



HAL
open science

Couverture d'un mot bidimensionnel par un motif chevauchant

Guilhem Gamard

► **To cite this version:**

Guilhem Gamard. Couverture d'un mot bidimensionnel par un motif chevauchant. Théorie et langage formel [cs.FL]. Université Montpellier, 2017. Français. NNT : 2017MONT027 . tel-01916701

HAL Id: tel-01916701

<https://theses.hal.science/tel-01916701>

Submitted on 8 Nov 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ACADÉMIE DE MONTPELLIER
UNIVERSITÉ DE MONTPELLIER

Thèse

POUR OBTENIR LE GRADE DE DOCTEUR

présentée par

Guilhem GAMARD

intitulée

**Couverture d'un mot bidimensionnel par
un motif chevauchant**

Spécialité	Informatique
Collège Doctoral	Languedoc-Roussillon
École Doctorale	Information, Structures, Systèmes
Unité de Recherche	Lirmm

Soutenue le vendredi 30 juin 2017, devant le jury composé de :

M.-P. BÉAL	Professeur	Université Paris-Est	Rapporteur
M. CROCHEMORE	Professeur	King's College	Président
E. JEANDEL	Professeur	Université de Lorraine	Rapporteur
A. PINLOU	Maître de conférences	Université Paul-Valéry Montpellier	Examineur
M. RAO	Chargé de recherches	CNRS et ENS de Lyon	Examineur
G. RICHOMME	Professeur	Université Paul-Valéry Montpellier	Directeur

Remerciements

Ma gratitude va en premier lieu à Gwenaël Richomme, pour son écoute, ses conseils toujours avisés, sa grande disponibilité et sa patience. Peu d'éléments de ce travail auraient pu voir le jour sans son aide ; toutefois, toutes les erreurs et faiblesses de ce manuscrit sont de mon fait.

Merci à Marie-Pierre Béal et à Emmanuel Jeandel d'avoir rapporté cette thèse et fourni de nombreux et intéressants retours ; ainsi qu'à Maxime Crochemore, Alexandre Pinlou et Michaël Rao, qui ont accepté de faire partie du jury de soutenance.

Il est nécessaire de mentionner ici mes collègues doctorants, notamment Florian Barbero, Julien Baste, Julien Destombes, François Dross, Valentin Garnero, Anaël Grandjean et Swan Rocher, qui ont fait du LIRMM un endroit Intéressant à vivre.

Les membres d'ESCAPE, passés et présents, m'ont souvent conseillé et créé un environnement scientifique riche et épanouissant. Merci en particulier à Bruno Durand, Andrei Romashchenko, Patrice Séébold et Alexander Shen pour les nombreuses et passionnantes discussions scientifiques qui ont jalonné la préparation de cette thèse, ainsi que pour toutes les idées qui en sont ressorties. Il en va de même pour Pascal Ochem, bien qu'il ne soit pas membre de l'équipe.

Merci à Daria Pchelina, ainsi qu'à Thomas Fattal, Abdulrhman Otman et Laurent Schadeck, pour des raisons qu'il n'est point besoin d'expliquer ici.

Enfin merci à mes parents, Nathalie et Patrice, sans qui cette thèse n'aurait jamais existé, *a minima* pour d'évidentes raisons biologiques ; ainsi qu'à mes frères, Sébastien et Thibaud.

Buy the ticket, take the ride.

— Hunter S. Thompson, *Fear and Loathing in Las Vegas*, 1971

Table des matières

0	Introduction	7
0.1	Contexte scientifique	7
0.2	Motivations et résumé des contributions	11
0.3	Organisation du manuscrit	12
1	Combinatoire des mots finis	15
1.1	Introduction aux mots finis	15
1.2	Périodicité	18
1.3	Première approche de la quasipériodicité	19
1.4	Forme normale des mots quasipériodiques	21
1.5	Compter les mots quasipériodiques	23
2	Combinatoire des mots infinis	27
2.1	Introduction aux mots infinis	27
2.2	Morphismes	31
2.3	Complexité en facteurs	33
2.4	Récurrence	36
2.5	Fréquences de facteurs	38
2.6	Quasipériodicité sur les mots infinis	41
3	Une méthode d'étude des quasipériodes	47
3.1	Détermination des quasipériodes	47
3.2	Exemple d'application : étude des quasipériodes d'un mot	49
3.3	Sur les quasipériodes gigognes	51
3.4	Sur la forme normale et les désstitutions	54
4	Pouvoir expressif de la quasipériodicité	59
4.1	Introduction à la dynamique symbolique	59
4.2	Les mots sturmiens	61
4.3	Caractérisations en termes de quasipériodes	63
4.4	Cas des mots bi-infinis	69
5	Combinatoire des mots en deux dimensions	73
5.1	Motivation : les pavages du plan	73
5.2	Définitions autour des mots 2D	75

5.3	L'entropie bidimensionnelle est bien définie	79
5.4	Un peu de combinatoire en deux dimensions	80
6	Quasipériodicité en deux dimensions	85
6.1	Quasipériodes forçant la périodicité	86
6.2	Configurations quasipériodiques irrégulières	89
6.3	Le cas de l'entropie	90
7	Quasipériodicité multi-échelles en 2D	97
7.1	Entropie topologique	97
7.2	Uniforme récurrence	98
7.3	Fréquences de facteurs	99
8	Perspectives	109
8.1	Problèmes d'algorithmique	109
8.2	Sur les mots biinfinis quasipériodiques	110
8.3	Définir des langages par des quasipériodes	111
8.4	Questions bidimensionnelles	112
8.5	Autres résultats	113
	Bibliographie	115
	Index	121

Je suis parti je ne sais où mais pas où je voulais aller.

— Alain Souchon (chanté par Laurent Voulzy), 1977

Chapitre 0

Introduction

0.1 Contexte scientifique

0.1.1 Mots finis

Les mots finis — aussi appelés chaînes de caractères — sont une structure de données fondamentale en informatique. Ils permettent de représenter du texte, ce qui leur donne un fort intérêt pour de nombreuses applications en programmation et en linguistique. Ils sont au centre de domaines tels que la théorie des langages et la théorie de l'information. Moyen de modélisation, ils sont au cœur d'études ou un outil pour de nombreux domaines orientés applications ou non : algèbre, linguistique, bio-informatique.

Les algorithmes résolvant des problèmes sur les mots finis jouent donc un rôle particulièrement important en théorie comme en pratique. La compression, la recherche de motifs et la reconstruction de mots à partir de contraintes données figurent parmi les tâches les plus courantes [3, 52]. Pour l'immense majorité des problèmes, il est possible de concevoir des algorithmes plus performants si on suppose que l'entrée possède une certaine régularité. Par exemple, il est plus rapide de rechercher un motif dans un mot périodique que dans un mot quelconque ; il est également très simple de compresser un mot périodique. Toutefois, les mots périodiques n'apparaissent pas souvent en pratique. Les algorithmiciens ont donc cherché des propriétés de régularité plus faibles, dans l'espoir de couvrir plus d'entrées, mais qui permettent tout de même d'améliorer leurs algorithmes.

C'est ainsi qu'au début des années 1990, Alberto Apostolico et Andrzej Ehrenfeucht ont introduit la notion de *quasipériodicité* [4]. La quasipériodicité est similaire à la périodicité, excepté que les quasipériodes ont le droit de se chevaucher, tandis que les périodes doivent toujours se suivre. Divers algorithmes ont été proposés pour détecter les quasipériodes d'un mot [34, 48], sous divers modèles : algorithmes *en ligne* [9], parallèles [10], sur des mots circulaires [29]... D'autres recherches visent à détecter les facteurs (sous-mots contigus) quasipériodiques dans une chaîne [4, 45]. La notion de *graine* est une légère généralisation de la notion de quasipériode ; la recherche de graines dans un mot est un problème courant de bio-informatique [14, 13]. Divers autres

problèmes de bio-informatique utilisent les mêmes outils combinatoires que l'étude de la quasipériodicité : le problème de couverture [46], ou encore le problème de la plus courte superchaîne [81].

Il a été montré que les mots quasipériodiques peuvent se décomposer en une forme normale [47, 71]. Celle-ci permet non seulement d'imaginer des algorithmes efficaces pour tester la quasipériodicité d'un mot, mais également d'énumérer les mots quasipériodiques [75], et donc de les compter, ce qui permet une analyse en moyenne de la complexité des algorithmes de détection. D'autres outils purement combinatoires permettent de démontrer l'optimalité de certains algorithmes [42].

0.1.2 Mots infinis

Les mots infinis sont une généralisation des mots finis. Ces objets abstraits se révèlent utiles dans de nombreuses modélisations, pour étudier par exemple les décimales d'un nombre réel, une source de données sans fin, ou encore l'exécution d'un phénomène à temps discret et à nombre fini d'états tel qu'une chaîne de Markov. On les rencontre donc en théorie des nombres, en théorie de l'information ainsi qu'en théorie des systèmes dynamiques [1, 7, 8, 35]. En voici quelques exemples.

$$314159265358979323846264338327950288419716939937510\dots \quad (1)$$

$$\blacksquare \square \blacksquare \square \square \square \blacksquare \square \blacksquare \square \square \square \square \square \square \square \blacksquare \square \blacksquare \square \square \blacksquare \square \blacksquare \dots \quad (2)$$

De la même façon qu'avec les mots finis, certains mots infinis sont plus réguliers que d'autres. Par exemple, la suite

$$aaaaaaaaa\dots$$

est la plus « régulière » possible ; elle possède en outre de très nombreuses propriétés. En particulier, elle est *périodique* et sa période est a . Le mot infini suivant, appelé le mot de Fibonacci

$$abaababaabaababaababaabaaba\dots$$

et dont la définition sera rappelée ultérieurement, est *quasipériodique*, le mot fini aba étant l'une de ses quasipériodes. En effet, le mot de Fibonacci est constitué d'une répétition *ad infinitum* du motif aba , dont certaines occurrences se chevauchent (elles partagent une lettre a) et d'autres s'enchaînent sans partager de lettres. La généralisation de la quasipériodicité aux mots infinis fut proposée par Solomon Marcus [66, 67]. Il posa initialement diverses questions, notamment sur les quasipériodes des mots sturmiens, auxquelles divers auteurs ont apporté des réponses que nous passons maintenant en revue.

Les mots sturmiens, introduits en 1940 [70], constituent la famille de mots la plus célèbre et la plus étudiée. Nous renvoyons le lecteur à [35, chapitre 6] et [61, chapitre 2] pour une introduction complète aux mots sturmiens. Sans rentrer dans les détails techniques, retenons simplement que les mots sturmiens sont moralement les plus réguliers possibles après les mots périodiques. Ils possèdent de nombreuses propriétés

intéressantes et se révèlent utiles dans l'étude de nombreux phénomènes, notamment Tetris [65], le jeu d'échecs [30], et en optimisation combinatoire [43]. Parmi les mots sturmiens, la sous-famille des *sturmiens standards* correspond aux plus réguliers.

Les mots sturmiens sont presque tous quasipériodiques, et il existe une caractérisation de ceux qui ne le sont pas. Les quasipériodes d'un mot sturmien donné, s'il en possède, font également l'objet d'une caractérisation [58, 59]. Ces caractérisations s'étendent à la famille plus générale des épisturmiens [41].

Les *morphismes de monoïdes libres* sont une notion naturelle de transformation d'un mot infini en un autre. Les morphismes préservent la quasipériodicité ; toutefois, certains morphismes ont la propriété plus forte de transformer des mots non quasipériodiques en mots quasipériodiques. Nous les appelons des morphismes faiblement quasipériodiques s'ils transforment *certain*s mots non quasipériodiques en mots quasipériodiques, et morphismes fortement quasipériodiques s'ils transforment *tous* les mots en mots quasipériodiques. Les morphismes fortement quasipériodiques possèdent une caractérisation décidable [57].

Enfin, la quasipériodicité fut étudiée sous l'angle des systèmes dynamiques [68]. Ceci permit de montrer que cette propriété n'implique aucune autre propriété de régularité classique, telle que l'uniforme récurrence ou l'entropie topologique nulle. En d'autres termes, il est possible de construire des exemples de mots quasipériodiques très irréguliers, très désordonnés. Pire encore, ces exemples ne sont pas exceptionnels, mais au contraire très nombreux. Toutefois, la quasipériodicité *multi-échelles*, c'est-à-dire le fait de posséder une infinité de quasipériodes, implique de bonnes propriétés de régularités sur les mots infinis. Cette notion nous intéressera donc tout particulièrement.

0.1.3 Pavages

Un pavage pourrait être décrit comme l'analogue d'un mot infini en deux dimensions qui respecte certaines règles locales. Concrètement, une *tuile* (ou devrait-on dire un *pavé*) est un carré dont les bords sont colorés (figure 1). Un pavage est une disposition de tuiles sur une grille, régulière et infinie dans toutes les directions, telle que deux bords de tuiles en contact ont toujours des couleurs qui correspondent (figure 2). Une même tuile peut apparaître plusieurs fois sur la grille. En général, nous partons d'un jeu de tuiles fini et nous étudions l'ensemble des pavages qu'il peut réaliser.

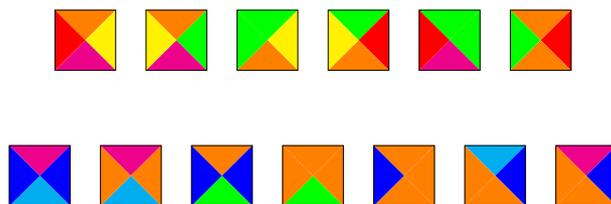


FIG. 1 : Exemple de jeu de tuiles (Kari-Culik [16])

Ce formalisme fut introduit par Hao Wang [82] en 1961, dans le but de développer un prouveur de théorèmes automatique. En effet, dans certains fragments de la logique du premier ordre, il est possible de transformer algorithmiquement une formule en

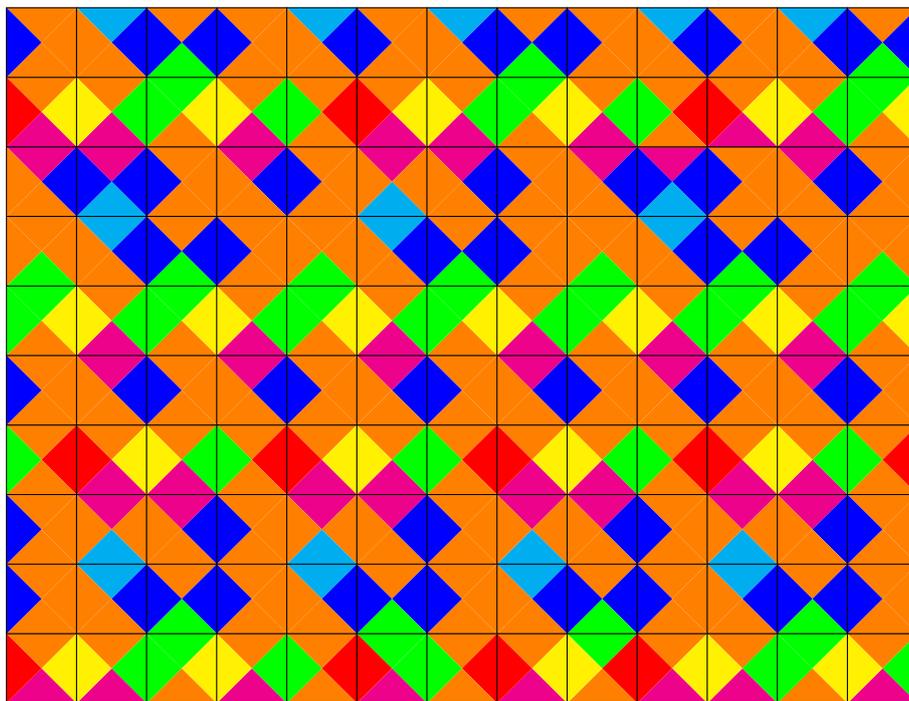


FIG. 2 : Exemple de jeu de pavage par les mêmes tuiles

un jeu fini de tuiles, de telle façon que la formule est satisfiable si et seulement si le jeu de tuiles génère un pavage du plan. L'étude combinatoire du problème du pavage du plan semblait a priori bien plus simple que celle de la logique du premier ordre, et il y avait un espoir que l'informatique naissante puisse résoudre le problème par recherche exhaustive. Ceci aurait permis de démontrer par ordinateur des théorèmes formalisables en logique du premier ordre.

Toutefois, la question de savoir si un jeu de tuiles fini pave le plan ou non fut montrée indécidable [6]. En effet, certains jeux de tuiles peuvent « embarquer » un calcul en leur sein, et paver revient alors à effectuer ce calcul [25]. Une question naturelle survient alors : qu'est-il possible d'embarquer dans un jeu de tuiles ? La difficulté ici réside dans le fait que les tuiles ne portent en elles qu'une information extrêmement locale. Il est par exemple difficile, en premier lieu, de concevoir un jeu de tuiles qui ne génère que des pavages apériodiques (de façon non-triviale, c'est-à-dire qu'il faut générer au moins un pavage du plan). De nombreux auteurs se sont attaqués à ce problème [16, 49, 50, 74, 76].

La question initiale peut se reformuler ainsi : « quelles propriétés globales peuvent être forcées par des règles locales [24] ? » Outre l'apériodicité, il est possible de concevoir des jeux de tuiles résistants aux erreurs [26], ou encore qui ne génèrent que des pavages de complexité algorithmique élevée [23].

Les pavages trouvent des applications en infographie [55], en logique, en théorie de l'information et en combinatoire [18]. Toutefois, la quasipériodicité sur les pavages n'a jamais été étudiée. En deux dimensions, les seuls travaux à notre connaissance sont un algorithme de détection des quasipériodes carrées à l'intérieur d'un bloc carré [15]. Le

but de cette thèse est de généraliser les travaux concernant la quasipériodicité, exposés dans la section précédente, aux pavages. Ceci offrirait des perspectives afin d’obtenir une meilleure compréhension combinatoire de certains d’entre eux.

0.2 Motivations et résumé des contributions

Parmi tous les jeux de tuiles apériodiques connus, celui introduit par Jarkko Kari et amélioré par Karel Culik en 1996 fut longtemps considéré comme le plus petit — 13 tuiles. Au cours de travaux précédant cette thèse, l’auteur chercha à comprendre la structure des pavages générés par le jeu de Kari-Culik. Les premiers résultats de ces travaux montraient que ces pavages n’étaient pas auto-similaires [21, 22], ce qui contrastait fortement avec tous les jeux de tuiles apériodiques découverts avant 1996.

Ces pavages ne possèdent, à première vue, aucune propriété de régularité classique : uniforme récurrence, entropie topologique nulle. . . Toutefois, ils possèdent une certaine régularité, qu’il est difficile de décrire avec les outils habituels de la théorie. Moralement, ces pavages arborent un mot sturmien sur chacune de leurs lignes. Toutefois, ces lignes ont parfois des *erreurs* ; elles ne sont parfois pas exactement des mots sturmiens, mais elles en sont très proches.

Les mots sturmiens sont fortement liés à la quasipériodicité [41, 59] ; nous développerons encore ce lien dans cette thèse. En outre, les pavages de Kari-Culik montrent un comportement ressemblant à de la quasipériodicité. Il a donc semblé approprié d’étudier la quasipériodicité sur les pavages plutôt que sur les mots infinis. En outre, la quasipériodicité peut également être vue comme une règle locale ; dans ce contexte, il est naturel de poser la question classique des pavages : « quelles propriétés globales peuvent être forcées par la quasipériodicité ? »

Toutefois, la quasipériodicité n’avait jamais été abordée sous cet angle, ni dans aucun autre formalisme proche. La première étape fut donc de généraliser nos connaissances actuelles au sujet de la quasipériodicité au cas de la dimension 2.

Nous avons ainsi généralisé tous les résultats de [68] aux mots infinis bidimensionnels (ou *configurations*, pour faire court). Plus précisément, nous avons montré qu’il est possible de construire de nombreuses configurations quasipériodiques irrégulières ; nous avons également montré que les configurations possédant une infinité de quasipériodes ont nécessairement une certaine forme de régularité. En développant ces résultats, nous avons construit divers outils combinatoires d’intérêt général pour l’étude des mots infinis à deux dimensions. Enfin, nous avons montré comment la quasipériodicité, vue comme une règle locale sur les configurations, pouvait forcer des propriétés globales, à la façon des pavages.

En voulant poursuivre ces travaux, nous nous aperçûmes que certaines questions qui nous semblaient naturelles en deux dimensions n’avaient pas été traitées en une dimension, nous travaillâmes donc à les résoudre. Nous avons mis au point une méthode pour déterminer l’ensemble des quasipériodes d’un mot infini et nous avons caractérisé les sturmiens standards en termes de quasipériodes.

Les contributions sont décrites plus en détails dans le résumé des chapitres qui figure à la section suivante.

0.3 Organisation du manuscrit

Les résultats ne sont pas présentés dans l'ordre chronologique : les théorèmes en dimension 2 ont été développés en premier, mais ils apparaissent dans les derniers chapitres. Ce document suit un plan allant du plus particulier au plus général, partant des rappels essentiels sur les mots finis pour aboutir à des résultats sur les mots infinis en deux dimensions.

Le chapitre 0 fournit une introduction générale au sujet, puis donne une description détaillée chapitre par chapitre de cet ouvrage.

Le chapitre 1 présente les définitions et outils essentiels sur les mots finis. Nous étudions tout particulièrement la *primitivité* et la *périodicité*, ainsi que les théorèmes liés. Ces notions seront les briques de base de nombreux raisonnements dans ce manuscrit. Nous présentons ensuite la *quasipériodicité*, notre principal objet d'étude. Nous exposons les résultats combinatoires de la littérature sur le sujet. En particulier, nous montrons comment tout mot quasipériodique peut se décomposer selon une certaine *forme normale*. La forme normale se généralise aux mots infinis et nous aidera à comprendre la combinatoire des mots quasipériodiques.

Le chapitre 2 présente les définitions et principaux résultats liés aux mots infinis. Nous nous focalisons sur les *propriétés de régularité* que peut posséder un mot infini, en donnant les définitions de trois notions abondamment étudiées : *l'uniforme récurrence*, *l'entropie topologique nulle* et la *convergence des fréquences de facteurs*. Ces trois propriétés sont orthogonales, et permettent donc de quantifier trois aspects différents de la « régularité » d'un mot. Nous abordons également les *morphismes de monoïdes libres*, un outil bien connu et très puissant qui nous sera très utile au chapitres 3 et 4.

Ensuite, nous présentons plus la quasipériodicité sur les mots infinis, et plus particulièrement deux résultats importants de la littérature. En premier lieu, pour presque toute quasipériode q , il existe des mots q -quasipériodiques arbitrairement « irréguliers » (en particulier, ils ne possèdent aucune des propriétés de régularité mentionnées précédemment). En second lieu, les mots possédant une *infinité* de quasipériodes, appelés les mots quasipériodiques multi-échelles, possèdent les trois propriétés de régularité mentionnées ci-avant.

Le chapitre 3 apporte une réponse à une question très simple et naturelle, mais qui n'avait pourtant pas été traitée dans la littérature. *Étant donné un mot infini arbitraire, quelles sont ses quasipériodes ?* En toute généralité, il est difficile de répondre ; tout dépend de la façon dont nous est présenté le mot. Étant donné que nous manipulons des objets infinis, il n'est pas possible de donner directement un algorithme. La question est donc plus proche de : *étant donné un mot infini arbitraire, que peut-on dire de ses quasipériodes ?* Nous donnons une méthode permettant de déterminer l'ensemble de quasipériodes d'un mot infini, en connaissant les préfixes spéciaux droits et les préfixes dont le carré apparaît dans le mot [37]. La combinatoire des mots dispose des outils nécessaires pour étudier ces préfixes particuliers dans un mot arbitraire, et ainsi caractériser l'ensemble de ses quasipériodes. Nous en faisons la démonstration

en étudiant les quasipériodes de plusieurs exemples, qui montrent au passage que la structure de l'ensemble des quasipériodes varie considérablement d'un mot à un autre.

Le chapitre 4 commence par présenter les mots sturmiens, une famille de mots bien connue et extrêmement étudiée. Ils sont parfois décrits comme les mots les plus réguliers après les périodiques. Nous mentionnerons les résultats de la littérature qui étudient les quasipériodes des mots sturmiens, puis nous montrerons comment les mots périodiques et certains mots sturmiens peuvent être caractérisés en termes de quasipériodes [37]. En d'autres termes, il s'agit exactement des mots dont l'ensemble des quasipériodes obéit à une condition donnée. La quasipériodicité a donc un certain pouvoir expressif, et il s'agit d'une propriété de régularité pertinente. Nous expliquons enfin pourquoi il est tentant de généraliser ces résultats au cas des mots bi-infinis, et pourquoi c'est une question plus difficile qu'il n'y paraît.

Le chapitre 5 est en quelque sorte l'alter ego du chapitre 1 dans le cadre des mots bidimensionnels et des pavages. Il présente les définitions, résultats et outils de base afin de faire de la combinatoire sur les mots bidimensionnels infinis. Un certain nombre de ces résultats sont des généralisations des théorèmes classiques de combinatoire des mots en une dimension. Toutefois ils sont récents et furent développés en partie par l'auteur, en collaboration avec Gwenaël Richomme, Jeffrey Shallit et Taylor J. Smith [40]. Ces deux derniers n'avaient pas de motivations liées à la quasipériodicité, ce qui montre que ces résultats ont un intérêt général.

Le chapitre 6 traite des mots infinis bidimensionnels (ou *configurations* pour faire court) quasipériodiques, c'est-à-dire possédant une quasipériode. Contrairement au cas des mots, nous ne savons pas décomposer une configuration quasipériodique en *forme normale*, tous les résultats deviennent donc bien plus délicats à prouver. Nous explorons la question suivante : une propriété de la quasipériode peut-elle forcer une propriété sur les configurations qui la possèdent ? Nous donnons une condition nécessaire et suffisante pour qu'une quasipériode q admette des mots q -quasipériodiques mais non périodiques. Nous montrons ensuite une famille de quasipériodes non triviales qui forcent les configurations correspondantes à avoir une entropie topologique nulle. Ce résultat contraste fortement avec le cas unidimensionnel, où une quasipériode seule ne pouvait rien forcer d'autre que la périodicité. Enfin, nous montrons que l'uniforme récurrence et la convergence des fréquences ne peuvent pas être forcées par une unique quasipériode [38].

Le chapitre 7 traite des configurations quasipériodiques multi-échelles, c'est-à-dire possédant une infinité de quasipériodes. Notons que, de façon similaire au cas des mots, la quasipériodicité multi-échelles sur les configurations implique une entropie topologique nulle et la convergence des fréquences de facteurs. Nous montrerons ce dernier résultat sans utiliser d'outils de théorie ergodique [36], contrairement à ce qui avait été fait dans la littérature dans le cas des mots unidimensionnels. Enfin, nous explorerons les liens entre quasipériodicité multi-échelles et uniforme récurrence, qui sont plus complexes que dans le cas unidimensionnel, pour des raisons d'origines. En effet, nos résultats sur les mots considéraient des suites indicées par \mathbb{N} , tandis que nos

résultats sur les configurations considèrent des suites indicées par \mathbb{Z}^2 . Les problèmes que nous rencontrons ici ne viennent pas du passage de la dimension 1 à la dimension 2, mais du passage de \mathbb{N} à \mathbb{Z} .

Le chapitre 8 fait office de conclusion. Nous y présentons une liste de problèmes ouverts et de perspectives de recherches qui n'ont pu être abordées dans cette thèse. Nous mentionnons également quelques résultats développés au cours de la préparation de cette thèse, mais qui n'ont pas été abordés dans ce manuscrit, car sans lien direct avec la notion de quasipériodicité.

Philosophy is divided into two parts : Computer Science and Nonsense.

— Alexander Shen, *Séminaire Eco-Escape*, Montpellier, 2011

Chapitre 1

Combinatoire des mots finis

Comme indiqué en introduction, les mots finis, sont une structure fondamentale en informatique. Les mots finis ne sont pas notre principal sujet d'étude — nous nous concentrerons sur les mots infinis, en une et deux dimensions — mais ils sont au cœur de tous nos raisonnements, y compris en deux dimensions. Ce chapitre présente donc les principales définitions et propriétés dont nous nous servons par la suite.

1.1 Introduction aux mots finis

Un *alphabet* est un ensemble fini non vide ; les éléments d'un alphabet sont appelés les *lettres*. Fixons un alphabet Σ . Pour tout entier naturel n , nous notons Σ^n l'ensemble des *mots de n lettres sur Σ* . Formellement, il s'agit du produit cartésien de n copies de Σ , dont nous notons les composantes par juxtaposition : $x = x_0x_1 \dots x_{n-1}$ en lieu et place de l'écriture habituelle $x = (x_0, x_1, \dots, x_{n-1})$. La première lettre est donc d'indice 0 et la dernière d'indice $n - 1$.

Posons

$$\Sigma^* = \bigcup_{n \in \mathbb{N}} \Sigma^n$$

l'ensemble des *mots finis* sur l'alphabet Σ . Si x est un élément de Σ^* , nous notons $|x|$ sa *longueur*, l'unique entier n tel que x soit élément de Σ^n . Remarquons que Σ^* , indépendamment de l'ensemble Σ , contient toujours *le mot vide*, noté ε , qui est l'unique mot vérifiant $|\varepsilon| = 0$. Voici des exemples de mots sur l'alphabet $\Sigma = \{a, b, c\}$.

$$aaaaaaaaa \tag{1.1}$$

$$abcabcabcabc \tag{1.2}$$

$$\tag{1.3}$$

$$ccabccabccabcca \tag{1.4}$$

$$b \tag{1.5}$$

$$abacabbaccabbbaecca \tag{1.6}$$

$$baabaabbabaabababaabaabaabbabababbaabbaba \tag{1.7}$$

Les deux premiers exemples sont frappants, car ils sont constitués de répétitions d'un même motif. Afin de formaliser cette caractéristique, nous définissons l'opération de *concaténation*. Si $x = x_0 \dots x_{n-1}$ et $y = y_0 \dots y_{m-1}$ sont deux mots de Σ^* , la concaténation de x et y , notée $x \cdot y$ ou plus souvent xy , est le mot de longueur $|x| + |y|$ défini par

$$xy = x_0 \dots x_{n-1}y_0 \dots y_{m-1}.$$

La concaténation n'est pas commutative, en particulier car les mots ab et ba sont différents. Toutefois, elle est associative : la relation $x \cdot (y \cdot z) = (x \cdot y) \cdot z$ est vérifiée pour tous mots x, y et z , nous noterons donc simplement xyz cette valeur. Enfin, le mot vide est un élément neutre : $x \cdot \varepsilon = \varepsilon \cdot x = x$ pour tout mot x . Ces propriétés font du triplet $(\Sigma^*, \cdot, \varepsilon)$ un *monoïde non commutatif*, appelé *le monoïde libre* sur Σ . Nous y reviendrons par la suite.

La concaténation se notant multiplicativement, il est naturel de définir, pour tout mot u et pour tout entier n ,

$$\begin{aligned} u^0 &= \varepsilon \\ u^{n+1} &= u \cdot u^n. \end{aligned}$$

En d'autres termes, u^n est la concaténation de u avec lui-même n fois. Ainsi, le mot de l'équation (1.1) peut s'exprimer a^8 et le mot de l'équation (1.2) peut se réécrire $(abc)^4$. Si u est un mot et n, m sont des entiers, la loi des puissances

$$(u^n)^m = u^{(n \times m)}$$

est vérifiée.

Si w est un mot qui peut s'écrire u^n , pour un mot fini u et un entier naturel n , nous disons que u est une *racine* de w . Un mot est toujours racine de lui-même, et les racines ne sont pas nécessairement uniques : le mot (1.1) admet $a, aa, aaaa$ et a^8 pour racines, tandis que le mot (1.2) admet $abc, abcabc$ et $(abc)^4$ pour racines. Un mot qui ne possède pas d'autre racine que lui-même est appelé *primitif*.

Considérons w un mot fini et u sa plus courte racine ; alors u est nécessairement primitif. En effet, nous aurions dans le cas contraire $w = u^n$ et $u = v^m$, donc par loi des puissances $q = v^{n \times m}$, pour v un mot fini et n, m des entiers supérieurs ou égaux à 2. Ceci contredirait la minimalité de la longueur de u . En conséquence, nous appellerons la plus courte racine d'un mot sa *racine primitive*.

La primitivité nous permet de caractériser les couples (x, y) qui commutent sous concaténation. Comme nous l'avons vu, si x, y sont des mots finis, nous avons généralement $xy \neq yx$. Il existe toutefois des exceptions : par exemple, si $x = abc$ et $y = abcabc$, alors $xy = yx = (abc)^3$. Le théorème suivant donne des conditions nécessaires et suffisantes pour que la concaténation commute.

Théorème 1.1 (Lyndon, Schutzenberger, 1962 [64]). *Soient x et y des mots finis non vides. Les affirmations suivantes sont équivalentes.*

1. *L'égalité $xy = yx$ est vérifiée.*
2. *Les mots x, y possèdent une racine commune z .*
3. *Il existe des entiers strictement positifs n, m tels que $x^n = y^m$.*

Nous aurons souvent besoin de manipuler des mots finis, munissons-nous donc d'un lexique approprié. Si f, p, u et s sont des mots satisfaisant l'équation

$$u = p f s, \tag{1.8}$$

nous disons que p est un *préfixe* de u , que s est un *suffixe* de u et que f est un *facteur* de u . Ce dernier terme provient d'une lecture algébrique de l'équation (1.8). Si $p \neq u$ (respectivement $s \neq u$), nous disons que p est un *préfixe propre* (respectivement un *suffixe propre*) de u . La *position* de f dans u est l'entier $|p|$ et nous disons que f *couvre* les positions $|p|$ à $|p| + |f| - 1$ (incluses). Deux facteurs *se chevauchent* s'ils couvrent au moins une position commune. Remarquons que les préfixes et les suffixes sont toujours des facteurs. Lorsqu'un facteur apparaît dans une position autre que préfixe ou suffixe, nous l'appelons un *facteur interne*. Les préfixes du mot (1.2), *abcabcabcabc*, sont

$$\begin{aligned} \varepsilon \quad & a \quad ab \quad abc \quad abca \quad abcab \quad abcabc \quad abcabca \quad abcabcab \quad abcabcabc \\ & abcabcabca \quad abcabcabcab \quad abcabcabcabc \end{aligned}$$

et les suffixes du mot (1.4), *ccabccabccabcca*, sont

$$\begin{aligned} \varepsilon \quad & a \quad ca \quad cca \quad bcca \quad abcca \quad cabcca \quad ccabcca \quad bccabcca \\ & abccabcca \quad cabccabcca \quad ccabccabcca \quad bccabccabcca \\ & abccabccabcca \quad cabccabccabcca \quad ccabccabccabcca. \end{aligned}$$

Le mot (1.7), *baabaabbabaabababaabaabaabbabababbaabbaba*, comprend 666 facteurs différents; nous laissons au lecteur le soin de les énumérer. Le facteur *bab* apparaît aux positions 15, 18, 20 et 22 de ce même mot, et les deux dernières occurrences se chevauchent car elles couvrent toutes les deux la position 22.

Comme nous venons de le constater, un facteur f peut apparaître plusieurs fois dans u ; nous parlons alors des *occurrences* de f dans u . Par de multiples occurrences, f peut être à la fois préfixe, facteur interne et suffixe de u . Ainsi, le facteur *abc* apparaît en position préfixe, interne et suffixe dans le mot (1.2). Le mot vide apparaît à toutes les positions de tous les mots. Enfin, le mot u est à la fois un préfixe et un suffixe de u lui-même, mais non un facteur interne. Nous notons $|u|_f$ le nombre d'occurrences de f dans u .

Par abus de notation, si u est un mot commençant par la lettre a , nous notons $a^{-1}u$ le mot u privé de sa première lettre. De la même façon, si u termine par la lettre a , nous notons ua^{-1} le mot u privé de sa dernière lettre.

Forts de ce vocabulaire, nous pouvons poursuivre notre introduction à la combinatoire des mots. Il existe un algorithme efficace permettant de tester si un mot est primitif ou non.

Proposition 1.2 (Folklore). *Un mot u est primitif si et seulement s'il n'est pas un facteur interne de u^2 .*

Démonstration. Procédons par contraposée et supposons qu'il existe des mots p, s non vides tels que $u^2 = pus$. Observons que $|u| = |p| + |s|$. Comme p est un préfixe et s un suffixe de u , nous déduisons que $u = ps$. L'équation $u^2 = pus$ se réécrit alors

$$psps = pus$$

et, en simplifiant par p à gauche et par s à droite, nous déduisons que $u = ps = sp$. Par le théorème 1.1, les mots s et p sont tous deux puissances d'un même mot z . Alors u est également puissance de z , donc non primitif.

Réciproquement, si u n'est pas primitif, alors $u = v^n$ pour un mot v non vide et un entier n supérieur ou égal à 2. Ainsi, $u^2 = v^{2n}$ et nous pouvons prendre $p = v$ et $s = v^{2n-2}$ comme solution de l'équation $u^2 = pus$. \square

Pour tester la primitivité d'un mot u , il suffit de chercher une occurrence de u dans u^2 . Cette opération s'effectue en temps linéaire [52]. De nombreuses opérations sur les chaînes de caractères, notamment de recherche de sous-chaînes et de compression, peuvent s'effectuer plus rapidement sur des entrées non primitives si l'on connaît la racine primitive : la combinatoire sert bien l'algorithmique.

1.2 Périodicité

Considérons à nouveau le mot (1.4), $ccabccabccabcca$. Bien que primitif, il possède une forte régularité. Si un mot w vérifie

$$w_i = w_{i+k}$$

pour tous les entiers i tels que $i, i+k \in \{0, \dots, |w| - 1\}$, alors nous disons que k est une *période* de w . Nous parlerons parfois de la période pour désigner le préfixe $w_0w_1w_2 \dots w_{k-1}$. Un mot w possédant une période $k \leq |w|/2$ est appelé *périodique*. Un tel mot peut s'écrire

$$w = u^n u',$$

avec un entier $n \geq 2$ et un préfixe u' de u . Le mot (1.4) admet 4 (ou $ccab$) comme période et se décompose en $(ccab)^3cca$. La propriété « être non primitif » implique « être périodique » : si $w = u^n$ (pour u non vide et $n \geq 2$) alors w admet u comme période avec $u' = \varepsilon$ dans l'équation (1.2). L'inverse n'est pas vrai : le mot (1.4) est périodique sans être une puissance.

Le lemme suivant donne une écriture en équations utile lorsque nous rencontrons un facteur périodique dans nos raisonnements.

Lemme 1.3 (Voir [63, prop. 1.3.4]). *Soient x, y et z des mots finis. L'égalité $xy = yz$ est vérifiée si et seulement s'il existe des mots u, v non vides et un entier naturel k tels que*

$$x = uv \quad y = (uv)^k u \quad z = vu.$$

Comme nous l'avons déjà mentionné, un mot fini peut admettre plusieurs périodes. Par exemple, le multiple d'une période (plus court que le mot considéré) est de nouveau une période : $(ab)^8$ admet ab , $(ab)^2$, et $(ab)^4$ comme périodes et comme racines ; il admet $(ab)^3$ comme période, mais pas comme racine. De façon plus générale, les périodes sont liées par des relations de divisibilité.

Théorème 1.4 (Fine, Wilf, 1965 [33]). *Soit w un mot admettant deux périodes m et n . Si $|w| \geq n + m - \text{pgcd}(m, n)$, alors $\text{pgcd}(m, n)$ est également une période de w .*

Nous possédons maintenant tous les résultats sur les mots finis utiles pour la suite de ce travail. Le lecteur désirant approfondir ses connaissances sur les mots pourra consulter un dictionnaire, ou à défaut l'ouvrage de référence par M. Lothaire [63].

Nous avons présenté la théorie des mots d'un point de vue combinatoire, car c'est le point de vue que nous adopterons dans la suite de ce travail. Toutefois, il existe également une approche algébrique : nous pouvons voir les mots finis comme des produits d'indéterminées sur un monoïde général, et chaque résultat sur les mots finis se traduit donc en un résultat d'algèbre. Par exemple, le théorème 1.1 peut parfaitement se lire algébriquement. Cette autre approche est davantage développée dans l'ouvrage [61], qui contient également de nombreux développements sur les mots finis.

Enfin, de nombreux résultats de combinatoire des mots peuvent être exploités pour la conception d'algorithmes efficaces sur les chaînes de caractères. L'ouvrage [62] donne une large introduction à diverses applications, notamment algorithmiques, de la discipline.

1.3 Première approche de la quasipériodicité

Entrons dans le vif du sujet. Comme nous l'avons mentionné, diverses notions de combinatoire des mots trouvent des applications en algorithmique du texte. Par exemple, la proposition 1.2 induit un algorithme efficace de test de primitivité. En outre, de nombreuses opérations sur les mots, telles que la compression, la recherche de motifs etc., peuvent être effectuées plus efficacement si l'entrée est garantie d'être périodique. Toutefois, peu de situations permettent d'assurer que l'entrée sera périodique. Les algorithmiciens ont donc travaillé à définir de nouvelles propriétés de régularité, plus faibles que la périodicité — avec l'espoir de couvrir un plus grand nombre d'entrées — mais apportant suffisamment d'informations sur le mot pour permettre la conception d'algorithmes rapides.

C'est dans ce contexte qu'en 1993, Apostolico et Ehrenfeucht ont introduit la notion de *quasipériodicité* [4]. Soient q et u deux mots finis différents. Nous disons que q est une *quasipériode* de u si et seulement si toute position n de u est couverte par une occurrence de q comme facteur de u . En d'autres termes, pour tout entier n entre 0 et $|u| - 1$, il existe un entier i compris entre 0 et $|q| - 1$ tel que

$$0 \leq n - i \leq n - i + |q| - 1 \leq |u| - 1 \text{ et} \\ q = u_{n-i}u_{n-i+1} \dots u_{n-i+|q|-1}$$

Ainsi, le mot

$$abaababaabaaba$$

admet deux quasipériodes, aba et $abaaba$ (voir figure 1.1). Insistons sur le fait qu'un mot n'est jamais quasipériode de lui-même. Un mot admettant au moins une quasipériode est dit *quasipériodique*; dans le cas contraire, il est appelé *superprimitif*.



FIG. 1.1 : Mise en évidence des quasipériodes aba et $abaaba$.

Si q est une quasipériode de u , alors q est nécessairement un préfixe et un suffixe de u . En effet, la première et la dernière lettre de u doivent être couvertes par des occurrences de q , et celles-ci ne peuvent pas « déborder » du mot selon notre définition. Le mot

$$baababaabaab,$$

obtenu en supprimant la première et la dernière lettre de l'exemple précédent, est superprimitif.

Considérons de nouveau l'exemple $abaababaabaaba$ et ses deux quasipériodes, aba et $abaaba$. Le mot aba est une quasipériode de $abaaba$. Il s'agit en fait d'une propriété générale.

Remarque 1.5. Soit u un mot fini et q_1, q_2, \dots, q_k ses quasipériodes, ordonnées par longueurs croissantes. Pour tout $i \in \{1, \dots, k-1\}$, le mot q_{i+1} est q_i -quasipériodique. En particulier, q_1 est l'unique quasipériode superprimitive de w .

Démonstration. Étant donné que q_i et q_{i+1} sont toutes deux préfixes et suffixes de w , le mot q_i est un préfixe et un suffixe de q_{i+1} (voir figure 1.2). Or w est q_i -quasipériodique, donc q_{i+1} doit être recouverte d'occurrences de q_i . \square

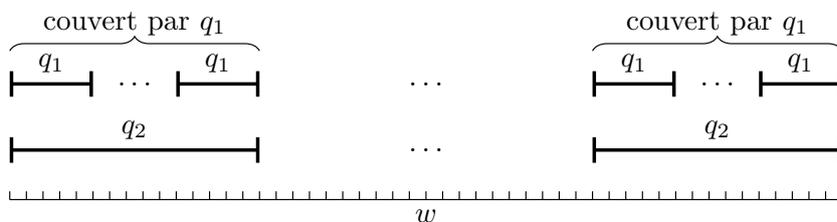


FIG. 1.2 : Un mot fini possède une unique quasipériode superprimitive.

1.4 Forme normale des mots quasipériodiques

En 2000, Laurent Mouchard démontre [71] que tout mot quasipériodique peut s'écrire sous une certaine *forme normale*. La forme normale se révélera extrêmement précieuse tout au long de notre travail, aussi voyons immédiatement les notions nécessaires à sa formulation.

Soit q un mot fini. Si r est à la fois un préfixe propre et un suffixe propre de q , nous disons que r est un *bord* de q . Ainsi q n'est jamais un bord de lui-même. Si $2|r| \leq |q|$, alors le mot q s'écrit :

$$q = r \cdot s \cdot r$$

avec s un mot fini. Toutefois, si $2|r| > |q|$, alors les deux occurrences de r se chevauchent et q n'admet pas de telle décomposition avec r . Ainsi, les bords du mot *abacabaccabacaba* sont ε, a, aba et *abacaba*, tandis que les bords de *ababab* sont ε, ab et *abab*. Tout mot puissance admet au moins ses racines, ainsi que les bords de ses racines, comme bords. Le mot vide est un bord de tout mot.

Si q est un mot et r un bord de q , alors les mots t et t' tels que

$$q = t \cdot r \quad \text{et} \quad q = r \cdot t'$$

sont appelés respectivement un *antibord gauche* et un *antibord droit* de q . Lorsque nous aurons besoin de préciser, nous écrirons que t est l'antibord gauche (et t' l'antibord droit) *par rapport à r* dans q . L'antibord gauche est toujours un préfixe de q et l'antibord droit un suffixe. Le mot vide n'est jamais un antibord. Nous appelons L_q l'ensemble des antibords gauches et R_q l'ensemble des antibords droits de q . Ainsi,

$$\begin{aligned} L_{aba} &= \{aba, ab\} & R_{aba} &= \{aba, ba\} \\ L_{abaaba} &= \{abaaba, abaab, aba\} & R_{abaaba} &= \{abaaba, baaba, aba\}. \end{aligned}$$

Voyons maintenant le théorème de forme normale de Mouchard.

Théorème 1.6 (Mouchard, 2000 [71]). *Soient q et u des mots finis. Le mot u est égal à q ou admet q comme sa plus petite quasipériode si et seulement s'il peut s'écrire*

$$u = q \cdot v_1 \cdots v_n$$

où les v_i sont tous des antibords droits de q .

De façon équivalente, u peut se décomposer sous la forme $v'_1 \cdots v'_n \cdot q$, où les v'_i sont des antibords gauches de q .

Dans les deux cas, la décomposition est unique.

Démonstration. Si u admet q comme quasipériode, toute occurrence de q , sauf la première, peut s'écrire $q = xy$ où x est la partie (éventuellement vide) partagée avec l'occurrence de q précédente et y le reste. Le mot x est un préfixe de q , mais aussi un suffixe puisqu'il apparaît à la fin de l'occurrence de q précédente; donc y est un antibord par définition. La décomposition s'obtient immédiatement par récurrence sur le nombre d'occurrences de q dans u .

Réciproquement, supposons que tout mot de la forme $u' = q \cdot v_1 \dots v_{n-1}$, avec les v_i des antibords de q , soit q -quasipériodique. Montrons que $u = q \cdot v_1 \dots v_{n-1} \cdot v_n$ l'est également. Comme u' est quasipériodique, il admet q comme suffixe et donc tous les bords de q comme suffixe. En particulier, le bord b tel que $q = bv_n$ est suffixe de u' , et donc $u = u'v_n$ admet q comme suffixe. Les $|v_n|$ dernières lettres sont bien couvertes par une occurrence de q , et l'hypothèse de récurrence implique que le reste du mot est également couvert par des occurrences de q , donc u est q -quasipériodique.

Enfin, supposons que u est q -quasipériodique et admet deux décompositions distinctes $u = qv_1 \dots v_n = qv'_1 \dots v'_n$, et montrons que q est quasipériodique (donc non superprimitif). Sans perte de généralité, nous supposons que $v_1 \neq v'_1$ et $|v'_1| < |v_1|$, autrement dit qu'il existe un mot w tel que $qv'_1w = qv_1$. Ce mot possède trois occurrences de q : la première termine à la position $|q| - 1$, la deuxième termine à la position $|q| + |v'_1| - 1$ et la troisième termine à la position $|q| + |v_1| - 1$. La deuxième occurrence partage donc plus de la moitié de ses lettres avec l'une des deux autres ; en d'autres termes, il existe des mots finis p et s tels que $pq = qs$ avec $|p| = |s| \leq |q|/2$. Le lemme 1.3 montre qu'il existe des mots finis x , y et un entier k tels que $p = xy$, $q = (xy)^kx$ et $s = yx$. Par un argument de longueur, nous avons $k \geq 2$ et le mot xyx est une quasipériode de q .

La plus petite quasipériode d'un mot est superprimitive (la relation « est quasipériode de » étant transitive), donc si q est la plus petite quasipériode de u , la décomposition doit être unique. \square

Par exemple, le mot $w = abaababaabaaba$ s'écrit respectivement sur L_{aba} et R_{aba} :

$$aba \cdot ab \cdot aba \cdot aba \cdot aba \quad aba \cdot aba \cdot ba \cdot aba \cdot aba$$

Rien, dans la première partie de la preuve du théorème 1.6, ne fait appel au fait que q est la plus petite quasipériode de u . Il est donc possible de décomposer de la même façon un mot quasipériodique u sur une autre quasipériode q . Par exemple, le même mot w se décompose sur L_{abaaba} et sur R_{abaaba} :

$$abaab \cdot aba \cdot abaaba \quad abaaba \cdot baaba \cdot aba$$

Si q est lui-même quasipériodique, la décomposition n'est pas nécessairement unique. Par exemple le mot $(aba)^4 = abaabaabaaba$ admet $abaaba$ pour quasipériode, et peut se décomposer comme $abaaba \cdot abaaba$ ou $aba \cdot aba \cdot abaaba$. Afin d'obtenir une forme normale, nous devons donc prendre v_1 de longueur maximale, puis v_2 de longueur maximale, et ainsi de suite de façon gloutonne.

Étant donnée une quasipériode q fixée, nous pouvons utiliser le théorème 1.6 afin de construire un test efficace de q -quasipériodicité. Il suffit de construire un automate déterministe reconnaissant les antibords de q , à la façon de l'algorithme de Knuth-Morris-Pratt [52], puis de tester l'entrée avec cet automate. Il existe des techniques bien plus puissantes : dans l'article original [4], Apostolico et Ehrenfeucht donnent un algorithme permettant de déterminer tous les *facteurs* quasipériodiques maximaux (ainsi que leurs quasipériodes respectives) d'un mot donné en temps linéaire. Ici la maximalité s'entend

comme le fait que la prolongation du facteur d'une lettre à gauche ou à droite casse la quasipériodicité de celui-ci. La présentation de cet algorithme constituerait toutefois une très longue digression, nous renvoyons donc le lecteur intéressé à l'article *op. cit.* D'autres algorithmes ont depuis été proposés pour détecter les quasipériodes d'un mot en temps quasilinéaire [53, 34], dont certains optimaux [42].

Le théorème 1.6 a une autre conséquence : si q est un mot dont le seul bord est vide, alors tout mot q -quasipériodique u est égal à une puissance de q .

Remarque 1.7. Soit q un mot fini. Il existe un mot q -quasipériodique primitif seulement si la racine primitive de q possède un bord non vide.

Démonstration. Si la racine primitive r de q ne possède aucun bord non vide, alors tous les bords de q sont des puissances de r , et donc les antibords également. Par le théorème 1.6, tout mot q -quasipériodique admet donc r comme racine. □

En particulier, si la quasipériode q est unaire, c'est-à-dire qu'elle est une puissance d'une lettre, alors tout mot q -quasipériodique u est également unaire. Ainsi, les quasipériodes $a, b, aa, bb, ab, ba, aaa, bbb, baa, bba, aab$ et abb forcent tous leurs mots quasipériodiques à être non primitifs. Les plus courtes quasipériodes qui autorisent des mots quasipériodiques primitifs sont donc aba et bab . Ces dernières deviendront donc nos quasipériodes de référence, nos exemples de choix ; ceci suggère *Barbara Ann* comme accompagnement musical à cette thèse.

Le théorème 1.6 met également en lumière la relation entre primitivité et superprimitivité. En effet, un mot q -quasipériodique w commence par qt , avec t un antibord droit de q ; en particulier, si $q = rt$, alors w commence par $r tt$ et contient donc un carré. Plus généralement, un mot w admettant une quasipériode q contient au moins autant de carrés que d'occurrences de q , c'est-à-dire au moins autant de carrés que d'éléments dans sa forme normale.

1.5 Compter les mots quasipériodiques

Le problème de recherche de facteurs carrés dans un mot est classique en algorithmique du texte (voir [78] et [62, chapitre 8]). La quasipériodicité — et la forme normale — permettent de générer facilement des mots primitifs contenant un grand nombre de carrés. Il suffit par exemple de concaténer des occurrences de aba et ba au hasard, en commençant par aba , pour générer de tels mots. La quasipériodicité peut donc servir à générer des exemples d'entrées défavorables pour certains algorithmes.

Dans ce contexte, la question du *comptage* des mots quasipériodiques devient pertinente, car le comptage permet de réaliser des analyses de complexité en moyenne. Nous allons donner une méthode pratique, reposant sur le théorème 1.6, pour déterminer le nombre de mots de longueur n qui admettent aba pour quasipériode. Nous nous basons sur l'article [75] de Ronney Polley et Ludwig Staiger, qui va bien plus loin et donne le comportement asymptotique du nombre de mots aba -quasipériodiques.

Par le théorème de Mouchard, l'ensemble des mots aba -quasipériodiques est l'ensemble des mots qui s'écrivent comme une concaténation d'éléments de $\{aba, ba\}$ commençant par aba . Soit Q l'ensemble des mots aba -quasipériodiques et E l'ensemble des mots acceptés par l'automate fini déterministe \mathcal{A} , représenté sur la figure 1.3. Alors $E = Q \cup \{aba\}$. Nous pouvons donc compter les mots acceptés par \mathcal{A} , et en déduire simplement le nombre de mots aba -quasipériodiques.

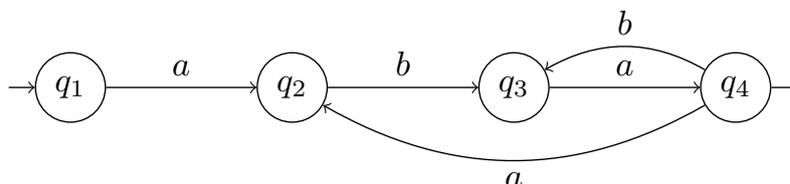


FIG. 1.3 : L'automate \mathcal{A} reconnaissant les mots aba -quasipériodiques.

Le nombre de mots de E de longueur n est exactement le nombre de chemins de longueur n reliant q_1 à q_4 dans \mathcal{A} . Définissons la *matrice d'adjacence* de \mathcal{A} , notée $\mathcal{M}_{\mathcal{A}}$, dont le coefficient (i, j) est le nombre de transitions de q_j vers q_i ,

$$\mathcal{M}_{\mathcal{A}} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

Une démonstration par récurrence sur n prouve sans difficulté que le vecteur

$$V_n = (\mathcal{M}_{\mathcal{A}})^n \cdot (1000)^t,$$

où X^t désigne la transposée de X , est tel que la $i^{\text{ème}}$ composante $V_n(i)$ est exactement le nombre de chemins de longueur n entre q_1 et q_i . Ainsi, le nombre de mots de E de longueur n est donné par

$$(0001) \cdot (\mathcal{M}_{\mathcal{A}})^n \cdot (1000)^t$$

Il s'agit du coefficient inférieur gauche de $(\mathcal{M}_{\mathcal{A}})^n$, cette matrice pouvant être calculée efficacement par exponentiation rapide.

La figure 1.4 montre le tracé de $Q \cap \Sigma^n$ pour n variant de 0 à 50 ; la tendance observée se poursuit pour les valeurs plus grandes. Une étude asymptotique du comportement de cette fonction passe par le calcul des valeurs propres de $(\mathcal{M}_{\mathcal{A}})$, que nous ne détaillerons pas ici. Pour des considérations pratiques, nous pouvons nous contenter d'observer que le nombre de mots aba -quasipériodiques croît exponentiellement, mais que la *proportion* de mots aba -quasipériodiques parmi tous les mots d'une longueur donnée tend très rapidement vers 0.

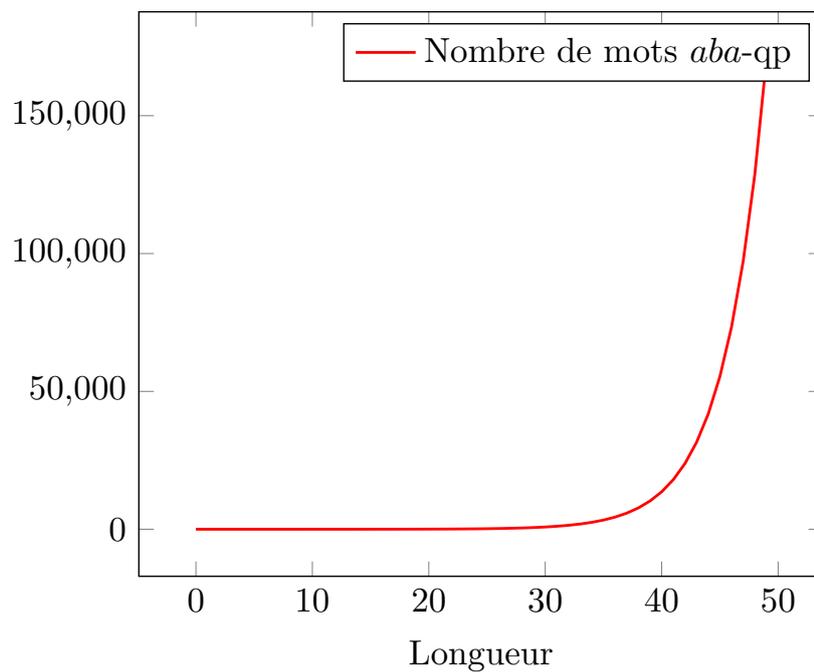
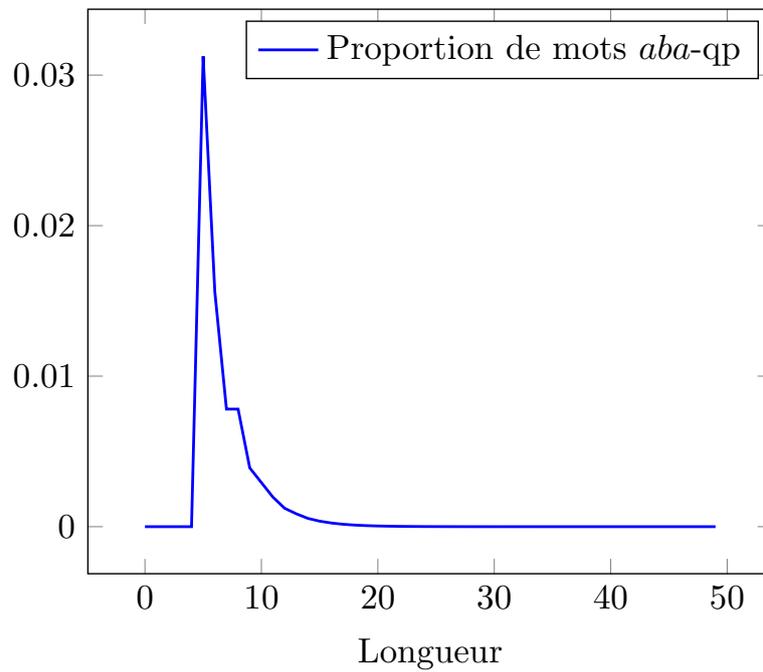


FIG. 1.4 : Nombre et proportion (parmi les mots binaires) de mots *aba*-quasipériodiques d'une longueur donnée

Words are merely tools and if you use the right ones you can actually put even your life in order, if you don't lie to yourself and use the wrong words.

— Hunter S. Thompson, dans une lettre à Larry Callen datée du 14 juillet 1958

Chapitre 2

Combinatoire des mots infinis

Entre 1906 et 1912, Axel Thue proposa d'étudier en toute généralité les chiffres d'un nombre réel arbitraire, écrit dans une base arbitraire également. En d'autres termes, il s'agit d'étudier les suites infinies de symboles choisis dans un alphabet fini, ce que nous appelons en langage moderne les *mots infinis*. Thue observa que certains *motifs* sont *inévitables*, c'est-à-dire qu'ils se retrouvent dans tous les mots infinis. D'autres, par opposition, sont *évitables* : il existe des mots infinis qui ne les comportent pas. Ses travaux furent les prémices à la résolution de nombreux problèmes, de la théorie des nombres [56] au jeu d'échecs [30] en passant par la géométrie [70]. De nos jours, l'étude des mots infinis fait partie de la *combinatoire des mots*, dont les ouvrages de référence sont les livres de Lothaire [63, 61, 62]. Cette discipline entretient des relations avec la bio-informatique, la théorie des langages formels et des automates, ainsi que la théorie des systèmes dynamiques et la théorie des nombres.

2.1 Introduction aux mots infinis

Étant entendu qu'un alphabet fini, non vide Σ est fixé, un *mot infini* est une fonction de \mathbb{N} , l'ensemble des entiers naturels, vers Σ . Ainsi,

$$\mathbf{w}_1 = aaaaaaaaaaaaaaaaaaaaaaaaaaaaaa \dots \quad (2.1)$$

$$\mathbf{w}_2 = abaabaabaabaabaabaabaaba \dots \quad (2.2)$$

$$\mathbf{w}_3 = bbaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa \dots \quad (2.3)$$

$$\mathbf{w}_4 = abaababaabaababaababaabaabaabaab \dots \quad (2.4)$$

$$\mathbf{w}_5 = ababbabbbabbbbabbbbabbbbba \dots \quad (2.5)$$

sont tous des (commencements de) mots infinis sur l'alphabet $\Sigma = \{a, b\}$. La lettre d'indice i d'un mot infini \mathbf{w} est naturellement notée $\mathbf{w}(i)$. Comme les mots infinis sont indexés par \mathbb{N} , l'indice de la première lettre est 0. Plus généralement, nous noterons $\mathbf{w}(i, \dots, i+n)$ le mot fini formé par la restriction de \mathbf{w} à l'intervalle entier $\{i, \dots, i+n\}$. Par exemple, si \mathbf{w} est le mot (2.3), alors $\mathbf{w}(1, \dots, 4) = bbaa$.

Dans cette section, nous décrivons les fondamentaux des mots infinis. La majorité des définitions et résultats exposés ici proviennent de l'article de Marston Morse et

Gustav Hedlund [69, 70]. Pour une référence complète, on consultera [63] ou les premiers chapitres de [1].

Contrairement aux mots finis, il n'est pas possible d'écrire explicitement un mot infini dans son intégralité¹. Toutefois, nous pouvons donner une formule — ou un algorithme — qui renvoie la $i^{\text{ème}}$ lettre du mot, étant donné un entier i . Ainsi, les quatre premiers exemples ci-dessus sont définis, dans l'ordre, par :

- $\mathbf{w}_1(i) = a$;
- $\mathbf{w}_2(i) = a$ si $i \equiv 0$ ou 2 modulo 3 , et b autrement ;
- $\mathbf{w}_3(i) = b$ si $i \leq 2$ et a autrement ;
- $\mathbf{w}_4(i) = a$ si $\lfloor \varphi(i+1) \rfloor - \lfloor \varphi i \rfloor = 2$, et b autrement ;

où $\varphi = \frac{1+\sqrt{5}}{2}$ est le nombre d'or et $\lfloor x \rfloor$ est la *partie entière* de x , c'est-à-dire le plus grand entier inférieur ou égal à x . Notons \mathbf{f} le mot \mathbf{w}_4 ; il est généralement appelé le mot de Fibonacci.

Pour décrire le mot \mathbf{w}_5 , il est plus commode de donner une suite de ses préfixes de plus en plus longs. Soit (u_n) la suite de mots définie par

$$\begin{aligned} u_0 &= \varepsilon \\ u_{n+1} &= u_n a b^n \end{aligned}$$

Il existe un unique mot infini dont tous les u_n sont des préfixes et ce mot infini est \mathbf{w}_5 . Voyons un autre exemple. La suite de mots $(s_n)_{n \in \mathbb{N}}$ spécifiée par

$$\begin{aligned} s_0 &= a \\ s_1 &= ab \\ s_{n+1} &= s_n s_{n-1} \end{aligned}$$

détermine uniquement un mot infini : il existe un unique mot dont s_i est un préfixe pour tout i . Il s'agit encore une fois de $\mathbf{w}_4 = \mathbf{f}$, le mot de Fibonacci. La preuve d'équivalence de ces deux définitions est donnée dans [61, chapitre 2].

Posons maintenant quelques définitions élémentaires ; ce sont des généralisations immédiates de leurs homologues sur les mots finis, une sensation de *déjà vu* pourrait donc se faire sentir. Un mot fini u est un *préfixe* de \mathbf{w} si

$$u = \mathbf{w}(0, 1, \dots, |u| - 1)$$

pour $|u| = n$ (rappelons que u_i désigne la lettre d'indice i dans u). De la même façon, un mot fini u est un *facteur* de \mathbf{w} s'il existe un entier k tel que :

$$u = \mathbf{w}(k, k+1, \dots, k+|u|-1)$$

Dans ce cas, l'entier k est la *position* de u et nous disons que u *couvre* les positions k à $k+|u|-1$. Un facteur u d'un mot infini \mathbf{w} est *spécial droit* s'il existe deux lettres

¹Comme aurait écrit Pierre de Fermat : « j'ai trouvé un mot infini véritablement merveilleux que cette marge est trop étroite pour contenir ».

distinctes a, b telles que ua et ub sont tous deux facteurs de \mathbf{w} . Symétriquement, u est *spécial gauche* si et seulement si au et bu sont facteurs de \mathbf{w} . Un facteur est *bispécial* lorsqu'il est à la fois spécial gauche et spécial droit. De plus, si aub est un facteur de \mathbf{w} pour un mot fini u et des lettres a, b , nous disons que aub est la *prolongation droite* de au et la *prolongation gauche* de ub . En d'autres termes, un facteur est spécial droit (resp. gauche) si et seulement s'il possède plus d'une prolongation droite (resp. gauche). Enfin, nous disons que ub est un *successeur* de au , qui inversement est un *prédécesseur* de ub . Dans un mot donné, un facteur peut avoir plusieurs successeurs ou prédécesseurs.

Ainsi, l'ensemble des préfixes du mot $\mathbf{w}_1 = aba\ aba\ aba\ \dots$ de l'équation (2.2) est donné par

$$(aba)^n \quad (aba)^n a \quad (aba)^n ab$$

pour n parcourant les entiers naturels, tandis que l'ensemble de ses facteurs, aussi appelé son *langage*, est donné par

$$\begin{array}{ccc} (aba)^n & (aba)^n a & (aba)^n ab \\ a(aba)^n & a(aba)^n a & a(aba)^n ab \\ ba(aba)^n & ba(aba)^n a & ba(aba)^n ab \end{array}$$

pour n parcourant les entiers naturels. Intuitivement, tout facteur de \mathbf{w}_1 peut commencer (resp. finir) à trois positions différentes au sein d'une occurrence de aba . Dans le mot de Fibonacci \mathbf{f} , représenté à l'équation (2.4), aba est spécial gauche et spécial droit à la fois, de même que ε et a . La lettre b , en revanche, n'est pas spéciale : elle est toujours précédée et suivie d'un a . Enfin, $baaba$ est spécial droit mais non spécial gauche, puisque b est toujours précédé de a . (Ces faits peuvent se démontrer par récurrence, en utilisant la définition en termes de la suite $(s_n)_n$.) Observons que tout suffixe d'un facteur spécial droit est lui-même spécial droit ; de la même façon, tout préfixe d'un facteur spécial gauche est lui-même spécial gauche. Ainsi, $aaba, aba, ba$ et a sont spéciaux droits dans (2.4).

Un mot infini \mathbf{u} est un *suffixe* de \mathbf{w} s'il existe un entier k tel que

$$\mathbf{u}(i) = \mathbf{w}(i + k)$$

pour tout entier i . Le suffixe est *propre* si cet entier k est strictement positif. Avec cette définition, nous pouvons dire qu'un mot infini est *périodique* s'il est un suffixe propre de lui-même, c'est-à-dire qu'il existe un entier (strictement positif) k vérifiant

$$\mathbf{w}(i) = \mathbf{w}(i + k)$$

pour tout entier i . Le mot $\mathbf{w}(0, \dots, k - 1)$ est appelé une *période* de \mathbf{w} , et *la* période est implicitement la plus courte. Nous confondrons parfois une période et sa longueur. Remarquons que, dans le contexte des mots infinis, les notions de *puissance* et de *périodicité* s'équivalent. Un mot infini possédant un suffixe périodique est appelé *ultimement périodique*.

Si u est un mot fini, nous notons u^ω le mot infini périodique de période u . Ainsi, le mot $\mathbf{w}_1 = aaaaaa\ \dots$ peut s'écrire a^ω et le mot $\mathbf{w}_2 = aba\ aba\ aba\ \dots$ s'écrit $(aba)^\omega$.

Le symbole ω fait ici référence au premier nombre ordinal transfini ; certains auteurs écrivent u^ω . En outre, si v est un mot fini et \mathbf{u} un mot infini, nous noterons $v \cdot \mathbf{u}$, ou parfois $v\mathbf{u}$, le mot infini \mathbf{w} qui vérifie

$$\begin{aligned} \mathbf{w}(k) &= v_k & \forall k \in \{0, \dots, |v| - 1\}, \\ \mathbf{w}(n + |v|) &= \mathbf{u}(n) & \forall n \in \mathbb{N}. \end{aligned}$$

En particulier, le mot ultimement périodique $\mathbf{w}_3 = bbb\ aaaaaa \dots$ s'écrit $bbb \cdot a^\omega$. Tout mot ultimement périodique s'écrit sous la forme uv^ω , pour u et v des mots finis.

Le *graphe de Rauzy* fournit une représentation visuelle intuitive de l'ensemble des facteurs d'une longueur donnée dans un mot, ainsi que des relations prédécesseur/successeur. Si \mathbf{w} est un mot infini, le graphe de Rauzy de \mathbf{w} à l'ordre n , noté $\mathcal{G}_{\mathbf{w}}(n)$, se définit comme suit. L'ensemble des sommets est l'ensemble des facteurs de longueur n dans \mathbf{w} . Il existe un arc de u vers v si et seulement si \mathbf{w} possède une occurrence de u à une certaine position k et une occurrence de v à la position $k + 1$. En d'autres termes, \mathbf{w} possède un facteur s'écrivant atb , pour a, b des lettres et t un mot fini, tel que $at = u$ et $tb = v$. En particulier, les arcs de $\mathcal{G}_{\mathbf{w}}(n)$ sont en bijection avec les facteurs de longueur $n + 1$ dans \mathbf{w} . Par convention, nous étiquetons l'arc entre u et v par la dernière lettre de v , c'est-à-dire b . Le graphe $\mathcal{G}_{\mathbf{f}}(3)$, pour \mathbf{f} le mot de Fibonacci (2.4) est donné par la figure 2.1.

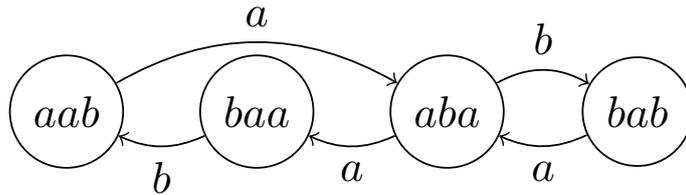


FIG. 2.1 : Graphe de Rauzy du mot de Fibonacci à l'ordre 3.

Considérons une suite $(u_n)_{n \in \mathbb{N}}$ de mots finis, dont les longueurs tendent vers l'infini et vérifiant la propriété *u_n est préfixe de u_{n+1} pour tout n* . Alors, pour tout entier naturel i , tous les u_n de longueur supérieure à i possèdent la même i -ème lettre, disons a_i . Comme les longueurs des u_n tendent vers l'infini, a_i est bien définie pour tout entier naturel i . Nous appelons alors le mot \mathbf{u} , défini par $\mathbf{u}(i) = a_i$, la *limite* de $(u_i)_{i \in \mathbb{N}}$. Nous disons également que la suite $(u_i)_{i \in \mathbb{N}}$ *converge* vers le mot \mathbf{u} . Il existe d'autres cas où une suite de mots finis converge vers un mot infini ; ils seront présentés en détail dans la section 4.1.

La donnée d'une suite calculable de préfixes finis est équivalente à la donnée d'un algorithme renvoyant la $i^{\text{ème}}$ lettre du mot. Toutefois, il est aisé de construire des exemples de suites de préfixes non calculables. Considérons par exemple la suite $(v_n)_{n \in \mathbb{N}}$ donnée par

$$\begin{aligned} v_0 &= \varepsilon \\ v_{n+1} &= \begin{cases} v_n a v_n & \text{si la } n\text{-ième machine de Turing s'arrête,} \\ v_n b v_n & \text{autrement,} \end{cases} \end{aligned}$$

pour une énumération standard des machines de Turing. Il est bien connu [80] qu'aucun algorithme ne permet de déterminer la i -ème lettre du mot \mathbf{v} défini comme la limite de (v_n) , mais cela ne freinera pas notre étude d'un tel mot. La détermination des premières lettres de \mathbf{v} est laissée en exercice au lecteur².

2.2 Morphismes

Voyons maintenant un outil utile pour la description et l'étude des mots infinis : les morphismes. Étant donnés deux alphabets Σ et T , une fonction μ de Σ^* vers T^* est un *morphisme* si, pour tous mots u et v de Σ^* , la relation

$$\mu(u \cdot v) = \mu(u) \cdot \mu(v) \quad (2.6)$$

est vérifiée. Nous disons que les morphismes sont les fonctions qui *commutent avec la concaténation*. Notons que cette propriété implique $\mu(\varepsilon) = \varepsilon$. De plus, un morphisme μ est uniquement déterminé par l'image des lettres. En effet, pour tout mot fini $u = u_1u_2 \dots u_n$, on évalue

$$\mu(u) = \mu(u_1)\mu(u_2) \dots \mu(u_n).$$

Ainsi, il existe un unique morphisme ϕ de $\{a, b\}^*$ vers $\{a, b\}^*$ qui vérifie

$$\begin{aligned} \phi(a) &= ab, \\ \phi(b) &= a. \end{aligned}$$

Nous l'appelons le *morphisme de Fibonacci*. Nous pouvons par exemple calculer :

$$\phi(ababaaba) = \phi(a)\phi(b)\phi(a)\phi(b)\phi(a)\phi(a)\phi(b)\phi(a) = abaabaababaab.$$

Observons que la composition de deux morphismes est toujours un morphisme. Il est donc aisé de produire de nouveaux morphismes à partir des anciens, en les composant entre eux ou en composant un morphisme avec lui-même. La composition étant associative, l'ensemble des morphismes sur un alphabet donné forme un monoïde non commutatif, avec l'identité dans le rôle de l'élément neutre.

Un morphisme est dit *uniforme* si les images des lettres sont toutes de la même longueur.

Tout morphisme s'étend de façon immédiate aux mots infinis. En effet, si \mathbf{w} est un mot infini et μ un morphisme, nous posons

$$\mu(\mathbf{w}) = \mu(\mathbf{w}(1)) \mu(\mathbf{w}(2)) \mu(\mathbf{w}(3)) \dots$$

dont la définition formelle (une limite de suite de mots finis) se déduit immédiatement de la notation. Nous possédons maintenant un moyen utile de créer un nouveau mot infini à partir d'un ancien : prendre l'image par un morphisme. Nous avons par exemple

$$\phi(bbb \cdot a^\omega) = aaa \cdot (ab)^\omega.$$

²L'auteur suggère que cet exercice constituerait un excellent système de détection de robots sur Internet (c.-à-d. de *captcha*).

Mais certains morphismes permettent également de construire un mot infini à partir d'une unique lettre. Considérons par exemple la suite $(F_n)_{n \in \mathbb{N}}$ spécifiée par

$$\begin{aligned} F_0 &= a \\ F_{n+1} &= \phi(F_n). \end{aligned}$$

Les premiers termes de la suite sont :

$$a \quad ab \quad aba \quad abaab \quad abaababa \quad abaababaabaab \quad abaababaabaababa \dots$$

Montrons par récurrence que, pour tout entier n , le mot F_n est un préfixe de F_{n+1} . Tout d'abord, nous vérifions directement que a est bien un préfixe de $F_1 = ab$. De plus, si F_n est un préfixe de F_{n+1} , alors ce dernier peut s'écrire

$$F_{n+1} = F_n u$$

pour un certain mot fini u non vide. En composant chaque membre par ϕ et en appliquant $F_{n+2} = \phi(F_{n+1})$, nous déduisons

$$F_{n+2} = \phi(F_n) \phi(u)$$

et $F_{n+1} = \phi(F_n)$ est bien un préfixe de F_{n+2} (et $\phi(u)$ est non vide). La suite F_n converge donc vers un mot infini, qui se trouve être de nouveau le mot de Fibonacci \mathbf{f} . Observons que la preuve précédente ne dépend pas du morphisme ϕ . Nous en déduisons le résultat suivant.

Proposition 2.1 (Folklore). *Soit τ un morphisme tel que $\tau(a)$ commence par a pour une certaine lettre a . Alors, la suite de mots $(t_n)_{n \in \mathbb{N}}$ définie par*

$$\begin{aligned} t_0 &= a \\ t_{n+1} &= \tau(t_n) \end{aligned}$$

possède la propriété que t_i est un préfixe de t_{i+1} pour tout entier i . Si $|t_n|$ tend vers l'infini, il existe une limite \mathbf{t} pour la suite (t_n) , que nous appelons le point fixe de τ en a car elle vérifie $\tau(\mathbf{t}) = \mathbf{t}$.

Dans ce qui suit, nous noterons τ^n l'itérée n fois du morphisme τ , c'est-à-dire que τ^0 est le morphisme identité et $\tau^{n+1} = \tau \circ \tau^n$. Cela contribuera à alléger les notations, en permettant par exemple d'écrire : « $\tau^n(a)$ est un préfixe de $\tau^{n+1}(a)$ pour tout entier n ». Par abus de notation, nous écrirons $\tau^\omega(a)$ pour la limite de la suite $(\tau^n(a))_{n \in \mathbb{N}}$. Le symbole ω fait toujours référence au nombre ordinal ω ; certains auteurs écrivent $\tau^\infty(a)$. Soulignons au passage que, pour un morphisme μ , les mots $\mu^\omega(a)$ et $\mu(a)^\omega$ sont en général distincts ; dans le cas de ϕ , les mots \mathbf{f} et $(ab)^\omega$ sont différents (à la position 4).

Les points fixes de morphismes constituent une classe de mots répandue et très étudiée. Lorsqu'un mot \mathbf{w} peut s'écrire comme le point fixe d'un morphisme, nous dirons qu'il est *purement morphique*.

Voyons un autre exemple. Le morphisme π défini par

$$\begin{aligned}\pi(a) &= ab \\ \pi(b) &= b\end{aligned}$$

admet un point fixe en a , dont on vérifie aisément qu'il s'agit de $a \cdot b^\omega$. En conséquence, tout mot ultimement périodique uv^ω peut s'écrire comme l'image par un morphisme du mot $\pi^\omega(a)$: si le morphisme η envoie a sur u et b sur v , alors

$$uv^\omega = \eta(\pi^\omega(a)).$$

Les mots pouvant s'écrire sous la forme $\mu(\nu^\omega(a))$, pour μ, ν des morphismes et a une lettre, sont appelés des *mots morphiques*. Les mots purement morphiques correspondent au cas où μ est le morphisme identité. Nous venons de montrer que tous les mots ultimement périodiques sont morphiques, mais il ne sont pas les seuls. Par exemple, le mot de Fibonacci n'est pas ultimement périodique et il est purement morphique, donc en particulier morphique.

Les mots morphiques sont une véritable usine à contre-exemples et à exemples. Nous les utiliserons intensivement au chapitre 3 dans notre étude des mots infinis quasipériodiques. De façon plus générale, le mot de Fibonacci est un exemple très commode en combinatoire des mots.

Un autre exemple célèbre est le mot de Thue-Morse, défini comme le point fixe du morphisme

$$\begin{aligned}a &\mapsto ab \\ b &\mapsto ba\end{aligned}$$

dont les premières lettres sont

$$abbabaabbaababbabaababba \dots$$

que nous recroiserons occasionnellement.

Les mots purement morphiques sont parfois appelés D0L et les mots morphiques des HD0L. En outre, les morphismes sont parfois appelés des *substitutions*. L'ouvrage [35] contient une étude en profondeur des substitutions, montrant ainsi la grande diversité de techniques et d'applications dont la théorie des morphismes est pourvue.

2.3 Complexité en facteurs

Nous connaissons à présent plusieurs manières de définir des mots infinis. Voyons comment les classer et les caractériser. Intuitivement, le mot infini

$$a^\omega = aaaaaaaaaa \dots$$

est le plus *simple*, plus *régulier* et plus *symétrique* possible. Comment quantifier cela ? Comment déterminer lequel d'entre deux mots est le plus régulier ? Nous disposons de deux outils pour cet usage : la *complexité en facteurs* et la *fonction de récurrence*.

Soit \mathbf{w} un mot infini. On note $\mathcal{L}(\mathbf{w})$ le langage de \mathbf{w} , c'est-à-dire, rappelons-le, l'ensemble de ses facteurs. On appelle *langage à l'ordre n* de \mathbf{w} , et on note $\mathcal{L}_n(\mathbf{w})$, l'ensemble des facteurs de longueur n de \mathbf{w} . La relation

$$\mathcal{L}(\mathbf{w}) = \bigcup_{n \in \mathbb{N}} \mathcal{L}_n(\mathbf{w})$$

est donc vérifiée. Tous les éléments du langage de \mathbf{w} sont des mots finis. Posons $P_{\mathbf{w}}$ la fonction définie par

$$P_{\mathbf{w}}(n) = \text{card}(\mathcal{L}_n(\mathbf{w})), \quad (2.7)$$

c'est-à-dire que $P_{\mathbf{w}}(n)$ est le nombre de facteurs de longueur n distincts apparaissant dans \mathbf{w} . La fonction $P_{\mathbf{w}}$ est appelée la *complexité en facteurs*, ou parfois simplement la *complexité* de \mathbf{w} . La lettre P provient de la terminologie historique *indice de permutation* [69], qui n'est plus employée de nos jours.

Si $\mathbf{u} = (aba)^\omega$, on observe que $P_{\mathbf{u}}(0) = 1, P_{\mathbf{u}}(1) = 2, P_{\mathbf{u}}(2) = 3, P_{\mathbf{u}}(3) = 3$ et plus généralement $P_{\mathbf{u}}(n) = 3$ pour tout $n \geq 2$. En effet, tout facteur assez grand de $(aba)^\omega$ est uniquement déterminé par la position à laquelle il commence à l'intérieur d'une occurrence de aba , et il n'y a que trois positions possibles.

Intuitivement, plus la croissance de la fonction $P_{\mathbf{w}}$ est rapide, plus le mot \mathbf{w} est considéré comme *complexe* ou *irrégulier*. Si \mathbf{w} est un mot sur k lettres, la valeur maximale de $P_{\mathbf{w}}$ est $P_{\mathbf{w}}(n) = k^n$, c'est-à-dire que \mathbf{w} contient tous les mots finis possibles. Un exemple de mot ayant une telle complexité est le mot de Champernowne, ici sur un alphabet binaire,

0 1 10 11 100 101 110 111 1000 1001 1010 1011 1100 1101 1110 1111 ...

qui s'obtient en concaténant le développement en base k de chaque entier naturel. Par opposition, le mot de plus faible complexité est le mot a^ω pour a une lettre. Il ne contient qu'un seul facteur de chaque longueur n , le mot a^n , et possède donc une complexité constante égale à 1.

Si un mot \mathbf{w} a une complexité en facteurs de la forme $P_{\mathbf{w}}(n) = \varepsilon^n$, nous disons que $\log_2 \varepsilon$ est l'*entropie topologique*, ou plus simplement l'*entropie*, de \mathbf{w} . Formellement, l'entropie est la quantité définie par

$$h(\mathbf{w}) = \lim_{n \rightarrow \infty} \frac{\log_2 P_{\mathbf{w}}(n)}{n}$$

Cette limite existe car la suite $(\log_2 P_{\mathbf{w}}(n))_n$ est sous-additive : $P_{\mathbf{w}}(m+n) \leq P_{\mathbf{w}}(m) \times P_{\mathbf{w}}(n)$ donc $\log_2 P_{\mathbf{w}}(m+n) \leq \log_2 P_{\mathbf{w}}(m) + \log_2 P_{\mathbf{w}}(n)$. Le lemme de Fekete [31] assure que toute suite de la forme $(u_n/n)_n$, pour u_n sous-additive, converge.

La terminologie *entropie* provient de la théorie de l'information et de la théorie des systèmes dynamiques ; il s'agit d'une notion bien plus générale, dont nous nous servirons uniquement pour évaluer la *dissymétrie* d'un mot. Par exemple, le mot de Champernowne a une entropie de 1.

Le théorème suivant est fondamental dans l'étude des mots infinis. Il montre en particulier que l'ultime périodicité s'exprime en termes de complexité.

Théorème 2.2 (Morse, Hedlund, 1938 [69]). *Soit \mathbf{w} un mot infini. Les affirmations suivantes sont équivalentes :*

1. *il existe un entier n tel que $P_{\mathbf{w}}(n) = P_{\mathbf{w}}(n+1)$;*
2. *le mot \mathbf{w} est ultimement périodique ;*
3. *la fonction $P_{\mathbf{w}}$ est bornée.*

Démonstration. Nous procédons par implications circulaires.

(1. \implies 2.) Supposons que $P_{\mathbf{w}}(n) = P_{\mathbf{w}}(n+1)$ pour un certain entier n . Chaque facteur de longueur n possède une unique prolongation droite, en d'autres termes aucun facteur de longueur n n'est spécial droit. Considérons maintenant un facteur de longueur n apparaissant aux positions i et j de \mathbf{w} , pour $i < j$. Alors

$$\mathbf{w}(i) = \mathbf{w}(j), \dots, \mathbf{w}(i+n-1) = \mathbf{w}(j+n-1),$$

et comme nul facteur de longueur n n'est spécial droit, les prochaines lettres sont également fixées, donc

$$\mathbf{w}(i+n) = \mathbf{w}(j+n), \mathbf{w}(i+n+1) = \mathbf{w}(j+n+1), \dots$$

et le mot $\mathbf{w}_i \mathbf{w}_{i+1} \dots$ est périodique, donc \mathbf{w} est bien ultimement périodique.

(2. \implies 3.) Si \mathbf{w} est ultimement périodique, il s'écrit nécessairement sous la forme $u \cdot v^\omega$, avec u et v des mots finis. Alors, le mot \mathbf{w} possède au plus $|u| + |v|$ facteurs d'une longueur donnée : chaque facteur commence dans u ou dans v .

(3. \implies 1.) Enfin, supposons que la fonction $P_{\mathbf{w}}$ soit bornée et montrons l'existence d'un entier n tel que $P_{\mathbf{w}}(n) = P_{\mathbf{w}}(n+1)$. Définissons la suite $(u_n)_{n \in \mathbb{N}}$ par $u_n = P_{\mathbf{w}}(n) - n$. Comme $P_{\mathbf{w}}(1) \geq 1$, nous avons $u_1 \geq 0$. De plus, $P_{\mathbf{w}}$ étant bornée, il existe un entier k tel que u_k soit strictement négatif. Enfin, tout facteur de longueur n ayant au moins une prolongation droite, nous avons

$$P_{\mathbf{w}}(n) \leq P_{\mathbf{w}}(n+1)$$

pour tout entier n . En soustrayant $n+1$ à chaque membre, nous obtenons

$$u_n - 1 \leq u_{n+1}$$

d'où nous concluons que $-1 \leq u_{n+1} - u_n$. Ainsi, la suite u_n ne peut décroître de plus de 1 à chaque rang (elle croît même parfois). Comme u_0 est positif ou nul et que u_k est strictement négatif, il existe nécessairement un entier n tel que $u_n = 0$. \square

Nous étudierons souvent la complexité en facteurs des mots que nous rencontrerons. Ceci nous permettra de nous assurer qu'ils ne sont pas ultimement périodiques, cas généralement considéré comme trop simple pour permettre une étude sérieuse, et de se faire une idée de la régularité du mot étudié.

Comme exemple de ce principe, le théorème suivant relie la notion de mot purement morphique avec la complexité en facteurs, montrant que les mots purement morphiques ne peuvent pas être *trop* complexes. Ceci montre que la propriété « être morphique » est une propriété de régularité et que nous pouvons considérer ces mots comme des cas simples, réguliers.

Théorème 2.3 (Pansiot, 1984 [73], voir aussi [12]). *Soit \mathbf{w} un mot purement morphique. Alors l'une des affirmations suivantes est vérifiée.*

$$\begin{aligned} P_{\mathbf{w}}(n) &\in \Theta(1), \\ P_{\mathbf{w}}(n) &\in \Theta(n), \\ P_{\mathbf{w}}(n) &\in \Theta(n \log \log n), \\ P_{\mathbf{w}}(n) &\in \Theta(n \log n), \\ P_{\mathbf{w}}(n) &\in \Theta(n^2). \end{aligned}$$

À titre de rappel, $f(n) \in \Theta(g(n))$ signifie qu'il existe deux réels α, β tels que $\alpha g(n) \leq f(n) \leq \beta g(n)$ dès que n est assez grand. Le théorème de Pansiot nous donne donc une information asymptotique sur la complexité en facteurs des mots purement morphiques, mais non une caractérisation. Il n'y a pas, à ce jour, de théorème similaire pour les mots morphiques non purement morphiques [12].

Les morphismes sont une construction importante que nous utiliserons souvent. Il nous est donc naturel de demander comment se comportent les différentes notions de régularité que nous rencontrons sous l'image par un morphisme.

Lemme 2.4 ([12]). *Soit \mathbf{w} un mot infini et μ un morphisme. La relation*

$$\frac{1}{\alpha} P_{\mathbf{w}}\left(\left\lfloor \frac{n}{\alpha} \right\rfloor\right) \leq P_{\mu(\mathbf{w})}(n) \leq P_{\mathbf{w}}(n)$$

est vérifiée avec α la quantité $\max_{a \in \Sigma} |\mu(a)|$.

2.4 Récurrence

La fonction de complexité ne fournit qu'une information partielle sur le mot \mathbf{w} étudié. Par exemple, un facteur n'apparaissant qu'une seule fois dans \mathbf{w} contribue de la même façon à la complexité qu'un facteur apparaissant une infinité de fois. Dans l'idée de donner davantage d'importance aux facteurs fréquents, nous posons les définitions suivantes. Un facteur u de \mathbf{w} est dit *récurrent* s'il apparaît une infinité de fois dans \mathbf{w} et *exceptionnel* autrement. Parmi les facteurs récurrents, certains peuvent apparaître plus fréquemment que d'autres. Nous disons qu'un facteur u d'un mot \mathbf{w} est *uniformément récurrent* si la distance entre deux occurrences consécutives de u dans \mathbf{w} est bornée. Les lettres se trouvant entre deux occurrences consécutives de u forment un *mot de retour* de u . Nous reviendrons sur les mots de retour lors de l'étude de la quasipériodicité sur les mots infinis.

Considérons la limite de la suite $(v_n)_{n \in \mathbb{N}}$ définie par

$$\begin{aligned} v_0 &= a, \\ v_{n+1} &= v_n \cdot b^n a, \end{aligned}$$

c'est-à-dire le mot

$$\mathbf{v} = ababbabbabbbabbbbabbbba \dots$$

Deux occurrences successives de b dans \mathbf{v} sont séparées par au plus deux positions, le facteur b est donc uniformément récurrent dans ce mot. Les occurrences successives de a sont de plus en plus espacées, mais il en existe tout de même une infinité, le facteur a est donc récurrent mais non uniformément. Enfin, il n'y a qu'une seule occurrence de aba dans \mathbf{v} , en préfixe, donc aba est exceptionnel, c'est-à-dire non récurrent dans \mathbf{v} .

Un mot dont tous les facteurs sont récurrents est appelé *récurrent*, et un mot dont tous les facteurs sont uniformément récurrents est appelé *uniformément récurrent*. Le mot \mathbf{v} défini ci-avant n'est donc pas récurrent (aba ne l'étant pas lui-même). Observons que ab^ω est un mot purement morphique mais non uniformément récurrent : le facteur a n'y apparaît qu'une seule fois. Ainsi, les mots morphiques ou purement morphiques ne sont pas nécessairement récurrents.

Voyons un exemple de mot uniformément récurrent : \mathbf{v} le mot de Thue-Morse, c'est-à-dire point fixe du morphisme ν défini par

$$\nu(a) = ab \quad \nu(b) = ba,$$

dont les premières lettres sont $abbabaabbaababbabaababba\dots$. Montrons par récurrence sur la longueur que tous ses facteurs sont uniformément récurrents. En premier lieu nous vérifions immédiatement que, dans l'image de n'importe quel mot par ν , les facteurs a et b sont uniformément récurrents : deux occurrences successives ne sont jamais séparées par plus de deux positions. Supposons maintenant que tous les facteurs de longueur strictement inférieure à n , pour un entier $n \geq 2$, sont uniformément récurrents. Soit w un facteur de longueur n de \mathbf{v} , et plus précisément $w = \mathbf{v}(i, \dots, i + n - 1)$ pour un entier naturel i . Considérons le facteur $w' = \mathbf{v}(i', \dots, i' + n')$, où

$$i' = \begin{cases} i & \text{si } i \text{ est pair,} \\ i - 1 & \text{autrement;} \end{cases} \quad n' = \begin{cases} n & \text{si } n \text{ est pair,} \\ n + 1 & \text{autrement;} \end{cases}$$

si bien que w' contient w , commence à une position paire de \mathbf{v} et possède une longueur paire. Rappelons que $\mathbf{v} = \nu(\mathbf{v})$, donc l'image inverse $x = \nu^{-1}(w')$ est bien définie et vérifie $|x| < n$. En particulier par hypothèse de récurrence, x est uniformément récurrent, et chaque occurrence de x induit une occurrence de w' , donc de w , dans \mathbf{v} . Donc w est également uniformément récurrent dans \mathbf{v} .

Certains mots sont récurrents, mais non uniformément. Par exemple, la limite \mathbf{u} de la suite $(u_n)_{n \in \mathbb{N}}$ définie par

$$\begin{aligned} u_0 &= a, \\ u_{n+1} &= u_n b^{2^n} u_n, \end{aligned}$$

dont les premières lettres sont $ababbababbbbababbaba\dots$ est récurrente : tout facteur de \mathbf{u} apparaît dans un u_i , et chaque u_i apparaît une infinité de fois. Toutefois, le mot \mathbf{u} contient des blocs arbitrairement longs de b , ce qui contredit l'*uniforme* récurrence, puisque deux occurrences consécutives de a (par exemple) peuvent se trouver arbitrairement éloignées.

Soit \mathbf{w} un mot uniformément récurrent et u un facteur de \mathbf{w} . Étant donné que u est uniformément récurrent, deux occurrences consécutives de u sont espacées d'au

plus k lettres, pour un certain entier k . Ceci implique que tout facteur de longueur $k + |u| - 1$ de \mathbf{w} contient une occurrence de u ; nous appelons la quantité minimale vérifiant cette propriété la *longueur englobante* de u (il s'agit d'une terminologie non standard introduite pour les besoins de cet exposé).

Soit n un entier naturel; nous pouvons calculer la longueur englobante de chaque facteur de longueur n , et considérer la longueur englobante maximale, que nous notons $R_{\mathbf{w}}(n)$. En d'autres termes, tout facteur de longueur $R_{\mathbf{w}}(n)$ de \mathbf{w} contient tous les facteurs de longueur n . Nous appelons $R_{\mathbf{w}}$ la *fonction de récurrence* de \mathbf{w} ; certains auteurs la nomment *fonction de quasipériodicité* [20], mais la « quasipériodicité » étudiée dans ces travaux n'est pas la notion qui fait l'objet de cette thèse. Contrairement à ce que son nom suggère, la fonction de récurrence n'est définie que pour les mots *uniformément* récurrents; la récurrence simple ne suffit pas (la fonction vaudrait $+\infty$ à partir d'un certain rang).

La récurrence et l'uniforme récurrence sont préservées par l'application d'un morphisme injectif.

2.5 Fréquences de facteurs

Tentons d'affiner encore nos notions de régularité. La complexité en facteurs présente le défaut que tout sous-mot, qu'il apparaisse une unique fois ou une infinité de fois, contribue de la même façon à la mesure de « complexité ». La fonction de récurrence rend l'analyse plus fine, mais elle ne s'applique qu'à certains mots particuliers (uniformément récurrents) et ne fait pas la distinction entre un facteur apparaissant une fois toutes les 5 lettres et un autre apparaissant une fois toutes les 5000 lettres. Intéressons-nous donc précisément à cela.

Étant donnés deux mots finis u et w , nous notons $|w|_u$ le nombre d'occurrences de u dans w , parfois appelé la « longueur de w en u ». La *fréquence* de u dans w est le rapport $\mathcal{F}_{\mathbf{w}}(u) = |w|_u/|w|$. Cette notion se généralise immédiatement aux mots infinis :

$$\mathcal{F}_{\mathbf{w}}(u) = \lim_{n \rightarrow \infty} \frac{|\mathbf{w}(0, 1, \dots, n-1)|_u}{n},$$

pour \mathbf{w} un mot infini. Cette limite n'existe pas toujours; le fait, pour un mot, d'avoir les fréquences de chaque facteur qui convergent est une propriété de régularité. Nous disons d'un tel mot qu'il *a des fréquences* pour ses facteurs, ou parfois qu'il est *uniquement ergodique*, en référence à une notion générale en théorie de systèmes dynamiques qui se traduit par le fait d'avoir des fréquences dans le cas des mots infinis [32].

Tout mot infini périodique v^ω a des fréquences pour tous ses facteurs, et il est très simple de les calculer. Si u est un facteur de v^ω et que $k = \lceil |u|/|v| \rceil$, nous avons

$$\mathcal{F}_{v^\omega}(u) = \frac{|v^\omega(0, 1, \dots, k|v| + |u| - 2)|_u}{k|v|}.$$

Intuitivement, pour calculer la fréquence de u dans v^ω , il suffit de calculer le nombre d'occurrences de u qui commencent dans les k premières périodes de v^ω . La démonstration est un simple calcul impliquant la définition de la périodicité. En appliquant ce résultat, nous obtenons ainsi $\mathcal{F}_{(aba)^\omega}(aa) = 1/3$.

Notre discussion pourrait laisser penser que la fréquence d'un facteur est toujours rationnelle, or ce n'est pas le cas. Pour nous en convaincre, tâchons de calculer la fréquence de la lettre a dans le mot de Fibonacci, c'est-à-dire la valeur $f_{\mathbf{f}}(a)$. Rappelons que $\mathbf{f} = \phi^\omega(a)$, avec $\phi(a) = ab$ et $\phi(b) = a$. En premier lieu, le mot \mathbf{f} s'écrivant sur un alphabet binaire, la fréquence de a donne immédiatement la fréquence de b par $f_{\mathbf{f}}(a) = 1 - f_{\mathbf{f}}(b)$. Il nous sera utile de calculer ces deux valeurs ensemble ; développons maintenant une machinerie appropriée à cette fin.

Soit w un mot fini sur un alphabet $\Sigma = \{a_1, a_2, \dots, a_k\}$ et μ un morphisme quelconque de Σ vers $\Sigma' = \{a_1, a_2, \dots, a_{k'}\}$. Nous définissons le *vecteur de Parikh* de w comme le vecteur dont la $i^{\text{ème}}$ coordonnée est le nombre d'occurrences de la lettre a_i dans w ; autrement dit, $\mathcal{C}(w) = (|w|_{a_1}, |w|_{a_2}, \dots, |w|_{a_k})$. Par exemple, $\mathcal{C}(abacaba) = (4, 2, 1)$. Nous définissons aussi la *matrice d'incidence* de μ par $\mathcal{M}(\mu)_{i,j} = |\mu(a_j)|_{a_i}$, c'est-à-dire que la j -ème colonne de cette matrice est le vecteur de Parikh de $\mu(a_j)$. Ainsi, $\mathcal{M}(\phi) = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$. La relation suivante

$$\mathcal{C}(\mu(w)) = \mathcal{M}(\mu) \times \mathcal{C}(w),$$

est valide pour tout mot fini w et tout morphisme μ .

Revenons au mot de Fibonacci. Le fait qu'il soit un mot morphique garantit que ses facteurs possèdent des fréquences [32]. Observons, par définition de la fréquence des facteurs et du mot de Fibonacci, que

$$f_{\mathbf{f}}(a) = \lim_{n \rightarrow \infty} f_{\phi^n(a)}(a).$$

En outre, si w est un mot fini, le normalisé du vecteur de Parikh $\mathcal{C}(w)/\|\mathcal{C}(w)\|_1$ possède pour i -ème composante la fréquence de a_i dans w . (La norme $\|\cdot\|_1$ est la somme des valeurs absolues des composantes du vecteur.) Nous cherchons donc à calculer

$$(f_{\mathbf{f}}(a), f_{\mathbf{f}}(b)) = \lim_{n \rightarrow \infty} \frac{\mathcal{M}(\mu)^n \times \mathcal{C}(a)}{\|\mathcal{M}(\mu)^n \times \mathcal{C}(a)\|_1}.$$

Il s'agit, par définition, d'un vecteur propre de $\mathcal{M}(\mu)$, dont toutes les composantes sont positives et dont la somme des composantes vaut 1. L'unique vecteur satisfaisant toutes ces conditions est $(1, \varphi - 1)/\varphi = (\varphi - 1, 2 - \varphi) \simeq (0, 618034; 0, 381966)$. Le nombre $\varphi = (1 + \sqrt{5})/2 \simeq 1, 618034$ n'est pas rationnel. Nous avons donc bien un facteur dont la fréquence est irrationnelle.

De la même façon, il existe des mots binaires dont la fréquence de a n'est pas nécessairement algébrique. Plus généralement, pour tout nombre réel x compris entre 0 et 1, il existe un mot binaire \mathbf{w} vérifiant $f_{\mathbf{w}}(a) = x$. Nous pouvons par exemple prendre le mot défini par

$$\mathbf{w}(i) = \lfloor (i + 1)x \rfloor - \lfloor ix \rfloor$$

en posant $a = 1, b = 0$. En effet :

$$f_{\mathbf{w}}(a) = \lim_{n \rightarrow \infty} \frac{\sum_{i=0}^{n-1} \mathbf{w}(i)}{n} = \lim_{n \rightarrow \infty} \frac{\sum_{i=0}^{n-1} \lfloor (i + 1)x \rfloor - \lfloor ix \rfloor}{n} = \lim_{n \rightarrow \infty} \frac{\lfloor nx \rfloor - \lfloor 0x \rfloor}{n} = x.$$

Nous pouvons donc aisément exhiber des mots dont certains facteurs ont des fréquences transcendantales, non calculables, etc. Nous présenterons les mots définis ainsi dans la section 4.1.

Enfin, comme nous l'avons mentionné, tous les mots n'ont pas des fréquences qui convergent pour leurs facteurs. Par exemple, le mot \mathbf{v} donné par

$$\begin{aligned}v_0 &= ab \\v_{n+1} &= v_n a^{3^{2n}} b^{3^{2n+1}}\end{aligned}$$

n'a pas de fréquence pour a . Cette dernière oscille entre $1/4$ et $3/4$ à mesure que l'on considère des préfixes de plus en plus grands, en se rapprochant arbitrairement près infiniment souvent de chacune de ses deux valeurs : elle ne converge pas mais possède deux valeurs d'adhérence.

Nous avons développé, au cours de cette section, de nombreuses *mesures* de régularité d'un mot : la croissance de sa complexité en facteurs, de sa fonction de récurrence éventuelle, convergence des fréquences de facteurs... Leurs définitions respectives découlent naturellement les unes des autres, de telle façon que nous pourrions croire qu'elles sont liées. Le lemme suivant énonce que les trois notions sont largement indépendantes : il est par exemple possible de produire des mots uniformément récurrents avec une complexité en facteurs exponentielle. Nous considérons cela comme une bonne nouvelle, car nos trois mesures de régularité sont complémentaires, et analysent des aspects différents d'un mot donné.

Lemme 2.5. *Soient x, y, z trois éléments de l'ensemble $\{\text{est, n'est pas}\}$. Alors il existe un mot qui x uniformément récurrent, qui y de complexité en facteurs polynomiale, et dont tout facteur a une fréquence qui z convergente.*

La fin de cette section est dédiée à la preuve du lemme 2.5. Nous donnons huit exemples de mots infinis afin de montrer que toutes les combinaisons des propriétés évoquées peuvent exister.

1. *Uniformément récurrent, complexité polynomiale, fréquences convergentes.* Les mots périodiques sont tous dans cette catégorie, ainsi que le mot de Fibonacci et le mot de Thue-Morse précédemment évoqués.
2. *Non-uniformément récurrent, complexité polynomiale, fréquences convergentes.* Il suffit de prendre n'importe quel mot \mathbf{w} de la catégorie précédente et de leur ajouter, au début, un facteur qui n'est pas dans le langage de \mathbf{w} . Par exemple, ba^ω satisfait à ces conditions.
3. *Uniformément récurrent, complexité exponentielle, fréquences convergentes.* Définissons la suite $(k_n)_{n \in \mathbb{N}}$ comme $k_0 = 3$ et $k_{n+1} = k_n!$. Le mot fini $B(n, i)$, pour n un entier et $1 \leq i \leq k_n$, est défini sur l'alphabet $\{1, 2, 3\}$ comme suit :

$$\begin{aligned}B(0, i) &= i \\B(n+1, i) &= B(n, \pi_i(1)) \dots B(n, \pi_i(k_n))\end{aligned}$$

avec π_i la i -ème permutation de $\{1, \dots, k_n\}$, en ordonnant par exemple les permutations lexicographiquement. Le mot infini $\mathbf{u} = B(0, 1)B(1, 1)B(2, 1)B(3, 1) \dots$ possède les propriétés voulues [32, proposition 7.4.3].

4. *Non-uniformément récurrent, complexité exponentielle, fréquences convergentes.* Il suffit de considérer le mot $4 \cdot \mathbf{u}$, où \mathbf{u} est le mot infini du point précédent et 4 une lettre n'apparaissant pas dans \mathbf{u} .
5. *Uniformément récurrent, complexité polynomiale, pas de fréquences convergentes.* Soit η le morphisme défini par $\eta(0) = 1$ et $\eta(1) = 0$, et $(k_n)_{n \in \mathbb{N}}$ la suite d'entiers définie au point 3. La suite de mots finis $(C_n)_{n \in \mathbb{N}}$, spécifiée par

$$\begin{aligned} C_0 &= 1 \\ C_{n+1} &= C_n \eta(C_n)^{k_n} \end{aligned}$$

converge vers un mot infini \mathbf{c} sur l'alphabet $\{0, 1\}$ qui possède les propriétés voulues [32, section 7.5.3.4].

6. *Non-uniformément récurrent, complexité polynomiale, pas de fréquences convergentes.* Il suffit de considérer le mot $2 \cdot \mathbf{c}$, où 2 est une lettre n'apparaissant pas dans \mathbf{c} , pour obtenir les propriétés voulues.
7. *Uniformément récurrent, complexité exponentielle, pas de fréquences convergentes.* La construction de ce mot étant relativement lourde, nous référons le lecteur à [32, début de la section 7.5.3.3] et ses références.
8. *Non-uniformément récurrent, complexité exponentielle, pas de fréquences convergentes.* Considérons la suite de tous les mots finis sur l'alphabet $\{0, 1\}$, ordonnée par longueur puis par ordre lexicographique pour une longueur donnée. La concaténation de tous les mots finis, dans cet ordre, donne un mot infini possédant tous les facteurs possibles sur $\{0, 1\}$. Ce mot possède les propriétés voulues. Notons toutefois qu'il est récurrent, bien que non-uniformément.

2.6 Quasipériodicité sur les mots infinis

À partir de 2002, Solomon Marcus généralise la notion de quasipériodicité aux mots infinis [66, 67]. Soit q un mot fini et \mathbf{w} un mot infini ; alors q est une *quasipériode* de \mathbf{w} si et seulement si toute position de \mathbf{w} est couverte par une occurrence de q . De façon équivalente, q est un préfixe de \mathbf{w} et $R_{\mathbf{w}}(q) < |q|$. Marcus pose de nombreux problèmes au sujet des mots infinis quasipériodiques. Des travaux ultérieurs par Amy Glen, Florence Levé et Gwenaél Richomme répondent [58, 59, 41] à ces questions. Ils étudient également les liens entre morphismes et mots infinis quasipériodiques, en introduisant la notion de *morphisme (fortement) quasipériodique*. Un morphisme μ est *fortement quasipériodique* lorsque l'image de tout mot infini par μ est quasipériodique. C'est notamment le cas du morphisme ϕ^2 défini par

$$\phi^2(a) = aba, \quad \phi^2(b) = ab,$$

autrement dit la composée du morphisme ϕ de Fibonacci avec lui-même. Étant donné que les images de a et de b commencent toutes deux par a , l'image de tout mot de longueur au moins 2 par ϕ^2 commencera par aba . De façon plus générale, dans le mot $\phi^2(\mathbf{w}_1\mathbf{w}_2\dots) = \phi^2(\mathbf{w}_1)\phi^2(\mathbf{w}_2)\dots$, l'image de chaque \mathbf{w}_i (pour i un entier naturel) est le préfixe d'une occurrence de aba . Étant donné que les images des lettres par ϕ^2 sont de longueurs plus courtes ou égales à aba , nous en concluons que le mot est aba -quasipériodique.

Il est possible de démontrer un résultat similaire au théorème de Mouchard sur les mots infinis. En particulier, un mot infini est aba -quasipériodique si et seulement s'il est l'image d'un mot infini par ϕ^2 . Par exemple, le mot de Fibonacci \mathbf{f} est aba -quasipériodique. Solomon Marcus et Thierry Monteil, en 2006, généralisent ce fait [68].

Théorème 2.6 (Marcus, Monteil, 2006 [68]). *Soit q un mot fini. Un mot infini \mathbf{w} est q -quasipériodique si et seulement s'il s'écrit*

$$\mathbf{w} = qu_1u_2u_3\dots$$

où les u_i sont tous des antibords droits de q . De façon équivalente, il s'écrit

$$\mathbf{w} = v_1v_2v_3\dots$$

où les v_i sont tous des antibords gauches de q .

Démonstration. Appelons b_1 le bord associé à u_1 dans q ; alors il existe un mot u'_1 tel que $q = b_1u_1 = u'_1b_1$. Donc $qu_1 = u'_1b_1u_1 = u'_1q$. En itérant cet argument, nous montrons que chaque antibord dans la décomposition est le suffixe d'une occurrence de q . Comme les antibords de q sont moins longs que lui-même, le mot q couvre bien toutes les positions dans \mathbf{w} . Un argument symétrique montre que cette propriété reste valide pour les antibords gauches.

Réciproquement, supposons que \mathbf{w} est un mot quasipériodique. Appelons $(q_n)_{n \in \mathbb{N}}$ la suite des positions des occurrences de q (en particulier, $q_0 = 0$). Par quasipériodicité, $q_{n+1} - q_n \leq |q|$ pour tout entier n . Considérons le factor $u_n = \mathbf{w}(q_n, \dots, q_{n+1} + |q| - 1)$. Nous avons des mots finis p_n, s_n tels que $u_n = qs_n = p_nq$. Observons que $\mathbf{w} = p_0p_1p_2\dots = qs_0s_1s_2\dots$. Comme $|p_n| = |s_n| \leq q_{n+1} - q_n \leq |q|$, les mots p_n et s_n sont respectivement les antibords gauches et droits de q associés au bord $b_n = \mathbf{w}(q_{n+1}, \dots, q_n + |q| - 1)$. \square

Soit q un mot fini sur un alphabet Σ et b_1, b_2, \dots, b_k ses antibords gauches, rangés par longueurs croissantes. Définissons un morphisme δ_q , de $\{1, \dots, k\}^*$ vers Σ^* par $\delta_q(i) = b_i$; ce morphisme δ_q est appelé le *morphisme dérivateur gauche* de q (une définition symétrique, avec les antibords droits, donne le *morphisme dérivateur droit*). Voici par exemple les morphismes dérivateurs de aba et $ababa$:

$$\begin{array}{ll} \delta_{aba} : \{1, 2\} \rightarrow \{a, b\}^* & \delta_{ababa} : \{1, 2, 3\} \rightarrow \{a, b\}^* \\ 1 \mapsto ab & 1 \mapsto ab \\ 2 \mapsto aba & 2 \mapsto abab \\ & 3 \mapsto ababa \end{array}$$

Le théorème 2.6 affirme qu'un mot infini \mathbf{w} est q -quasipériodique si et seulement si \mathbf{w} est l'image d'un mot infini par δ_q . En choisissant un mot q superprimitif (par exemple aba), le morphisme δ_q est injectif. En effet, s'il ne l'était pas, il existerait deux suites d'antibords de q $v_0v_1v_2\dots$ et $v'_0v'_1v'_2\dots$ différentes, c'est-à-dire que $v_i \neq v'_i$ pour un certain entier i , mais telles que $v_0 \cdot v_1 \cdot v_2 \cdot \dots = v'_0 \cdot v'_1 \cdot v'_2 \cdot \dots$ (en concaténant). Sans perte de généralité, supposons que $|v_i| < |v'_i|$; alors l'occurrence de q dont v'_i est le suffixe appartient au chevauchement de deux occurrences de q , celles dont v_i et v_{i+1} sont les suffixes. Les mêmes arguments que ceux employés dans la deuxième partie de la preuve du théorème 1.6 de Mouchard montrent ensuite que q est quasipériodique.

Nous pouvons donc considérer n'importe quel mot « désordonné », « asymétrique » et prendre son image par δ_q ; ceci donnera un mot q -quasipériodique tout aussi désordonné et asymétrique. La proposition suivante est une conséquence immédiate du lemme 2.5 et des remarques que nous venons d'énoncer.

Proposition 2.7. *Soit q un mot fini superprimitif. Alors il existe un mot q -quasipériodique qui n'est pas uniformément récurrent, dont les fréquences de certains facteurs ne convergent pas, et possédant une complexité en facteurs exponentielle.*

Nous voyons donc que la quasipériodicité n'est pas une bonne notion de symétrie. Il s'agit en fait d'un simple recodage sur un alphabet différent; moralement, les mots aba -quasipériodiques sont exactement les mots pouvant s'écrire sur l'alphabet $\{aba, ab\}$. La quasipériodicité est une propriété extrêmement locale. Par opposition, les propriétés de régularité dont nous disposons sur les mots infinis sont globales et donnent de l'information sur l'ensemble du mot considéré. Pour pallier ce problème, Marcus et Monteil [68] ont introduit la quasipériodicité multi-échelles. Un mot est *quasipériodique multi-échelles* si et seulement s'il admet une infinité de quasipériodes. Étant donné que toutes les quasipériodes sont des préfixes, ceci implique que les quasipériodes sont de plus en plus grandes. Ils démontrent le résultat suivant.

Théorème 2.8 (Marcus, Monteil, 2006 [68]). *Soit \mathbf{w} un mot infini quasipériodique multi-échelles. Alors \mathbf{w} est uniformément récurrent, possède une entropie topologique nulle et les fréquences de tous ses facteurs convergent.*

Démonstration. Soit \mathbf{w} un mot quasipériodique multi-échelles et $(q_n)_{n \in \mathbb{N}}$ la suite de ses quasipériodes, ordonnée par longueurs croissantes.

1. Le mot \mathbf{w} est uniformément récurrent. Les (q_n) sont des préfixes de \mathbf{w} , donc tout facteur de \mathbf{w} apparaît dans un q_n , et donc à intervalles bornés.
2. Le mot \mathbf{w} possède une entropie nulle. En effet, le nombre de facteurs de longueur $|q_n|$ est borné par $|q_n|^2$: tout facteur de longueur $|q_n|$ commence dans une occurrence de q_n (il y a donc $|q_n|$ possibilités) et dépend du chevauchement entre cette occurrence et la suivante (il y a au plus $|q_n|$ chevauchements possibles).
3. Le mot \mathbf{w} a les fréquences de tous ses facteurs qui convergent. La preuve de ce résultat est assez technique, et une version généralisée est présentée au chapitre 6 (théorème 7.4). Nous ne démontrons donc pas cette propriété pour l'instant. Le lecteur intéressé pourra se référer à [68].

□

Les mêmes auteurs montrent toutefois que la complexité en facteurs peut prendre des valeurs arbitrairement élevées entre deux longueurs de quasipériodes. Il y a donc, dans une certaine mesure, une liberté dans les mots quasipériodiques multi-échelles. Nous quantifierons plus précisément l'amplitude de cette liberté au chapitre 3, en montrant des mots infinis dont les ensembles de quasipériodes ont des structures diverses ; voir proposition 3.4 et théorèmes 3.6 et 3.11.

To define is to limit.

— Oscar Wilde, *Le portrait de Dorian Gray*, 1890

Chapitre 3

Une méthode d'étude des quasipériodes

Ce chapitre présente une méthode permettant, étant donné un mot infini, de déterminer l'ensemble de ses quasipériodes. Cette méthode n'est pas à proprement parler un algorithme, puisque l'entrée est en général infinie ; il s'agit plutôt d'un plan d'attaque, d'une protopreuve, dont la formalisation prend la forme d'une collection de propriétés utiles. Nous utiliserons ces dernières pour construire divers exemples extrémaux de mots quasipériodiques. En effet, il n'est pas possible, pour tout ensemble de mots finis Q , de trouver un mot infini \mathbf{w} dont l'ensemble de quasipériodes soit exactement Q . Nous explorerons donc diverses possibilités intéressantes pour Q , sans toutefois aller jusqu'à une caractérisation complète.

3.1 Détermination des quasipériodes d'un mot infini arbitraire

Commençons par une observation simple : si \mathbf{w} est un mot infini et q l'une de ses quasipériodes, alors q est nécessairement un préfixe de \mathbf{w} , puisqu'elle doit couvrir la première lettre. Appelons p_n le préfixe de longueur n de \mathbf{w} ; alors il est naturel de se demander, sachant que p_n est une quasipériode pour un certain entier n , si p_{n+1} et p_{n-1} sont à leur tour des quasipériodes. Nous donnons des conditions nécessaires et suffisantes pour que ce soit le cas.

Soit q un mot fini. Un *chevauchement* de q est un mot fini, q -quasipériodique, comprenant exactement deux occurrences de q . Si w est un chevauchement de q , la quantité $2|q| - |w|$ est appelée *l'étendue* du chevauchement ; il s'agit du nombre de lettres partagées par les deux occurrences de q . La quantité $|w| - |q|$ est appelée la *longueur de retour* de q dans w . Ces deux définitions sont illustrées par la figure 3.1. La terminologie « longueur de retour » est en fait générale, elle provient de l'étude des mots de retour ; le lecteur souhaitant approfondir ce point pourra consulter [27, 28]. Nous aurons besoin d'un lemme technique sur les chevauchements pour nos caractérisations.

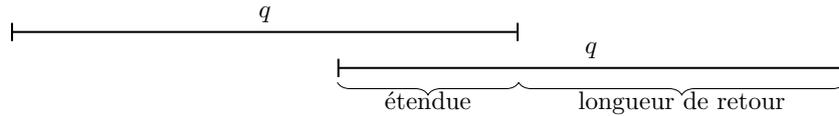


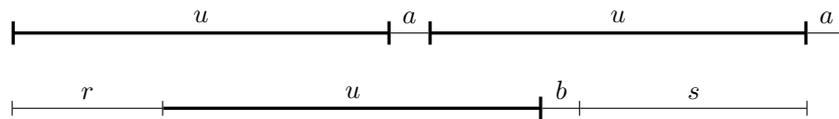
FIG. 3.1 : Illustration de l'étendue et du chevauchement

Lemme 3.1. *Soit u un mot fini et a, b deux lettres. Si w est un chevauchement de ua et si ub est un facteur de w , alors $a = b$.*

Démonstration. Le résultat est immédiat si ub est un préfixe ou un suffixe de w . Distinguons deux cas (figure 3.2). Si $w = uaua$, puisque ub est un facteur interne de w , il existe un mot non vide r et un mot s tels que $uaua = rubsa$. Observons que $u = rs$ (car $|u| = |rs|$ avec r un préfixe et s un suffixe de u) et que $ub = sar$ (car $rubsa = uaua = rsarsa$); en comptant les lettres, nous déduisons que $a = b$.

Si $|w| < 2|ua|$, alors il existe des mots r, s tels que $u = rs$ et $w = rsasa$. Puisque ub est un facteur interne de w , le mot sb est un facteur interne de $sasa$, et nous sommes réduits au cas précédent. \square

Premier cas.



Deuxième cas.

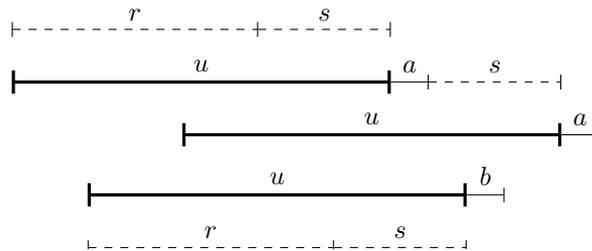


FIG. 3.2 : Illustration de la preuve du lemme 3.1

L'auteur remercie Patrice Séebold pour la preuve du lemme 3.1, qui a permis de simplifier la démonstration du théorème suivant.

Théorème 3.2. *Soit \mathbf{w} un mot infini, q un facteur de \mathbf{w} et a, b des lettres telles que qa est une quasipériode de \mathbf{w} et qab un facteur de \mathbf{w} .*

1. Le mot qab n'est **pas** une quasipériode si et seulement si qa est spécial droit.
2. Le mot q n'est **pas** une quasipériode si et seulement si $qaqa$ est facteur de \mathbf{w} et q n'est pas un facteur interne de qaq .

3.2. EXEMPLE D'APPLICATION : ÉTUDE DES QUASIPÉRIODES D'UN MOT 49

Démonstration. Les deux affirmations se prouvent séparément.

Affirmation 1. Si qa n'est pas spécial droit, toute occurrence de qa dans \mathbf{w} se prolonge en une occurrence de qab . À plus forte raison, le mot qab recouvre également w . Réciproquement, supposons que qab soit une quasipériode de \mathbf{w} et qu'il existe une lettre c , différente de b , telle que qac est un facteur de \mathbf{w} . Alors il existe une occurrence de qac facteur d'un chevauchement de qab , et le lemme 3.1 donne $b = c$, une contradiction.

Affirmation 2. Si $qaqa$ est facteur de \mathbf{w} et q n'est pas un facteur interne de qaq , alors la lettre a au milieu de qaq n'est pas couverte par q , qui n'est donc pas une quasipériode. Si q est facteur interne de qaq , alors les occurrences de qaq sont bien recouvertes par des occurrences de q qui est donc une quasipériode. Si $qaqa$ n'est pas facteur de \mathbf{w} , alors toutes les occurrences de qa partagent au moins une lettre et q recouvre \mathbf{w} . \square

Nous pouvons reformuler ce théorème de la façon suivante, qui donne la *protopreuve* annoncée pour l'étude de l'ensemble des quasipériodes d'un mot donné.

Corollaire 3.3. *Soit \mathbf{w} un mot infini et soit p_i son préfixe de longueur i . L'ensemble des longueurs des quasipériodes de \mathbf{w} est une union d'intervalles d'entiers disjoints. L'intervalle $[i, j]$ est un intervalle de longueurs de quasipériodes tel qu'il n'y a pas de quasipériodes de longueur $i - 1$ ni $j + 1$ (lorsque $j < \infty$) si, et seulement si :*

- $p_i p_i$ est un facteur de \mathbf{w} ;
- p_{i-1} n'est pas un facteur interne de $p_i p_{i-1}$;
- p_j est un facteur spécial droit de \mathbf{w} (si $j < \infty$) ; et
- pour tout entier k tel que $i \leq k < j$, le mot p_k n'est pas un facteur spécial droit de \mathbf{w} .

Ce corollaire est à la base des preuves des autres résultats de ce chapitre.

3.2 Exemple d'application : étude des quasipériodes d'un mot

Voyons comment ce théorème nous permet d'étudier l'ensemble des quasipériodes d'un mot donné.

Proposition 3.4. *Soit h le morphisme défini par $h(a) = abaaba$ et $h(b) = bababa$. L'ensemble des quasipériodes de $h^\omega(a)$ est $\{h^n(aba) \mid n \in \mathbb{N}\}$.*

La démonstration de cette proposition figure ci-après.

L'intérêt de cet exemple réside en ce qu'il n'admet *aucune quasipériode en dehors des* $h^n(aba)$. Remarquons que, pour tout entier naturel n , le mot $h^{n+1}(aba)$ admet $h^n(aba)$ pour quasipériode ; $h^\omega(a)$ ne possède donc qu'une seule quasipériode superprimitive. Nous commenterons ce fait plus en avant dans la section 3.3.

Pour référence, voici les 120 premières lettres de $h^\omega(a)$.

abaabababaabaabaabaababababaabaababababaabaababababaabaaba
bababaabaabaabaababababaabaabaabaababababaabaabaabaababababa...

La démonstration de la proposition 3.4 fait appel à une nouvelle notion. Soient μ un morphisme et v un mot fini. Nous disons que v est *synchronisant* pour μ lorsque, pour tout mot fini ou infini w et pour tout préfixe de $\mu(w)$ pouvant s'écrire uv , il existe un unique mot u' tel que $u = \mu(u')$. Les occurrences de $\mu(v)$ nous permettent donc de connaître notre position dans la préimage de manière unique.

Afin de vérifier qu'un mot v est synchronisant pour un morphisme injectif uniforme μ , il suffit d'énumérer tous les facteurs de longueur $2|v|$ des mots dans l'image de μ , puis de vérifier exhaustivement que v est préfixe (et seulement préfixe) de ces facteurs.

Démonstration de la proposition 3.4. Nous allons utiliser le corollaire 3.3 pour prouver ce résultat. Nous avons besoin de montrer divers faits intermédiaires à cette fin. Posons $\mathbf{w} = h^\omega(a)$.

Fait 1. *Les préfixes p tels que pp est un facteur de \mathbf{w} sont $\varepsilon, h^n(a), h^n(ab)$ et $h^n(aba)$, pour n parcourant les entiers naturels.*

Considérons p un préfixe arbitraire tel que pp soit facteur de \mathbf{w} et montrons par récurrence sur $|p|$ que p a l'une des formes voulues. Dans le cas $|p| \leq 6$, une vérification exhaustive montre que p est ε, a, ab, aba ou $h(a)$.

Supposons maintenant que $|p| > 6$; en particulier, le mot p possède le préfixe $abaabab$, qui est un mot synchronisant pour h (ceci peut être vérifié directement car h est uniforme et injectif). Étant donné que pp est un facteur de \mathbf{w} , le mot πpp est un préfixe de \mathbf{w} pour un certain mot fini π . Il existe donc des mots π' et p' tels que $h(\pi'p'p')$ soit un préfixe de \mathbf{w} , avec $p = h(p')$ et $\pi = h(\pi')$. Nous déduisons de l'injectivité de h que $\pi'p'p'$ est un préfixe de \mathbf{w} , qui vérifie $|\pi'p'p'| < |\pi pp|$. Par hypothèse de récurrence, $\pi'p'p'$ a l'une des formes voulues, donc $\pi pp = h(\pi'p'p')$ également.

Fait 2. *Les mots $h^n(a)$ et $h^n(ab)$ ne sont pas des quasipériodes de \mathbf{w} .*

Nous montrons par récurrence sur n que $h^n(a)$ n'est pas une quasipériode de \mathbf{w} . Une vérification directe montre que a et $h(a)$ ne sont pas des quasipériodes de \mathbf{w} . Supposons que $h^n(a)$ ne soit pas quasipériode et montrons que $h^{n+1}(a)$ ne l'est pas non plus. Nous raisonnons par contraposée : supposons donc que $h^{n+1}(a)$ soit une quasipériode et montrons que $h^n(a)$ en est également une. Appelons $(\pi_k)_{k \in \mathbb{N}}$ la suite des mots tels que $\pi_k h^{n+1}(a)$ est un préfixe de \mathbf{w} , ordonnés par longueur. Si $h^{n+1}(a)$ est une quasipériode, nous avons $|\pi_{k+1}| - |\pi_k| \leq |h^{n+1}(a)|$.

Le mot $abaabab$, synchronisant pour h , est un préfixe de $h^{n+1}(a)$; il existe donc une suite de mots $(\pi'_k)_{k \in \mathbb{N}}$ tels que $\pi'_k = h(\pi_k)$ pour tout entier naturel k . Observons que $\pi'_k h^n(a)$ est une suite de préfixes de \mathbf{w} qui vérifie $|\pi'_{k+1}| - |\pi'_k| \leq |h^n(a)|$. Ceci montre bien que $h^n(a)$ est une quasipériode de \mathbf{w} .

Un raisonnement en tout point similaire montre que $h^n(ab)$ n'est quasipériode de \mathbf{w} pour aucun entier naturel n .

Fait 3. *Les mots $h^n(aba)$ sont des quasipériodes de \mathbf{w} .*

En effet, l'image par h de tout mot commençant par a admet aba pour quasipériode. Le point fixe de h admet donc $h^n(aba)$ comme quasipériode pour tout entier naturel n .

Fait 4. *Les mots $h^n(aba)$ sont des facteurs spéciaux droits de \mathbf{w} .*

Nous vérifions ce fait par récurrence sur n . Une vérification directe montre que aba est un facteur spécial droit de \mathbf{w} . Étant donné que $h(a)$ et $h(b)$ ne commencent pas par la même lettre, si $h^n(aba)$ est spécial droit, alors $h^{n+1}(aba)$ l'est également pour tout entier naturel n .

Conclusion. Les mots $h^n(aba)$ sont les seuls préfixes de \mathbf{w} dont le carré apparaît dans \mathbf{w} qui sont également des quasipériodes. Ils sont tous spéciaux droits, donc le corollaire 3.3 permet de conclure que ce sont les seules quasipériodes de \mathbf{w} . \square

Ce schéma de preuve peut se généraliser aisément à d'autres cas. Étant donné un mot \mathbf{w} dont on connaît la liste des préfixes spéciaux droits et la liste des préfixes qui sont également des facteurs carrés, il s'agit d'étudier lesquels sont des quasipériodes. Les préfixes, ainsi que les facteurs spéciaux droits et carrés, font partie des objets classiquement étudiés en combinatoire des mots. Il est donc aisé d'étudier les quasipériodes de nombreux mots classiques du domaine.

Dans le cas particulier d'un mot morphique (ou comme dans notre exemple purement morphique) $\mathbf{w} = g \circ h^\omega(a)$, il est possible de « deviner » les petits préfixes spéciaux droits et dont les carrés sont facteurs en étudiant informatiquement un long préfixe de \mathbf{w} (qui est calculable). On conjecture alors que les préfixes que nous cherchons sont les images par $g \circ h^n$ de ceux-là, ce que l'on valide avec un raisonnement par récurrence. Il faut ensuite étudier lesquels de ces préfixes sont des quasipériodes : il est généralement possible d'exclure les non quasipériodes par une vérification directe ; pour les autres, un raisonnement par récurrence permet de conclure. Dans les deux raisonnements par récurrence, la difficulté se résume souvent à trouver un mot synchronisant permettant de conclure.

Cette méthode ne constitue nullement un algorithme ni un théorème, mais plutôt une *recette* à adapter à la situation.

3.3 Sur les quasipériodes gigognes

Les exemples les plus naturels de mots quasipériodiques multi-échelles sont les points fixes des morphismes dont les images des lettres sont les antibords d'un mot, comme par exemple (ici dans le cas de aba) :

$$\begin{aligned} \varphi : \{a, b\}^* &\rightarrow \{a, b\}^* \\ a &\mapsto aba \\ b &\mapsto ba \end{aligned}$$

Il est aisé de voir que $\varphi^\omega(a)$ possède la quasipériode aba , mais également $\varphi(aba)$, $\varphi^2(aba)$ et ainsi de suite. En particulier, si on appelle q_n le mot $\varphi^n(aba)$, alors q_{n+1} est q_n -quasipériodique (en tant que mot fini). Nous disons que $(q_n)_{n \in \mathbb{N}}$ est une suite de *quasipériodes gigognes*. Dans un mot fini, la suite des quasipériodes (par longueurs croissantes) est toujours une suite de quasipériodes gigognes, comme l'énonce la remarque 1.5. Cette propriété s'étend-elle aux mots multi-échelles ? Si non, l'ensemble

Lemme 3.8. *Pour tout entier naturel n , le mot $f^n(q)$ est superprimitif.*

Démonstration. Une vérification directe montre que le mot q est superprimitif. Supposons que $f^n(q)$ soit superprimitif et montrons que $f^{n+1}(q)$ l'est également. Nous raisonnons par contraposée : supposons que $f^{n+1}(q)$ soit quasipériodique et montrons que $f^n(q)$ l'est également.

Soit r une quasipériode de $f^{n+1}(q)$. Alors il existe un entier k et une suite de k couples de mots $(p_i, s_i)_{1 \leq i \leq k}$ tels que $p_i r s_i = f^{n+1}(q)$ pour tout i . Ordonnons-les par longueur des p_i . Nous établissons maintenant qu'il existe des mots r' et $(p'_i, s'_i)_{1 \leq i \leq k}$ tels que $p_i = f(p'_i)$, $r = f(r')$ et $s_i = f(s'_i)$. Observons que $f(a)ba$ est nécessairement un préfixe de $f^{n+1}(q)$. Une recherche exhaustive montre qu'aucun préfixe de $f(a)ba$ n'est une quasipériode de $f^{n+1}(q)$, donc $f(a)ba$ est un préfixe de r . En outre, le mot $f(a)ba$ est synchronisant pour f . Ceci implique l'existence de mots $(p'_i)_{1 \leq i \leq k}$ tels que $p_i = f(p'_i)$ pour tout i . Considérons maintenant $f^{n+1}(q) = p_k r s_k$ la dernière occurrence de r dans le mot ; s_k est donc vide (r doit couvrir la dernière lettre) et $f^{n+1}(q) = p_k r = f(p'_k) r$, d'où l'on déduit l'existence d'un r' tel que $r = f(r')$. Un raisonnement analogue montre alors l'existence de mots s'_i tels que $s_i = f(s'_i)$.

Observons maintenant que $f^n(q) = p'_i r' s'_i$ pour tout i ; en outre, $|p_{i+1}| - |p_i| < |r'|$ pour tout i , d'où l'on déduit que r' est une quasipériode de $f^n(q)$, achevant ainsi le raisonnement. \square

Lemme 3.9. *Soit r une quasipériode de $f^\omega(a)$; alors $r = f^n(q)$ pour un certain entier naturel n .*

Démonstration. En premier lieu, nous allons montrer que $f^n(q)$ est spécial droit. Ensuite, nous montrerons que toute quasipériode r de $f^\omega(a)$ telle que r^2 est facteur du mot est de la forme $f^n(q)$. Ceci nous permettra de conclure, grâce au théorème 3.3, que toutes les quasipériodes sont bien de cette forme.

Une vérification directe permet de voir que q est spécial droit dans $f^\omega(a)$. Étant donné que $f(a)$ et $f(b)$ commencent par des lettres différentes, si $f^n(q)$ est spécial droit dans $f^\omega(a)$, alors $f^{n+1}(q)$ l'est également. Ainsi, en raisonnant par récurrence, nous déduisons que tous les $f^n(q)$ sont spéciaux droits.

Supposons maintenant que r soit une quasipériode de $f^\omega(a)$ telle que r^2 soit facteur de $f^\omega(a)$. Nous allons montrer par récurrence sur $|r|$ que $r = f^n(q)$ pour un certain entier naturel n . Si $|r| \leq |f(a)ba|$, nous pouvons vérifier directement que $r = q$. Supposons maintenant que toutes les quasipériodes plus courtes que r soient de la forme voulue et montrons que r l'est également. Nous pouvons faire l'hypothèse que $|r| > |f(a)ba|$. Le mot $f(a)ba$ étant un préfixe de $f^\omega(a)$, il est donc également un préfixe de r . En outre, il est synchronisant pour f . Étant donné que r^2 est facteur de $f^\omega(a)$, ce dernier admet un préfixe de la forme πr^2 . Nous en déduisons l'existence de mots π', r' tels que $\pi = f(\pi')$ et $r = f(r')$.

Montrons maintenant que r' est une quasipériode de $f^\omega(a)$. Appelons $(p_i)_{i \in \mathbb{N}}$ les mots finis tels que $p_i r$ soient tous les préfixes de $f^\omega(a)$ se terminant par r , ordonnés par longueur. Comme r est une quasipériode, la relation $|p_{i+1}| - |p_i| \leq |r|$ est vérifiée. Le mot r ayant un préfixe synchronisant pour f , il existe des mots $(p'_i)_{i \in \mathbb{N}}$ tels que, pour tout entier i , le mot p'_i est le seul satisfaisant $p_i = f(p'_i)$. Ils vérifient en particulier

$|p'_{i+1}| - |p'_i| \leq |r'|$, donc r' est une quasipériode de $f^\omega(a)$. Nous en déduisons par hypothèse de récurrence que $r' = f^n(q)$ pour un certain entier naturel n , et $r = f(r') = f^{n+1}(q)$ est bien de la forme désirée. \square

Ce troisième lemme achève la preuve du théorème 3.6.

3.4 Sur la forme normale et les désubstitutions

Soit q un mot fini sur un alphabet Σ qui possède k antibords. Rappelons que le morphisme dérivateur gauche de q , noté δ_q , se définit comme suit : $\delta_q(i)$ est le $i^{\text{ème}}$ antibord gauche de q (par ordre de longueur croissant). Un mot infini \mathbf{w} sur Σ est q -quasipériodique si et seulement s'il existe un mot \mathbf{w}' sur $\{1, \dots, k\}$ tel que $\mathbf{w} = \delta_q(\mathbf{w}')$ (théorème 2.6). Nous appelons \mathbf{w}' le *mot dérivé à gauche* de \mathbf{w} . Nous avons déjà rencontré le morphisme dérivateur gauche dans la section 2.6 afin de construire des mots q -quasipériodiques avec diverses propriétés. Le morphisme η_q défini comme : « $\eta_q(i)$ est le $i^{\text{ème}}$ antibord droit de q (par ordre de longueur croissante) » est appelé le morphisme dérivateur droit de q . Un mot infini \mathbf{w} est q -quasipériodique si et seulement s'il existe un mot \mathbf{w}' tel que $\mathbf{w} = q\eta_q(\mathbf{w}')$; le mot \mathbf{w}' est appelé le *mot dérivé à droite* de \mathbf{w} .

La terminologie *quasipériodique multi-échelles* suggère la conjecture suivante : un mot \mathbf{w} est quasipériodique multi-échelles si et seulement s'il existe un mot dérivé à gauche de \mathbf{w} qui soit à son tour multi-échelles. Dans cette section, nous allons montrer qu'il n'en est rien : certains mots multi-échelles possèdent uniquement des mots dérivés superprimitifs. Appelons \mathbf{u} le mot défini par la proposition 3.4 et observons que \mathbf{u} satisfait aux conditions du lemme suivant.

Lemme 3.10. *Soit \mathbf{w} un mot quasipériodique multi-échelles dont toutes les quasipériodes sont spéciales droites. Pour toute quasipériode q de \mathbf{w} , le mot $\delta_q^{-1}(\mathbf{w})$ n'est pas quasipériodique.*

Démonstration. Soit q une quasipériode de \mathbf{w} et \mathbf{w}' le mot tel que $\mathbf{w} = \delta_q(\mathbf{w}')$. Supposons que r soit une quasipériode de \mathbf{w}' ; alors $\delta_q(r)$ est une quasipériode de \mathbf{w} . Par construction, $\delta_q(i)$ commence par la première lettre de q , appelons-la a , pour toute lettre i . Ainsi, $\delta_q(r)$ est toujours suivi de la lettre a dans \mathbf{w} , il n'est donc pas spécial droit : nous avons une contradiction avec l'hypothèse. \square

Le lemme précédent n'évoque que le mot dérivé à gauche. En effet, le mot dérivé à droite de \mathbf{u} est bien quasipériodique. Nous construisons maintenant un mot multi-échelles dont ni le dérivé à gauche, ni le dérivé à droite ne sont quasipériodiques. Soient f, g, λ, χ les morphismes définis comme suit.

$$f : \begin{cases} a \mapsto aba \\ b \mapsto ba \end{cases} \quad g : \begin{cases} a \mapsto aba \\ b \mapsto bba \end{cases} \quad \lambda : \begin{cases} a \mapsto aba \\ b \mapsto ab \end{cases} \quad \chi : \begin{cases} a \mapsto baa \\ b \mapsto bab \end{cases}$$

Théorème 3.11. *L'égalité $f(g^\omega(a)) = \lambda(\chi^\omega(b))$ est vérifiée ; appelons \mathbf{w} le mot qu'elle définit. Les quasipériodes de \mathbf{w} sont les mots de la forme $f(g^n(a)) = \lambda(\chi^n(b))$, pour n*

parcourant les entier naturels. Pour toute quasipériode q de \mathbf{w} , le mot dérivé à gauche est $g^\omega(a)$ et le mot dérivé à droite est $\chi^\omega(b)$; ces deux mots sont superprimitifs.

Le reste de cette section est dévolu à la preuve du théorème 3.11.

Posons $u_n = f(g^n(a))$ et $v_n = f(g^n(b))$.

Observation préliminaire. Soit π un mot fini. Si $\pi ababb$ est un préfixe de $g^\omega(a)$, alors il existe un mot fini π' tel que $\pi = g(\pi')$.

Étape 1. Soit x un préfixe non vide de $g^\omega(a)$; le mot xx est un facteur de $g^\omega(a)$ si et seulement si $x = g^n(a)$ ou bien $x = g^n(ab)$, pour n un entier naturel.

Une vérification directe montre que aa et $abab$ sont facteurs de $g^\omega(a)$. Une récurrence immédiate montre que, pour tout entier n , les mots $g^n(aa)$ et $g^n(abab)$ sont également des facteurs de $g^\omega(a)$. Réciproquement, soit x un préfixe de $g^\omega(a)$. Supposons que x est facteur de $g^\omega(a)$ et que $x \notin \{a, ab, g(a)\}$. Une vérification directe montre que $|x| \geq 6$ et que bb est facteur de x . Ainsi, il existe un mot x' et un préfixe strict u de $g(a)$ ou de $g(b)$ (c'est-à-dire que $u \in \{\varepsilon, a, ab, b, bb\}$) tels que $x = g(x')u$. Comme xx est facteur de $g^\omega(a)$, son préfixe $g(x')uababb$ est également un facteur de $g^\omega(a)$. Par l'observation préliminaire, nous avons $u = \varepsilon$ et $x = g(x')$, avec x' un préfixe de $g^\omega(a)$ et $x'x'$ un facteur de $g^\omega(a)$. Comme $|x'| < |x|$, nous pouvons itérer cet argument jusqu'à trouver $x \in \{a, ab, g(a)\}$.

Étape 2. Le mot $g^\omega(a)$ n'est pas quasipériodique.

Par le théorème 3.2, si $g^\omega(a)$ était quasipériodique, il aurait une quasipériode q qui serait un préfixe tel que qq serait facteur de $g^\omega(a)$. Par l'étape 1, nous avons $q = g^n(a)$ ou $q = g^n(ab)$ pour un certain entier n ; choisissons n minimal. Une vérification directe montre qu'aucun des mots de $\{a, ab, aba, ababba\}$ n'est une quasipériode de $g^\omega(a)$, donc $n \geq 2$ et q commence par $ababb$. Soit π un mot tel que πqq est un préfixe de $g^\omega(a)$. Par l'observation préliminaire et le fait que g est 3-uniforme, il existe des mots π' et q' tels que $\pi = g(\pi')$ et $q = g(q')$. Comme g est injectif, nous avons $q' = g^{n-1}(a)$ ou bien $q' = g^{n-1}(ab)$, avec q' un préfixe de $g^\omega(a)$ et $q'q'$ facteur de $g^\omega(a)$, ce qui contredit la minimalité de n .

Étape 3. Le mot \mathbf{w} est u_n -quasipériodique, pour tout entier n .

Il est immédiat que $\mathbf{w} \in u_n\{u_n, v_n\}^\omega$ pour tout n et que u_n et v_n commencent et finissent tous les deux par la lettre a . Nous montrons par récurrence que $u_0 = aba$, $v_n = a^{-1}u_n$ et $u_{n+1} = u_nv_nu_n$ pour tout n . Calculons.

$$\begin{aligned} u_{n+1} &= f(g^{n+1}(a)) = f(g^n(aba)) = u_nv_nu_n \\ v_{n+1} &= f(g^n(bba)) = v_nv_nu_n = (a^{-1}u_n)v_nu_n = a^{-1}u_{n+1} \end{aligned}$$

Donc $\mathbf{w} \in \{u_n, a^{-1}u_n\}$ et a et la dernière lettre de u_n , ce qui montre que \mathbf{w} est u_n -quasipériodique.

Étape 4. Soit x un préfixe non vide de $f(g^\omega(a))$. Le mot xx est facteur de $f(g^\omega(a))$ si et seulement si $x = a$, $x = ab$, $x = f(g^n(a))$ ou $x = f(g^n(ab))$ pour un entier n .

Si x a la forme prescrite, xx est facteur de $f(g^\omega(a))$ par l'étape 1. Réciproquement, supposons que xx est facteur de $f(g^\omega(a))$. Si $|x| \leq 6$, alors une vérification directe montre que $x = a$, $x = ab$, $x = aba = f(g^0(a))$, ou bien $x = ababa = f(g^0(ab))$. Si $|x| \geq 7$, $x = f(ab)abu$ pour un certain mot u . Par hypothèse, $f(ab)abuf(ab)abu$ est facteur de $f(g^\omega(a))$. Si π est un mot tel que $\pi f(ab)ab = \pi(ababaab)$ est un préfixe de \mathbf{w} , alors il existe π' tel que $\pi = f(\pi')$. Ainsi, $f(ab)abu = f(x')$ pour un certain mot x . Comme $f(a)$ et $f(b)$ commencent par des lettres différentes, $x'x'$ doit apparaître dans $g^\omega(a)$ et x' est un préfixe de $g^\omega(a)$. Par l'étape 1, $x' = g^n(a)$ ou bien $x' = g^n(ab)$ et x a bien la forme voulue.

Étape 5. Pour tout entier n , le mot u_n est un facteur spécial droit de \mathbf{w} .

Comme aa et ab sont des facteurs de $g^\omega(a)$, les mots $u_n u_n$ et $u_n v_n$ sont des facteurs de \mathbf{w} . Le mot u_n commence par a et v_n par b , donc $u_n a$ et $u_n b$ sont également des facteurs de \mathbf{w} .

Étape 6. Les quasipériodes de \mathbf{w} sont également les mots u_n , pour n parcourant les entiers naturels.

Observons que a et ab ne sont pas des quasipériodes de \mathbf{w} . Par les étapes 4, 5 et le théorème 3.2, il reste simplement à démontrer que $f(g^n(ab))$ n'est pas une quasipériode de \mathbf{w} , quel que soit n . La vérification est immédiate pour $n = 0$ et $n = 1$. Supposons donc que $f(g^n(ab))$ est une quasipériode pour $n \geq 2$. Le mot $f(g^n(ab))$ commence par $f(g(a)) = ababaa$. Ainsi, pour tout mot π tel que $\pi f(g^n(ab))$ est un préfixe de $f(g^\omega(a))$, il existe π' tel que $\pi = f(\pi')$. Il en découle que $\pi' g^n(ab)$ est un préfixe de $g^\omega(a)$. Le mot \mathbf{w} est quasipériodique si et seulement s'il existe une suite de mots $(\pi_k)_{k \in \mathbb{N}}$, rangés par longueurs croissantes, tels que $\pi_0 = \varepsilon$, et pour tout entier k , que $\pi_k f(g^n(ab))$ soit préfixe de \mathbf{w} et $|\pi_{k+1}| - |\pi_k| \leq f(g^n(ab))$. Soit $(\pi'_k)_{k \in \mathbb{N}}$ la suite des mots tels que π'_k est l'unique mot vérifiant $\pi_k = f(\pi'_k)$, pour tout k . Nous avons $\pi_0 = \varepsilon$ et, pour tout k , le mot $\pi'_k g^n(ab)$ est un préfixe de $g^\omega(a)$. De plus, $|\pi'_{k+1}| - |\pi'_k| \leq |g^n(ab)|$. Ceci implique que $g^n(ab)$ est une quasipériode de $g^\omega(a)$, ce qui contredit l'étape 2.

Étape 7. Les mots u_n et v_n sont primitifs, pour tout n .

Supposons dans un premier temps que $u_n = x^k$, pour un certain entier $k \geq 2$. Par l'étape 4, nous avons $u_n = a^k$, $u_n = (ab)^k$ ou bien pour un certain entier ℓ nous avons $u_n = f(g^\ell(a^k))$ ou $u_n = f(g^\ell((ab)^k))$. Les deux premières possibilités ne peuvent se réaliser car $u_n = aba$ ou bien u_n commence par $ababaa$. Dans les deux autres cas, étant donné que $u_n = f(g^n(a))$ et que f, g sont injectifs, nous avons $g^{n-\ell}(a) = a^k$ ou $g^{n-\ell}(a) = (ab)^k$. Ces deux égalités ne peuvent être vraies car $k \geq 2$ implique $n - \ell \geq 1$, et $g^{n-\ell}(a)$ est aba ou commence par $ababb$. Le cas $v_n = x^k$ mène également à une contradiction car v_n commence par la lettre b .

Étape 8. Pour tout entier n , le mot $f(g^n(aa))$ contient exactement deux occurrences de $f(g^n(a))$, et de plus pour tout entier ℓ , le mot $f(g^n(ab^\ell))$ contient exactement $\ell + 1$

occurrences de $f(g^n(a))$.

Par l'étape 7, le mot u_n est primitif et la proposition 1.2 implique que $f(g^n(aa)) = u_n u_n$ ne peut contenir u_n comme facteur interne. Il contient donc exactement deux occurrences de u_n .

Soit $\ell \geq 0$; observons que $f(g^n(ab^\ell)) = a(v_n)^{\ell+1}$. Comme v_n est primitif, par la proposition 1.2 il n'est pas facteur interne de $v_n v_n$ et $f(g^n(ab^\ell))$ contient exactement $\ell + 1$ occurrences de v_n . Le mot v_n finit par la lettre a et la preuve de l'étape 3 montre que $u_n = av_n$, donc $f(g^n(ab^\ell))$ contient exactement $\ell + 1$ occurrences de u_n .

Étape 9. Soit q une quasipériode de \mathbf{w} ; le mot dérivé droit de \mathbf{w} par q est $g^\omega(a)$.

Soit $n \geq 1$. Rappelons que $\mathbf{w} \in u_n \{u_n, v_n\}^\omega$ avec $u_n = f(g^n(a))$ et $v_n = f(g^n(b))$. L'étape 8 implique que $u_n u_n$ contient exactement deux occurrences de u_n , et pour tout entier ℓ que $u_n v_n^\ell$ contient exactement $\ell + 1$ occurrences de u_n . De plus, la preuve de l'étape 3 a montré que $v_n = a^{-1}u_n$. Ainsi le morphisme dérivateur droit de u_n est $(f \circ g^n)$, et comme g^n est injectif sur les mots infinis, le mot dérivé associé à u_n est nécessairement $g^\omega(a)$.

Étape 10. Pour tous entiers n et ℓ , nous avons $f(g^n(ab^\ell)) = \lambda(\chi^n(b^\ell a))$.

Nous le montrons par induction. Pour $n = 0$, nous calculons.

$$f(ab^\ell) = aba(ba)^\ell = (ab)^\ell aba = \lambda(b^\ell a)$$

Soit n un entier naturel. Nous avons la relation suivante.

$$f(g^{n+1}(ab^\ell)) = f(g^n(aba(bba)^\ell)) = f(g^n(ab))f(g^n((abb)^\ell))f(g^n(a))$$

Donc, par hypothèse d'induction,

$$f(g^{n+1}(ab)) = \lambda(\chi^n(ba))\lambda(\chi^n((bba)^\ell))\lambda(\chi^n(a)) = \lambda(\chi^n(ba(bba)^\ell a)) = \lambda(\chi^{n+1}(b^\ell a))$$

et la relation est démontrée.

Étape 11. La relation $f(g^\omega(a)) = \lambda(\chi^\omega(b))$ est vérifiée.

Par l'étape 10, nous pouvons calculer.

$$\begin{aligned} f(g^\omega(a)) &= \lim_{n \rightarrow \infty} f(g^n(a)) = \lim_{n \rightarrow \infty} \lambda(\chi^n(a)) = \lim_{n \rightarrow \infty} \lambda(\chi^{n-1}(baa)) = \lim_{n \rightarrow \infty} \lambda(\chi^n(b)) \\ &= \lambda(\chi^\omega(b)) \end{aligned}$$

Étape 12. Soit q une quasipériode de \mathbf{w} ; le mot dérivé gauche de \mathbf{w} par q est le mot $\chi^\omega(b)$.

Nous avons $\mathbf{w} \in u_n \{u_n, v_n\}^\omega$, et pour tout entier naturel n , la relation $u_n = f(g^n(a)) = \lambda(\chi^n(a))$. Une récurrence immédiate montre que $u_n = av_n = \lambda(\chi^n(b))a = af(g^n(b))$. Nous avons donc $\mathbf{w} \in \{\lambda(\chi^n(a)), \lambda(\chi^n(b))\}^\omega$. Comme $u_n v_n^\ell = (\lambda(\chi^n(b)))^\ell u_n$ contient exactement $\ell + 1$ occurrences de u_n et $u_n u_n$ contient exactement 2 occurrences de u_n , nous en déduisons que le morphisme dérivateur gauche de $\lambda(\chi^n(a))$ est $(\lambda \circ \chi^n)$. L'injectivité de $(\lambda \circ \chi^n)$ et l'égalité $\mathbf{w} = \lambda(\chi^\omega(b))$ impliquent que le mot dérivé gauche de \mathbf{w} est $\chi^\omega(b)$.

Étape 13. Le mot $\chi^\omega(b)$ est superprimitif.

Ceci est une conséquence du fait que $\chi^\omega(b)$ ne possède aucun carré comme préfixe. Supposons, par contradiction, que u soit le plus petit mot tel que uu est préfixe de $\chi^\omega(b)$. Si $|u| \equiv 0 \pmod 3$, alors $u = \chi(v)$ pour un mot v , et comme χ est uniforme et injectif, vv est un préfixe de $\chi^\omega(b)$, ce qui contredit la minimalité de u . Si $|u| \equiv 1 \pmod 3$, alors $u = \chi(v)b$ pour un mot v . Mais alors l'image d'une lettre par χ devrait commencer par bb , ce qui n'arrive pas dans la définition de χ . Enfin, si $|u| \equiv 2 \pmod 3$, alors $u = \chi(v)ba$ pour un certain mot v . L'image d'un certain mot par χ devrait alors commencer par $baba$, ce qui est impossible par définition de χ .

Ceci conclut la preuve du théorème 3.11 : la quasipériodicité multi-échelles n'est pas directement liée à la structure des mots dérivés.

*Science is what we understand well enough to explain to a computer.
Art is everything else we do.*

— Donald Knuth, *A = B*, 1996

Chapitre 4

Pouvoir expressif de la quasipériodicité

Le distinguo entre quasipériodicité simple (une seule quasipériode) et multi-échelles (une infinité de quasipériodes) est introduit dans l'article de Marcus et Monteil [68]. Ils montrent comment construire des mots quasipériodiques avec de mauvaises propriétés de symétrie, et en concluent que la quasipériodicité simple n'est pas une bonne notion pour le cas des mots infinis. Ils introduisent ensuite la *quasipériodicité multi-échelles*, c'est-à-dire le fait d'avoir une infinité de quasipériodes. Ils montrent qu'un mot infini quasipériodique multi-échelles est nécessairement uniformément récurrent, d'entropie topologique nulle et ont les fréquences de leurs facteurs qui convergent. Ils soulignent également que les systèmes dynamiques constituent une grille de lecture intéressante pour la notion de quasipériodicité.

Dans ce chapitre, nous approfondissons leurs résultats. Nous commençons par donner une introduction au point de vue dynamique sur les mots, ses motivations et les principales notions associées. Nous présentons ensuite une famille de mots très célèbre : les *mots sturmiens*, définis intuitivement comme les mots *les plus réguliers non ultimement périodiques*. Nous montrons ensuite comment les mots périodiques, ainsi que les mots sturmiens standards (une sous-famille importante des mots sturmiens) se caractérisent par des conditions sur leurs ensembles de quasipériodes. En d'autres termes, les caractères « périodique » et « sturmien standard » s'expriment en termes de propriétés sur l'ensemble des quasipériodes. Nous affinons ainsi le propos de Marcus et Monteil : non seulement une infinité de quasipériodes offre une grande régularité, mais en outre une plus grande « densité » de quasipériodes implique une plus grande régularité.

4.1 Introduction à la dynamique symbolique

Les mots infinis se révèlent utiles pour l'étude des systèmes dynamiques. Un *système dynamique discret* est un couple (X, T) où X est un espace topologique, et T une fonction continue de X vers X . Étant donné un point $x \in X$, la théorie des systèmes dynamiques s'intéresse à l'orbite de x sous T , autrement dit la suite $x, T(x), T^2(x), \dots$. La notion naturelle de transformation entre systèmes dynamiques est la conjugaison :

si (X, T) et (X', T') sont deux systèmes, une *conjugaison* est une fonction continue f de X vers X' , telle que $f \circ T = T' \circ f$. De nombreuses propriétés intéressantes sur les systèmes dynamiques sont invariantes par conjugaison ; pour étudier un tel système, il est donc possible de chercher une conjugaison avec un système plus simple, puis d'étudier ce dernier.

C'est ici que les mots infinis jouent un rôle. Soit Σ un alphabet et $\Sigma^{\mathbb{N}}$ l'ensemble des mots infinis sur Σ ; nous donnons une structure topologique à $\Sigma^{\mathbb{N}}$. Si \mathbf{w} et \mathbf{w}' sont deux mots infinis, nous disons que la *distance* entre \mathbf{w} et \mathbf{w}' est donnée par

$$d(\mathbf{w}, \mathbf{w}') = 2^{-\min\{n \in \mathbb{N} \mid \mathbf{w}(n) \neq \mathbf{w}'(n)\}}$$

et $d(\mathbf{w}, \mathbf{w}) = 0$. En d'autres termes, plus \mathbf{w} et \mathbf{w}' possèdent un long préfixe commun, plus la distance entre eux sera petite. La symétrie de d est immédiate, et l'inégalité triangulaire se démontre aisément.

Nous pouvons donc considérer sur $\Sigma^{\mathbb{N}}$ la topologie induite par la distance d . Dans cette topologie, les boules ouvertes (ou fermées, ce sont les mêmes) sont données par les *cylindres*. Si u est un mot fini, le cylindre de u , noté $[u]$, est l'ensemble des mots admettant u comme préfixe. Par définition, tous les cylindres sont des boules ouvertes et réciproquement.

Cette topologie permet déjà de donner un sens formel à la notion de *convergence* de suites de mots. Nous disons qu'une suite de mots infinis converge si elle converge dans la topologie ci-dessus. Si $(u_n)_{n \in \mathbb{N}}$ est une suite de mots finis, nous disons qu'elle converge vers un mot infini \mathbf{w} si et seulement si $(|u_n|)$ tend vers l'infini et $(u_n^\omega)_{n \in \mathbb{N}}$ converge vers \mathbf{w} en tant que suite de mots infinis. La convergence est en particulier assurée lorsque u_{n+1} est préfixe de u_n pour tout n ; nous évoquons ce cas particulier au chapitre 1.

Pour obtenir un système dynamique, nous avons également besoin d'une transformation continue. Notons σ l'opération qui supprime la première lettre d'un mot infini : $\sigma(\mathbf{w})(n) = \mathbf{w}(n+1)$. Cette fonction est continue, car pour tout ε positif et pour tous $x, y \in \Sigma^{\mathbb{N}}$, si $d(x, y) < \varepsilon/2$, alors $d(\sigma(x), \sigma(y)) < \varepsilon$. En outre, cette transformation très simple met en correspondance les notions de systèmes dynamiques avec les notions de combinatoire des mots. Par exemple, l'entropie topologique est définie de façon générale sur tous les systèmes dynamiques, mais dans le cadre de $(\Sigma^{\mathbb{N}}, \sigma)$ elle correspond exactement à la définition en termes de complexité en facteurs.

Si X est une partie de $\Sigma^{\mathbb{N}}$, fermée (pour la topologie que nous avons définie) et stable par σ , nous disons que (X, σ) est un *système symbolique* (en anglais *subshift*). Il s'agit d'un cas particulier de système dynamique discret. Tous les systèmes symboliques sont des *sous-systèmes* de $(\Sigma^{\mathbb{N}}, \sigma)$, ce dernier étant appelé le *full shift* en anglais. L'étude des systèmes symboliques revient à l'étude de la combinatoire des mots ; nos résultats peuvent donc systématiquement se traduire dans un contexte dynamique.

Nous pouvons construire un système symbolique à partir de n'importe quel mot infini \mathbf{w} . Soit S l'orbite de \mathbf{w} , c'est-à-dire $S = \{\mathbf{w}, \sigma(\mathbf{w}), \sigma^2(\mathbf{w}), \dots\}$ et \bar{S} l'adhérence de S . L'ensemble \bar{S} est un fermé — puisque l'adhérence est l'intersection de tous les fermés qui contiennent S — invariant par σ , puisque S l'est et que prendre l'adhérence n'ajoute que les points d'accumulation des suites de S . Nous disons que (\bar{S}, σ) est le *système engendré* par le mot \mathbf{w} .

Tout comme les mots infinis, les systèmes dynamiques peuvent posséder des propriétés de symétrie. Par exemple, un système symbolique est dit *minimal* s'il ne contient strictement aucun autre système. Le théorème suivant caractérise cette propriété en termes de combinatoire des mots et en termes topologiques.

Théorème 4.1 (Initialement Morse, Hedlund, 1938 [69]; voir aussi [7]). *Un système symbolique \mathbf{w} est minimal si et seulement s'il est engendré par un mot uniformément récurrent, si et seulement si tout élément du système possède une orbite dense (qui intersecte tous les ouverts).*

Tous les systèmes symboliques ne sont pas nécessairement engendrés par un mot. Le *full shift* en est un premier exemple. On peut également considérer pour X l'ensemble des mots infinis sur $\{a, b\}$ qui ne contiennent pas le facteur ab . Nous avons dans ce cas $X = \{a^\omega, b^\omega\} \cup \{b^n a \mid n \in \mathbb{N}\}$. Cet ensemble est clairement stable par σ , et il est également fermé : toute suite d'éléments de X a toutes ses valeurs d'adhérence dans X .

Généralisons cet exemple. Soit F un ensemble de mots finis ; l'ensemble X des mots infinis dont aucun facteur n'est dans F , autrement dit $X = \{\mathbf{w} \mid \mathcal{L}(\mathbf{w}) \cap F = \emptyset\}$ est toujours invariant par σ et fermé (voir par exemple [7]). Nous disons que l'ensemble X est *défini par l'ensemble des facteurs interdits* F . Notons que $\Sigma^{\mathbb{N}}$ est défini en prenant comme ensemble de facteurs interdits l'ensemble vide.

Théorème 4.2 (Voir par exemple [7]). *Le couple (X, σ) est un système symbolique si et seulement si X est défini par des facteurs interdits.*

L'étude des systèmes symboliques revient donc à l'étude des ensembles de mots définissables par des facteurs interdits. L'ouvrage [60] constitue une introduction approfondie à l'étude des systèmes symboliques du point de vue de leurs facteurs interdits. Nous référons également le lecteur aux ouvrages [35] et [8], ainsi que les articles originaux [69, 70] pour en savoir plus au sujet des systèmes symboliques.

4.2 Les mots sturmiens

Commençons par un corollaire classique du théorème 2.2 de Morse-Hedlund.

Lemme 4.3. *Soit \mathbf{w} un mot infini. S'il existe un entier n tel que $P_{\mathbf{w}}(n) \leq n$, alors \mathbf{w} est ultimement périodique.*

Démonstration. Soit \mathbf{w} un mot non ultimement périodique. Nous avons $P_{\mathbf{w}}(0) = 1$, puisque le mot vide est l'unique facteur de longueur 0 et qu'il est facteur de tous les mots. De plus, pour tout entier n , nous avons $P_{\mathbf{w}}(n+1) > P_{\mathbf{w}}(n)$; en effet, la fonction $P_{\mathbf{w}}$ ne peut décroître (quel que soit \mathbf{w}) et le théorème 2.2 affirme que si $P_{\mathbf{w}}(n+1) = P_{\mathbf{w}}(n)$ pour un certain n , alors \mathbf{w} serait ultimement périodique. Une récurrence immédiate montre donc que, dans ces conditions, $P_{\mathbf{w}}(n) \geq n + 1$. \square

En conséquence, la plus « petite » fonction de complexité possible pour un mot infini non ultimement périodique est $P_{\mathbf{w}}(n) = n + 1$. Il se trouve qu'il existe des mots atteignant cette complexité.

Théorème 4.4 (Voir par exemple [12, Section 4.10.3]). *Le mot de Fibonacci \mathbf{f} , défini comme $\phi^\omega(a)$ pour $\phi(a) = ab$ et $\phi(b) = a$, vérifie $P_{\mathbf{f}}(n) = n + 1$.*

Un mot \mathbf{w} vérifiant $P_{\mathbf{w}}(n) = n + 1$ pour tout n est appelé un *mot sturmien*. Ils furent étudiés par Morse et Hedlund dans [70], et font depuis l'objet d'une abondante littérature. Nous renvoyons le lecteur à [61, chapitre 2] et à [35, chapitre 6] pour un survol complet.

Les mots sturmiens peuvent être définis de nombreuses façons équivalentes. Nous disons qu'un mot \mathbf{w} sur un alphabet binaire $\{a, b\}$ est *équilibré* lorsque, pour tous facteurs u, v de même longueur dans u , le nombre d'occurrences de a diffère au plus de 1; en d'autres termes, $-1 \leq |u|_a - |v|_a \leq 1$. Les morphismes suivants sont également importants dans l'étude des mots sturmiens (L pour *left* et R pour *right*).

$$L_a : \begin{cases} a & \mapsto a \\ b & \mapsto ab \end{cases} \quad L_b : \begin{cases} a & \mapsto ba \\ b & \mapsto b \end{cases} \quad R_a : \begin{cases} a & \mapsto a \\ b & \mapsto ba \end{cases} \quad R_b : \begin{cases} a & \mapsto ab \\ b & \mapsto b \end{cases}$$

Théorème 4.5 (Voir par exemple [35, chapitre 6]). *Soit \mathbf{w} un mot infini. Les conditions suivantes sont équivalentes :*

1. \mathbf{w} est sturmien, c'est-à-dire que $P_{\mathbf{w}}(n) = n + 1$;
2. \mathbf{w} est équilibré et non ultimement périodique;
3. il existe une suite de mots infinis $(\mathbf{w}_n)_{n \in \mathbb{N}}$ telle que $\mathbf{w}_0 = \mathbf{w}$ et $\mathbf{w}_n = \mu_n(\mathbf{w}_{n+1})$, avec $\mu_n \in \{L_a, L_b, R_a, R_b\}$, pour tout n et $(\mu_n)_{n \in \mathbb{N}}$ infiniment souvent dans $\{L_a, R_a\}$ et dans $\{L_b, R_b\}$.
4. il existe $\alpha \in [0; 1] \setminus \mathbb{Q}$ et $\beta \in [0; 1]$ vérifiant l'une des deux relations suivantes

$$\forall i \in \mathbb{N}, \mathbf{w}(i) = \lceil \alpha(i+1) + \beta \rceil - \lceil \alpha i + \beta \rceil \quad (4.1)$$

$$\forall i \in \mathbb{N}, \mathbf{w}(i) = \lfloor \alpha(i+1) + \beta \rfloor - \lfloor \alpha i + \beta \rfloor \quad (4.2)$$

quitte à recoder $\{0, 1\}$ sur $\{a, b\}$.

Outre leurs nombreuses définitions équivalentes (le théorème précédent n'est pas exhaustif), les mots sturmiens possèdent de nombreuses propriétés remarquables : ils sont par exemple tous uniformément récurrents, les fréquences de leurs facteurs convergent, et leur entropie topologique est nulle (presque par définition : $P_{\mathbf{w}}(n) = n + 1$). En outre, la suite $(\mu_n)_n$ évoquée dans la condition 3 du théorème précédent est liée au développement en fraction continue du réel α dans la condition 4. Ce réel α est appelé la *pente* du sturmien. Deux sturmiens \mathbf{w} et \mathbf{w}' de même pente possèdent les mêmes facteurs, et il existe un entier n tel que $\mathbf{w} = \sigma^n(\mathbf{w}')$ ou bien $\mathbf{w}' = \sigma^n(\mathbf{w})$. Les mots sturmiens ont fait l'objet d'innombrables publications scientifiques (voir les références de [35, chapitre 6] et [61, chapitre 2]). Certains mots sturmiens possèdent des propriétés encore plus remarquables.

Théorème 4.6. *Soit \mathbf{w} un mot sturmien. Les conditions suivantes sont équivalentes :*

1. *tout préfixe de \mathbf{w} est un facteur spécial gauche ;*
2. *la suite $(\mu_n)_{n \in \mathbb{N}}$ (voir théorème 4.5) vérifie $\mu_n \in \{L_a, L_b\}$ pour tout n et μ_n n'est pas ultimement constante ;*
3. *le réel β (voir théorème 4.5) est égal à 0.*

Nous disons d'un sturmien satisfaisant aux conditions du théorème 4.6 qu'il est *standard*.

L'article original de Marcus [67] sur la quasipériodicité des mots infinis posait diverses questions au sujet des mots sturmiens. En réponse, les travaux de Levé et Richomme donnèrent une caractérisation des mots sturmiens quasipériodiques, ainsi que des quasipériodes d'un mot sturmien quasipériodique arbitraire [59]. Ces résultats furent ensuite étendus à des classes plus larges de mots, telles que les épisturmiens (une généralisation sur des alphabets plus grands) [41].

Pour exprimer ces caractérisations, nous aurons besoin de nouvelles notions. Supposons que notre alphabet Σ (dans le cas des sturmiens, $\Sigma = \{a, b\}$) est muni d'un ordre total sur ses lettres. L'*ordre lexicographique* sur $\Sigma^{\mathbb{N}}$ est défini comme suit : $\mathbf{w} \leq \mathbf{w}'$ si et seulement s'il existe un mot fini u , des mots infinis \mathbf{v} et \mathbf{v}' et des lettres ℓ, ℓ' tels que $\mathbf{w} = u\ell\mathbf{v}$ et $\mathbf{w}' = u\ell'\mathbf{v}'$, avec $\ell \leq \ell'$. Un mot infini est *de Lyndon* si et seulement s'il est strictement inférieur, dans l'ordre lexicographique, à tous ses suffixes. (Une notion homonyme existe sur les mots finis, mais nous n'en aurons pas besoin dans cette thèse.) La notion de mot de Lyndon dépend donc de l'ordre choisi sur l'alphabet.

Théorème 4.7 (Levé, Richomme, 2007 [59]). *Soit \mathbf{w} un mot sturmien. Les conditions suivantes sont équivalentes :*

1. *\mathbf{w} n'est pas quasipériodique ;*
2. *\mathbf{w} se décompose sur $\{L_a, R_b\}$ ou bien sur $\{L_b, R_a\}$;*
3. *il existe un ordre total sur $\{a, b\}$ tel que \mathbf{w} est un mot de Lyndon.*

De façon plus générale, nul mot infini de Lyndon n'est quasipériodique. Tous les mots sturmiens qui possèdent au moins une quasipériode en possèdent en fait une infinité ; en d'autres termes, tout mot sturmien est soit de Lyndon, soit quasipériodique multi-échelles. Considérons α un nombre dans $[0; 1] \setminus \mathbb{Q}$; il existe une infinité non dénombrable de sturmiens de pente α , et exactement deux d'entre eux sont des mots de Lyndon. Nous pouvons donc dire informellement que l'immense majorité des mots sturmiens sont quasipériodiques multi-échelles. Les mots de Lyndon peuvent être regardés comme des cas particuliers.

4.3 Caractérisations en termes de quasipériodes

Non content de fournir une méthode d'étude des quasipériodes d'un mot infini arbitraire, le théorème 3.2 appelle à l'étude de deux cas particuliers. Le premier est le

cas où un mot \mathbf{w} possède un seul intervalle de longueurs de quasipériodes de la forme $[i; \infty)$. Cela signifie que tous les préfixes, à partir d'une certaine longueur i , sont des quasipériodes. Par le corollaire 3.3, le mot \mathbf{w} ne possède alors pas de préfixe spécial droit de longueur supérieure à i ; le théorème de Morse et Hedlund implique que ce mot est ultimement périodique. Le fait que \mathbf{w} soit quasipériodique multi-échelles permet de conclure facilement qu'il est en fait purement périodique. Réciproquement, on observe sans difficulté que tout mot purement périodique admet tout préfixe plus long que sa période minimale comme quasipériode. Nous venons donc de prouver la proposition suivante.

Proposition 4.8. *Soit \mathbf{w} un mot infini. Alors \mathbf{w} est périodique si et seulement s'il existe un entier n tel que tout préfixe plus long que n soit une quasipériode.*

Ce résultat, en dépit de sa simplicité, nous ouvre une voie intéressante. En effet, il montre que la périodicité s'exprime en termes de quasipériodicités. Cette idée, bien que naturelle, n'avait pas été développée dans les travaux antérieurs à cette thèse. Nous explorons donc les autres cas extrêmes du corollaire 3.3. En particulier, dans un mot infini, la continuation d'un préfixe spécial droit n'est jamais une quasipériode.

Question. Existe-t-il un mot infini \mathbf{w} :

- qui soit aperiodique, c'est-à-dire qui possède une infinité de préfixes spéciaux droits ;
- dont tout préfixe qui n'est pas la continuation d'un facteur spécial droit soit une quasipériode ?

La deuxième condition exprime, moralement, que l'ensemble de quasipériodes de \mathbf{w} est « maximal », autrement dit que tout préfixe doit être une quasipériode à moins que cela ne soit explicitement interdit par le corollaire 3.3. En se référant à la figure 4.1, nous demandons un mot dont tous les préfixes, sauf ceux noircis, sont des quasipériodes. De façon surprenante, de tels mots existent et coïncident exactement avec une famille bien connue.

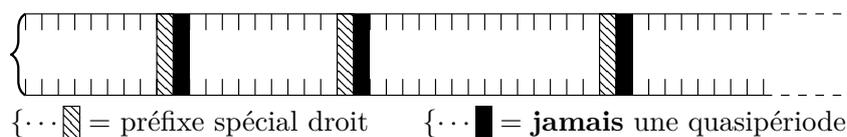


FIG. 4.1 : Préfixes d'un mot quasipériodique

Théorème 4.9. *Soit \mathbf{w} un mot aperiodique et p_i le préfixe de longueur i de \mathbf{w} . Alors \mathbf{w} est un sturmien standard si et seulement s'il possède une infinité de quasipériodes et satisfait la condition suivante : pour tout entier $n \geq 1$, le préfixe p_n est une quasipériode si et seulement si p_{n-1} n'est pas spécial droit.*

Nous traiterons les deux implications de ce théorème séparément. Chacune possède un intérêt propre : l'une donne une caractérisation simple et complète des quasipériodes des mots sturmiens standard. L'autre donne une nouvelle définition des sturmiens standards en termes de quasipériodicités, se plaçant ainsi dans la continuité de la proposition 4.8.

Lemme 4.10. *Soit \mathbf{w} un mot sturmien standard et p_i le préfixe de longueur i de \mathbf{w} . Soit n un entier strictement positif. Alors p_n est une quasipériode de \mathbf{w} si et seulement si p_{n-1} n'est pas bispécial.*

Démonstration. Si p_{n-1} est bispécial, alors p_n ne peut pas être une quasipériode par le théorème 3.2. Réciproquement, si p_{n-1} n'est pas bispécial, appelons ℓ et r l'unique facteur spécial gauche et spécial droit de longueur $n-1$ dans \mathbf{w} ; comme \mathbf{w} est standard, $\ell = p_{n-1}$. Soit $\mathcal{G} = \mathcal{G}_{\mathbf{w}}(n-1)$ le graphe de Rauzy de \mathbf{w} à l'ordre $n-1$. Ce graphe est l'union de trois chemins orientés arc-disjoints : un de ℓ vers r , que nous appelons C , et deux de r vers ℓ , que nous appelons D et E (cf. figure 4.2). Comme $\ell \neq r$, nous avons $|C| \geq 1$, $|D| \geq 1$ et $|E| \geq 1$. Par définition d'un sturmien, $\mathcal{G}_{\mathbf{w}}(n-1)$ possède n sommets et $n+1$ arcs ; nous avons donc $|C| + |D| \leq n$ et $|C| + |E| \leq n$. De plus, par définition d'un sturmien standard, le premier arc de C est p_n . En observant la morphologie du graphe, nous constatons que tout chemin commençant par p_n rencontre à nouveau p_n après n arcs au maximum, donc p_n est une quasipériode de \mathbf{w} . \square

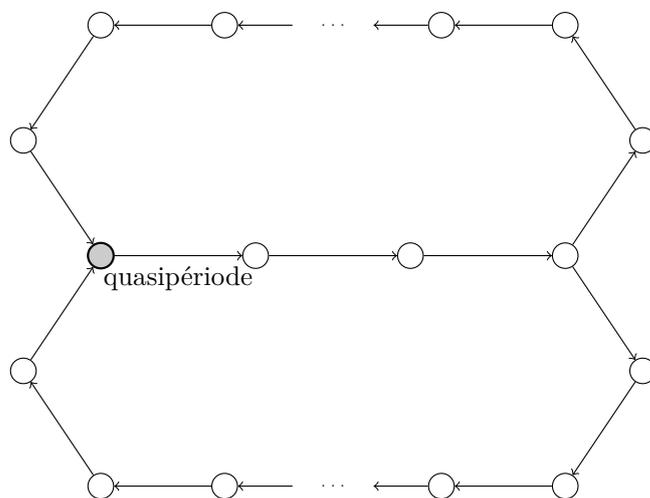


FIG. 4.2 : Graphe de Rauzy générique d'un mot Sturmien standard

Avant de prouver l'autre direction du théorème 4.9, nous avons besoin d'un résultat technique sur un morphisme bien particulier.

Lemme 4.11. *Soient \mathbf{w} un mot infini et q un mot fini sur l'alphabet $\{a, b\}$ et c une lettre de cet alphabet.*

1. *Le mot q est quasipériode de \mathbf{w} si et seulement si $L_a(q)$ et $L_a(q)a$ le sont dans $L_a(\mathbf{w})$.*
2. *Un mot u est spécial droit dans \mathbf{w} si et seulement si $L_a(u)aa$ l'est dans $L_a(\mathbf{w})$.*

Rappelons que $L_a(a) = a$ et $L_a(b) = ab$.

Démonstration. Les affirmations se prouvent séparément.

Affirmation 1. Si q est une quasipériode de \mathbf{w} , il est immédiat que $L_a(q)$ est une quasipériode de $L_a(\mathbf{w})$. Toute occurrence de $L_a(q)$ dans $L_a(\mathbf{w})$ est suivie de $L_a(a)$ ou de $L_a(b)$, qui commencent tous deux par a . Le facteur $L_a(q)$ n'est pas spécial droit dans $L_a(\mathbf{w})$, et $L_a(q)a$ est également quasipériode de $L_a(\mathbf{w})$. Réciproquement, supposons que $L_a(q)$ et $L_a(q)a$ sont des quasipériodes de $L_a(\mathbf{w})$. Considérons la suite $(\pi_k)_{k \in \mathbb{N}}$ de préfixes de $L_a(\mathbf{w})$, ordonnés par longueur croissante, tels que pour tout k , le mot $\pi_k L_a(q)$ soit préfixe de $L_a(\mathbf{w})$. Comme a est synchronisant pour L_a et que $L_a(q)$ commence toujours par a , il existe une suite $(\pi'_k)_{k \in \mathbb{N}}$ vérifiant $L_a(\pi'_k) = \pi_k$ pour tout entier k . Sachant que $|\pi_{k+1}| - |\pi_k| \leq L_a(q)$ pour tout entier k , une démonstration par récurrence montre que $|\pi'_{k+1}| - |\pi'_k| \leq q$ pour tout entier k .

Affirmation 2. Le mot u est spécial droit dans \mathbf{w} si et seulement si ua et ub apparaissent dans \mathbf{w} , si et seulement si $L_a(u)aa$ et $L_a(u)ab$ apparaissent dans $L_a(\mathbf{w})$, si et seulement si $L_a(u)a$ est spécial droit. \square

Lemme 4.12. *Soit \mathbf{w} un mot infini apériodique, quasipériodique multi-échelles et p_i le préfixe de longueur i de \mathbf{w} . Supposons que p_n soit une quasipériode de \mathbf{w} si et seulement si p_{n-1} n'est pas spécial droit. Alors \mathbf{w} est sturmien standard.*

Démonstration. Appelons a la première lettre de \mathbf{w} et B l'ensemble des lettres autres que a qui composent \mathbf{w} .

Fait 1. *Tout facteur non vide de \mathbf{w} dans B^* est de longueur 1.*

En premier lieu, tout facteur dans B^* a sa longueur bornée par celle de la plus petite quasipériode de \mathbf{w} , puisque toutes les quasipériodes de \mathbf{w} commencent par a . Soit x un facteur dans B^* parmi ceux de longueur maximale et p le plus court préfixe de \mathbf{w} finissant par x . Par maximalité de x , le facteur p est toujours suivi par la lettre a , il n'est donc pas spécial droit. Ainsi, par hypothèse, pa est une quasipériode de \mathbf{w} . Par construction, p commence par a et finit par ax , et ne contient aucune autre occurrence de x . Les bords de pa sont donc les mots ε et a . Ses antibords sont p et pa . Par le théorème 2.6, \mathbf{w} se décompose sur $p \cdot \{p, ap\}^\omega$. Soient π le mot et b la lettre tels que $p = \pi b$. Le mot \mathbf{w} se décompose donc sur $\pi b \{\pi b, a\pi b\}^\omega$. Comme il n'y a qu'une seule occurrence de x dans p , en position suffixe, il n'y en a aucune dans π . De plus, π commence par la lettre a .

Supposons maintenant que π ne soit pas spécial droit. Par hypothèse, p serait une quasipériode de \mathbf{w} ; or par la définition de x et de p , le mot p ne peut être facteur interne de pap , nous aurions donc $\mathbf{w} = p^\omega$, une contradiction avec l'apériodicité de \mathbf{w} .

Donc π est spécial droit et il existe une lettre c , différente de b , telle que πc apparaît dans \mathbf{w} .

Nous distinguons deux cas. Si πc est facteur de $\pi b \pi b$, alors le lemme 3.1 implique que $c = b$: une contradiction. Supposons donc que πc soit un facteur de $\pi b a \pi b$. Le mot x est suffixe de $p = \pi b$, et π (donc πc) n'admet pas x pour facteur. Ainsi, πc est facteur interne de $x a \pi b$. Comme π commence par a et que x est écrit sur B^* , nous avons nécessairement $\pi c = a \pi$, ce qui implique $a = c$ et $\pi = a^{|\pi|}$. Nous en déduisons $|x| = 1$.

Fait 1 bis. Il existe un mot apériodique \mathbf{w}' tel que $\mathbf{w} = L_a(\mathbf{w}')$. Nous posons p'_n le préfixe de longueur n de \mathbf{w}' . (Ce résultat est une reformulation directe du fait 1.)

Fait 1 ter. *Le mot \mathbf{w} est binaire.*

Supposons que \mathbf{w} contienne trois différentes lettres, a, b et c . La lettre a est la première de \mathbf{w} ; considérons sans perte de généralité que b apparaisse avant c pour la première fois. Soit π le plus court préfixe de \mathbf{w} terminant par un b . Par le fait 1, la lettre b est toujours suivie de la lettre a , donc le mot πa est un préfixe de \mathbf{w} . Étant donné que πa ne contient pas c , il ne peut être une quasipériode de \mathbf{w} (celles-ci doivent contenir toutes les lettres). Par hypothèse, π est spécial droit; c'est une contradiction avec le fait que b n'est pas spécial droit.

Fait 2. *La plus courte quasipériode de \mathbf{w} est un préfixe de la forme $a^k b a$, pour $k \geq 1$.*

Par les faits 1 et 1 ter, toute occurrence de b est suivie d'une occurrence de a . En particulier, le préfixe $a^k b$ n'est pas spécial droit, donc par hypothèse sur \mathbf{w} , le mot $a^k b a$ est une quasipériode de \mathbf{w} . Comme \mathbf{w} est apériodique, $a^k b$ n'est pas une quasipériode de \mathbf{w} (le seul antibord de $a^k b$ est le mot lui-même). En outre, toutes les quasipériodes de \mathbf{w} possèdent plusieurs lettres différentes, sans quoi \mathbf{w} serait unaire, et donc périodique.

Fait 3. *Pour tout entier n , le mot p'_{n+1} (cf. fait 1 bis) est une quasipériode de \mathbf{w}' si et seulement si p'_n n'est pas spécial droit dans \mathbf{w}' .*

Soit p un préfixe de \mathbf{w}' et c la lettre telle que pc est encore un préfixe de \mathbf{w}' . Supposons que p ne soit pas spécial droit dans \mathbf{w}' et montrons que pc est une quasipériode de ce mot. Par le lemme 4.11 (affirmation 1), nous devons montrer que $L_a(pc)$ et $L_a(pc)a$ sont des quasipériodes de $\mathbf{w} = L_a(\mathbf{w}')$. En utilisant l'hypothèse sur \mathbf{w} , il s'agit de montrer que les préfixes de $L_a(pc)$ et $L_a(pc)a$ de longueurs respectives $|L_a(pc)| - 1$ et $|L_a(pc)a| - 1$ ne sont pas spéciaux droits dans \mathbf{w} . Nous distinguons deux cas.

Si $c = a$, nous avons $L_a(pc) = L_a(p)a$ et $L_a(pc)a = L_a(p)aa$. Le mot $L_a(p)$ n'est pas spécial droit dans \mathbf{w} car les images de chaque lettre par L_a commencent par a . Comme p n'est pas spécial droit dans \mathbf{w}' , le mot $L_a(p)a$ ne l'est pas non plus dans \mathbf{w} par le lemme 4.11 (affirmation 2).

Si $c = b$, alors $L_a(pc) = L_a(p)ab$ et $L_a(pc)a = L_a(p)aba$. Comme p n'est pas spécial droit dans \mathbf{w}' , le mot $L_a(p)a$ n'est pas spécial droit dans \mathbf{w} par le lemme 4.11 (affirmation 2). Le mot $L_a(p)ab = L_a(pb)$ n'est pas non plus spécial droit dans \mathbf{w} , puisque l'image de chaque lettre commence par a .

Supposons maintenant que pc soit une quasipériode de \mathbf{w}' et montrons que p n'est pas spécial droit. Par le lemme 4.11 (affirmation 1), les mots $L_a(pc)$ et $L_a(pc)a$ sont des quasipériodes de \mathbf{w} . L'hypothèse sur \mathbf{w} implique que $L_a(p)a$ n'est pas spécial droit dans \mathbf{w} . Le lemme 4.11 (affirmation 2) montre alors que p n'est pas spécial droit dans \mathbf{w}' .

Fait 4. *Le mot \mathbf{w}' possède une infinité de quasipériodes.*

Par le fait 2, il existe un entier k tel que a^kba est une quasipériode de \mathbf{w} , donc \mathbf{w} se décompose sur $a^kb\{a^kb, a^{k+1}b\}$. Comme \mathbf{w} est apériodique, il existe un entier $i \geq 1$ tel que $(a^kb)^i a^{k+1}b$ est un préfixe de \mathbf{w} . Soit j le plus grand entier tel que $((a^kb)^i a^{k+1}b)^j$ soit un préfixe de \mathbf{w} ; l'existence de j est assurée car \mathbf{w} est apériodique. Soit $(u_n)_{n \in \mathbb{N}}$ la suite des préfixes, ordonnés par longueur croissante, qui commencent par $(a^kb)^i a^{k+1}b$ et qui terminent par $((a^kb)^i a^{k+1}b)^{j-1} (a^kb)^i a^{k+1}$; le fait que \mathbf{w} soit multi-échelles implique l'uniforme récurrence (théorème 2.8), donc $(u_n)_{n \in \mathbb{N}}$ est bien une suite infinie. Soit n un entier arbitraire. Comme a^{k+2} n'est pas facteur de \mathbf{w} , toute occurrence de u_n est suivie de la lettre b . Par le choix de j , le mot $u_n b u_n b$ n'est pas un facteur de \mathbf{w} . Deux occurrences consécutives de $u_n b$ doivent donc toujours partager au moins $|(a^kb)^i a^{k+1}b|$ lettres. Soit u'_n l'unique mot vérifiant $L_a(u'_n) = u_n$; ce mot est quasipériode de \mathbf{w}' . Nous avons donc exhibé une infinité de quasipériodes de \mathbf{w}' .

Conclusion. Nous avons prouvé que \mathbf{w} est un mot binaire (disons sur l'alphabet $\{a, b\}$); qu'il existe un mot multi-échelles \mathbf{w}' vérifiant $\mathbf{w} = L_x(\mathbf{w}')$ pour une lettre x ; que ce mot \mathbf{w}' satisfait aux hypothèses du présent lemme. En conséquence, nous pouvons itérer l'argument sur \mathbf{w}' , et ainsi de suite. Le mot \mathbf{w} est donc décomposable sur $\{L_a, L_b\}$, et ainsi un sturmien standard (théorème 4.6). \square

Nous pouvons reformuler les deux résultats de cette section d'une manière plus concise en définissant la *fonction de quasipériodicité*. (Une notion homonyme existe dans [20] mais elle est complètement différente, puisque dans l'article en question, « quasipériodicité » désigne en fait l'uniforme récurrence.) Si \mathbf{w} est un mot infini, nous notons $Q_{\mathbf{w}}(n)$ le nombre de quasipériodes de longueur n dans \mathbf{w} ; la fonction $Q_{\mathbf{w}}$ est la fonction de quasipériodicité de \mathbf{w} . Notons que $Q_{\mathbf{w}}$ prend ses valeurs dans $\{0, 1\}$. Toutefois, cette notion permet d'exprimer les caractérisations des mots périodiques et sturmiens d'une façon plus concise.

Corollaire 4.13. *Soit \mathbf{w} un mot infini.*

1. *Le mot \mathbf{w} est périodique si et seulement si $Q_{\mathbf{w}}$ ne s'annule pas à partir d'un certain rang.*
2. *Le mot \mathbf{w} est un sturmien standard si et seulement si les intervalles d'entiers sur lesquels $Q_{\mathbf{w}}$ s'annule sont tous de longueur 1, sauf le premier.*

Ce corollaire exprime que les sturmiens standards sont *les mots apériodiques ayant le plus de quasipériodes*. Nous trouvons ici un parallèle frappant avec la complexité en facteurs et le théorème de Morse-Hedlund. En effet nous expliquions, au début de ce

chapitre (lemme 4.3 et la discussion qui suit) que les mots sturmiens sont *les mots apériodiques ayant le moins de facteurs*. Ceci nous laisse penser que la fonction de quasipériodicité est une bonne source d’informations au sujet de la régularité d’un mot infini, au même titre que la fonction de complexité. Pour mener ce parallèle plus loin, nous aimerions nous débarrasser de la condition *standard* et parler des mots sturmiens en général. Cette ambition n’est pas encore réalisée à l’heure actuelle, mais la section suivante décrit des pistes allant dans ce sens.

4.4 Cas des mots bi-infinis

Pour généraliser le corollaire 4.13 à tous les mots sturmiens, nous devons considérer des mots bi-infinis. Soit Σ un alphabet. Un *mot biinfini* est simplement une fonction de \mathbb{Z} vers Σ ; ce sont des mots qui n’ont pas d’origine. Les définitions que nous avons données jusqu’à présent s’étendent aux mots bi-infinis, à l’exception de la notion de préfixe.

Un mot bi-infinis \mathbf{w} est *sturmien* si et seulement s’il vérifie $P_{\mathbf{w}}(n) = n + 1$ pour tout n et s’il est uniformément récurrent. Dans le cas bi-infinis, nous devons demander explicitement l’uniforme récurrence afin d’exclure des mots tels que

... aaaaaabaaaaa ...

qui possèdent bien $n+1$ facteurs de longueur n , mais qui ne correspondent pas vraiment aux mots sturmiens infinis à droite.

Moralement, tout mot sturmien infini à droite est une « coupure » d’un mot sturmien bi-infinis de même pente. Un sturmien standard correspond à la coupure au « meilleur » endroit possible, tandis qu’un sturmien de Lyndon est la coupure au « pire » endroit possible. Nous pouvons donc penser, intuitivement, que les sturmiens bi-infinis ne souffriront pas des problèmes de « coupure » et devraient mieux se comporter vis-à-vis de la quasipériodicité.

Toutefois, le passage au bi-infinis pose problème pour la quasipériodicité. Le théorème 3.2 demeure vrai sur les mots bi-infinis, car sa preuve est entièrement locale et ne repose pas sur le fait que le mot ambiant soit infini à droite ou bi-infinis. Toutefois, l’observation que *toute quasipériode d’un mot est en particulier un préfixe de ce mot* est invalidée dans notre nouveau cadre, puisqu’il n’y a plus de préfixes à proprement parler. L’observation qu’il y a au maximum une quasipériode d’une longueur donnée repose sur la précédente : elle est donc également caduque. Il peut y avoir plusieurs quasipériodes d’une longueur donnée, et nous pouvons nous inspirer du théorème 3.2 pour les caractériser.

Théorème 4.14. *Soit \mathbf{w} un mot bi-infinis et q une quasipériode de \mathbf{w} .*

- *Un successeur de q est une quasipériode de \mathbf{w} si et seulement si q n’est **pas** spécial droit.*
- *Un prédécesseur de q est une quasipériode de \mathbf{w} si et seulement si q n’est **pas** spécial gauche.*

Démonstration. Nous prouvons uniquement la première affirmation, l'autre étant parfaitement symétrique.

Si q possède un unique successeur r , chaque occurrence de q est immédiatement suivie d'une occurrence de r qui recouvre donc \mathbf{w} . Réciproquement, soient a, b, c des lettres et q' le mot tels que $q = aq'$ et $q'b, q'c$ soient deux facteurs de \mathbf{w} différents. Si $q'b$ est une quasipériode de \mathbf{w} , alors il existe une occurrence de $q'c$ facteur d'un chevauchement de deux occurrences de $q'b$, et le lemme 3.1 montre que $b = c$: une contradiction avec $q'b \neq q'c$. \square

De façon équivalente au théorème 4.14, si q possède un unique successeur (respectivement prédécesseur), ce dernier est une quasipériode ; si q possède plusieurs successeurs (respectivement prédécesseurs), aucun d'entre eux n'est une quasipériode.

Le théorème 3.2 (qui continue à fonctionner pour les bi-infinis), combiné avec le théorème 4.14, généralise notre méthode d'étude des quasipériodes aux mots bi-infinis. Toutefois, la méthode s'avère bien moins pratique dans ce cas, puisque la détermination des *facteurs* spéciaux droits et carrés est moins aisée que la détermination des *préfixes* possédant ces propriétés. Il y a bien plus de facteurs que de préfixes, ce qui rend l'étude exhaustive manuelle très fastidieuse, et de plus il y a moins de techniques disponibles dans la littérature.

Le théorème 4.14 admet une interprétation en termes de graphe de Rauzy : les quasipériodes de longueur n d'un mot \mathbf{w} donné s'organisent en *chemins*, dont le premier sommet est spécial gauche (possède plusieurs arcs entrants) et le dernier sommet est spécial droit (possède plusieurs arcs sortants), sans autres sommets spéciaux, comme représenté sur la figure 4.3. Ce schéma est intéressant, car l'évolution aux ordres supérieurs (c'est-à-dire aux longueurs de facteurs supérieures) d'un chemin de quasipériodes dans le graphe de Rauzy aux ordres supérieurs est très prévisible : un chemin de longueur k à l'ordre n donnera un chemin de longueur $k - 1$ à l'ordre $n + 1$, et toutes les branches allant du sommet spécial droit au sommet spécial gauche sont allongées de 1 lors du passage de n à $n + 1$. À un moment donné, le chemin aura longueur 0, ce qui s'interprète combinatoirement.

Corollaire 4.15. *Soit \mathbf{w} un mot biinfini et q une quasipériode de \mathbf{w} ; alors il existe une quasipériode u bispéciale, de longueur minimale qui contient q comme facteur. Tout facteur r de u tel que $|q| \leq |r| \leq |u|$ est une quasipériode de \mathbf{w} .*

En étudiant les graphes de Rauzy de quelques exemples de mots multi-échelles, nous pourrions imaginer que tout mot quasipériodique possède au plus une quasipériode spéciale droite (et une quasipériode spéciale gauche) de chaque longueur. Pour obtenir la quasipériode bispéciale du corollaire précédent, il suffirait donc de partir de la quasipériode spéciale gauche de même longueur que q , puis de considérer ses successeurs jusqu'à obtenir une quasipériode qui est également spéciale droite. Toutefois, il est possible de produire des contre-exemples. Soit $q = ababa^2baba^2ba$ et $r = aba^2baba^2baba$. Le morphisme h défini par $h(a) = q$ et $h(b) = a^{-1}q$ permet de construire des mots dont q et r sont tous deux des quasipériodes spéciales droites. Par exemple, supposons que \mathbf{w} est le mot biinfini spécifié par $\mathbf{w}(n) = a$ si $n \leq 0$ et $\mathbf{w}(n) = b$ sinon ; alors $h(\mathbf{w})$ admet

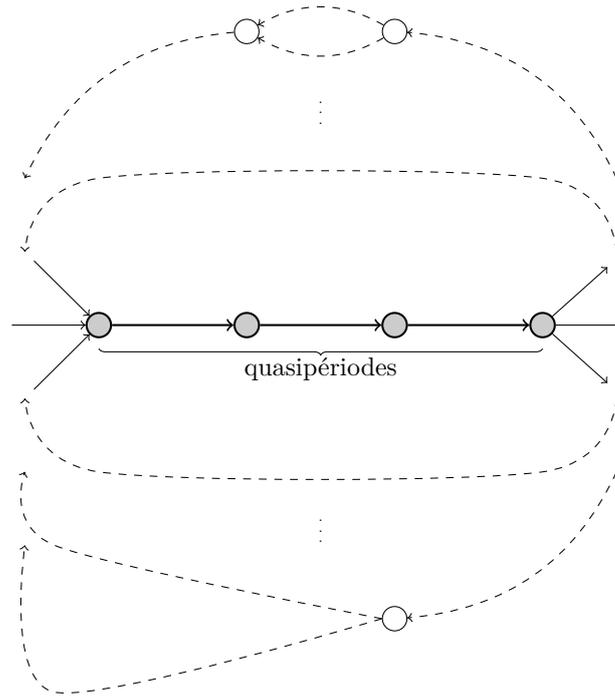


FIG. 4.3 : Portrait-robot du graphe de Rauzy d'un mot biinfini quasipériodique

q et r comme quasipériodes bispéciales.

$$\mathbf{w} = \dots aaaaaa \cdot bbbbbb \dots$$

$$h(\mathbf{w}) = {}^\omega q \cdot (a^{-1}q)^\omega$$

Par l'équation précédente, il est clair que \mathbf{w} est q -quasipériodique et spécial droit. Ci-après nous écrivons $h(\mathbf{w})$, où la barre $|$ marque la frontière entre ${}^\omega q$ et $(a^{-1}q)^\omega$ et les espaces marquent les limites des occurrences de q et de $a^{-1}q$.

$$\dots aba \ ababaababaaba \ ababaababaaba \ | \ babaababaaba \ babaababaaba \ ba \ \dots \quad (4.3)$$

Mettons maintenant en exergue les occurrences de r :

$$\dots \overbrace{aba \ ababaababaaba}^r \ \overbrace{ababaababaaba}^r \ | \ \overbrace{babaababaaba}^r \ \overbrace{babaababaaba}^r \ ba \ \dots \quad (4.4)$$

Ce comportement se poursuit à gauche et à droite, le mot r est donc quasipériode de \mathbf{w} . La première occurrence de r sur la figure est suivie de a , tandis que la troisième est suivie de b , ce dernier est donc spécial droit.

Conjecture. Soit \mathbf{w} un mot biinfini et deux quasipériodes q, r de même longueur. Les deux quasipériodes bispéciale minimales qui contiennent respectivement q et r sont de même longueur.

Si cette conjecture était démontrée, il deviendrait aisé de prouver un théorème analogue au corollaire 4.13 sur les mots bi-infinis, qui caractériserait tous les mots sturmiens.

All art is quite useless.

— Oscar Wilde, preface of *The picture of Dorian Gray*, 1890

Chapitre 5

Combinatoire des mots en deux dimensions

Dans les chapitres suivants, nous nous intéresserons à généraliser tous les résultats concernant la quasipériodicité au cas des mots infinis bidimensionnels. Le présent chapitre motive cette étude, donne les définitions utiles pour travailler en deux dimensions et présente quelques outils de combinatoire des mots 2D. Ces outils, qui nous seront utiles dans les chapitres ultérieurs, ont été développés au cours de la préparation cette thèse en collaboration avec d'autres auteurs.

5.1 Motivation : les pavages du plan

En 1961, Hao Wang [82] introduit les tuiles qui portent son nom : une *tuile de Wang* est un carré dont les côtés sont colorés. Les couleurs sont choisies dans un ensemble fini C , fixé *a priori*. Soit T est un ensemble de tuiles ; un *pavage* du plan par T est un recouvrement du plan par des copies d'éléments de T , sans rotation ni symétrie possible, de telle façon que deux côtés adjacents aient toujours la même couleur. Le *problème Domino* demande, étant donné en entrée un ensemble de tuiles T , s'il existe un pavage du plan par T . Wang réduit le problème de décidabilité des formules *AEA* au problème domino, conjecturant que ce dernier est décidable. Une forme *AEA* est une formule pouvant s'écrire $\forall x \exists y \forall z \Phi_{x,y,z}$, où $\Phi_{x,y,z}$ est une formule sans symboles de fonction et sans quantificateurs.

Wang propose l'algorithme suivant.

Algorithme de Wang.

1. $n \leftarrow 1$
2. Essayer de paver un carré de taille $n \times n$ de toutes les façons possibles
3. Si on peut paver le carré de telle façon que les couleurs des côtés opposés sont égales, alors renvoyer VRAI
4. Si on ne peut pas paver le carré, alors renvoyer FAUX
5. Sinon, $n \leftarrow n + 1$ et recommencer à l'étape 2.

En effet, si un jeu de tuiles T peut paver un carré de telle façon que les côtés opposés

sont égaux, alors il est possible de paver le plan de façon périodique en répétant ce carré à l'infini. À l'inverse, s'il existe une taille n telle que l'on ne peut pas paver un carré $n \times n$, il est clair que le jeu de tuiles ne peut pas paver le plan.

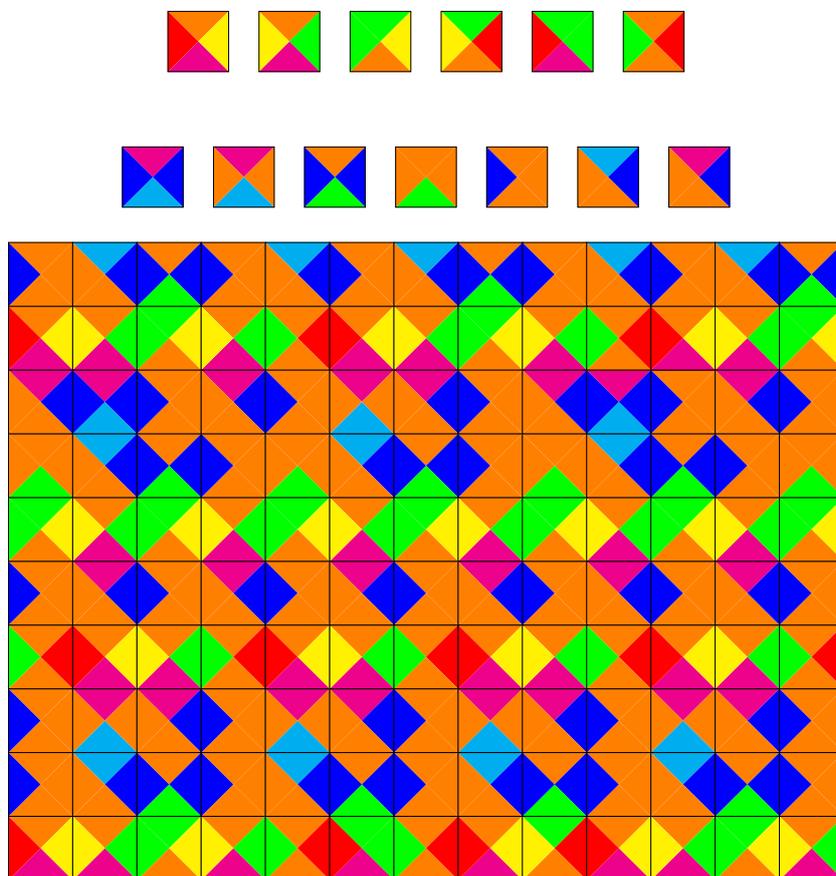


FIG. 5.1 : Un exemple de jeu de tuiles et de pavage

Le seul cas où cet algorithme échoue est le suivant : le jeu de tuile T permet effectivement de paver le plan, mais *jamais* de façon périodique. Sans connaissance particulière du problème, ce cas de figure semble improbable. En outre, le théorème de compacité semble intuitivement aller dans le sens de la conjecture de Wang.

Théorème 5.1 (Compacité, Wang, 1961). *Un jeu de tuile T pave le plan si et seulement s'il pave toute partie finie du plan.*

Démonstration. Si T pave le plan, il pave clairement toute partie finie du plan par restriction.

Réciproquement, si T pave toute partie finie du plan, il pave en particulier tous les carrés $(2n + 1) \times (2n + 1)$ centrés sur l'origine ; appelons $(c_n)_{n \in \mathbb{N}}$ cette suite de pavages partiels. Soit $(u_n)_{n \in \mathbb{N}}$ une suite de tous les couples d'entiers relatifs, autrement dit $n \mapsto u_n$ est une bijection de \mathbb{N} vers \mathbb{Z}^2 , telle que pour tout n si $u_n = (x, y)$ alors $|x| \leq n$ et $|y| \leq n$. Nous pouvons par exemple prendre une spirale démarrante à l'origine. Comme T est fini et (c_n) infini, il existe une infinité d'éléments de c_n portant la même

tuile en position u_0 ; appelons $(c_n^{(0)})$ cette sous-suite. De la même façon, il existe une infinité d'éléments de $(c_n^{(0)})$ s'accordant sur la tuile en position u_1 ; appelons $(c_n^{(1)})$ cette suite, etc. Par le lemme de König (tout arbre infini de degré fini possède une branche infinie, cf. [54]), nous pouvons itérer cet argument infiniment et ainsi définir un pavage valide du plan. Comme nous extrayons notre pavage de configurations qui respectent les règles locales de coloration, il les respecte également. \square

Toutefois en 1964, Robert Berger, étudiant de Wang, démontre [6] que le problème Domino est indécidable en embarquant une machine de Turing dans les tuiles. En conséquence, l'algorithme de Wang ne termine pas sur toutes les entrées, et il existe donc bel et bien des jeux de tuiles qui pavent le plan, mais jamais de façon périodique. Nous les appelons des *jeux de tuiles apériodiques*. Depuis 1964, de nombreux auteurs ont produit de tels jeux de tuiles, de plus en plus petits ; les plus célèbres sont ceux de Raphael Robinson [76] et Roger Penrose [74]. le jeu de tuiles apériodique minimal, à 11 tuiles, fut découvert en 2015 par Emmanuel Jeandel et Michael Rao [49], et une vérification exhaustive par ordinateur leur a permis de montrer qu'il n'en n'existait aucun plus petit.

L'engouement pour les jeux de tuiles apériodiques trouve ses racines dans la question suivante : *comment des règles locales peuvent forcer une propriété globale ?* [24] Le fait que les règles locales à la Wang permettent de forcer l'apériodicité fut une surprise. Il est possible de forcer d'autres propriétés, notamment en embarquant des calculs (une machine de Turing) dans les tuiles [25] ; on obtient ainsi des jeux de tuiles qui garantissent une complexité de Kolmogorov élevée pour leurs pavages du plan [23], qui permettent de corriger les erreurs [26], ou encore qui garantissent une complexité en facteurs élevée [19]. En dehors de cette question, les tuiles de Wang ont des applications en combinatoire [18], en infographie [55], en logique [82] et en théorie des systèmes dynamiques discrets. En effet, il est bien connu que les systèmes symboliques de types finis sont équivalents aux jeux de tuiles de Wang ; voir par exemple [77].

Certains jeux de tuiles apériodiques, tel que le jeu à 13 tuiles de Kari-Culik [16], sont encore mal compris. Ce dernier n'est pas minimal (au sens des systèmes dynamiques) et admet une entropie topologique strictement positive [22]. Toutefois, tous les pavages qu'il permet de construire possèdent, intuitivement, une forte structure, qui doit se traduire par des propriétés de régularité, de symétrie. C'est l'une des motivations pour l'étude de la quasipériodicité en deux dimensions : chercher de nouvelles propriétés de régularité afin d'étudier ces pavages.

5.2 Définitions élémentaires autour des mots bidimensionnels

Soit Σ un alphabet fini. Pour un ensemble D quelconque, une *configuration* (ou parfois une *coloration*) de D est une fonction de D vers Σ . Ainsi, les mots infinis sont les configurations de \mathbb{N} , et les mots bi-infinis les configurations de \mathbb{Z} . Une configuration de \mathbb{Z}^2 est, en quelque sorte, un mot biinfini à deux dimensions : une grille infinie dont chaque case comporte une lettre de Σ . Le terme « mot bidimensionnel » étant assez

lourd, nous convenons qu'une *configuration* dont nous ne précisons pas le domaine sera toujours une configuration de \mathbb{Z}^2 . En outre, notre alphabet favori pour ce chapitre et les suivants sera $\Sigma = \{\square, \blacksquare\}$, ce qui présente l'avantage de libérer des lettres pour les notations tout en donnant des figures plus claires.

Si i, j sont des entiers relatifs et h, w des entiers naturels, l'ensemble $R = \{i, \dots, i + h - 1\} \times \{j, \dots, j + w - 1\}$ est appelé un *rectangle*. Nous le notons aussi $\text{rect}(i, j, h, w)$. Le couple (h, w) est la *taille* (ou les *dimensions*) du rectangle, h sa *hauteur* et w sa *largeur*. Nous posons $\text{height}(R) = h$ et $\text{width}(R) = w$. L'*aire* de R , notée $|R|$, est la quantité $h \times w$. Tout rectangle de hauteur ou de largeur nulle est l'ensemble vide. Les points $(i, j), (i + h - 1, j), (i + h - 1, j + w - 1), (i, j + w - 1)$ sont les *coins* du rectangle.

Un *bloc* est une configuration d'un rectangle ; les blocs sont les analogues des mots finis. En particulier, la configuration de domaine vide est un bloc, que nous appelons *le bloc vide*. La taille, la hauteur, la largeur et les coins d'un bloc sont ceux de son domaine. Nous considérons deux blocs identiques à translation près comme égaux ; lorsque nous voulons les différencier, nous parlons d'*occurrences*. Par exemple, les configurations de $\{0, 1, 2\}$ et de $\{7, 8, 9\}$ sur l'alphabet $\{\square, \blacksquare\}$ définies par

$$\begin{aligned} b : 0 &\mapsto \square, 1 \mapsto \blacksquare, 2 \mapsto \square \\ b' : 7 &\mapsto \square, 8 \mapsto \blacksquare, 9 \mapsto \square \end{aligned}$$

sont le même bloc, mais pas la même occurrence. Ainsi, la configuration représentée sur la figure 5.2 comporte plusieurs occurrences du bloc $\blacksquare\square\blacksquare$.

Les *coordonnées* d'une occurrence sont les coordonnées de son coin inférieur gauche. Si b est un bloc de taille $m \times n$ qui possède une occurrence aux coordonnées (i, j) dans \mathbf{w} , nous disons que b *couvre* les positions de $\text{rect}(i, j, m, n)$ dans \mathbf{w} . Nous notons $\mathbf{w}[i, j]$ la lettre aux coordonnées i, j dans \mathbf{w} et $\mathbf{w}[(i, j), \dots, (i + h - 1, j + w - 1)]$ la restriction de \mathbf{w} à $\text{rect}(i, j, h, w)$. Attention : avec nos définitions, les ordonnées croissantes vont vers le haut (s'il n'y a pas de coordonnées négatives, l'origine est en bas à gauche).

Nous notons $\Sigma^{n \times m}$ l'ensemble des blocs de taille (n, m) , et $\Sigma^{\geq n \times m}$ l'ensemble des blocs dont la hauteur est supérieure ou égale à n et la largeur à m , puis enfin Σ^{**} l'ensemble des blocs sur Σ .

Soient b, b' deux blocs de tailles respectives (h, w) et (h', w') . Si $h = h'$, alors la *concaténation horizontale* de b et b' , notée $b \oplus b'$ est le bloc de domaine $\text{rect}(0, 0, h, w + w')$ comportant une occurrence de b en $(0, 0)$ et une occurrence de b' en $(0, w)$. Symétriquement, si $w = w'$, alors la *concaténation verticale* de b et b' , notée $b \ominus b'$, est le bloc de domaine $\text{rect}(0, 0, h + h', w)$ comportant une occurrence de b en $(0, 0)$ et une occurrence de b' en $(h, 0)$; en d'autres termes, b est *en dessous* de b' . Pour tout bloc b et tous entiers m, n , nous définissons $b^{m \times n}$ comme la concaténation de $m \times n$ copies de b . Formellement, $b^{1 \times 1} = b$ et $b^{(m+1) \times 1} = b^{m \times 1} \oplus b$ et $b^{m \times (n+1)} = b^{m \times n} \oplus b^{m \times 1}$. Si un bloc b peut s'écrire $a^{m \times n}$, avec a un bloc et m, n des entiers naturels, alors a est une *racine* de b . Si b n'admet aucune racine, hormis lui-même, nous disons que b est *primitif*.

Soient q et u des blocs finis, différents, avec q non vide. Nous disons que u est un *bord diagonal* de q lorsque u apparaît dans deux coins opposés de q . Dans cette thèse, nous emploierons le terme *bord* au lieu de *bord diagonal* ; toutefois, certains auteurs utilisent la terminologie *bord* pour les facteurs qui apparaissent dans les quatre coins

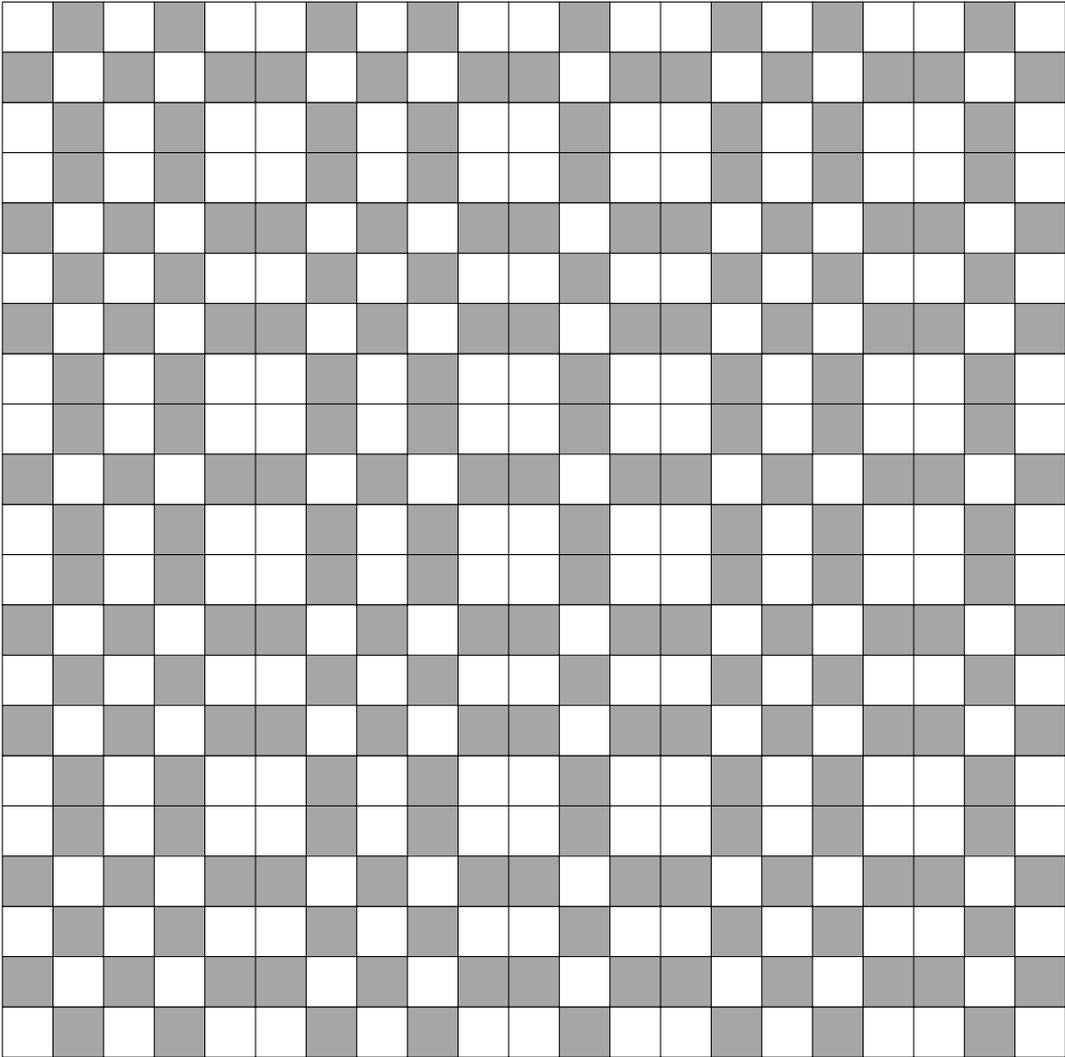


FIG. 5.2 : Exemple de configuration de \mathbb{Z}^2

à la fois [15]. Si u possède la même largeur que q , nous dirons qu'il s'agit d'un bord *pleine largeur*; s'il possède la même hauteur, c'est un bord *pleine hauteur*. La figure 5.3 illustre ces notions.

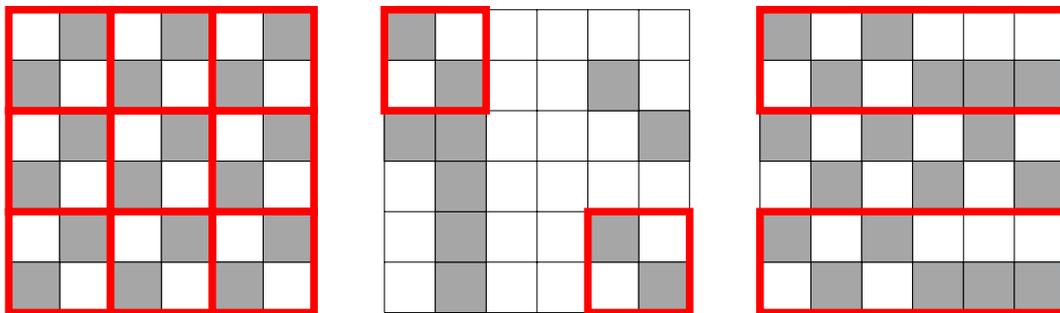


FIG. 5.3 : Exemple de racine, de bord et de bord pleine largeur

Soit \mathbf{w} une configuration et b un bloc; b est un *facteur* de \mathbf{w} si et seulement s'il apparaît dans \mathbf{w} . Le *langage* de \mathbf{w} , noté $\mathcal{L}(\mathbf{w})$, est l'ensemble de ses facteurs. Pour tous entiers naturels n, m , nous posons $\mathcal{L}_{n \times m}(\mathbf{w}) = \mathcal{L}(\mathbf{w}) \cap \Sigma^{n \times m}$ et $\mathcal{L}_{\geq n \times m}(\mathbf{w}) = \mathcal{L}(\mathbf{w}) \cap \Sigma^{\geq n \times m}$. La *complexité en facteurs* de \mathbf{w} (ou plus simplement sa *complexité*) est la fonction définie par $P_{\mathbf{w}}(m, n) = \text{card } \mathcal{L}_{m \times n}(\mathbf{w})$. L'entropie topologique d'une configuration se définit par

$$\liminf_{n \rightarrow \infty} \frac{\log_2 P_{\mathbf{w}}(n, n)}{n^2}.$$

En effet, il y a maintenant $|\Sigma|^{n^2}$ blocs de taille $n \times n$ possibles, nous devons donc regarder la base d'une double exponentielle. Cette limite existe toujours, comme nous le montrerons dans la section 5.3.

Soit \mathbf{w} une configuration. Si (x, y) est un vecteur d'entiers non nul tel que, pour tous i, j on a $\mathbf{w}[i, j] = \mathbf{w}[i + x, j + y]$, alors (x, y) est appelé un *vecteur de périodicité* de \mathbf{w} . Une configuration admettant au moins un vecteur de périodicité est appelée *périodique*. Dans le cas unidimensionnel, périodicité et complexité en facteurs sont reliées par le théorème 2.2 de Morse-Hedlund. En deux dimensions, toutefois, les choses sont moins simples [72].

Conjecture 5.2 (Nivat, 1997). *Soit \mathbf{w} une configuration. S'il existe des entiers m, n tels que $P_{\mathbf{w}}(m, n) \leq mn$, alors \mathbf{w} est périodique.*

La réciproque est fautive : il existe des mots périodiques ayant une complexité plus grande que $P_{\mathbf{w}}(m, n) = mn$. Il existe des contre-exemples à la conjecture en dimension supérieure (configurations de \mathbb{Z}^d pour $d \geq 3$) [79]. La conjecture de Nivat fait l'objet de travaux de recherche actifs [51]; le résultat suivant est le dernier d'une longue série de bornes de plus en plus proches de la conjecture [17].

Théorème 5.3 (Cyr, Kra, 2012). *Soit \mathbf{w} une configuration. S'il existe des entiers m, n tels que $P_{\mathbf{w}}(m, n) \leq mn/2$, alors \mathbf{w} est périodique.*

Toutefois, les techniques utilisées pour prouver ce théorème sont assez lourdes, sans commune mesure avec la démonstration du théorème de Morse-Hedlund. Notons que la conjecture de Nivat était l'une des motivations pour développer la quasipériodicité sur les configurations, dans l'espoir que cette étude nous donne de nouveaux outils combinatoires pour aborder cette question.

De même que les mots, une configuration peut être *régulière* (au sens intuitif) de bien des façons, la périodicité n'étant qu'une notion parmi d'autres. Soit \mathbf{w} une configuration. Si, pour tout entier m , il existe un entier n tel chaque élément de $\mathcal{L}_{n \times n}(\mathbf{w})$ contienne au moins une occurrence de chaque élément de $\mathcal{L}_{m \times m}(\mathbf{w})$, nous disons que \mathbf{w} est *uniformément récurrente*, ou parfois *minimale* (en référence aux systèmes symboliques, cf. section 4.1). La fonction qui, à tout entier m , associe le plus petit n correspondant, est appelée la *fonction de récurrence* de \mathbf{w} . Les liens entre quasipériodicité et uniforme récurrence sont moins clairs en deux dimensions qu'en une.

Enfin, si u et v sont des blocs, nous notons $|u|_v$ le nombre d'occurrences de v dans u et $f_v(u)$ la *fréquence* de v dans u , c'est-à-dire la quantité $|u|_v/|u|$. Dans le cas d'une configuration \mathbf{w} , la fréquence d'un bloc v se définit par

$$f_v(\mathbf{w}) = \lim_{n \rightarrow \infty} \frac{|\mathbf{w}[(-n, -n) \dots (+n, +n)]|_v}{(2n + 1)^2}$$

si la limite existe. Lorsque tous les facteurs d'une configuration possèdent une fréquence bien définie, nous disons que la configuration a des *fréquences qui convergent*, ou plus simplement qu'elle a des *fréquences*. De la même façon que dans le cas unidimensionnel, les fréquences des facteurs d'une configuration ne convergent pas toujours.

5.3 L'entropie bidimensionnelle est bien définie

Nous avons affirmé que l'entropie topologique d'une configuration est bien définie. Toutefois, contrairement au cas des mots infinis, il n'est plus possible de conclure simplement à partir du lemme de Fekete. La preuve habituelle passe donc par la théorie des systèmes dynamiques, qui permet de démontrer que l'entropie de tout système dynamique converge (et qu'une configuration de \mathbb{Z}^2 est bien un système dynamique). Toutefois, il existe un argument plus simple dans le cas des mots bidimensionnels; il fut redéveloppé par Julien Destombes, Andrei Romashchenko, Anaël Grandjean et l'auteur.

Proposition 5.4 (Folklore). *Soit \mathbf{w} une configuration de \mathbb{Z}^d et $P_{\mathbf{w}}(n)$ le nombre de facteurs de \mathbf{w} (c'est-à-dire de configurations différentes de l'hypercube $\{0, \dots, n-1\}^d$ qui apparaissent dans \mathbf{w}). La limite*

$$\lim_{n \rightarrow \infty} \frac{\log_2 P_{\mathbf{w}}(n)}{n^d}$$

existe.

Appelons h_n la quantité $\frac{\log_2 P_{\mathbf{w}}(n)}{n^d}$.

Étape 1. *La limite de h_{2^n} existe.* Un hypercube de taille $2n$ est constitué de 2^d hypercubes de taille n , nous avons donc $P_{\mathbf{w}}(2n) \leq (P_{\mathbf{w}}(n))^{2^d}$. À partir de là,

$$\frac{\log_2 P_{\mathbf{w}}(2n)}{(2n)^d} \leq \frac{2^d \log_2 P_{\mathbf{w}}(n)}{(2n)^d}$$

et donc $h_{2n} \leq h_n$, ce qui permet de conclure par récurrence que $h_{2^{n+1}} \leq h_{2^n}$. La suite h_{2^n} est donc décroissante et admet 0 pour borne inférieure, nous en déduisons qu'elle converge vers un certain réel e .

Étape 2. *Nous avons $\limsup_{n \rightarrow \infty} h_n \leq e$.* Soit $\varepsilon > 0$; nous montrons qu'il existe un entier N tel que, pour tout $n > N$, nous avons $h_n < e + \varepsilon$. Soit un entier ℓ que $h_{2^\ell} < e + \varepsilon/2$, un entier $n \gg 2^\ell$ et k l'entier tel que $k2^\ell < n < (k+1)2^\ell$. En découpant l'hypercube de taille $(k+1)2^\ell$ en $(k+1)^d$ hypercubes de taille 2^ℓ , nous bornons : $P_{\mathbf{w}}(n) \leq (P_{\mathbf{w}}(2^\ell))^{(k+1)^d}$. Ainsi nous avons

$$\frac{\log_2 P_{\mathbf{w}}(n)}{n^d} \leq (k+1)^d \times \frac{\log_2 P_{\mathbf{w}}(2^\ell)}{(2^\ell)^d} \times \frac{(2^\ell)^d}{n^d}$$

d'où $h_n \leq \left(\frac{k+1}{k}\right)^d h_{2^\ell}$. Comme $\left(\frac{k+1}{k}\right)^d$ tend vers 1 et que h_{2^ℓ} est une constante, nous avons $h_n > h_{2^\ell} + \varepsilon/2 < e + \varepsilon$ pour tout n assez grand.

Étape 3. *Nous avons $\liminf_{n \rightarrow \infty} h_n \geq e$.* Soit $\varepsilon > 0$; nous montrons que, pour tout entier n , nous avons $h_n > e - \varepsilon$. Soit L tel que, pour tout $\ell > L$, nous avons $h_{2^\ell} > e - \varepsilon/2$. Prenons un $\ell > L$ tel que $2^\ell \gg n$ et k un entier tel que $kn \leq 2^\ell < (k+1)n$. En découpant l'hypercube de taille $(k+1)n$ en $(k+1)^d$ hypercubes de taille n , on borne $P_{\mathbf{w}}(2^\ell) \leq (P_{\mathbf{w}}(n))^{(k+1)^d}$. Nous pouvons donc calculer

$$\frac{\log_2 P_{\mathbf{w}}(2^\ell)}{(2^\ell)^d} \leq (k+1)^d \times \frac{\log_2 P_{\mathbf{w}}(n)}{n^d} \times \frac{n^d}{(2^\ell)^d}$$

d'où nous tirons $h_{2^\ell} \leq \left(\frac{k+1}{k}\right)^d h_n$. Comme $\left(\frac{k+1}{k}\right)^d$ tend vers 1 et que h_{2^ℓ} est une constante, nous avons $h_n > h_{2^\ell} - \varepsilon/2 > e - \varepsilon$.

Conclusion. Nous avons $\liminf h_n = \limsup h_n = e$, donc la limite de h_n est bien définie.

5.4 Un peu de combinatoire en deux dimensions

Nous généralisons maintenant aux blocs les résultats élémentaires présentés dans la section 1.1. Ces résultats constituent les briques de base de nombreux raisonnements ultérieurs, y compris au-delà du cadre de cette thèse : ils ont été en partie développés par Jeffrey Shallit et Taylor J. Smith, qui n'avaient pas de motivations en lien avec la quasipériodicité. L'autre partie fut développée par Gwenaël Richomme et l'auteur pour les besoins des démonstrations présentées dans les chapitres suivants.

Théorème 5.5. *Soient a et b deux blocs non vides. Les affirmations suivantes sont équivalentes :*

1. *il existe des entiers $m_1, m_2, n_1, n_2 > 0$ tels que $a^{m_1 \times n_1} = b^{m_2 \times n_2}$;*
2. *il existe un bloc fini, non vide c et des entiers $m_1, m_2, n_1, n_2 > 0$ tels que $a = c^{m_1 \times n_1}$ et $b = c^{m_2 \times n_2}$;*
3. *il existe des entiers $m_1, m_2, n_1, n_2 > 0$ tels que $a^{m_1 \times n_1} \circ b^{m_2 \times n_2} = b^{m_2 \times n_2} \circ a^{m_1 \times n_1}$, où $\circ \in \{\ominus, \oplus\}$.*

Démonstration. Nous démontrons les trois implications séparément.

(1 \implies 2). Soient a et b des blocs dont les tailles respectives sont $h_1 \times w_1$ et $h_2 \times w_2$, tels que $a^{m_1 \times n_1} = b^{m_2 \times n_2}$. Nous avons donc $m_1 h_1 = m_2 h_2$ et $n_1 w_1 = n_2 w_2$. Soit $a' = a^{m_1 \times 1}$ et $b' = b^{m_2 \times 1}$, de telle façon que $a'^{1 \times n_1} = b'^{1 \times n_2}$. En considérant a' et b' comme des mots unidimensionnels sur l'alphabet $\Sigma^{m_1 h_1 \times 1}$ avec la concaténation horizontale, le théorème 1.1 montre qu'il existe un bloc r et des entiers n_3, n_4 tels que $a' = r^{1 \times n_3}$ et $b' = r^{1 \times n_4}$. Soit $\rho = \text{width}(r)$ et posons $a'' = a[(0, 0) \dots (h_1 - 1, \rho - 1)]$ et $b'' = b[(0, 0) \dots (h_2 - 1, \rho - 1)]$, de telle façon que $a = a''^{1 \times n_3}$ et $b = b''^{1 \times n_4}$. En considérant les ρ premières colonnes de a' et b' , nous observons que $a''^{m_1 \times 1} = b''^{m_2 \times 1}$. Par le théorème 1.1 de nouveau, il existe un mot c et des entiers m_3, m_4 tels que $a'' = c^{m_3 \times 1}$ et $b'' = c^{m_4 \times 1}$, d'où $a = c^{m_3 \times n_3}$ et $b = c^{m_4 \times n_4}$.

(2 \implies 3). Si $a = c^{m_1 \times n_1}$ et $b = c^{m_2 \times n_2}$, alors nous pouvons écrire

$$\begin{aligned} a^{m_2 \times n_2} \circ b^{m_1 \times n_1} &= c^{m_1 m_2 \times n_1 n_2} \circ c^{m_2 m_1 \times n_2 n_1} \\ &= c^{m_2 m_1 \times n_2 n_1} \circ c^{m_1 m_2 \times n_1 n_2} \\ &= b^{m_1 \times n_1} \circ a^{m_2 \times n_2}. \end{aligned}$$

(3 \implies 1). Sans perte de généralité, supposons que \circ désigne la concaténation horizontale. Posons (h_1, w_1) les dimensions de a et (h_2, w_2) celles de b . Par hypothèse, il existe des entiers m_1, m_2, n_1, n_2 tels que $a^{m_1 \times n_1} \circ b^{m_2 \times n_2} = b^{m_2 \times n_2} \circ a^{m_1 \times n_1}$. Donc, pour tous entiers i, j , nous avons

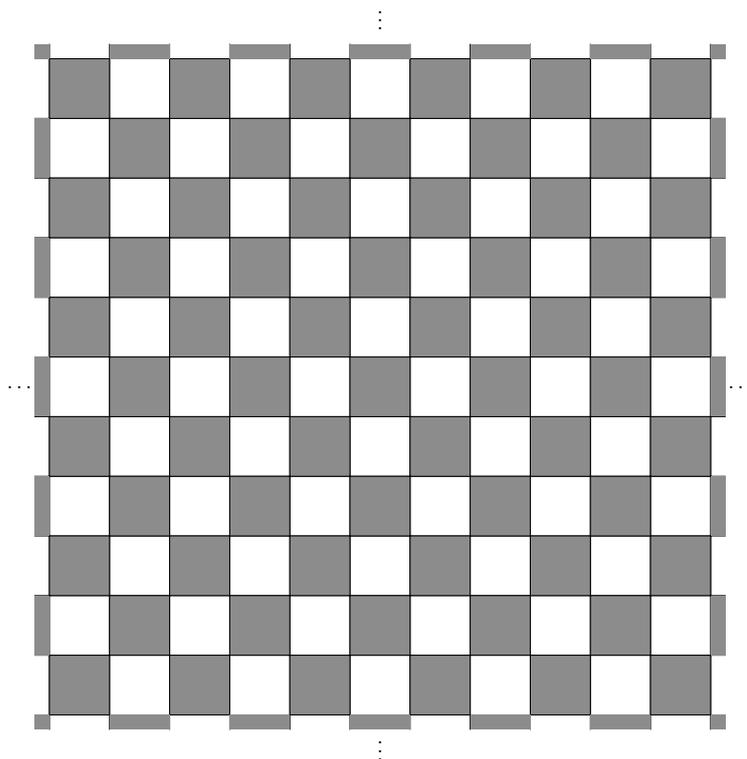
$$a^{m_1 \times i n_1} \circ b^{m_2 \times j n_2} = b^{m_2 \times j n_2} \circ a^{m_1 \times i n_1}.$$

Si $i = w_2 n_2$ et $j = w_1 n_1$, alors $i w_1 n_1 = j w_2 n_2 = \text{width}(a^{m_1 \times i n_1}) = \text{width}(b^{m_2 \times j n_2})$ et le résultat suit. \square

L'implication (1 \implies 2) se reformule comme suit.

Corollaire 5.6. *Si un bloc possède deux racines distinctes r_1 et r_2 , alors il existe un bloc r qui est racine de r_1 et de r_2 .*

Tout bloc b possède donc une unique racine minimale pour la relation « est racine de ». Appelons r cette racine ; alors r est primitive, sans quoi la racine primitive de r serait aussi racine de b , et r ne serait pas minimale. Réciproquement, par le corollaire précédent, b ne peut admettre deux racines primitives distinctes. Nous parlerons donc de *la* racine primitive d'un bloc. Toutefois, l'unicité de la racine primitive ne se

FIG. 5.4 : Configuration périodique de \mathbb{Z}^2

généralise pas aux configurations de \mathbb{Z}^2 : comme l'a montré Nicolas Bacquey [5], la configuration périodique $(\begin{smallmatrix} \square & \blacksquare \\ \blacksquare & \square \end{smallmatrix})^{\mathbb{Z}^2}$, représentée sur la figure 5.4, possède deux racines primitives : $\begin{smallmatrix} \square & \blacksquare \\ \blacksquare & \square \end{smallmatrix}$ et $\begin{smallmatrix} \blacksquare & \square \\ \square & \blacksquare \end{smallmatrix}$.

Le résultat suivant généralise la proposition 1.2 (qui caractérise la primitivité des mots finis). Il fut développé par J. Shallit et T. J. Smith.

Lemme 5.7 (Shallit, Smith, 2017 [40]). *Soit a un bloc de taille $h \times w$; appelons ℓ_i la racine primitive de la $i^{\text{ème}}$ ligne de a et c_j la racine primitive de sa $j^{\text{ème}}$ colonne. Si (p, q) est la taille de la racine primitive de a , alors $p = \text{ppcm}(|c_0|, \dots, |c_{w-1}|)$ et $q = \text{ppcm}(|\ell_0|, \dots, |\ell_{h-1}|)$.*

Démonstration. Soit r la racine primitive de a , le couple (p, q) sa taille, et appelons ℓ, c les quantités $\text{ppcm}(|\ell_0|, \dots, |\ell_{h-1}|)$, $\text{ppcm}(|c_0|, \dots, |c_{w-1}|)$. Pour tout $i \in \{0, \dots, h-1\}$, la $i^{\text{ème}}$ ligne de a admet q comme période, donc $|\ell_i|$ divise q . *A fortiori*, ℓ divise q . Si $\ell < q$, alors $a[(0, 0) \dots (h-1, \ell-1)]$ serait une racine de r , qui ne serait donc pas primitif : contradiction. Nous avons donc bien $q = \ell$. L'équation $p = c$ se montre de façon analogue. \square

Ce lemme induit également un algorithme pour tester la primitivité (et déterminer la racine primitive) d'un bloc en temps linéaire : en effet, le calcul de la racine primitive de chaque ligne et colonne se fait en temps linéaire, de même que le calcul du plus petit multiple commun.

La remarque suivante peut être vue comme une version bidimensionnelle du théorème 1.4 de Fine et Wilf.

Remarque 5.8. Si une configuration \mathbf{w} possède deux vecteurs de périodicité non colinéaires, alors elle en possède un horizontal et un vertical.

Démonstration. Si \vec{u} et \vec{v} sont deux vecteurs de périodicité de \mathbf{w} , ainsi en est-il de toute combinaison linéaire (à coefficients entiers) de \vec{u} et \vec{v} . Comme ces derniers sont non nuls et non colinéaires, il est possible d'en trouver une combinaison linéaire qui forme un vecteur vertical, et une autre qui forme un vecteur horizontal. \square

Une *figure* est une configuration d'une partie finie de \mathbb{Z}^2 . Son domaine n'est pas nécessairement convexe, ni même connexe. Les blocs sont des cas particuliers de figures. Soit Q un ensemble de blocs, f une figure et $R = \{r_1, \dots, r_k\}$ un ensemble de rectangles ; alors R est un *recouvrement* de f par Q si et seulement si R est une partition du domaine de f telle que pour tout i , la restriction de f à r_i est un élément de Q . La figure 5.5 montre deux recouvrements d'une figure par trois blocs. Le résultat suivant fut montré de façon indépendante par Shallit et Smith d'une part, et par Richomme et l'auteur d'autre part.

Théorème 5.9. *Soit f une figure et $Q = \{a, b\}$ un ensemble de deux blocs. Si f admet deux recouvrements différents par Q , alors a et b admettent une racine commune c .*

Démonstration. Soit f une figure qui possède deux recouvrements différents U et V par $\{a, b\}$. Supposons, en vue d'une contradiction, que a et b n'ont aucune racine commune. Sans perte de généralité, nous pouvons supposer que f est d'aire minimale (card(dom(f)) est minimal), puis que la quantité $|a| + |b|$ est minimale pour cette aire de f .

Considérons (i, j) l'élément de dom(f) minimisant i , et parmi eux celui qui minimise j . En d'autres termes (i, j) est position la plus en bas, et parmi elles la position la plus à gauche de dom(f). Cette position doit être couverte par a dans l'un des recouvrements et par b dans l'autre ; si tel n'était pas le cas, nous pourrions retirer cette occurrence de a (ou b) à f en continuant à satisfaire nos conditions, donc f ne serait pas d'aire minimale.

Quitte à permuter les rôles de a et de b , nous pouvons considérer que $\text{height}(a) \geq \text{height}(b) = h$ et appeler a' la restriction de a à ses h lignes inférieures. Soit w le plus grand entier tel que $\text{rect}(i, j, h, w) \subseteq \text{dom}(f)$. Posons $L = \text{rect}(i, j, h, w)$. Étant donné que f possède deux recouvrements différents par $\{a, b\}$, la restriction de f à L possède deux recouvrements par a' et b , qui diffèrent à la position (i, j) . En voyant $f(L)$, a' et b comme des mots unidimensionnels sur l'alphabet $\Sigma^{h \times 1}$, le théorème 1.1 de Lyndon-Schutzenberger montre qu'il existe un bloc c qui est racine de a' et b . Comme a et b n'ont aucune racine commune, c n'est pas racine de a ; mais alors a et b se décomposent sur $\{a', c\}$, et nous déduisons de U et V deux recouvrements différents de f par $\{a', c\}$. Nous avons $|c| \leq |b|$ et $|a'| < |a|$ (sans quoi $\text{height}(a) = h$ et $a' = a$, donc c serait racine de a), ce qui montre que la quantité $|a| + |b|$ n'est pas minimale : une contradiction. \square

Ce théorème devient faux si Q possède trois éléments ou plus, comme observé dans [44]. La figure 5.5 donne un contre-exemple : elle montre une configuration d'un domaine fini, recouvert de deux façon différentes par trois blocs qui ne partagent aucune racine commune.

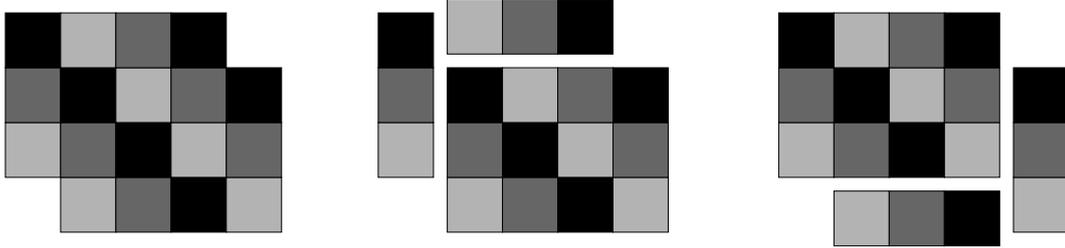


FIG. 5.5 : Contre-exemple au théorème 5.9 dans le cas $|Q| > 2$, voir [5]

NOTHING EASIER THAN GENERALIZATION

— Jean-Yves Béziau, 2011 [11]

Chapitre 6

Quasipériodicité en deux dimensions

Nous souhaitons maintenant étudier la quasipériodicité des configurations de \mathbb{Z}^2 . La première étape, en toute logique, est d'établir un théorème de forme normale. Celui-ci permet de comprendre la structure des mots quasipériodiques et de créer divers exemples intéressants, notamment pour montrer l'indépendance de la quasipériodicité avec diverses propriétés de régularité.

Toutefois, la combinatoire des configurations de \mathbb{Z}^2 est plus complexe que celle de \mathbb{Z} , et nous n'avons à ce jour aucun théorème de forme normale en deux dimensions. Cela tient au fait que les quasipériodes n'ont aucune raison d'être *alignées* sur un réseau quelconque. Il est donc difficile, étant donnée une occurrence d'une quasipériode, de définir l'occurrence au-dessus et l'occurrence à sa droite de façon unique. En effet, dans le cas général, aucune définition naturelle n'est satisfaisante : soit les opérations *au-dessus* et *à droite* ne commutent pas, soit les contraintes que doit respecter une configuration pour représenter une forme normale sont telles qu'il est tout aussi difficile de travailler sur la configuration quasipériodique directement.

Nous allons donc étudier les propriétés de la quasipériodicité sur les configurations de \mathbb{Z}^2 sans passer par une forme normale. Nous cherchons, comme dans le cas unidimensionnel, à construire des exemples de configurations quasipériodiques irrégulières, par exemple avec une haute entropie. Toutefois, sans théorème de forme normale, nous sommes contraints de travailler avec des *gadgets*, dont la preuve de bon fonctionnement nécessite un certain travail combinatoire.

Soient q et u des blocs. Nous disons que q est une *quasipériode* de u si et seulement si toute position de u est couverte par une occurrence de q . La même définition s'étend si u est une configuration. Dans tous les cas, nous ne considérons que des quasipériodes rectangulaires.

Il existe un algorithme en temps linéaire qui calcule toutes les quasipériodes carrées d'un bloc carré [15]. Nous allons toutefois nous intéresser ici aux configurations, qui ont un domaine infini, et donc nous éloigner des aspects algorithmiques. Notons tout de même que les questions de quasipériodicité sur les blocs sont en lien avec des questions de théorie des codes en deux dimensions ; voir par exemple [2] et ses références.

6.1 Quasipériodes forçant la périodicité

Nous nous intéressons à la métaquestion suivante : si q est un bloc possédant une certaine propriété X , toute configuration quasipériodique possède-t-elle une propriété Y ? Réciproquement, nous pouvons fixer Y et nous demander s'il existe une propriété X la forçant. Voici un exemple de telle question : quelles sont les quasipériodes qui forcent leurs configurations quasipériodiques à être périodiques? Si a est une lettre, l'ensemble des configurations a -quasipériodiques se réduit au singleton $\{a^{\mathbb{Z}^2}\}$, dont l'étude ne présente pas beaucoup d'intérêt. Nous souhaitons donc exclure de telles quasipériodes de la suite de notre étude, et à cette fin nous devons les caractériser.

Rappelons que, dans le cas unidimensionnel, il existe un mot apériodique q -quasipériodique si et seulement si la racine primitive de q possède un bord non vide, comme l'énonce la remarque 1.7. Ce résultat se généralise en deux dimensions. Dans un bloc q , un *bord diagonal* est un bloc apparaissant dans deux coins opposés de q .

Théorème 6.1. *Soit q un bloc fini. Il existe une configuration q -quasipériodique, non périodique si et seulement si la racine primitive de q a un bord diagonal non vide.*

Cette section est consacrée à la preuve du théorème 6.1.

En premier lieu, supposons que \mathbf{w} soit une configuration à la fois q -quasipériodique et apériodique. Il existe alors deux occurrences de q qui se chevauchent, sans quoi \mathbf{w} serait q -périodique. De plus, le chevauchement n'est pas une puissance de la racine primitive de q , car si tous les chevauchements étaient ainsi, la racine primitive serait période de \mathbf{w} . Nous en déduisons que la racine primitive de q possède un bord diagonal non vide.

Montrons maintenant la réciproque : supposons que r , la racine primitive de q , possède un bord diagonal non vide b , et construisons une configuration apériodique admettant q comme quasipériode. Considérons pour cela les quatre tuiles α, β, γ et δ de la figure 6.1. Chaque rectangle est une occurrence de q , et le chevauchement est systématiquement b .

Soit $A = \{a_1, a_2, a_3, a_4\}$ et μ la fonction de $A^{\mathbb{Z}^2}$ vers $\Sigma^{\mathbb{Z}^2}$ définie par

$$\mu(a_1) = \alpha \quad \mu(a_2) = \beta \quad \mu(a_3) = \delta \quad \mu(a_4) = \gamma.$$

Si la configuration donnée en entrée obéit à certaines conditions de régularité, la fonction μ se comporte comme un morphisme avec une forme inhabituelle de concaténation. Sur la figure 6.1, chaque tuile possède des *ancres*, qui sont les lettres signalées par les carrés gris. La concaténation horizontale s'effectue en superposant l'ancre droite de la première tuile et l'ancre inférieure gauche de la seconde. De la même façon, la concaténation verticale s'effectue en superposant l'ancre supérieure de la première tuile avec l'ancre inférieure gauche de la seconde. Notons \cup l'union de deux fonctions de domaines disjoints, c'est-à-dire que $(u \cup v)[x, y] = u[x, y]$ si $(x, y) \in \text{dom}(u)$ et $(u \cup v)[x, y] = v[x, y]$ si $(x, y) \in \text{dom}(v)$. De plus, l'opérateur $S_{n,m}$ (comme *shift*) est la translation par le vecteur (n, m) . Alors nous avons

$$\begin{aligned} \mu(a_i \cup u) &= \mu(a_i) \cup (S_{\text{height}(b), 4 \cdot \text{width}(q)} \circ \mu)(u), \\ \mu\left(\begin{smallmatrix} u \\ v \end{smallmatrix}\right) &= \mu(u) \cup (S_{4 \cdot \text{height}(q), \text{width}(b)} \circ \mu)(v). \end{aligned}$$

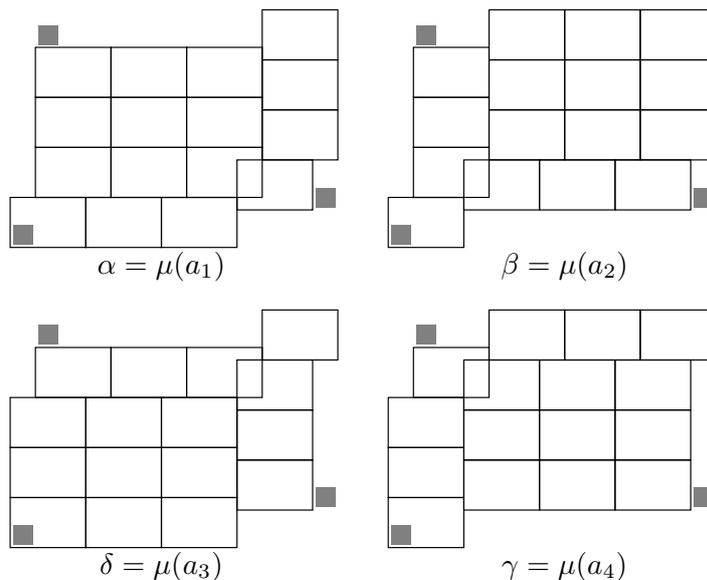


FIG. 6.1 : Quatre tuiles pour construire un mot q -quasipériodique. Chaque rectangle est une occurrence de q .

Si u est un bloc, l'ancre inférieure gauche de $\mu(u[i, j])$ a pour coordonnées

$$(4j \text{ height}(q) + i \text{ height}(b); 4i \text{ width}(q) + j \text{ width}(b)).$$

La figure 6.2 illustre le fonctionnement de μ .

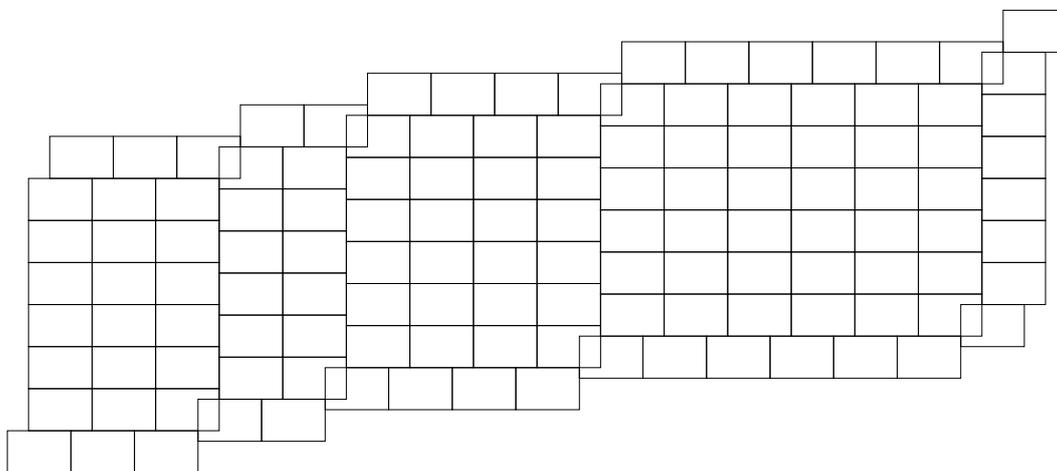


FIG. 6.2 : Représentation de $\mu \begin{pmatrix} a_3 & a_4 & a_4 & a_3 \\ a_1 & a_2 & a_2 & a_1 \end{pmatrix}$, chaque rectangle est une occurrence de q .

Nous disons qu'une configuration sur l'alphabet A est *idoine* si elle satisfait les conditions suivantes :

1. chaque ligne est soit sur l'alphabet $\{a_1, a_2\}$ soit sur l'alphabet $\{a_3, a_4\}$;

2. chaque colonne est soit sur l'alphabet $\{a_1, a_3\}$ soit sur l'alphabet $\{a_2, a_4\}$.

Observons que, si \mathbf{w} est idoine, alors $\mu(\mathbf{w})$ est bien définie. Chaque lettre de $\mu(a_1a_2)$, $\mu(a_2a_1)$, $\mu(a_3a_4)$ et $\mu(a_4a_3)$ est bien définie, de même que chaque lettre de $\mu\left(\begin{smallmatrix} a_1 \\ a_3 \end{smallmatrix}\right)$, $\mu\left(\begin{smallmatrix} a_3 \\ a_1 \end{smallmatrix}\right)$, $\mu\left(\begin{smallmatrix} a_2 \\ a_4 \end{smallmatrix}\right)$, $\mu\left(\begin{smallmatrix} a_4 \\ a_2 \end{smallmatrix}\right)$. En effet, nulle position ne reçoit deux lettres différentes, et l'image de $\mu(\mathbf{w})$ ne possède aucun « trou » ; plus formellement :

- pour tous entiers i, j, j_1, j_2 tels que $j_1 \leq j \leq j_2$, si (i, j_1) et (i, j_2) sont dans le domaine de \mathbf{w} , alors (i, j) y appartient également ;
- pour tous entiers j, i, i_1, i_2 tels que $i_1 \leq i \leq i_2$, si (i_1, j) et (i_2, j) sont dans le domaine de \mathbf{w} , alors (i, j) y appartient également.

Montrons maintenant que μ préserve l'apériodicité : si \mathbf{w} est apériodique, alors $\mu(\mathbf{w})$ l'est également.

Lemme 6.2. *Soient x, y deux éléments différents de l'ensemble $\{\alpha, \beta, \gamma, \delta\}$. Alors une occurrence de x et une occurrence de y ne peuvent se chevaucher si leurs ancres coïncident.*

En d'autres termes, les situations montrées par la figure 6.3 ne se produisent jamais.

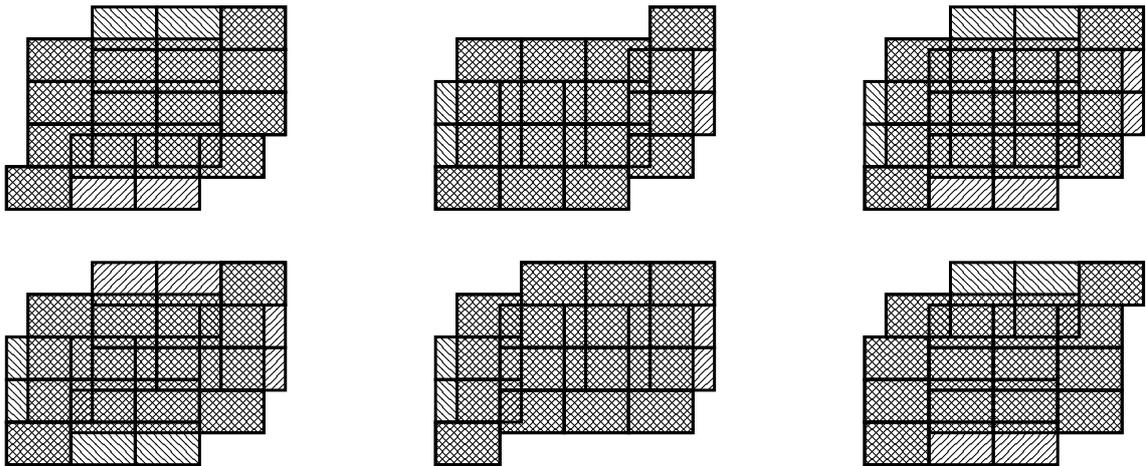


FIG. 6.3 : Tous les chevauchements possibles pour $\alpha, \beta, \gamma, \delta$.

Démonstration. Il y a six valeurs possibles pour l'ensemble $\{x, y\}$. La preuve fonctionne toujours de la même façon, à symétrie près, nous traiterons donc uniquement le cas $x = \alpha$ et $y = \beta$ (en haut à gauche sur la figure 6.3). Nous appelons q le bloc utilisé pour la construction des tuiles, r sa racine primitive, et b un bord diagonal de r .

Supposons que x et y puissent se superposer. Alors il existe trois occurrences de q verticalement alignées, q_1, q_2 et q_3 , telles que q_1 est facteur de q_2q_3 . (Par exemple, voir la deuxième colonne d'occurrences q sur la figure 6.3.) Si nous regardons q_1, q_2 et q_3 comme des mots sur l'alphabet $\Sigma^{1 \times \text{width}(q)}$, nous voyons qu'il existe des mots x, x' sur $\Sigma^{1 \times \text{width}(q)}$ tels que $q_1 = xx'$ et $q_2 = q_3 = x'x$ (la concaténation est effectuée du haut

vers le bas). Par le théorème 1.1, les mots x et x' (et donc q) sont des puissances d'un même mot s sur l'alphabet $\Sigma^{1 \times \text{width}(q)}$.

Le bloc s est un préfixe vertical de q et de x . Il existe donc trois occurrences de s , verticalement alignées, s_1, s_2 et s_3 , telles que s_1 est facteur interne de s_2s_3 . Il existe donc des mots y, y' sur l'alphabet $\Sigma^{\text{height}(s) \times 1}$ tels que $s_1 = yy'$ et $s_2 = s_3 = y'y$ et $\text{width}(y') = \text{width}(b)$. Par le théorème 1.1 à nouveau, il existe un mot t sur l'alphabet $\Sigma^{\text{height}(s) \times 1}$ tel que y et y' (et donc s) sont des puissances de t .

Soient k, ℓ les entiers vérifiant $q = s^k$ (en concaténation verticale) et $s = t^\ell$ (en concaténation horizontale). Alors $q = t^{k \times \ell}$ et t est une racine de q . De plus,

$$\text{width}(t) \leq \text{width}(y') = \text{width}(b) \quad \text{et} \quad \text{height}(t) = \text{height}(s) \leq \text{height}(b)$$

Ainsi, $|t| \leq |b|$, ce qui implique une contradiction : soit $t = b$, ce qui est impossible car b est un bord de (donc un bloc strictement plus petit que) la racine primitive de q , soit $t \subset b \subseteq r$, et dans ce cas t est la racine primitive de q , qui ne peut donc être r . \square

Lemme 6.3. *Soit q un bloc, r sa racine primitive et b un bord diagonal non vide de r . Soit \mathbf{w} une configuration apériodique et idoine. Alors $\mu(\mathbf{w})$ est apériodique et q -quasipériodique.*

Démonstration. Par construction, $\mu(\mathbf{w})$ est q -quasipériodique pour tout \mathbf{w} . Supposons que $\mu(\mathbf{w})$ a un vecteur de périodicité $p \in \mathbb{Z}^2$ non nul. Nous démontrons que, dans ces conditions, \mathbf{w} est périodique.

Soit $a \in \mathbb{Z}^2$ les coordonnées d'une ancre d'une tuile de $\mu(\mathbf{w})$. Pour tout $i \in \mathbb{Z}$, posons $t_i = a + ip$. Comme les tuiles contiennent au plus $16 \times \text{width}(q) \times \text{height}(q)$ lettres, par principe des tiroirs, il existe deux coordonnées t_i et t_j qui se trouvent à la même position relative de l'ancre de leurs tuiles respectives. La différence de coordonnées entre ces deux ancres est donc un multiple de p .

Appelons T_i (respectivement T_j) la tuile couvrant la position t_i (respectivement t_j). Comme T_i est la translation par $(j-i) \times p$ de T_j , par le lemme 6.2, elles sont occurrences d'une même tuile. De la même façon, le voisin du dessus et de droite de T_i et de T_j sont également égaux. En itérant cet argument, nous déduisons que \mathbf{w} lui-même est périodique. \square

Ceci achève la preuve du théorème 6.1. Nous appelons *quasipériode triviale* toute quasipériode dont la racine primitive admet le bloc vide comme seul bord diagonal. Si une configuration possède une quasipériode triviale, alors elle est nécessairement périodique. A contrario, toute quasipériode non triviale admet une configuration apériodique.

6.2 Configurations quasipériodiques irrégulières

Poursuivons la généralisation des résultats élémentaires du cas unidimensionnel aux configurations de \mathbb{Z}^2 . Comme le montre la proposition 2.7, la quasipériodicité en une dimension n'est pas une bonne notion de symétrie : il existe toutes sortes de mots,

très simples ou très irréguliers, qui sont quasipériodiques. Sans grande surprise, cette constatation se retrouve dans le cas bidimensionnel : rajouter une dimension ajoute de la liberté pour créer du désordre.

Proposition 6.4. *Soit q une quasipériode non triviale. Il existe une configuration q -quasipériodique qui n'est pas uniformément récurrente, et il existe une configuration q -quasipériodique dont les fréquences de certains facteurs ne convergent pas.*

Démonstration. Nous reprenons les définitions et notations de la section précédente. Considérons $\mathbf{w} \in A^{\mathbb{Z}^2}$, représenté sur la figure 6.4 et défini comme suit :

- $\mathbf{w}[0, 0] = a_4$;
- $\mathbf{w}[i, 0] = a_2$ pour $j \in \mathbb{Z}^*$;
- $\mathbf{w}[0, j] = a_3$ pour $i \in \mathbb{Z}^*$;
- $\mathbf{w}[i, j] = a_1$ pour $i, j \in \mathbb{Z}^*$.

Le mot \mathbf{w} est idoine, donc $\mu(\mathbf{w})$ est bien défini, q -quasipériodique, et ne contient $\mu\left(\begin{smallmatrix} a_1 & a_2 & a_1 \\ a_3 & a_4 & a_3 \\ a_1 & a_2 & a_1 \end{smallmatrix}\right)$ qu'une seule fois. En effet, il s'agit du seul facteur de cette taille comprenant une densité de chevauchements sur b aussi élevée, comme le montre la figure 6.5. Il n'est donc pas uniformément récurrent.

Considérons maintenant $\mathbf{u} \in A^{\mathbb{Z}^2}$ construit en plaçant une copie du mot de Champernowne (symétrisé pour en obtenir une version biinfinie) sur l'alphabet $\{a_1, a_2\}$ sur chaque ligne. La fréquence du facteur indiqué sur la figure 6.6 est proportionnelle à la fréquence de a_1a_2 dans \mathbf{u} , qui n'est pas définie. Donc ce facteur n'a pas de fréquence dans $\mu(\mathbf{u})$. \square

6.3 Le cas de l'entropie

De façon surprenante, le résultat précédent ne s'étend pas à l'entropie topologique. Le même type de preuve ne peut plus fonctionner, car toute configuration idoine doit nécessairement avoir une entropie topologique nulle. En effet, dans une telle configuration, tout bloc de taille $n \times n$ est uniquement déterminé par sa frontière en haut et à droite, or il y a au plus 4^{2n+1} possibilités pour cette frontière. La suite $(\log(4^{2n+1})/n^2)_{n \in \mathbb{N}}$ tend vers zéro, donc l'entropie de la configuration considérée est nulle.

Attelons-nous maintenant à construire un mot q -quasipériodique d'entropie positive. Dans un cas précis, nous pouvons simplement importer la preuve unidimensionnelle.

Remarque 6.5. Soit q un bloc de taille $h \times w$ qui possède au moins un bord b de pleine largeur (ou de pleine hauteur) qui ne soit pas puissance de la racine primitive de q . Alors il existe une configuration q -quasipériodique d'entropie topologique strictement positive.

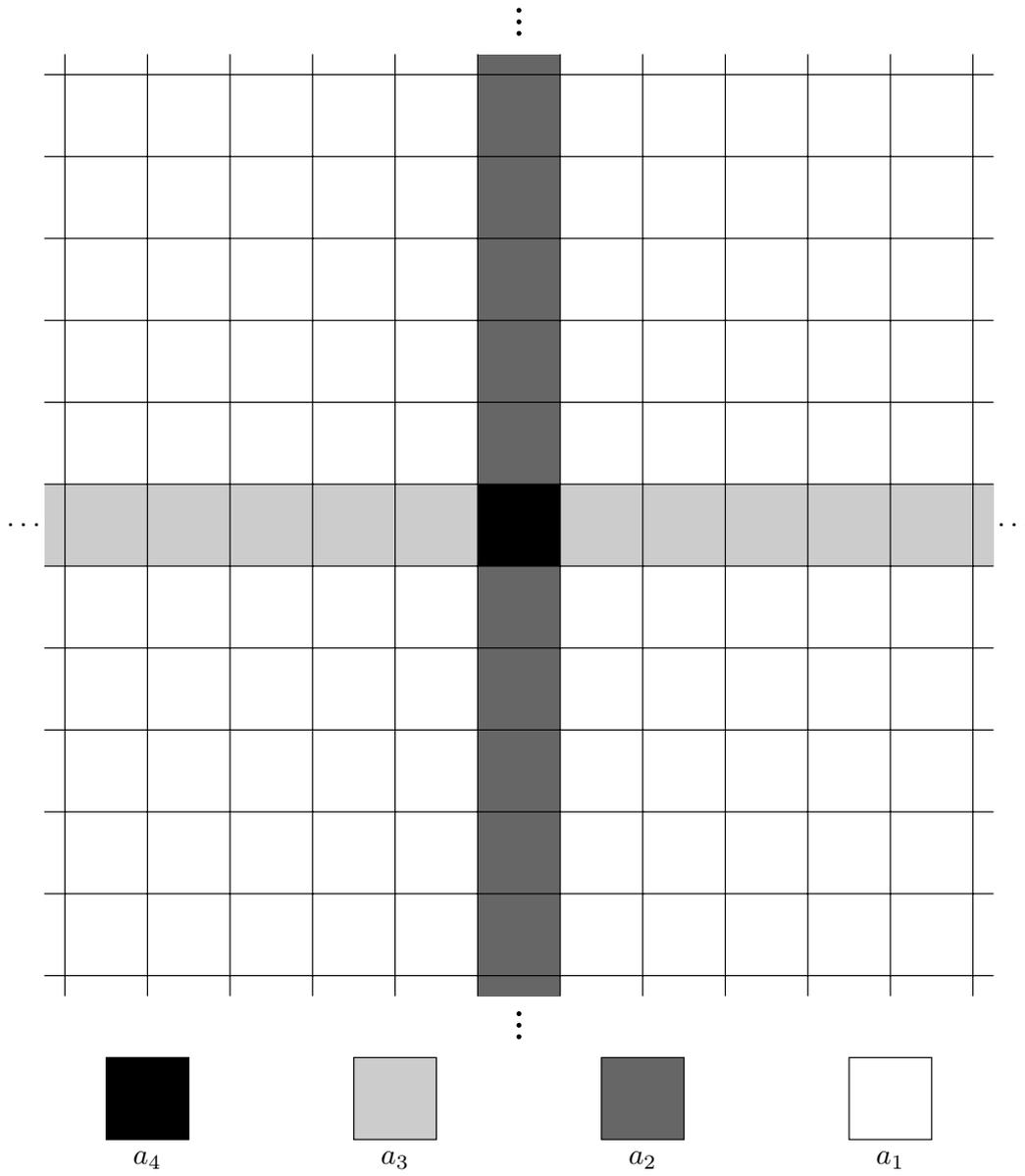


FIG. 6.4 : Illustration de la preuve de la proposition 6.4

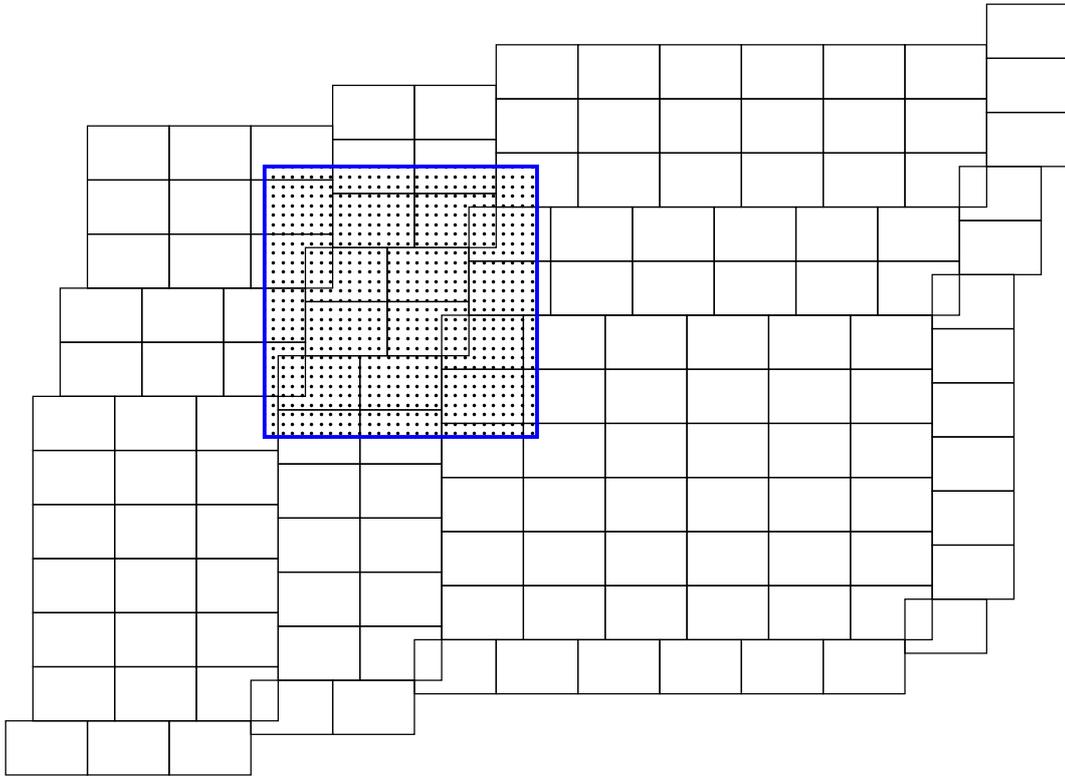


FIG. 6.5 : Facteur non récurrent dans $\mu(\mathbf{w})$, preuve de la proposition 6.4

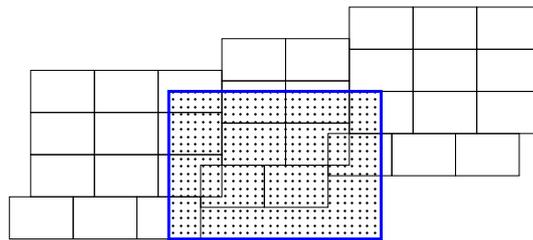


FIG. 6.6 : Facteur de $\mu(a_1a_2)$ dont la fréquence ne converge pas dans $\mu(\mathbf{u})$, preuve de la proposition 6.4

Démonstration. Supposons sans perte de généralité que b soit de pleine largeur et appelons h' sa hauteur. Alors il existe deux blocs différents z_1, z_2 , de taille $(h + \text{ppcm}(h, h - h')) \times w$, tous deux couverts par des occurrences de q : le bloc z_1 est une concaténation de l'antibord de b dans q , et le bloc z_2 est une concaténation verticale d'occurrences de q . Le morphisme par blocs défini par $h(1) = z_1$ et $h(2) = z_2$ est injectif ; l'image d'une configuration d'entropie topologique positive par h a de nouveau une entropie topologique positive, et admet q comme quasipériode puisque z_1 et z_2 sont tous deux q -quasipériodiques. \square

Nous montrons maintenant que la construction d'un mot q -quasipériodique avec une entropie topologique strictement positive requiert des conditions supplémentaires sur q : la non trivialité, telle que définie à la fin de la section précédente, ne suffit plus. Il existe d'autres cas qui forcent l'entropie topologique à être nulle.

Théorème 6.6. *Soit q un bloc de taille $h \times w$ et \mathbf{w} une configuration q -quasipériodique. Si q possède un coin dont le seul bord est vide, alors \mathbf{w} a une entropie topologique nulle.*

Montrons maintenant le théorème 6.6. En premier lieu, nous pouvons supposer que tous les bords sont dans le coin supérieur gauche/inférieur droit de q . Les coins opposés n'ont pas de bords, et en particulier q n'a aucun bord pleine hauteur ni pleine largeur.

Pour deux entiers y, x , nous notons $\text{occ}(y, x)$ l'occurrence de q dans \mathbf{w} qui couvre la position (y, x) . S'il y a plusieurs occurrences, nous choisissons celle d'ordonnée minimale ; en cas d'égalité, nous choisissons celle d'abscisse minimale parmi elles. Rappelons que les *coordonnées* d'une occurrence sont les coordonnées de son coin inférieur gauche.

Considérons des entiers y, x et une occurrence de domaine $\text{rect}(y, x, h, w)$. Nous définissons les opérateurs

$$\begin{aligned} \text{right}(y, x) &= \text{occ}(y + h - 1, x + w) \\ \text{above}(y, x) &= \text{occ}(y + h, x + w - 1) \end{aligned}$$

qui donnent intuitivement l'occurrence à droite et l'occurrence au-dessus de celle en (x, y) . De plus, si y', x', y'', x'' sont les valeurs qui satisfont $\text{right}(y, x) = (y', x')$ et $\text{above}(y, x) = (y'', x'')$, nous définissons les quantités

$$d_R(y, x) = y' - y, \quad d_A(y, x) = x'' - x.$$

La figure 6.7 illustre ces définitions.

Étant donné que q n'a pas de bord dans son coin supérieur droit, les occurrences (y, x) , $\text{right}(y, x)$ et $\text{above}(y, x)$ ne se chevauchent pas. Les coordonnées de $\text{right}(y, x)$ sont donc $(y + d_R(y, x), x + w)$, et celles de $\text{above}(y, x)$ sont $(y + h, x + d_A(y, x))$. Si $d_A(y, x) > 0$ et $d_R(y, x) > 0$ à la fois, alors $(d_A(y, x), d_R(y, x))$ est la taille d'un bord de q dans le coin supérieur gauche. Seuls les trois cas représentés sur la figure 6.7 sont possibles :

1. $d_R(y, x) = 0$ et $0 \leq d_A(y, x) < w$;
2. $d_A(y, x) = 0$ et $0 \leq d_R(y, x) < h$;

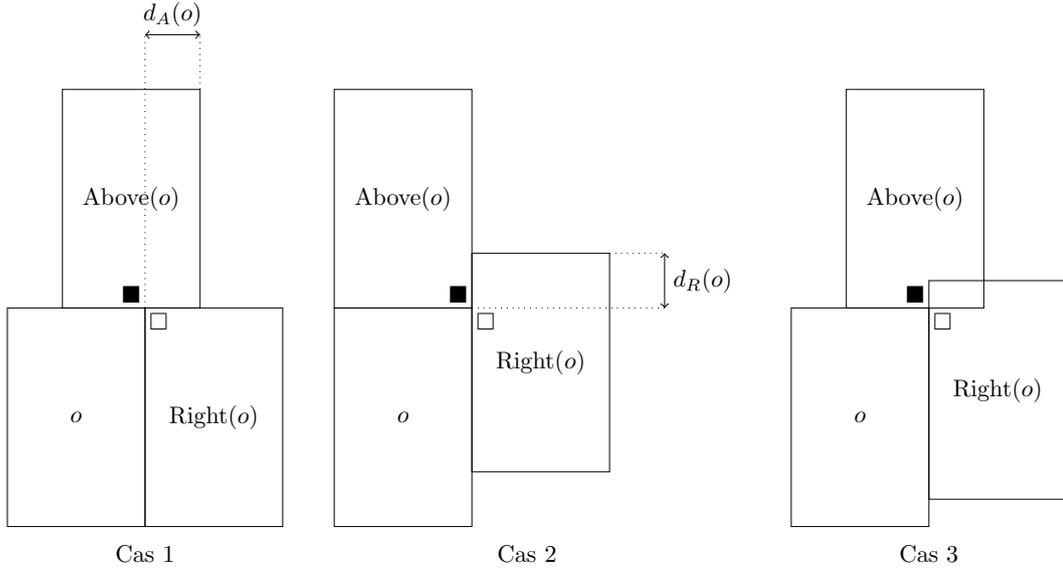


FIG. 6.7 : Illustration de above, right, d_A and d_R . Les marques \square et \blacksquare montrent les lettres couvertes par $\text{right}(o)$ et $\text{above}(o)$, respectivement.

3. $(d_A(y, x), d_R(y, x))$ est la taille d'un bord non vide du coin supérieur gauche de q .

Lemme 6.7. Soient \mathbf{w} et q comme dans le théorème 6.6. Soient (x, y) les coordonnées d'une occurrence de q dans \mathbf{w} . Alors,

$$(\text{right} \circ \text{above})(y, x) = (\text{above} \circ \text{right})(y, x) = \text{occ}(y + h + d_R(x, y), x + w + d_A(x, y)).$$

Démonstration. Soient y', x' les entiers définis par

$$x' = x + w + d_A(x, y) \quad \text{et} \quad y' = y + h + d_R(y, x).$$

L'occurrence $\text{occ}(y', x')$ a nécessairement (y', x') comme coordonnées pour son coin inférieur gauche : sans cela, l'occurrence (y', x') chevaucherait soit $\text{right}(y, x)$, soit $\text{above}(y, x)$ par le coin supérieur droit, ce qui est exclu (cf. figure 6.8). Ainsi, le coin supérieur droit de $\text{occ}(y', x')$ a pour coordonnées

$$(y + 2h + d_R(x, y) - 1, x + 2w + d_A(x, y) - 1).$$

Observons que

$$\begin{aligned} (\text{right} \circ \text{above})(y, x) &= \text{occ}(y + d_R(x, y) + h, x + 2w - 1) && \text{et} \\ (\text{above} \circ \text{right})(y, x) &= \text{occ}(y + 2h - 1, x + w + d_A(x, y)). \end{aligned}$$

Ces deux points sont couverts par $\text{occ}(y', x')$, ce qui peut se vérifier directement sur la figure 6.8 ou plus formellement par des calculs de coordonnées

$$\begin{aligned} x + w + d_A(x, y) &\leq x + 2w - 1 \leq x + 2w + d_A(x, y) - 1, \\ y + h + d_R(y, x) &\leq y + 2h - 1 \leq y + 2h + d_R(y, x) - 1. \end{aligned}$$

De plus, $\text{occ}(y', x')$ est l'occurrence le plus en bas et le plus à gauche couvrant ces deux points. Toute occurrence plus bas ou plus à gauche chevaucherait $\text{above}(y, x)$ ou $\text{right}(y, x)$ par le coin supérieur droit, ce qui est exclu. Donc, par définition de l'opérateur occ , l'occurrence $\text{occ}(y', x')$ est bien celle recherchée. \square

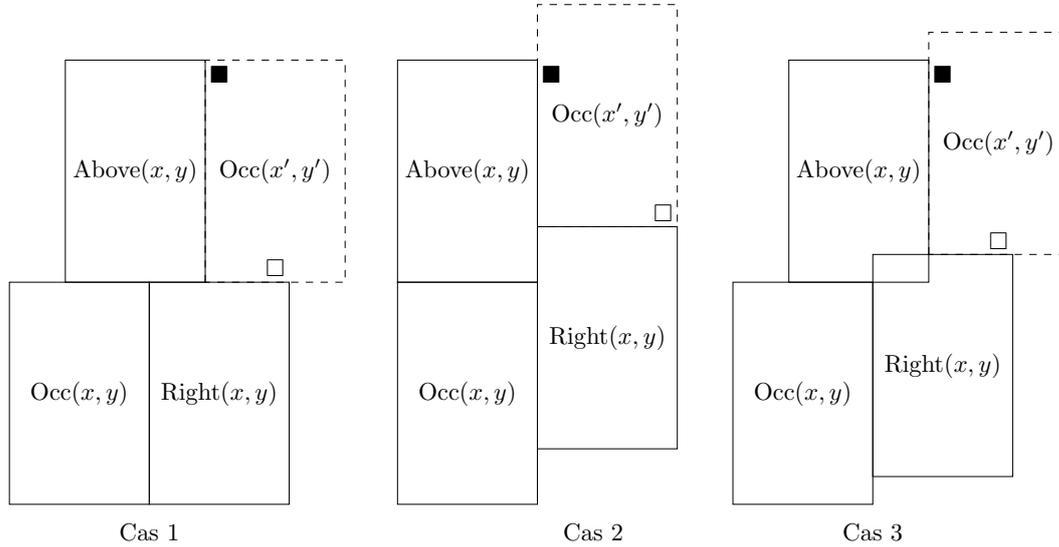


FIG. 6.8 : Illustration de la preuve du lemme 6.7. La marque \blacksquare indique $(y + 2h - 1, x + w + d_A(y, x))$ et \square indique $(y + d_R(x, y) + h, x + 2w - 1)$.

Démonstration du théorème 6.6. Soit n un entier et B un bloc arbitraire de \mathbf{w} de taille $n \times n$, dont les coordonnées sont (y, x) . Soient u_1, \dots, u_k la suite des occurrences de q qui couvrent la frontière inférieure de B , c'est-à-dire les positions $\{(y, x), \dots, (y, x + n - 1)\}$. De même, soit v_1, \dots, v_ℓ la suite des occurrences de q qui couvrent la frontière gauche de B , c'est-à-dire les positions $\{(y, y), \dots, (y + n - 1, x)\}$. Observons que $u_1 = v_1$.

Par le lemme 6.7, B est uniquement déterminé par $u_1, \dots, u_k, v_1, \dots, v_\ell$ et les coordonnées de B relativement à celles de u_1 . Nous bornons le nombre de possibilités pour B . Pour tout entier i compris entre 1 et $k - 1$, nous avons soit $u_{i+1} = \text{right}(u_i)$, soit $\text{above}(u_{i+1}) = \text{right}(u_i)$, il y a donc $2^k < 2^n$ possibilités pour la suite (u_i) . De la même façon, pour tout entier j compris entre 1 et $\ell - 1$, nous avons soit $v_{j+1} = \text{above}(v_j)$, soit $\text{right}(v_{j+1}) = \text{above}(v_j)$, donc au plus $2^\ell < 2^n$ possibilités pour la suite (v_j) . Nous avons donc au plus $|q| \times 2^{2n}$ possibilités pour B . Observons que

$$\lim_{n \rightarrow \infty} \frac{\log_2(|q| \times 2^{2n})}{n^2} = \lim_{n \rightarrow \infty} \frac{2n + \log_2 |q|}{n^2} \rightarrow 0,$$

et donc \mathbf{w} a une entropie topologique nulle. \square

Reste donc à explorer le dernier cas, en dehors de la remarque 6.5 et du théorème 6.6 : que se passe-t-il si un bloc q possède des blocs dans tous les coins, mais pas de bord pleine largeur/hauteur ? Après d'infructueuses tentatives de construire un mot q -quasipériodique d'entropie non nulle, nous formulons la conjecture suivante.

Conjecture 6.8. *Soit q un bloc qui possède des bords dans tous les coins, mais pas de bord pleine largeur ou pleine hauteur. Alors toute configuration q -quasipériodique est d'entropie nulle.*

Idée de preuve. Considérer \mathbf{w} une configuration q -quasipériodique, c un carré $n \times n$ arbitraire, et définir u comme l'union de toutes les occurrences de q qui recouvrent une position de c dans \mathbf{w} . Montrer que la frontière du domaine de u détermine uniquement les lettres de u , par des techniques analogues à la preuve du théorème 5.9.

Si cette conjecture se révélait vraie, nous aurions le résultat suivant. *Soit q un bloc. Il existe une configuration q -quasipériodique d'entropie strictement positive si et seulement si q possède un bord pleine largeur (ou pleine hauteur) qui n'est pas une puissance de sa racine primitive.* En d'autres termes, il n'existerait qu'une seule façon de « créer de l'entropie » avec une quasipériode : en construisant des lignes (ou des colonnes) d'occurrences, indépendantes les unes des autres, dont les successions chevauchement/non chevauchement contiennent de l'entropie. Intuitivement, la seule façon de créer de l'entropie en deux dimensions serait d'émuler le comportement unidimensionnel.

A mathematician is a device turning coffee into theorems.

— Folklore, attribué à Alfréd Rényi

Chapitre 7

Quasipériodicité multi-échelles en deux dimensions

Dans ce chapitre, nous définissons la *quasipériodicité multi-échelles* sur les configurations de \mathbb{Z}^2 . Le passage des configurations de \mathbb{N} à \mathbb{Z} avait posé problème pour l'uniforme récurrence ; notre objet d'étude étant les configurations de \mathbb{Z}^2 , nous verrons que les mêmes problèmes se posent. Toutefois, dans les configurations multi-échelles, l'entropie topologique est nulle et les fréquences des facteurs convergent. Nous montrerons ce dernier point sans utiliser d'outils de théorie ergodique, afin de minimiser les prérequis nécessaires à la lecture de ce chapitre et de montrer comment travailler combinatoirement sur les blocs.

En une dimension, la quasipériodicité multi-échelles se définit comme le fait d'avoir une infinité de quasipériodes, mais nous ne pouvons pas généraliser naïvement cette définition. En effet, dans ce cas, toute configuration dont chaque ligne (ou chaque colonne) est quasipériodique multi-échelles, avec le même ensemble de quasipériodes, serait à son tour quasipériodique multi-échelles. Nous sommes ici particulièrement attachés à étudier les phénomènes spécifiques aux configurations en dimension 2, nous allons donc exclure ce cas de figure. Une configuration \mathbf{w} est appelée *quasipériodique multi-échelles* si et seulement si, pour tout entier naturel n , la configuration \mathbf{w} possède une quasipériode dont la hauteur et la largeur sont supérieures à n .

7.1 Entropie topologique

Comme nous l'avons mentionné, la quasipériodicité multi-échelles sur les mots infinis implique diverses propriétés : uniforme récurrence, unique ergodicité, et faible complexité en facteurs donc entropie topologique nulle (théorème 2.8). Nous cherchons maintenant à généraliser ces résultats aux mots en deux dimensions.

Proposition 7.1. *Toute configuration quasipériodique multi-échelles a une entropie topologique nulle.*

Démonstration. Soit \mathbf{w} une configuration multi-échelles et q une quasipériode de taille $n \times m$ de \mathbf{w} . Sans perte de généralité, supposons $n \leq m$. Soit s un bloc arbitraire de

taille $m \times m$ de \mathbf{w} . Ce bloc est couvert d'occurrences de q (qui peuvent dépasser), et la position relative de ces occurrences et de s détermine uniquement s .

Nous avons besoin d'au plus $4m$ occurrences de q pour couvrir s . En effet, chaque occurrence de q doit avoir l'un de ses coins dans s . Si deux occurrences o, o' de q ont toutes deux leurs coins inférieurs droits sur la même ligne de s , avec o' le plus à droite, alors $\text{dom}(o) \cap \text{dom}(s) \subseteq \text{dom}(o') \cap \text{dom}(s)$ et l'occurrence de o n'apporte aucune information supplémentaire sur s par rapport à o' . Étant donné qu'un rectangle possède 4 coins et que s possède m lignes, s est uniquement déterminé par $4m$ occurrences de q . Chacune de ces m occurrences est uniquement déterminée par sa position relative au bord gauche de s . Ainsi, le nombre $P_{\mathbf{w}}(m, m)$ de possibilités pour s est borné par m^{4m} . Comme $\log(m^{4m})/m^2 = 4m \log(m)/m^2$ tend vers 0 quand m tend vers l'infini, l'entropie de \mathbf{w} est nulle. \square

7.2 Uniforme récurrence

Proposition 7.2. *Toute configuration de \mathbb{N}^2 qui est quasipériodique multi-échelles est uniformément récurrente.*

Démonstration. Chaque quasipériode apparaît à l'origine de la configuration et les quasipériodes grandissent dans les deux directions, donc tout facteur apparaît dans une quasipériode et le résultat suit. \square

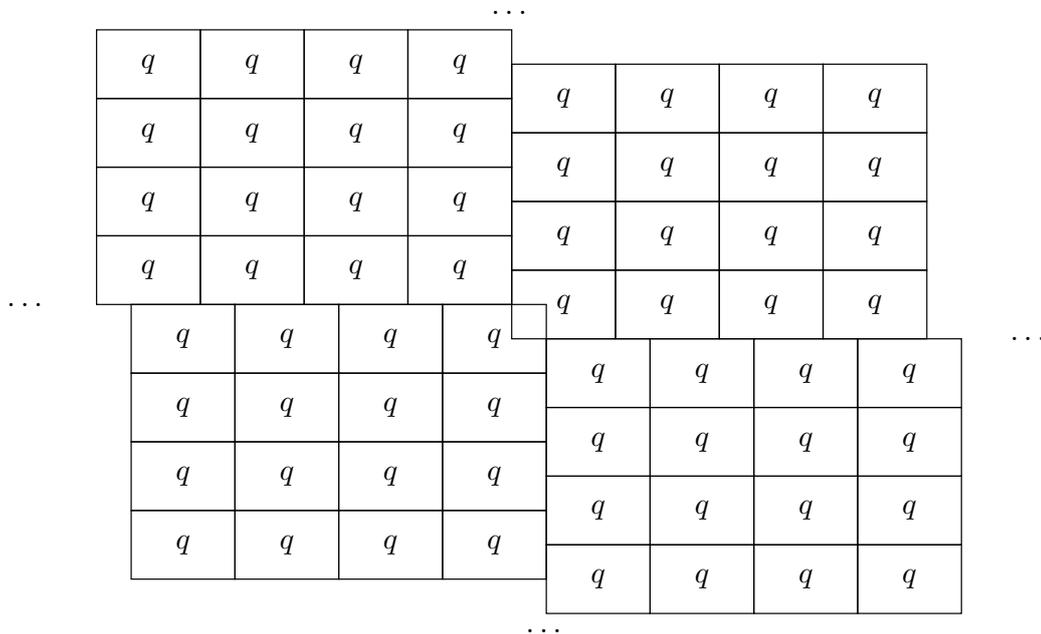


FIG. 7.1 : Configuration multi-échelles, mais non récurrente

Toutefois, il existe des configurations de \mathbb{Z}^2 qui ne sont pas uniformément récurrentes. La figure 7.1 en présente une, pour une quasipériode q dont la racine primitive

possède un bord diagonal non vide, par exemple $q = \begin{smallmatrix} b & b & a \\ a & b & b \end{smallmatrix}$. Le facteur central

$$\begin{matrix} b & b & b & b & a \\ b & b & a & b & b \\ a & b & b & b & b \end{matrix}$$

n'apparaît qu'une seule fois, et n'est donc pas récurrent ni uniformément récurrent. Nous pouvons introduire une version forte de la quasipériodicité multi-échelles qui force l'uniforme récurrence : une configuration \mathbf{w} est *fortement* multi-échelles si tout facteur de \mathbf{w} est un facteur d'une quasipériode de \mathbf{w} . En d'autres termes, pour tout facteur u de \mathbf{w} , il existe une quasipériode q de \mathbf{w} telle que u est facteur de q . Considérons par exemple $r = \begin{smallmatrix} \square & \blacksquare & \square \\ \blacksquare & \square & \blacksquare \\ \square & \blacksquare & \square \end{smallmatrix}$ et h le morphisme défini par

$$\begin{aligned} h(\square) &= r^{1 \times 3} \\ &= \begin{matrix} \square & \blacksquare & \square & \square & \blacksquare & \square & \square & \blacksquare & \square \\ \blacksquare & \square & \blacksquare & \blacksquare & \square & \blacksquare & \blacksquare & \square & \blacksquare \\ \square & \blacksquare & \square & \square & \blacksquare & \square & \square & \blacksquare & \square \end{matrix} \\ h(\blacksquare) &= \text{damier } 3 \times 9 \\ &= \begin{matrix} \square & \blacksquare & \square & \blacksquare & \square & \blacksquare & \square & \blacksquare & \square \\ \blacksquare & \square & \blacksquare & \square & \blacksquare & \square & \blacksquare & \square & \blacksquare \\ \square & \blacksquare & \square & \blacksquare & \square & \blacksquare & \square & \blacksquare & \square \end{matrix} \end{aligned}$$

Ce morphisme admet un point fixe $h^\omega(\square)$, qui est une configuration de \mathbb{Z}^2 . (Pour obtenir le point fixe, nous positionnons le centre des $h^n(\square)$ à l'origine, de telle façon que $h^{n+1}(\square)$ soit une extension de $h^n(\square)$.) Cette configuration admet $r, h(r), h^2(r), \dots$ comme quasipériodes. En outre, tout facteur de cette configuration est un facteur de $h^n(\square)$ pour un certain n , donc en particulier un facteur de $h^n(r)$. Cette configuration est donc fortement multi-échelles.

Proposition 7.3. *Une configuration \mathbf{w} est fortement multi-échelles si et seulement si elle est multi-échelles et uniformément récurrente.*

Démonstration. La propriété de forte quasipériodicité multi-échelles implique l'uniforme récurrence par définition. Réciproquement, considérons \mathbf{w} une configuration quasipériodique multi-échelles qui soit uniformément récurrente, et B un bloc de \mathbf{w} . Alors il existe un entier n tel que tout bloc de taille $n \times n$ contienne B . Par définition de la quasipériodicité multi-échelles, il existe une quasipériode de taille supérieure à $n \times n$; cette dernière doit nécessairement contenir une occurrence de B . \square

7.3 Fréquences de facteurs

Penchons-nous maintenant sur les fréquences des facteurs des configurations multi-échelles. Cette section est consacrée à la preuve du résultat suivant.

Théorème 7.4. *Toute configuration de \mathbb{Z}^2 quasipériodique multi-échelles a des fréquences qui convergent pour tous ses facteurs.*

Rappelons que si u, v sont des blocs finis, alors $|u| = \text{width}(u) \times \text{height}(u)$ et $|v|_u$ est le nombre d'occurrences de u dans v . La fonction f_u est définie par $f_u(v) = |v|_u/|v|$. Si \mathbf{w} est une configuration de \mathbb{Z}^2 , alors

$$f_u(\mathbf{w}) = \lim_{n \rightarrow \infty} f_u(\mathbf{w}[(-n, -n), \dots, (+n, +n)]).$$

Notre objectif est de montrer que, si \mathbf{w} est une configuration quasipériodique multi-échelles, alors $f_u(\mathbf{w})$ est bien définie pour tout bloc u ; en particulier, nous devons montrer que la limite ci-dessus converge. Rappelons que $\mathcal{L}_{\geq k \times k}(\mathbf{w})$ est l'ensemble des blocs de \mathbf{w} dont la largeur et la hauteur sont plus grandes que k . Pour les besoins de la preuve, nous étendons les définitions de $|v|$ et $|v|_u$ aux figures de domaine non nécessairement rectangulaire. Alors $|v| = \text{card}(\text{dom}(v))$ et $|v|_u$ est le nombre d'occurrences complètes de u dans v .

Lemme 7.5. *Soit v un bloc fini et v_1, v_2 une partition de v , c'est-à-dire deux figures telles que $\text{dom } v_1 \cap \text{dom } v_2 = \emptyset$ et $v_1 \cup v_2 = v$. Soit u un bloc. Si v_1 est un bloc dans un coin de v , alors :*

$$|v_1|_u \leq |v_1|_u + |v_2|_u \leq |v|_u \leq |v_1|_u + |v_2|$$

Démonstration. Les deux premières inégalités sont immédiates par les propriétés de v_1, v_2 . Pour la troisième inégalité, observons que toute occurrence de u dans v est soit entièrement contenue dans v_1 , soit dans v_2 , soit se trouve à cheval entre les deux. Les occurrences du premier type sont au nombre de $|v_1|_u$ par définition; les autres occurrences ont toutes un coin dans v_2 , et il ne peut y avoir plus de telles occurrences que de lettres dans v_2 . \square

Lemme 7.6. *Soient x, y, z des réels strictement positifs et $(u_n)_{n \in \mathbb{N}}$ une suite de blocs telle que, pour tout n , nous avons $u_n \in \Sigma^{>n \times n}$. Alors la relation*

$$\lim_{n \rightarrow \infty} \frac{|u_n|}{x \text{width}(u_n) + y \text{height}(u_n) + z} \rightarrow +\infty.$$

est satisfaite.

Démonstration. Comme x, y sont positifs, les deux fonctions suivantes sont croissantes.

$$w \mapsto \frac{wh}{xw + yh + z} \quad h \mapsto \frac{wh}{xw + yh + z}$$

De plus, $\text{width}(u_n) > n$ et $\text{height}(u_n) > n$ pour tout entier n , donc la relation suivante est vérifiée pour des valeurs de n arbitrairement grandes.

$$\frac{n^2}{(x + y)n + z} < \frac{|u_n|}{x \text{width}(u_n) + y \text{height}(u_n) + z}$$

Le résultat suit par le fait que la fonction $n \mapsto n^2/(n \times (x + y) + z)$ n'a pas de borne supérieure. \square

Lemme 7.7. *Soit \mathbf{w} une configuration quasipériodique multi-échelles et u un bloc de \mathbf{w} dont la fréquence ne converge pas. Alors il existe un réel $\varepsilon > 0$, un réel $t \in [0; 1]$, un ensemble infini \mathcal{V} de facteurs de \mathbf{w} et un ensemble infini \mathcal{Q} de quasipériodes de \mathbf{w} qui satisfont l'une ou l'autre des relations suivantes*

$$\begin{aligned} \forall q \in \mathcal{Q}, \forall v \in \mathcal{V}, f_u(v) < t - \varepsilon < t < f_u(q) \text{ ou alors} \\ \forall q \in \mathcal{Q}, \forall v \in \mathcal{V}, f_u(v) > t + \varepsilon > t > f_u(q), \end{aligned}$$

et les ensembles $\mathcal{L}_{\geq k \times k}(\mathbf{w}) \cap \mathcal{V}$ et $\mathcal{L}_{\geq k \times k}(\mathbf{w}) \cap \mathcal{Q}$ sont non vides pour tout entier k .

Démonstration. La suite $(f_u(\mathbf{w}[(-n, -n), \dots, (+n, +n)]))_{n \in \mathbb{N}}$ prend ses valeurs dans le compact $[0; 1]$. Puisqu'elle ne converge pas, elle possède deux valeurs d'adhérence ℓ_1 et ℓ_2 telles que $\ell_1 < \ell_2$. Posons $t = (\ell_1 + \ell_2)/2$ et $\varepsilon = (\ell_2 - \ell_1)/4$. Définissons

$$\begin{aligned} B^- &= \{v \mid v \in \mathcal{L}(\mathbf{w}) \text{ et } f_u(v) < t\}, \text{ et} \\ B^+ &= \{v \mid v \in \mathcal{L}(\mathbf{w}) \text{ et } f_u(v) \geq t\}. \end{aligned}$$

L'ensemble $\{B^-, B^+\}$ forme une partition de $\mathcal{L}(\mathbf{w})$, donc par le principe des tiroirs, il y a soit une infinité de quasipériodes dans B^- , soit une infinité de quasipériodes dans B^+ . (Si les deux ensembles contiennent une infinité de quasipériodes, nous considérons l'un des deux au choix.) Nous distinguons ces deux cas.

Cas 1. Si B^- contient une infinité de quasipériodes, posons \mathcal{Q} l'ensemble des quasipériodes appartenant à B^- . Il existe une sous-suite de $(\mathbf{w}[(-n, -n), \dots, (+n, +n)])$, disons v_n , telle que $f_u(v_n)$ converge vers $\ell_2 = t + 2\varepsilon$. Il existe donc un entier n_0 tel que, pour tout $n > n_0$, nous avons $f_u(v_n) > t + \varepsilon$. Prenons alors $\mathcal{V} = \{v_n \mid n > n_0\}$.

Cas 2. Symétriquement, si B^+ contient une infinité de quasipériodes, appelons \mathcal{Q} l'ensemble des quasipériodes appartenant à B^+ . Il existe alors une sous-suite de $(\mathbf{w}[(-n, -n), \dots, (+n, +n)])$, disons (v_n) , qui converge vers $\ell_1 = t - 2\varepsilon$. Il existe donc un entier n_0 tel que, pour tout $n > n_0$, nous avons $f_u(v_n) > t - \varepsilon$. Prenons alors $\mathcal{V} = \{v_n \mid n > n_0\}$. \square

Lemme 7.8. *Soit \mathbf{w} une configuration quasipériodique multi-échelles et u un facteur de \mathbf{w} . Supposons qu'il existe un réel $\varepsilon > 0$ et un réel $t \in [0; 1]$ tels que les deux conditions*

$$\begin{aligned} \forall k \in \mathbb{N}, \exists v \in \mathcal{L}_{\geq k \times k}(\mathbf{w}) \text{ tel que } f_u(v) \leq t - \varepsilon, \\ \forall k \in \mathbb{N}, \exists q \in \mathcal{L}_{\geq k \times k}(\mathbf{w}) \cap \mathcal{Q}(\mathbf{w}) \text{ tel que } f_u(q) > t, \end{aligned}$$

sont satisfaites. Alors la relation

$$\forall k \in \mathbb{N}, \exists v' \in \mathcal{L}_{\geq k \times k}(\mathbf{w}) \text{ et } f_u(v') \leq t - \frac{11}{10}\varepsilon$$

est vérifiée également.

Démonstration. Soit k un entier ; nous montrons qu'il existe un bloc v' approprié.

Soit v un élément de $\mathcal{L}(\mathbf{w})$ tel que $f_u(v) \leq t - \varepsilon$ et q une quasipériode de \mathbf{w} telle que $f_u(q) > t$. Posons

$$\begin{aligned} \alpha &= 2 \text{ width}(q) \text{ height}(v) + 2 \text{ height}(q) \text{ width}(v) + 4|q|, \\ \beta &= 2k \times (\text{width}(q) + \text{height}(q)). \end{aligned}$$

En utilisant le lemme 7.6, nous pouvons choisir une quasipériode q qui satisfait les relations suivantes,

$$\text{width}(q) > 10k \quad (7.1)$$

$$\text{height}(q) > 10k \quad (7.2)$$

$$|q| > 10\beta \quad (7.3)$$

$$t|q| > (t - \frac{\varepsilon}{10})(|q| + \beta) \quad (7.4)$$

puis un facteur v qui satisfait la relation

$$\frac{\alpha}{|v|} < \frac{\varepsilon}{10} \quad (7.5)$$

en observant que cette dernière est équivalente à $|v|/\alpha > 10/\varepsilon$.

Soit x un bloc, d'aire minimale, dont la largeur est multiple de $2 \text{width}(q)$ et la hauteur multiple de $2 \text{height}(q)$, tel que v apparaît dans le coin inférieur gauche de x . Comme x est minimal, nous avons $\text{width}(x) - \text{width}(v) < 2 \text{width}(q)$ et $\text{height}(x) - \text{height}(v) < 2 \text{height}(q)$ (cf. figure 7.2), donc $|v| \leq |x| \leq |v| + \alpha$ (par définition de α). En particulier, $|x|_u \leq |v|_u + \alpha$ par le lemme 7.5, et nous pouvons écrire

$$f_u(x) \leq \frac{|v|_u + \alpha}{|x|}$$

en utilisant $|v| \leq |x|$, l'équation (7.5) et $f_u(v) < t - \varepsilon$, cette relation se réécrit

$$f_u(x) \leq t - \frac{9\varepsilon}{10}. \quad (7.6)$$

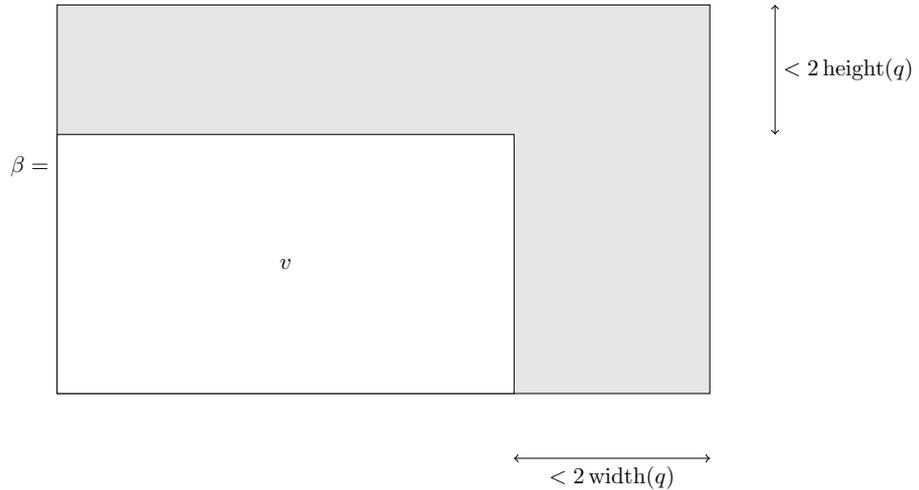


FIG. 7.2 : Anatomie d'un bloc x .

La largeur et la hauteur de x étant des multiples du double de la largeur et de la hauteur de q , nous pouvons découper x en blocs de taille $2 \text{width}(q) \times 2 \text{height}(q)$.

Appelons-les $z_{i,j}$, pour $i \in \{1, \dots, n\}$ et $j \in \{1, \dots, m\}$. Étant donné que $f_u(x) \leq t - \frac{9\varepsilon}{10}$, il existe un bloc y dans l'ensemble des $z_{i,j}$ tel que $f_u(y) \leq t - \frac{9\varepsilon}{10}$. Si tel n'était pas le cas, en appliquant le lemme 7.5, nous obtiendrions

$$f_u(x) \geq \sum_{1 \leq i \leq n} \sum_{1 \leq j \leq m} \frac{|z_{i,j}|}{|x|} f_u(z_{i,j}) > \left(t - \frac{9\varepsilon}{10}\right) \sum_i \sum_j \frac{|z_{i,j}|}{|x|} = t - \frac{9\varepsilon}{10} \geq f_u(x),$$

une contradiction car l'une des inégalités est stricte.

Décomposons y en cinq blocs : une occurrence de q et quatre blocs m_1, m_2, m_3, m_4 , comme montré sur la figure 7.3. Par la définition des m_i et le fait que $\text{width}(q), \text{height}(q) > 10k$, nous avons

$$\begin{aligned} \text{width}(m_1) &\geq \text{width}(q) > 10k \\ \text{width}(m_3) &\geq \text{width}(q) > 10k \\ \text{width}(m_2) + \text{width}(m_4) &= \text{width}(q) > 10k \\ \text{height}(m_2) &\geq \text{height}(q) > 10k \\ \text{height}(m_4) &\geq \text{height}(q) > 10k \\ \text{height}(m_1) + \text{height}(m_3) &= \text{height}(q) > 10k \end{aligned}$$

Nous pourrions potentiellement avoir soit $\text{width}(m_2) < k$, soit $\text{width}(m_4) < k$, mais pas les deux. Si $\min(\text{width}(m_2), \text{width}(m_4)) < k$, posons σ la valeur parmi $\{2, 4\}$ telle que $\text{width}(m_\sigma) < k$; autrement, posons $m_0 = \varepsilon$ et $\sigma = 0$. De la même façon, nous pourrions avoir $\text{height}(m_1) < k$ ou $\text{height}(m_3) < k$, mais pas les deux. Si $\min(\text{height}(m_1), \text{height}(m_3)) < k$, posons τ l'unique valeur parmi $\{1, 3\}$ qui satisfait $\text{height}(m_\tau) < k$; autrement, nous posons $\tau = 0$. Définissons $q' = q \cup m_\sigma \cup m_\tau$.

Observons que $\text{height}(m_\sigma) \leq 2 \text{height}(q)$ et $\text{width}(m_\tau) \leq 2 \text{width}(q)$, donc par définition de β nous avons $|q| \leq |q'| = |q| + |m_\sigma| + |m_\tau| \leq |q| + \beta$. De plus $|q|/|y| = 1/4$, et par l'équation (7.3), $|q| + \beta \leq 11|q|/10$; nous pouvons écrire

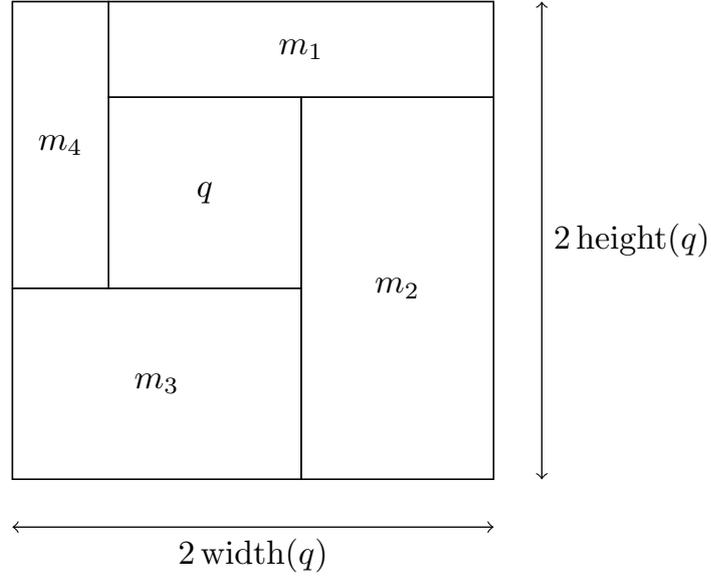
$$\frac{1}{4} \leq \frac{|q'|}{|y|} \leq \frac{11}{40}. \quad (7.7)$$

De plus $|q|_u \leq |q'|_u$, car q' contient q . En utilisant $f_u(q) > t$ et l'équation (7.4), nous pouvons écrire

$$f_u(q') \geq \frac{|q|_u}{|q'|} = f_u(q) \frac{|q|}{|q'|} > t \frac{|q|}{|q| + \beta} > t - \frac{\varepsilon}{10}.$$

Considérons à présent une nouvelle décomposition de y en $q', m'_1, m'_2, m'_3, m'_4$ comme suit : si $\text{width}(m_i) > k$ et $\text{height}(m_i) > k$, alors nous posons $m'_i = m_i$; autrement, m'_i est le bloc vide. Sans perte de généralité, supposons que m'_1 est non vide (il existe toujours au moins deux m'_i non vides) et tel que $f_u(m'_1)$ est minimal parmi les m'_i non vides. Soit $\mu = |m'_1| + |m'_2| + |m'_3| + |m'_4|$. Nous avons $|y| = |q'| + \mu$; en appliquant le lemme 7.5, nous écrivons

$$f_u(y) \geq \frac{|q'|_u + \mu f_u(m'_1)}{|y|} = f_u(q') \frac{|q'|}{|y|} + f_u(m'_1) \frac{\mu}{|y|}$$

FIG. 7.3 : Anatomie du bloc y .

Supposons, en vue d'une contradiction, que $f_u(m'_1) \geq t - \frac{11\varepsilon}{10}$, et réécrivons la relation ci-dessus à l'aide de l'équation (7.7) et de $|y| = |q'| + \mu$ en

$$f_u(y) \geq t \left(\frac{\mu + |q'|}{|y|} \right) - \varepsilon \left(\frac{3}{4} \times \frac{11}{10} + \frac{11}{40} \times \frac{1}{10} \right) = t - \frac{341}{400} \varepsilon$$

Donc $f_u(y) \geq t - \frac{341}{400} \varepsilon > t - \frac{9}{10} \varepsilon \geq f_u(y)$: une contradiction. Nous avons donc $f_u(m'_1) < t - \frac{11}{10} \varepsilon$, et m'_1 convient pour v' . \square

Nous allons prouver un lemme symétrique, qui permet d'obtenir des fréquences « de plus en plus grandes ». Tout d'abord, nous avons besoin d'un résultat technique supplémentaire.

Lemme 7.9. *Soit u un bloc fini, et v un autre bloc fini pouvant être décomposé en des blocs v_1, \dots, v_n disjoints tels que $\text{width}(v_i) > \text{width}(u)$ et $\text{height}(v_i) > \text{height}(u)$ pour tout i entre 1 et n . Alors,*

$$f_u(v) \leq \sum_{i=1}^n \frac{|v_i|_u + \text{width}(v_i) \times \text{height}(u) + \text{height}(v_i) \times \text{width}(u)}{|v|}$$

Démonstration. Chaque occurrence de u dans v possède son coin inférieur gauche dans l'un des v_i . Chaque occurrence est soit entièrement contenue dans v_i , soit déborde. Il y a au plus $|v_i|_u$ et au plus $\text{width}(v_i) \times \text{height}(u) + \text{height}(v_i) \times \text{width}(u)$ occurrences de chaque sorte, d'où la borne. \square

Le lemme suivant est presque symétrique au lemme 7.8, nous ne donnerons donc dans la preuve que les différences par rapport à celui-ci.

Lemme 7.10. *Soit \mathbf{w} une configuration quasipériodique multi-échelles et u un bloc de \mathbf{w} . Supposons qu'il existe en réel $\varepsilon > 0$ et un réel $t \in [0; 1]$ tels que les deux conditions*

$$\begin{aligned} \forall k \in \mathbb{N}, \exists v \in \mathcal{L}_{\geq k \times k}(\mathbf{w}) \text{ tel que } f_u(v) &\geq t + \varepsilon \\ \forall k \in \mathbb{N}, \exists q \in \mathcal{L}_{\geq k \times k}(\mathbf{w}) \cap \mathcal{Q}(\mathbf{w}) \text{ tel que } f_u(q) &< t \end{aligned}$$

sont satisfaites, alors la relation

$$\forall k \in \mathbb{N}, \exists v' \in \mathcal{L}_{\geq k \times k}(\mathbf{w}) \text{ et } f_u(v') \geq t + \frac{11}{10}\varepsilon$$

est vérifiée.

Démonstration. Soient q une quasipériode et v un facteur de \mathbf{w} vérifiant $f_u(q) < t$ et $f_u(v) \geq t + \varepsilon$. Posons

$$\begin{aligned} \alpha &= 2(\text{width}(q) \times \text{height}(v) + \text{height}(q) \times \text{width}(v)) + 4|q| \\ \beta &= 2k \times (\text{width}(q) + \text{height}(q)) \\ \gamma &= 2(\text{width}(q) \text{height}(u) + \text{height}(q) \text{width}(u)) \\ \delta &= 3(\text{width}(q) \times \text{height}(u) + \text{height}(q) \times \text{width}(u)) \end{aligned}$$

Grâce au lemme 7.6, nous pouvons choisir q assez grand pour vérifier les relations suivantes,

$$\text{width}(q) > 10k \tag{7.8}$$

$$\text{height}(q) > 10k \tag{7.9}$$

$$\frac{\beta}{|q|} < \frac{\varepsilon}{10} \tag{7.10}$$

$$\frac{\gamma}{4|q|} < \frac{\varepsilon}{1000} \tag{7.11}$$

$$\frac{\delta}{4|q|} < \frac{\varepsilon}{400} \tag{7.12}$$

puis v assez grand pour vérifier les relations suivantes,

$$\text{width}(v) > 2 \text{width}(q) \tag{7.13}$$

$$\text{height}(v) > 2 \text{height}(q) \tag{7.14}$$

$$(t + \varepsilon) \frac{|v|}{|v| + \alpha} > t + \frac{9\varepsilon}{10} \tag{7.15}$$

(Il faut observer que la relation (7.15) est équivalente à $|v|/\alpha > (t + 9\varepsilon/10)/(\varepsilon/10)$.)

Considérons x un bloc d'aire minimale, de largeur multiple de $2 \text{width}(q)$ et de hauteur multiple de $2 \text{height}(q)$, tel que v apparaît dans le coin inférieur gauche de x . Comme $|x| \leq |v| + \alpha$, nous avons $|x| < 4|v|$. De plus, $v \subseteq x$, donc $|v|_u \leq |x|_u$, et la relation (7.15) permet d'écrire

$$f_u(x) \geq \frac{|v|_u}{|x|} \geq (t + \varepsilon) \frac{|v|}{|v| + \alpha} > t + \frac{9\varepsilon}{10}.$$

Décomposons x en blocs $z_{i,j}$, tels que $\text{width}(z_{i,j}) = 2 \text{width}(q)$ et $\text{height}(z_{i,j}) = 2 \text{height}(q)$ pour $1 \leq i \leq n$ et $1 \leq j \leq m$. En particulier, $nm = |x|/(4|q|)$. Étant donné que $f_u(x) > t + \frac{9\varepsilon}{10}$, il existe des entiers i, j tels que $f(z_{i,j}) > t + \frac{89\varepsilon}{100}$. Si tel n'était pas le cas, le lemme 7.9 permettrait d'écrire

$$\begin{aligned} f_u(x) &\leq \sum_{i=1}^n \sum_{j=1}^m \frac{|z_{i,j}|_u + \gamma}{|x|} \\ &\leq t + \frac{89}{100}\varepsilon + \sum_{i=1}^n \sum_{j=1}^m \frac{\gamma}{|x|} \\ &= t + \frac{89}{100}\varepsilon + \left(\frac{|x|}{4|q|} \times \frac{\gamma}{|x|} \right) \end{aligned}$$

et la relation (7.11) permet de conclure que $f_u(x) < t + \frac{9}{10}\varepsilon$, une contradiction. Posons donc y un bloc parmi les $z_{i,j}$ tel que $f_u(y) > t + \frac{89}{100}\varepsilon$, décomposons y en q, m_1, m_2, m_3, m_4 comme dans la preuve du lemme 7.8, et définissons q' comme dans la preuve de ce même lemme. Étant donné que $|q'| \geq |q|$, le lemme 7.5, la relation $f_u(q) < t$ et l'équation (7.10) permettent d'établir que

$$f_u(q') < \frac{|q|_u}{|q'|} + \frac{\beta}{|q'|} < t + \frac{\varepsilon}{10}.$$

De plus, $1/4 \leq |q'|/|y| \leq 11/40$.

Décomposons y en $q', m'_1, m'_2, m'_3, m'_4$ de la même façon que dans le lemme 7.8. Soit $\mu = |m'_1| + |m'_2| + |m'_3| + |m'_4|$. Nous avons $|y| = 4|q| = |q'| + \mu$. Sans perte de généralité, supposons que m'_1 est non vide et maximal pour f_u parmi les m'_i .

Montrons maintenant que,

$$|y|_u \leq |q'|_u + |m'_1|_u + |m'_2|_u + |m'_3|_u + |m'_4|_u + \delta \quad (7.16)$$

c'est-à-dire que δ est un majorant du nombre d'occurrences de u dans y qui s'étalent sur plusieurs composantes. En premier lieu, voyons y décomposé sur $\{q, m_1, m_2, m_3, m_4\}$, comme sur la figure 7.3. La longueur totale des frontières horizontales est $\text{width}(m_2) + \text{width}(m_4) + 2 \text{width}(q)$, et la longueur totale des frontières verticales est $\text{height}(m_1) + \text{height}(m_3) + 2 \text{height}(q)$; comme $\text{width}(m_2) + \text{width}(m_4) = \text{width}(q)$, et $\text{height}(m_1) + \text{height}(m_3) = \text{height}(q)$, ces longueurs des frontières sont respectivement bornées par $3 \text{width}(q)$ et $3 \text{height}(q)$. Ainsi, le nombre total d'occurrences de u à cheval sur deux composantes est borné par $3 \times (\text{height}(u) \times \text{width}(q) + \text{width}(u) \times \text{height}(q))$. La décomposition $\{q', m'_1, m'_2, m'_3, m'_4\}$ est produite à partir de $\{q, m_1, m_2, m_3, m_4\}$ en fusionnant des composantes, la longueur des frontières est donc inférieure ou égale. Le nombre d'occurrences de u à cheval est donc, lui aussi, inférieur ou égal, ce qui prouve la relation 7.16.

Nous pouvons donc écrire

$$f_u(y) \leq f_u(m'_1) \times \frac{\mu}{|y|} + f_u(q') \times \frac{|q'|}{|y|} + \frac{\delta}{|y|}$$

Supposons, par contradiction, que $f_u(m'_1) \leq t + \frac{11\varepsilon}{10}$. Alors,

$$t + \frac{89\varepsilon}{100} < f_u(y) < (t + \frac{11}{10}\varepsilon) \times \frac{\mu}{|y|} + (t + \frac{1}{10}\varepsilon) \times \frac{|q'|}{|y|} + \frac{\delta}{|y|}$$

d'où l'on tire

$$t + \frac{89}{100}\varepsilon < f_u(y) < t + \frac{342\varepsilon}{400} < t + \frac{89}{100}\varepsilon,$$

une contradiction. Nous avons donc $f_u(m'_1) > t + \frac{11\varepsilon}{10}$; posons $v' = m'_1$ et le lemme est démontré. \square

Nous sommes maintenant prêts pour la preuve du théorème 7.4.

Démonstration du théorème 7.4. Supposons que \mathbf{w} soit une configuration quasipériodique multi-échelles et que u soit un facteur de \mathbf{w} dont la fréquence ne converge pas. Par le lemme 7.7, il existe des réels $t \in [0; 1]$ et $\varepsilon > 0$, un ensemble infini de blocs \mathcal{V} et un ensemble infini de blocs \mathcal{Q} tels que :

- soit $\forall v \in \mathcal{V}, \forall q \in \mathcal{Q}, f_u(v) > t + \varepsilon > t > f_u(q)$;
- soit $\forall v \in \mathcal{V}, \forall q \in \mathcal{Q}, f_u(v) < t < t - \varepsilon < f_u(q)$;

de plus, $\mathcal{V} \cap L_{\geq k \times k}(\mathbf{w})$ et $\mathcal{Q} \cap \mathcal{L}_{\geq k \times k}(\mathbf{w})$ sont non vides pour tout k .

Cas 1 : $f_u(v) < t < t - \varepsilon < f_u(q)$. Le lemme 7.8 affirme qu'il existe un ensemble infini de blocs \mathcal{V}' , avec les mêmes propriétés que v , tel que $\forall v' \in \mathcal{V}', f_u(v') < t - \frac{11}{10}\varepsilon$. En appliquant le lemme de nouveau à \mathcal{V}' , et ainsi de suite, nous obtenons des facteurs dont les fréquences tendent vers $-\infty$, donc en particulier sont inférieures à 0 : une contradiction.

Cas 2 : $f_u(v) > t + \varepsilon > t > f_u(q)$. Le lemme 7.10 affirme qu'il existe un ensemble infini de blocs \mathcal{V}' , avec les mêmes propriétés que v , tel que $\forall v' \in \mathcal{V}', f_u(v') > t + \frac{11}{10}\varepsilon$. En appliquant le lemme de nouveau à \mathcal{V}' , et ainsi de suite, nous obtenons des facteurs dont les fréquences tendent vers $+\infty$, donc en particulier sont supérieures à 1 : une contradiction. \square

A comathematician is a device turning cotheorems into ffee.

— Folklore

Chapitre 8

Perspectives

Ce court chapitre conclut la thèse en récapitulant des problèmes laissés ouverts après les travaux présentés dans ce manuscrit. Certains ont déjà été signalés au cours du texte, d'autres non.

8.1 Problèmes d'algorithmique

Dans le chapitre 3, nous avons fourni une méthode d'étude des quasipériodes d'un mot fini arbitraire. Cette méthode n'est pas à proprement parler un algorithme, puisque les objets que nous manipulons sont, par essence, infinis. Toutefois, certains mots infinis admettent des représentations finies, par exemple les mots morphiques. La question suivante est donc naturelle : existe-t-il un algorithme qui, étant donné un mot morphique \mathbf{w} et un mot fini q , teste si q est une quasipériode de \mathbf{w} ?

Si un mot infini est défini par $\mathbf{w} = f(g^\omega(a))$, pour une lettre a et des morphismes f, g non effaçants, il est possible de calculer l'ensemble des facteurs d'une longueur donnée de \mathbf{w} en temps fini. En calculant les facteurs de longueur $|q| + 2$, nous obtenons le graphe de Rauzy $\mathcal{G}_{\mathbf{w}}(|q| + 1)$, grâce auquel il est possible de tester si q est quasipériode de \mathbf{w} ou non. Il est donc possible d'imaginer un algorithme qui teste si un mot fini est quasipériode d'un mot morphique, ou de façon équivalente, d'énumérer les quasipériodes d'un mot morphique par ordre croissant de longueur. En poussant l'idée plus loin, une possibilité serait d'automatiser le travail effectué dans le chapitre 3. Par exemple, si les problèmes suivants s'avèrent décidables, une partie de l'étude du chapitre 3 pourrait être automatisée pour certaines familles de mots infinis.

Problème 1. Entrée : un mot morphique $\mathbf{w} = f(g^\omega(a))$, un morphisme h et un mot fini x . Sortie : booléen qui indique si les mots de la forme $h^n(x)$ sont tous des quasipériodes de \mathbf{w} .

Problème 2. Entrée : un mot morphique $\mathbf{w} = f(g^\omega(a))$, un morphisme h et un mot fini x . Sortie : booléen qui indique si le langage $\{h^n(x) \mid n \in \mathbb{N}\}$ est exactement l'ensemble des quasipériodes de \mathbf{w} .

Le problème 2 semble difficile à résoudre algorithmiquement, car il nécessite une stratégie pour garantir que le mot \mathbf{w} ne possède *aucune* quasipériode hors de celles demandées. Nous ne disposons pas de stratégies automatisables à l'heure actuelle; toutefois, la réduction du problème 2 au problème de correspondance de Post (ou au problème de l'arrêt) semble également malaisée.

Ces questions sont possiblement plus simples si l'on se restreint aux mots purement morphiques. Toutes sortes de variantes peuvent être imaginées, à partir de toute représentation finie d'un mot infini ou d'un langage.

Le problème suivant est, en quelque sorte, la réciproque du problème 2 dans le contexte des automates finis, au lieu des morphismes. Un automate fini \mathcal{A} *accepte* un mot infini \mathbf{w} si et seulement si \mathbf{w} est l'étiquette d'un chemin infini de \mathcal{A} .

Problème 3. Entrée : un automate fini \mathcal{A} et un mot fini q . Sortie : un booléen qui indique si tout mot infini accepté par \mathcal{A} possède une quasipériode q ?

Ce problème est en lien avec des problèmes ouverts dans [57] au sujet des morphismes quasipériodiques : un morphisme est *faiblement quasipériodique* s'il envoie au moins un mot infini non quasipériodique sur un mot quasipériodique; un morphisme est *fortement quasipériodique* s'il envoie tout mot infini sur un mot quasipériodique. La question concerne la caractérisation des morphismes faiblement quasipériodiques.

8.2 Sur les mots biinfinis quasipériodiques

Dans le chapitre 4, nous avons caractérisé les mots périodiques et les mots sturmiens avec des conditions sur l'ensemble de leurs quasipériodes. Afin de généraliser ces caractérisations, nous avons entrepris l'étude des mots biinfinis quasipériodiques. Ce domaine laisse place à de nombreuses questions non résolues.

La question principale ici est de prouver ou réfuter l'affirmation suivante : *un mot biinfini \mathbf{w} est sturmien si et seulement si les intervalles d'entiers où sa fonction de quasipériodicité s'annule sont tous de longueur 1, sauf le premier*. Pour cela, la première étape est d'étudier la structure des quasipériodes des mots biinfinis; l'obstacle principal ici est l'existence de mots biinfinis possédant deux quasipériodes spéciales droites de même longueur, ce qui invalide de nombreuses tentatives de raisonnements combinatoires pour résoudre cette question. Cela ne rend pas la conjecture fautive pour autant, et des travaux se poursuivent pour tenter de la résoudre, encore à l'heure où ces lignes sont rédigées.

Nous avons également mentionné que les mots biinfinis quasipériodiques multi-échelles ne sont pas tous uniformément récurrents; il serait intéressant de *caractériser ceux qui en le sont pas*, afin de pouvoir les exclure des études futures. Une partie d'entre eux sont ultimement périodiques, de la forme :

...babababa ababababa...

Toutefois, les mots multi-échelles non uniformément récurrents ne sont pas tous ultimement périodiques. Considérons par exemple le mot \mathbf{w} , défini comme la limite de la

suite :

$$\begin{aligned} q_0 &= aabaa \\ q_1 &= a^{-1}q_0a^{-2}q_0a^{-1}q_0 \\ q_{n+1} &= q_n(a^{-1}a_n)(a^{-2}q_n)(a^{-1})q_n \end{aligned}$$

Le symétrisé de \mathbf{w} , c'est-à-dire le mot biinfini \mathbf{w}' satisfaisant $\mathbf{w}'(-n-1) = \mathbf{w}'(n) = \mathbf{w}(n)$ pour tout entier naturel n , n'est pas uniformément récurrent : le facteur $aaaa$ n'apparaît qu'en position -2 . Toutefois, un raisonnement par récurrence montre que q_{n+1} est q_n -quasipériodique pour tout n ; un autre montre que q_n est un palindrome pour tout n . En conséquence, q_n est une quasipériode de \mathbf{w}' pour tout n .

8.3 Définir des langages par des quasipériodes

Il existe une généralisation naturelle de la quasipériodicité qui n'a pas été, à notre connaissance, étudiée dans la littérature. Un mot infini \mathbf{w} pourrait avoir deux quasipériodes, q_1 et q_2 , au sens suivant : chaque position de \mathbf{w} est couverte soit par q_1 , soit par q_2 au choix. Il s'agit en quelque sorte d'une disjonction de quasipériodes. Rien ne nous force d'ailleurs à nous limiter à deux : un mot infini pourrait avoir un ensemble fini arbitraire de quasipériodes. Dans ce cas, nous pouvons exiger sans surcoût que toutes les quasipériodes aient même longueur.

L'analogie de la variante multi-échelles, dans ce contexte, est comme suit : il existe une infinité d'entiers u_0, u_1, \dots pour lesquels \mathbf{w} possède un ensemble de quasipériodes (au sens du paragraphe précédent) de longueur u_i . Nous nous rapprochons ici de la notion de « facteurs interdits » utilisée pour définir les systèmes symboliques (voir théorème 4.2). Il s'agit, en quelque sorte, d'une notion de « facteurs autorisés » qui n'est pas exactement le complémentaire de la précédente.

Sébastien Ferenczi et Thierry Monteil ont étudié une version plus générale de ce concept : la K -déconnectabilité (voir [32]). Un mot (ou un système symbolique) \mathbf{w} est K -déconnectable s'il existe une infinité d'entiers u_0, u_1, \dots et un réel α tels que, pour tout entier i , il existe un ensemble de K quasipériodes (au sens des paragraphes précédents) de longueur u_i dans \mathbf{w} , excepté que les quasipériodes ont le droit d'être espacées de αu_i lettres. Autrement dit, la notion est analogue à la quasipériodicité multi-échelles du paragraphe précédent, excepté que les occurrences des quasipériodes ont le droit d'être légèrement espacées. La terminologie « déconnectable » provient de l'interprétation de cette propriété sur les graphes de Rauzy de \mathbf{w} . Ferenczi et Monteil ont démontré le théorème suivant.

Théorème 8.1 (Ferenczi, Monteil, 2010 [32]). *Si un système symbolique (X, σ) est K -déconnectable, alors il admet au plus K mesures ergodiques.*

Sans rentrer dans le détail des mesures ergodiques, disons simplement que si \mathbf{w} est un mot infini K -déconnectable, alors pour tout facteur u de \mathbf{w} , la suite dont la limite définit la fréquence de u dans \mathbf{w} possède au plus K valeurs d'adhérence. Les mesures

ergodiques sont un objet d'intérêt dans l'étude des systèmes dynamiques généraux ; il est donc utile de pouvoir les étudier sur les systèmes symboliques.

Il serait intéressant d'avoir une méthode, ou même un algorithme, permettant de passer d'une représentation en facteurs interdits à une représentation en quasipériodes généralisées (au sens des paragraphes précédents). Nous pourrions ensuite l'utiliser pour étudier les mesures ergodiques de nombreux systèmes symboliques, et plus généralement utiliser tout résultat concernant les quasipériodes généralisées sur n'importe quel système symbolique.

8.4 Questions bidimensionnelles

L'une des questions les plus étudiées au cours de la préparation de cette thèse, mais néanmoins toujours ouverte à l'heure où ces lignes sont rédigées, est celle de la forme normale des configurations de \mathbb{Z}^2 quasipériodiques. En ayant connaissance des travaux sur la quasipériodicité des mots infinis, il s'agit de la première question que nous pouvons nous poser en démarrant l'étude sur \mathbb{Z}^2 . La décomposition sur \mathbb{Z}^2 n'a aucune raison de se faire sur des blocs rectangulaires et la notion d'antibords ne se généralise pas bien en deux dimensions. Il est possible de construire des configurations quasipériodiques avec de « mauvaises » propriétés en emboîtant des occurrences d'une quasipériode, comme nous le montrons au chapitre 6. Mais toutes les configurations ne sont pas capturées par cette construction, loin s'en faut.

La forme normale pourrait nous aider par exemple à déterminer si les codages de rotations en deux dimensions sont tous quasipériodiques — les codages de rotations faisant figure de généralisation des mots sturmiens en deux dimensions. Cette question est la généralisation de celle de Marcus [67], dont la réponse serait une généralisation du chapitre 4. Toutefois, les problèmes rencontrés lors du passage de \mathbb{N} à \mathbb{Z} risquent de s'accroître lors du passage de \mathbb{Z} à \mathbb{Z}^2 .

Une autre question fut laissée en suspens au chapitre 6 : quelles sont les quasipériodes rectangulaires qui forcent l'entropie topologique de leurs configurations à être nulle ? Nous en avons un sous-ensemble (celles dont les bords sont tous dans la même diagonale, théorème 6.6), mais y en a-t-il d'autres ? Il s'agit d'une information importante, car nous observons ici une forte divergence entre 1D et 2D. En une dimension, les quasipériodes forcent soit la périodicité, soit rien. En deux dimensions, certaines quasipériodes forcent l'entropie nulle, mais pas la convergence de la fréquence des facteurs ni l'uniforme récurrence. Nous nous demandons donc, plus généralement, quelles sont les (combinaisons de) propriétés qui peuvent — ou non — être forcées par des quasipériodes.

Ceci nous mène à voir la quasipériodicité comme une règle locale, à la façon de ce que nous évoquions pour les mots infinis dans la section 8.3. Nous pourrions envisager des liens avec la théorie des pavages à la Wang. Trouver, étant donné une quasipériode q , un jeu de tuiles dont les pavages sont en bijection avec les quasipériodes, par une bijection préservant dans une certaine mesure la structure combinatoire des objets considérés.

8.5 Autres résultats

Pour finir, l'étude de la quasipériodicité permet de développer un savoir-faire combinatoire qui s'applique ensuite à d'autres champs. Nous mentionnons très brièvement quelques résultats qui furent développés au cours de la préparation de cette thèse, mais sans lien avec la quasipériodicité.

En premier lieu, l'une des motivations à l'étude des quasipériodes en deux dimensions était la recherche d'une meilleure compréhension du jeu de tuiles de Kari-Culik. Bruno Durand, Anaël Grandjean et l'auteur ont montré que le système engendré par le jeu de tuiles de Kari-Culik était d'entropie topologique strictement positive [21]. Plus récemment, les mêmes ont montré que l'entropie demeure positive lorsque certaines restrictions sont rajoutées aux pavages considérés. Notamment, certains blocs spécifiques, servant de *gadgets* dans la preuve d'entropie positive initiale, étaient interdits. [22]

Le deuxième résultat concerne la combinatoire des mots. Un *motif* est un mot dont chaque lettre doit être comprise comme une variable désignant un mot fini. Par exemple, le motif uu désigne un carré; le motif uvu désigne un mot avec un bord, et $uvvu$ désigne un chevauchement. Un mot infini \mathbf{w} peut *éviter* un motif, ce qui signifie que nul facteur de \mathbf{w} ne correspond au motif considéré. Par exemple, le mot de Thue-Morse évite uuu et $uvvu$. Il existe une famille de motifs, appelés les *formules circulaires*, dont Pascal Ochem, Gwenaël Richomme, Patrice Séébold et l'auteur ont montré qu'elle était évitable. Le mot infini les évitant fut construit comme un mot morphique, en partie grâce à une recherche exhaustive informatique [39].

La perspective ne permet pas aux aveugles de voir.

Georg Christoph Lichtenberg, Aphorismes

Bibliographie

- [1] Jean-Paul ALLOUCHE and Jeffrey SHALLIT. *Automatic Sequences: Theory, Applications, Generalizations*. Cambridge University Press, 2003.
- [2] Marcella ANSELMO, Dora GIAMMARRESI, and Maria MADONIA. “Nonexpandable Nonoverlapping Sets of Pictures”. In : *Theor. Comput. Sci.* 657 (2017), pp. 127–136.
- [3] Alberto APOSTOLICO and Maxime CROCHEMORE. “String pattern matching for a deluge survival kit”. In : *Handbook of Massive Data Sets*. Ed. by J. ABELLO, P.M. PARDALOS, and M.G.C. RESENDE. Kluwer Academic Publishers, 2001.
- [4] Alberto APOSTOLICO and Andrzej EHRENFEUCHT. “Efficient detection of quasiperiodicities in strings”. In : *Theor. Comput. Sci.* 119 (1993), pp. 247–265.
- [5] Nicolas BACQUEY. “Primitive roots of bi-periodic infinite pictures”. In : *Words 2015*. Words 2015, Local Proceedings. Kiel, Germany, Sept. 2015.
- [6] Robert BERGER. “The Undecidability of the Domino Problem”. PhD thesis. Harvard University, 1964. See also the 66th Memoir of the American Mathematical Society, same author and title, 1966.
- [7] Valérie BERTHÉ and Michel RIGO, eds. *Combinatorics, Automata and Number Theory*. Vol. 135. Encyclopedia of Mathematics and Its Applications. Cambridge University Press, 2010.
- [8] Valérie BERTHÉ and Michel RIGO, eds. *Combinatorics, Words and Symbolic Dynamics*. Vol. 159. Encyclopedia of Mathematics and Its Applications. Cambridge University Press, 2016.
- [9] Dany BRESLAUER. “An on-line string superprimitivity test”. In : *Inf. Process. Lett.* 44.6 (1992), pp. 345–347.
- [10] Dany BRESLAUER. “Testing String Superprimitivity in Parallel”. In : *Inf. Process. Lett.* 49.5 (1994), pp. 235–241.
- [11] Jean-Yves BÉZIAU. “Musical Hexagon”. Slides du Séminaire MaMuPhi du 5 novembre 2011, École Normale Supérieure de Paris. 2011.
- [12] Julien CASSAIGNE and François NICOLAS. “Factor complexity”. In : ed. by Valérie BERTHÉ and Michel RIGO. Vol. 135. Encyclopedia of Mathematics and Its Applications. Cambridge University Press, 2010, pp. 179–261.

- [13] Michalis CHRISTOU, Maxime CROCHEMORE, and Costas S. ILIOPOULOS. “On left and right seeds of a string”. In : *J. Discrete Algorithms* 17 (2012), pp. 31–44.
- [14] Michalis CHRISTOU, Maxime CROCHEMORE, Costas S. ILIOPOULOS, Marcin KUBICA, Solon P. PISSIS, Jakub RADOSZEWSKI, Wojciech RYTTER, Bartosz SZREDER, and Tomasz WALÈN. “Efficient seed computation revisited”. In : *Theor. Comput. Sci.* 483 (2013), pp. 181–181.
- [15] Maxime CROCHEMORE, Costas S. ILIOPOULOS, and Maureen KORDA. “Two-Dimensional Prefix String Matching and Covering on Square Matrices”. In : *Algorithmica* 20.4 (1998), pp. 353–373.
- [16] Karel CULÍK. “An aperiodic set of 13 Wang tiles”. In : *Discrete Math.* 160.1-3 (1996), pp. 245–251.
- [17] Van CYR and Bryna KRA. “Nonexpansive \mathbb{Z}^2 -subdynamics and Nivat’s conjecture”. In : *Trans. Amer. Math. Soc.* 367 (2015), pp. 6487–6537.
- [18] Erik D. DEMAINE and Martin L. DEMAINE. “Jigsaw Puzzles, Edge Matching, and Polyomino Packing: Connections and Complexity”. In : *Graphs and Combinatorics* 23 (2007). (Supplement-1), pp. 195–208.
- [19] Julien DESTOMBES. *Mémoire de stage de master*. 2016.
- [20] Bruno DURAND. “Tilings and Quasiperiodicity”. In : *Theor. Comput. Sci.* 221.1-2 (1999), pp. 61–75.
- [21] Bruno DURAND, Guilhem GAMARD, and Anaël GRANDJEAN. “Aperiodic Tilings and Entropy”. In : *Developments in Language Theory - 18th International Conference, DLT 2014, Ekaterinburg, Russia, August 26-29, 2014. Proceedings.* 2014, pp. 166–177.
- [22] Bruno DURAND, Guilhem GAMARD, and Anaël GRANDJEAN. “Aperiodic tilings and entropy”. In : *Theor. Comput. Sci.* 666 (2017), pp. 36–47.
- [23] Bruno DURAND, Leonid A. LEVIN, and Alexander SHEN. “Complex tilings”. In : *J. Symb. Log.* 73.2 (2008), pp. 593–613.
- [24] Bruno DURAND, Leonid A. LEVIN, and Alexander SHEN. “Local rules and global order, or aperiodic tilings”. In : *Math. Intelligencer* 27.1 (2004), pp. 64–68.
- [25] Bruno DURAND, Andrei ROMASHCHENKO, and Alexander SHEN. *Fixed Point and Aperiodic Tilings*. Tech. rep. TR08-030. ECCO, 2008.
- [26] Bruno DURAND, Andrei ROMASHCHENKO, and Alexander SHEN. “Fixed-point tile sets and their applications”. In : *J. Comput. System Sci.* (2012).
- [27] Fabien DURAND. “A Characterization of Substitutive Sequences Using Return Words”. In : *Discrete Math.* 179 (1998), pp. 89–101.
- [28] Fabien DURAND. *Récurrence en Dynamique Topologique d’Entropie Nulle*. Habilitation à diriger des recherches. Université de Picardie Jules Verne, 2004. (Version longue).

- [29] Art M. DUVAL and William F. SMYTH. “Covering a circular string with substring of fixed length”. In : *Int. J. Found. Comput. Sci.* 7.1 (1996), pp. 87–93.
- [30] Max EUWE. “Mengentheoretische Betrachtungen über das Schachspiel”. In : *Proc. Konin. Akad. Wetenschappen* 32.5 (1929), pp. 633–642.
- [31] Michael FEKETE. “Über die Verteilung der Wurzeln bei gewissen algebraischen Gleichungen mit ganzzahligen Koeffizienten”. In : *Math. Z.* 17.1 (1923), pp. 228–249.
- [32] Sébastien FERENCZI and Thierry MONTEIL. “Infinite words with uniform frequencies, and invariant measures”. In : ed. by Valérie BERTHÉ and Michel RIGO. Vol. 135. *Encyclopedia of Mathematics and Its Applications*. Cambridge University Press, 2010, pp. 387–423.
- [33] Nathan J. FINE and Herbert S. WILF. “Uniqueness theorems for periodic functions”. In : *Proc. Amer. Math. Soc.* 16.1 (1965), pp. 109–109.
- [34] Tomàs FLOURI, Costas S. ILIOPOULOS, Tomasz KOCIUMAKA, Solon P. PISSIS, Simon J. PUGLISI, William F. SMYTH, and Wojciech TYCZYŃSKI. “New and Efficient Approaches to the Quasiperiodic Characterisation of a String”. In : *PSC*. 2012.
- [35] N. Pytheas FOGG. *Substitutions in dynamics, arithmetics and combinatorics*. Lecture Notes in Mathematics 1794. Édité par Valérie Berthé, Sébastien Ferenczi, Christian Mauduit et Anne Siegel. Springer-Verlag, 2002.
- [36] Guilhem GAMARD and Gwenaël RICHOMME. “Coverability and multi-scale coverability on infinite pictures”. In : *J. Comput. System Sci.* (2017). arXiv:1506.08375, under review after minor modifications.
- [37] Guilhem GAMARD and Gwenaël RICHOMME. “Determining Sets of Quasiperiods of Infinite Words”. In : *41st International Symposium on Mathematical Foundations of Computer Science, MFCS 2016, August 22-26, 2016 - Kraków, Poland*. 2016, 40:1–40:13.
- [38] Guilhem GAMARD and Gwenaël RICHOMME. “Coverability in Two Dimensions”. In : *Language and Automata Theory and Applications — 9th International Conference, LATA 2015, Nice, France, March 2-6, 2015, Proceedings*. 2015, pp. 402–413.
- [39] Guilhem GAMARD, Pascal OCHEM, Gwenaël RICHOMME, and Patrice SÉÉBOLD. “Avoidability of Circular Formulas”. In : *Theor. Comput. Sci* (2016), submitted.
- [40] Guilhem GAMARD, Gwenaël RICHOMME, Jeffrey SHALLIT, and Taylor J. SMITH. “Periodicity in rectangular arrays”. In : *Inf. Process. Lett.* 118 (2017), pp. 58–63.
- [41] Amy GLEN, Florence LEVÉ, and Gwenaël RICHOMME. “Quasiperiodic and Lyndon episturmian words”. In : *Theor. Comput. Sci.* 409.3 (2008), pp. 578–600.
- [42] Richard GROULT and Gwenaël RICHOMME. “Optimality of some algorithms to detect quasiperiodicities”. In : *Theor. Comput. Sci.* 411.34–36 (2010), pp. 3110–3122.

- [43] Arie HORDIJK and D. A. van der LAAN. “Bounds for Deterministic Periodic Routing Sequences”. In : *Integer Programming and Combinatorial Optimization, 8th International IPCO Conference, Utrecht, The Netherlands, June 13-15, 2001, Proceedings*. 2001, pp. 236–250.
- [44] Mari HUOVA. “A note on defect theorems for 2-dimensional words and trees”. In : *J. Automata Lang. Combin.* 14 (2009), pp. 203–209.
- [45] Costas S. ILIOPOULOS and Laurent MOUCHARD. “An $O(n \log n)$ Algorithm for Computing all Maximal Quasiperiodicities in Strings”. In : *Combinatorics, Computation and Logic. Proceedings of DMTCs’99 and CATS’99*. 1999, pp. 262–272.
- [46] Costas S. ILIOPOULOS and Laurent MOUCHARD. “Quasiperiodicity and string covering”. In : *Theor. Comput. Sci.* 218.1 (1999), pp. 205–216.
- [47] Costas S. ILIOPOULOS and Laurent MOUCHARD. “Quasiperiodicity: from detection to normal forms”. In : *J. Automata Lang. Combin.* 4.3 (1999), pp. 213–228.
- [48] Costas S. ILIOPOULOS and K PARK. “An optimal $O(\log \log n)$ -time algorithm for parallel superprimitivity testing”. In : *j. k. i. s. s.* 21.8 (1994), pp. 1400–1404.
- [49] Emmanuel JEANDEL and Michael RAO. *An aperiodic set of 11 Wang tiles*. ArXiv 1506.06492. 2015.
- [50] Jarkko KARI. “A small aperiodic set of Wang tiles”. In : *Discrete Mathematics* 160.1-3 (1996), pp. 259–264.
- [51] Jarkko KARI and Michal SZABADOS. “An Algebraic Geometric Approach to Nivat’s Conjecture”. In : *Automata, Languages, and Programming - 42nd International Colloquium, ICALP 2015, Kyoto, Japan, July 6-10, 2015, Proceedings, Part II*. Ed. by Magnús M. HALLDÓRSSON, Kazuo IWAMA, Naoki KOBAYASHI, and Bettina SPECKMANN. Vol. 9135. Lecture Notes in Computer Science. Springer, 2015, pp. 273–285. ISBN : 978-3-662-47665-9.
- [52] Donald E. KNUTH. *Sorting and Searching*. Vol. 3. The Art of Computer Programming. Second edition. Addison-Wesley, 1998.
- [53] Tomasz KOCIUMAKA, Marcin KUBICA, Jakub RADOSZEWSKI, Wojciech RYTTER, and Tomasz WALEN. “A linear time algorithm for seeds computation”. In : *Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2012, Kyoto, Japan, January 17–19, 2012*. Ed. by Yuval RABANI. 2012, pp. 1095–1112.
- [54] Dénes KÖNIG. “Über eine Schlussweise aus dem Endlichen ins Unendliche”. In : *Acta Sci. Math.* 3.2-3 (1927), pp. 121–130.
- [55] Ares LAGAE. *Wang Tiles in Computer Graphics*. Synthesis Lectures on Computer Graphics and Animation. Morgan & Claypool Publishers, 2009.
- [56] Pierre LECOMTE and Michel RIGO. “Abstract numeration systems”. In : ed. by Valérie BERTHÉ and Michel RIGO. Vol. 135. Encyclopedia of Mathematics and Its Applications. Cambridge University Press, 2010, pp. 123–178.

- [57] Florence LEVÉ and Gwenaél RICHOMME. “On Quasiperiodic Morphisms”. In : *Proceedings of the 9th international conference on Words*. Ed. by J. KARHUMÄKI, A. LEPISTÖ, and L. ZAMBONI. 2013, pp. 181–192.
- [58] Florence LEVÉ and Gwenaél RICHOMME. “Quasiperiodic infinite words: some answers”. In : *Bull. Europ. Assoc. Theoret. Comput. Sci.* 84 (2004), pp. 128–238.
- [59] Florence LEVÉ and Gwenaél RICHOMME. “Quasiperiodic Sturmian words and morphisms”. In : *Theor. Comput. Sci.* 372.1 (2007), pp. 15–25.
- [60] Douglas LIND and Brian MARCUS. *An Introduction to Symbolic Dynamics and Coding*. Cambridge University Press, 1995.
- [61] M. LOTHAIRE. *Algebraic Combinatorics on Words*. Vol. 90. Encyclopedia of Mathematics and its Applications. Cambridge University Press, 2002.
- [62] M. LOTHAIRE. *Applied Combinatorics on Words*. Vol. 105. Encyclopedia of Mathematics and its Applications. Cambridge University Press, 2005.
- [63] M. LOTHAIRE. *Combinatorics on Words*. Vol. 17. Encyclopedia of Mathematics and its Applications. Reprinted in the *Cambridge Mathematical Library*, Cambridge University Press, UK, 1997. Addison-Wesley, 1983.
- [64] Roger LYNDON and Marcel-Paul SCHUTZENBERGER. “The equation $a^M = b^N c^P$ in a free group”. In : *Michigan Math. J.* 9 (1962), pp. 289–298.
- [65] Jean MAIRESSE and Laurent VUILLON. “Asymptotic behavior in a heap model with two pieces”. In : *Theor. Comput. Sci.* 270.1-2 (2002), pp. 525–560.
- [66] Solomon MARCUS. “Bridging Two Hierarchies of Infinite Words”. In : *J.UCS* 8.2 (2002), pp. 292–296.
- [67] Solomon MARCUS. “Quasiperiodic infinite words”. In : *Bull. Eur. Assoc. Theor. Comput. Sci.* 82 (2004), pp. 170–174.
- [68] Thierry MONTEIL and Solomon MARCUS. *Quasiperiodic infinite words: multi-scale case and dynamical properties*. ArXiv:math/0603354v1. 2006.
- [69] Marston MORSE and Gustav A. HEDLUND. “Symbolic Dynamics”. In : *Amer. J. Math.* 60.4 (1938), pp. 815–866.
- [70] Marston MORSE and Gustav A. HEDLUND. “Symbolic Dynamics II Sturmian Trajectories”. In : *Amer. J. Math.* 62.1 (1940), pp. 1–42.
- [71] Laurent MOUCHARD. “Normal forms of quasiperiodic strings”. In : *Theor. Comput. Sci.* 249.2 (2000), pp. 313–324.
- [72] Maurice NIVAT. Invited talk at ICALP, Milan. 1997.
- [73] Jean-Jacques PANSIOT. “A propos d’une conjecture de F. Dejean sur les répétitions dans les mots”. In : *Discrete Appl. Math.* 7 (1984), pp. 297–311.
- [74] Roger PENROSE. “Pentaplexity: A class of Non-Periodic Tilings of the Plane”. In : *Eureka* 39 (1978).

- [75] Ronney POLLEY and Ludwig STAIGER. “Quasiperiods, Subword Complexity and the Smallest Pisot Number”. In : *J. Autom. Lang. Comb.* 21.1-2 (2016), pp. 93–106.
- [76] Raphael M. ROBINSON. “Undecidability and Nonperiodicity for Tilings of the Plane”. In : *Invent. Math.* 12.3 (1971).
- [77] Klaus SCHMIDT. “Multi-Dimensional Symbolic Dynamical Systems”. In : *Codes, Systems, and Graphical Models*. Ed. by Brian MARCUS and Joachim ROSENTHAL. Springer New York, 2001, pp. 67–82.
- [78] Axel THUE. “Über unendliche Zeichenreihen”. In : *Kristiania Videnskapsselskaps Skrifter Klasse I. Mat.-Naturv* 7 (1906), pp. 1–22.
- [79] Robert TIJDEMAN and Josemir W. SANDER. “The complexity of functions on lattices”. In : *Theor. Comput. Sci.* 246.1–2 (2000), pp. 195–225.
- [80] Alan TURING. “On Computable Numbers, with an Application to the Entscheidungsproblem”. In : *Proc. London Math. Soc.* 42 (1937), pp. 230–265.
- [81] Jonathan S. TURNER. “Approximation Algorithms for the Shortest Common Superstring Problem”. In : *Inf. Comput.* 83.1 (1989), pp. 1–20.
- [82] Hao WANG. “Proving theorems by Pattern Recognition II”. In : *Bell Systems technical journal* 40 (1961), pp. 1–41.

*[Face à cet autodafé,] Cohen le barbare était consterné.
 Quiconque menait une existence rude à la belle étoile savait la valeur d'un livre bien
 épais qui faisait au moins une saison de feux de camp quand on arrachait les pages à
 bon escient. Plus d'une vie avait dû son salut par une nuit de neige à une poignée de
 petit bois mouillé et un livre bien sec. Si l'on avait envie de fumer et que l'on ne
 trouvait pas de pipe, un livre faisait toujours l'affaire.
 Cohen savait bien que des gens écrivaient dans des livres. Il avait toujours considéré
 que c'était gâcher du papier en frivolités.*

— Terry Pratchett, *Le huitième sortilège*, 1983,
 traduction française de Patrick Couton, 1993.

Index

- a^{-1} , 17
- accepter, 110
- algorithme, 83, 85
- alphabet, 15
- antibord, 21
 - droit, 21
 - gauche, 21
- bispécial, 29
- bord, 21
 - diagonal, 76
- chevauchement, 17, 47
- concaténation, 16
- conjugaison, 60
- convergence, 30, 60
- coordonnées, 76
- couvrir une position, 17, 29, 76
- cylindre, 60
- D0L, 33
- δ_q , 43
- distance, 60
- entropie
 - topologique, 34
- ergodique
 - uniquement, 39
- étendue, 47
- éviter, 113
- facteur, 17, 28
 - interne, 17
 - propre, 17
- facteurs interdits, 61
- figure, 83
- fonction de quasipériodicité, 68
- forme normale, 21
- formule circulaire, 113
- fortement multi-échelles, 99
- fréquence, 38
- full shift, 60
- Graphe de Rauzy, 30
- HD0L, 33
- langage, 29
- lettre, 15
- longueur, 15
 - de retour, 47
- matrice
 - d'adjacence, 24
 - d'incidence, 39
- monoïde libre, 16
- morphisme, 31
 - de Fibonacci, 31
 - uniforme, 31
- morphisme dérivateur
 - gauche, 43
- morphisme dérivateur
 - droit, 43, 54
- mot
 - infini, 27
- mot sturmien
 - bi-infinis, 69
 - standard, 63
- mot vide, 15
- mot biinfini, 69
- mot de Lyndon, 63
- mot dérivé
 - à gauche, 54
 - à droite, 54

Vous pouvez maintenant éteindre votre télévision et reprendre une activité normale.

— PPDA, les guignols de l'info

Résumé. Nous étudions dans cette thèse la notion de quasipériodicité, introduite par Apostolico et Ehrenfeucht au début des années 1990, puis étendue aux mots infinis par Solomon Marcus au début des années 2000. Un mot (fini ou infini) w est quasipériodique s’il peut être couvert par des occurrences, éventuellement chevauchantes, d’un autre mot, fini, appelé sa quasipériode. En 2006, Solomon Marcus et Thierry Monteil ont introduit la notion plus forte de quasipériodicité multi-échelles : le fait d’avoir une infinité de quasipériodes.

Dans un premier temps, nous donnons des résultats sur les mots infinis en une dimension. Nous donnons notamment une approche permettant de déterminer les quasipériodes d’un mot infini à partir de ses facteurs carrés et de ses facteurs spéciaux. Nous montrons ensuite que la famille des mots périodiques, ainsi que celle des mots standard sturmiens, peuvent être caractérisées en termes de quasipériodicité multi-échelles.

Dans un deuxième temps, nous étudions la quasipériodicité des mots infinis bidimensionnels. Nous montrons que, contrairement au cas unidimensionnel où la quasipériodicité ne force aucune propriété forte des mots infinis, il existe des quasipériodes q qui forcent les mots 2D q -quasipériodiques à être d’entropie nulle. Nous montrons également que la quasipériodicité multi-échelles en deux dimensions force l’existence de fréquences uniformes pour les facteurs.

Abstract. We study the notion of quasiperiodicity, initially introduced by Apostolico and Ehrenfeucht at the beginning of the 1990’s, then extended to infinite words by Solomon Marcus at the beginning of the 2000’s. A (finite or infinite) word w is quasiperiodic if it can be covered by (possible overlapping) occurrences of another finite word, called its quasiperiod. In 2006, Solomon Marcus and Thierry Monteil introduced a stronger notion: multi-scale quasiperiodicity, the property of having infinitely many quasiperiods.

First we give results about infinite words in one dimension. Most notably we provide a method to determine the quasiperiods of an infinite word from its square and special factors. We show that the family of periodic words and standard Sturmian words are characterizable in terms of multi-scale quasiperiodicity.

Then we study quasiperiodicity of two-dimensional infinite words. We show that, by contrast with the one-dimensional case where quasiperiodicity do not force any property on infinite words, there exist quasiperiods q which force 2D q -quasiperiodic words to have zero entropy. We also show that multi-scale quasiperiodicity in two dimension force the existence of uniform frequencies for factors.
