradiction.
($\Longleftarrow$) Supposons maintenant que pgcd$(n, m) = d > 1$. Écrivons $n = dp$ et $m = dq$. Pour tout $0 \leq i < m$, $in + 1 = idp + 1 \not\equiv 0 \mod dq$. Donc $C(n, 1)$ et $C(m, 1)$ sont superposables et un des décalages possibles est $k = 1$. ■

Lemme 4.2.2 *Soit $C(n, \alpha) \in \{a < x\}^*$. Pour toute position i d'un a dans $\widetilde{C}(n, \alpha)$, il existe $j \in \mathbb{N}$ tel que*

$$i\alpha < jn \leq (i + 1)\alpha.$$

Preuve Par le Lemme 2.2.8, on sait que $\widetilde{C}(n, \alpha) = C(n, n - \alpha) \in \{x < a\}^*$. En utilisant la généralisation de la Définition 2.2.3 aux puissances de mots de Christoffel en remplaçant respectivement a, x, α et β par $x, a, n - \alpha$ et α, on trouve que $\widetilde{C}(n, \alpha)[i] = a$ si et seulement si $(i+1)\alpha \mod n \leq i\alpha \mod n$. Donc i est la position d'une lettre a si et seulement si $(i+1)\alpha \mod n \leq i\alpha \mod n$. Mais $(i + 1)\alpha \mod n \leq i\alpha \mod n$ si et seulement s'il existe un multiple de n entre $i\alpha$ et $(i + 1)\alpha$ inclusivement. On conclut en remarquant que cette condition est satisfaite si et seulement si $i\alpha < jn \leq (i + 1)\alpha$. ■

Lemme 4.2.3 *Soient $C(n, \alpha) \in \{a < x\}^*$ et $C(n, \beta) \in \{b < x\}^*$, des mots ou des puissances de mots de Christoffel de même longueur. Si $\alpha|\beta$, alors l'ensemble des positions des a dans $C(n, \alpha)$ est un sous-ensemble de l'ensemble des positions des b dans $C(n, \beta)$.*

Preuve Raisonnons sur les mots miroirs. Comme $\alpha|\beta$, on écrit $\beta = q\alpha$, $q \in \mathbb{N}$. Soit i la position d'un a dans $\widetilde{C}(n, \alpha)$. Alors par le Lemme 4.2.2, il existe $j \in \mathbb{N}$ tel que $i\alpha < jn \leq (i + 1)\alpha$. En multipliant chacun des membres de l'inéquation par $\frac{\beta}{\alpha} = \frac{q\alpha}{\alpha} = q \in \mathbb{N}$, on trouve

$$i(q\alpha) < (jq)n \leq (i + 1)(q\alpha)$$

$$i\beta < (jq)n \leq (i + 1)\beta,$$

avec $jq \in \mathbb{N}$. Donc i est aussi la position d'un b dans $\widetilde{C}(n, \beta)$. ■

Théorème 4.2.4 (Th. 2 de (Simpson, 2004)) *Soient deux mots de Christoffel $C(n,\alpha) \in \{a < x\}^*$, $C(m,\beta) \in \{b < x\}^*$ et posons $p = \text{pgcd}(m,n)$. Alors $C(n,\alpha)$ et $\gamma^k C(m,\beta)$ sont superposables exactement si et seulement si $C(p,\alpha)$ et $\gamma^k C(p,\beta)$ le sont aussi.*

Preuve ($\Longleftarrow$) Soit A (resp. A'), l'ensemble des positions des a dans $C(p,\alpha)$ (resp. dans $C(n,\alpha)$) et B (resp. B'), l'ensemble des positions des b dans $\gamma^k C(p,\beta)$ (resp. dans $\gamma^k C(m,\beta)$). Supposons que $C(p,\alpha)$ et $C(p,\beta)$ sont superposables avec un décalage k. On a donc que

$$A + p\mathbb{Z} \cap B - k + p\mathbb{Z} = \emptyset.$$

Comme $\text{pgcd}(m,n) = p$, posons $n = pi$. Alors $C(n,\alpha) = C(pi,\alpha)$. D'autre part, on a $(C(p,\alpha))^i = C(pi,\alpha i)$. Comme $C(pi,\alpha)$ et $C(pi,\alpha i)$ ont la même longueur et que $\alpha | \alpha i$, le Lemme 4.2.3 s'applique et on conclut que l'ensemble des positions des a dans $C(pi,\alpha)$ est inclus dans l'ensemble des positions des a dans $C(pi,\alpha i) = (C(p,\alpha))^i$. Cela implique que l'ensemble des positions des a dans ${}^{\omega}C(pi,\alpha)^{\omega}$ est inclus dans l'ensemble des positions des a dans ${}^{\omega}(C(p,\alpha))^{\omega}$. Autrement dit, $A' + n\mathbb{Z} \subseteq A + p\mathbb{Z}$. De façon semblable, on montre que $B' + m\mathbb{Z} \subseteq B + p\mathbb{Z}$. Comme $A + p\mathbb{Z} \cap B - k + p\mathbb{Z} = \emptyset$, il en découle que $A' + p\mathbb{Z} \cap B' - k + p\mathbb{Z} = \emptyset$, donc que $C(n,\alpha)$ et $\gamma^k C(m,\beta)$ sont superposables.

($\Longrightarrow$) (Par l'absurde) Supposons que $C(n,\alpha)$ et $\gamma^k C(m,\beta)$ sont superposables. Comme $p = \text{pgcd}(m,n)$, posons $n = pi$ et $m = pj$, avec $i \perp j$. Comme $\alpha \perp n$ (resp. $\beta \perp m$), on a $\alpha \perp i$ (resp. $\beta \perp j$). Supposons que $C(p,\alpha)$ et $\gamma^k C(p,\beta)$ ne se superposent pas. Il existe donc r, ℓ tels que la r-ième occurrence de a dans $C(p,\alpha)$ apparaisse à la même position que la ℓ-ième occurrence de b dans $\gamma^k C(p,\beta)$. En terme de suites de Beatty (voir Chapitre III), cela se traduit par

$$\left\lfloor \frac{p}{\alpha} r \right\rfloor = \left\lfloor \frac{p}{\beta} \ell \right\rfloor - k.$$

Cela implique que pour $s \in \mathbb{N}$,

$$\left\lfloor \frac{p}{\alpha}(r + s\alpha) \right\rfloor = \left\lfloor \frac{p}{\beta}(\ell + s\beta) \right\rfloor - k. \tag{4.1}$$

Il reste à montrer qu'il existe un s' tel que

$$i|(r + s'\alpha) \qquad \text{et} \qquad j|(\ell + s'\beta). \tag{4.2}$$

En effet, on aurait alors que les positions données en (4.1) du membre de gauche et du membre de droite sont respectivement des positions des mots $C(pi, \alpha) = C(n, \alpha)$ et $\gamma^k C(pj, \beta) = \gamma^k C(m, \beta)$. D'où une contradiction, comme nous avions supposé $C(n, \alpha)$ et $\gamma^k C(m, \beta)$ superposables.

Comme $i \perp j$, $G = \mathbb{Z}/i\mathbb{Z} \times \mathbb{Z}/j\mathbb{Z}$ est un groupe cyclique. Ce groupe est engendré par (α, β). En effet, si $s(\alpha, \beta) = 0$ dans G, alors $s\alpha \equiv 0 \mod i$ et $s\beta \equiv 0 \mod j$, et donc $i|s$ et $j|s$, et donc, $ij|s$. Donc l'ordre de (α, β) est ij. Ainsi, $\forall (n_1, n_2) \in G$, $\exists t$ tel que $(n_1, n_2) = t(\alpha, \beta)$. Autrement dit, pour tout $n_1, n_2 \in \mathbb{Z}$, $\exists t \in \mathbb{Z}$ tel que $n_1 \equiv t\alpha \mod i$ et $n_2 \equiv t\beta \mod j$. En prenant $n_1 = r$, $n_2 = \ell$ et $t = -s'$, on conclut. ■

Corollaire 4.2.5 *Si les mots de Christoffel $C(n, \alpha)$ et $C(m, \beta)$ se superposent, alors $m \not\perp n$ et $\alpha + \beta \leq p$, où $p = \text{pgcd}(m, n)$.*

Preuve Posons $p = \text{pgcd}(m, n)$. Par le Théorème 4.2.4, on sait que $C(n, \alpha)$ et $C(m, \beta)$ sont superposables si et seulement si $C(p, \alpha)$ et $C(p, \beta)$ le sont. Cela implique que si $C(n, \alpha)$ et $C(m, \beta)$ sont superposables, alors $\alpha + \beta \leq p$. Et comme $\alpha, \beta > 0$, on a $1 < \alpha + \beta \leq p$, et donc, $m \not\perp n$. ■

4.3 Superposition de mots de Christoffel de même longueur

À partir de maintenant, nous ne considérons que des mots de Christoffel de même longueur, puisque le Théorème 4.2.4 nous permet ensuite de généraliser nos résultats. Soient $C(n, \alpha)$ et $C(n, \beta)$. Alors on étudie deux cas.

4.3.1 Cas particulier : si $\alpha|\beta$

Dans cette section, nous étudions la superposition des mots de même longueur avec $\alpha|\beta$. Nous montrons une propriété nécessaire et suffisante pour que de tels mots de Christoffel soient superposables (Théorème 4.3.4). Ensuite, nous donnons un critère que doit satisfaire le décalage

pour permettre la superposition (Corollaire 4.3.3), puis nous explicitons un décalage qui sera toujours valide (Corollaire 4.3.5). Pour terminer la section, nous montrons de quelle façon un mot de Christoffel peut être vu comme la superposition de plusieurs mots de Christoffel (Théorème 4.3.6). Ces propriétés nous seront utiles pour prouver le résultat principal de la prochaine section : un critère nécessaire et suffisant pour que deux mots de Christoffel quelconques de même longueur soient superposables.

Lemme 4.3.1 *Soit le mot de Christoffel* $C(n,\alpha) \in \{a < x\}^*$. *Si* $\alpha\overline{\alpha} \equiv -1 \mod n$, *alors* $C(n,\alpha) = \gamma^{\overline{\alpha}}\widetilde{C}(n,\alpha)$.

Preuve Par le Lemme 4.1.1, on sait que les positions modulo n des a dans $C(n,\alpha)$ sont données par l'ensemble

$$A = \{0, \overline{\alpha}, 2\overline{\alpha}, \ldots, (\alpha-1)\overline{\alpha}\}.$$

D'autre part, les positions modulo n des a dans le mot miroir $\widetilde{C}(n,\alpha)$ sont données par

$$\begin{aligned}\widetilde{A} &= \{n-1, n-1-\overline{\alpha}, n-1-2\overline{\alpha}, \ldots, n-1-(\alpha-1)\overline{\alpha}\} \\ &= \{-1, -1-\overline{\alpha}, -1-2\overline{\alpha}, \ldots, -1-(\alpha-1)\overline{\alpha}\}.\end{aligned}$$

Finalement, en prenant le conjugué $\gamma^{\overline{\alpha}}\widetilde{C}(n,\alpha)$, on trouve que les positions des a sont données par

$$\begin{aligned}\gamma^{\overline{\alpha}}\widetilde{A} &= \{-1-\overline{\alpha}, -1-\overline{\alpha}-\overline{\alpha}, -1-2\overline{\alpha}-\overline{\alpha}, \ldots, -1-(\alpha-1)\overline{\alpha}-\overline{\alpha}\} \\ &= \{-1-\overline{\alpha}, -1-2\overline{\alpha}, -1-3\overline{\alpha}, \ldots, -1-\alpha\overline{\alpha}\} \\ &= \{(\alpha-1)\overline{\alpha}, (\alpha-2)\overline{\alpha}, \ldots, , \overline{\alpha}, 0\} = A.\end{aligned}$$

■

Définition 4.3.2 Soient $I = [a,b]$ et $I' = [c,d]$, deux intervalles d'entiers. On dit que I est *à gauche* de I' si $a < c$.

Lemme 4.3.3 *Soient* $C(n,\alpha) \in \{a < x\}^*$ *et* $C(n,\beta) \in \{b < x\}^*$, *deux mots de Christoffel de même longueur avec* $\beta = q\alpha$ *et* $q \in \mathbb{N}$. *Soit* $\ell \in [0, n-1]$. *Les conditions suivantes sont équivalentes.*

i) $C(n,\alpha)$ *et* $\gamma^{\ell\overline{\beta}}C(n,\beta)$ *se superposent exactement.*

ii) $\ell \notin [-(\alpha-1)q, \beta[\mod n$.

iii) $C(n,\alpha)$ *et* $\gamma^{\overline{\beta}(1+\ell)}\widetilde{C}(n,\beta)$ *se superposent exactement.*

Preuve Montrons d'abord que les différences entre les positions des a du mot $C(n,\alpha)$ et celles des b dans $C(n,\beta)$ forment un intervalle entier de cardinalité $(2\alpha-1)q$. Les positions des a dans $C(n,\alpha)$ sont $\{0,\overline{\alpha},\ldots,(\alpha-1)\overline{\alpha}\}$ et celles des b dans $C(n,\beta)$ sont $\{0,\overline{\beta},\ldots,(\beta-1)\overline{\beta}\}$, où $\alpha\overline{\alpha} \equiv -1 \mod n$ et $\beta\overline{\beta} \equiv -1 \mod n$. On a $i\overline{\alpha} \equiv iq\overline{\beta} \mod n$. En effet, comme $q\alpha = \beta$, en multipliant de chaque côté par $\overline{\alpha}\overline{\beta}$, on obtient $\overline{\alpha} = q\overline{\beta}$. Donc les différences entre les positions des lettres forment l'ensemble

$$E = \{j\overline{\beta} - i\overline{\alpha}\}_{\substack{0\le j<\beta \\ 0\le i<\alpha}} = \{(j-iq)\overline{\beta}\}_{\substack{0\le j<\beta \\ 0\le i<\alpha}}. \tag{4.3}$$

Regardons les nombres $j - iq$. Pour $i = 0$, cela correspond à l'intervalle $[0,\beta[$, pour $i = 1$, à l'intervalle $[-q,\beta-q[$, et de manière générale, cela correspond à l'intervalle $[-iq,\beta-iq[$. Comme $q > 0$, on remarque que les intervalles se déplacent vers la gauche : l'intervalle $[-iq,\beta-iq[$ est à gauche de l'intervalle $[-(i-1)q,\beta-(i-1)q[$.

On a $\beta = q\alpha \implies \beta \ge q \implies \beta - iq \ge q - iq = -(i-1)q$. Donc l'union de deux intervalles consécutifs est aussi un intervalle et alors, l'union de ces α intervalles forme l'intervalle $[-(\alpha-1)q,\beta[$, qui a la cardinalité

$$\beta - (-(\alpha-1)q) = \beta + \alpha q - q = \alpha q + \alpha q - q = (2\alpha-1)q. \tag{4.4}$$

$C(n,\alpha)$ et $\gamma^k C(n,\beta)$ se superposent exactement si et seulement si le décalage k ne fait pas partie de l'ensemble E, c'est-à-dire si et seulement s'il existe $\ell \notin [-(\alpha-1)q,\beta[$ tel que $k = \ell\overline{\beta}$. Donc il existe $\ell \notin [-(\alpha-1)q,\beta[$ si et seulement si $C(n,\alpha)$ et $\gamma^{\ell\overline{\beta}}C(n,\beta)$ se superposent exactement. D'où i) $\iff$ ii). Par ailleurs, par le Lemme 4.3.1, on sait que $C(n,\beta) = \gamma^{\overline{\beta}}\widetilde{C}(n,\beta)$. En remplaçant $C(n,\beta)$ par cette valeur dans $\gamma^{\ell\overline{\beta}}C(n,\beta)$, on obtient

$$\gamma^{\ell\overline{\beta}}C(n,\beta) = \gamma^{\ell\overline{\beta}}\gamma^{\overline{\beta}}\widetilde{C}(n,\beta) = \gamma^{\overline{\beta}(\ell+1)}\widetilde{C}(n,\beta).$$

D'où i) $\iff$ iii). ∎

Corollaire 4.3.4 (Th. 4 et Cor. 5 de (Simpson, 2004)) *Soient* $C(n,\alpha) \in \{a < x\}^*$ *et* $C(n,\beta) \in \{b < x\}^*$ *tels que* $\beta = q\alpha$, $q \in \mathbb{N}$. *Alors* $C(n,\alpha)$ *et* $C(n,\beta)$ *se superposent si et seulement si* $(2\alpha-1)q < n$.

Preuve Par l'équation (4.4), on sait que Card$(E) = (2\alpha-1)q$. D'autre part, les mots $C(n,\alpha)$ et $C(n,\beta)$ se superposent si et seulement s'il existe un décalage $0 \leq k < n$ tel que les positions des α occurrences de a dans $C(n,\alpha)$ forment un ensemble disjoint de l'ensemble des positions des β occurrences de b dans $C(n,\beta)$. Un tel décalage k existe si et seulement si l'ensemble E (de l'équation (4.3)) est de cardinalité au plus $n-1$. D'où la condition nécessaire et suffisante $(2\alpha-1)q \leq n-1 < n$. ■

Du Lemme 4.3.3, il est aussi possible de déduire un décalage qui sera toujours valide pour deux mots de Christoffel superposables de même longueur, avec $\alpha|\beta$.

Corollaire 4.3.5 *Soit* $C(n,\alpha)$ *et* $C(n,\beta)$, *deux mots de Christoffel superposables tels que* $\beta = q\alpha$, $q \in \mathbb{N}$. *Alors* $C(n,\alpha)$ *et* $\gamma^{(1-r)}\widetilde{C}(n,\beta)$ *se superposent exactement, où* $\alpha r \equiv 1 \mod n$.

Preuve Par le Corollaire 4.3.3, il suffit de montrer qu'il existe $\ell \notin [-(\alpha-1)q, \beta[\mod n$ tel que

$$\overline{\beta}(1+\ell) \equiv 1-r \mod n.$$

En multipliant de chaque côté par β, on obtient

$$-1-\ell \equiv \beta - \beta r \mod n$$

$$\Longleftrightarrow$$

$$\ell \equiv \beta r - \beta - 1 \mod n.$$

Il suffit donc de montrer que modulo n, on a

$$\beta r - \beta - 1 \notin [-(\alpha-1)q, \beta[= [-\beta+q, \beta[$$

$$\Longleftrightarrow$$

$$\beta r \notin [q+1, 2\beta+1[.$$

Or, $\beta r \equiv q \mod n$, car $\beta = q\alpha$ et $\alpha r \equiv 1 \mod n$. Il suffit donc de montrer que

$$q + n \geq 2\beta + 1 \iff n \geq 2\alpha q + 1 - q = (2\alpha - 1)q + 1.$$

On conclut en utilisant le Corollaire 4.3.4. ■

Théorème 4.3.6 (Th. 6 de (Simpson, 2004)) *Soit le mot de Christoffel $C(n, q\alpha) \in \{a < x\}^*$. Alors l'ensemble des positions des a dans $C(n, q\alpha)$ est la réunion des ensembles*

$$\{0, \overline{\alpha}, ..., (\alpha - 1)\overline{\alpha}\} + k\overline{q\alpha},$$

pour $0 \leq k < q$. De plus, le mot de Christoffel $C(n, q\alpha)$ est le résultat de la superposition exacte des q conjugués de $C(n, \alpha)$ suivants :

$$C(n, \alpha), \gamma^{-\overline{q\alpha}} C(n, \alpha), \ldots, \gamma^{-(q-1)\overline{q\alpha}} C(n, \alpha).$$

Preuve Par le Lemme 4.1.1, l'ensemble des positions des a dans $C(n, q\alpha)$ est

$$\{0, \overline{q\alpha}, 2\overline{q\alpha}, \ldots, (q\alpha - 1)\overline{q\alpha}\} = \bigcup_{j=0}^{q\alpha-1} j\overline{q\alpha}.$$

En séparant les positions selon leur résidu modulo q, on obtient

$$\bigcup_{j=0}^{q\alpha-1} j\overline{q\alpha} = \bigcup_{k=0}^{q-1} \bigcup_{i=0}^{\alpha-1} (iq + k)\overline{q\alpha}. \tag{4.5}$$

Comme $q\alpha\overline{q\alpha} \equiv -1 \mod n$, on a $q\overline{q\alpha} \equiv \overline{\alpha} \mod n$. Ainsi, en remplaçant $q\overline{q\alpha}$ par $\overline{\alpha}$ dans l'équation (4.5), on obtient

$$\bigcup_{j=0}^{q\alpha-1} j\overline{q\alpha} = \bigcup_{k=0}^{q-1} \bigcup_{i=0}^{\alpha-1} i\overline{\alpha} - k\overline{q\alpha}. \tag{4.6}$$

On peut alors réécrire le membre de droite de l'équation (4.6) comme

$$\bigcup_{k=0}^{q-1} \{0, \overline{\alpha}, 2\overline{\alpha}, \ldots, (\alpha - 1)\overline{\alpha}\} + k\overline{q\alpha}.$$

De plus, les ensembles

$$\{0, \overline{\alpha}, \ldots, (\alpha-1)\overline{\alpha}\},$$

$$\{\overline{q\alpha}, \overline{q\alpha}+\overline{\alpha}, \ldots, \overline{q\alpha}+(\alpha-1)\overline{\alpha}\},$$

$$\ldots,$$

$$\{(q-1)\overline{q\alpha}, (q-1)\overline{q\alpha}+\overline{\alpha}, \ldots, (q-1)\overline{q\alpha}+(\alpha-1)\overline{\alpha}\}$$

correspondent respectivement aux positions des a dans les mots de Christoffel $C(n,\alpha)$, $\gamma^{-\overline{q\alpha}}C(n,\alpha)$, $\ldots, \gamma^{-(q-1)\overline{q\alpha}}C(n,\alpha)$. ■

4.3.2 Cas général

Dans cette section, nous traitons le cas général de superposition de deux mots de Christoffel de même longueur. Pour ce faire, nous considérons les mots de Christoffel écrits sous la forme $C(n, q\alpha)$ et $C(n, q\beta)$, avec $\alpha \perp \beta$ et $q \in \mathbb{N}$.

Notation 4.3.7 Pour $0 \leq i < \alpha$, on note V_i l'intervalle entier

$$V_i = [(-q+1)\beta, q\beta - 1] + i\overline{\alpha}\beta.$$

Proposition 4.3.8 *Les mots de Christoffel $C(n, q\alpha) \in \{a < x\}^*$ et $C(n, q\beta) \in \{b < x\}^*$ avec $\alpha \perp \beta$ se superposent si et seulement si l'union*

$$\bigcup_{i=0}^{\alpha-1} V_i \tag{4.7}$$

n'est pas un système complet de résidus modulo n.

Preuve Par le Théorème 4.3.6, en interchangeant q et α, on trouve que $C(n, q\alpha)$ est la superposition exacte des α conjugués $C(n,q), \gamma^{-\overline{q\alpha}}C(n,q), \ldots, \gamma^{-(\alpha-1)\overline{q\alpha}}C(n,q)$. On peut donc écrire que l'ensemble des positions des a dans $C(n, q\alpha)$ est $\bigcup_{i=0}^{\alpha-1} \mathrm{pos}_a(\gamma^{-i\overline{q\alpha}}C(n,q))$, où pos_a désigne les positions des lettres a. De plus, par le Lemme 4.3.3, en remplaçant α, q et β par respectivement q, β et $q\beta$, on trouve que $C(n,q)$ et $\gamma^{\ell\overline{q\beta}}C(n, q\beta)$ se superposent exactement si et seulement s'il existe $\ell \notin [-(q-1)\beta, q\beta[\mod n$. De manière plus générale, $\gamma^{-i\overline{q\alpha}}C(n,q)$

et $\gamma^{\ell\overline{q\beta}}C(n,q\beta)$ se superposent exactement si et seulement si $C(n,q)$ et $\gamma^{\ell\overline{q\beta}+i\overline{q\alpha}}C(n,q\beta)$ se superposent exactement. Pour retrouver la forme du Lemme 4.3.3 iii), réécrivons $\ell\overline{q\beta}+i\overline{q\alpha}$. On a

$$\begin{aligned}\ell\overline{q\beta}+i\overline{q\alpha} &= \ell\overline{q\beta}-\overline{q\beta}q\beta i\overline{q\alpha}\\ &= \overline{q\beta}(\ell+q\beta i\overline{q}\,\overline{\alpha})\\ &= \overline{q\beta}(\ell-i\overline{\alpha}\beta).\end{aligned}$$

On a maintenant la forme du Lemme 4.3.3 iii) et on peut dire que $\gamma^{\ell\overline{q\beta}+i\overline{q\alpha}}C(n,q\beta)$ et $C(n,q)$ se superposent exactement si et seulement s'il existe $\ell-i\overline{\alpha}\beta\notin[-(q-1)\beta,q\beta[\mod n$. Cela est équivalent à dire qu'il existe un $\ell\notin[-(q-1)\beta,q\beta[+i\overline{\alpha}\beta=V_i$. Mais on doit avoir que $\ell\notin V_i$ pour tout $0\leq i<\alpha$. Ainsi, les mots $C(n,q\alpha)$ et $C(n,q\beta)$ se superposent si $\bigcup_{i=0}^{\alpha-1}V_i$ n'est pas un système complet de résidus modulo n. ■

Corollaire 4.3.9 *Il existe* $\ell\notin\bigcup_{i=0}^{\alpha-1}V_i \mod n$*, un élément hors de l'union des intervalles, si et seulement si* $C(n,q\alpha)$ *et* $\gamma^{(\ell+1)\overline{q\beta}}\widetilde{C}(n,q\beta)$ *se superposent exactement.*

Preuve On sait par la Proposition 4.3.8 que les mots de Christoffel $C(n,q\alpha)$ et $C(n,q\beta)$ se superposent si et seulement si l'union des V_i n'est pas un système complet de résidus modulo n. Dans la preuve, on montre qu'il doit exister un $\ell\notin[-(q-1)\beta,q\beta[+i\overline{\alpha}\beta$ pour $0\leq i<\alpha$ et qu'alors, tous les mots $\gamma^{-i\overline{q\alpha}}C(n,q)$ se superposent exactement avec le mot $\gamma^{\ell\overline{q\beta}}C(n,q\beta)$. Comme $C(n,q\alpha)$ est le résultat de la superposition exacte des α conjugués de $C(n,q)$ suivants

$$C(n,q),\gamma^{-\overline{q\alpha}}C(n,q),\ldots,\gamma^{-(\alpha-1)\overline{q\alpha}}C(n,q),$$

on a que $C(n,q\alpha)$ et $\gamma^{\ell\overline{q\beta}}C(n,q\beta)$ se superposent exactement si et seulement s'il existe $\ell\notin[-(q-1)\beta,q\beta[+i\overline{\alpha}\beta$ pour $0\leq i<\alpha$. Par le Lemme 4.3.1, on sait que $C(n,q\beta)=\gamma^{\overline{q\beta}}\widetilde{C}(n,q\beta)$. Ainsi, on a que $C(n,q\alpha)$ et $\gamma^{\ell\overline{q\beta}}\gamma^{\overline{q\beta}}\widetilde{C}(n,q\beta)=\gamma^{(\ell+1)\overline{q\beta}}\widetilde{C}(n,q\beta)$ se superposent exactement. ■

Lemme 4.3.10 *Soient* $\alpha,\beta\in\mathbb{N}-\{0\}$*, avec* $\alpha\perp\beta$*. Soit l'équation*

$$x\alpha+y\beta = n-2\alpha\beta(q-1), \tag{4.8}$$

avec $q,\alpha,\beta\perp n$ *et* $q\geq 1$.

i) *L'équation (4.8) a toujours une solution* $x, y \in \mathbb{Z}$.

ii) *Elle a toujours une solution avec* $1 \leq y \leq \alpha$ *et cette solution est unique.*

iii) *Si l'équation (4.8) est satisfaite, alors* $\alpha \perp (\alpha - y)$.

Preuve Comme $\alpha \perp \beta$, en utilisant le théorème de Bezout, on trouve que l'équation (4.8) a toujours une solution $x, y \in \mathbb{Z}$. Supposons maintenant qu'il existe deux solutions telles que $1 \leq y, y' \leq \alpha$. On aurait alors $x\alpha + y\beta = x'\alpha + y'\beta$, et donc, $\alpha(x - x') = \beta(y' - y)$. Mais $\alpha \perp \beta$ implique que $\alpha|(y' - y)$. Comme $1 \leq y, y' \leq \alpha$, cela est impossible. De plus, en utilisant l'équation (4.8), on trouve que $\alpha \perp y$. En effet, on peut réécrire l'équation (4.8) comme

$$\alpha(x + 2\beta(q - 1)) = n - y\beta$$

et puisque $\alpha \perp n$, il en découle que $\alpha \perp y$. Finalement, $\alpha \perp (\alpha - y)$. ■

Posons $z = \alpha - y$. Soit $i \in [0, \alpha - 1]$, l'une des valeurs possibles de z, comme $z = \alpha - y$ et que $1 \leq y \leq \alpha$. Puisque $\alpha \perp z$, il existe un unique $r(i) \in \mathbb{N}$ tel que $i \equiv r(i)z \mod \alpha$. Pour $0 \leq r < \alpha$, posons

$$M(r) = r(x + (2q - 1)\beta) - \left\lfloor \frac{zr}{\alpha} \right\rfloor \beta. \tag{4.9}$$

Dans ce qui suit, nous utilisons les fonctions $r(i)$ et $M(r(i))$ pour obtenir un nouvel ordre pour les intervalles V_i.

Remarque 4.3.11 Soit $a = bq + r$, la division euclidienne de a par b, où $r < b$ et $a, b, q, r \in \mathbb{N}$. On a $r = a \mod b$ et $q = \left\lfloor \frac{a}{b} \right\rfloor$. Ainsi,

$$a = bq + r \iff a - bq - r = 0 \iff a - b\left\lfloor \frac{a}{b} \right\rfloor - (a \mod b) = 0.$$

Lemme 4.3.12 (Lemme 7 de (Simpson, 2004)) *Pour* $i \in [0, \alpha - 1]$,

$$M(r(i)) \equiv -i\overline{\alpha}\beta \mod n.$$

Preuve Considérons $\alpha(M(r(i)) + i\overline{\alpha}\beta)$, pour un i fixé. Dans ce qui suit, afin d'alléger la notation, nous écrirons r pour désigner $r(i)$. En utilisant l'équation (4.8) et la définition de $M(r)$, on obtient :

$$\begin{aligned}
\alpha\left(M(r) + i\overline{\alpha}\beta\right) &\equiv \alpha r\left(x + (2q-1)\beta\right) - \alpha\left\lfloor\frac{zr}{\alpha}\right\rfloor\beta + \alpha i\overline{\alpha}\beta \\
&\equiv r\left(x\alpha + 2\alpha\beta(q-1) + \alpha\beta\right) - \alpha\beta\left\lfloor\frac{zr}{\alpha}\right\rfloor - i\beta \\
&\equiv r(-y\beta + \alpha\beta) - \alpha\beta\left\lfloor\frac{zr}{\alpha}\right\rfloor - i\beta \\
&\equiv \beta\left((\alpha - y)r - \alpha\left\lfloor\frac{zr}{\alpha}\right\rfloor - i\right) \\
&\equiv \beta\left(zr - \alpha\left\lfloor\frac{zr}{\alpha}\right\rfloor - i\right) \mod n.
\end{aligned}$$

Il suffit de remplacer a et b dans la remarque précédente par respectivement zr et α, et de se rappeler que $i = zr \mod \alpha$. On a alors que le terme entre parenthèse vaut 0, et donc que $\alpha(M(r) + i\overline{\alpha}\beta) \equiv 0 \mod n$. Comme $\alpha \perp n$, on a $M(r) + i\overline{\alpha}\beta \equiv 0 \mod n$ et on conclut. ■

Les Lemmes 4.3.13, 4.3.14, 4.3.15, 4.3.16 et 4.3.17 ne sont pas des résultats originaux, puisqu'ils apparaissent de façon sous-entendue dans la preuve du Théorème 8 dans (Simpson, 2004). Nous remercions d'ailleurs l'auteur pour les correspondances éclairantes que nous avons eues avec lui.

Lemme 4.3.13 *Soient $I_0, I_1, \ldots, I_{r-1}$, r intervalles de même longueur et $n \in \mathbb{N}$ fixé, tels que*

i) $\max(I_0) - \min(I_{r-1}) \geq n - 1 \geq 1$;

ii) *pour $0 \leq j < r-1$, si I_{j+1} est à gauche de I_j, alors $I_{j+1} \cup I_j$ est un intervalle.*

Alors $\bigcup_{j=0}^{r-1} I_j$ est un système complet de résidus modulo n.

Preuve Supposons que l'intervalle I_{r-1} ne soit pas situé à gauche de I_0. Alors comme $\max(I_0) - \min(I_{r-1}) \geq n-1$, cela implique que $I_0 \cup I_{r-1}$ est un intervalle et que $I_0 \cap I_{r-1} = [\min(I_{r-1}), \max(I_0)]$. Il en découle que $\text{Card}(I_0 \cap I_{r-1}) \geq n$ et donc que $\bigcup_{j=0}^{r-1} I_j$ est un système complet de résidus modulo n. Supposons maintenant que l'intervalle I_{r-1} est situé à gauche de l'intervalle I_0. Par ii), il existe bien des intervalles consécutifs qui se déplacent vers la gauche. La condition ii) nous assure aussi que tous les entiers entre I_{r-1} et I_0 font partie de l'union

des j intervalles. Comme $\max(I_0) - \min(I_{r-1}) \geq n-1$, le nombre d'entiers compris entre le début de l'intervalle I_{r-1} et la fin de l'intervalle I_0 est au moins n. Dans tous les cas, $\bigcup_{j=0}^{r-1} I_j$ est bien un système complet de résidus modulo n. ■

Lemme 4.3.14 *Soient $I_0, I_1, \ldots, I_{r-1}$ des intervalles finis dans $\mathbb{Z}$ et soit I le plus petit intervalle qui les contient tous. On suppose que*

i) $I \setminus \bigcup_{j=0}^{r-1} I_j$ *n'est pas vide ;*

ii) *si* $x \in \bigcup_{j=0}^{r-1} I_j$ *et* $y \in I \setminus \bigcup_{j=0}^{r-1} I_j$, *alors* $|y-x| < n$.

Alors $\bigcup_{j=0}^{r-1} I_j$ *ne contient pas tous les entiers modulo* n.

Preuve Il existe $y \in I \setminus \bigcup_{j=0}^{r-1} I_j$. Pour un tel y, il n'existe pas de $x \in \bigcup_{j=0}^{r-1} I_j$ tel que $y \equiv x \mod n$ (ce qui démontre le lemme). En effet, on aurait $y - x = kn$, avec $k \in \mathbb{Z}$. Comme $y \notin \bigcup_{j=0}^{r-1} I_j$ et $x \in \bigcup_{j=0}^{r-1} I_j$, on a $k \neq 0$ (sinon $x = y$). Donc $|k| \geq 1 \implies |y-x| \geq n$, qui contredirait ii). ■

Lemme 4.3.15 *Soit $z = \alpha - y$ et $0 \leq r < \alpha$. Alors* $\left\lfloor \frac{z(r+1)}{\alpha} \right\rfloor - \left\lfloor \frac{zr}{\alpha} \right\rfloor \in \{0,1\}$.

Preuve Posons $zr = i\alpha + t$, avec $i \in \mathbb{N}$ et $0 \leq t < \alpha$. Alors

$$\begin{aligned} \left\lfloor \frac{z(r+1)}{\alpha} \right\rfloor - \left\lfloor \frac{zr}{\alpha} \right\rfloor &= \left\lfloor \frac{i\alpha + t + z}{\alpha} \right\rfloor - \left\lfloor \frac{i\alpha + t}{\alpha} \right\rfloor \\ &= i + \left\lfloor \frac{t+z}{\alpha} \right\rfloor - i - \left\lfloor \frac{t}{\alpha} \right\rfloor \\ &= \left\lfloor \frac{t+z}{\alpha} \right\rfloor - \left\lfloor \frac{t}{\alpha} \right\rfloor. \end{aligned}$$

Comme $1 \leq y \leq \alpha$ et $0 \leq t < \alpha$, on a $0 \leq t + z < 2\alpha$ et donc,

$$\left\lfloor \frac{t+z}{\alpha} \right\rfloor - \left\lfloor \frac{t}{\alpha} \right\rfloor = \left\lfloor \frac{t+z}{\alpha} \right\rfloor - 0 \leq 1.$$

■

Lemme 4.3.16 *Soit $M(r)$ tel que défini à l'équation (4.9). Alors*

i) $M(0) = 0$ *;*

ii) $M(\alpha - 1) = n - x - 2\beta(q-1) - i$, *où* $i = 0$ *si* $y \neq \alpha$ *et* $i = \beta$ *sinon.*

Preuve En utilisant la définition de $M(r)$, on obtient

i) $M(0) = 0(x + (2q-1)\beta) - \left\lfloor \frac{z \cdot 0}{\alpha} \right\rfloor \beta = 0.$

ii) Si $y \neq \alpha$, alors

$$\begin{aligned} M(\alpha-1) &= (\alpha-1)(x+(2q-1)\beta) - \left\lfloor \frac{z(\alpha-1)}{\alpha} \right\rfloor \beta && (4.10)\\ &= \alpha x + (2q-1)\alpha\beta - x - (2q-1)\beta - z\beta + \beta && (4.11)\\ &= \alpha x - z\beta + (2q-1)\alpha\beta - x + \beta - (2q-1)\beta \\ &= \alpha x + y\beta - \alpha\beta + (2q-1)\alpha\beta - x + \beta - (2q-1)\beta \\ &= n - 2\alpha\beta(q-1) + 2(q-1)\alpha\beta - x + \beta - (2q-1)\beta \\ &= n - x + \beta - (2q-1)\beta = n - x - 2\beta(q-1) && (4.12) \end{aligned}$$

Si $y = \alpha$, alors

$$M(\alpha - 1) = n - x - 2\beta(q-1) - \beta. \tag{4.13}$$

On passe de l'équation (4.10) à l'équation (4.11) en utilisant le fait que

$$\left\lfloor \frac{z(\alpha-1)}{\alpha} \right\rfloor \beta = z\beta + \left\lfloor \frac{-z}{\alpha} \right\rfloor \beta = z\beta - \beta,$$

comme $0 < z = \alpha - y < \alpha$, puisque $1 \leq y < \alpha$. Si $y = \alpha$, alors $z = 0$ et on a plutôt que

$$\left\lfloor \frac{z(\alpha-1)}{\alpha} \right\rfloor \beta = 0,$$

d'où l'équation (4.13). ∎

Lemme 4.3.17 *Soit $M(r)$ tel que défini à l'équation (4.9). Alors*

i) $M(r+1) - M(r) = x + (2q-1)\beta - \beta\left(\left\lfloor \frac{z(r+1)}{\alpha} \right\rfloor - \left\lfloor \frac{zr}{\alpha} \right\rfloor\right)$ *;*

ii) *si* $x \leq 0$, *alors* $M(r+1) - M(r) \leq \beta(2q-1)$.

Preuve On a $M(r+1) - M(r)$

$$\begin{aligned} &= \left((r+1)(x+(2q-1)\beta) - \left\lfloor \frac{z(r+1)}{\alpha} \right\rfloor \beta\right) - \left(r(x+(2q-1)\beta) - \left\lfloor \frac{zr}{\alpha} \right\rfloor \beta\right) \\ &= x + (2q-1)\beta - \beta\left(\left\lfloor \frac{z(r+1)}{\alpha} \right\rfloor - \left\lfloor \frac{zr}{\alpha} \right\rfloor\right), \end{aligned}$$

qui est $\leq \beta(2q-1)$ si $x \leq 0$. ■

Théorème 4.3.18 (Th. 8 de (Simpson, 2004)) *Soient les mots de Christoffel $C(n, q\alpha)$ et $C(n, q\beta)$ avec $\alpha \perp \beta$. Alors, $C(n, q\alpha)$ et $C(n, q\beta)$ sont superposables si et seulement s'il existe $x, y \in \mathbb{N} - \{0\}$ tels que*

$$x\alpha + y\beta = n - 2\alpha\beta(q-1). \tag{4.14}$$

Preuve Considérons les mots de Christoffel $C(n, q\alpha)$ et $C(n, q\beta)$. Par le Lemme 4.3.10, on sait qu'il existe x, y qui satisfont (4.14), avec $1 \leq y \leq \alpha$. On veut montrer que $C(n, q\alpha)$ et $C(n, q\beta)$ sont superposables si et seulement si $x > 0$.

($\Longrightarrow$) Supposons que $x \leq 0$. Considérons l'union des intervalles donnée à l'équation (4.7) :

$$\bigcup_{i=0}^{\alpha-1} \{[(-q+1)\beta, q\beta - 1] + i\overline{\alpha}\beta\}.$$

La Proposition 4.3.8 nous indique que $C(n, q\alpha)$ et $C(n, q\beta)$ sont superposables si et seulement si cette union n'est pas un système complet de résidus modulo n. Par le Lemme 4.3.12, on sait que cet ensemble est égal modulo n à l'union

$$\bigcup_{r=0}^{\alpha-1} \{[(-q+1)\beta, q\beta - 1] - M(r)\}. \tag{4.15}$$

Posons $I_r = [(-q+1)\beta, q\beta - 1] - M(r)$, pour $0 \leq r < \alpha$. Alors, $\max(I_0) = q\beta - 1 - M(0) = q\beta - 1$ et $\min(I_{\alpha-1}) = (-q+1)\beta - M(\alpha-1)$. Ainsi, on a

$$\begin{aligned} \max(I_0) - \min(I_{\alpha-1}) &= q\beta - 1 - ((-q+1)\beta - M(\alpha-1)) \\ &= q\beta - 1 + q\beta - \beta + n - x - 2\beta(q-1) - i \\ &= \beta + n - x - 1 - i \end{aligned}$$

où $i \in \{0, \beta\}$. Comme $x \leq 0$, $-x$ est non négatif. Donc, $\max(I_0) - \min(I_{\alpha-1}) \geq n-1$.

De plus, pour tout $0 \leq r < \alpha - 1$, $I_r \cup I_{r+1}$ est un intervalle. En effet, supposons que $I_r \cup I_{r+1}$ n'est pas un intervalle. Cela signifierait que $M(r+1) - M(r) > \text{lg}(I_r) + 1$. Donc il suffit que $M(r+1) - M(r) \leq \text{lg}(I_r) + 1$ pour que $I_r \cup I_{r+1}$ soit un intervalle. Par le Lemme 4.3.17, on a $M(r+1) - M(r) \leq \beta(2q-1)$.

Par ailleurs, tous les intervalles sont de longueur

$$\text{lg}(I_r) = q\beta - 1 - (-q+1)\beta = 2q\beta - \beta - 1 = \beta(2q-1) - 1.$$

On a donc bien que $M(r+1) - M(r) \leq \text{lg}(I_r) + 1$. En appliquant le Lemme 4.3.13, on conclut que si $x \leq 0$, les mots ne sont pas superposables.

($\Longleftarrow$) Supposons maintenant que $x > 0$ et montrons qu'alors, les mots se superposent. Il suffit donc de montrer que si $x > 0$, alors $\bigcup_{r=0}^{\alpha-1} \{[-(q-1)\beta, q\beta - 1] - M(r)\}$ ne contient pas tous les entiers résidus modulo n.

Rappelons que $I_r = [-(q-1)\beta, q\beta - 1] - M(r)$, pour $0 \leq r < \alpha$. Puisque $x > 0$ et $q \geq 1$, en utilisant les Lemmes 4.3.15 et 4.3.17, on a

$$\begin{aligned} M(r+1) - M(r) &= x + (2q-1)\beta - \beta\left(\left\lfloor \frac{z(r+1)}{\alpha} \right\rfloor - \left\lfloor \frac{zr}{\alpha} \right\rfloor\right) \\ &\geq x + (2q-1)\beta - \beta \\ &= x + 2q\beta - 2\beta \\ &= x + 2\beta(q-1) \geq x > 0. \end{aligned}$$

Donc les intervalles I_r se décalent vers la gauche, pour $0 \leq r < \alpha$. Ils ont tous la même cardinalité

$$\text{Card}(I_r) = \text{lg}(I_r) + 1 = \beta(2q-1) - 1 + 1 = \beta(2q-1).$$

Supposons que $I = \bigcup_{r=0}^{\alpha-1} I_r$ ne soit pas un intervalle. Alors la condition i) du Lemme 4.3.14 est satisfaite. Pour la condition ii), il suffit de prendre $y = \min(I_0) - 1$ (puisque $\max(I \setminus \cup I_j) \leq \min(I_0) - 1$) et $x = \min(I)$ et de vérifier que $y - x < n$. On a $x = -(q-1)\beta - M(\alpha - 1)$ et $y = -(q-1)\beta - 1$.

Donc

$$y-x = (-(q-1)\beta - 1) - (-(q-1)\beta - (n - x - 2\beta(q-1) - i)) = n - x - 1 - 2\beta(q-1) - i,$$

avec $i \in \{0, \beta\}$. Comme $x > 0$, ce nombre est bien $< n$. Par le Lemme 4.3.14, on conclut que la réunion de ces intervalles ne contient pas tous les entiers modulo n. ■

4.4 Nombre de superpositions de mots de Christoffel de même longueur

Les résultats de cette section n'apparaissent pas dans (Simpson, 2004). Ils se déduisent de cet article, mais ils n'y sont pas mentionnés.

Définition 4.4.1 Soient $C(n, \alpha)$ et $C(n, \beta)$, deux mots de Christoffel superposables. On définit le *nombre de superpositions* de ces deux mots comme

$$\text{Card}\left(\{k \in [0, n-1] \,|\, C(n,\alpha) \text{ et } \gamma^k C(n,\beta) \text{ sont exactement superposables}\}\right).$$

Dans cette section, nous donnons le nombre exact de superpositions entre deux mots de Christoffel de même longueur.

D'abord, quelques résultats sont préalables.

Corollaire 4.4.2 *(du Lemme 4.3.14 et de sa preuve) Si les deux conditions du Lemme 4.3.14 sont satisfaites, alors*

i) *les éléments de* $I \setminus \bigcup_{j=0}^{r-1} I_j$ *sont tous distincts modulo* n *;*

ii) *si* $\text{Card}(I) \geq n$*, alors modulo* n *les éléments de* $I \setminus \bigcup_{j=0}^{r-1} I_j$ *coïncident avec ceux de l'ensemble des éléments qui ne sont pas, modulo* n*, dans* $\bigcup_{j=0}^{r-1} I_j$ *;*

iii) *si* $\text{Card}(I) < n$*, alors modulo* n *les éléments de* $I \setminus \bigcup_{j=0}^{r-1} I_j$ *coïncident avec ceux de l'ensemble des éléments qui ne sont pas, modulo* n*, dans* $\bigcup_{j=0}^{r-1} I_j$ *auxquels on ajoute les éléments*

$$\{\min(I) - (n - \text{Card}(I)), \ldots, \min(I) - 2, \min(I) - 1\},$$

qui sont au nombre de $n - \text{Card}(I)$.

Preuve

i) Si $x, y \in I \setminus \bigcup_{j=0}^{r-1} I_j$, alors $|x - y| < n$. En effet, sans perte de généralité, supposons $y > x$. Alors $y \leq \max(I \setminus \bigcup_{j=0}^{r-1} I_j)$ et $x > \min(I)$, car ce dernier est dans $\bigcup_{j=0}^{r-1} I_j$. Donc $y - x < \max(I \setminus \bigcup_{j=0}^{r-1} I_j) - \min(I)$. Par le Lemme 4.3.14 ii), on a que

$$y - x < \max(I \setminus \bigcup_{j=0}^{r-1} I_j) - \min(I) < n,$$

donc que $y - x < n - 1$.

ii) I contient tous les éléments $\mod n$. Par ii) du Lemme 4.3.14, aucun élément de $\bigcup_{j=0}^{r-1} I_j$ n'est congru à un élément de $I \setminus \bigcup_{j=0}^{r-1} I_j$. D'où l'assertion.

iii) Les éléments rajoutés ne sont congrus à aucun élément dans I. D'où l'assertion. ∎

Lemme 4.4.3 *Soit le mot de Christoffel* $C(n, j) \in \{a < b\}^*$. *Alors*

$$C(n,j)[i] = \begin{cases} a & \text{si } \left\lfloor \frac{n-j}{n}(i+1) \right\rfloor - \left\lfloor \frac{n-j}{n} i \right\rfloor = 0 \\ b & \text{si } \left\lfloor \frac{n-j}{n}(i+1) \right\rfloor - \left\lfloor \frac{n-j}{n} i \right\rfloor = 1. \end{cases}$$

Preuve Découle du graphe de Cayley des mots de Christoffel. Faire la différence des parties entières correspond à vérifier si l'on a dépassé ou pas un multiple de n. Si la différence des parties entières est nulle, alors on n'a pas dépassé un multiple de n. ∎

Proposition 4.4.4 *Soient* $C(n, q\alpha)$ *et* $C(n, q\beta)$, *deux mots de Christoffel superposables avec* $y \leq \alpha$ *et* $\alpha \perp \beta$. *Le nombre de superpositions est*

i) xy, *si* $x \leq \beta$;

ii) $x\alpha + y\beta - \alpha\beta$, *si* $x > \beta$;

où x, y est la solution de l'équation (4.14).

Preuve Rappelons que selon le Théorème 4.3.18, si deux mots sont superposables, alors $x > 0$. Notons I le plus petit intervalle qui contient l'union des intervalles donné à l'équation (4.15). Calculons d'abord Card(I).

$$\begin{aligned}
\text{Card}(I) &= \max(I) - \min(I) + 1 \\
&= \max(I_0) - \min(I_{\alpha-1}) + 1 \\
&= (q\beta - 1) - ((-q+1)\beta - M(\alpha-1)) + 1 \\
&= q\beta - 1 + q\beta - \beta + (n - x - 2\beta(q-1) - i) + 1 \\
&= 2q\beta - \beta + n - x - 2\beta(q-1) - i \\
&= n - x + \beta - i,
\end{aligned}$$

où $i \in \{0, \beta\}$.

i) Supposons $x \leq \beta$ et $y \neq \alpha$. Alors Card$(I) = n - x + \beta \geq n$. Par le Corollaire 4.4.2 ii), l'ensemble complémentaire mod n de $\bigcup_{j=0}^{\alpha-1} I_j$ a la cardinalité du nombre d'éléments situés entre I_0 et I_1, I_1 et I_2, etc. Le nombre d'éléments situés entre I_r et I_{r+1} est $M(r+1) - M(r) - (2q-1)\beta$, soit la distance entre le début des deux intervalles moins la cardinalité d'un intervalle. On a

$$M(r+1) - M(r) - (2q-1)\beta = x - \beta\left(\left\lfloor \frac{z(r+1)}{\alpha} \right\rfloor - \left\lfloor \frac{zr}{\alpha} \right\rfloor\right). \qquad (4.16)$$

Il y a un trou entre deux intervalles si l'équation (4.16) est > 0 et cette valeur correspond au nombre d'entiers contenus dans le trou. Comme $x \leq \beta$, ce sera le cas pour tout r tel que $\left\lfloor \frac{z(r+1)}{\alpha} \right\rfloor - \left\lfloor \frac{zr}{\alpha} \right\rfloor = 0$. En utilisant le Lemme 4.4.3, avec $j = y$, $i = r$ et $n = \alpha$, on trouve que ce sera le cas pour exactement y valeurs de r. Il y a donc xy superpositions.

ii) Supposons $x > \beta$. Alors Card$(I) = n - x + \beta < n$. On a toujours que $\left\lfloor \frac{z(r+1)}{\alpha} \right\rfloor - \left\lfloor \frac{zr}{\alpha} \right\rfloor = 0$ pour y valeurs de r. De plus, comme $0 \leq r < \alpha$, il y a $(\alpha - 1)$ trous contenant chacun

$$x - \beta\left(\left\lfloor \frac{z(r+1)}{\alpha} \right\rfloor - \left\lfloor \frac{zr}{\alpha} \right\rfloor\right)$$

entiers. Ainsi,

$$\left\lfloor \frac{z(r+1)}{\alpha} \right\rfloor - \left\lfloor \frac{zr}{\alpha} \right\rfloor = 1$$

pour $\alpha - 1 - y = z - 1$ valeurs. Donc $x - \beta \left(\left\lfloor \frac{z(r+1)}{\alpha} \right\rfloor - \left\lfloor \frac{zr}{\alpha} \right\rfloor \right) = x - \beta$ pour $(z-1)$ valeurs de r.

En utilisant le Corollaire 4.4.2 iii), on sait qu'il y a $n - \text{Card}(I)$ autres valeurs possibles. On fait la somme et on obtient le nombre de superpositions :

$$\begin{aligned} xy + (x-\beta)(z-1) + n - \text{Card}(I) &= xy + (x-\beta)(\alpha - y - 1) + n - (n - x + \beta) \\ &= xy + x\alpha - \alpha\beta - xy + \beta y - x + \beta + n - n + x - \beta \\ &= x\alpha + y\beta - \alpha\beta. \end{aligned}$$

iii) Supposons $x \leq \beta$ et $y = \alpha$. Alors $\text{Card}(I) = n - x < n$. On est donc dans une situation semblable à ii). Par contre, une particularité apparaît :

$$\left\lfloor \frac{z(r+1)}{\alpha} \right\rfloor - \left\lfloor \frac{zr}{\alpha} \right\rfloor = 0$$

entre tous les intervalles, puisque $y = \alpha$ et donc, $z = \alpha - y = 0$. Comme il y a $(y-1) = (\alpha - 1)$ trous entre I_0 et $I_{\alpha-1}$, en utilisant le Corollaire 4.4.2 iii), on trouve que le nombre de superpositions est donné par

$$x(y-1) + n - \text{Card}(I) = x(y-1) + n - (n-x) = xy.$$

■

Remarque 4.4.5 À partir du Lemme 4.3.10, on a supposé que x et y est la solution de l'équation (4.14) telle que $y \leq \alpha$. Dans la preuve de la Proposition 4.4.4, nous utilisons ces résultats. Il serait possible de tout réécrire ces résultats en considérant plutôt la solution telle que $x \leq \beta$. Nous obtiendrions ainsi un résultat similaire à la Proposition 4.4.4, avec les conditions $y \leq \alpha$ et $y > \alpha$.

Le Théorème 4.4.6 est une généralisation du Corollaire 4.3.5 pour n'importe quelles valeurs de q, α, β.

Théorème 4.4.6 *Soient $C(n, q\alpha)$ et $C(n, q\beta)$, avec $\alpha \perp \beta$, deux mots de Christoffel qui se superposent. Alors $C(n, q\alpha)$ et $\gamma^{1-r}\widetilde{C}(n, q\beta)$ se superposent exactement, où $qr \equiv 1 \mod n$.*

Preuve Par le Corollaire 4.3.9, on sait que $C(n, q\alpha)$ et $\gamma^{(\ell+1)\overline{q\beta}}\widetilde{C}(n, q\beta)$ se superposent si et seulement si $\exists \ell \notin \bigcup_{i=0}^{\alpha-1} V_i \mod n$. Il suffit donc de montrer qu'il existe un $\ell \notin \bigcup_{i=0}^{\alpha-1} V_i \mod n$ tel que $(\ell+1)\overline{q\beta} = 1 - r$. Autrement dit, on veut montrer qu'il existe ℓ à l'extérieur des intervalles tel que

$$\begin{aligned}(\ell+1)\overline{q\beta} &= 1 - r \\ \ell\overline{q\beta} &= 1 - r - \overline{q\beta} \\ \ell &= -q\beta(1 - r - \overline{q\beta}) = -q\beta + rq\beta - 1 = \beta - 1 - q\beta.\end{aligned}$$

Montrons que $\beta - 1 - q\beta$ n'est jamais dans l'union des V_i.

Si $\alpha = 1$, alors l'union des V_i est l'intervalle $[-q\beta + \beta, q\beta - 1]$. Alors $\beta - q\beta - 1$ est l'élément précédent l'intervalle et comme les mots se superposent, l'intervalle est de longueur < n, donc $\beta - q\beta - 1 \mod n$ ne fait pas partie de l'intervalle. Si $\alpha > 1$, considérons les intervalles I_0 et I_1. Il y a des éléments entre ces deux intervalles, car

$$M(1) - M(0) - (2q-1)\beta = x + (2q-1)\beta - \left\lfloor \frac{z}{\alpha} \right\rfloor \beta - 0 - (2q-1)\beta = x > 0,$$

comme $z = \alpha - y < \alpha$. Nous allons montrer que l'élément $\beta - 1 - q\beta$ est entre I_0 et I_1 :

$$\begin{aligned}]\max(I_1), \min(I_0)[&=]q\beta - 1 - (x + (2q-1)\beta), (-q+1)\beta[\qquad (4.17) \\ &=]q\beta - 1 - x - 2q\beta + \beta, -q\beta + \beta[\\ &=]\beta - q\beta - 1 - x, -q\beta + \beta[. \qquad (4.18)\end{aligned}$$

Et donc, $\beta - q\beta - 1$ est bien entre I_0 et I_1. Pour conclure, il suffit de montrer que cet élément n'apparaît pas dans un autre intervalle. Cela sera vrai si $(\beta - q\beta - 1) - \min(I_{\alpha-1}) < n$. Vérifions :

$$\begin{aligned}(\beta - q\beta - 1) - \min(I_{\alpha-1}) &= \beta - q\beta - 1 - ((-q+1)\beta - (n - x - 2\beta(q-1) - i)) \\ &= n - 2\beta(q-1) - x - 1 - i < n\end{aligned}$$

où $i \in \{0, \beta\}$. ∎

4.5 Généralisation des résultats à des mots de longueurs différentes

Dans cette section, nous utilisons le Théorème 4.2.4 pour généraliser les résultats des Sections 4.3.2 et 4.4 pour deux mots de Christoffel quelconques.

Théorème 4.5.1 *(Morikawa, 1985a; Simpson, 2004) Soient les mots de Christoffel $C(n, q\alpha)$ et $C(m, q\beta)$ avec $\alpha \perp \beta$. Alors $C(n, q\alpha)$ et $C(m, q\beta)$ sont superposables si et seulement s'il existe $x, y \in \mathbb{N} - \{0\}$ tels que*

$$x\alpha + y\beta \quad = \quad p - 2\alpha\beta(q-1), \tag{4.19}$$

où $p = \text{pgcd}(m, n)$.

Preuve Posons $p = \text{pgcd}(m, n)$. On sait par le Théorème 4.2.4 que $C(n, q\alpha)$ et $C(m, q\beta)$ sont superposables si et seulement si $C(p, q\alpha)$ et $C(p, q\beta)$ le sont aussi. On conclut en utilisant le Théorème 4.3.18 qui nous assure que $C(p, q\alpha)$ et $C(p, q\beta)$ sont superposables si et seulement s'il existe de $x, y \in \mathbb{N} - \{0\}$ satisfaisant l'équation (4.19). ∎

Lemme 4.5.2 *Si $m > n$ et $C(p, \alpha)$ et $\gamma^{-k}C(p, \beta)$ se superposent exactement, alors $C(n, \alpha)$ et $\gamma^{-k+ip}C(m, \beta)$ se superposent aussi exactement, où $p = \text{pgcd}(n, m)$ et $0 \leq i < \frac{m}{p}$.*

Preuve Le Théorème 4.2.4 nous indique que $C(p, \alpha)$ et $\gamma^{-k}C(p, \beta)$ se superposent exactement si et seulement si $C(n, \alpha)$ et $\gamma^{-k}C(m, \beta)$ se superposent exactement. Mais $C(p, \alpha)$ et $\gamma^{-k}C(p, \beta)$ se superposent exactement si et seulement si $C(p, \alpha)$ et $\gamma^{-k+ip}C(p, \beta)$ se superposent aussi exactement. Ces $-k + ip$ correspondront à des décalages différents, pour $0 \leq i < \frac{m}{p}$, pour des mots de longueur au plus m. ∎

Proposition 4.5.3 *Soient $C(n, q\alpha)$ et $C(m, q\beta)$, deux mots de Christoffel superposables, avec $\alpha \perp \beta$, $p = \text{pgcd}(m, n)$ et $m > n$. Le nombre de superpositions est*
i) $xy\dfrac{m}{p}$, *si $x \leq \beta$;*
ii) $(x\alpha + y\beta - \alpha\beta)\dfrac{m}{p}$, *si $x > \beta$;*
où $x, y \in \mathbb{N} - \{0\}$ est la solution de $x\alpha + y\beta = p - \alpha\beta(q-1)$ telle que $y \leq \alpha$.

Preuve Découle de la Proposition 4.4.4 et du Lemme 4.5.2. ■

Théorème 4.5.4 *Soient $C(n, q\alpha)$ et $C(m, q\beta)$, avec $\alpha \perp \beta$ et $p = pgcd(m, n)$, deux mots de Christoffel qui se superposent. Alors $C(n, q\alpha)$ et $\gamma^{-(r-1)+ip}\widetilde{C}(m, q\beta)$ se superposent exactement, où $qr \equiv 1 \mod p$ et $0 \leq i < \frac{m}{p}$.*

Preuve Découle du Théorème 4.4.6 et du Lemme 4.5.2. ■

4.6 Autres résultats concernant la superposition de deux mots de Christoffel

Dans cette section, nous donnons d'abord une condition suffisante pour que deux mots de Christoffel se superposent sous quelques hypothèses. Nous terminons la section par un résultat concernant le mot obtenu suite à la superposition de deux mots de Christoffel.

Théorème 4.6.1 *Soient $u = C(n, \alpha) \in \{a < z\}^*$, $v = C(n, \beta) \in \{b < z\}^*$, des mots de Christoffel. Il existe $x, y \in \mathbb{N} - \{0\}$ tels que $\alpha x + \beta y = n$ si et seulement si u et $\widetilde{v}$ se superposent.*

Preuve ($\Longrightarrow$) Supposons qu'il existe $x, y \in \mathbb{N} - \{0\}$ tels que $\alpha x + \beta y = n$. Considérons les mots de Christoffel $u' = C(n, x\alpha)$ et $v' = C(n, y\beta)$. Comme $\alpha x + \beta y = n$, ces deux mots sont complémentaires, c'est-à-dire que u' et $\widetilde{v'}$ se superposent exactement. En utilisant le Lemme 4.2.3, on conclut que u et $\widetilde{v}$ se superposent aussi exactement.

($\Longleftarrow$) Supposons que u et $\widetilde{v}$ se superposent. Posons $d = \text{pgcd}(\alpha, \beta)$. Par le Lemme 4.2.3, $C(n, d) \in \{a < z\}^*$ et $\widetilde{C}(n, d) \in \{z < b\}^*$ se superposent aussi. Montrons maintenant que u et $\widetilde{v}$ ne se superposent que si $d|n$. Si d ne divise pas n, alors $C(n, d)$ s'écrit comme un produit de az^i et de az^{i+1}, il commence par az^i et finit par az^{i+1}. De plus, $\widetilde{C}(n, d)$ finit par bz^ib. Il y a donc un conflit entre un a et un b, puisque

$$C(n, d) = paz^iz \text{ et } \widetilde{C}(n, d) = p'bz^ib.$$

Donc $d|n$.

Mais comme $d = \text{pgcd}(\alpha, \beta)$, on a que $d|\alpha$ et $d|\beta$. Ainsi, $\text{pgcd}(n, \alpha) = d$ et $\text{pgcd}(n, \beta) = d$. Comme $C(n, \alpha)$ et $C(n, \beta)$ sont des mots de Christoffel, $\alpha \perp n$ et $\beta \perp n$. Ainsi, $d = 1$. En appliquant le Théorème 4.5.1 en prenant $m = n$, $q = 1$, on a $p = 1$ et donc, $C(n, \alpha)$ et $C(n, \beta)$ se superposent si et seulement s'il existe $x, y \in \mathbb{N} - \{0\}$ tels que $x\alpha + y\beta = n$. ■

Définition 4.6.2 Soit le mot $w \in \mathcal{A}^*$. Le mot $w' \in \mathcal{A}^*$ obtenu suite à la *décimation* $D_{p/q}$ de w par rapport à la lettre a, avec $p \leq q$, est le mot w' pour lequel p occurrences sur q de la lettre a sont supprimées. Lorsque $p/q > 0$, on supprime en partant de la droite et sinon, en partant de la gauche. On supprime les p premières occurrences rencontrées sur q.

Exemple 4.6.3 En appliquant une décimation $D_{1/3}$ sur le mot $w = aabaababababa$ par rapport à la lettre a, on obtient $w' = abababab$. Puis, en effectuant la décimation $D_{-1/2}$ sur w' par rapport à la lettre b, on obtient $w'' = aabaab$.

Théorème 4.6.4 *Soient $u = C(n, \alpha) \in \{a < z\}^*$ et $C(n, \beta) \in \{b < z\}^*$, deux mots de Christoffel superposables avec $\alpha \perp \beta$. Soit v le conjugué de $C(n, \beta)$ qui se superpose exactement à u. Soit le nouveau mot w défini par*

$$w[i] = \begin{cases} a & \text{si } u[i] = a \\ b & \text{si } v[i] = b \\ z & \text{sinon.} \end{cases}$$

Soit w' le mot obtenu à partir de w en supprimant les lettres z. Alors, w' est le mot de Christoffel de pente β/α.

Preuve Soient $x, y \in \mathbb{N} - \{0\}$ tels que $\alpha x + \beta y = n$. Par le Théorème 4.3.18, on sait que de tels x, y existent. Considérons le mot de Christoffel t avec αx occurrences de la lettre a et βy occurrences de la lettre b. Effectuons la décimation $D_{(i-1)/i}$ sur le mot t par rapport à la lettre a. On enlève ainsi les $(\alpha i - \alpha)$ lettres a en trop. Ensuite, effectuons la décimation $D_{-(j-1)/j}$ sur le mot obtenu sur la lettre b. On enlève ainsi les $(\beta j - \beta)$ lettres b en trop. Comme la décimation préserve les mots de Christoffel (Borel, 2001), le mot w' ainsi obtenu est un mot de Christoffel de longueur $\alpha + \beta$ avec α occurrences de la lettre a et β occurrences de la lettre b. ■

Exemple 4.6.5 Soient $u = C(13,4) = axxaxxaxxaxxx$ et $C(13,3) = bxxxbxxxbxxxx$. Ces mots sont superposables. En effet, il suffit de prendre le conjugué de $v = \widetilde{C}(13,3) = xxxxbxxxbxxxb$. On trouve alors $w = axxabxaxbaxxb$ et $z = aababab$. D'autre part, $4x + 3y = 13$ a pour solution $x = 1$ et $y = 3$. Considérons le mot de Christoffel $t = C(13,4) = abbabbabbabbb$. Effectuons la décimation $D_{0/1}$ sur t par rapport à la lettre a : on ne supprime aucun a. Puis on effectue la décimation $D_{-2/3}$ par rapport à la lettre b sur le mot obtenu : à partir de la gauche, on supprime 2 occurrences de b sur 3. On obtient alors le mot $aababab = C(7,4)$.

4.7 Problème de la monnaie

Dans la section précédente, nous avons vu que deux mots de Christoffel u et $\tilde{v}$ de longueur n se superposent si et seulement s'il existe des entiers α, β tels que $\alpha x + \beta y = n$. Dans cette section, la forme $\alpha x + \beta y$ intervient à nouveau : nous prouvons, en utilisant la représentation géométrique des mots de Christoffel, des résultats classiques de Sylvester concernant le problème de la monnaie, aussi connu sous le nom de *problème de Frobenius*.

D'abord, rappelons ce qu'est le problème de la monnaie.

Définition 4.7.1 (Weisstein, 2007) Soient n entiers $0 < a_1 < \ldots < a_n$ avec $n \geq 2$ qui représentent n différentes valeurs de pièces de monnaie et tels que pgcd$(a_1, a_2, \ldots, a_n) = 1$. Les montants d'argent pouvant être obtenus à partir de ces pièces de monnaie sont donnés par

$$\sum_{i=1}^{n} a_i x_i,$$

où les $x_i \in \mathbb{N}$ représentent le nombre de pièces a_i utilisées. Le *problème de la monnaie* consiste à déterminer le plus grand entier $N = g(a_1, a_2, \ldots, a_n)$ non représentable avec les pièces de monnaie $a_1, a_2, \ldots, a_n$. Cet entier est appelé le *nombre de Frobenius* .

Si $a_1 = 1$, toute somme d'argent peut être représentée. Par contre, dans le cas général, seuls quelques montants peuvent être représentés. Par exemple, avec des pièces de $2, 5$ et 10, il est impossible de représenter 1 et 3, alors que toutes les autres quantités sont représentables. Donc $g(2,5,10) = 3$.

Proposition 4.7.2 *(Sylvester, 1884) Le plus grand entier non représentable avec les deux pièces de monnaie a et b est*

$$g(a,b) \quad = \quad (a-1)(b-1)-1. \tag{4.20}$$

La Proposition 4.7.3 apparaît dans (Weisstein, 2007), mais l'auteur initial nous est inconnu.

Proposition 4.7.3 *Le nombre d'entiers non représentables avec a et b est donné par*

$$\frac{(a-1)(b-1)}{2}. \tag{4.21}$$

Corollaire 4.7.4 *Le nombre d'éléments du sous-monoïde engendré par a et b qui sont inférieurs à $(a-1)(b-1)$ est*

$$\frac{(a-1)(b-1)}{2}.$$

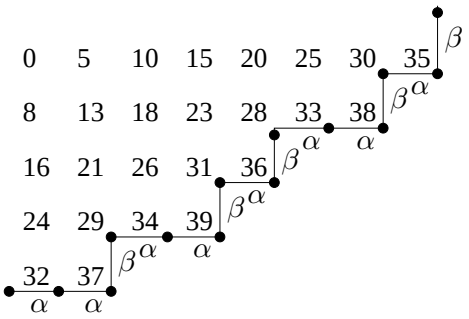
Preuve On sait par la Proposition 4.7.2 qu'à partir de $(a-1)(b-1)$ inclusivement, tous les entiers sont représentables. Donc les $\frac{(a-1)(b-1)}{2}$ entiers non représentables donnés dans la Proposition 4.7.3 sont nécessairement inférieurs à $(a-1)(b-1)$. Comme la moitié des $(a-1)(b-1)$ éléments inférieurs à $(a-1)(b-1)$ (en incluant le 0) ne sont pas représentables avec a et b, il en reste exactement la même quantité qui le soit. ■

Dans ce qui suit, nous allons voir qu'il est possible de prouver le Corollaire 4.7.4 en utilisant la représentation géométrique des mots de Christoffel et les graphes de Cayley.

Théorème 4.7.5 *Soient $a, b \in \mathbb{N}$. Considérons le quart de plan défini par $x \geq 0$ et $y \leq 0$, et ayant en la coordonnée $(x,-y)$ la valeur $xb+ya$. En ne dessinant que les coordonnées $(x,-y)$ entières telles que $xb+ya < ab$, la frontière obtenue se code par un mot de Christoffel ayant exactement a occurrences de la lettre α et b occurrences de la lettre β.*

Avant d'en faire la preuve, voici un exemple.

Exemple 4.7.6 Pour $a = 8$, $b = 5$, on a $ab = 40$. On obtient alors :



Si l'on associe x à un déplacement vers la droite et y à un déplacement vers le haut et si l'on commence au coin inférieur gauche, alors la frontière inférieure se code par le mot $\alpha\alpha\beta\alpha\alpha\beta\alpha\beta\alpha\alpha\beta\alpha\beta$: c'est bien le mot de Christoffel avec 8 occurrences de α et 5 occurrences de β.

Preuve (du Théorème 4.7.5) Considérons le graphe de Cayley du mot de Christoffel ayant a occurrences de la lettre α et b occurrences de la lettre β, avec $\alpha < \beta$. On obtient alors le graphe de Cayley représenté linéairement par

$$0 \to b \to 2b \bmod (a+b) \to \ldots \to ib \bmod (a+b) \to \ldots \to (a+b-1)b \bmod (a+b) \to 0$$

Dans ce graphe de Cayley, s'il existe $k \in \mathbb{N}$ tel que

$$ib < k(a+b) \leq (i+1)b,$$

alors

$$(i+1)b \mod (a+b) = (ib \mod (a+b)) - a. \tag{4.22}$$

Sinon, on a

$$(i+1)b \mod (a+b) = (ib \mod (a+b)) + b. \tag{4.23}$$

Considérons le graphe de Cayley précédent auquel on ajoute la valeur $ab - a - b$. Comme les valeurs du graphe de Cayley initial ne dépassaient pas $a + b$, les valeurs du nouveau graphe de Cayley ne dépassent pas $a + b + ab - a - b = ab$. Cela correspond exactement à prendre le chemin le plus à droite et le plus en bas tel que la valeur à la coordonnée $(x, -y)$ ne dépassent pas ab. En effet, on fait $+b$ (voir l'équation (4.23) : déplacement vers la droite) si l'on ne dépasse pas la valeur ab, sinon on fait $-a$ (voir l'équation (4.22) : déplacement vers le haut). ■

Dans l'exemple précédent, le graphe de Cayley est

$$0 \to 5 \to 10 \to 2 \to 7 \to 12 \to 4 \to 9 \to 1 \to 6 \to 11 \to 3 \to 8 \to 0$$

Le nouveau graphe de Cayley obtenu en ajoutant $ab - a - b = 27$ est

$$27 \to 32 \to 37 \to 29 \to 34 \to 39 \to 31 \to 36 \to 28 \to 33 \to 38 \to 30 \to 35 \to 27$$

et il correspond bien à la frontière décrite dans l'Exemple 4.7.6.

Voici une nouvelle preuve du Corollaire 4.7.4 qui utilise le résultat du Théorème 4.7.5.

Preuve (du Corollaire 4.7.4) Si l'on exclut les nombres qui forment la frontière dans le Théorème 4.7.5, en utilisant le graphe de Cayley vu précédemment, on obtient qu'il reste exactement les $xa + yb$ qui sont inférieurs à $(a-1)(b-1)$. Le nombre total de cases dans le rectangle est ab et comme il faut retrancher la frontière qui contient $a + b - 1$ éléments, puis diviser par 2, on obtient : $\dfrac{ab - (a+b-1)}{2} = \dfrac{(a-1)(b-1)}{2}$. ■

Nous avons donc réussi à exprimer en terme de combinatoire des mots, sous quelles conditions deux mots de Christoffel se superposent. Dans l'affirmative, nous allons plus loin que dans les travaux de (Morikawa, 1985a; Simpson, 2004) et nous donnons le décalage nécessaire pour qu'il y ait superposition. Afin d'éventuellement prouver la conjecture de Fraenkel, il faudrait généraliser ce résultat à la superposition de plus de deux mots de Christoffel.

Chapitre V

MOTS LISSES EXTRÉMAUX

En 1965, Kolakoski (Kolakoski, 1965) propose le problème suivant : *Quelle règle permet de construire la suite* $221121221221121122121121221 1 \cdots$ *? Quelle en est la* n*-ième lettre ? Cette suite est-elle périodique ?* Un an plus tard, la solution à ce problème est donnée dans (Kolakoski et Ucoluk, 1966). Les auteurs montrent que cette suite, notée K, est non périodique et ils en donnent la règle de construction : la suite décrit les longueurs des blocs successifs qu'elle possède. En effet,

$$\underbrace{22}_{2}\underbrace{11}_{2}\underbrace{2}_{1}\underbrace{1}_{1}\underbrace{22}_{2}\underbrace{1}_{1}\underbrace{22}_{2}\underbrace{11}_{2}\underbrace{2}_{1}\underbrace{11}_{2}\underbrace{22}_{2}\cdots$$

Ils prouvent que pour une première lettre fixée, cette construction donne une suite unique : si la première lettre est 1, alors la suite obtenue est $1K$, c'est-à-dire la suite K précédée de la lettre 1.

Cette solution est aussi attribuée à W. Bluger, H. B. Corstius, P. Cull, J. Dix, R. F. Jackson, N. Miller, J. Nadas, C. E. Olson et plusieurs autres. Dès son introduction, le *mot de Kolakoski* a donc suscité beaucoup d'intérêt auprès des mathématiciens. Entre autres, dans (Dekking, 1981), l'auteur s'intéresse aux suites qui s'engendrent d'elles-même. Puis, Weakley (Weakley, 1989) montre que la fonction de complexité de K, c'est-à-dire le nombre de facteurs de longueur n, est bornée de façon polynomiale. En étudiant les mots obtenus par itération alternante de morphismes, les auteurs de (Culik, Karhumäki et Lepistö, 1992) obtiennent le mot de Kolakoski. Carpi (Carpi, 1993; Carpi, 1994) montre que le mot de Kolakoski ne contient qu'un nombre fini de carrés, ce qui implique, en inspectant toutes les possibilités, que K est sans cube. Dans (Dekking, 1981; Dekking, 1997), l'auteur introduit de nombreuses conjectures au sujet de K

dont la plupart demeurent encore non prouvées malgré l'intérêt qu'on lui porte. Entre autre, il conjecture que K est récurrent, que l'ensemble de ses facteurs est fermé sous l'image miroir et sous complémentation et que la fréquence des lettres est $1/2$. Dans (Brlek et Ladouceur, 2003), un lien est établi entre la complexité palindromique et la récurrence de K : la présence d'un nombre infini de palindromes impliquerait la récurrence de K. Plus récemment, une formule récursive pour K a été donnée dans (Steinsky, 2006).

La classe des *mots lisses infinis* sur l'alphabet $\mathcal{A} = \{1, 2\}$, notée $\mathcal{K}$, est fortement reliée au mot de Kolakoski, puisqu'elle en est une généralisation. D'abord introduits dans (Brlek et Ladouceur, 2003), puis étudiés entre autres dans (Brlek, del Lungo et Ladouceur, 2002; Bergeron-Brlek et al., 2003; Berthé, Brlek et Choquette, 2005; Jamet et Paquin, 2005; Brlek, Jamet et Paquin, 2008; Brlek et al., 2006; Brlek, Melançon et Paquin, 2007), les mots lisses forment une famille infinie de mots pour laquelle beaucoup de propriétés n'ont pas encore été étudiées. Dans (Brlek et Ladouceur, 2003), les propriétés palindromiques des mots lisses sont décrites. Puis dans (Brlek et al., 2006), les auteurs montrent que les mots lisses ne contiennent pas de carré. Ils y conjecturent aussi que toutes les conjectures de Dekking au sujet de la structure de K se généralisent à la classe des mots lisses infinis sur $\{1, 2\}$. Dans (Berthé, Brlek et Choquette, 2005), les auteurs généralisent la classe des mots lisses infinis à un alphabet général à deux lettres $\{a < b\}$, avec $a, b \in \mathbb{N} - \{0\}$, et même, à des alphabets numériques à plus de deux lettres. Ils y établissent aussi une caractérisation du mot de Fibonacci par un mot ultimement périodique. Dans les articles (Brlek, Jamet et Paquin, 2008; Brlek, Melançon et Paquin, 2007), les auteurs s'intéressent aux mots lisses extrémaux, c'est-à-dire le plus petit et le plus grand de la classe selon l'ordre lexicographique. Dans le premier article, les auteurs se limitent aux mots lisses sur les alphabets $\{1, 2\}$, puis dans (Brlek, Melançon et Paquin, 2007), ils étudient les mots lisses extrémaux sur des alphabets plus généraux. Dans (Jamet et Paquin, 2005), les auteurs s'intéressent aux surfaces discrètes représentables par un pavage lisse, c'est-à-dire un pavage ne contenant que des mots lisses. Ces trois articles (Brlek, Jamet et Paquin, 2008; Brlek, Melançon et Paquin, 2007; Jamet et Paquin, 2005) font l'objet de ce chapitre et des deux suivants.

Dans un premier temps, nous définissons les mots lisses sur l'alphabet $\{1, 2\}$ tels qu'initiale-

ment définis, puis nous montrons comment un pavage lisse du quart du plan peut être associé à un mot lisse. Ensuite, nous définissons la version finie des mots lisses : facteurs, préfixes et suffixes lisses, puis nous introduisons les graphes de De Bruijn, qui s'avèrent être fort utiles pour représenter la structure d'un mot lisse. Par la suite, nous entrons dans le vif du sujet : l'étude des mots lisses extrémaux. Dans un premier temps, nous définissons ce que sont les mots lisses extrémaux, puis nous montrons comment les construire. Nous utilisons d'abord un algorithme naïf, puis nous montrons comment l'améliorer en utilisant les graphes de De Bruijn et finalement, nous présentons un troisième algorithme qui utilise la bijection entre les mots lisses infinis et les mots infinis. Par la suite, nous étudions leur structure en calculant les dérivées successives de ces mots en utilisant l'opérateur Δ, l'opérateur de codage par blocs, puis en analysant leur factorisation de Lyndon.

5.1 Mots lisses

Avant de définir la famille des mots lisses, quelques résultats et définitions sont préalables.

Lemme 5.1.1 *Soit $\mathcal{A} = \{1, 2\}$. Alors tout mot non vide $w \in \mathcal{A}^\infty$ s'écrit de façon unique comme un produit de facteurs*

$$w = \begin{cases} 1^{i_0}2^{i_1}1^{i_2}\cdots & \text{si } w[0] = 1 \\ 2^{i_0}1^{i_1}2^{i_2}\cdots & \text{si } w[0] = 2 \end{cases}$$

avec $i_k > 0$, pour $k \geq 0$.

Preuve Il suffit de prendre i_k comme étant la longueur maximale du $(k+1)$-ième bloc de lettres. ■

Exemple 5.1.2 Le mot $u = 112112222212212121$ s'écrit comme $u = 1^2 2^1 1^2 2^5 1^1 2^2 1^1 2^1 1^1 2^1$, le mot $v = 221212222$ s'écrit comme $v = 2^2 1^1 2^1 1^1 2^4$ et $w = (112)^\omega$ s'écrit comme $w = (1^2 2^1)^\omega$.

Définition 5.1.3 La *fonction de codage par blocs* est définie par l'opérateur $\Delta : \mathcal{A}^\infty \longrightarrow \mathbb{N}^\infty$,

$$\Delta(w) = i_0 i_1 i_2 \cdots = \prod_{k \geq 0} i_k,$$

où les i_k correspondent aux exposants de l'unique factorisation décrite dans le Lemme 5.1.1. Par convention, $\Delta(\varepsilon) = 0$.

La *fonction de codage par blocs* est utilisée dans plusieurs applications comme moyen de compression des données. Par exemple, la première étape qu'effectuent les télécopieurs afin de compresser les données consiste en l'application de Δ sur chaque ligne de pixels. Cette fonction s'est révélée fort utile lors de l'énumération des facteurs de la suite de Thue-Morse (Brlek, 1989).

Remarque 5.1.4 L'opérateur Δ introduit ici n'a aucun lien avec le Δ des suites directrices des suites épisturmiennes. À partir de maintenant, le symbole Δ fera toujours référence à celui de la Définition 5.1.3.

Exemple 5.1.5 Soit $u = 12212211$. Alors $u = 1^1 2^2 1^1 2^2 1^2$ et donc, $\Delta(u) = [1, 2, 1, 2, 2]$.

Afin d'avoir une notation plus compacte, la ponctuation est supprimée si cela n'occasionne aucune ambiguïté. Ainsi, pour l'Exemple 5.1.5, $\Delta(u) = 12122$.

Exemple 5.1.6

a) Soit $v = 221212222221$ qui s'écrit comme $v = 2^2 1^1 2^1 1^1 2^6 1^1$. Alors $\Delta(v) = 211161$.

b) Soit $w = 122111222211111$ qui s'écrit comme $w = 1^1 2^2 1^3 2^4 1^5$. Alors $\Delta(w) = 12345$.

c) Soit $x = (1121112)^\omega$ qui s'écrit comme $x = (1^2 2^1 1^3 2^1)^\omega$. Alors $\Delta(x) = (2131)^\omega$.

Remarquons que dans l'Exemple 5.1.5, l'alphabet du mot initial est le même que celui du mot obtenu suite à l'application de l'opérateur Δ. Par contre, il peut arriver comme dans l'Exemple 5.1.6 que les alphabets ne coïncident pas.

Le lemme suivant apparaît dans (Brlek et al., 2006), mais aucune preuve n'y est donnée de façon explicite.

Lemme 5.1.7 *La fonction Δ est une contraction, c'est-à-dire que pour tout mot $w \in \mathcal{A}^*$,*

$$\lg(\Delta(w)) \leq \lg(w)\,, \tag{5.1}$$

et l'égalité est obtenue si $w \in \{\varepsilon, 2\} \cdot (12)^* \cdot \{\varepsilon, 1\}$.

Preuve Puisque la fonction Δ associe à un mot w le mot formé par la concaténation des longueurs de ses blocs, pour chaque bloc de longueur ≥ 1, une seule lettre est associé. Ainsi, $\lg(\Delta(w)) \leq \lg(w)$. L'égalité sera donc obtenue si tous les blocs sont de longueur 1, c'est-à-dire si w à l'une des 4 formes suivantes : $(12)^n$, $2(12)^n1$, $(12)^n1$ ou $2(12)^n$ pour $n \in \mathbb{N}$. L'ensemble de ces mots correspond à l'ensemble $\{\varepsilon, 2\} \cdot (12)^* \cdot \{\varepsilon, 1\}$. ■

Il est possible d'itérer l'opérateur Δ sur un mot fini $w \in \{1,2\}^*$ jusqu'à ce qu'un mot de longueur 1 soit obtenu, ou jusqu'à ce que l'alphabet du mot obtenu diffère de l'alphabet initial. Dans ce chapitre, à moins de le préciser, nous utilisons toujours l'alphabet à deux lettres $\mathcal{A} = \{1,2\}$.

Exemple 5.1.8 Soit $w = 12211211$. En appliquant successivement l'opérateur Δ, on obtient
$\Delta^0(w) = 12211211$
$\Delta^1(w) = 12212$
$\Delta^2(w) = 1211$
$\Delta^3(w) = 112$
$\Delta^4(w) = 21$
$\Delta^5(w) = 11$
$\Delta^6(w) = 2.$

Exemple 5.1.9 Soit $u = 2221112121212$. En appliquant successivement l'opérateur Δ, on obtient
$\Delta^0(w) = 2221112121212$
$\Delta^1(w) = 331111111.$

Exemple 5.1.10 Soit le mot $v = 112212211221211$.En appliquant successivement l'opérateur Δ, on obtient
$\Delta^0(w) = 112212211221211$
$\Delta^1(w) = 221222112$
$\Delta^2(w) = 21321.$

Nous pouvons maintenant définir l'objet d'étude de ce chapitre et du suivant : la famille des *mots lisses*

Définition 5.1.11 L'ensemble des *mots lisses* finis ou infinis sur l'alphabet $\mathcal{A} = \{1, 2\}$, noté $\mathcal{K}$, est défini par

$$\mathcal{K} = \{w \in \mathcal{A}^\infty \mid \forall k \in \mathbb{N}, \Delta^k(w) \in \mathcal{A}^\infty\}.$$

Ainsi, parmi les mots des exemples 5.1.8, 5.1.9 et 5.1.10, le seul mot lisse est celui du premier exemple : $w = 12211211$. Il est plutôt difficile de montrer qu'un mot infini est lisse. Nous verrons ultérieurement comment construire des mots lisses infinis.

Afin d'étudier les mots lisses, rappelons les propriétés suivantes de l'opérateur Δ.

Proposition 5.1.12 *(Brlek et Ladouceur, 2003) Pour tout $u \in \mathcal{A}^*$ et pour tout $p \in \text{pal}(\mathcal{A}^*)$, l'opérateur Δ satisfait les conditions*

i) $\Delta(\widetilde{u}) = \widetilde{\Delta(u)}$;

ii) $\Delta(\overline{u}) = \Delta(u)$;

iii) $\Delta(p) \in \text{pal}(\mathcal{A}^*)$.

Exemple 5.1.13 Soient $w = 12112112$, $v = 21221221$ et $p = 2212122$. On a

a) $\Delta(\widetilde{w}) = \Delta(21121121) = 121211$ et $\widetilde{\Delta(w)} = \widetilde{112121} = 121211$;

b) $\Delta(\overline{v}) = \Delta(12112112) = 112121$ et $\Delta(v) = 112121$;

c) $p \in \text{pal}(\mathcal{A}^*)$ et $\Delta(p) = 21112 \in \text{pal}(\mathcal{A}^*)$;

d) $\Delta(w) = 112121 = \Delta(v)$ et $w = \overline{v}$.

L'Exemple 5.1.13 d) illustre le fait que Δ n'est pas bijective, comme $\Delta(v) = \Delta(\overline{v})$. Il est toutefois possible de définir pour des alphabets numériques à deux lettres $\{a < b\}$ des fonctions pseudo-inverses de la façon suivante.

Définition 5.1.14 Soit $u \in \mathcal{A}^\infty$ et fixons $\mathcal{A} = \{a < b\}$. Les fonctions pseudo-inverses sont définies par

$$\Delta_a^{-1}, \Delta_b^{-1} : \mathcal{A}^\infty \longrightarrow \mathcal{A}^\infty$$

telles que

$$\Delta_\alpha^{-1}(u) = \alpha^{u[0]}\overline{\alpha}^{u[1]}\alpha^{u[2]}\overline{\alpha}^{u[3]}\cdots, \quad \text{pour } \alpha \in \{a, b\}.$$

Exemple 5.1.15 Soit $u = 11211$. Alors $\Delta_2^{-1}(u) = \Delta_2^{-1}(11211) = 212212$.

Exemple 5.1.16 Soit $v = 12211212$. Alors $\Delta_1^{-1}(v) = \Delta_1^{-1}(12211212) = 122112122122$.

Lemme 5.1.17 *La fonction pseudo-inverse Δ^{-1} commute sous l'image miroir, c'est-à-dire que pour $u \in \mathcal{A}^*$, $\alpha, \beta \in \mathcal{A}$,*

$$\widetilde{\Delta_\alpha^{-1}(u)} = \Delta_\beta^{-1}(\widetilde{u}) \tag{5.2}$$

où $\beta = \alpha$ si $\lg(u)$ est impaire et $\beta = \overline{\alpha}$ si $\lg(u)$ est paire.

Preuve Soit $u = u[0]u[1]u[2]\cdots u[n]$ et supposons $\mathcal{A} = \{1, 2\}$. Si $\lg(u)$ est impaire, alors

$$\Delta_1^{-1}(w) = 1^{u[0]}2^{u[1]}\cdots 2^{u[n-1]}1^{u[n]}$$

et $\widetilde{\Delta_1^{-1}(u)} = 1^{u[n]}2^{u[n-1]}\cdots 2^{u[1]}1^{u[0]}$. D'autre part,

$$\Delta_1^{-1}(\widetilde{u}) = \Delta_1^{-1}(u[n]u[n-1]\cdots u[1]u[0]) = \widetilde{\Delta_1^{-1}(w)}.$$

Si $\lg(u)$ est paire, alors

$$\Delta_1^{-1}(w) = 1^{u[0]}2^{u[1]}\cdots 1^{u[n-1]}2^{u[n]}$$

et $\widetilde{\Delta_1^{-1}(u)} = 2^{u[n]}1^{u[n-1]}\cdots 2^{u[1]}1^{u[0]}$. D'autre part,

$$\Delta_2^{-1}(\widetilde{u}) = \Delta_2^{-1}(u[n]u[n-1]\cdots u[1]u[0]) = \widetilde{\Delta_1^{-1}(w)}.$$

■

Exemple 5.1.18 Soit le mot $u = 1213431$. Alors
$\Delta_1^{-1}(u) = 1^1 2^2 1^1 2^3 1^4 2^3 1^1 = 122122211112221$ et $\Delta_2^{-1}(u) = 211211122221112$.

Cet exemple nous donne l'intuition du résultat qui suit.

Lemme 5.1.19 *Soit $u \in \mathcal{A}^\infty$. Alors $\Delta_1^{-1}(u) = \overline{\Delta_2^{-1}(u)}$.*

Preuve Découle directement du fait que $\Delta(u) = \Delta(\overline{u})$. ■

Définition 5.1.20 (Brlek et Ladouceur, 2003) L'ensemble des *mots k-différentiables* sur l'alphabet $\mathcal{A}$, noté $\Delta^k_{\mathcal{A}}$, est défini par

$$\Delta^k_{\mathcal{A}} = \{w \in \mathcal{A}^+ | (\Delta^k(w) = 2) \wedge (\forall j, 1 \leq j \leq k-1, \Delta^j(w) \in \mathcal{A}^+)\}$$

et alors, $\Delta^+_{\mathcal{A}} = \bigcup_{k\geq 1} \Delta^k_{\mathcal{A}}$. Autrement dit, un mot fini w est k-différentiable s'il est lisse et que $\Delta^k(w) = 2$.

Proposition 5.1.21 *(Brlek et Ladouceur, 2003) L'ensemble des mots lisses finis satisfait les propriétés de fermeture*

$$u \in \Delta^k_{\mathcal{A}} \iff \overline{u}, \widetilde{u} \in \Delta^k_{\mathcal{A}}, \ \forall k \geq 0 \tag{5.3}$$

$$u \in \mathcal{K} \iff \overline{u} \in \mathcal{K}. \tag{5.4}$$

La propriété (5.3) signifie qu'un mot u est k-différentiable si et seulement si son complément et son image miroir le sont aussi. La propriété (5.4) nous indique qu'un mot u est lisse si et seulement si son complément $\overline{u}$ l'est aussi.

Exemple 5.1.22 Le mot $u = 11211221211$ est lisse. Plus particulièrement, u est 6-différentiable. En effet,

$\Delta^0(u) = 11211221211$

$\Delta^1(u) = 2122112$

$\Delta^2(u) = 11221$

$\Delta^3(u) = 221$

$\Delta^4(u) = 21$

$\Delta^5(u) = 11$

$\Delta^6(u) = 2.$

Par la Proposition 5.1.21, on sait que le mot $\widetilde{u} = 22122112122$ est aussi un mot lisse et que

$$\Delta^6(\widetilde{u}) = \Delta^6(11212211211) = 2.$$

Définition 5.1.23 Le *mot de Kolakoski* (Kolakoski, 1965)

$$K = 2211212212211211221211212211211212212211212212112112 2\cdots$$

est le point fixe sous l'opérateur Δ et ayant 2 comme première lettre.

Construisons le mot de Kolakoski. On sait qu'il commence par 2. Donc $K[0] = 2$ et cela implique $\Delta(K)[0] = 2$. Ainsi, en appliquant $\Delta_2^{-1}(2)$, on trouve que $K[0,1] = 22$. Ainsi, on a aussi $\Delta(K)[0,1] = 22$ et en appliquant $\Delta_2^{-1}(22)$, on trouve $K[0,3] = 2211$. Ainsi de suite, la longueur du préfixe de K augmente et on peut obtenir un préfixe de K aussi long que désiré.

Dans la famille $\mathcal{K}$ des mots lisses, l'opérateur Δ a exactement deux points fixes et ce sont les deux mots infinis suivants :

$$\Delta(K) = K \quad \text{et} \quad \Delta(1 \cdot K) = 1 \cdot K.$$

5.2 Pavage lisse du quart de plan et bijection Φ

Dans cette section, nous introduisons les deux outils nécessaires pour construire des mots infinis qui sont lisses : la bijection Φ entre les mots lisses infinis et les mots infinis sur un alphabet $\mathcal{A}$ et le pavage lisse d'un quart de plan.

Définition 5.2.1 Soit $w \in \mathcal{A}^\infty$ un mot lisse. Alors il existe une bijection

$$\Phi : \mathcal{K} \longrightarrow \mathcal{A}^\infty$$

qui associe de façon unique à tout $w \in \mathcal{K}$ un mot $\Phi(w) \in \mathcal{A}^\infty$. Cette bijection est définie par

$$\Phi(w) = \Delta^0(w)[0]\Delta^1(w)[0]\Delta^2(w)[0]\Delta^3(w)[0]\cdots = \prod_{i\geq 0} \Delta^i(w)[0].$$

Exemple 5.2.2 Dans l'Exemple 5.1.8, on a $\Phi(12211211) = 1111212$ et $u = 1111212$ n'est pas un mot lisse, puisque $\Delta(u) = 4111$.

Remarquons qu'en général, le mot obtenu par la fonction Φ n'est pas lisse.

Exemple 5.2.3 Pour le mot de Kolakoski K, on a $\Delta^i(K)[0] = 2$ pour tout i, puisque K est un point fixe pour la fonction Δ. Ainsi, $\Phi(K) = 2^\omega$.

Par convention, pour un mot fini $w \in \{1,2\}^*$, $\Phi(w)$ termine par un 2. Sinon, on obtiendrait une queue infinie de 1.

Comme Φ est une fonction bijective, nous devons définir sa fonction inverse. Pour ce faire, des définitions et lemmes sont préalables.

Définition 5.2.4 Soit $u \in \mathcal{A}^k$, alors $\Phi^{-1}(u) = w_k$, où

$$w_n = \begin{cases} u[k-1], & \text{si } n = 1; \\ \Delta^{-1}_{u[k-n]}(w_{n-1}), & \text{si } 2 \leq n \leq k. \end{cases}$$

Exemple 5.2.5 Soit le mot w pour lequel $u = \Phi(w) = 1212222$. On a $k = 7$, $w_1 = u[6] = 2$,
$w_2 = \Delta^{-1}_{u[5]}(w_1) = \Delta_2^{-1}(2) = 22$,
$w_3 = \Delta^{-1}_{u[4]}(w_2) = \Delta_2^{-1}(22) = 2211$,
$w_4 = \Delta^{-1}_{u[3]}(w_3) = \Delta_2^{-1}(2211) = 221121$,
$w_5 = \Delta^{-1}_{u[2]}(w_4) = \Delta_1^{-1}(221121) = 112212112$,
$w_6 = \Delta^{-1}_{u[1]}(w_5) = \Delta_2^{-1}(112212112) = 2122112112122$,
$w_7 = \Delta^{-1}_{u[1]}(w_6) = \Delta_1^{-1}(2122112112122) = 11211221211212212211 = \Phi^{-1}(u) = w$.

Définition 5.2.6 Soit w, un mot lisse fini. Le *pavage partiel du quart de plan* associé à w est construit de façon à ce que la i-ième ligne consiste en $\Delta^i(s)$.

Exemple 5.2.7 Soit le mot lisse $w = 2122121122122$. Le pavage partiel du quart de plan associé à w est
$\Delta^0(w) = 2122121122122$
$\Delta^1(w) = 112112212$
$\Delta^2(w) = 212211$
$\Delta^3(w) = 1122$
$\Delta^4(w) = 22$
$\Delta^5(w) = 2$.

Exemple 5.2.8 Considérons maintenant le mot $\Phi(w) = 212122$ auquel la lettre 2 est ajoutée. En appliquant Δ^{-1} en partant de la fin de $\Phi(w)$ et en écrivant de bas en haut les mots obtenus, le pavage partiel du quart de plan suivant est obtenu.

$\Delta^0(w) = 2122121122122 \cdot 12112$

$\Delta^1(w) = 112112212 \cdot 1121$

$\Delta^2(w) = 212211 \cdot 211$

$\Delta^3(w) = 1122 \cdot 12$

$\Delta^4(w) = 22 \cdot 11$

$\Delta^5(w) = 2 \cdot 2$

$\Delta^6(w) = 2.$

Les Exemples 5.2.7 et 5.2.8 suggèrent que pour une lettre x et un mot w, $\Phi^{-1}(w)$ est préfixe de $\Phi^{-1}(wx)$. Ce résultat est vrai en général.

Lemme 5.2.9 (Lemme du collage) *(Berthé, Brlek et Choquette, 2005) Soient $u, v \in \Delta_{\mathcal{A}}^*$. S'il existe m tel que pour tout i, $0 \leq i \leq m$, la dernière lettre de $\Delta^i(u)$ est différente de la première lettre de $\Delta^i(v)$ et $\Delta^i(u) \neq 1$, $\Delta^i(v) \neq 1$, alors*

i) $\Phi(uv) = \Phi(u)[0, m] \cdot \Phi \circ \Delta^{m+1}(uv)$;

ii) $\Delta^i(uv) = \Delta^i(u)\Delta^i(v)$.

Corollaire 5.2.10 *Soit $u \in \mathcal{A}^*$ et $x \in \mathcal{A}$. Alors $\Phi^{-1}(u)$ est préfixe de $\Phi^{-1}(ux)$.*

Le Corollaire 5.2.10 nous permet de définir la fonction inverse de Φ pour un mot infini de la façon suivante.

Définition 5.2.11 Soit $u \in \mathcal{A}^\omega$. On définit $\Phi^{-1}(u) = \lim\limits_{k\to\infty} w_k = \lim\limits_{k\to\infty} \Phi^{-1}(u[0, k-1])$, où w_k est le même que celui de la Définition 5.2.4.

Soit u un mot infini. Afin d'obtenir un préfixe de $\Phi^{-1}(u)$ arbitrairement long, il suffit de prendre un préfixe de u aussi long que nécessaire. Par le Corollaire 5.2.10, la limite existe et donc, la fonction Φ^{-1} est bien définie pour les mots infinis. Cette bijection apparaît dans la thèse de Lamas (Lamas, 1995) pour la classification des mots infinis. Dans la même période, Dekking

(Dekking, 1997) utilise indépendamment cette bijection pour montrer l'existence de mots satisfaisant $\Delta^n(w) = w$ pour tout n. Pour $n = 1$, il s'agit des mots de Kolakoski.

Corollaire 5.2.12 *(du Corollaire 5.2.10) Tout mot lisse fini u est prolongeable en un mot lisse plus long.*

Preuve Il suffit d'ajouter une lettre au mot $\Phi(u)$ et d'appliquer Φ^{-1} à ce mot. ■

Exemple 5.2.13 Soit le mot lisse w tel que $\Phi(w) = 12122122212222\cdots$. Les préfixes de w de longueurs croissantes obtenus par Φ^{-1} sont

$\Phi^{-1}(1) = 1$

$\Phi^{-1}(12) = 11$

$\Phi^{-1}(121) = 11$

$\Phi^{-1}(1212) = 112$

$\Phi^{-1}(12122) = 112112212$

$\Phi^{-1}(121221) = 112112212$

$\Phi^{-1}(1212212) = 11211221211$

...

Dans l'exemple précédent, on peut remarquer que malgré l'information supplémentaire obtenue en ajoutant la lettre 1 à un mot fini, la longueur du mot obtenu par Φ^{-1} n'augmente pas. Pour utiliser toute l'information, il faudrait *forcer les coups*.

Définition 5.2.14 Inspiré de l'opérateur Δ^{-1}, l'opérateur δ^{-1} est défini comme

$$\delta_\alpha^{-1}(x) = \Delta_\alpha^{-1}(x) \cdot \beta,$$

où $x \in \mathcal{A}^*$, $\alpha, \beta \in \mathcal{A}$ et β est la lettre complémentaire à la dernière lettre de $\Delta_\alpha^{-1}(x)$.

La définition précédente découle du Lemme du collage : forcément, la lettre qui suit $\Delta_\alpha^{-1}(x)$ doit être différente de la dernière lettre de $\Delta_\alpha^{-1}(x)$. Cette définition nous permet de construire des préfixes de mots lisses infinis plus long qu'en utilisant la fonction Δ^{-1}.

Définition 5.2.15 L'opérateur ϕ^{-1} est défini comme l'analogue de Φ^{-1}, avec δ^{-1} plutôt que Δ^{-1}.

Exemple 5.2.16 Reprenons l'Exemple 5.2.13 avec l'opérateur ϕ^{-1}. Le résultat obtenu est

$\phi^{-1}(1) = 1$

$\phi^{-1}(12) = 112$

$\phi^{-1}(121) = 1121$

$\phi^{-1}(1212) = 112112212$

$\phi^{-1}(12122) = 1121122121121$

$\phi^{-1}(121221) = 112112212112122112112$

$\phi^{-1}(1212212) = 1121122121121221121121221211221221121$

...

Définition 5.2.17 Un *pavage lisse du quart de plan*, aussi appelé par abus de langage *pavage lisse du plan*, consiste au quart du plan pour lequel la i-ième ligne est $\Delta^i(w)$, où w est un mot lisse infini fixé.

Proposition 5.2.18 *Pour un alphabet $\mathcal{A}$ fixé, l'ensemble des mots lisses infinis est en bijection avec l'ensemble des mots infinis.*

Preuve Découle de la fonction bijective Φ. ∎

Corollaire 5.2.19 *Il existe un nombre infini de mots lisses infinis.*

Preuve Découle de la Proposition 5.2.18 : comme il y a un nombre infini de mots infinis sur un alphabet donné, le nombre de mots lisses infinis sur cet alphabet est lui aussi infini. ∎

Corollaire 5.2.20 *Tout mot infini $u \in \mathcal{A}^\omega$ décrit un unique pavage lisse du plan.*

Preuve Découle de la Proposition 5.2.18. ∎

Exemple 5.2.21 Considérons le mot $u = 1112^\omega$. Ce mot détermine le pavage lisse suivant

$\mathbf{1}2212112212211211212212211212212112212212112122112122121\cdots$

$\mathbf{1}2112212212112122112112212112112212212112212211211212212121122\cdots$

$\mathbf{1}1221211211222122112122121122122121121122121121221221121121212\cdots$

$\mathbf{2}2112122122112112212112122112112122122112122121121122122112121\cdots$

$\mathbf{2}2112122122112112212112122112112122122112122121121122122112121\cdots$

$\mathbf{2}2112122122112112212112122112112122122112122121121122122112121\cdots$

$\mathbf{2}2112122122112112212112122112112122122112122121121122122112121\cdots$

$\mathbf{2}2112122122112112212112122112112122122112122121121122122112121\cdots$

$\mathbf{2}2112122122112112212112122112112122122112122121121122122112121\cdots$

$\mathbf{2}2112122122112112212112122112112122122112122121121122122112121\cdots$

...

Proposition 5.2.22 *(Bergeron-Brlek et al., 2003) Soit $w \in \mathcal{A}^\omega$, un mot lisse infini. Si $\Phi(w)$ est périodique de période minimale de longueur p, alors $\Delta^i(w) = \Delta^{i+np}(w)$, pour tout $n \in \mathbb{N}$. Si w est un mot lisse fini tel que $w \in \Delta^k_{\mathcal{A}}$, alors $\Delta^{i+np}(w)$ est toujours préfixe de $\Delta^i(w)$, pour $i + np \leq k$.*

5.3 Facteurs lisses

Afin de définir la notion de facteur lisse, nous devons introduire un nouvel opérateur similaire à Δ noté D. Cet opérateur a d'abord été défini dans (Dekking, 1981) et a ensuite été utilisé pour l'énumération des mots C^∞ dans (Weakley, 1989).

Définition 5.3.1 L'opérateur $D : \mathcal{A}^* \longrightarrow \mathbb{N}^*$ est défini comme

$$D(w) = \begin{cases} \varepsilon & \text{si } \Delta(w) = 1 \text{ ou } w = \varepsilon, \\ \Delta(w) & \text{si } \Delta(w) = 2x2 \text{ ou } \Delta(w) = 2, \\ 2x & \text{si } \Delta(w) = 2x1, \\ x2 & \text{si } \Delta(w) = 1x2, \\ x & \text{si } \Delta(w) = 1x1. \end{cases}$$

Appliquer l'opérateur D correspond à éliminer le préfixe 1 et/ou le suffixe 1 du mot obtenu suite à l'application de l'opérateur Δ, s'ils existent. Notons que l'opérateur D est seulement défini sur les mots finis.

Tout comme l'opérateur Δ, il est possible d'itérer D. L'itération arrête lorsque l'alphabet change, ou lorsque le mot obtenu est ε.

Exemple 5.3.2 Soit $w = 211211221$. Alors
$D^1(w) = 2122$
$D^2(w) = 12$
$D^3(w) = \varepsilon$.

Définition 5.3.3 Un mot fini w est appelé *un facteur lisse* s'il existe $k \in \mathbb{N}$ tel que $D^k(w) = \varepsilon$ et que pour tout $j \leq k-1$, $D^j(w) \in \mathcal{A}^*$.

Le mot de l'Exemple 5.3.2 est un facteur lisse.

De façon analogue à un facteur lisse, nous définissons maintenant ce qu'est un *préfixe lisse* et un *suffixe lisse*. Pour ce faire, deux nouveaux opérateurs sont nécessaires.

Définition 5.3.4 (Brlek, Melançon et Paquin, 2007) La *dérivée à droite* est définie par la fonction $D_r : \mathcal{A}^* \to \mathbb{N}^*$ telle que

$$D_r(w) = \begin{cases} \varepsilon & \text{si } \Delta(w) = 1 \text{ ou } w = \varepsilon, \\ \Delta(w) & \text{si } \Delta(w) = x2, \\ x & \text{si } \Delta(w) = x1. \end{cases}$$

Définition 5.3.5 (Brlek, Melançon et Paquin, 2007) De façon semblable, la *dérivée à gauche* est définie par la fonction $D_\ell : \mathcal{A}^* \to \mathbb{N}^*$ telle que

$$D_\ell(w) = \begin{cases} \varepsilon & \text{si } \Delta(w) = 1 \text{ ou } w = \varepsilon, \\ \Delta(w) & \text{si } \Delta(w) = 2x, \\ x & \text{si } \Delta(w) = 1x. \end{cases}$$

Définition 5.3.6 Un mot w est *lisse à droite* (resp. *lisse à gauche*) s'il existe k tel que $D_r^k(w) = \varepsilon$ (resp. $D_\ell^k(w) = \varepsilon$) et pour tout $j \leq k-1$, $D_r^j(w) \in \mathcal{A}^*$ (resp. $D_\ell^j(w) \in \mathcal{A}^*$). Un mot lisse à droite (resp. à gauche) est alors appelé un *préfixe lisse* (resp. *suffixe lisse*).

Les noms *préfixe lisse* et *suffixe lisse* proviennent du fait que ce sont respectivement les opérateurs D_r et D_ℓ qui permettent de dire si le mot est respectivement préfixe d'un mot lisse et suffixe d'un mot lisse. En effet, il est facile de se convaincre qu'un préfixe lisse est toujours prolongeable à droite en un mot lisse et qu'un suffixe lisse est toujours prolongeable à gauche en un mot lisse.

Exemple 5.3.7 Soit $w = 1122121$. En appliquant successivement les opérateurs D_r et D_ℓ sur le mot w, on obtient

$D_r^0(w) = 1122121$ $\quad D_\ell^0(w) = 1122121$

$D_r^1(w) = 2211$ $\quad D_\ell^1(w) = 22111$

$D_r^2(w) = 22$ $\quad D_\ell^2(w) = 23.$

$D_r^3(w) = 2$

Le mot w est lisse à droite, mais n'est pas lisse à gauche, comme $D_\ell^2(w) \notin \mathcal{A}^*$.

Exemple 5.3.8 Soit $v = 12122112$. Alors $D_r^2(v) = 32$ et $D_\ell^4(v) = 1$. Ainsi, le mot v n'est pas lisse à droite, mais est lisse à gauche.

Remarque 5.3.9 Certains facteurs lisses ne sont ni lisse à droite, ni lisse à gauche. Par exemple, prenons le mot $u = 1212212211212$. En appliquant successivement l'opérateur D, on obtient

$D^1(u) = 11212211$

$D^2(u) = 21122$

$D^3(u) = 22$

$D^4(u) = 2.$

Ainsi, le mot u est lisse, mais n'est ni lisse à droite, ni lisse à gauche puisque

$D_r^1(u) = 111212211$ $\quad D_\ell^1(u) = 112122111$

$D_r^2(u) = 31122$ $\quad D_\ell^2(u) = 21123.$

La notion de facteur lisse est donc moins restrictive que les notions de préfixes et suffixes lisses.

Pour terminer cette section, rappelons certaines conjectures et propriétés des mots lisses infinis qui découlent de (Brlek et al., 2006),

Conjecture 5.3.10 *Tous les mots lisses infinis contiennent tous les facteurs finis lisses.*

Conjecture 5.3.11 *La fréquence des lettres* 1 *et* 2 *dans un mot lisse* $w \in \{1,2\}^\omega$ *est* $1/2$*.*

Proposition 5.3.12 *Il n'existe aucun mot lisse périodique.*

5.4 Graphe de De Bruijn

Dans cette section, nous introduisons les graphes de De Bruijn, aussi connus sous le nom de graphes de Rauzy, qui nous permettront ensuite de représenter les facteurs d'une longueur donnée pour un mot fixé.

Définition 5.4.1 Le *graphe de De Bruijn* d'ordre k associé au mot $w \in \mathcal{A}^\infty$ est un graphe orienté ayant des sommets étiquetés par les facteurs de $F_k(w)$ et ayant une arête étiquetée $\alpha \in \mathcal{A}$ partant du sommet y et allant au sommet z si et seulement si $z = y[1, k-1]\alpha$.

Exemple 5.4.2 Soit le *mot infini de Fibonacci* $\mathcal{F}$ défini par

$$\mathcal{F} = \lim_{n\to\infty} \mathcal{F}_n \quad \text{où} \quad \mathcal{F}_0 = 2, \quad \mathcal{F}_1 = 1 \quad \text{et} \quad \mathcal{F}_n = \mathcal{F}_{n-1}\mathcal{F}_{n-2}, \quad \forall n \geq 2.$$

Les premières valeurs de $\mathcal{F}_n$ sont $\mathcal{F}_2 = 12, \mathcal{F}_3 = 121, \mathcal{F}_4 = 12112, \mathcal{F}_5 = 12112121, \ldots$ et

$$\mathcal{F} = 12112121121121211212112112121121121211212112112121121211212112112\cdots$$

Il est bien connu que $\mathcal{F}$ est une suite sturmienne : donc pour tout n, $P(n) = n + 1$. On sait que $\mathcal{F}$ n'a que 4 facteurs de longueur 3 : $112, 121, 211, 212$. Ainsi, son graphe de De Bruijn d'ordre 3 est le graphe à 4 sommets illustré à la Figure 5.1.

Le principal désavantage de ce type de graphe est qu'il devient grand très rapidement. En effet, le nombre de sommets d'un graphe de De Bruijn d'ordre k d'un mot w est l'ensemble de ses facteurs de longueur k et ce nombre est de l'ordre de $\mathcal{O}(2^k)$. Il est toutefois possible de diminuer le nombre de sommets d'un graphe de De Bruijn en supprimant certains sommets et arêtes *superflus* et le graphe résultant est appelé le *graphe de De Bruijn réduit*.

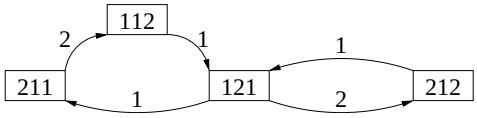


Figure 5.1 Graphe de De Bruijn d'ordre 3 du mot de Fibonacci.

Notation 5.4.3 Soit un graphe G. Pour un sommet x, pred(x) (resp. succ(x)) désigne l'ensemble des sommets y tel qu'il existe une arête qui relie le sommet y au sommet x (resp. de x à y). De plus, label(y, x) dénote l'étiquette de l'arête reliant le sommet y au sommet x et $d^+(x)$ désigne le nombre d'arêtes sortant du sommet x.

Définition 5.4.4 Soit $G = (S, A)$, le graphe de De Bruijn d'ordre k du mot $w \in \mathcal{A}^\infty$, où S est l'ensemble des sommets et $A \subseteq S \times \mathcal{A} \times S$ est l'ensemble des arêtes. Le *graphe de De Bruijn réduit* associé à G est le nouveau graphe $G' = (S', A')$, où $S' \subseteq S$ et $A' \subseteq S' \times \mathcal{A}^* \times S'$. Le graphe G' est obtenu de G en enlevant récursivement des arêtes et sommets *superflus* de G. C'est-à-dire

$$S' = S - \{x \in S : d^+(x) = 1 \text{ et } \text{succ}(x) \neq x\}$$

et pour tout sommet x supprimé, avec $z \in$ pred(x) et $y \in$ succ(x), les arêtes $(z, \text{label}(z, x), x)$ et $(x, \text{label}(x, y), y)$ sont supprimées de A et $(z, (\text{label}(z, x)\text{label}(x, y), y)$ y est ajoutée. Ce nouvel ensemble forme A'.

La Figure 5.2 illustre la suppression d'un sommet.

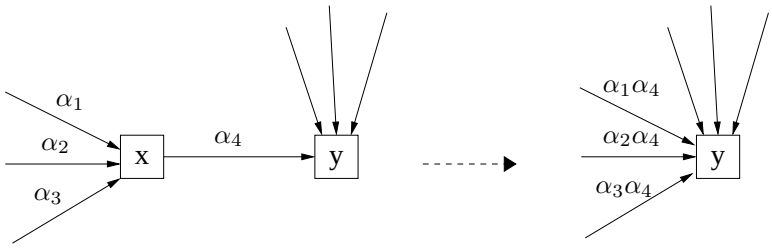


Figure 5.2 Suppression du sommet x d'un graphe de De Bruijn.

Remarque 5.4.5 L'ordre dans lequel les sommets sont supprimés n'a pas d'influence sur le graphe de De Bruijn réduit obtenu, puisque la concaténation des mots est associative. Remar-

quons que dans le graphe réduit, les étiquettes sont des mots de $\mathcal{A}^*$. Par exemple, le graphe de De Bruijn réduit pour le mot de Fibonacci est donné par la Figure 5.3

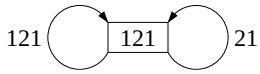


Figure 5.3 Graphe de De Bruijn réduit d'ordre 3 associé au mot de Fibonacci.

Les graphes de De Bruijn nous seront utiles dans ce qui suit, puisqu'ils permettent de représenter la structure d'un mot.

5.5 Calcul des mots extrémaux

Avant de montrer comment construire les mots extrémaux, voici leur définition.

Définition 5.5.1 Le *mot lisse infini minimal* (resp. *maximal*), noté m (resp. M) est le plus petit (resp. grand) mot lisse infini selon l'ordre lexicographique.

Lemme 5.5.2 *Pour deux mots $u, v \in \{a < b\}^\omega$, on a $u < v \iff \overline{u} > \overline{v}$.*

Preuve Par définition de $<$ sur les mots infinis, on a $u < v$ si et seulement si l'on peut écrire $u = pau'$ et $v = pbv'$, avec $a < b$. Supposons donc que $u = pau'$ et $v = pbv'$. Alors $\overline{v} = \overline{p}\overline{a}\overline{v'}$ et $\overline{u} = \overline{p}\overline{b}\overline{u'}$. Alors $\overline{u} > \overline{v}$. ∎

Le mot minimal et le mot maximal sont fortement liés. En effet, on a le lemme suivant.

Lemme 5.5.3 $M = \overline{m}$.

Preuve Le Lemme 5.5.2 nous indique que $u < v \iff \overline{u} > \overline{v}$. Ainsi, si $m < x$ pour tout mot $x \neq m$, alors on a aussi $\overline{m} > \overline{x}$. D'où $M = \overline{m}$. ∎

Cela signifie donc que le calcul de m nous donne directement le mot $M = \overline{m}$: il suffit d'échanger l'ordre des lettres sur l'alphabet. Dans ce qui suit, nous nous intéressons seulement au calcul du mot minimal, comme le mot maximal est ensuite obtenu directement.

5.5.1 Algorithme naïf

L'algorithme le plus simple qui calcule un préfixe du mot minimal m d'une longueur n donnée est l'algorithme naïf qui suit.

Algorithme 5.5.4

DÉBUT lisseMinimal

Entrée : longueurMax ;

0 : $m := 1$;

1 : **Boucle**

2 : **Si** estLisse $(m \cdot 1)$ **alors** m := $m \cdot 1$;

3 : **Sinon** m := $m \cdot 2$;

4 : **Fin si** ;

5 : **SORTIR QUAND** longueur(m)=longueurMax ;

6 : **Fin boucle**

FIN lisseMinimal

La condition *estLisse* est vérifiée en utilisant la dérivée à droite D_r et nous assure que le préfixe calculé est un préfixe d'au moins un mot lisse infini. On obtient alors

$$m[0, 23] = 112112212112122112112122$$

pour longueurMax $= 24$ et conséquemment,

$$M[0, 23] = 221221121221211221221211.$$

L'algorithme naïf consiste à construire le plus petit préfixe lisse possible pour une longueur k donnée. On tente de le prolonger par la lettre 1, on vérifie si le mot ainsi construit est toujours un préfixe lisse ou pas. Si le mot est toujours un préfixe lisse, alors on a ajouté la bonne lettre. Sinon, on remplace le 1 par un 2. Puis on répète. Comme l'opérateur D_r vérifie qu'on a un préfixe lisse, ce préfixe sera toujours prolongeable et donc, l'algorithme peut toujours nous fournir un préfixe de m aussi long que souhaité.

Proposition 5.5.5 *Si la Conjecture 5.3.11 est vraie, alors l'Algorithme 5.5.4 calcule un préfixe de m de longueur n en $\mathcal{O}(n^2 \log(n))$ opérations.*

Preuve Pour chaque lettre ajoutée au préfixe de m, on doit vérifier si le préfixe ainsi obtenu est le préfixe d'un mot lisse en utilisant l'opérateur D_r. En supposant que la Conjecture 5.3.11 sur la fréquence des lettres dans un mot lisse soit vraie, il y aurait en moyenne autant de 1 que de 2 dans les préfixes lisses. En appliquant la dérivée D_r sur un préfixe de longueur k, on obtient donc un mot d'une longueur $2k/3$ et le nombre de dérivées est $\log_{3/2} k$. On effectue $k \log_{3/2} k$ opérations pour un préfixe de longueur k. Ainsi, pour un préfixe de longueur n, le nombre d'opérations à effectuer est donné par

$$\sum_{k=1}^{n} k \log_{3/2} k < n^2 \log n.$$

D'où la conclusion. ■

5.5.2 Algorithme naïf amélioré

Afin de réduire le nombre d'applications de l'opérateur D_r dans l'Algorithme 5.5.4, il est naturel de vouloir ajouter plus d'une lettre à chaque étape. Pour ce faire, nous utilisons le graphe de De Bruijn réduit. En effet, en considérant le graphe de De Bruijn réduit du mot m pour un ordre donné, les prolongements possibles du préfixe lisse apparaissent plus clairement. Le nouvel algorithme proposé consiste à fixer l'ordre du graphe de De Bruijn à un $k \in \mathbb{N}$ et à chaque étape, selon les k dernières lettres du préfixe lisse, on connaît les prolongements possibles. On tente de mettre le plus petit des prolongements, on vérifie s'il est lisse et sinon, on tente de mettre le deuxième plus petit prolongement, et ainsi de suite. Formellement, on a l'algorithme suivant.

Algorithme 5.5.6

DÉBUT lisseMinimal2

Entrée : m, longueurMax, etatActuel ;

0 : **Si** longueur(m) $\geq$longueurMax RETOURNER(m) ;

1 : **Sinon**

2 : **Pour** $s \in$ succ(etatActuel) **faire** # selon l'ordre lexicographique croissant

3 : nouveau_m := m·label(etatActuel,s) ;

4 : **Si** estLisse(nouveau_m) **alors**

5 : lisseMinimal (nouveau_m, longueurMax, s) ;

6 : **Fin si** ;

7 : **Fin pour** ;

8 : **RETOURNER**(m) ;

9 : **Fin si** ;

FIN lisseMinimal2

Par exemple, pour trouver un préfixe de longueur 100 du mot lisse infini minimal, on utilise l'algorithme lisseMinimal avec les paramètres (112112, 100, etatInitial). La condition *estLisse* est encore vérifiée en utilisant l'opérateur de dérivée à droite D_r.

Le graphe de De Bruijn réduit d'ordre 6 pour le mot minimal m est représenté à la Figure 5.4.

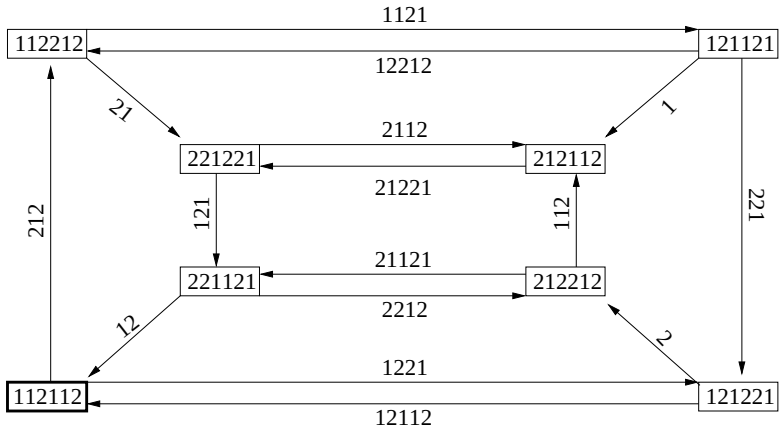


Figure 5.4 Graphe de De Bruijn d'ordre 6 réduit pour le mot m.

Dans le graphe d'ordre 6, il y a 8 sommets et la longueur moyenne des étiquettes des arêtes est de 3.375. Le Tableau 5.1 illustre la croissance du nombre de sommets et des longueurs des étiquettes des arêtes en fonction de l'ordre du graphe de De Bruijn réduit.

Ordre	Nombre de sommets	Longueur moyenne des étiquettes
1	2	1
2	2	1.5
3	4	2.25
4	4	2.25
5	4	2.25
6	8	3.375
10	12	4.667
15	16	5.062
20	32	7.594
25	36	8.611
50	128	17.09

Tableau 5.1 Longueur moyenne des étiquettes.

Proposition 5.5.7 *(Weakley, 1989) Le nombre de sommets d'un graphe de De Bruijn d'un mot lisse varie de façon polynomiale en fonction de son ordre, puisque le nombre de facteurs de longueur* n *est dans* $\mathcal{O}(n^{\frac{\log 3}{\log 2}})$.

La réduction du graphe de De Bruijn fait diminuer le nombre de sommets du graphe sans toutefois en faire diminuer la classe de complexité. Le Tableau 5.1 suggère que la longueur moyenne des étiquettes, c'est-à-dire le nombre moyen de lettres ajoutées à chaque itération de l'algorithme, croît de façon logarithmique en fonction de l'ordre du graphe. Ainsi, augmenter l'ordre du graphe augmente la longueur moyenne des étiquettes des arêtes, donc le nombre de lettres ajoutées par itération (amélioration de la performance), mais augmente aussi le nombre de sommets (diminution de la performance). Comme la longueur moyenne des étiquettes varie de façon logarithmique et le nombre de sommets de façon polynomiale, cela signifie que pour un ordre plus grand, il y a plus de comparaisons à effectuer pour trouver le bon prolongement et alors, la performance obtenue par la croissance de la longueur moyenne est négligeable. Dans le but d'obtenir un algorithme plus efficace que l'Algorithme 5.5.4, on doit donc trouver le meilleur rapport entre le nombre de sommets et la longueur moyenne des étiquettes des arêtes. L'algo-

rithme a été programmé en *Maple* pour des graphes d'ordre 1, 10, 20, 30, 40, 50, 55, 60 et 65. Expérimentalement, pour un ordre plus petit ou égal à 50, l'augmentation de l'ordre améliore la performance en temps de l'algorithme. Pour un ordre plus grand que 50, l'algorithme devient moins rapide. L'algorithme le plus performant inspiré de l'algorithme naïf serait donc celui qui utilise un graphe de De Bruijn réduit d'ordre 50.

5.5.3 Algorithme utilisant la bijection Φ

En utilisant la bijection Φ, il est possible de trouver un algorithme plus performant pour calculer m. Comme m est en bijection avec $\Phi(m)$, l'algorithme consiste à construire un préfixe du mot $\Phi(m)$ plutôt que de construire un préfixe du mot m.

Algorithme 5.5.8
DÉBUT lisseMinimal3
Entrée : longueurMax ;
$0:$ phim$= 1$;
$1:$ **Tant que** longueur(phim)$<$ longueurMax **faire**
$2:$ **Si** phiInv(phim$\cdot 1$) $<$ phiInv(phim$\cdot 2$) **alors** phim := phim$\cdot 1$;
$3:$ **Sinon** phim := phim$\cdot 2$;
$4:$ **Fin si** ;
$5:$ **Fin tant que** ;
$6:$ **RETOURNER** phiInv(phim) ;
FIN lisseMinimal3

Dans cet algorithme, phim désigne le mot préfixe de $\Phi(m)$, phiInv est la fonction ϕ^{-1} qui calcule les coups forcés.

Proposition 5.5.9 *Si la Conjecture 5.3.11 est vraie, alors l'Algorithme 5.5.8 s'effectue en $\mathcal{O}(n^2)$ opérations, où n est la longueur du préfixe de m calculé.*

Preuve En supposant la Conjecture 5.3.11 vraie, on a que pour un préfixe p de m de longueur k, $\Phi(p)$ est de longueur environ $\log_{3/2} k$. Considérons $\Phi(p)$. Afin d'ajouter une nouvelle lettre

à $\Phi(p)$, on doit calculer $\phi^{-1}(\Phi(p)\cdot 1)$ et $\phi^{-1}(\Phi(p)\cdot 2)$. Construire le mot obtenu par la fonction ϕ^{-1} prend $\displaystyle\sum_{i=1}^{\log_{3/2} k}\left(\frac{3}{2}\right)^i$ opérations. Il faut faire ce calcul pour les deux lettres possibles et ensuite comparer les deux mots de longueur k ainsi obtenus. La comparaison se fait en $2k$ opérations. Ainsi, le coût en nombre d'opérations pour ajouter la k-ième lettre au préfixe de m est

$$2\sum_{i=1}^{\log_{3/2} k}\left(\frac{3}{2}\right)^i + 2k.$$

Pour terminer, il suffit de faire varier k de 1 à n afin de construire un mot $\Phi(p)$ (resp. p) de longueur $\log_{3/2} k$ (resp. n). On a alors

$$\begin{aligned}
\sum_{k=1}^{n}(2\sum_{i=1}^{\log_{3/2} k}\left(\frac{3}{2}\right)^i + 2k) &= 2\sum_{k=1}^{n}\sum_{i=1}^{\log_{3/2} k}\left(\frac{3}{2}\right)^i + 2\sum_{k=1}^{n} k \\
&= 2\sum_{k=1}^{n}\left(\frac{3/2(1-(3/2)^{\log_{3/2} k})}{1-3/2}\right) + n(n+1) \\
&= 2\sum_{k=1}^{n}(-3(1-k)) + n(n+1) \\
&= 6\sum_{k=1}^{n}(k-1) + n(n+1) \\
&= 6n(n+1)/2 - 6n + n(n+1)
\end{aligned}$$

Le calcul d'un préfixe de longueur n du mot m avec l'Algorithme 5.5.8 s'effectue donc en $\mathcal{O}(n^2)$. ∎

5.6 Étude des dérivées successives et des factorisations de Lyndon de m et M

Dans cette section, nous étudions la structure de m et M en analysant leurs dérivées successives et leur factorisation de Lyndon. Les dérivées successives des mots lisses nous permettent de construire le mot $\Phi(m)$ (resp. $\Phi(M)$) et comme Φ est une fonction bijective, si le mot $\Phi(m)$ (resp. $\Phi(M)$) a des propriétés régulières, cela nous permettrait de mieux caractériser m (resp. M). Par ailleurs, l'étude des factorisations de Lyndon est motivée par la propriété de minimalité et de maximalité des mots extrémaux. Intuitivement, leur factorisation de Lyndon utiisant des propriétés d'ordre lexicographique pourraient avoir des propriétés intéressantes.

En calculant les premières dérivées de m en utilisant l'opérateur Δ sur m, on obtient

$\Delta^0(m) = 1121122121121221121121222121122121122112121122121122121122112122112\cdots$
$\Delta^1(m) = 2122112112212112112212212112212211211212212112212212112112212211212212111\cdots$
$\Delta^2(m) = 1122122112122121122122121121122121122121122112121221122112212112122122112\cdots$
$\Delta^3(m) = 2212211211221211212211211212212112212211211212211211221211211221221211211\cdots$
$\Delta^4(m) = 2122122112112212112112212212112212211211212212112112212211212212211211212221\cdots$
$\Delta^5(m) = 1121221221121221211221221121121221221121221211221221121122121121122122121\cdots$
$\Delta^6(m) = 2112122112112212212112122112112212112112212212112112212112112122122112122122\cdots$
$\Delta^7(m) = 1211221222121121221221121221211211221211212211211221221121122121121122\cdots$
$\Delta^8(m) = 1122121121221221121121221121122121122121122121122121122121121122122112\cdots$
$\Delta^9(m) = 2211211212212112212211211212211211221211212212211211212211211221221211211\cdots$
$\Delta^{10}(m) = 2212112112212212112212211211212212112212212112112212112121221121122121211\cdots$
$\Delta^{11}(m) = 2112122121122122121121122121122112112212212112122112112122122112112212\cdots$
$\Delta^{12}(m) = 1211211221211212211211221221211211221211212212211211221211211221221211211\cdots$
$\Delta^{13}(m) = 1121221121122122121121221121121221221121221211211221221211221122112122122\cdots$
$\Delta^{14}(m) = 2112212212112112212112122112112122121122112122112121212\cdots$
$\Delta^{15}(m) = 1221211212211211221211211221221121\cdots$
$\Delta^{16}(m) = 1211211221221121212211\cdots$
$\Delta^{17}(m) = 1121221221121221\cdots$
$\Delta^{18}(m) = 2112122112\cdots$

Comme m et M sont complémentaires, $\Delta^i(m) = \Delta^i(M)$ pour $i \geq 1$. Ainsi, on obtient la proposition suivante.

Proposition 5.6.1 $\Phi(m) = 12122121122211211121122\cdots$ *et* $\Phi(M) = 21^{-1}\Phi(m)$

Intuitivement, par leur définition, les mots extrémaux devraient avoir des propriétés intéressantes par rapport à l'ordre lexicographique. Dans ce qui suit, nous étudions donc les factorisations de Lyndon des mots extrémaux. Les premiers facteurs de la factorisation de m sont

$$
\begin{aligned}
\ell_0(m) &= 112112212112122, \\
\ell_1(m) &= 112112\underline{1}22121122122, \\
\ell_2(m) &= 1121121221\underline{1}2112212112122122112112212211211221221121122121121\cdots \\
\ell_3(m) &= 1121121221121122121121221221121121221121122122121121221121121\underline{1}221221\cdots \\
\ell_4(m) &= 1121121221121122121121221221121121221121122122121121221121121221221\cdots \\
\ell_5(m) &= 1121121221121122121121\underline{1}221221121221211211221221211221221121221211211\cdots \\
\ell_6(m) &= 1121121221121122121121122122112121221121122122121122122112\underline{1}2122121\cdots \\
\ell_7(m) &= 1121121221121122121121122122112121221121122122121122122112112122121\cdots \\
\ell_8(m) &= 1121121221121122121121122122112121221121121122121\underline{1}12122112112212212112122121121\cdots
\end{aligned}
$$

$$
\begin{aligned}
\ell_9(m) &= 1121121221121122121121122122112121221121122121121221121122112212212112\underline{1}\cdots\\
\ell_{10}(m) &= 11211212211211221211211221221121221211211221211212211211221221211211 2\cdots\\
\ell_{11}(m) &= 11211212211211221211211221221121221211211221211212211211221221211211 2\cdots\\
\ell_{12}(m) &= 11211212211211221211211221221121221211211221211212211211221221211211 2\cdots
\end{aligned}
$$

et les longueurs sont respectivement

$$15, 18, 180, 910, 382, 1948, 2961, 490, 1703, 2359, 2194, 4679, 7278.$$

Les premiers facteurs de la factorisation de Lyndon de M sont

$$
\begin{aligned}
\ell_0(M) &= 2,\\
\ell_1(M) &= 2,\\
\ell_2(M) &= \underline{1}22,\\
\ell_3(M) &= 1\underline{1}2122121122122122,\\
\ell_4(M) &= 112122,\\
\ell_5(M) &= 1121\underline{1}22122121122122112122122,\\
\ell_6(M) &= 112112212212112122,\\
\ell_7(M) &= 112112\underline{1}22121122122112122122112112122121122122121122112212112122122112\cdots\\
\ell_8(M) &= 112112212212112122,\\
\ell_9(M) &= 1121121221\underline{1}2112212112122122112112212112112212212211221221121\cdots\\
\ell_{10}(M) &= 1121121221121122121121221221121112212211211221222121\underline{1}2211211212212211\cdots\\
\ell_{11}(M) &= 1121121221121122121121\underline{1}22122112122121122122121121122122112212212211\cdots\\
\ell_{12}(M) &= 1121121221121122121121122122112122121121\underline{1}12212212112212211221221221122\cdots\\
\ell_{13}(M) &= 1121121221121122121121122122112122121121122121\underline{1}12122112112212212112112\cdots,
\end{aligned}
$$

et les longueurs sont respectivement

$$1, 1, 3, 17, 6, 27, 19, 80, 18, 180, 268, 1753, 2107, 816.$$

Les lettres soulignées représentent la première position pour laquelle le facteur diffère du facteur précédent. Ces factorisations suggèrent que le préfixe commun de deux mots de Lyndon consécutifs dans la factorisation de Lyndon de m et de M est de longueur croissante.

Cela laisse donc supposer l'existence d'un mot qui soit un suffixe de m et de M qui soit minimal.

Définition 5.6.2 On note s, le *suffixe lisse minimal,* c'est-à-dire le plus petit mot selon l'ordre lexicographique qui soit un suffixe d'un mot lisse.

On construit le mot s en utilisant l'Algorithme 5.5.4 modifié comme suit : à chaque fois qu'on ajoute une nouvelle lettre au préfixe de s, on vérifie si s est toujours un suffixe lisse en appliquant l'opérateur de dérivée à gauche D_ℓ. On obtient ainsi le suffixe lisse

$$s = 1121121221121122121121122122112121221121122121121221121122121\cdots$$

Lemme 5.6.3 *Le mot s n'est pas un mot lisse.*

Preuve En effet,

$\Delta(s) = 2121122122112121221221121121\cdots$

$\Delta^2(s) = 11122122112122121\cdots$

$\Delta^3(s) = 3212211211\cdots$

Donc s est un suffixe lisse, mais n'est pas un mot lisse. ■

Proposition 5.6.4 *Le suffixe lisse minimal s est un mot de Lyndon infini.*

Preuve Par la définition de mot de Lyndon infini, il suffit de montrer que s est strictement plus petit que tous ses suffixes. Comme s est le suffixe lisse minimal, s'il existe un suffixe plus petit que lui-même, il y a contradiction. S'il existe un suffixe qui lui est égal, alors cela impliquerait que s est périodique (voir Figure 5.5), ce qui est impossible, par la Proposition 5.3.12. ■

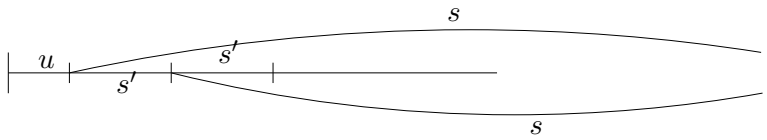


Figure 5.5 Schéma de la preuve de la Proposition 5.6.4.

Le Tableau 5.2 indique la position i pour laquelle le mot s diffère de $\ell_k(m)$ et $\ell_k(M)$. Dans ce tableau, $?$ signifie que nous n'avons pas calculé suffisament de facteurs pour connaître la position. Le calcul d'un préfixe de longueur 50 000 n'était pas suffisant pour obtenir le 15-ième facteur de la factorisation de Lyndon de m. Observons que dans ce tableau, pour $k \to \infty$, $\ell_k(m)$ et $\ell_k(M)$ semblent converger vers s. On pourrait être tenté de conjecturer que m et M terminent par le suffixe s, mais c'est impossible, comme l'énonce le lemme qui suit.

k	Position dans $f_k(m)$	Position dans $f_k(M)$
1	7	2
2	11	2
3	23	2
4	23	5
5	23	5
6	46	7
7	46	7
8	46	11
9	68	11
10	80	23
11	80	23
12	80	40
13	142	46
14	142	80
15	?	248

Tableau 5.2 Positions pour lesquelles $s[i] \neq f_k(m)[i]$ et $s[i] \neq f_k(M)[i]$.

Lemme 5.6.5 *m et M ne peuvent pas tout deux avoir le suffixe s.*

Preuve Supposons que $m = us$ et $M = vs$, avec $\lg(u) < \lg(v)$. On peut alors écrire $up = v$, pour un préfixe p de s. Mais si s commence par p et que $M = \overline{m}$, alors $m = up\overline{p}s'$, pour $s' \in \mathcal{A}^\omega$. Cela implique que M a vp comme préfixe. Ainsi de suite, comme l'illustre la Figure 5.6, on trouve que $m = u(p\overline{p})^\omega$ et $M = v(p\overline{p})^\omega$. Donc m et M sont ultimement périodiques. Par la Proposition 5.3.12, on sait que c'est imposible. Si $\lg(u) > \lg(v)$, la preuve est similaire. Si $\lg(u) = \lg(v)$, alors on trouve que $s = \overline{s}$, ce qui est impossible. ■

Le Lemme 5.6.5 donne l'idée de la conjecture suivante.

Conjecture 5.6.6 $\lim_{i\to\infty} \ell_i(m) = s$ *ou (exclusif)* $\lim_{i\to\infty} \ell_i(M) = s$.

Étudions maintenant plus particulièrement la factorisation de Lyndon de m et M. Rappelons

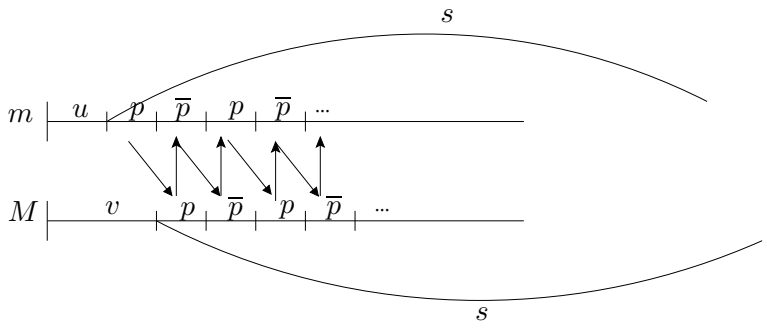


Figure 5.6 Schéma de la preuve du Lemme 5.6.5.

d'abord un résultat sur le nombre de carrés.

Lemme 5.6.7 *(Brlek et al., 2006) Le nombre de facteurs carrés des mots lisses est fini :* $\text{Card}(\text{Carrés}(\mathcal{K})) = 46$.

Corollaire 5.6.8 *(Brlek et al., 2006) Les mots lisses sont sans cube.*

Il en découle que $\text{Card}(\text{Carrés}(m)) \leq 46$ et parmi les 46 carrés possibles, ceux qui sont des carrés de mots de Lyndon sont

$$1^2, (112)^2, (112112212)^2, (1121122122121122122112122122)^2, (122)^2 \text{ et } 2^2.$$

Proposition 5.6.9 *Soit $m = \ell_1(m)\ell_2(m)\cdots$, la factorisation de Lyndon de m. Alors*
i) *$\ell_1(m)$ est le seul facteur de Lyndon qui est un préfixe lisse ;*
ii) *$\ell_i(m) > \ell_{i+1}(m)$, pour tout i.*

Preuve i) Par la définition de factorisation de Lyndon, on sait que $\ell_1(m) \geq \ell_i(m)$, pour $i > 0$. Précédemment, on a calculé les premiers facteurs de la factorisation de Lyndon de m et on a obtenu que $\ell_1(m) > \ell_2(m)$. Ainsi, pour tout $i \geq 1$, $\ell_0(m) > \ell_i(m)$. Mais par la construction de m, aucun préfixe lisse n'est plus petit que m. Par conséquent, un préfixe lisse plus petit que $\ell_1(m)$ n'existe pas et alors, pour tout $\ell_i(m) < m$, $\ell_i(m)$ ne peut être un préfixe lisse.

ii) Supposons maintenant qu'il existe $i \geq 2$ tel que $\ell_i(m) = \ell_{i+1}(m)$. Ainsi, cela signifierait qu'il y a un facteur puissance carré d'un mot de Lyndon. En ordonnant les facteurs de mots

lisses qui sont des carrés de mots de Lyndon, on obtient

$$1^2 < (112)^2 < (112112212)^2 < (1121122122121122122112122 12)^2 < (122)^2 < 2^2. \quad (5.5)$$

Comme les seuls carrés $\leq \ell_1(m)$ sont $(112)^2$ et 1^2, pour avoir $\ell_i(m) = \ell_{i+1}(m)$, il faudrait que $\ell_i(m) = 112$ ou $\ell_i(m) = 1$. Dans les deux cas, cela impliquerait que m est ultimement périodique : contradiction. ■

On obtient un résultat semblable pour le mot lisse maximal M.

Proposition 5.6.10 *Dans la factorisation de Lyndon de M,*
i) $\ell_0(M)$, $\ell_1(M)$, $\ell_2(M)$, $\ell_4(M)$, $\ell_5(M)$, $\ell_6(M)$ *sont les seuls préfixes lisses ;*
ii) $\ell_i(M) > \ell_{i+1}(M)$, *pour tout* $i \geq 1$.

Preuve i) On vérifie directement que les facteurs de Lyndon $\ell_i(M)$, pour $i \leq 6$, sont lisses ou pas en utilisant l'opérateur D_r. Pour $i > 6$, il suffit de vérifier que $\ell_7(M) < \ell_0(m)$.

ii) En observant que $\ell_7(M) < (112112212)^2$ dans la formule (5.5), on conclut en utilisant le même raisonnement que dans la preuve de la Proposition 5.6.9. ■

Proposition 5.6.11 *Le mot lisse infini minimal m n'est pas un mot de Lyndon.*

Preuve Le mot lisse infini minimal a le suffixe $1121121221211\cdots$ plus petit que lui-même et commençant à la position 15 . ■

Si tous les mots lisses infinis contiennent tous les facteurs lisses finis comme le prétend la Conjecture 5.3.10, alors aucun mot lisse infini n'est un mot de Lyndon. En effet, supposons qu'un mot lisse infini w contiennent tous les facteurs lisses et supposons que w soit un mot de Lyndon infini. Remarquons qu'un mot peut être un facteur lisse, mais non pas un mot lisse. L'inverse est faux. Il est donc plus restrictif d'être un mot lisse qu'un facteur lisse. On peut ainsi se convaincre qu'il serait toujours possible de trouver un facteur lisse plus petit que w, et donc que w ne pourrait pas être un mot de Lyndon. La Proposition 5.6.11 vient donc renforcer la Conjecture 5.3.10.

Récemment, les mots lisses ont été généralisés à différents alphabets à deux lettres ou plus. Dans le chapitre suivant, nous montrons comment le changement des lettres de l'alphabet influence les propriétés des mots lisses extrémaux et contrairement aux mots lisses extrémaux sur l'alphabet $\{1, 2\}$, nous verrons que certaines régularités apparaissent.

Chapitre VI

MOTS LISSES EXTRÉMAUX GÉNÉRALISÉS

Dans ce chapitre, nous utilisons la notion introduite dans (Berthé, Brlek et Choquette, 2005) de mot lisse généralisé à différents alphabets. Ainsi, nous considérons les mots lisses extrémaux sur différents alphabets à deux lettres. Dans un premier temps, nous étudions les mots lisses extrémaux sur l'alphabet $\{1, 3\}$. Il apparaît que de remplacer le 2 pour un 3 dans l'alphabet change grandement la structure des mots lisses extrémaux et même, que cela leur donne des propriétés beaucoup plus intéressantes. En observant ainsi que de considérer les mots lisses extrémaux sur l'alphabet $\{1, 3\}$ plutôt que sur $\{1, 2\}$ en influence grandement les propriétés, nous nous intéressons ensuite aux mots lisses extrémaux sur des alphabets de lettres de même parité. Ainsi, pour un alphabet ordonné $\{a < b\}$ fixé, avec $a, b \in \mathbb{N} - \{0\}$ et a, b de même parité, nous étudions les mots lisses extrémaux. Nous donnons des algorithmes rapides pour les construire, nous montrons des propriétés de récurrence, de fermeture pour les ensembles de facteurs et nous fournissons une formule pour la fréquence des lettres pour certains types d'alphabets. Nous nous intéressons ensuite aux factorisations de Lyndon et montrons que contrairement au mot lisse minimal sur l'alphabet $\{1, 2\}$, il existe des mots lisses minimaux qui sont des mots de Lyndon. Nous montrons qu'il existe un lien entre les mots minimaux sur des alphabets pairs et certains mots de Kolakoski généralisés. Ainsi, nous montrons des propriétés pour les mots de Kolakoski généralisés qui sont toujours des problèmes ouverts sur l'alphabet $\{1, 2\}$. Les résultats de ce chapitre font l'objet de la publication (Brlek, Jamet et Paquin, 2008).

6.1 Quelques définitions et résultats préalables

Rappelons une propriété utile sur les mots de Lyndon.

Lemme 6.1.1 (Chapitre 5.1 de (Lothaire, 1983)) *Soient $u, v \in \mathcal{L}$. Alors $uv \in \mathcal{L}$ si et seulement si $u < v$.*

Lemme 6.1.2 *Soient $u, v \in \mathcal{L}$ et supposons $u < v$. Alors $uv < v$.*

Preuve Si $u \notin \text{Pref}(v)$, alors c'est évident. Supposons donc que $u \in \text{Pref}(v)$ et écrivons $v = u^k v'$, avec k maximal. Ainsi, u ne peut pas être préfixe de v'. Comme v est un mot de Lyndon, v est plus petit que tous ses suffixes. Plus particulièrement, $v < v'$. Ainsi, $u < v < v'$. Comparons $v = u^k v'$ et $uv = u^{k+1} v'$: la comparaison se fait entre v' et u. Comme $u < v'$ et $u \notin \text{Pref}(v')$, on conclut que $uv < v$. ■

Corollaire 6.1.3 *Soient $u, v \in \mathcal{L}$, avec $u < v$. Alors uv^n, $u^n v \in \mathcal{L}$, pour tout $n \geq 0$.*

Preuve Il suffit d'appliquer le Lemme 6.1.1 n fois. En effet, si $u, v \in \mathcal{L}$ et $u < v$, alors $uv \in \mathcal{L}$. Ainsi, $u, uv \in \mathcal{L}$ et $u < uv$ et donc, $u^2 v \in \mathcal{L}$. De la même façon, considérons $uv, v \in \mathcal{L}$. Par le Lemme 6.1.2, $uv < v$, et par le Lemme 6.1.1, on a que $uv^2 \in \mathcal{L}$. ■

À moins d'indication contraire, dans ce chapitre, nous ne considérons que des alphabets $\mathcal{A} = \{a < b\}$, avec $a, b \in \mathbb{N} - \{0\}$.

Toutes les notions vues dans le Chapitre V se généralisent naturellement aux mots lisses sur un alphabet $\{a < b\}$. Ainsi, la fonction de codage par blocs reste bien définie si l'on change l'alphabet $\{1, 2\}$ pour $\{1, 3\}$, $\{2, 3\}$, $\{3, 7\}$, etc.

Exemple 6.1.4 Soit $w = 223322333222333$. En appliquant successivement l'opérateur Δ, on obtient

$\Delta^0(w) = 223322333222333$

$\Delta^1(w) = 222333$

$\Delta^2(w) = 33$

$\Delta^3(w) = 2$.

Exemple 6.1.5 Soit $u = 22444242222444$. En appliquant successivement l'opérateur Δ, on obtient

$\Delta^0(u) = 22444242222444$

$\Delta^1(u) = 231143$

$\Delta^2(u) = 11211$

$\Delta^3(u) = 212$

$\Delta^4(u) = 111$

$\Delta^5(u) = 3.$

La famille des mots lisses peut donc être généralisée à différents alphabets, comme l'ont fait les auteurs de (Berthé, Brlek et Choquette, 2005).

Définition 6.1.6 Soit un alphabet ordonné numérique $\mathcal{A}$. On définit la famille des *mots lisses infinis sur l'alphabet* $\mathcal{A}$ par l'ensemble

$$\mathcal{K}_{\mathcal{A}} = \{w \in \mathcal{A}^{\omega} \mid \forall k \in \mathbb{N}, \Delta^k(w) \in \mathcal{A}^{\omega}\}.$$

Définition 6.1.7 Soit un alphabet ordonné numérique $\mathcal{A}$. Un mot fini $w \in \mathcal{A}^*$ est *lisse* s'il existe k tel que $\lg(\Delta^k(w)) = 1$ et $\Delta^k(w) \leq \beta$, où β est la lettre maximale de $\mathcal{A}$, et pour tout $j < k$, on a $\Delta^j(w) \in \mathcal{A}^*$.

Ainsi, le mot w de l'Exemple 6.1.4 est un mot lisse sur l'alphabet $\{2, 3\}$, tandis que le mot u de l'Exemple 6.1.5 n'est pas un mot lisse sur l'alphabet $\{2, 4\}$, mais l'est sur l'alphabet $\{1, 2, 3, 4\}$.

Les fonctions pseudo-inverses de Δ demeurent bien définies pour un alphabet à 2 lettres autre que l'alphabet $\{1, 2\}$.

Exemple 6.1.8 Soit $u = 44554544445$. Alors

$$\Delta_4^{-1}(u) = 444455554444455555444455555444455554444555544444$$

et

$$\Delta_5^{-1}(u) = 555544445555544444555544444555544445555444455555.$$

Définition 6.1.9 Les *mots de Kolakoski généralisés* sont définis comme étant les points fixes sous l'opérateur Δ. Pour un alphabet à deux lettres $\{a, b\}$, il y a exactement deux points fixes : $K_{(a,b)}$ et $K_{(b,a)}$, où $K_{(a,b)}$ le mot de Kolakoski généralisé sur l'alphabet $\{a, b\}$ et commençant par la lettre a.

Exemple 6.1.10 Le mot de Kolakoski sur l'alphabet $\mathcal{A} = \{1, 2\}$ et commençant par la lettre 2 est $K = K_{(2,1)}$. Le mot de Kolakoski $K_{(2,3)} = 2233222333223322333222\cdots$ et $K_{(3,1)} = 3331113331313331113331\cdots$.

La bijection Φ définie au Chapitre V se généralise naturellement à tout alphabet numérique ordonné à deux lettres.

Exemple 6.1.11 Pour le mot $w = 13331113331333111333131333111333133113331$, on a

$\Delta^0(w) = 13331113331333111333131333111333133113331$

$\Delta^1(w) = 1333133311133313331$

$\Delta^2(w) = 131333131$

$\Delta^3(w) = 1113111$

$\Delta^4(w) = 313$

$\Delta^5(w) = 111$

$\Delta^6(w) = 3.$

Donc w est un mot lisse sur l'alphabet $\{1, 3\}$ et $\Phi(w) = 1111313$.

Exemple 6.1.12 Soit le mot w pour lequel $u = \Phi(w) = 3533$. On retrouve w en appliquant Φ^{-1}, donc en itérant Δ^{-1} sur les lettres du mot u.

$\Delta^0(w) = 33333555553333355533355533333555533333$

$\Delta^1(w) = 555333555$

$\Delta^2(w) = 333$

$\Delta^3(w) = 3.$

Tout comme pour l'alphabet $\{1, 2\}$, il est possible de *forcer les coups* pour un alphabet $\{a < b\}$ fixé.

Définition 6.1.13 Soit l'alphabet $\mathcal{A} = \{a < b\}$ et $x \in \mathcal{A}^*$. Alors

$$\delta_\alpha^{-1}(x) = \Delta_\alpha^{-1}(x) \cdot \beta^a,$$

où β est la lettre complémentaire de la dernière lettre de $\Delta_\alpha^{-1}(x)$ et $\alpha, \beta \in \mathcal{A}$. La fonction ϕ^{-1} est définie de façon analogue à Φ, mais utilise δ^{-1} plutôt que Δ^{-1}.

Exemple 6.1.14 Dans l'exemple précédent, si l'on applique ϕ^{-1}, on obtient

$\delta^0(w) = 33333555553333355533355533333555553333355533355533355533333$
$55555333335555533333555333555333555333333555533333555$

$\delta^1(w) = 55533355533333555553333355$

$\delta^2(w) = 333555$

$\delta^3(w) = 3.$

Les opérateurs D, D_ℓ et D_r se généralisent aussi à des alphabets à deux lettres comme l'énoncent les trois définitions suivantes.

Définition 6.1.15 Pour un alphabet $\{a < b\}$, on définit l'opérateur D comme

$$D(w) = \begin{cases} \varepsilon & \text{si } \Delta(w) < b \text{ ou } w = \varepsilon, \\ \Delta(w) & \text{si } \Delta(w) = bxb \text{ ou } \Delta(w) = b, \\ bx & \text{si } \Delta(w) = bxu, \\ xb & \text{si } \Delta(w) = uxb, \\ x & \text{si } \Delta(w) = uxv, \end{cases}$$

où u et v sont des blocs de longueur $< b$. Appliquer l'opérateur D correspond à éliminer de $\Delta(w)$ les blocs préfixes ou suffixes de longueur plus petite que la lettre maximale, s'ils existent.

Définition 6.1.16 Un mot fini w est appelé *un facteur lisse* s'il existe k tel que $D^k(w) = \varepsilon$ et $\forall j < k,\ D^j(w) \in \mathcal{A}^*$.

Définition 6.1.17 La *dérivée à droite* est définie par la fonction $D_r : \mathcal{A}^* \to \mathbb{N}^*$ telle que

$$D_r(w) = \begin{cases} \varepsilon & \text{si } \Delta(w) < b \text{ ou } w = \varepsilon, \\ \Delta(w) & \text{si } \Delta(w) = xb, \\ x & \text{si } \Delta(w) = xu, \end{cases}$$

où u est un bloc de longueur $< b$.

Définition 6.1.18 La *dérivée à gauche* est définie par la fonction $D_\ell : \mathcal{A}^* \to \mathbb{N}^*$ telle que

$$D_\ell(w) = \begin{cases} \varepsilon & \text{si } \Delta(w) < b \text{ ou } w = \varepsilon, \\ \Delta(w) & \text{si } \Delta(w) = bx, \\ x & \text{si } \Delta(w) = ux, \end{cases}$$

où u est un bloc de longueur $< b$.

Définition 6.1.19 Un mot w est *lisse à droite* (resp. *lisse à gauche*) sur l'alphabet $\mathcal{A}$ s'il existe k tel que $D_r^k(w) = \varepsilon$ (resp. $D_\ell^k(w) = \varepsilon$) et pour tout $j < k$, $D_r^j(w) \in \mathcal{A}^*$ (resp. $D_\ell^j(w) \in \mathcal{A}^*$). Un mot lisse à droite est alors appelé un *préfixe lisse* et un mot lisse à gauche est appelé un *suffixe lisse*.

Exemple 6.1.20 Soit $w = 223223322333222$. En appliquant successivement les opérateurs D_r et D_ℓ sur le mot w, on obtient

$D_r^0(w) = 223223322333222$ $\quad D_\ell^0(w) = 223223322333222$

$D_r^1(w) = 2122233$ $\quad D_\ell^1(w) = 122233$

$D_r^2(w) = 113$ $\quad D_\ell^2(w) = 32$

$D_r^3(w) = 2$ $\quad D_\ell^3(w) = 1.$

Dès les premières dérivées, il est clair que le mot w n'est ni lisse à droite, ni lisse à gauche sur l'alphabet $\{2, 3\}$, mais qu'il l'est sur l'alphabet $\{1, 2, 3\}$.

Exemple 6.1.21 Soit $v = 33344433344443333444433344433$. On obtient

$D_r^0(v) = 33344433344443333444433344433$ $\quad D_\ell^0(v) = 33344433344443333444433344433$

$D_r^1(v) = 33344433$ $\quad D_\ell^1(v) = 33444333$

$D_r^2(v) = 33$ $\quad D_\ell^2(v) = 33.$

Le mot v est donc un préfixe lisse et un suffixe lisse sur l'alphabet $\{3, 4\}$.

Définition 6.1.22 Le *mot lisse infini minimal* (resp. *maximal*) sur l'alphabet $\{a < b\}$, noté $m_{\{a,b\}}$ (resp. $M_{\{a,b\}}$) est le plus petit (resp. grand) mot lisse infini selon l'ordre lexicographique.

L'algorithme 5.5.4 du Chapitre V qui calcule le mot minimal se généralise pour différents alphabets de la façon suivante.

Algorithme 6.1.23

DÉBUT lisseMinimalGeneral

Entrée : $\mathcal{A} = \{a, b\}$, longueurMax ;

0 : $m_{\{a,b\}} := a$;

1 : **Boucle**

2 : **Si** estLisse $(m_{\{a,b\}} \cdot a)$ **alors** $m_{\{a,b\}}$:= $m_{\{a,b\}} \cdot a$;

3 : **Sinon** $m_{\{a,b\}}$:= $m_{\{a,b\}} \cdot b$;

4 : **Fin si** ;

5 : **SORTIR QUAND** longueur($m_{\{a,b\}}$)=longueurMax ;

6 : **Fin boucle ;**

FIN lisseMinimalGeneral

Dans cet algorithme, la condition *estLisse* est vérifiée avec l'opérateur D_r. Notons que l'Algorithme 6.1.23 ne dépend pas de la parité des lettres de l'alphabet. Pour différents alphabets et pour longueurMax $= 47$, on obtient les mots

$$m_{\{1,2\}}[0, 46] = 11211221211212211211212212112212211211212211211,$$
$$M_{\{1,2\}}[0, 46] = 21211221211212211211212212112212211211212211211,$$
$$m_{\{1,3\}}[0, 46] = 11131113131113111313111313111311131311131113131,$$
$$M_{\{1,3\}}[0, 46] = 33133331313331333131333131333133331333133313313,$$
$$m_{\{2,4\}}[0, 46] = 22224444222244442244224422224444222244442244224,$$
$$m_{\{3,5\}}[0, 46] = 33333555553333355533355533333555553333355533355,$$
$$m_{\{2,3\}}[0, 46] = 22233322233223322233322233223333222233223332233,$$
$$m_{\{3,4\}}[0, 46] = 33334444333344433344433334444333344433344433334.$$

Si la Conjecture 5.3.11 est vraie, le calcul d'un préfixe de longueur n de $m_{\{a,b\}}$ avec l'Algorithme 6.1.23 nécessite $\mathcal{O}(n^2 \log(n))$ opérations. Afin de réduire le nombre d'applications de l'opérateur D_r, il serait pertinent d'ajouter plus d'une lettre à chaque étape, comme au Chapitre V, en utilisant les graphes de De Bruijn réduits. Toutefois, nous ne nous y attardons pas, puis-

qu'il existe des algorithmes beaucoup plus efficaces pour les alphabets de même parité, comme nous le montrons dans les sections suivantes.

6.2 Mot de Fibonacci et mot lisse minimal sur l'alphabet $\{1,3\}$

Dans cette section, nous étudions les mots lisses extrémaux sur l'alphabet $\{1,3\}$. Nous montrons que ces mots ont une structure plus régulière que les mots lisses extrémaux sur l'alphabet $\{1,2\}$. De plus, nous montrons que les mots lisses extrémaux sur l'alphabet $\{1,3\}$ sont étonnamment reliés au mot de Fibonacci.

Pour débuter, rappelons que le mot infini de Fibonacci $\mathcal{F}$ est défini comme

$$\mathcal{F} = \lim_{n\to\infty} \mathcal{F}_n \quad \text{où} \quad \mathcal{F}_0 = 2, \quad \mathcal{F}_1 = 1, \quad \text{et} \quad \forall n \geq 2,\ \mathcal{F}_n = \mathcal{F}_{n-1}\mathcal{F}_{n-2}.$$

Dans (Berthé, Brlek et Choquette, 2005), les auteurs montrent que $\mathcal{F}$ n'est pas lisse sur l'alphabet $\mathcal{A} = \{1,2\}$, mais l'est sur l'alphabet $\mathcal{A} = \{1,2,3\}$. Le pavage lisse décrivant $\mathcal{F}$ est

$$\begin{aligned}
\Delta^0(\mathcal{F}) &= 12112121121121211212112112121121121121121121121121121121121121121121\cdots\\
\Delta^1(\mathcal{F}) &= 11211121211121112121112121112111212111211121211121211121112112112112112\cdots\\
\Delta^2(\mathcal{F}) &= 21311131311131113131113131113111313111311131311131311131113113113113113\cdots\\
\Delta^3(\mathcal{F}) &= 11131113131113111313111313111311131311131113131113131113111311311131311\cdots\\
\Delta^4(\mathcal{F}) &= 31311131311131113131113131113111313111311131311131311131113113113113113\cdots\\
\Delta^5(\mathcal{F}) &= 11131113131113111313111313111311131311131113131113131113111311311131311\cdots\\
\Delta^6(\mathcal{F}) &= 31311131311131113131113131113111313111311131311131311131113113113113113\cdots\\
\Delta^7(\mathcal{F}) &= 11131113131113111313111313111311131311131113131113131113111311311131311\cdots\\
\Delta^8(\mathcal{F}) &= 31311131311131113131113131113111313111311131311131311131113113113113113\cdots\\
\Delta^9(\mathcal{F}) &= 11131113131113111313111313111311131311131113131113131113111311311131311\cdots\\
\Delta^{10}(\mathcal{F}) &= 31311131311131113131113131113111313111311131311131311131113113113113113\cdots
\end{aligned}$$

Proposition 6.2.1 *(Berthé, Brlek et Choquette, 2005) Soit $\mathcal{F}$, le mot de Fibonacci. Alors*

$$\Phi(\mathcal{F}) = 112(13)^\omega.$$

Proposition 6.2.2 *(Berthé, Brlek et Choquette, 2005) Soit $\mathcal{F}$ le mot de Fibonacci. Alors*

i) $33, 31313$ *ne sont pas facteurs de $\Delta^k(\mathcal{F})$, pour $k \geq 0$;*

ii) $\mathcal{F}$ *et* $\Delta(\mathcal{F}) \in \{1,2\}^\omega$ *;*

iii) $\Delta^2(\mathcal{F}) = 2u$*, où* $u \in \{1,3\}^\omega$ *;*

iv) $\Delta^k(\mathcal{F}) \in \{1,3\}^\omega$ *pour* $k \geq 3$*.*

En utilisant l'Algorithme 6.1.23 décrit dans la section précédente, on obtient les premières lettres du mot lisse infini minimal sur l'alphabet $\mathcal{A} = \{1,3\}$, soit

$$m_{\{1,3\}} = 11131113131113111313111313111311131311131113131113131113\cdots$$

Ce mot coïncide avec $\Delta^3(\mathcal{F})$. En effet, on a le Théorème 6.2.3.

Théorème 6.2.3 $m_{\{1,3\}} = \Delta^3(\mathcal{F})$*, où* $\mathcal{F}$ *est le mot de Fibonacci.*

Afin de prouver ce théorème, certains lemmes sont requis.

Lemme 6.2.4 *Soit* $\mathcal{F}_n$*, le* n*-ième mot de Fibonacci. Alors pour* $n \geq 3$*,* $\mathcal{F}_n$ *admet les factorisations*

i) $\mathcal{F}_n = \mathcal{F}_k \prod_{i=k-1}^{n-2} \mathcal{F}_i$*, pour* $1 \leq k < n$ *;*

ii) $\mathcal{F}_n = \left(\prod_{i=0}^{k-1} \mathcal{F}_{n-2i-1}\right) \mathcal{F}_{n-2k}$*, pour* $1 \leq k \leq n/2$*,* k *un entier.*

Preuve Rappelons que par définition, $\mathcal{F}_n = \mathcal{F}_{n-1}\mathcal{F}_{n-2}$.

i) Pour $n = 3$ et $k = 1$, on a $\mathcal{F}_3 = \mathcal{F}_1\mathcal{F}_0\mathcal{F}_1$. Si $k = 2$, on a $\mathcal{F}_3 = \mathcal{F}_2\mathcal{F}_1$. Supposons que la factorisation existe pour tout $j \leq n-1$. Alors, par définition, $\mathcal{F}_n = \mathcal{F}_{n-1}\mathcal{F}_{n-2} = (\mathcal{F}_{n-2}\mathcal{F}_{n-3})\mathcal{F}_{n-2}$. Par hypothèse de récurrence, on a

$$\mathcal{F}_n = \mathcal{F}_k \left(\prod_{i=k-1}^{n-4} \mathcal{F}_i\right) \mathcal{F}_{n-3}\mathcal{F}_{n-2} = \mathcal{F}_k \prod_{i=k-1}^{n-2} \mathcal{F}_i.$$

ii) Pour $n = 3$ et $k = 1$, on a $\mathcal{F}_3 = \mathcal{F}_2\mathcal{F}_1$. Supposons que la factorisation existe pour tout $j \leq n-1$. Alors, $\mathcal{F}_n = \mathcal{F}_{n-1}\mathcal{F}_{n-2} = \mathcal{F}_{n-1}\mathcal{F}_{n-3}\mathcal{F}_{n-4}$ et par hypothèse de récurrence,

$$\mathcal{F}_n = \mathcal{F}_{n-1}\mathcal{F}_{n-3}\mathcal{F}_{n-4} = \mathcal{F}_{n-1}\mathcal{F}_{n-3}\left(\prod_{i=0}^{k-1} \mathcal{F}_{n-4-2i-1}\right)\mathcal{F}_{n-4-2k} = \left(\prod_{i=0}^{k-1} \mathcal{F}_{n-2i-1}\right)\mathcal{F}_{n-2(k+2)},$$

avec $2k \leq n-4$, donc $2(k+2) \leq n$. ∎

Lemme 6.2.5 *Les mots finis de Fibonacci $\mathcal{F}_n$ satisfont, pour $n \geq 2$, les conditions*
i) $\Delta^3(\mathcal{F}_{2n+1} \cdot 1^{-1}) = \Delta^3(\mathcal{F}_{2n} \cdot 2^{-1}) \cdot \Delta^3(2 \cdot \mathcal{F}_{2n-1} \cdot 1^{-1})$;
ii) $\Delta^3(\mathcal{F}_{2n} \cdot 2^{-1}) = \Delta^3(\mathcal{F}_{2n-1} \cdot 1^{-1}) \cdot \Delta^3(1 \cdot \mathcal{F}_{2n-2} \cdot 2^{-1})$.

Preuve Par un calcul direct, on a

$$\begin{aligned}
\Delta^3(\mathcal{F}_5 1^{-1}) &= \Delta^3(1211212) = \Delta^2(112111) = \Delta^1(213) = 111,\\
\Delta^3(1\mathcal{F}_4 2^{-1}) &= \Delta^3(11211) = \Delta^2(212) = \Delta^1(111) = 3,\\
\Delta^3(\mathcal{F}_4 2^{-1}) &= \Delta^3(1211) = \Delta^2(112) = \Delta^1(21) = 11,\\
\Delta^3(2\mathcal{F}_3 1^{-1}) &= \Delta^3(212) = \Delta^2(111) = \Delta^1(3) = 1,\\
\Delta^3(\mathcal{F}_3 1^{-1}) &= \Delta^3(12) = \Delta^2(11) = \Delta^1(2) = 1,\\
\Delta^3(1\mathcal{F}_2 2^{-1}) &= \Delta^3(11) = \Delta^2(2) = \Delta^1(1) = 1.
\end{aligned}$$

Les deux conditions sont bien vérifiées pour $n = 2$.
La récurrence $\mathcal{F}_n = \mathcal{F}_{n-1} \cdot \mathcal{F}_{n-2}$ implique que

$$\mathcal{F}_n \cdot a^{-1} = \mathcal{F}_{n-1} \cdot \mathcal{F}_{n-2} \cdot a^{-1} = (\mathcal{F}_{n-1} \cdot b^{-1}) \cdot (b \cdot \mathcal{F}_{n-2} \cdot a^{-1}),$$

où $a, b \in \{1, 2\}$. Il suffit alors de montrer que les conditions du *Lemme du collage* sont satisfaites dans les deux cas.

i) Vérifions que la dernière lettre de $\Delta^i(\mathcal{F}_{2n} \cdot 2^{-1})$ diffère de la première de $\Delta^i(2 \cdot \mathcal{F}_{2n-1} \cdot 1^{-1})$, pour $0 \leq i \leq 3$. Du Lemme 6.2.4 ii), on sait que pour tout $n \geq 3$, $\mathcal{F}_{2n}$ a le suffixe $\mathcal{F}_6$. De plus, pour $n \geq 3$, on sait par le Lemme 6.2.4 i) que $\mathcal{F}_{2n}$ a $\mathcal{F}_6$ pour préfixe. Comme $\mathcal{F}_6 = 1211212112112$, le mot $\mathcal{F}_{2n} \cdot 2^{-1}$ a le suffixe $s = 121121211211$ et le mot $2 \cdot \mathcal{F}_{2n-1} \cdot 1^{-1}$ a le préfixe $p = 21211212$. En appliquant les opérateurs D_ℓ et D_r généralisés sur l'alphabet $\{1, 2, 3\}$, à s et p respectivement, on obtient

$$\begin{aligned}
{D_\ell}^0(s) &= 121121211211 & {D_r}^0(p) &= \mathbf{2}1211212\\
{D_\ell}^1(s) &= 1211121\mathbf{2} & {D_r}^1(p) &= \mathbf{1}11211\\
{D_\ell}^2(s) &= 1311\mathbf{1} & {D_r}^2(p) &= \mathbf{3}1\\
{D_\ell}^3(s) &= 1\mathbf{3} & {D_r}^3(p) &= \mathbf{1}.
\end{aligned}$$

Ainsi, les conditions du *Lemme du collage* sont satisfaites.

ii) De façon semblable, on obtient que pour tout $n \geq 4$, le mot $\mathcal{F}_{2n-1} \cdot 1^{-1}$ a le suffixe $s = 12112121121121211212$ et le mot $1 \cdot \mathcal{F}_{2n-2} \cdot 2^{-1}$ a le préfixe $p = 1121121211211$. En appliquant encore les opérateurs D_ℓ et D_r sur s et p respectivement, on obtient le résultat. ■

Lemme 6.2.6 *(Berthé, Brlek et Choquette, 2005)* $\Delta^3(2\mathcal{F}) = \Delta^3(\mathcal{F})$.

Preuve (du Théorème 6.2.3) Procédons par récurrence pour montrer que pour tout $n \geq 2$, $\Delta^3(\mathcal{F}_{2n-1}1^{-1})$ est minimal. Vérifions pour $n = 2$:

$$\Delta^3(\mathcal{F}_3 \cdot 1^{-1}) = \Delta^3(12) = \Delta^2(11) = \Delta(2) = 1$$

est bien minimal. Vérifions un cas de base moins trivial, disons pour $n = 4$.

$$\begin{aligned}
\mathcal{F}_7 \cdot 1^{-1} &= 121121211211212112121 \cdot 1^{-1} = 12112121121121211212,\\
\Delta(\mathcal{F}_7 \cdot 1^{-1}) &= 1121112121112111,\\
\Delta^2(\mathcal{F}_7 \cdot 1^{-1}) &= 213111313,\\
\Delta^3(\mathcal{F}_7 \cdot 1^{-1}) &= 1113111.
\end{aligned}$$

Supposons maintenant que $\Delta^3(\mathcal{F}_{2k-1} \cdot 1^{-1})$ est minimal pour tout $2 \leq k \leq n$. Par le Lemme 6.2.5,

$$\Delta^3(\mathcal{F}_{2(n+1)-1} \cdot 1^{-1}) = \Delta^3(\mathcal{F}_{2n+1} \cdot 1^{-1}) = \Delta^3(\mathcal{F}_{2n} \cdot 2^{-1}) \cdot \Delta^3(2 \cdot \mathcal{F}_{2n-1} \cdot 1^{-1}).$$

Par hypothèse de récurrence, $\Delta^3(\mathcal{F}_{2n} \cdot 2^{-1})$ est minimal. Il suffit de montrer que $\Delta^3(2 \cdot \mathcal{F}_{2n-1} \cdot 1^{-1})$ est minimal, comme la concaténation de deux mots minimaux est certainement minimale. Par le Lemme 6.2.6, comme $\Delta^3(\mathcal{F}_{2n-1} \cdot 1^{-1})$ est minimal, alors $\Delta^3(2 \cdot \mathcal{F}_{2n-1} \cdot 1^{-1})$ est aussi minimal. Pour conclure, il suffit de prendre la limite

$$\lim_{n\to\infty} \Delta^3(\mathcal{F}_{2n-1} \cdot 1^{-1}) = \Delta^3(\mathcal{F}).$$

Cette dernière existe, comme $\mathcal{F}_{n-1}$ est préfixe de $\mathcal{F}_n$. ■

Puisque $\Delta^3(\mathcal{F}) = m_{\{1,3\}}$, les propriétés suivantes découlent directement de (Berthé, Brlek et Choquette, 2005).

Corollaire 6.2.7 *Les mots lisses infinis extrémaux sur l'alphabet* $\mathcal{A} = \{1, 3\}$ *satisfont aux conditions suivantes*

i) $\Delta^k(m_{\{1,3\}}) = \Delta^{k+2}(m_{\{1,3\}}),\ \forall k \geq 0$ *;*

ii) $\Phi(m_{\{1,3\}}) = (13)^\omega$ *et* $\Phi(M_{\{1,3\}}) = 3(31)^\omega$ *;*

iii) $m_{\{1,3\}}$ *ne contient pas les facteurs* 33 *et* 31313 *et* $M_{\{1,3\}}$ *ne contient pas les facteurs complémentaires* 11 *et* 13131 *;*

iv) *posons* $m_{\{1,3\}} = 11u$*, alors* $\Delta(m_{\{1,3\}}) = 3u$*.*

Proposition 6.2.8 *Le transducteur illustré à la Figure 6.1 calcule le mot lisse infini minimal* $m_{\{1,3\}}$ *en temps linéaire.*

Preuve Découle de la la propriété iv) du Corollaire 6.2.7. ∎

Rappelons qu'un transducteur est un automate fini à une bande et deux pointeurs : un pour lire et l'autre pour écrire. Les transitions de l'automate étiquetées de la forme α/β signifient qu'on lit α, qu'on écrit β, puis qu'on se déplace à l'état suivant.

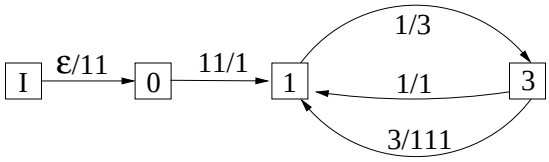


Figure 6.1 Transducteur engendrant le mot lisse minimal sur l'alphabet $\{1, 3\}$.

Le Tableau 6.1 décrit comment le mot $m_{\{1,3\}}$ est construit en utilisant le transducteur.

La fait que le mot de Fibonacci et le mot lisse infini minimal sur l'alphabet $\{1, 3\}$ soient dans la même orbite sous l'opérateur Δ est surprenant et soulève d'intéressantes questions. Est-ce que les propriétés du mot minimal sur $\{1, 3\}$ se généralisent à tout alphabet impair ? La section suivante répond à cette question.

6.3 Mots lisses extrémaux sur un alphabet à deux lettres impaires

Dans cette section, nous considérons un alphabet à deux lettres $\mathcal{A} = \{a < b\}$, avec a et b des entiers positifs impairs. Introduisons d'abord un lemme fort utile pour la suite.

Lu	Écrit	Préfixe de $m_{\{1,3\}}$
ε	11	11
11	1	111
1	3	1113
3	111	1113111
1	3	11131113
1	1	111311131
1	3	1113111313
3	111	1113111313111
1	3	1131113131113
...	...	...

Tableau 6.1 Construction du mot lisse minimal à l'aide d'un transducteur.

Lemme 6.3.1 *Pour tout $u \in \mathcal{A}^+, \Phi^{-1}(u)$ est un palindrome de longueur impaire.*

Preuve Soit $w = \Phi^{-1}(u)$. Procédons par récurrence sur la longueur de u. Si $n = \lg(u) = 1$, alors $w = \beta \in \mathcal{A}$, qui est un palindrome. Si $n = 2$, alors $u = \alpha\beta$, avec $\alpha, \beta \in \{a, b\}$. Ainsi, $\Phi^{-1}(u) = w = \alpha^\beta$ est un palindrome de longueur β. Comme a et b sont impaires, il en découle que w a une longueur impaire. Supposons maintenant que l'énoncé est vrai pour tout u tel que $\lg(u) \leq k$. Posons $u' \in \mathcal{A}^k$. Donc $w = \Phi^{-1}(u')$ est un palindrome de longueur impaire. Posons $\lg(w) = 2j + 1$. On peut donc écrire $w = w' \cdot w[j] \cdot \widetilde{w'}$, $w' \in \mathcal{A}^*$ et alors,

$$\Delta_\alpha^{-1}(w) = \Delta_\alpha^{-1}(w' \cdot w[j] \cdot \widetilde{w'}),$$

pour $\alpha \in \mathcal{A}$. Il y a deux cas à considérer : si $\lg(w')$ est impair,

$$\Delta_\alpha^{-1}(w) = \Delta_\alpha^{-1}(w') \cdot \Delta_{\overline{\alpha}}^{-1}(w[j]) \cdot \Delta_\alpha^{-1}(\widetilde{w'}) = \Delta_\alpha^{-1}(w') \cdot \Delta_{\overline{\alpha}}^{-1}(w[j]) \cdot \widetilde{\Delta_\alpha^{-1}(w')},$$

et si $\lg(w')$ est paire,

$$\Delta_\alpha^{-1}(w) = \Delta_\alpha^{-1}(w') \cdot \Delta_\alpha^{-1}(w[j]) \cdot \Delta_{\overline{\alpha}}^{-1}(\widetilde{w'}) = \Delta_\alpha^{-1}(w') \cdot \Delta_\alpha^{-1}(w[j]) \cdot \widetilde{\Delta_\alpha^{-1}(w')}.$$

Rappelons que dans les équations précédentes, on a utilisé le Lemme 5.1.17. Comme $\forall \alpha, \beta \in \mathcal{A}$, et $\Delta_\alpha^{-1}(\beta) = \alpha^\beta$ est de longueur impaire, alors $\Delta_\alpha^{-1}(w)$ est de longueur impaire. ■

Le Théorème 6.3.2 montre que les mots lisses extrémaux sur un alphabet impair a une structure beaucoup plus simple que ceux sur l'alphabet $\{1, 2\}$ et que le Corollaire 6.2.7 ii) se généralise.

Théorème 6.3.2 $\Phi(m_{\{a,b\}}) = (ab)^\omega$.

Preuve Procédons par récurrence sur la longueur des préfixes de $u = \Phi(m_{\{a,b\}})$. Remarquons d'abord que $m_{\{a,b\}}$ commence par la lettre minimale a. On peut facilement vérifier que $\Phi^{-1}(ab) = a^b < a^a b \cdot w = \Phi^{-1}(aax)$, pour tout $x \in \Sigma, w \in \Sigma^*$. Supposons maintenant que $\Phi^{-1}((ab)^k)$ est minimal pour tout $k \leq n$. Par le Lemme 6.3.1, toutes les lignes du pavage (voir Figure a)) obtenu par $\Phi^{-1}((ab)^n)$ sont de longueur impaire. Donc $\Phi^{-1}((ab)^n)$ commence et termine par la même lettre, soit la lettre a.

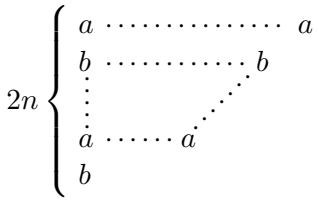


Figure a)

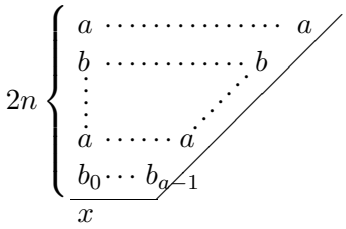


Figure b)

Soit x, la $(2n+1)$-ième lettre de $\Phi(m_{\{a,b\}})$. Que $x = a$ ou $x = b$, la $2n$-ième ligne commence par au moins a occurrences de la lettre b. Comme a, b sont impairs, par le Lemme 6.3.1, toutes les lignes commencent et terminent par la même lettre, en alternant. La Figure b) illustre cette propriété. Les indices dans la figure comptent le nombre de lettres. Par exemple, $b_0 \cdots b_{a-1}$ signifie qu'il y a a occurrences de la lettre b.

Si $x = a$, alors $\Delta_b^{-1}(a) = b^a$ et la $2n$-ième ligne a le préfixe $b^a x = b^a a$. Si $x = b$, alors $\Delta_b^{-1}(b) = b^a b^{b-a}$ et la $2n$-ième ligne commence par $b^a x = b^a b$. Dans les deux cas, cela signifie que la $2n$-ième ligne commence par $b^a x$ (voir Figure c)).

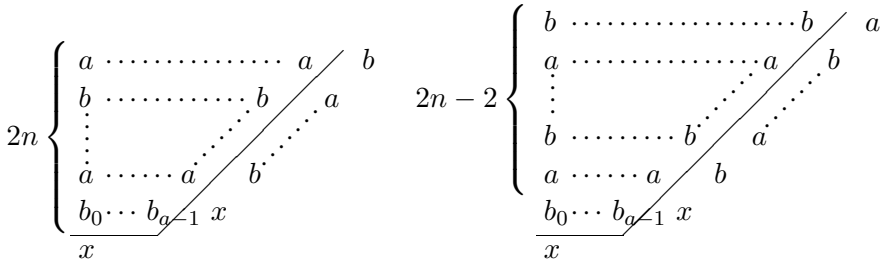


Figure c) Figure d)

Par le Lemme du collage, on a $\Phi^{-1}((ab)^n x) = \Phi^{-1}((ab)^n a) \cdot \Phi^{-1}((ba)^{n-1}bx)s$, pour un $s \in \Sigma^*$. On déduit alors que la lettre x doit être celle qui rend $\Phi^{-1}((ba)^{n-1}bx)$ minimal. Dans la Figure d), on considère $\Phi^{-1}((ba)^{n-1}bx)$. Pour que ce mot soit minimal, $\Phi^{-1}((ab)^{n-1}x)$ doit être minimal. Par l'hypothèse de récurrence, on a donc $x = a$. Il en découle que si $\Phi^{-1}((ab)^n)$ est minimal, alors $\Phi^{-1}((ab)^n a)$ l'est aussi.

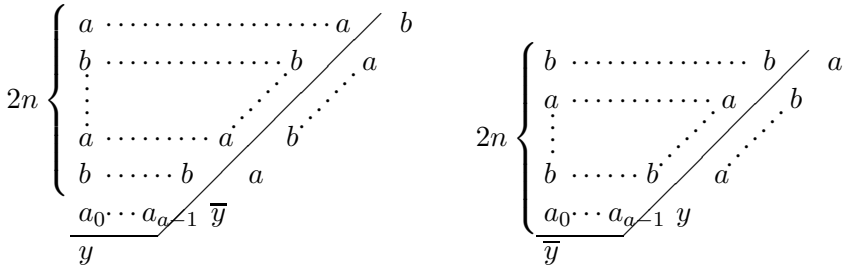


Figure e) Figure f)

La Figure e) montre que la prochaine lettre, disons y, est celle qui rend $\Phi^{-1}((ba)^n\overline{y})$ minimal. La Figure f) considère ce mot et illustre le fait que $\overline{y}$ doit être tel que $\Phi^{-1}((ab)^{n-1}ay)$ est minimal. Par l'hypothèse de récurrence, on obtient $y = b$ et on conclut. ■

Corollaire 6.3.3 $\Phi(M_{\{a,b\}}) = b(ba)^\omega$.

Preuve On obtient directement Φ pour le mot lisse maximal en utilisant l'égalité $\Delta(m_{\{a,b\}}) = \Delta(M_{\{a,b\}})$. ■

Le Théorème 6.3.2 et le Corollaire 6.3.3 montrent bien que les propriétés de périodicité de $\Phi(m_{\{1,3\}})$ se généralisent à tout alphabet impair. Cette périodicité de $\Phi(m_{\{a,b\}})$ nous donne un algorithme linéaire qui construit le mot minimal et donc, le mot lisse maximal pour les alphabets impairs.

Corollaire 6.3.4 *Soit $\alpha \in \mathcal{A} = \{a, b\}$. Le transducteur de la Figure 6.2 engendre $m_{\{a,b\}}$.*

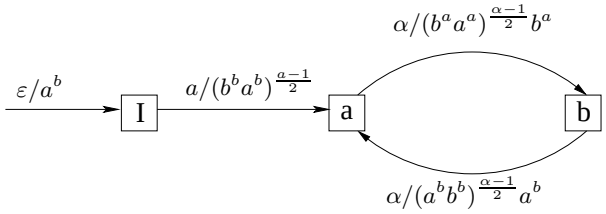


Figure 6.2 Transducteur engendrant $m_{\{a,b\}}$ pour un alphabet impair.

En échangeant les lettres a et b dans le transducteur précédent, on obtient directement le transducteur pour le mot lisse maximal.

Le Corollaire 6.2.7 iii) implique que $F(m_{\{1,3\}})$ n'est pas fermé par complémentation. La structure palindromique des mots lisses (voir Lemme 6.3.1) nous permet toutefois d'établir le résultat suivant.

Proposition 6.3.5 *Pour tout mot lisse infini w, l'ensemble $F(w)$ est fermé sous l'image miroir et w est récurrent.*

Preuve Soit f un facteur fini de w. Alors $w = ufv$ pour $u, f \in \mathcal{A}^*$ et $v \in \mathcal{A}^\omega$. Comme tout mot lisse w a, par le Lemme 6.3.1, des préfixes palindromiques arbitrairement longs, il existe un préfixe palindromique p de w commençant par uf, et ainsi, contenant $\widetilde{uf}$ et on obtient la fermeture sous l'image miroir. Pour la propriété de récurrence, une étape supplémentaire est nécessaire. Comme p contient f et $\widetilde{f}$, tout préfixe palindromique plus long q contient nécessairement deux occurrences de f et $\widetilde{f}$. Remarquons que p est alors préfixe et suffixe de q. Donc f apparaît au moins deux fois dans q. La Figure 6.3 illustre cette preuve. ∎

6.3.1 Factorisations de Lyndon

Dans cette section, nous montrons que la Conjecture 5.3.10 ne se généralise pas pour les mots lisses sur un alphabet $\mathcal{A} \neq \{1, 2\}$. Avant de le prouver, débutons par un résultat négatif.

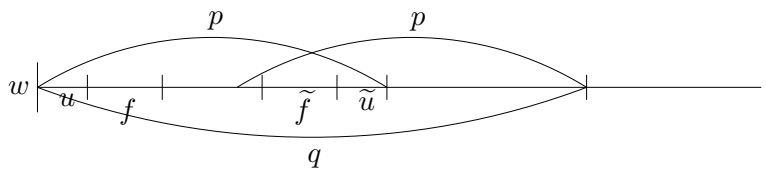


Figure 6.3 Schéma de la preuve de la Proposition 6.3.5.

Lemme 6.3.6 *Si $a \neq 1$, alors $m_{\{a,b\}} \notin \mathcal{L}_\infty$.*

Preuve En calculant $\Phi^{-1}((ab)^2)$, on obtient le pavage fini
$[(a^b b^b)^{\frac{a-1}{2}} \underline{a}^b (b^a a^a)^{\frac{a-1}{2}} b^a]^{\frac{b-1}{2}} (a^b b^b)^{\frac{a-1}{2}} a^b$
$(b^a a^a)^{\frac{b-1}{2}} b^a$
a^b
b.
Par la Proposition 5.2.22, on sait que $\Phi^{-1}((ab)^2)$ est préfixe de $m_{\{a,b\}}$. Le mot lisse minimal admet donc un suffixe de la forme $a^b b^a a s'$, avec $s' \in \mathcal{A}^\omega$. Comme $a^b b^a a s' < a^b b^b s$, $m_{\{a,b\}} \notin \mathcal{L}_\infty$. ∎

Exemple 6.3.7 Le mot $m_{\{3,5\}} = 333335555\underline{3}33335553335553333355555\cdots$ a le suffixe $s = 33333555333\cdots$ qui est plus petit que lui, alors $m_{\{3,5\}} \notin \mathcal{L}_\infty$.

Dans le Lemme 6.3.6, on suppose que $a \neq 1$ afin de s'assurer que le mot commence par $a^b b^b$ (sinon $\frac{a-1}{2} = 0$). Si $a = 1$, la situation est bien différente et on va montrer que $m_{\{1,b\}} \in \mathcal{L}_\infty$. Avant d'en faire la preuve, quelques résultats au sujet des préfixes lisses sont préalables. Dans ce qui suit, pour $k \geq 1$, on pose

$$w_{2k} = \Phi^{-1}((1b)^k) \quad \text{et} \quad w_{2k+1} = \Phi^{-1}(b(1b)^k). \tag{6.1}$$

Proposition 6.3.8 *Soit $\mathcal{A} = \{1, b\}$. Alors les conditions suivantes sont vérifiées.*
i) $w_n = (w_{n-2} \cdot w_{n-3})^{\frac{b-1}{2}} \cdot w_{n-2}$, *pour tout* $n \geq 4$.
ii) $w_{2k}w_{2k-1},\ w_{2k}w_{2k+1} \in \mathcal{L}$, *pour tout* $k \geq 1$.
iii) $w_{2k-2}w_{2k-1} \geq w_{2k}$ *et* $w_{2k} \notin \mathrm{Pref}(w_{2k-2}w_{2k-1})$, *pour tout* $k \geq 2$.

Preuve Procédons par récurrence. i) Par un calcul direct, on obtient $w_1 = b$, $w_2 = 1^b$, $w_3 = (b1)^{\frac{b-1}{2}}b$ et $w_4 = (1^b b)^{\frac{b-1}{2}}1^b$. Comme $w_4 = (1^b \cdot b)^{\frac{b-1}{2}} \cdot 1^b = (w_2 \cdot w_1)^{\frac{b-1}{2}} \cdot w_2$, l'énoncé est vrai pour $n = 4$. Supposons maintenant que $w_m = (w_{m-2}w_{m-3})^{\frac{b-1}{2}}w_{m-2}$, pour tout $m \leq n$. Alors, comme la fonction Δ^{-1} est distributive et puisqu'en vertu du Lemme 6.3.1 les w_i sont des palindromes de longueur impaire, on a

$$\begin{aligned} w_{n+1} &= \Delta_\alpha^{-1}(w_n) \\ &= \Delta_\alpha^{-1}\left((w_{n-2}w_{n-3})^{\frac{b-1}{2}}w_{n-2}\right) \\ &= \Delta_\alpha^{-1}\left((w_{n-2}w_{n-3})^{\frac{b-1}{2}}\right)\Delta_\alpha^{-1}(w_{n-2}) \\ &= \left(\Delta_\alpha^{-1}(w_{n-2})\Delta_{\overline{\alpha}}^{-1}(w_{n-3})\right)^{\frac{b-1}{2}}\Delta_\alpha^{-1}(w_{n-2}) \\ &= (w_{n-1}w_{n-2})^{\frac{b-1}{2}}w_{n-1}, \end{aligned}$$

avec $\alpha = b$ si n pair, $\alpha = 1$ sinon.

ii) Par (6.1), on a que $w_2w_1 = 1^b b,\ w_2w_3 = 1^b(b1)^{\frac{b-1}{2}}b \in \mathcal{L}$. Donc l'énoncé est vrai pour $k = 1$. Supposons maintenant que $w_{2k}w_{2k-1},\ w_{2k}w_{2k+1} \in \mathcal{L}$ pour tout $k \leq n$.

1. Par i), $w_{2n+2}w_{2n+1} = (w_{2n}w_{2n-1})^{\frac{b-1}{2}} \cdot w_{2n}w_{2n+1}$. En utilisant l'hypothèse de récurrence, $w_{2n}w_{2n-1},\ w_{2n}w_{2n+1} \in \mathcal{L}$, donc $w_{2n+2}w_{2n+1} = u^{\frac{b-1}{2}}v$, où $u, v \in \mathcal{L}$ avec $u \in \text{Pref}(v)$ implique que $u < v$. On conclut en utilisant le Corollaire 6.1.3.

2. Par i), $w_{2n+2}w_{2n+3} = w_{2n+2}w_{2n+1} \cdot (w_{2n}w_{2n+1})^{\frac{b-1}{2}}$. En utilisant respectivement 1. et l'hypothèse de récurrence, on déduit que $w_{2n+2}w_{2n+1},\ w_{2n}w_{2n+1} \in \mathcal{L}$. Alors, $w_{2n+2}w_{2n+3} = uv^{\frac{b-1}{2}}$, où $u, v \in \mathcal{L}$, $v \in \text{Suff}(u)$ implique que $u < v$. On conclut à nouveau en utilisant le Corollaire 6.1.3.

iii) Pour $k = 2$, $w_2w_3 = 1^b(b1)^{\frac{b-1}{2}}b = 1^b b1bs$ et $w_4 = (1^b b)^{\frac{b-1}{2}}1^b = 1^b b11s'$, $s, s' \in \mathcal{A}^*$. Ainsi, le lemme est vérifié pour $k = 2$. Supposons maintenant que l'énoncé est vrai pour tout $k \leq n$. Alors par i),

$$w_{2n}w_{2n+1} = w_{2n}w_{2n-1}(w_{2n-2}w_{2n-1})^{\frac{b-1}{2}}$$

et

$$w_{2n+2} = w_{2n}(w_{2n-1}w_{2n})^{\frac{b-1}{2}}.$$

Par hypothèse de récurrence, $w_{2n-2}w_{2n-1} \geq w_{2n}$. ∎

Exemple 6.3.9 Soit $\mathcal{A} = \{1, 5\}$. Alors $\Phi(m_{\{1,5\}}) = (15)^\omega$ et $w_1 = 5$, $w_2 = 11111$, $w_3 = 51515$. La Proposition 6.3.8 i) indique que

$$w_4 = (w_2 w_1)^{\frac{5-1}{2}} w_2 = w_2 w_1 w_2 w_1 w_2 = 11111511111511111.$$

Remarquons que $w_2 w_1 = 111115 \in \mathfrak{L}$ et $w_2 w_3 = 1111151515 \in \mathfrak{L}$.

Proposition 6.3.10 *Soit $\mathcal{A} = \{1, b\}$ et soit L_n la factorisation de Lyndon de w_n. Alors pour $n \geq 4$,*

$$L_n = \begin{cases} \left(\bigodot_{i=1}^{\frac{b-1}{2}} w_{n-2} w_{n-3} \right) \cdot L_{n-2}, & \textit{si } n \textit{ est pair} \\ L_{n-2} \cdot \left(\bigodot_{i=1}^{\frac{b-1}{2}} w_{n-3} w_{n-2} \right), & \textit{si } n \textit{ est impair}. \end{cases}$$

Preuve (Par récurrence sur n) On a $w_1 = b$, $w_2 = 1^b$, $w_3 = (b1)^{\frac{b-1}{2}} b$, $w_4 = (1^b b)^{\frac{b-1}{2}} 1^b$, $w_5 = ((b1)^{\frac{b-1}{2}} b 1^b)^{\frac{b-1}{2}} (b1)^{\frac{b-1}{2}} b$ et les factorisations de Lyndon de ces mots sont respectivement

$$L_1 = b, \quad L_2 = \bigodot_{i=1}^{b} 1, \quad L_3 = b \bigodot_{i=1}^{\frac{b-1}{2}} (1b), \quad L_4 = \bigodot_{i=1}^{\frac{b-1}{2}} (1^b b) \bigodot_{i=1}^{b} 1$$

et

$$L_5 = b \bigodot_{i=1}^{\frac{b-1}{2}} (1b) \bigodot_{i=1}^{\frac{b-1}{2}} \left(1^b (b1)^{\frac{b-1}{2}} b \right).$$

Comme $L_4 = \left(\bigodot_{i=1}^{\frac{b-1}{2}} w_2 w_1 \right) \cdot L_2$ et $L_5 = L_3 \cdot \left(\bigodot_{i=1}^{\frac{b-1}{2}} w_2 w_3 \right)$, la formule pour L_n est vérifiée pour $n = 4, 5$.

Supposons maintenant qu'elle soit vraie pour tout $m < n$.

i) si n est pair : par la Proposition 6.3.8 i), $w_n = (w_{n-2} w_{n-3})^{\frac{b-1}{2}} w_{n-2}$; par la Proposition 6.3.8 ii), $w_{n-2} w_{n-3} \in \mathfrak{L}$ avec w_{n-2} comme préfixe propre. Il en découle que $w_{n-2} \geq w_{n-3} w_{n-2}$.

ii) si n est impair : par la Proposition 6.3.8 i), $w_n = w_{n-2} (w_{n-3} w_{n-2})^{\frac{b-1}{2}}$; par la Proposition 6.3.8 ii), $w_{n-2} w_{n-3} \in \mathfrak{L}$. Comme w_{n-2} est suffixe du mot de Lyndon $w_{n-3} w_{n-2}$, on a $w_{n-2} > w_{n-3} w_{n-2}$. Pour terminer, il suffit de montrer que le dernier facteur de L_{n-2}

est plus petit que $w_{n-3}w_{n-2}$. Par hypothèse de récurrence, le dernier facteur de L_{n-2} est le mot de Lyndon $w_{n-5}w_{n-4}$. Par la Proposition 6.3.8 iii), on sait que $w_{n-5}w_{n-4} > w_{n-3}$ et $w_{n-3} \notin \text{Pref}(w_{n-5}w_{n-4})$. ■

On peut maintenant énoncer notre résultat concernant la factorisation de Lyndon du mot lisse minimal de la forme $m_{\{1,b\}}$.

Théorème 6.3.11 *Soit $\mathcal{A} = \{1, b\}$. Alors*
i) $m_{\{1,b\}} \in \mathcal{L}_\infty$ *;*
ii) *la factorisation de Lyndon de $\Delta(m_{\{1,b\}})$ est une suite infinie de mots de Lyndon finis.*

Preuve Il suffit de prendre la limite $n \to \infty$ des énoncés de la Proposition 6.3.10. ■

6.3.2 Fréquences des lettres

Dans cette sous-section, nous donnons une formule pour la fréquence des lettres dans les mots lisses extrémaux sur l'alphabet $\mathcal{A} = \{1, b\}$. Remarquons que pour un alphabet impair $\mathcal{A} = \{a, b\}$ avec un a quelconque, nous ne connaissons toujours pas la fréquence des lettres de $m_{\{a,b\}}$.

Théorème 6.3.12 *Soit $\mathcal{A} = \{1, b\}$. Alors la fréquence de la lettre b dans $m_{\mathcal{A}}$ est*

$$f_b(m_{\mathcal{A}}) = \frac{1}{\sqrt{2b-1}+1}. \tag{6.2}$$

Preuve Par les Propositions 5.2.22 et 6.3.8 i), w_{2n} est un préfixe de $m_{\{1,b\}}$ pour tout $n \geq 1$ et on a la définition récursive suivante pour $m_{\{1,b\}}$:

$$\begin{aligned} w_1 &= b;\ w_2 = 1^b;\ w_3 = (b1)^{\frac{b-1}{2}}b; \\ w_k &= (w_{k-2}w_{k-3})^{\frac{b-1}{2}}w_{k-2}; \\ m_{\{1,b\}} &= \lim_{n\to\infty} w_{2n}. \end{aligned}$$

En posant $f_n = |w_n|_b$ et $g_n = |w_n|_1$, la définition récursive de w_n nous donne une récurrence pour f_n et g_n. Ainsi,

$$f_n = \frac{b-1}{2}(f_{n-2} + f_{n-3}) + f_{n-2} = \frac{b+1}{2}f_{n-2} + \frac{b-1}{2}f_{n-3}, \tag{6.3}$$

avec les conditions initiales $f_1 = 1$, $f_2 = 0$, $f_3 = \frac{b+1}{2}$ et

$$g_n = \frac{b+1}{2} g_{n-2} + \frac{b-1}{2} g_{n-3}, \tag{6.4}$$

avec les conditions initiales $g_1 = 0$, $g_2 = b$ et $g_3 = \frac{b-1}{2}$. Afin de compléter la preuve, il suffit de résoudre les récurrences.

Le polynôme caractéristique associé à la récurrence f_n donnée en (6.3) est

$$z^3 - \frac{b+1}{2} z - \frac{b-1}{2} = 0,$$

qui se réécrit comme

$$(z+1)\left(z - \frac{1+\sqrt{2b-1}}{2}\right)\left(z - \frac{1-\sqrt{2b-1}}{2}\right) = 0.$$

Si $b \neq 5$: Comme les racines du polynômes sont -1, $\frac{1+\sqrt{2b-1}}{2}$ et $\frac{1-\sqrt{2b-1}}{2}$, il en découle que $f_n = c_1(-1)^n + c_2(\frac{1+\sqrt{2b-1}}{2})^n + c_3(\frac{1-\sqrt{2b-1}}{2})^n$, avec $c_1, c_2, c_3 \in \mathbb{R}$. En utilisant les conditions initiales, on trouve

$$c_1 = \frac{2}{b-5}, \quad c_2 = \frac{b+\sqrt{2b-1}}{\sqrt{2b-1}(1+b+2\sqrt{2b-1})}, \quad c_3 = -\frac{b-2+\sqrt{2b-1}}{\sqrt{2b-1}(b-5)}.$$

On a donc une formule close pour f_n.

En procédant de la même façon avec la récurrence (6.4), on trouve que pour $b \neq 5$,

$$g_n = c_1'(-1)^n + c_2'\left(\frac{1+\sqrt{2b-1}}{2}\right)^n + c_3'\left(\frac{1-\sqrt{2b-1}}{2}\right)^n,$$

avec

$$c_1' = -\frac{b+1}{b-5}, \quad c_2' = \frac{b\sqrt{2b-1}+2b-1}{\sqrt{2b-1}(1+b+2\sqrt{2b-1})}$$

et

$$c_3' = \frac{2b-1+\sqrt{2b-1}(b-2)}{\sqrt{2b-1}(b-5)}.$$

La fréquence de b est donc donnée par

$$\lim_{n\to\infty} \frac{f_{2n}}{f_{2n}+g_{2n}} = \frac{1}{\sqrt{2b-1}+1}.$$

Si $b = 5$: Les racines du polynôme f_n sont $-1, -1$ et 2 et alors, $f_n = c_1(-1)^n + c_2 n(-1)^n + c_3(2)^n$. Les racines de g_n sont les mêmes. Finalement, en utilisant les conditions initiales, on trouve $c_1 = \frac{-2}{9}$, $c_2 = \frac{-1}{3}$, $c_3 = \frac{2}{9}$, $c_1' = \frac{1}{3}$, $c_2' = 1$ et $c_3' = \frac{2}{3}$. Alors, $\lim_{n\to\infty} \dfrac{f_{2n}}{f_{2n} + g_{2n}} = \dfrac{1}{4}$, qui est égal à $\frac{1}{\sqrt{2\cdot 5 - 1} + 1}$. ■

6.4 Mots lisses extrémaux sur un alphabet à deux lettres paires

Dans cette section, nous supposons que $\mathcal{A} = \{a < b\}$ est un alphabet contenant deux entiers pairs.

Lemme 6.4.1 *Si $w \in \mathcal{A}^+$ est lisse, alors pour tout $\alpha \in \mathcal{A}$,* $\lg(\Delta_\alpha^{-1}(w))$ *est paire.*

Preuve Posons $\lg(w) = n \geq 1$. En appliquant Δ_α^{-1} à w, on obtient

$$\begin{aligned}\Delta_\alpha^{-1}(w) &= \Delta_\alpha^{-1}(w[0]w[1]\cdots w[n-1]) \\ &= \Delta_\alpha^{-1}(w[0])\Delta_{\overline{\alpha}}^{-1}(w[1])\cdots\Delta_\beta^{-1}(w[n-1]) \\ &= \alpha^{w[0]}\overline{\alpha}^{w[1]}\ldots\beta^{w[n-1]}\end{aligned}$$

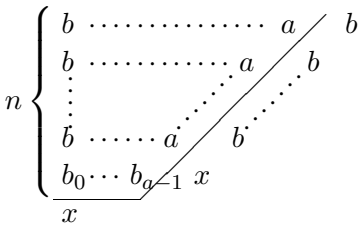
où $\beta = \alpha$ si n est impair et $\beta = \overline{\alpha}$ si n est pair. Comme $\lg(\Delta_\alpha^{-1}(w)) = \sum_{i=0}^{n-1} w[i]$, on conclut. ■

Le théorème suivant montre que tout comme pour les alphabets impairs, les mots lisses extrémaux w sur un alphabet pair sont caractérisés par la périodicité de $\Phi(w)$.

Théorème 6.4.2 $\Phi(M_{\{a,b\}}) = b^\omega$.

Preuve Procédons par récurrence sur la longueur des préfixes de $u = \Phi(M_{\{a,b\}})$. Le mot $M_{\{a,b\}}$ commence nécessairement par le préfixe b^b et donc, $\Phi(M_{\{a,b\}})[0] = b$. Supposons maintenant que $\Phi^{-1}(b^k)$ est maximal, pour tout $k \leq n$. Posons $u = b^n x$ et $v = \Delta_b^{-1}(x)$. Si $\Phi(M_{\{a,b\}})[n] = x$, alors $v = \Delta_b^{-1}(x) = b^x$ et conséquemment, $v[x] = x$. Par le Lemme 6.4.1, chaque préfixe est de longueur paire et alors, il termine par la lettre a. Maintenant, en utilisant le Lemme du collage (Lemme 5.2.9), la lettre x doit être celle qui rend le mot $\Phi^{-1}(b^{n-1}x)$ maximal. Par hypothèse de récurrence, on conclut que $x = b$. ■

On a la situation suivante :



De l'égalité $\Delta(m_{\{a,b\}}) = \Delta(M_{\{a,b\}})$ découle le corollaire suivant.

Corollaire 6.4.3 $\Phi(m_{\{a,b\}}) = ab^\omega$.

Ainsi, $M_{\{a,b\}} = \Delta(m_{\{a,b\}})$. De plus, comme $\Phi(M_{\{a,b\}}) = b^\omega$, $M_{\{a,b\}}$ correspond au mot de Kolakoski généralisé $K_{(b,a)}$. Cette dernière propriété nous donne directement un algorithme linéaire qui engendre les préfixes du mot lisse minimal pour un alphabet pair et donc, du mot maximal en échangeant les lettres de l'alphabet. Cet algorithme est représenté par le transducteur de la Figure 6.4, où $\alpha \in \{a, b\}$.

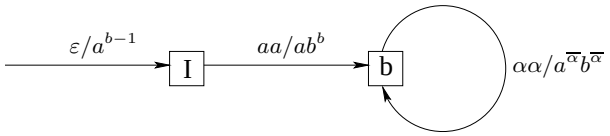


Figure 6.4 Transducteur engendrant $m_{\{a,b\}}$ pour un alphabet pair.

Proposition 6.4.4 *La fréquence des lettres dans les mots lisses extrémaux sur un alphabet pair est* $1/2$.

Preuve Comme ce transducteur a deux cycles (un pour chaque lettre) avec le même état initial, tout chemin infini non périodique va dans ces deux cycles. Comme un nombre égal de a et de b sont écrits dans chaque cycle, la fréquence de chacune des lettres est $\frac{1}{2}$. ■

Ce résultat est fort surprenant puisque la fréquence des lettres des mots de Kolakoski $K_{(1,2)}$ et $K_{(2,1)}$ sont toujours inconnues. En effet, la meilleure borne connue est 0.50084 et a été démontrée par Chvátal (Chvátal, 1994), qui a construit un algorithme pour approximer la fréquence des lettres du mot de Kolakoski. Tout récemment, dans (Steinsky, 2006), l'auteur a mis en doute

cette conjecture en calculant les fréquences des lettres du mot de Kolakoski de longueur 100 à 300 millions. La courbe qu'il obtient ne semble pas converger vers $1/2$. Par la suite, on trouve dans (Monteil, Thomasse et Tisserand, 2007) le calcul des fréquences pour des longueurs de 100 milliards. Pour une telle longueur, une situation similaire à celle soulevée par Steinsky est observée.

Un lemme analogue au Lemme 6.3.1 exhibant la structure palindromique des préfixes des mots lisses sur un alphabet impair, montre que dans le cas d'un alphabet pair, les préfixes des mots lisses sont des répétitions.

Lemme 6.4.5 *Pour tout $u \in \mathcal{A}^{\geq 2}$ lisse, il existe $p \in \mathcal{A}^{2m}$ tel que $\Phi^{-1}(u) = p^{\frac{u[k]}{2}}$.*

Preuve Soit $w = \Phi^{-1}(u)$. Procédons par récurrence sur la longueur de u. Si $\lg(u) = 2$, alors $u = \alpha\beta$, $\alpha, \beta \in \mathcal{A}$ et $w = \alpha^{\beta} = (\alpha\alpha)^{\frac{\beta}{2}}$, ainsi $p = \alpha\alpha$. Supposons maintenant que l'énoncé est vrai pour tout u tel que $\lg(u) \leq k$. Soit $v \in \mathcal{A}^*$ tel que $\lg(v) = k + 1$. Alors $\Phi^{-1}(v) = \Delta^{-1}_{v[0]}(\Phi^{-1}(v[1,k]))$ et par hypothèse de récurrence, pour un p de longueur paire,

$$\Phi^{-1}(v) = \Delta^{-1}_{v[0]}(p^{\frac{u[k]}{2}}),$$

qui peut se réécrire comme $\Phi^{-1}(v) = (\Delta^{-1}_{v[0]}(p))^{\frac{u[k]}{2}}$. On conclut en appliquant le Lemme 6.4.1. ∎

Cette propriété peut être utilisée pour montrer que les mots lisses extrémaux sont récurrents, en adaptant la preuve donnée dans le cas des alphabets pairs. En fait, pour ces alphabets, la propriété de récurrence est vraie pour tout mot lisse, incluant les mots de Kolakoski $K_{(a,b)}$ et $K_{(b,a)}$.

Théorème 6.4.6 *Les mots lisses sont récurrents.*

Preuve Soit $u \in \mathcal{A}^{\omega}$ et posons $w = \Phi^{-1}(u)$. Soient $f \in F(w)$ et n tel que $p = \Phi^{-1}(u[0, n-1])$ contient f comme facteur. Posons $q = \Phi^{-1}(u[0,(n+1)])$, $\alpha = u[n]$ et $\beta = u[n+1]$. Par définition,

$$\Delta^n(q) = \Delta^{-2}_{\alpha, u[n-1]}(\beta) = \Delta^{-1}_{u[n-1]}(\alpha\alpha) \cdot x,$$

où $\Delta^{-1}_{u[n-1]}(\alpha\alpha)$ termine par la lettre $\overline{u[n-1]}$ et $x \in \mathcal{A}^*$. Soit q' le préfixe de q tel que $\Delta^{n+1}(q') = \alpha\alpha$. Alors $w = q'w'$ pour un mot w' et en utilisant le lemme du collage, on a que pour tout k tel que $0 \leq k \leq n$,

$$\Delta^k(w) = \Delta^k(q') \cdot \Delta^k(w').$$

Par le Lemme 6.4.1, si $\Delta^k(q')$ commence par $u[k-1]$, alors il termine par $\overline{u[k-1]}$. Il en découle que $\Delta^k(w')$ commence par $u[k-1]$ et par conséquent, w' contient nécessairement une autre occurrence de p, donc de f. ■

Proposition 6.4.7 *$F(m_{\{a,b\}})$ et $F(M_{\{a,b\}})$ ne sont pas fermés sous l'image miroir, ni sous la complémentation.*

Preuve Considérons le préfixe $p = (b^b a^b)^{b/2}$ de $M_{\{a,b\}}$ et supposons que son image miroir $\widetilde{p} = (a^b b^b)^{b/2} \in F(M_{\{a,b\}})$. Alors, $\Delta(\widetilde{p}) = b^b$ serait un facteur de $\Delta(M_{\{a,b\}})$ codant $\widetilde{p}$ dans $M_{\{a,b\}}$. Par le Lemme 6.4.1, tout facteur a^a, a^b, b^a, b^b dans $\Delta(M_{\{a,b\}})$ code un facteur de $M_{\{a,b\}}$ commençant par b et terminant par a. Contradiction. Pour la non fermeture sous la complémentation, il suffit d'observer que pour ce préfixe p, on a $\widetilde{p} = \overline{p}$. ■

6.4.1 Factorisation de Lyndon

Nous allons maintenant montrer que le mot lisse minimal sur un alphabet pair est un mot de Lyndon infini. D'abord, quelques lemmes sont préalables.

Lemme 6.4.8 *Soit $w_n = \Phi^{-1}(b^n)$. Alors, pour $n \geq 3$, $w_n = (v_1^{b/2} v_2^{b/2})^{b/2}$, avec $v_2 < v_1$, $v_2 \notin \text{Pref}(v_1)$ et $\lg(v_1)$, $\lg(v_2)$ paires.*

Preuve Procédons par récurrence sur n. On peut facilement calculer

$$w_1 = b, \quad w_2 = b^b \quad \text{et donc,} \quad w_3 = (b^b a^b)^{b/2} = ((bb)^{b/2}(aa)^{b/2})^{b/2}.$$

Comme $aa < bb$, $aa \notin \text{Pref}(bb)$, $\lg(aa)$ et $\lg(bb)$ sont paires, la propriété est vérifiée pour $n = 3$. Supposons maintenant que l'énoncé est vrai pour tout $3 \leq k \leq n$. Alors

$$w_{n+1} = \Delta_b^{-1}(w_n) = \Delta_b^{-1}((v_1^{b/2} v_2^{b/2})^{b/2}) = [(\Delta_b^{-1}(v_1))^{b/2}(\Delta_b^{-1}(v_2))^{b/2}]^{b/2},$$

avec $\lg(\Delta_b^{-1}(v_1))$ et $\lg(\Delta_b^{-1}(v_2))$ paires, par le Lemme 6.4.1. Par la Proposition 5.2.22, v_2 est préfixe de $\Delta_b^{-1}(v_2)$ et v_1 est préfixe de $\Delta_b^{-1}(v_1)$. Comme $v_2 < v_1$ et v_2 n'est pas préfixe de v_1, $\Delta_b^{-1}(v_2) < \Delta_b^{-1}(v_1)$ et $\Delta_b^{-1}(v_2) \notin \text{Pref}(\Delta_b^{-1}(v_1))$. ■

Notation 6.4.9 Soit $w_n = (v_1^{b/2} v_2^{b/2})^{b/2}$. Alors, $\forall n \geq 3$, $\underline{w_n}$ désigne le mot $(v_1^{a/2} v_2^{a/2})^{b/2}$.

Lemme 6.4.10 *Soit* $w_n = \Phi^{-1}(b^n)$. *Pour* $n \geq 4$,

i) $w_n = (w_{n-1}\underline{w_{n-1}})^{b/2}$;

ii) $\underline{w_{n-1}} < w_{n-1}$, $\underline{w_{n-1}} \notin \text{Pref}(w_{n-1})$ *et* $\lg(w_{n-1})$, $\lg(\underline{w_{n-1}})$ *sont paires ;*

iii) $\overline{u}_1, \overline{u}_2 \in \mathcal{L}$, *où* $w_n = u_1^{b/2}$ *et* $\underline{w_n} = u_2^{b/2}$.

Preuve Procédons par récurrence sur n. Par calcul direct, on obtient

$$w_1 = b, \quad w_2 = b^b, \quad w_3 = (b^b a^b)^{b/2} \quad \text{et} \quad w_4 = ((b^b a^b)^{b/2}(b^a a^a)^{b/2})^{b/2}.$$

Comme $\underline{w_3} = \underline{(b^b a^b)^{b/2}} = (b^a a^a)^{b/2}$, on a bien $w_4 = (w_3 \cdot \underline{w_3})^{b/2}$. Ainsi, i) est vérifié pour $n = 4$. Comme $\underline{w_3} = (b^a a^a)^{b/2} < (b^b a^b)^{b/2} = w_3$, $\underline{w_3} \notin \text{Pref}(w_3)$, $\lg(w_3) = b^2$ et $\lg(\underline{w_3}) = ab$, ii) est aussi vérifiée. Finalement, $w_4 = (u_1)^{b/2}$, où $u_1 = (b^b a^b)^{b/2}(b^a a^a)^{b/2}$, $\underline{w_4} = (u_2)^{b/2}$, $u_2 = (b^b a^b)^{a/2}(b^a a^a)^{a/2}$ et $\overline{u_1}, \overline{u_2} \in \mathcal{L}$. Supposons maintenant que les trois énoncés sont vrais pour tout $4 \leq k \leq n$.

i) $w_{n+1} = \Delta_b^{-1}(w_n) = \Delta_b^{-1}((w_{n-1}\underline{w_{n-1}})^{b/2})$. Par hypothèse de récurrence, $\lg(w_{n-1})$ et $\lg(\underline{w_{n-1}})$ sont paires. On a donc

$$w_{n+1} = \Delta_b^{-1}((w_{n-1}\underline{w_{n-1}})^{b/2}) = [\Delta_b^{-1}(w_{n-1})\Delta_b^{-1}(\underline{w_{n-1}})]^{b/2}.$$

Posons $w_{n-1} = (v_3^{b/2} v_4^{b/2})^{b/2}$. Alors, $\underline{w_{n-1}} = (v_3^{a/2} v_4^{a/2})^{b/2}$,

$$w_n = \Delta_b^{-1}(w_{n-1}) = [(\Delta_b^{-1}(v_3))^{b/2}(\Delta_b^{-1}(v_4))^{b/2}]^{b/2}$$

et

$$\Delta_b^{-1}(\underline{w_{n-1}}) = [(\Delta_b^{-1}(v_3))^{a/2}(\Delta_b^{-1}(v_4))^{a/2}]^{b/2} = \underline{w_n}.$$

Ainsi, $w_{n+1} = (w_n \underline{w_n})^{b/2}$.

ii) Par le Lemme 6.4.8 et par i), $w_n = (v_1^{b/2} v_2^{b/2})^{b/2}$ et $\underline{w_n} = (v_1^{a/2} v_2^{a/2})^{b/2}$ avec $v_1^{b/2} = w_{n-1}$ et $v_2^{b/2} = \underline{w_{n-1}}$. Par hypothèse de récurrence, $\underline{w_{n-1}} < w_{n-1}$ et $\underline{w_{n-1}} \notin \text{Pref}(w_{n-1})$. Comme $v_2 \notin \text{Pref}(v_1)$ et $v_2 < v_1$, on a

$$\underline{w_n} = v_1^{a/2} v_2^{a/2} s < v_1^{a/2} v_1^{(b-a)/2} s' = w_n, \ \ s, s' \in \mathcal{A}^*.$$

On a aussi que $\underline{w_n} \notin \text{Pref}(w_n)$ et leurs longueurs sont respectivement $ab(v_1 + v_2)/2$ et $b^2(v_1 + v_2)/2$, qui sont paires.

iii) En utilisant i) et ii), on a $w_{n+1} = (w_n \underline{w_n})^{b/2} = u_3^{b/2}$, avec $\underline{w_n} < w_n$. Alors, $\overline{u_3} = \overline{w_n \underline{w_n}} = \overline{w_n} \cdot \overline{\underline{w_n}}$. Par le Lemme 5.5.2, on a $\underline{w_n} < w_n \iff \overline{w_n} < \overline{\underline{w_n}}$. De plus, par hypothèse, $w_n = u_1^{b/2}, \underline{w_n} = u_2^{b/2}$, avec $\overline{u_1}, \overline{u_2} \in \mathcal{L}$. En utilisant le Corollaire 6.1.3, on obtient que $\overline{u_3} \in \mathcal{L}$. Considérons maintenant $\overline{u_4}$ satisfaisant $\underline{w_{n+1}} = u_4^{b/2}$. En utilisant le Lemme 6.4.8, on sait que $u_3 = v_1^{b/2} v_2^{b/2}$ et alors, que $u_4 = v_1^{a/2} v_2^{a/2}$, avec $v_2 < v_1$. Ainsi, $\overline{u_4} = \overline{v_1}^{a/2} \overline{v_2}^{a/2}$, avec $\overline{v_1} < \overline{v_2}$. On conclut en utilisant le Corollaire 6.1.3. ∎

Théorème 6.4.11 $m_{\{a,b\}} \in \mathcal{L}_\infty$.

Preuve Par le Théorème 6.4.2, $\Phi(M_{\{a,b\}}) = b^\omega$. Posons $w_n = \Phi^{-1}(b^n)$. Alors,

$$M_{\{a,b\}} = \lim_{n \to \infty} w_n.$$

Par le Lemme 6.4.10, on sait que $w_n = (w_{n-1} \underline{w_{n-1}})^{b/2}$, avec $\overline{w_{n-1} \underline{w_{n-1}}} \in \mathcal{L}$. Comme

$$m_{\{a,b\}} = \overline{M_{\{a,b\}}} = \lim_{n \to \infty} (\overline{w_n}),$$

$\lg(w_n) < \lg(w_{n+1})$ et par la Proposition 5.2.22, on conclut. ∎

Dans ce chapitre, nous avons montré certaines des conjectures de Dekking pour les mots extrémaux pour des alphabets de même parité : récurrence, fréquences des lettres, fermeture sous image miroir et sous complémentation. Déterminer les fréquences des lettres pour les mots extrémaux sur des alphabets impairs demeure un problème intéressant, tout comme pour les mots extrémaux sur des alphabets de différentes parités. Les propriétés combinatoires des mots extrémaux sur des alphabets de différentes parités restent très mystérieuses, tout comme la fonction de complexité (nombre de facteurs de longueur n) pour tous les mots extrémaux. Il serait aussi intéressant d'étudier les mots extrémaux sur des alphabets à plus de deux lettres.

Chapitre VII

SURFACES DISCRÈTES ET MOTS LISSES

Parmi les nombreuses caractérisations des suites sturmiennes vues au Chapitre II, la plus commune en géométrie discrète est l'approximation de droites. Des généralisations des suites sturmiennes ont été introduites, entre autres, les suites épisturmiennes et les suites k-équilibrées. Plusieurs auteurs ont généralisé les suites sturmiennes à deux dimensions : il suffit de penser aux travaux de (Berthé et Vuillon, 2000; Arnoux, Berthé et Siegel, 2004; Berthé et Tijdeman, 2004; Durand, Guerziz et Koskas, 2004). Dans (Jamet, 2004), l'auteur introduit les $(1,1,1)$-surfaces discrètes comme étant la généralisation naturelle des plans discrets introduits dans (Berthé et Vuillon, 2000; Arnoux, Berthé et Siegel, 2004) et il donne un critère pour déterminer quelles suites à deux dimensions sur l'alphabet à 3 lettres $\{1,2,3\}$ codent une $(1,1,1)$-surface discrète. Comme il est possible d'associer un pavage du quart de plan autant à un mot lisse infini qu'à une $(1,1,1)$-surface discrète, nous nous sommes interrogés sur la relation entre les pavages lisses et les $(1,1,1)$-surfaces discrètes.

Dans ce chapitre, nous caractérisons donc les $(1,1,1)$-surfaces discrètes introduites dans (Jamet, 2004) qui sont décrites par un pavage lisse. Nous rappelons d'abord les définitions préalables concernant les surfaces discrètes. Puis, nous décrivons la bijection entre une surface discrète et un pavage du plan. Nous montrons finalement que les seuls mots lisses engendrant de telles surfaces sont les mots $K_{(3,1)}$, $1K_{(3,1)}$ et $2K_{(3,1)}$, où

$$K_{(3,1)} = 3331113331313331113331333 1\cdots$$

est le mot de Kolakoski généralisé sur l'alphabet $\{1,3\}$ commençant par la lettre 3.

Remarque 7.0.12 Dans ce chapitre, la notation $[a, b]$ désignera l'intervalle continu de a à b inclusivement. De plus, $\Pi : \mathbb{R}^3 \longrightarrow \{\vec{x} \in \mathbb{R}^3 \mid (\vec{e_1} + \vec{e_2} + \vec{e_3}, \vec{x}) = 0\}$ est la projection orthogonale dans le plan $\mathcal{P}_0$ d'équation $(\vec{e_1} + \vec{e_2} + \vec{e_3}, \vec{x}) = 0$.

Dans ce chapitre, pour un alphabet $\mathcal{A}$ fixé, on dit que $u \in \mathcal{A}^{\mathbb{Z}^2}$ est un *pavage du plan*. Ce pavage est constitué d'un nombre infini de mots bi-infinis placés les uns en-dessous des autres.

Les résultats de ce chapitre font l'objet de l'article (Jamet et Paquin, 2005).

7.1 Surfaces discrètes

Dans cette section, nous définissons ce que sont les plans discrets et les $(1, 1, 1)$-surfaces discrètes, puis nous rappelons certains résultats les concernant qui nous seront utiles pour la suite.

Définition 7.1.1 Soit $\{\vec{e_1}, \vec{e_2}, \vec{e_3}\}$, la base canonique de l'espace euclidien $\mathbb{R}^3$. Un élément de $\mathbb{Z}^3$ est appelé un *voxel*.

Le nom *voxel* provient de la contraction de *volumetric pixel* : un pixel en trois dimensions.

Définition 7.1.2 Le *cube unité fondamental* $\mathcal{C}$ est l'ensemble défini par

$$\mathcal{C} = \left\{ x_1\vec{e_1} + x_2\vec{e_2} + x_3\vec{e_3} \mid (x_1, x_2, x_3) \in [0, 1]^3 \right\}.$$

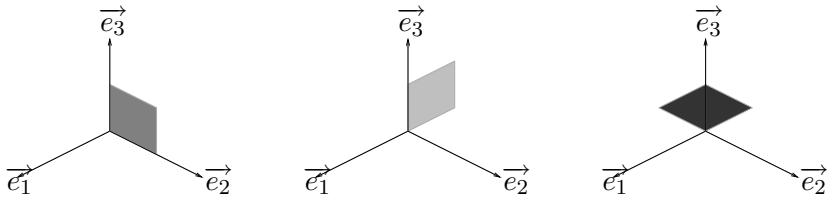
Définition 7.1.3 Soit $\vec{x} \in \mathbb{Z}^3$. Le *cube unité pointé par* $\vec{x}$ est l'ensemble $\vec{x} + \mathcal{C}$: le cube unité fondamental auquel la translation $\vec{x}$ a été appliquée.

Notation 7.1.4 On note $\mathcal{P}$, le plan d'équation $(\vec{v}, \vec{x}) = \mu$ avec $\vec{v} \in \mathbb{R}^3_+$, $\mu \in \mathbb{R}$ et $(\vec{v}, \vec{x}) = v_1x_1 + v_2x_2 + v_2x_3$ est le produit scalaire usuel des vecteurs $\vec{v}$ et $\vec{x}$.

Définition 7.1.5 Soit $\mathcal{P}$, un plan d'équation $(\vec{v}, \vec{x}) = \mu$. Soit $\mathcal{C}_\mathcal{P}$, l'union des cubes unités pointés par un voxel $\vec{x} \in \mathbb{Z}^3$ et intersectant le demi-espace ouvert $(\vec{v}, \vec{x}) < \mu$. On appelle *plan discret associé à l'ensemble* $\mathcal{P}$, le sous-ensemble $\mathfrak{P}_\mathcal{P} = \overline{\mathcal{C}_\mathcal{P}} \setminus \overset{\circ}{\mathcal{C}}_\mathcal{P}$ de $\mathcal{C}_\mathcal{P}$, où $\overline{\mathcal{C}_\mathcal{P}}$ (resp. $\overset{\circ}{\mathcal{C}}_\mathcal{P}$) est la fermeture (resp. l'intérieur) de l'ensemble $\mathcal{C}_\mathcal{P}$ dans $\mathbb{R}^3$, avec sa topologie usuelle.

Définition 7.1.6 On définit les 3 *faces fondamentales* telles qu'illustrées à la Figure 7.1 par

$$\begin{aligned} E_1 &= \{x_2\overrightarrow{e_2} + x_3\overrightarrow{e_3} \mid (x_2, x_3) \in [0,1[^2\}, \\ E_2 &= \{-x_1\overrightarrow{e_1} + x_3\overrightarrow{e_3} \mid (x_1, x_3) \in [0,1[^2\}, \\ E_3 &= \{-x_1\overrightarrow{e_1} - x_2\overrightarrow{e_2} \mid (x_1, x_2) \in [0,1[^2\}. \end{aligned}$$



(a) Face de type 1. (b) Face de type 2 (c) Face de type 3

Figure 7.1 Les 3 faces fondamentales.

Définition 7.1.7 Soit $\vec{x} \in \mathbb{Z}^3$ et $k \in \{1,2,3\}$. L'ensemble $\vec{x} + E_k$ est appelé une *face pointée de type* k. Le vecteur $\vec{x}$ est appelé le *sommet distingué* de $\vec{x} + E_k$.

On a alors le théorème suivant.

Théorème 7.1.8 (Arnoux, Berthé et Siegel, 2004) *Soit un plan discret $\mathfrak{P}_\mathcal{P}$ et soit $\mathcal{V}_\mathcal{P}$ l'ensemble de ses sommets. Supposons que le plan $\mathcal{P}$ admette un vecteur normal $\vec{v} \in \mathbb{R}^3_+$.*

i) *L'ensemble $\mathfrak{P}_\mathcal{P}$ est partitionné en faces pointées.*

ii) *Les fonctions $\Pi_{|\mathcal{V}_\mathcal{P}} : \mathcal{V}_\mathcal{P} \longrightarrow \Pi\left(\mathbb{Z}^3\right)$ et $\Pi_{|\mathfrak{P}_\mathcal{P}} : \mathfrak{P}_\mathcal{P} \longrightarrow \left\{\vec{x} \in \mathbb{R}^3 \mid (\vec{e_1} + \vec{e_2} + \vec{e_3}, \vec{x}) = 0\right\}$ sont bijectives.*

Définissons maintenant les $(1,1,1)$-surfaces discrètes.

Définition 7.1.9 (Jamet, 2004) Une union disjointe $\mathfrak{S} \subseteq \mathbb{R}^3$ de faces pointées est une $(1,1,1)$-*surface discrète* si la fonction

$$\begin{array}{rccc} \Pi_{|\mathfrak{S}} & : & \mathfrak{S} & \longrightarrow & \mathcal{P}_0 \\ & & \vec{x} & \mapsto & \Pi(\vec{x}) \end{array}$$

est bijective. L'ensemble $\mathcal{V}_\mathfrak{S} = \mathfrak{S} \cap \mathbb{Z}^3$ est appelé *l'ensemble des sommets* de $\mathfrak{S}$.

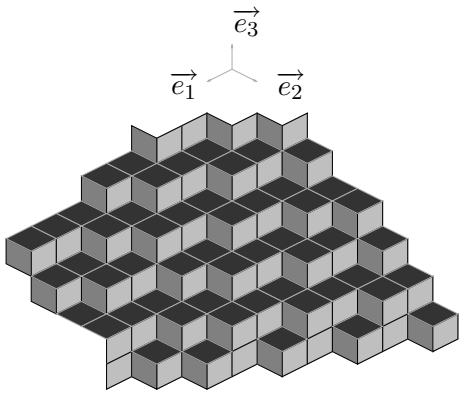


Figure 7.2 Un morceau de surface discrète.

Pour la suite, afin d'alléger le texte, nous appelons les $(1,1,1)$-surfaces discrètes des *surfaces discrètes*. La Figure 7.2 donne un exemple de surface discrète.

Avant de montrer comment associer un pavage du plan sur l'alphabet $\{1,2,3\}$ à n'importe quelle surface discrète $\mathfrak{S}$, rappelons un lemme technique.

Lemme 7.1.10 *Soit $\mathfrak{S}$ une surface discrète. Alors les propriétés suivantes sont vérifiées.*

i) *La fonction*

$$\begin{array}{rccc} \Pi_{|\mathcal{V}_{\mathfrak{S}}} & : & \mathcal{V}_{\mathfrak{S}} & \longrightarrow & \Gamma = \Pi\left(\mathbb{Z}^3\right) \\ & & \vec{x} & \mapsto & \Pi(\vec{x}) \end{array}$$

est bijective.

ii) *Tout sommet $\vec{x}$ de $\mathcal{V}_{\mathfrak{P}}$ est le sommet distingué d'une unique face pointée.*

Le Lemme 7.1.10 permet d'associer un pavage du plan sur l'alphabet $\{1,2,3\}$ à n'importe quelle surface discrète $\mathfrak{S}$ comme suit.

Proposition 7.1.11 *Soit $\Gamma = \Pi(\mathbb{Z}^3) = \mathbb{Z}\Pi(\vec{e_1}) \oplus \mathbb{Z}\Pi(\vec{e_2})$. On identifie Γ et $\mathbb{Z}^2$ par l'isomorphisme*

$$\begin{array}{rccc} \Psi & : & \mathbb{Z}^2 & \longrightarrow & \Gamma \\ & & (m,n) & \mapsto & m\Pi(\vec{e_1}) + n\Pi(\vec{e_2}). \end{array}$$

À toute surface discrète $\mathfrak{S}$, on associe le pavage de plan $u \in \{1,2,3\}^{\mathbb{Z}^2}$ tel que $\forall (m,n) \in \mathbb{Z}^2$, $\forall k \in \{1,2,3\}$,

$$u_{m,n} = k \text{ si et seulement si } \Pi^{-1}_{|\mathcal{V}_{\mathfrak{S}}}(m\Pi(\vec{e_1}) + n\Pi(\vec{e_2})) \text{ est du type } k \text{ dans } \mathfrak{S}.$$

En d'autres termes, $u_{m,n} = k$ *si la pré-image des points* $m\Pi(\vec{e_1}) + n\Pi(\vec{e_2})$ *est du type* k *dans* $\mathfrak{S}$.

La Figure 7.3 illustre comment un morceau de surface discrète est en bijection avec un pavage du plan.

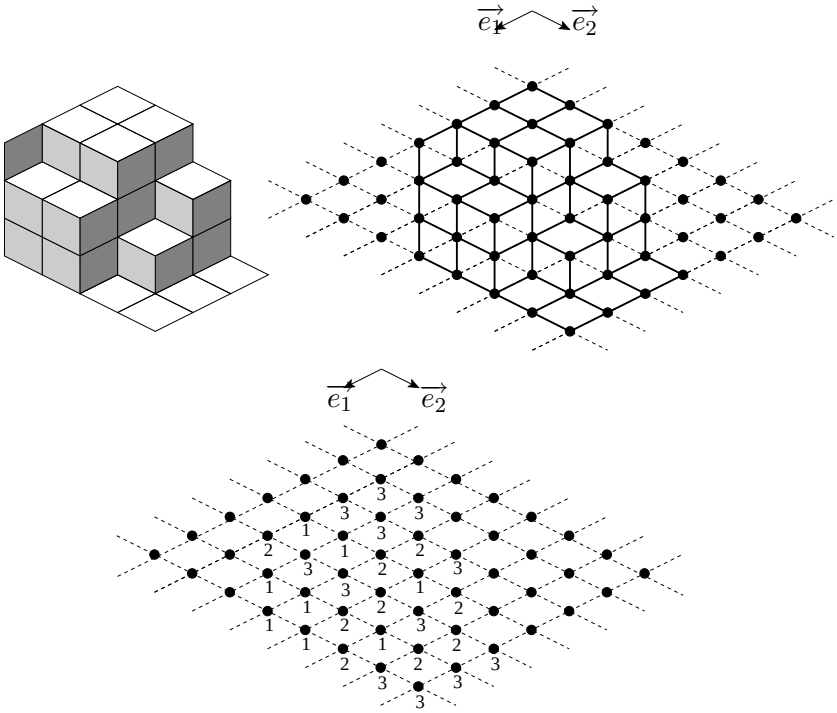


Figure 7.3 Exemple de surface discrète associée à un pavage de plan.

Il est naturel de se demander sous quelles conditions un pavage $u \in \{1,2,3\}^{\mathbb{Z}^2}$ code une surface discrète. La section suivante rappelle ces conditions.

7.2 Reconnaissance de surfaces discrètes

Après avoir rappelé plusieurs définitions, nous énonçons dans cette section le théorème fournissant une condition nécessaire et suffisante pour déterminer si un pavage $u \in \{1,2,3\}^{\mathbb{Z}^2}$ code une surface discrète ou pas.

Définition 7.2.1 Soit $\mathcal{A}$ un alphabet fini. Soit Ω un sous-ensemble fini de $\mathbb{Z}^2$. Une fonction $w : \Omega \longrightarrow \mathcal{A}$ est appelé un *motif fini pointé sur l'alphabet* $\mathcal{A}$.

Définition 7.2.2 Une *forme* $\overline{\Omega}$ de $\mathbb{Z}^2$ est la classe d'équivalence de $\Omega \subseteq \mathbb{Z}^2$ pour la relation d'équivalence

$$\Omega \sim \Omega' \iff \exists (v_1, v_2) \in \mathbb{Z}^2,\ \Omega_1 = \Omega_2 + (v_1, v_2).$$

Définition 7.2.3 Soit Ω un sous-ensemble fini de $\mathbb{Z}^2$. Un *motif fini* $\overline{w}$ de forme $\overline{\Omega}$ est la classe d'équivalence d'un motif fini pointé $w : \Omega \longrightarrow \mathcal{A}$ pour la relation d'équivalence suivante : pour toute paire $w : \Omega \longrightarrow \mathcal{A}$ et $w' : \Omega \longrightarrow \mathcal{A}$ de motifs finis pointés, $w \sim w'$ si et seulement s'il existe un élément $(v_1, v_2) \in \mathbb{Z}^2$ tel que

$$\Omega = \Omega' + (v_1, v_2) \text{ et } \forall (m, n) \in \Omega,\ w_{m,n} = {w'}_{m+v_1, n+v_2}.$$

Dans le but de simplifier les notations, à partir de maintenant, nous notons les motifs finis w au lieu de $\overline{w}$ et les formes Ω au lieu de $\overline{\Omega}$.

Définition 7.2.4 Soit $u \in \mathcal{A}^{\mathbb{Z}^2}$, un pavage du plan sur l'alphabet $\mathcal{A}$ et soit $w : \Omega \longrightarrow \mathcal{A}$ un motif pointé fini. Une *occurrence* de w dans u est un élément $(m_0, n_0) \in \mathbb{Z}^2$ tel que pour tout $(m, n) \in \Omega$, $w_{m,n} = u_{m_0+m, n_0+n}$.

Définition 7.2.5 L'ensemble des motifs finis apparaissant dans u est appelé le *langage* de u et est dénoté $\mathcal{L}(u)$. Pour une forme Ω donnée, l'ensemble des motifs finis de forme Ω apparaissant dans u est appelé le Ω-*langage* de u et est dénoté $\mathcal{L}_\Omega(u)$.

Définition 7.2.6 Soit Ω une forme. La fonction de Ω-*complexité de* u notée P_Ω est définie par

$$\begin{array}{rccc} P_\Omega & : & \mathcal{A}^{\mathbb{Z}^2} & \longrightarrow & \mathbb{N} \cup \{\infty\} \\ & & u & \mapsto & \text{Card}(\mathcal{L}_\Omega(u)). \end{array}$$

Définition 7.2.7 La *forme équerre* est la classe d'équivalence des ensembles

$$\{(m, n); (m, n+1); (m+1, n+1)\},$$

pour $(m, n) \in \mathbb{Z}^2$, pour la relation $\sim$ décrite à la Définition 7.2.3.

La Figure 7.4 illustre différents mots de la forme équerre apparaissant dans un pavage u donné.

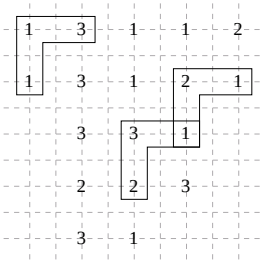


Figure 7.4 Exemples de mots équerres apparaissant dans un mot $u \in \{1,2,3\}^{\mathbb{Z}^2}$.

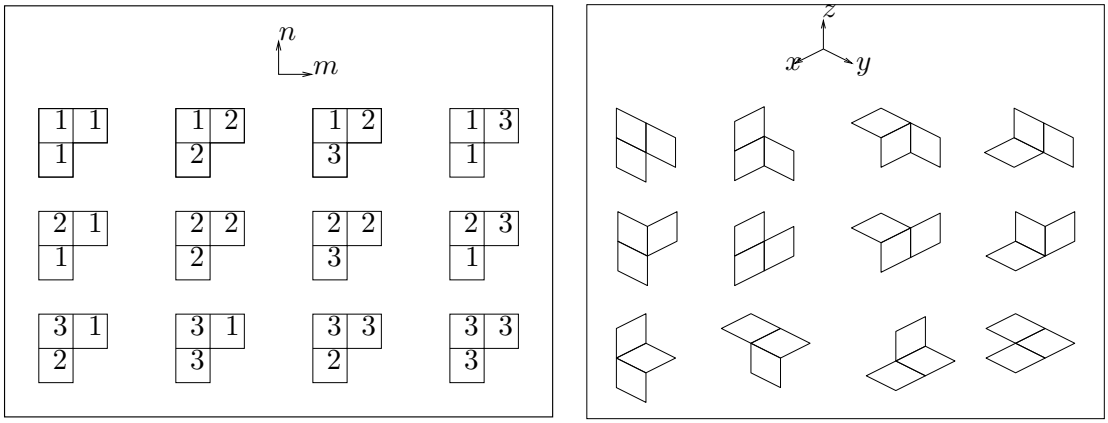


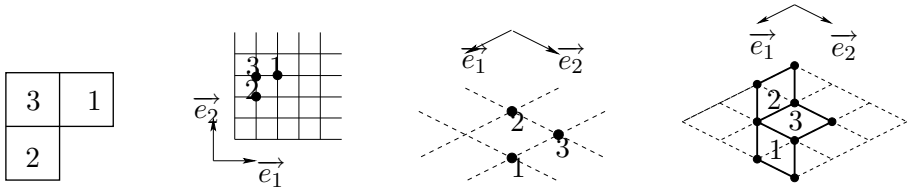
Figure 7.5 Gauche : Les mots équerres autorisés. Droite : La représentation en 3 dimensions des mots équerres autorisés.

Théorème 7.2.8 (Jamet, 2004) *Soit $u \in \{1,2,3\}^{\mathbb{Z}^2}$. Le pavage u décrit une surface discrète si et seulement si le langage équerre de u est inclus dans l'ensemble de motifs de la Figure 7.5.*

La Figure 7.6 donne un exemple de bijection entre un mot équerre et sa représentation en 3 dimensions.

7.3 Mots lisses infinis décrivant des morceaux de surfaces discrètes

Notre objectif est maintenant de caractériser les mots lisses infinis sur l'alphabet $\{1,2,3\}$ pour lesquels le pavage du quart de plan associé décrit un morceau de surface discrète. Nous montrons qu'il n'existe que trois pavages lisses décrivant des morceaux de surfaces discrètes.

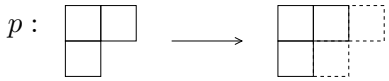


(a) (b) (c) (d)

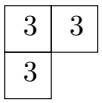
Figure 7.6 Exemple de la bijection d'un motif équerre et de sa représentation en 3 dimensions.

Lemme 7.3.1 *Tout mot équerre peut se prolonger vers la droite en utilisant l'ensemble des mots équerres de la Figure 7.5.*

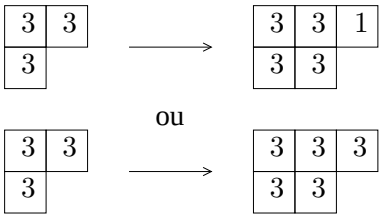
Cette construction se représente par



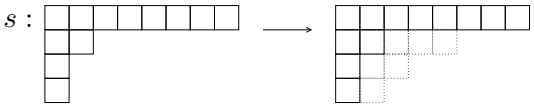
Exemple 7.3.2 Le motif



peut se prolonger vers la droite en utilisant les motifs de la Figure 7.5 de seulement deux façons :

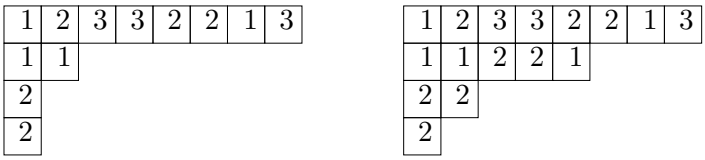


Lemme 7.3.3 *Tout pavage peut se prolonger en utilisant la définition de préfixe lisse. Cette construction se représente par*



et est définie par $R_i = D_r(R_{i-1})$, où R_i désigne la i-ième rangée du pavage et D_r dénote la dérivée à droite d'un mot lisse.

Exemple 7.3.4 En utilisant la condition s, le pavage de la Figure 7.7 (a) est prolongé comme l'illustre la Figure 7.7 (b).



(a) (b)

Figure 7.7 Exemple de prolongement qui utilise la condition s.

Voici maintenant le résultat principal de ce chapitre.

Théorème 7.3.5 *Soit un mot lisse infini $w \in \{1,2,3\}^\omega$. Le pavage du quart de plan associé à w décrit un morceau d'une surface discrète si et seulement si $w \in \{K_{(3,1)}, 1K_{(3,1)}, 2K_{(3,1)}\}$.*

Preuve Par élimination et en utilisant les mots équerres autorisés de la Figure 7.5 et la condition s, on obtient que les seuls mots équerres pouvant être le coin supérieur gauche du pavage lisse décrivant un morceau de surface discrète sont ceux de la Figure 7.8.

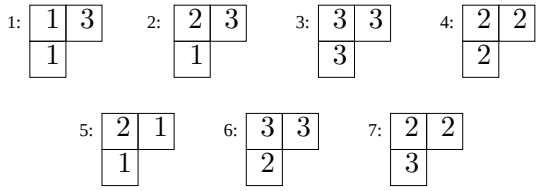


Figure 7.8 Les motifs possibles pour commencer le pavage.

Les 5 autres motifs sont nécessairement exclus, puisqu'ils ne respectent pas la condition s. Nous procédons ensuite en étudiant chacun des 7 cas. Le premier cas est illustré à la Figure 7.9.

Le motif initial peut être prolongé de 4 façons en utilisant les motifs autorisés de la Figure 7.5. Dans les deux premières prolongations du mot initial, la condition s n'est pas satisfaite. Ces dernières sont donc rejetées. Dans la troisième prolongation, si l'on applique la condition s,

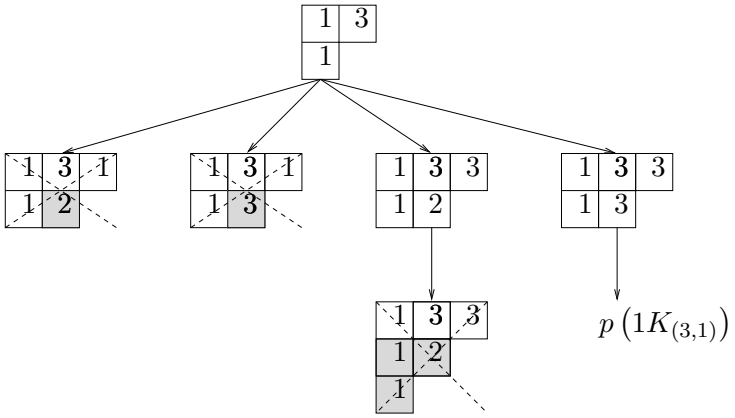


Figure 7.9 Étude du premier cas.

on obtient un motif interdit. Dans la dernière prolongation possible, le pavage associé au mot $1K_{(3,1)}$ est obtenu et on le note $p(1K_{(3,1)})$.

Les autres cas se traitent de façon semblable. Les Figures 7.11, 7.12, 7.13, 7.14, 7.15 et 7.16 les illustrent. L'étiquette de la flèche indique quelle condition a été utilisée pour prolonger le motif.

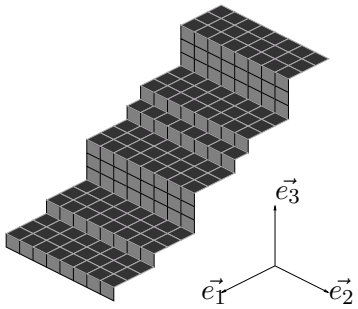


Figure 7.10 Une surface discrète associée au mot $K_{(3,1)}$.

Les seuls cas pour lesquels il existe un pavage lisse décrivant un morceau de surface discrète sont les 3 premiers. Pour les 4 autres, on obtient des contradictions avec les conditions s et p. ■

La Figure 7.10 illustre un morceau de la surface discrète associée à $K_{(3,1)}$.

Comme l'association des 3 faces fondamentales aux lettres 1, 2 et 3 est arbitraire, un problème naturel est de se demander quel serait l'effet d'une permutation des lettres sur notre résultat ? Est-il toujours valide ? Si non, il serait bien de caractériser, pour chaque permutation, les mots lisses infinis décrivant un morceau de surface discrète. Un autre problème intéressant est d'utili-

ser des mots lisses bi-infinis et de caractériser ceux qui ont un pavage lisse décrivant une surface discrète et non pas seulement un morceau.

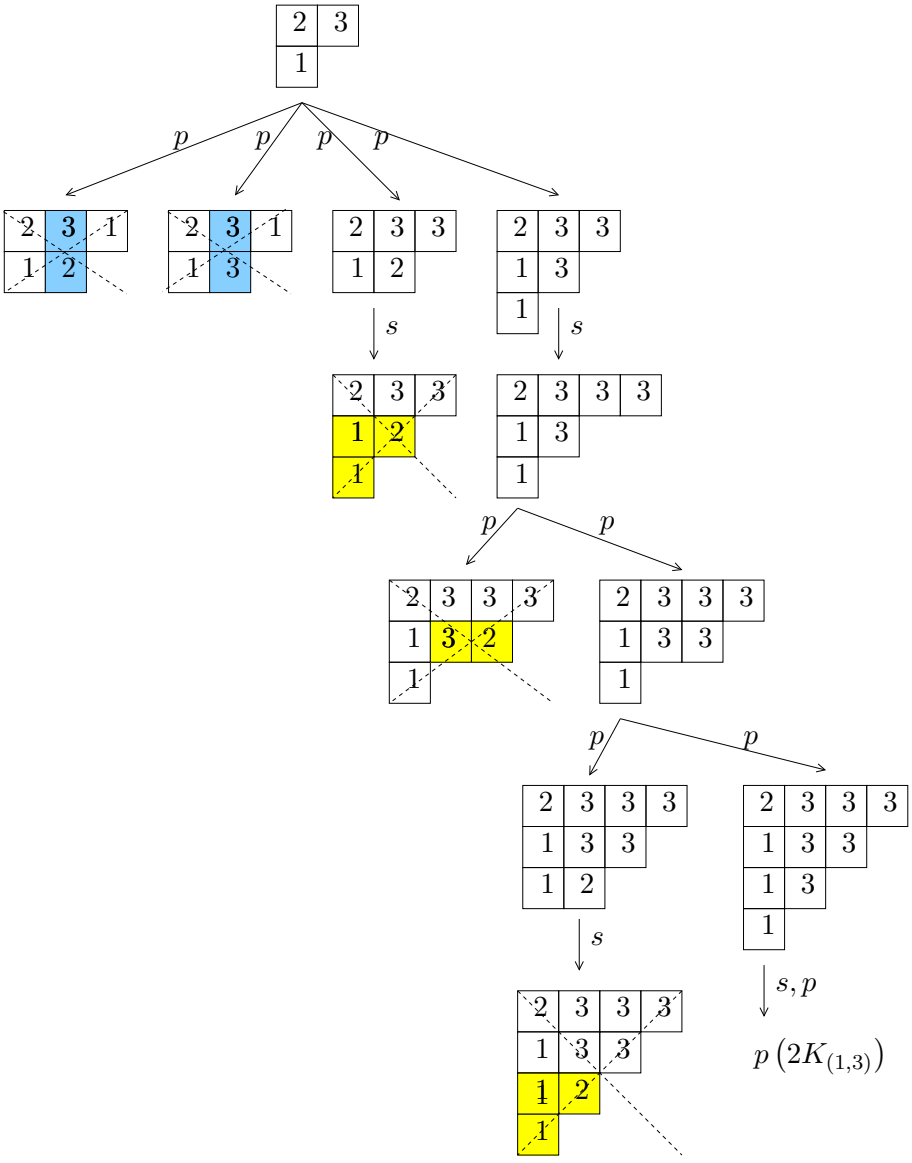


Figure 7.11 Cas 2.

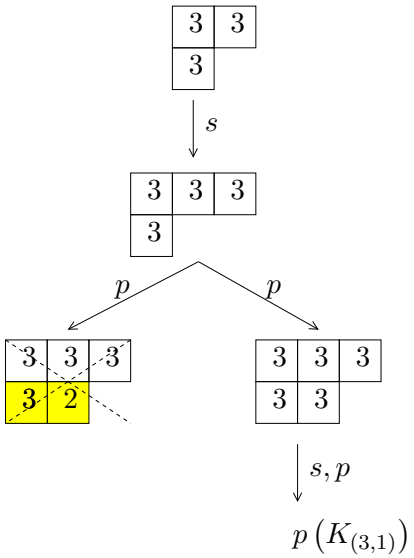


Figure 7.12 Cas 3.

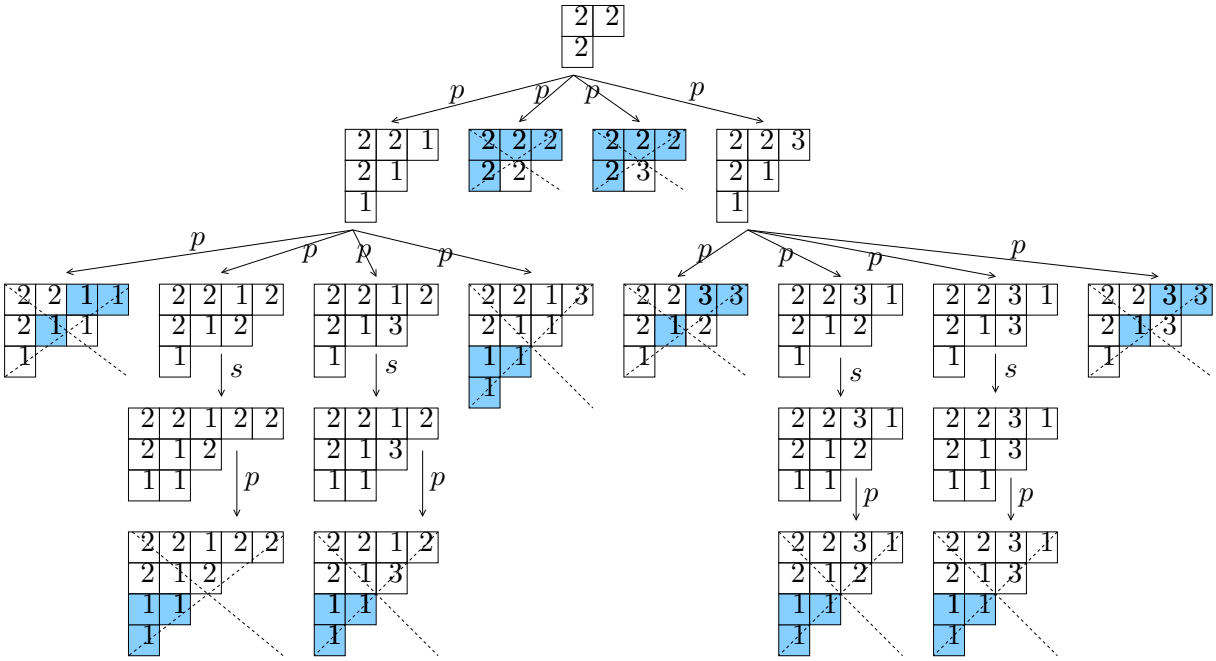


Figure 7.13 Cas 4.

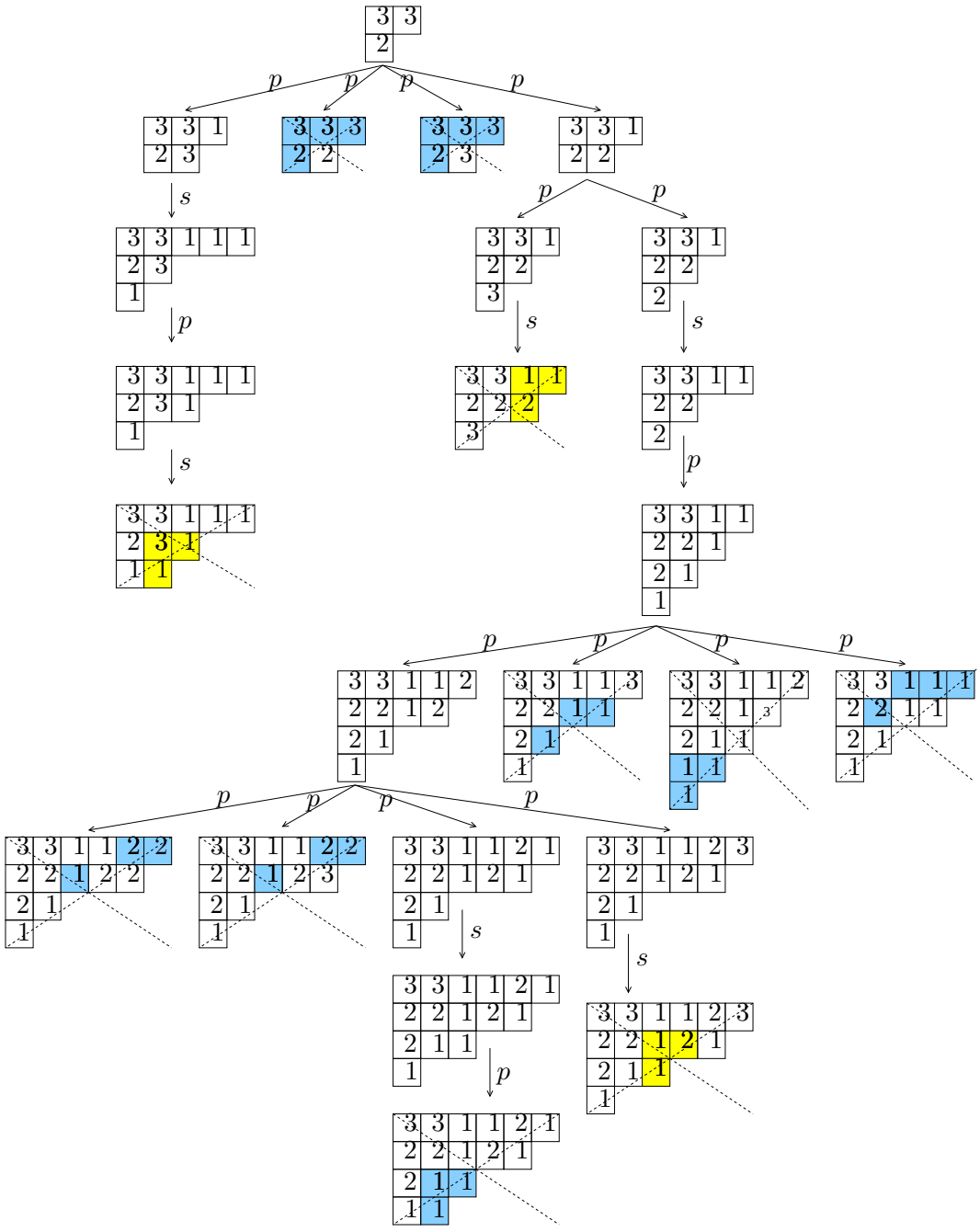


Figure 7.14 Cas 6.

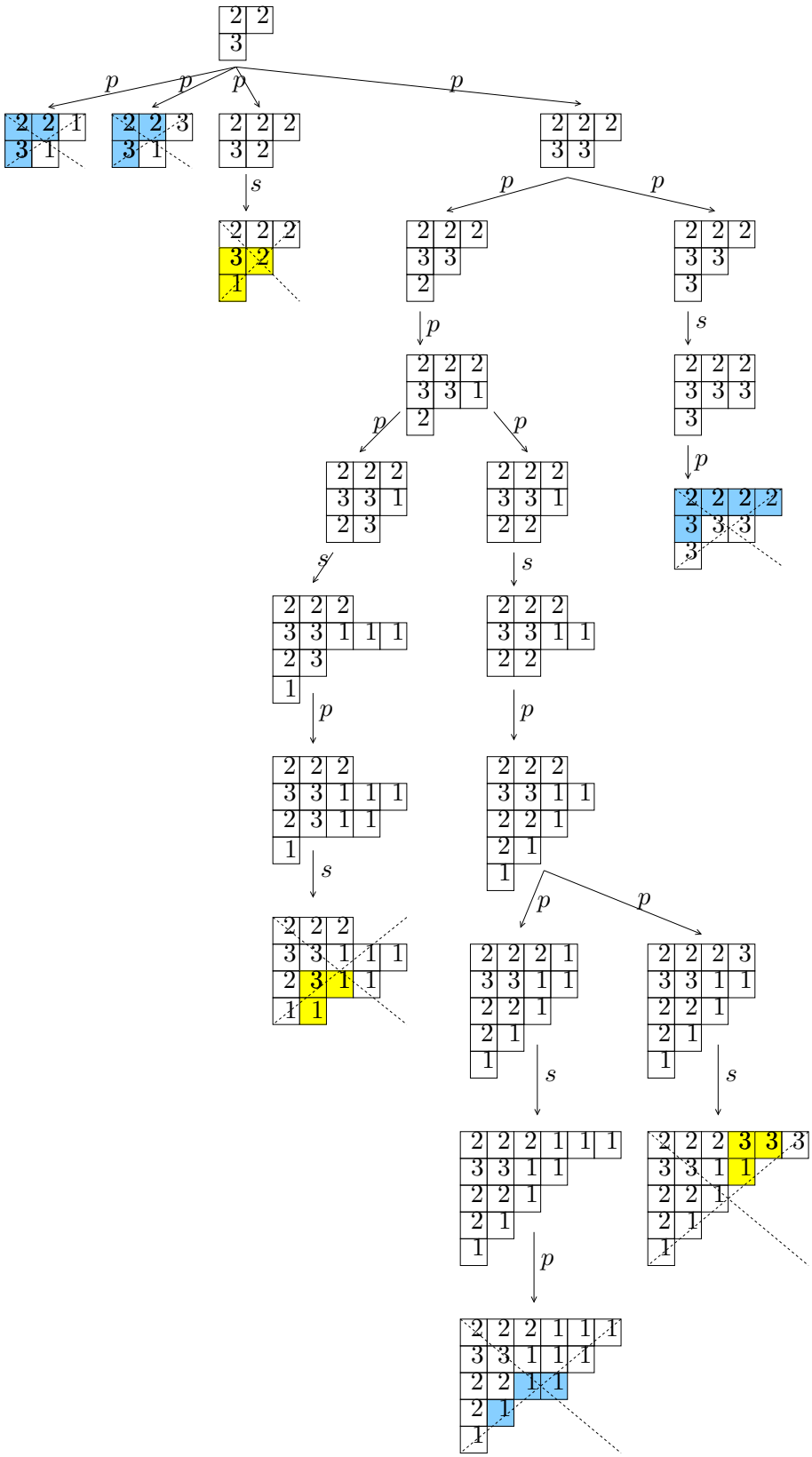


Figure 7.15 Cas 7.

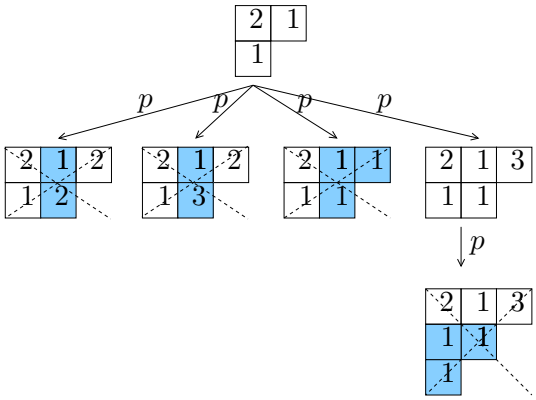


Figure 7.16 Cas 5.

CONCLUSION

Dans ce travail, nous avons étudié deux types de mots : les mots équilibrés et les mots lisses. Pour la première famille de mots, nous avons d'une part généralisé les mots de Christoffel sur un alphabet à 3 lettres et plus, et d'autre part, nous nous sommes intéressés à la conjecture de Fraenkel. Pour la deuxième famille de mots, nous voulions caractériser les mots lisses extrémaux et trouver des façons rapides pour les construire. Nous avions aussi pour objectif de déterminer l'intersection entre la famille des mots lisses infinis et les surfaces discrètes. Nous expliquons ici dans quelle mesure nos objectifs ont été atteints en rappelant les principaux résultats obtenus et nous énonçons certains problèmes ouverts.

Nous avons d'abord défini une nouvelle famille de mots : les mots épichristoffels. Nous avons vu que ces mots semblaient être la bonne généralisation des mots de Christoffel. En effet, tout comme les mots de Christoffel, nous avons montré qu'un mot épichristoffel est primitif et s'écrit toujours de façon unique comme le produit de deux palindromes. Par ailleurs, pour un k-tuplet de fréquences donné, nous avons fourni un algorithme permettant de déterminer si ces fréquences décrivent un mot épichristoffel ou pas. Rappelons que dans le cas de mots de Christoffel, il suffit de vérifier si les fréquences sont premières entre elles. De plus, nous avons montré que tout comme pour les mots de Christoffel, si le k-tuplet décrit les fréquences d'un mot épichristoffel, alors ce dernier est unique et nous avons donné un algorithme pour le construire. Nous avons aussi montré que certains résultats concernant les mots de Christoffel ne se généralisent pas toujours directement. Par exemple, le résultat de (de Luca et de Luca, 2006) ne fonctionne que dans un sens pour les mots épichristoffels : si un mot w est tel que sa racine fractionnaire est dans une classe épichristoffelle, alors w est facteur d'une suite épisturmienne. Pour finir, nous avons prouvé que w est dans une classe épichristoffelle si et seulement si tous ses conjugués sont facteurs d'une même suite épisturmienne. Nous avons trouvé une construction analogue à celle des graphes de Cayley pour les mots de Christoffel, mais malheureusement,

cette dernière ne fonctionne que pour certains mots épichristoffels. Il serait intéressant de parvenir à caractériser cette sous-classe de mots épichristoffels. De plus, il reste encore beaucoup de propriétés des mots de Christoffel qui n'ont pas été généralisées aux mots épichristoffels.

Par la suite, nous nous sommes intéressés aux suites satisfaisant la conjecture de Fraenkel. Dans un premier temps, remarquant que les suites satisfaisant la conjecture de Fraenkel semblent toutes être des suites épisturmiennes, nous avons caractérisé les suites épisturmiennes équilibrées et nous avons montré que parmi ces suites, les seules ayant des fréquences de lettres toutes différentes sont les suites de Fraenkel. Pour prouver la conjecture, il suffirait de montrer que les suites satisfaisant la conjecture de Fraenkel sont des suites épisturmiennes.

Dans un deuxième temps, nous voulions approcher de la conjecture de Fraenkel en étudiant la superposition de mots de Christoffel, comme nous avons montré que les suites satisfaisant la conjecture s'obtiennent par la superposition de k mots de Christoffel. Pour ce faire, nous avons utilisé les résultats connus de (Simpson, 2004) sur les suites de Beatty et que nous avons traduits en termes de mots de Christoffel. Ainsi, nous avons refait les preuves de Simpson en explicitant les détails manquant de ses preuves. Nous avons ainsi obtenu une condition nécessaire et suffisante pour que deux mots de Christoffel se superposent. Notre résultat original dans cette section est le nombre de superpositions de mots de Christoffel. Nous avons non seulement la condition qui permet de dire si deux mots se superposent, mais nous avons en plus le nombre de façons de le faire. Par ailleurs, la traduction des résultats de Simpson nous a permis de trouver de nouvelles propriétés concernant les mots de Christoffel. Par exemple, un mot de Christoffel de la forme $C(n, q\alpha)$ s'obtient par la superposition de q conjugués de $C(n, \alpha)$. Il serait intéressant de généraliser la condition nécessaire et suffisante de superposition de deux mots à trois mots, dans l'espoir de trouver une jolie condition pour la superposition de k mots et ainsi s'approcher de la conjecture de Fraenkel.

Le deuxième type de mots sur lequel nous avons travaillé est la famille des mots lisses. Nous nous sommes intéressés plus particulièrement aux mots lisses extrémaux, c'est-à-dire le plus petit et le plus grand selon l'ordre lexicographique. Nous avons d'abord étudié les mots lisses extrémaux sur les alphabets $\{1, 2\}$ et $\{1, 3\}$. Pour le premier alphabet, le meilleur algorithme pour le construire que nous ayons trouvé est un algorithme nécessitant $\mathcal{O}(n^2)$ opérations, où

n est la longueur du préfixe du mot extrémal construit. Pour le deuxième alphabet, après avoir montré que $\Phi(m_{\{1,3\}}) = (13)^\omega$, où Φ est la bijection naturelle entre les mots lisses et les mots, cette régularité de la structure des mots extrémaux nous fournit un algorithme qui les construit en temps linéraire. De plus, nous avons prouvé que le mot minimal sur l'alphabet $\{1, 3\}$ est dans le même orbite que le mot de Fibonnaci sous l'opérateur de codage par blocs Δ. La différence entre les deux alphabets nous a donné l'idée d'étudier les mots lisses extrémaux pour des alphabets à deux lettres de même parité. Ainsi, nous avons montré que pour un alphabet à deux lettres impaires, nous avons toujours $m_{\{a,b\}} = (ab)^\omega$ et ainsi, il est possible de construire les mots extrémaux sur cet alphabet en temps linéaire. Nous avons aussi prouvé une formule récursive pour construire les mots extrémaux sur l'alphabet $\{1, b\}$ et cette récursion nous a permis de trouver la fréquence des lettres dans les mots extrémaux. Nous avons ensuite prouvé que l'ensemble des facteurs de tout mot lisse sur un alphabet impair est fermé sous l'image miroir, et donc, que tout mot lisse est récurrent. Finalement, nous avons montré que pour l'alphabet impair $\{a < b\}$, le mot minimal est un mot de Lyndon si et seulement si $a = 1$. Pour un alphabet à 2 lettres paires, nous avons prouvé que $\Phi(m_{\{a,b\}}) = ab^\omega$ et par conséquent, nous avons obtenu un algorithme linéaire pour construire les mots lisses extrémaux. De cet algorithme, nous avons déduit la fréquence $1/2$ pour les lettres des mots extrémaux. Plus encore, nous avons montré que les mots lisses sur un alphabet pair sont récurrents, malgré que l'ensemble des facteurs ne soit pas fermé sous l'image miroir. Nous avons aussi montré que les mots lisses minimaux sur un alphabet pair sont tous des mots de Lyndon. Il est intéressant de rappeler que nous avons montré que les mots maximaux coïncident avec les mots de Kolakoski généralisés. Cela implique que les propriétés des mots maximaux s'appliquent aux mots de Kolakoksi généralisés, alors que ces dernières sont encore non prouvées pour le mot de Kolakoski sur l'alphabet $\{1, 2\}$. Il serait intéressant de voir s'il existe d'aussi jolies propriétés pour des alphabets de parité différentes, avec la plus petite lettre paire. De plus, comme nous avons vu que certains mots lisses minimaux sont des mots de Lyndon infinis, la question suivante est naturelle : existe-t'il des mots lisses non minimaux qui sont des mots de Lyndon infinis ?

Pour terminer, nous voulions caractériser l'intersection des surfaces discrètes et des mots lisses. Comme une surface discrète se décrit par un pavage du plan sur l'alphabet $\{1, 2, 3\}$, nous vou-

lions savoir quels sont les pavages lisses du quart de plan décrivant un morceau de surface discrète. Pour ce faire, nous avons utilisé la caractérisation de (Jamet, 2004) des pavages décrivant une surface discrète et la définition des mots lisses. Nous avons construit un mot lisse, donc un mot décrivant un pavage lisse, en vérifiant à chaque étape s'il satisfaisait les conditions de Jamet. Par élimination, nous avons montré qu'il n'existe que trois mots lisses pour lesquels le pavage associé décrit un morceau de surface discrète. Comme l'alphabet $\{1, 2, 3\}$ servant à coder la surface discrète a été fixé de façon arbitraire, il serait pertinent de voir l'impact sur notre résultat du renommage des lettres de l'alphabet.

BIBLIOGRAPHIE

Allouche, J.-P. et J. Shallit. 2003. *Automatic sequences*. Cambridge : Cambridge University Press. Theory, applications, generalizations.

Altman, E., B. Gaujal, et A. Hordijk. 1998. Balanced sequences and optimal routing. Report TW-97-08, Leiden University, The Netherlands.

Arnoux, P., V. Berthé, et A. Siegel. 2004. « Two-dimensional iterated morphisms and discrete planes », *Theoret. Comput. Sci.*, vol. 319, no. 1-3, p. 145–176.

Arnoux, P., C. Mauduit, I. Shiokawa, et J.-I. Tamura. 1994. « Complexity of sequences defined by billiard in the cube », *Bull. Soc. Math. France*, vol. 122, no. 1, p. 1–12.

Arnoux, P. et G. Rauzy. 1991. « Représentation géométrique de suites de complexité $2n + 1$ », *Bull. Soc. Math. France*, vol. 119, no. 2, p. 199–215.

Bang, T. 1957. « On the sequence $[n\alpha], n = 1, 2, \ldots$. Suplementary note to the preceding paper by th. skolem », *Math. Scand*, vol. 5, p. 69–76.

Barát, J. et P. P. Varjú. 2003. « Partitioning the positive integers to seven Beatty sequences », *Indag. Math. (N.S.)*, vol. 14, no. 2, p. 149–161.

Beatty, S. 1926. « Problem 3173 », *Amer. Math. Monthly*, vol. 33, p. 159.

Bender, E. A., O. Patashnik, et J. H. Rumsey. 1994. « Pizza slicing, Phi's and the Riemann hypothesis », *Amer. Math. Monthly*, vol. 101, no. 4, p. 307–317.

Bergeron-Brlek, A., S. Brlek, A. Lacasse, et X. Provençal. 2003. « Patterns in smooth tilings ». In *Proceedings of WORDS'03*. T. 27, série *TUCS Gen. Publ.*, p. 370–381. Turku Cent. Comput. Sci., Turku.

Bernoulli, J. 1772. « Sur une nouvelle espèce de calcul », *Recueil pour les astronomes*, vol. 1, p. 255–284.

Berstel, J. 2002. « Recent results on extensions of Sturmian words », *Internat. J. Algebra Comput.*, vol. 12, no. 1-2, p. 371–385. International Conference on Geometric and Combinatorial Methods in Group Theory and Semigroup Theory (Lincoln, NE, 2000).

Berstel, J. et A. de Luca. 1997. « Sturmian words, Lyndon words and trees », *Theoret. Comput. Sci.*, vol. 178, no. 1-2, p. 171–203.

Berthé, V., S. Brlek, et P. Choquette. 2005. « Smooth words over arbitrary alphabets », *Theoret. Comput. Sci.*, vol. 341, p. 293–310.

Berthé, V., A. de Luca, et C. Reutenauer. 2007. « On an involution of Christoffel words and sturmian morphisms », *European Journal of Combinatorics*.

Berthé, V. et R. Tijdeman. 2004. « Lattices and multi-dimensional words », *Theoret. Comput. Sci.*, vol. 319, no. 1-3, p. 177–202.

Berthé, V. et L. Vuillon. 2000. « Tilings and rotations on the torus : a two-dimensional generalization of sturmian sequences », *Discr. Math.*, vol. 223, no. 1-3, p. 27–53.

Bombieri, E. et J. E. Taylor. 1986. « Which distributions of matter diffract ? An initial investigation », *J. Phys.*, vol. 47, p. 19–28.

Borel, J.-P. 2001. « Image par homographie de mots de Christoffel », *Bull. Belg. Math. Soc. Simon Stevin*, vol. 8, no. 2, p. 241–255. Journées Montoises d'Informatique Théorique (Marne-la-Vallée, 2000).

Borel, J.-P. et F. Laubie. 1993. « Quelques mots sur la droite projective réelle », *Journal de Théorie des Nombres de Bordeaux*, vol. 5, p. 23–51.

Borel, J.-P. et C. Reutenauer. 2005. « Palindromic factors of billard words », *Theoret. Comput. Sci.*, vol. 340, no. 2, p. 334–348.

———. 2006. « On Christoffel classes », *RAIRO-Theoretical Informatics and Applications*, vol. 40, p. 15–28.

Bresenham, J. E. 1991. « Algorithm for computer control of a digital plotter », *IBM Systems J.*, vol. 4, p. 25–30.

Brlek, S. 1989. « Enumeration of factors in the Thue-Morse word », *Discrete Appl. Math.*, vol. 24, p. 83–96.

Brlek, S., A. del Lungo, et A. Ladouceur. 2002. « The class of smooth tilings ». In *Journées Montoises d'Informatique Théorique (Montpellier, France) 9–11 Septembre 2002*.

Brlek, S., S. Dulucq, A. Ladouceur, et L. Vuillon. 2006. « Combinatorial properties of smooth infinite words », *Theoret. Comput. Sci.*, vol. 352, p. 306–317.

Brlek, S., D. Jamet, et G. Paquin. 2008. « Smooth words on 2-letter alphabets having same parity », *Theoret. Comput. Sci.*, vol. 393, p. 166–181.

Brlek, S. et A. Ladouceur. 2003. « A note on differentiable palindromes », *Theoret. Comput. Sci.*, vol. 302, p. 167–178.

Brlek, S., G. Melançon, et G. Paquin. 2007. « Properties of the extremal infinite smooth words », *Discrete Math. Theor. Comput. Sci.*, vol. 9, no. 2, p. 33–49 (electronic).

Brown, T. C. 1993. « Descriptions of the characteristic sequence of an irrational », *Canad. Math. Bull.*, vol. 36, p. 15–21.

Carpi, A. 1993. « Repetitions in the Kolakovski sequence », *Bulletin of the EATCS*, vol. 50, p. 194–196.

———. 1994. « On repeated factors in C^∞-words », *Information Processing Letters*, vol. 52, p. 289–294.

Cassaigne, J., S. Ferenczi, et L. Q. Zamboni. 2000. « Imbalances in Arnoux-Rauzy sequences », *Ann. Inst. Fourier (Grenoble)*, vol. 50, no. 4, p. 1265–1276.

Castelli, M., F. Mignosi, et A. Restivo. 1999. « Fine and Wilf's theorem for three periods and a generalization of Sturmian words », *Theoret. Comput. Sci.*, vol. 218, no. 1, p. 83–94.

Christoffel, E. B. 1875. « Observatio arithmetica », *Math. ann.*, vol. 6, p. 145–152.

Chvátal, V. 1994. Notes on the Kolakoski sequence. Rapport, DIMACS Techn. Rep.

Connell, I. G. 1959. « Some properties of Beatty sequences I », *Canad. Math. Bull.*, vol. 2, p. 190–197.

———. 1960. « Some properties of Beatty sequences II », *Canad. Math. Bull.*, vol. 3, p. 17–22.

Coven, E. M. 1974. « Sequences with minimal block growth II », *Math. Syst. Th.*, vol. 8, p. 376–382.

Coven, E. M. et G. A. Hedlund. 1973. « Sequences with minimal block growth », *Math. Syst. Th.*, vol. 7.

Culik, II, K., J. Karhumäki, et A. Lepistö. 1992. *Alternating iteration of morphisms and the Kolakovski sequence*. Coll. « Lindenmayer systems », p. 93–106. Berlin : Springer.

de Luca, A. 1997a. « Combinatorics of standard sturmian words », *Structures in Logic and Computer Science*, vol. 1261.

———. 1997b. « Sturmian words : structure, combinatorics and their arithmetics », *Theoret. Comput. Sci.*, vol. 183.

de Luca, A. et A. de Luca. 2006. « Some characterizations of finite Sturmian words », *Theoret. Comput. Sci.*, vol. 356, no. 1-2.

Dekking, F. 1980–1981. On the structure of self generating sequences. Séminaire de théorie des nombres de Bordeaux, exposé 31.

Dekking, F. M. 1997. *What is the long range order in the Kolakoski sequence ?* Coll. « The mathematics of long-range aperiodic order (Waterloo, ON, 1995) ». T. 489, série *NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci.*, p. 115–125. Dordrecht : Kluwer Acad. Publ.

Droubay, X., J. Justin, et G. Pirillo. 2001. « Episturmian words and some constructions of de Luca and Rauzy », *Theoret. Comput. Sci.*, vol. 255, p. 539–553.

Durand, F., A. Guerziz, et M. Koskas. 2004. « Words and morphisms with Sturmian erasures », *Bull. Belg. Math. Soc. Simon Stevin*, vol. 11, no. 4, p. 575–588.

Erdos, P. et R. L. Graham. 1980. Old and new problems and results in combinatorial number theory. Monogr. de L'enseignement Math., Genève.

Fraenkel, A. S. 1969. « The bracket function and complementary sets of integers », *Canad. J. Math.*, vol. 21, p. 6–27.

———. 1973. « Complementing and exactly covering sequences », *J. Combin. Theory Ser. A*, vol. 14, p. 8–20.

Fraenkel, A. S., J. Levitt, et M. Shimshoni. 1972. « Characterization of the set of values $f(n) = [n\alpha], n = 1, 2, \cdots$ », *Discr. Math.*, vol. 2, p. 335–345.

Gessel, I. M. et C. Reutenauer. 1993. « Counting permutations with given cycle structure and descent set », *J. Combin. Theory Ser. A*, vol. 64, no. 2, p. 189–215.

Glen, A. 2007. « Powers in a class of $\mathcal{A}$-strict standard episturmian words », *Theoret. Comput. Sci.*, vol. 380, no. 3, p. 330–354.

Glen, A. 2008. « A characterization of fine words over a finite alphabet », *Theoret. Comput. Sci.*, vol. 391, p. 51–60.

Glen, A., J. Justin, et G. Pirillo. 2008. « Characterizations of finite and infinite episturmian words via lexicographic orderings », *European J. Combin.*, vol. 29, no. 1, p. 45–58.

Graham, R. 1973. « Covering the positive integers by disjoint sets of the form $\{[n\alpha + \beta] : n = 1, 2, \ldots\}$ », *J. Combin. Theory Ser. A*, vol. 15, p. 354–358.

Graham, R. et K. O'Bryant. 2005. « A discrete Fourier kernel and Fraenkel's tiling conjecture », *Acta Arith.*, vol. 118, no. 3, p. 283–304.

Graham, R. L. 1963. « On a theorem of Uspensky », *Amer. Math. Monthly*, vol. 70, p. 407–409.

Hedlund, G. A. 1944. « Sturmian minimal sets », *Amer. J. Math.*, vol. 66, p. 605–620.

Hubert, P. 2000. « Well balanced sequences », *Theoret. Comput. Sci.*, vol. 242, no. 1–2, p. 91–108.

Jamet, D. 2004. « On the language of discrete planes and surfaces ». In *Proc. of the Tenth International Workshop on Combinatorial Image Analysis*, p. 227–241. Springer-Verlag.

Jamet, D. et G. Paquin. 2005. « Discrete surfaces and infinite smooth words ». In *FPSAC'05 - 17th annual International conference on Formal Power Series and Algebraic Combinatorics (Taormina, Italie) June 20–25*.

Justin, J. 2000. « On a paper by M. Castelli, F. Mignosi, a. restivo », *Theor. Inform. Appl.*, vol. 34, no. 5, p. 373–377.

———. 2005. « Episturmian morphisms and a Galois theorem on continued fractions », *Theor. Inform. Appl.*, vol. 39, p. 207–215.

Justin, J. et G. Pirillo. 2002. « Episturmian words and episturmian morphisms », *Theoret. Comput. Sci.*, vol. 276, no. 1-2, p. 281–313.

———. 2004. « Episturmian words : shifts, morphisms and numeration systems », *Internat. J. Found. Comput. Sci.*, vol. 15, no. 2, p. 329–348.

Justin, J. et L. Vuillon. 2000. « Return words in Sturmian and episturmian sequences », *Theor. Inform. Appl.*, vol. 34, p. 343–356.

Kassel, C. et C. Reutenauer. 2007. « Sturmian morphisms, the braid group B_4, Christoffel words and bases of f_2 », *Annali di Matematica Pura ed Applicata*, vol. Series IV 186, p. 317–339.

Kolakoski, W. 1965. « Self generating runs, Problem 5304 », *Amer. Math. Monthly*, vol. 72, p. 674.

Kolakoski, W. et N. Ucoluk. 1966. « Problems and Solutions : Solutions of Advanced Problems : 5304 », *Amer. Math. Monthly*, vol. 73, no. 6, p. 681–682.

Lamas, P. 1995. « Contribution à l'étude de quelques mots infinis ». Mémoire de maîtrise, Université du Québec à Montréal.

Lothaire, M. 1983. *Combinatorics on words*. Addison Wesley, Reading MA.

———. 2002. *Algebraic Combinatorics on words*. Cambridge University Press.

———. 2005. *Applied Combinatorics on Words*. Cambridge University Press.

Markov, A. 1882. « Sur une question de Jean Bernouilli », *Math. ann.*, vol. 19, p. 27–36.

Monteil, T., S. Thomasse, et A. Tisserand. 2007. « Brute force Kolakoski », *http ://www.lirmm.fr/ monteil/blog/lire.php/Brute%20force%20Kolakoski*.

Morikawa, R. 1982. « On eventually covering families generated by the bracket function », *Bull. Fac. Liberal Arts Nagasaki Univ.*, vol. 23, p. 17–22.

———. 1983. « On eventually covering families generated by the bracket function II », *Bull. Fac. Liberal Arts Nagasaki Univ.*, vol. 24, p. 1–9.

———. 1984. « On eventually covering families generated by the bracket function III », *Bull. Fac. Liberal Arts Nagasaki Univ.*, vol. 25, p. 1–11.

———. 1985a. « Disjoint sequences generated by the bracket function », *Bull. Fac. Liberal Arts Nagasaki Univ.*, vol. 26, no. 1, p. 1–13.

———. 1985b. « Disjoint sequences generated by the bracket function II », *Number Theory and Combinatorics. Japan 1984 (Tokyo, Okayama and Kyoto, 1984)*, p. 305–321.

———. 1985c. « On eventually covering families generated by the bracket function IV », *Bull. Fac. Liberal Arts Nagasaki Univ.*, vol. 25, p. 1–8.

———. 1988. « Disjoint sequences generated by the bracket function III », *Bull. Fac. Liberal Arts Nagasaki Univ.*, vol. 28, no. 2, p. 1–24.

———. 1989. « Disjoint sequences generated by the bracket function IV », *Bull. Fac. Liberal Arts Nagasaki Univ.*, vol. 30, no. 1, p. 1–10.

———. 1992. « Disjoint sequences generated by the bracket function V », *Bull. Fac. Liberal Arts Nagasaki Univ.*, vol. 32, no. 2, p. 181–185.

———. 1993. « Disjoint sequences generated by the bracket function VI », *Bull. Fac. Liberal Arts Nagasaki Univ.*, vol. 34, no. 1, p. 1–23.

———. 1995. « On eventually covering families generated by the bracket function V », *Bull. Fac. Liberal Arts Nagasaki Univ.*, vol. 36, p. 1–17.

Morse, M. et G. A. Hedlund. 1938. « Symbolic dynamics », *Amer. J. Math*, vol. 60, p. 815–866.

———. 1940. « Symbolic dynamics ii - Sturmian trajectories », *Amer. J. Math*, vol. 62, p. 1–42.

Niven, I. 1963. « Diophantine approximations », *Interscience tracts in pure and applied mathematics*, vol. 14.

Paquin, G. et L. Vuillon. 2007. « A characterization of balanced episturmian sequences », *Electron. J. Combin.*, vol. 14, no. 1, p. Research Paper 33, 12 pp. (electronic).

Pytheas-Fogg, N. 2002. *Substitutions in Dynamics, Arithmetics and Combinatorics*. LNM 1794, Springer, Berlin.

Queffélec, M. 1987. *Substitution Dynamical Systems - A spectral analysis*. T. 1294. Lecture Notes Math., Springer-Verlag.

Rauzy, G. 1985. *Mots infinis en arithmétique in : Automata on Infinite Words*. T. 192. Lecture Notes in Computer Science.

Richomme, G. 2007. « Conjugacy of morphisms and Lyndon decomposition of standard Sturmian words », *Theoret. Comput. Sci.*, vol. 380, no. 3, p. 393–400.

Risley, R. N. et L. Q. Zamboni. 2000. « A generalization of Sturmian sequences : Combinatorial structure and transcendence », *Acta Arith.*, vol. 95, no. 2, p. 167–184.

Séébold, P. 1991. « Fibonacci morphisms and Sturmian words », *Theoret. Comput. Sci.*, vol. 88, p. 367–384.

Séébold, P. 1996. *On the conjugation of standard morphisms*. Coll. « Mathematical foundations of computer science 1996 (Cracow) ». T. 1113, série *Lecture Notes in Comput. Sci.*, p. 506–516. Berlin : Springer.

———. 1998. « On the conjugation of standard morphisms », *Theoret. Comput. Sci.*, vol. 195, no. 1, p. 91–109. Mathematical foundations of computer science (Cracow, 1996).

Simpson, R. 1991. « Disjoint covering systems of rational Beatty sequences », *Discr. Math.*,

vol. 92, p. 361–369.

———. 2004. « Disjoint beatty sequences », *Electronic journal of combinatorial number theory*, vol. 4, p. A12.

Siromoney, R., L. Matthew, V. R. Dare, et K. G. Subramanian. 1994. « Infinite lyndon words », *Information Processing Letters*, vol. 50, p. 101–104.

Skolem, T. 1957. « On certain distributions of integers in pairs with given differences », *Math. Scand.*, vol. 5, p. 57–68.

Steinsky, B. 2006. « A recursive formula for the Kolakoski sequence A000002 », *J. Integer Seq.*, vol. 9, no. 3, p. Article 06.3.7, 5 pp. (electronic).

Storlarsky, K. B. 1976. « Beatty sequences, continued fractions, and certain shift operators », *Canad. Math. Bull.*, vol. 19, p. 473–482.

Sylvester, J. J. 1884. « Question 7382 », *Mathematical Questions from the Educational Times*, vol. 41, no. 21.

Thue, A. 1906. « Uber unendliche zeichenreihen », *Norske Vid. Selsk. Skr. I Math-Nat K1*, vol. 7, p. 1–22.

———. 1910. « Die losung eines spezialfalles eines generellen logischen problems », *Norske Vid. Selsk. Skr. I Math-Nat. K1*, vol. 8.

———. 1912. « Uber die gegenseitige loge gleicher teile gewisser zeichenreihen », *Norske Vid. Selsk. Skr. I Math-Nat. K1*, vol. 1, p. 1–67.

———. 1914. « Probleme uber veranderungen von zeichenreihen nach gegebenen regeln », *Norske Vid. Selsk. Skr. I Math-Nat. K1*, vol. 10.

Tijdeman, R. 1996a. « On complementary triples of sturmian bisequences », *Indag. Math.*, vol. 7, p. 419–424.

———. 1996b. On disjoint pairs of Sturmian bisequences. Report W96-02, Leiden University, The Netherlands.

———. 1998. « Interwinings of periodic sequences », *Indag. Math.*, vol. 9, p. 113–122.

———. 2000a. « Exact covers of balanced sequences and Fraenkel's conjecture », *Algebraic number theory and Diophantine analysis (Graz, 1998), de Gruyter, Berlin*, p. 467–483.

———. 2000b. « Fraenkel's conjecture for six sequences », *Discr. Math.*, vol. 222, p. 223–234.

Uspensky, J. V. 1927. « On a problem arising out of the theory of a certain game », *Amer. Math. Monthly*, vol. 34, p. 516–521.

Vuillon, L. 2003. « Balanced words », *Bull. Belg. Math. Soc. Simon Stevin*, vol. 10, p. 787–805.

Weakley, W. D. 1989. « On the number of C^∞-words of each length », *J. Combin. Theory Ser.*

A, vol. 51, no. 1, p. 55–62.

Weisstein, E. W. 2007. « Coin problem », *Mathworld – A Wolfram web resource, http ://mathworld.wolfram.com/CoinProblem.html*.

Ziccardi, G. 1995. « Parole Sturmiane ». Mémoire de maîtrise, Università degli Studi di Roma.

INDEX