



**HAL**  
open science

**Analyse de systèmes dynamiques par discrétisation.  
Exemples d'applications en théorie des nombres et en  
biologie moléculaire**

Anne Siegel

► **To cite this version:**

Anne Siegel. Analyse de systèmes dynamiques par discrétisation. Exemples d'applications en théorie des nombres et en biologie moléculaire. Mathématiques [math]. Université Rennes 1, 2008. tel-00358996

**HAL Id: tel-00358996**

**<https://theses.hal.science/tel-00358996>**

Submitted on 5 Feb 2009

**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.

---

HABILITATION À DIRIGER DES RECHERCHES

présentée devant

**L'Université de Rennes 1**  
**Spécialité : Mathématiques et Informatique**

par

Anne Siegel

Analyse de systèmes dynamiques par discrétisation. Exemples  
d'applications en théorie des nombres et en biologie moléculaire.

**soutenue le 8 décembre 2008 devant le jury composé de :**

Pierre Arnoux, professeur, Université de la Méditerranée  
Marie-Pierre Béal, professeur, Université de Marne La Vallée  
Valérie Berthé, DR CNRS, Université de Montpellier  
Jérôme Buzzi, DR CNRS, Université Paris-Sud  
Hidde de Jong, DR INRIA, Centre Grenoble Rhône-Alpes  
Alejandro Maass, professeur, Université du Chili [rapporteur]  
Jacques Nicolas, DR INRIA, Centre Rennes Bretagne Atlantique  
Elisabeth Pécou, professeur, Université de Nice

**au vu des rapports de :**

Marie-Pierre Béal, professeur, Université de Marne La Vallée  
Hidde de Jong, DR INRIA, Centre Grenoble Rhône-Alpes  
Alejandro Maass, professeur, Université du Chili  
Boris Solomyak, professeur, University of Washington, Seattle

---

---

# Remerciements

Ce travail est avant tout le fruit de huit années de rencontres variées... Remercier tous les gens qui y ont contribué en peu de mots est une tâche ardue.

Mes premiers et plus sincères remerciements doivent aller à Pierre Arnoux, Valérie Berthé et Jacques Nicolas.

Pierre Arnoux m'a fait découvrir les systèmes dynamiques et a encadré mes premiers pas dans la recherche ; il a gardé un oeil sur moi depuis ma thèse, même si les pistes que j'ai explorées (biologie, topologie ou théorie des nombres) ne sont pas ses thèmes préférés. En plus des règles du métier, il m'a appris à prendre le temps d'identifier l'essentiel derrière une problématique (ce que je ne fais pas encore assez). Il est pour moi un modèle d'intégrité, et je reste admirative de sa capacité à décrire l'essentiel d'un article obscur de 30 pages en moins de 10 lignes limpides.

Valérie Berthé a elle aussi encadré mes premiers pas de chercheuse ; elle est devenue une amie très proche. Son imagination, son immense culture dans différents domaines, ses qualités d'écoute et de communicante font d'elle une grande scientifique dont j'espère être un jour à la hauteur.

Jacques Nicolas m'a accueillie dans le projet Symbiose, à une époque où j'étais une jeune mathématicienne très culottée récemment recrutée en informatique au CNRS. J'étais tentée par la biologie via l'informatique, alors que mes compétences dans ces deux domaines étaient très limitées. Il m'a laissé ma chance... En plus de ses qualités de chercheur, son soutien constant, ses conseils avisés, sa vision de la science et sa gestion tout en doigté de l'équipe, en respectant la personnalité de chacun (y compris la mienne!) m'ont offert un cadre rare pour m'épanouir en canalisant mon enthousiasme quelque peu débordant.

Mon plus grand respect et mes profonds remerciements vont aux rapporteurs qui ont accepté de s'attaquer à ce document, dont les contours sont bien variés. Par leurs résultats à l'interface de différents domaines, ils sont pour moi des modèles : Alejandro Maass dont le parcours scientifique et les thèmes de recherches font le grand écart entre la théorie ergodique et la bioinformatique ; Hidde de Jong, si grand nom de la modélisation des systèmes biologiques ; Marie-Pierre Béal si grand nom de l'informatique théorique ; Boris Solomyak, dont les travaux et la gentillesse ont inspiré tant de chercheurs... Je suis aussi très honorée que Jérôme Buzzi et Elisabeth Pécou (quand trouvera-t-on donc le temps de travailler ensemble ?) aient accepté de participer à mon jury.

Je dois ensuite remercier le groupe de *systémiciens* avec qui tout le projet autour de la biologie a été développé, et tous essentiels à sa réalisation : Michel Le Borgne qui m'a beaucoup appris en informatique et biologie (entre autre que je suis allergique à la programmation) ; Ovidiu Radulescu, à l'intuition débordante ; Philippe Veber, toujours prêt à explorer de nouvelles voies avec enthousiasme, même quand il s'agit de relire ce

mémoire (et sauf quand il s'agit de rédiger sa thèse); Carito Guziolowski, venue du bout du monde, à la volonté de fer, et imbattable pour faire parler des données de manière intelligente; Pierre Blavy, toujours bricoleur, mais d'une rigueur absolue lorsqu'il s'agit de proposer une conclusion biologique; François Moreews, expert en Java et en interactions, Annabel Bourdé et Gregory Ranchy, toujours présents pour trouver une solution logicielle à mes "*ce serait bien si...*"; Sandrine Laguarrigue et Nathalie Théret, nos *biologistes préférées* qui ont eu la patience de nous transmettre leurs intuitions et leurs connaissances en biologie pour que nous essayions de produire des résultats exploitables; Et les derniers arrivés, Tatiana Bauromatova et Sylvain Blachon, qui n'hésitent pas depuis leur arrivée à nous titiller sur l'applicabilité réelle de nos travaux ce qui oblige (heureusement) à réfléchir encore et encore!

Au sein de ce groupe, les caractères ont toujours été bien trempés, les opinions ont pu diverger, et les relations ont été parfois houleuses. Mais l'énergie, l'enthousiasme et globalement la bonne humeur ne se sont jamais démenties et sont au coeur du travail présenté ici.

Je dois encore mentionner les collaborateurs qui m'ont fait découvrir de nouveaux horizons scientifiques et géographiques: Joerg Thuswaldner, qui m'a introduit à la topologie et m'a permis de découvrir l'opéra viennois; Shigeki Akiyama et ses numérations non entières, ainsi que l'équipe de Japonais (Shunji Ito, Maki Furukado, Hiromi Ei) grâce auxquels le Japon est un univers un peu moins fermé; Arnaud Hilion et la géométrie des groupes libres (même si décidément j'ai du mal avec la géométrie); Boris Adamczewski (frère ou demi-frère selon la généalogie scientifique?) qui m'a toujours impressionnée par sa rigueur et ses intuitions; Wolfgang Steiner, lui aussi d'une rigueur bien nécessaire pour recadrer mes rédactions un peu brouillonnes. Tous ont contribué à élargir mon champ scientifique, et ont ainsi laissé leur trace dans ce document.

Ensuite, mon amitié va à toute l'équipe Symbiose et son ambiance incroyable avec ses pots, pauses café, repas, groupes de travail, séminaire Aubade, et activités musicales (cette année, je me mets à la batucada). Chose fascinante, cette ambiance n'a jamais empiété sur le sérieux et le professionnalisme de l'équipe. Avec une attention particulière à François Coste (mon alter ego dans l'équipe - souvent en opposition mais toujours amicalement) qui porte une grande partie de cette animation (sociale et scientifique). Prochain sur la liste des HDR?

Ces longs remerciements ne seraient pas complets sans y inclure l'IRISA, le CNRS et l'INRIA. Ces six dernières années, je n'ai pu que m'incliner devant l'efficacité des personnels de l'IRISA, dont j'ai profité voire abusé (très chère Marie-Noëlle...), ainsi que devant la bonne humeur qui règne dans ce laboratoire. J'espère sincèrement que l'alchimie (si rare ailleurs!) entre les personnels de différentes structures et cultures résistera aux mouvements actuels. Le CNRS a toujours soutenu mes recherches et mes prises de risque; l'INRIA m'a permis de profiter de son dynamisme. Là encore, je croise les doigts pour que ces deux instituts puissent continuer à se développer en toute complémentarité.

Enfin, Stéphane, ma moitié, père et époux (presque) parfait... Depuis plus de quinze ans, il s'est adapté un peu partout à mes envies de nouveautés. Son soutien, son équilibre, sa disponibilité, et la joie de vivre au sein de notre famille ont été des piliers de ce travail (son dos en sait quelque chose). Sans parler de sa patience, ces derniers mois de rédaction en particulier. Je lui suis infiniment redevable de supporter mes lubies parfois excessives.



*A mes loulous Quentin, Eymeric et Alexandre  
(Toujours en forme, jamais épuisés...)*



---

# Table des matières

<b>Avant-propos</b>	<b>9</b>
<b>1 Pourquoi coder un système dynamique ?</b>	<b>13</b>
1.1 De l’astronomie à la dynamique symbolique . . . . .	13
1.1.1 Rappels historiques . . . . .	13
1.1.2 Des exemples simples... et moins simples . . . . .	15
1.1.3 Dynamique symbolique . . . . .	16
1.2 Cas classiques en dynamique symbolique . . . . .	17
1.2.1 Le cas markovien . . . . .	18
1.2.2 Systèmes auto-induits et représentations par substitutions . .	21
1.2.3 Retour à la dynamique symbolique . . . . .	24
1.3 La topologie de fractals pour aborder différentes questions interdis- ciplinaires . . . . .	26
1.3.1 Quasi-cristaux et espaces de pavages . . . . .	27
1.3.2 Géométrie discrète et génération de plans discrets . . . . .	29
1.3.3 Étude des systèmes de numération en base non entière . . . .	31
1.4 Utilisation de systèmes dynamiques en sciences expérimentales . . .	33
1.4.1 Spécificités de la biologie moléculaire . . . . .	34
1.4.2 Systèmes dynamiques en biologie moléculaire . . . . .	36
1.4.3 Comment construire des réseaux sur lesquels appliquer des simulations dynamiques? . . . . .	40
1.5 Plan du document . . . . .	42
<b>2 Discrétiser de (simples ?) additions ou multiplications</b>	<b>43</b>
2.1 Les enjeux de la dynamique symbolique . . . . .	43
2.1.1 Systèmes dynamiques symboliques . . . . .	43
2.1.2 Deux situations orthogonales où la dynamique symbolique a montré son utilité . . . . .	46
2.1.3 Les questions classiques en dynamique symbolique . . . . .	47
2.1.4 Multiplications par des nombres simples . . . . .	48
2.1.5 Multiplier par un réel $\beta$ ; beta-numérations . . . . .	50
2.1.6 Additions et systèmes auto-induits . . . . .	52
2.2 Multidimensionnel : quel candidat pour une partition? . . . . .	54
2.2.1 Multiplier par une matrice . . . . .	55



---

2.2.2	Des multiplications par une matrice vers des additions sur un tore . . . . .	56
2.2.3	Additions sur un tore . . . . .	58
2.2.4	Les fractals de Rauzy : différentes approches . . . . .	60
2.2.5	Construction du fractal de Rauzy . . . . .	61
2.2.6	Propriété topologique fondamentale : autosimilarité . . . . .	63
2.2.7	Retour à la dynamique symbolique . . . . .	67
2.3	Obtenir (enfin) une bonne dynamique symbolique . . . . .	70
2.3.1	Des pavages périodiques aux pavages autosimilaires . . . . .	71
2.3.2	Différentes conditions de pavage . . . . .	73
2.3.3	Graphes de frontière et condition décidable de pavage . . . . .	77
2.3.4	Description de la frontière . . . . .	80
2.4	Perspectives de travail : le cas non Pisot . . . . .	82
<b>3</b>	<b>Utiliser les graphes de frontière</b>	<b>85</b>
3.1	Topologie des fractals de Rauzy . . . . .	86
3.1.1	Où la topologie des fractals s'interprète de manières variées . . . . .	86
3.1.2	Graphes de pavages et propriétés topologiques . . . . .	87
3.1.3	Exemple : homéomorphisme à un disque . . . . .	90
3.1.4	Perspectives d'étude . . . . .	92
3.2	Théorie des nombres. Développement en bases non-entières : le cas non unitaire . . . . .	93
3.2.1	Fractals dans le cas non unitaire . . . . .	94
3.2.2	Nombres dont le $\beta$ -développement est fini . . . . .	97
3.2.3	Nombres dont le développement est périodique . . . . .	99
3.2.4	Perspectives de travail . . . . .	104
3.3	Géométrie discrète : engendrement de plans discrets . . . . .	105
3.3.1	Plans discrets . . . . .	106
3.3.2	Substitutions généralisées . . . . .	107
3.3.3	Le cas purement substitutif . . . . .	110
3.3.4	Approximations de plans discrets tridimensionnels : l'école japonaise . . . . .	112
3.3.5	Un nouveau mode d'engendrement : règles locales pour des substitutions bidimensionnelles . . . . .	115
3.3.6	Engendrer des plans discrets en s'appuyant sur des règles locales . . . . .	118
3.3.7	Perspectives . . . . .	122
3.4	Invariants pour des systèmes dynamiques . . . . .	123
3.4.1	Dynamique des systèmes substitutifs non unitaires . . . . .	123
3.4.2	Vers des partitions de Markov pour des endomorphismes du tore . . . . .	124
3.4.3	Dynamique symbolique des homéomorphismes de surfaces. Automorphismes de groupes libres . . . . .	128

---

<b>4</b>	<b>Appréhender le fonctionnement d'un système dynamique en biologie moléculaire</b>	<b>133</b>
4.1	Dynamique en biologie moléculaire . . . . .	134
4.1.1	Systèmes dynamiques et séquences d'ADN ? . . . . .	134
4.1.2	Dynamique au niveau cellulaire . . . . .	135
4.1.3	Plusieurs stratégies pour modéliser la dynamique d'un réseau à un niveau qualitatif . . . . .	137
4.2	Prédictions pour des réseaux décrits par des réactions . . . . .	140
4.2.1	Existence des états stationnaires dans un modèle dynamique	140
4.2.2	Condition d'unicité . . . . .	144
4.2.3	Prédictions et comparaison de modèles . . . . .	146
4.2.4	Discussion . . . . .	148
4.3	Étude de la réponse d'un système à une influence . . . . .	149
4.3.1	Modèles différentiels décrits par un graphe d'influence . . . . .	149
4.3.2	Réponse linéaire d'un système à un influence extérieure . . . . .	152
4.3.3	Transport et métabolisme du lactose chez <i>E. coli</i> . . . . .	155
4.3.4	Régulation du métabolisme des acides gras . . . . .	157
4.3.5	Ajustement numérique pour analyser l'influence des voies . . . . .	158
4.4	Conclusion . . . . .	159
<b>5</b>	<b>De la dynamique à l'aide au raisonnement par l'analyse de contraintes</b>	<b>161</b>
5.1	Première question : comment construire des réseaux sur lesquels appliquer des simulations dynamiques ? . . . . .	161
5.1.1	Différentes approches pour la construction de modèles . . . . .	162
5.1.2	Formalisation : deux questions et une règle intuitive . . . . .	164
5.1.3	De l'intuition à une contrainte satisfaite par une régulation . . . . .	165
5.1.4	Extension de la contrainte à d'autres cadres . . . . .	168
5.2	Une approche pour le diagnostic d'un réseau . . . . .	170
5.2.1	Construire le système de contraintes décrivant un réseau et les observations disponibles . . . . .	170
5.2.2	Exploiter le système de contraintes : invariants et modules spécifiques . . . . .	171
5.2.3	Cohérence dans la littérature : vers une contrainte plus restreinte mais tout aussi intuitive ? . . . . .	172
5.2.4	Calcul des invariants et des diagnostics . . . . .	174
5.2.5	Schéma d'utilisation et validation . . . . .	176
5.2.6	Etude (en cours) du rôle de la molécule EWS/FLI1 dans le sarcome d'Ewing . . . . .	181
5.3	Une nouvelle question : identifier le signe des régulations . . . . .	183
5.3.1	Inférence de signes . . . . .	183
5.3.2	Diagnostic : identifier des interactions avec des effets multiples	185
5.3.3	Prédiction : utiliser la redondance pour tenir compte des incertitudes ? . . . . .	187

---

5.3.4	Combien de jeux expérimentaux sont-ils nécessaires pour procéder à une inférence ? . . . . .	189
5.4	De nombreuses questions à explorer : vers une aide au raisonnement pour le biologiste ? . . . . .	191
5.4.1	Différentes questions autour des plans expérimentaux . . . . .	191
5.4.2	Comment sélectionner des sommets pour valider un système ?	193
5.4.3	Plan expérimentaux autour du contrôle d'un système . . . . .	196
5.4.4	Identifier des groupes d'intérêt dans un réseau . . . . .	197
5.5	Synthèse des perspectives . . . . .	199
	<b>Bibliographie</b>	<b>200</b>
	<b>Activités scientifiques</b>	<b>214</b>
5.6	Publications . . . . .	215
5.7	Communications . . . . .	217
5.8	Encadrement et enseignements . . . . .	219
5.9	Coopérations et contrats . . . . .	220
5.10	Animation de la recherche . . . . .	221
	<b>Résumé</b>	<b>222</b>

---

# Avant-propos

**Pluridisciplinarité?** Ce travail présente des contributions théoriques et pratiques à la théorie des codages symboliques de systèmes dynamiques. Les applications qui vont être discutées sont variées puisqu'elles concernent différents champs mathématiques et la modélisation en biologie moléculaire. Cette diversité s'explique avant tout par mon parcours, puisque j'ai une formation de mathématiques pures (études de systèmes dynamiques symboliques), et j'ai ensuite été recrutée au CNRS par la section 07 (informatique) pour travailler sur les applications des systèmes dynamiques dans différents domaines.

Tout en continuant mes travaux en mathématiques sur l'étude de domaines adéquats pour discrétiser un système dynamique (voir chapitre 2), je me suis donc intéressée aux applications des systèmes dynamiques en informatique théorique, en particulier pour mieux caractériser les écritures de nombres en base non entière (chapitre 3, section 2) et proposer une stratégie d'engendrement et d'étude de plans discrets (chapitre 3, section 3). En parallèle, je me suis aussi beaucoup investie au sein du projet de bioinformatique Symbiose à l'IRISA pour comprendre comment les systèmes dynamiques peuvent soutenir les chercheurs en biologie moléculaire dans la modélisation et la compréhension des phénomènes dynamiques au sein de la cellule. Ces travaux sont décrits dans les chapitre 4 et 5.

**Questions de recherches** Le lecteur pourra donc se demander quel est l'homogénéité entre ces différents travaux. À mon sens, l'ambition qui unit ces travaux est de montrer comment on peut utiliser des méthodes de discrétisation de systèmes dynamiques pour exploiter au mieux les informations disponibles sur le système.

Un premier objectif est d'exhiber des informations au sujet d'une dynamique que l'on connaît explicitement et les traduire en propriétés concrètes. Un deuxième objectif est de produire de la connaissance sur une dynamique ou un modèle lorsqu'on ne le connaît pas explicitement. J'ai abordé ces deux questions sur deux grandes classes de systèmes dynamiques.

1. *L'étude des automorphismes et des additions sur un tore.* Inspirés par les cas unidimensionnels (beta-numération, étude des suites sturmiennes), la question principale qui se pose est de trouver un domaine fondamental pour le tore dans lequel les trajectoires de la dynamique considérée se codent par des systèmes symboliques *simples* (décrits par des mots interdits -sous-décalage de type fini- ou des compositions de processus d'itération -substitutions). Un

bon candidat pour ces partitions est donné par une structure dont la base est fractale, introduit par Rauzy dans les années 1980. Le chapitre 1 porte sur mon travail pour unifier les formalismes existant autour de ce fractal, et introduire une approche décidable pour s'assurer que ce fractal est un bon domaine pour un codage dans le cas où l'automorphisme du tore considéré admet une unique direction dilatante (le cas Pisot).

Dans le cas où le domaine considéré induit un bon codage, nous verrons au chapitre 2 (section 2.1) comment les approches que j'ai développées permettent de caractériser les propriétés topologiques du fractal. La fin du chapitre 2 détaille comment ces propriétés topologiques s'exploitent dans les différents domaines d'informatique théorique où les automorphismes et les additions sur un tore peuvent être considérés :

- en théorie des nombres, les propriétés topologiques du domaine de Markov permettent de caractériser les propriétés des développements finis ou purement périodiques de rationnels en base non entière.
- en géométrie discrète, ces propriétés s'interprètent en termes de conditions pour l'engendrement de plans discrets par des méthodes itératives.

2. *L'étude de systèmes dynamiques de grande échelle en biologie moléculaire.* Il s'avère que les données et les connaissances sur les modèles relatifs aux régulations transcriptionnelles dans une cellule sont souvent trop partielles pour leur appliquer les méthodes usuellement utilisées pour la modélisation de systèmes expérimentaux. Je travaille sur l'élaboration d'un formalisme permettant effectivement d'interpréter les observations en biologie moléculaire dans un cadre dynamique pour aider à la correction de modèles, et, dans le futur, à la mise en place de plans expérimentaux. Au vu de la qualité des données, les aspects dynamiques sont alors remplacés par des considérations sur les déplacements d'états stationnaires, et analyser les données revient à formaliser des contraintes (accessibles pour les méthodes actuelles de résolutions de contraintes) portant sur des ensembles discrets ; ces contraintes doivent décrire les questions des biologistes sur leurs systèmes. Au sein du projet Symbiose, nous avons ainsi montré comment les notions de corrections de modèles et de diagnostic de réseaux grande échelle peuvent être abordées.

**Spécificité : limiter ses ambitions pour aboutir à des applications décidables.** Directement liée à l'objectif tourné vers des applications, un point commun des ces travaux est que toutes les approches proposées se sont adaptées à l'état des connaissances actuelles, en adoptant une approche algorithmique.

En mathématiques, on souhaiterait idéalement prouver une conjecture générale qui dit que tous les systèmes symboliques engendrés par une substitution Pisot sont mesurablement isomorphes à une addition sur un tore. Pour l'instant, la preuve de ce résultat semble inaccessible. Les méthodes que j'ai développées, inspirées par l'informatique, consistent donc à vérifier cette propriété au cas par cas, avec une approche algorithmique basée sur la construction et l'étude de graphes. Lorsque la propriété est vérifiée, on peut concrètement exploiter cette information pour l'étude des écritures de nombres ou la structure des discrétisations de plans.

---

De même, en biologie moléculaire, nous sommes partis du constat que même si des informations de plus en plus nombreuses sont mises à disposition sur le comportement des systèmes biologiques, elles ne sont pas en nombre suffisant pour modéliser la complexité de la dynamique des systèmes usuellement considérés par les chercheurs en biologie (en particulier ceux qui travaillent sur des modèles eucaryotes). Nous nous sommes donc volontairement restreints à la compréhension des déplacements d'équilibres avec des approches de résolution de contraintes, avec l'ambition d'aider au diagnostic et à la mise en place de plans expérimentaux.

**Plan du document** Par contre, la diversité des applications visées implique que les travaux présentés ne sont pas du tout homogènes quant à la formalisation des concepts abordés. En fait, la formalisation va aller en décroissant, depuis l'étude de systèmes dynamiques en théorie ergodique (chapitre 2) à une discussion sur l'aide à l'expérimentation que peut apporter la résolution de contraintes.

Le premier chapitre est un chapitre de revue bibliographique sur les méthodes de discrétisation pour l'étude de systèmes dynamiques, axée sur les enjeux des systèmes symboliques : rechercher une partition de l'espace des états qui permette de décrire les trajectoires des points avec des suites symboliques qui vérifient une contrainte la plus simple possible. Ces enjeux sont d'abord discutés autour de l'existence de partitions de Markov pour certains systèmes, et ses conséquences en géométrie, théorie des nombres, physique théorique et en informatique (via la géométrie discrète). Ensuite, on évoquera la manière dont les systèmes dynamiques sont effectivement utilisés en biologie moléculaire, où les codages symboliques associés à une approche de vérification de modèle ont apporté des résultats substantiels.

Le second chapitre aborde la problématique de la dynamique symbolique dans le cadre des automorphismes et des endomorphismes du tore. Nous expliquons pourquoi la recherche d'une bonne partition pour décrire ces applications (domaine de Markov) est intimement liée à des questions de pavages par des compacts autosimilaires sur lesquels existe une dynamique d'échange de morceaux. Nous commencerons par résumer les connaissances sur ce domaine, et présenterons finalement une contribution à la détermination explicite de partitions de Markov, via l'étude des voisinages de pièces fractales dans un pavage.

Le troisième chapitre porte sur les applications des représentations des automorphismes du tore par des domaines de Markov à bord fractal. Nous discuterons en quoi des résultats précis sur la topologie des fractals permettent de répondre à des questions en théorie des nombres, en géométrie discrète, et en perspective, en géométrie des groupes libres.

Dans le quatrième chapitre, nous changeons de domaine d'application. On suppose maintenant que l'on dispose d'observations sur un système régi par une dynamique au sujet de laquelle on dispose d'informations très partielles. On souhaite exploiter au maximum cette information. La question qui se pose est donc de comprendre le comportement d'un système biologique à partir de connaissances et d'observations limitées. Nous montrons qu'en s'appuyant sur des méthodes algébriques combinées avec des études de déplacements d'états stationnaires, il est possible d'élucider la fonction de différentes régulations génétiques et de faire des prédic-

tions sur les effets de mutations spécifiques. Dans ce chapitre, la question biologique étudiée est la régulation du métabolisme des acides gras. Nous utilisons différents niveaux d'abstraction et différentes méthodes algébriques pour comprendre ce système dynamique.

Dans le cinquième chapitre, nous formalisons et étendons les raisonnements du chapitre 4 pour étudier la question de la correction de modèle. Ici, la problématique n'est plus de comprendre le fonctionnement d'un système mais de s'inspirer des approches dynamiques développées au chapitre 4 pour tester si l'ensemble des connaissances disponibles sur un système est suffisant pour expliquer un comportement observé. La démarche consiste à construire un système de contraintes qui sont vérifiées par les variations des concentrations des molécules du système, et d'interpréter différentes questions des biologistes sur un système comme des propriétés spécifiques des solutions du système de contraintes. En perspective, nous discutons des futures applications de cette méthode pour l'élaboration d'un plan expérimental en fonction des différents objectifs de l'expérimentateur : valider un modèle, contrôler un système, et comprendre son fonctionnement.

**Perspectives** Tout au long de ce document, nous allons discuter des différentes perspectives de ces travaux. Pour les synthétiser, mentionnons :

- En systèmes dynamiques, la construction de partition de Markov pour des endomorphismes du tore ou des automorphismes du tore non positifs, en considérant des espaces de représentation  $p$ -adiques ou des morphismes du groupe libre. En application, l'objectif est de proposer de nouveaux invariants (s'appuyant sur la topologie de fractals) pour les espaces de pavages ou les dynamiques associées à des automorphismes de groupes libres.
- En théorie des pavages, proposer explicitement des processus d'engendrement des plans discrets, associés à des critères topologiques pour contrôler la forme des motifs engendrés. En particulier, cette question est fondamentale pour avancer dans l'étude des approximations simultanées de nombres réels. Plus théoriquement, cela permettrait de décrire explicitement les additions sur un tore par des compositions de substitutions, et donc d'avancer dans la résolution de la conjecture dite *S-adique*.
- En biologie moléculaire, proposer des approches utilisables par les biologistes pour orienter leurs plans expérimentaux et plus généralement faciliter le développement de boucles de raisonnement/expérimentations sur des plateformes grande-échelle et automatiques.
- Plus théoriquement, en biologie, avancer dans la compréhension du fonctionnement des systèmes biologiques, via la caractérisation de fonctions pour des sous-groupes de produits regroupés en modules.

---

# Chapitre 1

## Pourquoi coder un système dynamique ?

Déterminisme (Laplace [Lap86]) : *Nous devons envisager l'état présent de l'univers comme l'effet de son état antérieur et comme la cause de celui qui va suivre. Une intelligence qui pour un instant donné connaîtrait toutes les forces dont la nature est animée et la situation respective des êtres qui la composent, si d'ailleurs elle était assez vaste pour soumettre ces données à l'analyse, embrasserait dans la même formule le mouvement des plus grands corps de l'univers et ceux du plus léger atome : rien ne serait incertain pour elle, l'avenir comme le passé serait présent à ses yeux. L'esprit humain offre, dans la perfection qu'il a su donner à l'astronomie, une faible esquisse de cette intelligence. Ses découvertes en mécanique et en géométrie, jointes à celles de la pesanteur universelle, l'ont mis à portée de comprendre dans les mêmes expressions analytiques les états passés et futurs du système du monde. En appliquant la même méthode à quelques autres objets de ses connaissances, il est parvenu à ramener à des lois générales les phénomènes observés, et à prévoir ceux que les circonstances données doivent faire éclore.*

Déterminisme (20ème siècle) : *L'univers est un système dynamique.*

### 1.1 De l'astronomie à la dynamique symbolique

#### 1.1.1 Rappels historiques

**Astrologie** Depuis toujours, l'homme essaie de prévoir l'avenir, d'une part en recherchant des régularités dans les événements (jour, année, saisons, lune), d'autre part en supposant qu'elles se répètent. L'hypothèse implicite est que pour prédire le futur d'un système, il est *suffisant* de connaître son passé. Aujourd'hui encore, un lieu commun veut que l'histoire, l'économie voire la psychologie se comportent de manière cyclique et qu'il suffit de bien étudier le passé pour comprendre le futur... Deux idées différentes sont ici mêlées : il existe des lois pour décrire un système,



et ces lois peuvent amener à des comportements cycliques qui sont identifiables par une bonne connaissance de l'ensemble du passé.

Galilée et Newton ont révolutionné ce point de vue : en mécanique classique, pour prévoir la trajectoire d'un point, il suffit de connaître sa position, sa vitesse initiale, et l'ensemble des forces auquel il est soumis. Il n'est pas nécessaire de connaître tout son passé. Des succès spectaculaires comme la prédiction du retour de la comète de Halley ont validé cette approche. L'hypothèse implicite est ici qu'à chaque système en évolution, on peut associer un *état*, qui contient toutes les informations pertinentes sur le système, et qui permet de prévoir son évolution future dès lors qu'on utilise des *lois*. Une des ambitions majeures de la physique est ainsi d'exhiber des lois d'évolution pertinentes.

**Poincaré et les systèmes dynamiques** Les prévisions sur le futur d'un système se font en général au moyen *d'équations différentielles*, que les mathématiciens ont cherché à résoudre de manière toujours plus complexe. Or, à la fin du 19<sup>ème</sup> siècle, Poincaré s'est aperçu que dans certains cas, il est *impossible* de trouver des solutions explicites à ces équations [Poi87, BG97]. Un nouveau point de vue a alors été proposé : connaissant l'état du système à un moment donné, on veut connaître l'état de ce système une unité de temps plus tard. Formellement, on considère l'ensemble de tous les états possibles, souvent appelé  $X$  et une règle d'évolution,  $T : X \rightarrow X$ , qui à un état  $x$  (par exemple, la donnée de la position et la vitesse du soleil, la terre et la lune) associe un état  $T(x)$  (la position des astres un jour plus tard). On obtient ainsi le concept de *système dynamique*  $(X, T)$ . Formellement, le futur d'un point  $x$  est décrit par les points  $x, T(x), T(T(x)), T(T(T(x)))...$  On l'appelle *orbite* de  $x$ .

Étudier un système dynamique est bien différent de l'étude des fonctions apprise au lycée. Dans ce dernier cadre, nous avons tous recherché la manière dont  $T(x)$  *varie* avec  $x$  : le principal objet d'étude est alors le graphe de la fonction  $T$ , en particulier ses variations, ses points d'inflexion ou son comportements aux points limites. Dans le cadre des systèmes dynamiques, on étudie l'*évolution* d'un point  $x$ , en recherchant des régularités dans la suite des futurs  $T(x), T(T(x)), T(T(T(x)))...$  C'est cette orbite de points qui constitue l'objet principal d'une étude dynamique. Les questions typiques sont alors : les éléments d'une orbite restent-ils dans un sous-ensemble de  $X$  ? Peuvent-ils y revenir régulièrement ? Si deux points  $x_1$  et  $x_2$  sont proches, leurs orbites restent-elles proches ? Peuvent-elles même se rejoindre ?

**Naissance d'un domaine mathématique** La notion de déterminisme telle qu'introduite par Laplace dit alors tout simplement que l'univers est un système dynamique, au sens où le futur est exactement prédit par une loi et un ensemble d'états habilement définis. Même si ce déterminisme a des limites, comme l'a montré la mécanique quantique (qui dit que l'état d'une particule ne peut pas être déterminée de manière certaine) ou les travaux de Prigogine en thermodynamique (qui portent sur les structures dissipatives et l'irréversibilité des phénomènes temporels), l'hypothèse déterministe reste valable dans de nombreux cadres (et en premier lieu en mécanique céleste), et les mathématiciens ont entamé des travaux pour obtenir

---

une classification des systèmes en fonction de leurs comportements et leur avenir à long terme, semblable à la classification des éléments en chimie.

Or, dès Poincaré, il est apparu que la multiplicité des trajectoires possibles rendait une telle classification illusoire. L'échec de cette ambition a ouvert un immense champ d'étude pour les mathématiciens, dans deux directions : d'une part, étudier de nombreux exemples ; d'autre part, trouver des concepts qui permettent de comprendre le comportement de ces systèmes, à la fois d'un point de vue théorique (caractérisation de différents types d'ordre et de chaos) et d'un point de vue pratique, en définissant des quantités qui décrivent partiellement un système dynamique, analogues à la masse ou au potentiel pour les systèmes physiques.

### 1.1.2 Des exemples simples... et moins simples

**Les systèmes finis** Les systèmes dynamiques les plus simples sont sans conteste les systèmes dynamiques finis, où  $X$  est un ensemble fini et  $T$  est surjective sur  $X$ . Ces systèmes sont bien compris : puisque l'ensemble des états est fini, l'orbite de chaque point doit forcément revenir à un moment ou à un autre sur un état déjà traversé ; le déterminisme de la dynamique implique alors que la trajectoire devient périodique. On dit que ces systèmes sont *pré-périodiques*. Par exemple, dans un ordinateur, dont les cases mémoires sont en nombre fini, on ne peut représenter qu'un nombre fini de nombres. Ainsi, pour un système dynamique calculé sur un ordinateur, toutes les orbites sont finies... Dans la pratique, elles sont très souvent de période gigantesque ; mais il arrive parfois, à cause des approximations faites lors d'un calcul numérique, qu'une orbite apparaisse finie alors qu'on sait que le système théorique que l'on étudie n'a pas d'orbite finie, ce qui induit des erreurs d'analyse.

**Les systèmes linéaires sur un espace vectoriel** Un peu plus complexes sont les systèmes dynamiques linéaires, où  $X$  est un espace euclidien  $\mathbb{R}^n$  et  $T$  est donnée par l'action d'une matrice  $\mathbf{M}$  sur  $\mathbb{R}^n$ . Dans ce cas, une classification complète des dynamiques existe : le comportement du système dépend essentiellement des valeurs propres de la matrice  $\mathbf{M}$ , qui permettent de décomposer l'espace en différents sous-espaces sur lesquels  $\mathbf{M}$  agit de manière simple (contraction, dilatation ou rotation) ; on parle de *théorie spectrale* des applications linéaires. Concrètement, on peut dans de nombreux cadres approximer une application quelconque par une application linéaire (via le théorème d'Hartman-Grobman et ses différentes généralisations) ; dans ce cas, la théorie spectrale permet de comprendre la dynamique. Ces dynamiques sont assez simples : l'origine est toujours fixe, et dans la plupart des cas, l'orbite d'un point soit devient périodique après un certain temps, soit se rapproche d'un point périodique, soit s'éloigne vers l'infini, en suivant éventuellement une spirale [HP05, BDV05]. Comme ce sont les dynamiques les plus simples à comprendre et à calculer, on a longtemps cru qu'il s'agissait des seules dynamiques possibles.

**Les systèmes chaotiques** Et pourtant, force a été de constater que les cas les plus fréquents dans la nature sont hautement non-linéaires. On entre alors dans

l'univers du *chaos*. Grossièrement, on parle de chaos lorsque deux points distants (si petite soit la distance qui les sépare) peuvent avoir deux futurs totalement distincts. Cependant que la signification du mot chaos est très fortement dépendante du domaine dans lequel il est employé [Bla, ASY97, Dev86]. Dans tous les cas ; une abondance de trajectoires est possible au voisinage de chaque point ; les notions d'entropie ou d'exposants de Lyapounov ont été développées pour comparer et quantifier ces trajectoires. Il faut par contre noter que, même si la trajectoire d'un point est imprévisible, l'abondance de trajectoires produit souvent une régularité *statistique* qui est exploitable pour étudier *en moyenne* le comportement du système. Les invariants associés aux systèmes dynamiques, l'entropie en premier lieu, permettent de décrire ces régularités.

### 1.1.3 Dynamique symbolique

**Codage d'un système** Pour trouver ces invariants, et plus généralement pour comprendre la dynamique des systèmes, Hadamard a suggéré l'idée de découper l'espace des états en un nombre fini de morceaux, chaque morceau ayant un *nom* (généralement un chiffre ou une lettre de l'alphabet) et d'associer à tout point du système la suite des noms des morceaux dans lesquels se trouve ses images par la dynamique [Had98].

Par exemple, dire que *abbcad...* *code* le point  $x$  signifie que  $x$  se trouve dans le morceau de nom  $a$ , son image  $T(x)$  dans le morceau  $b$ , l'image  $T(T(x))$  à nouveau dans le morceau  $b$ , par contre  $T^3(x)$  se trouve dans  $c$ ,  $T^4(x)$  revient dans  $a$ ,  $T^5(x)$  visite un nouveau morceau de nom  $d...$

Finalement, étant donnée une partition de l'espace d'états, on associe à tout point  $x$  un mot infini de symboles qui décrit sa trajectoire. On parlera de *codage* de la trajectoire de  $x$ , ou plus simplement de codage de  $x$ .

Quel est l'intérêt de cette transformation ? On avait au départ une dynamique donnée par une application difficile à décrire sur un espace d'états généralement simple. Or, via cette opération de codage, la dynamique devient simple, puisque coder  $T(x)$  revient à prendre le codage de  $x$  et lui enlever le premier terme ! On a donc transformé une application complexe en une application très simple... Mais l'espace des états s'est complexifié, puisqu'il s'agit maintenant d'un ensemble de mots infinis dont il reste à comprendre la structure. Autrement dit, on ne représente plus exactement la fonction de départ, ce qui permet de dépasser les limites des approches analytiques, tout en conservant certaines propriétés essentielles de la dynamique. La *dynamique symbolique* est ainsi le domaine qui cherche à identifier les propriétés des codages de systèmes dynamiques.

**Validation de l'approche** Hadamard a proposé cette approche pour étudier les propriétés de récurrence des systèmes : à quelles conditions peut-on assurer que la trajectoire d'un point dans un système revient aussi près que possible du point initial ? Il cherchait en particulier à savoir s'il existait un exemple de surface connexe à courbure négative constante admettant une géodésique récurrente non périodique. Pour cela, il a proposé de coder chaque géodésique par des 0 ou 1, en fonction des

morceaux de la frontière de la surface que la géodésique traversait. Hadamard a ainsi montré que trouver une bonne géodésique revenait à montrer l'existence d'un mot infini non périodique composé de 0 et de 1 tel que tout motif de longueur  $n$  se répète en un temps fini (dépendant seulement de  $n$ ) dans ce mot [Had98]. Vingt-cinq ans plus tard, Morse a démontré qu'un tel mot existe, en considérant le mot infini qui est fixé par la transformation  $0 \rightarrow 01, 1 \rightarrow 10$  (nous reviendrons sur ce concept de transformation très bientôt) [Mor21].

À partir de codages symboliques, on décrit un certain nombre de propriétés du système de départ. La notion d'entropie topologique, par exemple, est basée sur ce concept : on considère d'abord une partition du système des états ; on fixe un entier  $n$ , et on calcule le nombre  $H(n)$  de codages existant pour les trajectoires de taille  $n$ , c'est-à-dire pour les ensembles de points de la forme  $\{x, T(x) \dots T^{n-1}(x)\}$ . L'entropie topologique relativement à la partition est la limite de la quantité  $(\log H(n))/n$ . L'entropie du système dynamique est le supremum de toutes les entropies, pris sur l'ensemble des partitions finies. Un premier théorème montre que cette entropie existe, un autre théorème dit qu'elle est atteinte pour n'importe quelle partition pour laquelle deux points différents ne peuvent pas avoir le même codage [AKM65]. Finalement, pour calculer une entropie, il *suffit* de connaître tous les codages des points de la trajectoire pour une bonne partition, et de compter le nombre de motifs dans ces codages...

Depuis, le principal enjeu de la dynamique symbolique est de rechercher une partition de l'espace des états qui décrit les trajectoires des points avec des suites symboliques qui vérifient une contrainte la plus simple possible. Fondamentalement, cela revient à abstraire de manière si possible "simple" la dynamique de départ pour dépasser les limites des approches analytiques, tout en conservant certaines propriétés essentielles du point de vue de la dynamique.

**Domaine d'application : des applications linéaires sur des espaces qui ne sont plus vectoriels** Nous avons vu que les applications linéaires sur des espaces vectoriels sont prévisibles et non chaotiques. Pour introduire du désordre, il faut donc enlever une de ces deux hypothèses. D'un point de vue physique, on préférera rester dans un cadre vectoriel mais considérer des applications non linéaires et introduire des méthodes analytiques ; nous discuterons de cette question dans la section 1.4. Cependant, d'un point de vue mathématique, il est préférable de travailler sur des espaces compacts. Dans la suite de cette section, nous allons ainsi considérer des applications linéaires non plus à valeur dans des espaces vectoriels mais à valeurs dans un espace quotient compact d'un espace vectoriel. Nous allons voir comment ce passage au quotient permet de produire du chaos, et décrire ce chaos à l'aide de dynamique symbolique.

## 1.2 Cas classiques en dynamique symbolique

Ainsi, l'objet principal d'étude de la dynamique symbolique est l'ensemble des mots infinis que l'on obtient par codage d'un système dynamique.

### 1.2.1 Le cas markovien

**Partition markovienne** Le cas le plus simple d'un point de vue combinatoire est celui où les mots infinis sont exactement définis par un ensemble de mots finis interdits qui correspondent à des enchaînements interdits (on parle de *sous-décalage de type fini*). Une partition d'un espace qui permet de coder un système dynamique en un sous-décalage de type fini est appelée *partition markovienne*. L'existence d'une partition markovienne pour un système signifie que l'on peut décomposer l'espace des états en un nombre fini de morceaux, et que l'appartenance à ces morceaux est suffisante pour connaître le futur immédiat d'un point.

**Applications en théorie des nombres** L'exemple le plus naturel est donné par la multiplication par 2 ramenée sur le tore  $\mathbb{R}/\mathbb{Z}$ , en bijection avec l'intervalle  $[0, 1[$  (notée  $T_2$ ) : on considère un réel  $x \in [0, 1[$ , on le multiplie par 2 et on ne retient finalement que la partie après la virgule de cette quantité, c'est-à-dire  $T_2(x) = 2x - E(2x)$  (où  $E(y)$  désigne la partie entière de  $y$ ). Même si cette application semble proche de la multiplication par 2, dont la dynamique est bien comprise et prévisible, sa dynamique n'est absolument pas prévisible : elle est d'entropie non nulle et markovienne.

Cette application, et ses cousines (où 2 est remplacé par un réel positif  $\beta$ ) a fait l'objet de nombreuses études combinatoires, dynamiques ou ergodiques. La raison en est simple : cette application est le fondement de la *numération binaire*, ses généralisations sont liées aux *beta-numérations*. En effet, si on décompose l'espace des états (ici l'intervalle  $[0, 1[$ ) en deux parties naturelles :  $[0, 1/2[$  et  $[1/2, 1[$ , et on code sa dynamique comme expliqué précédemment, on montre que le codage d'un point  $x$  n'est rien d'autre que son développement binaire. L'espace des codages est bien markovien, puisqu'il n'y a aucun mot interdit dans les développements binaires des réels, et on en déduit que l'application  $T_2$  est d'entropie  $\log 2$ . Nous discuterons ce point plus en détails dans les sections 2.1.4 et 2.1.5.

On peut cependant se demander quel est l'apport réel de ce point de vue dynamique dans les systèmes de numération. En fait, un certain nombre de théorèmes de type probabilistes ont été établis concernant les systèmes dynamiques, portant en particulier sur les propriétés de retour près d'un point de départ (on est alors dans le domaine de la théorie ergodique). Ces théorèmes, associés à la vision dynamique des systèmes de numération, produisent des propriétés statistiques sur les écritures de nombres. On montre par exemple que dans le développement binaire de presque tous les nombres réels, la fréquence de 0 et de 1 est égale à  $1/2$ ; en fait, cette fréquence est uniforme quelle que soit la base (entière) dans laquelle on développe pratiquement tous les réels [Bor09, Lan94]. de manière plus concrète, ces approches sont aussi utilisées pour l'analyse dynamique d'algorithmes arithmétiques, en considérant de l'analyse spectrale basée sur de opérateurs de transfert [Val03].

Notons cependant qu'un piège se trouve dans le terme *presque tous* : déterminer si la fréquence de 1 dans le développement binaire de  $\sqrt{2}$  est uniforme dans n'importe quelle base est une question toujours ouverte [K06, Bor50, Wal06].

**Applications en géométrie** Plus généralement, la notion de partition markovienne a été utilisée en géométrie. D’abord, implicitement, par Smale, sous la forme d’un ensemble de Cantor invariant pour des difféomorphismes de la sphère [Sma67]. Ensuite, l’existence de partitions de Markov a été démontrée pour une classe assez large de systèmes dynamiques, en particulier les automorphismes hyperboliques des tores et les difféomorphismes pseudo-Anosov des surfaces [Bow75].

Puisque nous allons les mentionner de nombreuses fois dans la suite, rappelons que les endomorphismes de tores multidimensionnels sont les généralisations naturelles des multiplications sur un tore de dimension 1 : il s’agit d’associer à un élément  $\mathbf{x}$  d’un tore  $\mathbb{T}^n = \mathbb{R}^n / \mathbb{Z}^n$  la valeur du produit  $\mathbf{M}\mathbf{x}$  modulo le réseau  $\mathbb{Z}^n$ , où  $\mathbf{M}$  est une matrice de taille  $n$  à coefficients entiers. L’endomorphisme est un *automorphisme* lorsqu’il est inversible sur le tore, c’est-à-dire qu’il est de déterminant  $\pm 1$ . Cet automorphisme est hyperbolique si la matrice  $\mathbf{M}$  n’admet aucune valeur propre de module 1. Le résultat de Bowen dans [Bow75] montre que ces automorphismes admettent toujours une partition de Markov, qui dans ce cas, est donnée par un domaine fondamental de  $\mathbb{Z}^n$  et à un découpage de ce domaine. On parle alors de *domaine de Markov*.

Entre autres applications de cette existence, citons par exemple le fait que les automorphismes hyperboliques d’un tore de dimension deux sont mesurablement isomorphes si et seulement s’ils partagent la même entropie [Adl98, AW70].

**Trouver des domaines de Markov explicites pour des automorphismes du tore ?** Ainsi, les systèmes dynamiques, associés à une dynamique markovienne et à la théorie ergodique fournissent de nombreux résultats probabilistes (aussi appelés *métriques*) sur les trajectoires du système. Le problème réside dans l’obtention de résultats précis sur un point donné. Dans ce cadre, la connaissance explicite de partitions markoviennes est une étape clé pour mieux comprendre la structure des ensembles sur lesquels les théorèmes ergodiques ne sont pas valables.

Cependant, de nombreuses difficultés apparaissent pour l’obtention de partitions de Markov, et dès les exemples les plus simples que sont les automorphismes du tore. La difficulté est la suivante : alors qu’en dimension deux, des domaines de Markov existent pour les automorphismes du tore sous la forme de rectangles, en dimension supérieure à trois, on sait que la partie contractante du bord des éléments d’un domaine de Markov ne peut pas être lisse [Bow78, Caw91]. Il est donc vain d’espérer construire une partition avec des objets géométriques simples.

Pour aborder cette questions, les chercheurs se sont d’abord concentrés sur une sous-classe des automorphismes du tore : il s’agit des automorphismes hyperboliques du tore pour lesquels la matrice  $\mathbf{M}$  a une unique direction dilatante, qui sont appelés *automorphismes Pisot* puisque la valeur propre dominante de la matrice  $\mathbf{M}$  est un nombre de Pisot (tous ses conjugués de Galois sont de module strictement inférieur à 1).

Pour ces automorphismes Pisot du tore, des constructions ont été proposées en s’appuyant sur des propriétés arithmétiques de l’automorphisme : de bons candidats pour un domaine de Markov peuvent être construits à partir de points homoclines ; il s’agit de considérer un élément  $\mathbf{x}$  tel que l’orbite de  $\mathbf{x}$  dans le futur (c’est-à-dire

$\mathbf{M}^n \mathbf{x} \bmod \mathbb{T}^n$ ) et son orbite dans le passé ( $\mathbf{M}^{-n} \mathbf{x} \bmod \mathbb{T}^n$ ) convergent toutes les deux exponentiellement vite vers  $\mathbf{0}$  [ES97, KV98]. Cependant, les domaines ainsi construits ne sont pas tout à fait exacts, dans la mesure où on ne sait pas montrer qu'ils séparent les points (au sens où les orbites de deux points doivent avoir des codages différents). Nous rediscuterons ce point dans la section 2.2.

### Une approche géométrique : des partitions de Markov aux conditions de pavages

En parallèle avec des approches arithmétiques, différents auteurs ont adopté une approche géométrique, en proposant explicitement des domaines de  $\mathbb{R}^n$  candidats pour être une partition de Markov. Pour construire ces domaines, un constat a été fait : puisqu'on considère un automorphisme du tore  $\mathbf{M}$  qui admet une unique valeur propre dilatante, la géométrie de l'application est très spécifique ; il existe une unique droite sur laquelle l'action de  $\mathbf{M}$  est une homothétie dilatante, et un hyperplan sur laquelle la matrice agit comme une contraction.

Si on considère maintenant un domaine  $\mathcal{D}$  candidat pour une partition de Markov de l'action de  $\mathbf{M}$  sur le tore, ce domaine est en particulier un domaine fondamental de  $\mathbb{Z}^n$  et il pave  $\mathbb{R}^n$  régulièrement. En particulier, son intersection avec un hyperplan à coordonnées entière induit un pavage périodique de l'hyperplan : nous considérerons l'hyperplan diagonal, orthogonal au vecteur  $(1, \dots, 1)$ .

De plus, puisqu'il s'agit d'une partition de Markov, le domaine se décompose en sous-domaines dont les images par la matrice sont des unions de sous-domaines. Or, l'hyperplan contractant de  $\mathbf{M}$  est stable par l'action de  $\mathbf{M}$ . Il va donc "respecter" les décompositions en sous-domaines. On en déduit que les intersections entre les domaines de la forme  $\mathcal{D} + \mathbf{x}$ , pour  $\mathbf{x} \in \mathbb{Z}^n$  et cet hyperplan contractant forment un pavage de l'hyperplan qui est stable par l'action d'une contraction. On va parler de pavage *autosimilaire*. Cette stabilité implique que le nombre de pièces d'intersection est fini, alors qu'il est général infini lorsqu'on coupe un pavage avec des directions rationnelles par un hyperplan à direction irrationnelle. L'autosimilarité implique aussi que le bord des intersections ne va pas pouvoir être lisse, comme cela a été montré dans [Bow78, Caw91].

Les approches géométriques développées dans [Pra99, IO93] partent de ce constat pour proposer un domaine de Markov. Ils recherchent un domaine autosimilaire du plan contractant qui reproduit l'action de la matrice souhaitée. Ils construisent ensuite un domaine de Markov sous la forme d'une suspension adéquate du domaine dans le plan contractant le long de la direction dilatante. On montre que cette construction vérifie toutes les propriétés souhaitées pour les partitions de Markov à une unique condition : il faut que le domaine initialement construit dans l'hyperplan contractant induise un pavage autosimilaire de ce plan [IR06], ou de manière équivalente, un pavage périodique (Thèse de doctorat, 2000). La question de la détermination d'une partition de Markov pour les automorphismes du tore avec une unique direction dilatante est donc finalement ramenée à la vérification d'un pavage pour un domaine particulier d'un hyperplan.

**Des pavages aux additions sur un tore** Un "détail" a été passé sous silence jusqu'à maintenant : comment construire précisément un domaine autosimilaire

convenable dans l'hyperplan contractant ? Sur ce sujet, les approches dynamiques ont là encore inspiré les mathématiciens. En effet, si le domaine autosimilaire recherché dans un hyperplan contractant existe, on peut y définir une application naturelle : on considère l'action qui part du domaine autosimilaire (dans l'hyperplan contractant), suit la direction dilatante, et revient dans l'hyperplan contractant modulo  $\mathbb{Z}^n$ . Dans le cas d'un domaine de Markov, cette action est bien définie, et puisqu'il y a pavage périodique, il va s'agir d'un échange de morceaux qui est isomorphe en mesure à une addition sur un tore.

Ces raisonnements intuitifs ont mis en évidence que la construction explicite d'un domaine de Markov est intimement liée à l'étude des additions sur des tores. Avant de revenir aux partitions de Markov, nous allons donc faire une digression sur les additions sur un tore.

### 1.2.2 Systèmes auto-induits et représentations par substitutions

Du point de vue de la dynamique symbolique, le cas des additions sur un tore peut-être considéré comme "orthogonal" au cas markovien : il s'agit de systèmes dynamique rigides et prévisibles (par exemple, ils sont d'entropie nulle), mais qui ne sont tout de même pas pré-périodiques. On est alors dans une situation paradoxale : le système n'est pas très simple, puisqu'il y a une infinité de trajectoires différentes, mais ces trajectoires ne sont quand même pas très variées, puisque l'entropie est nulle. Ceci sous-entend qu'il y a une régularité cachée au fond du système et, comme nous allons le voir, des techniques d'inductions sont utilisées pour décrire ces régularités.

**Additions sur l'intervalle ; suites sturmiennes** Étant donné un nombre  $\alpha \in [0, 1[$ , l'addition  $R_\alpha$  sur le tore de dimension 1 associée à tout  $x \in [0, 1[$  la partie placée après la virgule de  $x + \alpha$ , c'est-à-dire  $R_\alpha(x) = x + \alpha - E(x + \alpha)$ .

Pour comprendre les trajectoires des points par cette application, Morse et Hedlund ont adopté une approche de dynamique symbolique, une partition naturelle étant donnée par les intervalles de continuité de  $R_\alpha$  [MH38, MH40]. Ils ont ainsi montré que les mots infinis obtenus par codage ont une caractérisation combinatoire très simple : dès que  $\alpha$  est irrationnel, il s'agit des mots qui admettent  $n + 1$  sous-mots différents de longueur  $n$  pour chaque entier  $n$  [MH38, MH40]. De plus, chaque point de  $[0, 1[$ , sauf un ensemble dénombrable de points, a un codage unique (distinct de tous les autres codages). On en déduit en particulier que l'entropie de ces applications est nulle.

Les codages sont appelés *mots sturmiens* ou *suites de coupure*, et elles ont une interprétation géométrique très simple : il s'agit de dessiner un quadrillage du plan par des droites verticales et horizontales en suivant le réseau  $\mathbb{Z}^2$ . Ensuite, il suffit de partir de 0, de suivre la direction de la droite  $Y = \alpha X + \rho$  en restant sur le quadrillage. Lorsqu'on traverse une direction verticale, on retient un 0, lorsqu'on traverse une direction horizontale, on retient la lettre 1 [Fog02, Chapitre 6]. Le mot infini ainsi obtenu est précisément le codage de l'orbite de  $\rho$  par l'addition  $R_\alpha$ .



Les différents points de vue combinatoire, dynamique et géométrique, les propriétés de minimalité et la convexité des applications sous-jacentes, font que ces suites sont utilisées dans de nombreux domaines, allant de la recherche de répétitions dans les séquences d'ADN [DR04] à l'optimisation de l'allocation de ressources dans les réseaux [GHVdL07].

**Induction** La méthode utilisée par Morse et Hedlund pour montrer que toutes les suites de complexité  $n+1$  sont des codages de rotation fait appel à une méthode assez classique en systèmes dynamiques, qui consiste à *induire* sur un sous-ensemble du système, ce qui revient à étudier la manière dont une trajectoire revient sur un sous-ensemble bien choisi de l'espace des états. Dans l'esprit de la dynamique symbolique, la première étape est toujours de partitionner le système un nombre de morceaux finis. Cependant, au lieu de chercher les trajectoires d'un point, on s'intéresse ensuite à la manière dont l'orbite d'un point  $x$  de la première pièce de la partition revient dans cette pièce (on parle d'*application de premier retour*). Lorsque l'application de premier retour a une forme semblable à l'application considérée initialement, on procède à une partition du nouveau système, et on calcule une nouvelle application de premier retour. À chaque étape, l'information qui est conservée est le lien entre l'application et son application de premier retour ; il s'agit des noms des pièces de la partition traversées avant de revenir dans la pièce de départ.

Avec cette méthode d'induction, Morse et Hedlund ont ainsi montré que tous les mots sturmiens s'obtiennent en composant des applications élémentaires sur des mots, appelées *substitutions* ou *morphismes itérés*, qui consistent à remplacer les lettres d'un alphabet par des mots finis, comme par exemple la substitution  $\sigma_0(0) = 01, \sigma_0(1) = 1$ . Plus précisément, le langage d'un mot sturmien de pente  $\alpha$  est déterminé par les mots finis de la forme  $\sigma_0^{a_1} \sigma_1^{a_2} \sigma_0^{a_3} \dots \sigma_0^{a_{2n+1}}(0)$ , où  $a_1, a_2, \dots$  sont donnés par l'algorithme d'Euclide appliqué sur  $\alpha$  (il s'agit de son *développement en fractions continues*) et  $\sigma_0, \sigma_1$  sont deux substitutions qui décrivent le lien entre une addition et son application de premier retour. Nous reviendrons en détail sur ces concepts dans la section 2.1.4.

Cette décomposition en substitutions a été abondamment exploitée dans la littérature ; elle est à la base d'algorithmes rapides (en  $\mathcal{O}(\log n)$ ) pour retrouver les  $n$  premiers termes du codage discret d'une droite à partir de son angle [MH38, MH40, Fog02].

**Additions multidimensionnelles : le cas auto-induit de Tribonacci** Cette approche d'induction est encore valable lorsqu'on déplace par translations plusieurs intervalles de  $[0, 1[$  (*échanges d'intervalles*) [Kea75, Vee78].

Par contre, lorsqu'on considère une addition sur un tore de dimension 2, les choses se passent beaucoup moins bien. Le problème est qu'il n'y a plus de partition naturelle du tore sur laquelle procéder à une induction. Pour progresser sur ce sujet, Rauzy a détaillé le cas de l'addition du *nombre de Tribonacci*. On considère un nombre complexe  $\alpha$  qui vérifie  $\alpha^3 = \alpha^2 + \alpha + 1$  et on considère l'application  $T$  définie sur  $\mathbb{C}$  par  $z \rightarrow z + \alpha \bmod \mathbb{Z} + i\mathbb{Z}$ . Rauzy a exhibé une partition du tore  $\mathbb{C} \bmod \mathbb{Z} + i\mathbb{Z}$  qui montre que  $T$  est auto-induite, c'est-à-dire qu'elle est conju-

guée à son application de premier retour [Rau82]. On en déduit qu’une dynamique symbolique pour l’addition de  $\alpha$  est engendrée par le point fixe de la substitution  $\sigma(1) = 12, \sigma(2) = 13, \sigma(3) = 1$  (appelée *substitution de Tribonacci*).

L’ensemble construit pour procéder au codage de l’addition est appelé *fractal de Rauzy* ou plus tard *tuile centrale*. Il s’agit d’un compact, autosimilaire (puisqu’il est solution d’un GIFS au sens de [MW88]), à bord fractal, d’intérieur non vide. Il est aussi l’adhérence de son intérieur et homéomorphe à un disque [Rau82, Mes98].

**Différents domaines d’applications** L’intérêt de ce fractal n’est pas seulement symbolique; il a en fait engendré des travaux dans un grand nombre de domaines. Une raison “philosophique” est que le fractal de Rauzy ne fait que représenter géométriquement le processus symbolique d’itération décrit par la substitution  $\sigma(1) = 12, \sigma(2) = 13, \sigma(3) = 1$ , elle-même très fortement liée au polynôme  $X^3 = X^2 + X + 1$ . Dès lors qu’un objet mathématique (système de numération, norme sur  $\mathbb{R}^2$ , automorphisme du tore...) a un lien avec cette règle d’itération, il va être visualisé ou étudié d’un point de vue géométrique via le fractal de Rauzy.

Ainsi, en théorie des nombres, Rauzy a déduit de sa construction que le triplet  $(1, 1/|\alpha|^2, 1/|\alpha|^4)$  est très mal approximé par des rationnels de même dénominateur pour une certaine norme, et que les meilleures approximations possibles sont données par la suite récurrente  $T_{n+3} = T_{n+2} + T_{n+1} + T_n$ , ouvrant une porte dans le domaine de l’approximation simultanée ( voir [HM06] et [Lot05, Chapitre 10]). Thurston a aussi utilisé le fractal dans le cadre de l’étude des systèmes de numération à base non-entière [Thu89]. Les premiers exemples de partitions de Markov pour des automorphismes de  $\mathbb{R}^3$  se sont aussi appuyés sur ce fractal [IO93, Pra99].

**Généralisation de la construction** Motivés par ces diverses applications, de nombreux travaux ont cherché à généraliser la construction du fractal de Rauzy dans différents domaines. Plusieurs méthodes de constructions sont apparues.

- La première est inspirée par la construction initiale de Rauzy [Rau82] et considère des séries entières (voir [Mes98, Mes00]). C’est cette approche qui a inspirées les théories sur les points homoclines de [Ver92, Sch00a]. Nous détaillerons cette approche dans la section 2.2.5.
- La seconde approche s’appuie sur les propriétés d’autosimilarité du fractal et des substitutions généralisées pour le construire par approximations successives; elle a été développée par l’école japonaise menée par S. Ito [IK91, AI01]. Dans ce cas là, on considère une substitution dont la matrice d’incidence admet une unique valeur propre dominante et qui est en plus de déterminant 1 (cas Pisot irréductible unitaire). Nous en discuterons à la section 2.3.2.
- La troisième approche suivie à partir des beta-numérations dans [Thu89] consiste à représenter de manière compacte l’ensemble des entiers pour une beta-numération. Dans ces cas là, on ne considère plus une substitution mais un nombre de Pisot  $\beta$  (tous ses conjugués algébriques sont de module inférieur à 1), auquel on peut associer une substitution (cas Pisot réductible). Nous en discuterons à la section 2.3.2.

---

**Contribution : unification et extension des différentes approches** J'ai écrit un certain nombre de publications dans des revues pour expliciter le lien entre ces différentes méthodes et montrer que l'approche initiale de Rauzy peut-être généralisée convenablement pour englober tous les autres cas.

- Avec V. Canterini, nous avons d'abord construit des fractals pour les substitutions irréductibles et unitaires ((Trans. AMS, 2001) et (JTN Bord., 2001)). et exploité ces fractals pour étudier les propriétés spectrales des systèmes substitutifs (voir section 2.2.7).

Ce travail m'a donné l'occasion de rédiger des synthèses sur ces propriétés spectrales ((Pytheas Fogg Intro, LNM, 2002), (Pytheas Fogg, LNM, 2002), (EWT, 2003)).

- Avec V. Berthé, nous avons construit des fractals pour toutes les substitutions réductibles dont la valeur propre dominante est un nombre de Pisot. Nous avons aussi montré que ces fractals sont des ensembles d'entiers pour des systèmes de numération associés à une substitution (Integers, 2005) (voir section 2.3.2 et les applications à la section 3.2) .
- Dans (Erg. Th. Dyn. Sys., 2003), j'ai montré comment cette approche permet de modéliser les systèmes engendrés par des substitutions non unitaires, qui n'avaient jamais fait l'objet de représentation géométrique auparavant et était simplement considérés d'un point de vue formel sous la forme de points homoclines (voir section 3.2.1, les application en théorie des nombres tout au long de la section 3.2 et les applications en théorie spectrale à la section 3.4.1).

### 1.2.3 Retour à la dynamique symbolique

**Deux classes de recouvrements associés à un fractal de Rauzy** Par construction, un fractal de Rauzy peut-être associé à plusieurs recouvrements de l'espace dans lequel il vit. On parle de recouvrement dans la mesure où on sait que positionner des copies du fractal à des endroits fixes permet de recouvrir tout l'espace considéré.

- Un des recouvrements est périodique : on place une copie du fractal sur tous les sommets d'un réseau ;
- un des recouvrements est autosimilaire : on place toutes les sous-pièces du fractal sur les sommets d'un ensemble localement fini de points ; les emplacements sont tels que lorsqu'on subdivise le fractal et qu'on renormalise, on retrouve exactement les mêmes configurations globales.

**Pavages et applications** La question qui se pose alors est de savoir si ce recouvrement est un pavage, c'est-à-dire si les intersections entre les tuiles sont de mesure nulle. Comme nous le détaillerons à la section 2.3, cette propriété de pavage s'interprète différemment en fonction des applications considérées : on obtiendra une partition de Markov pour un automorphisme du tore avec une unique direction dilatante dès lors que le recouvrement autosimilaire est un pavage, ce qui, dans ce cas, est équivalent au fait que le recouvrement périodique est un pavage. On pourra décrire une addition sur un tore par le système symbolique engendré par une sub-

stitution dès lors que la substitution a des propriétés algébriques (condition Pisot) et que le recouvrement périodique est un pavage.

Il existe une pléthore de conditions de pavages. Elles sont détaillées dans la section 2.3. Ces conditions ont été proposées par différents auteurs en fonction des points de vues considérés : ces conditions s’expriment en termes combinatoires [BK06, IR06, Liv87], de beta-développements [Hol96, Aki02], d’annulations sur des approximations de la frontière [EI05] ou de propriété d’autosimilarité [Mes98, FFIW06]. J’ai moi-même utilisé cette dernière méthode pour montrer que des systèmes substitutifs sont à spectre purement discret dans (Ann. Inst. Fourier, 2004).

Il faut cependant noter que toutes ces approches ont finalement été développées de manière ad hoc sans prendre en compte les différents points de vue existant sur les fractals de Rauzy. En particulier, aucune n’est réellement décidable, et le cas des recouvrements périodiques est souvent laissé de côté.

**Contribution : une condition de pavages unifiée et décidable** Ma contribution dans ce contexte a été de considérer cette question du pavage de manière globale et décidable. J’ai commencé à explorer cette idée dans (Ann. Inst. Fourier, 2004), en considérant des pavages périodiques. En collaboration avec J. Thuswaldner, dans une monographie soumise à *Mémoires de la SMF* (Monographie S. & T., 2008), nous avons continué ce travail pour uniformiser les approches et proposer une condition décidable qui englobe les cas irréductibles et réductibles, pour des recouvrements autosimilaires comme périodiques.

Notre approche consiste à construire un graphe orienté fini qui décrit exactement les pièces qui ont un contact avec le fractal de Rauzy dans le recouvrement considéré. La construction de ce graphe s’appuie sur les propriétés d’autosimilarité du fractal et la description des localisations des pièces dans les recouvrements par des ensembles localement finis. On construit ainsi un *graphe de frontière* dont les sommets sont les emplacements des pièces adjacentes au fractal de Rauzy. Calculer la mesure des intersections entre un fractal et ses voisins revient alors à calculer la valeur propre dominante de la matrice d’adjacence du graphe de frontière, ce qui est totalement décidable. L’intérêt de cette méthode par rapport aux autres décrites précédemment est son “universalité” : elle englobe tous les cas traités par les conditions existantes, les traite de manière décidable, et les étend à des cas inaccessibles jusqu’à présent, en particulier le cas des pavages périodiques par des fractals dans le cas Pisot réductible. Cette approche est détaillée à la section 2.3.3.

**Le cas général ?** Même si de nombreux points restent à éclaircir, on peut considérer que le cas auto-induit correspondant à un nombre de Pisot est maintenant bien compris pour les additions dans les espaces euclidiens. Or, il s’agit là d’une classe très particulière de l’ensemble des additions sur un tore. En effet, si une addition est auto-induite et peut-être représentée par un fractal de Rauzy, son paramètre principal vérifie des propriétés algébriques assez restrictives : il s’agit d’un nombre de Pisot unitaire, valeur propre d’une matrice à coefficients positifs. La généralisation de l’étude symbolique des endomorphismes et des additions sur un tore doit donc être considérée dans de multiples directions. Nous discutons certaines généralisations

dans différentes parties de ce manuscrit.

- La question des paramètres algébriques mais non unitaires sera étudiée sous l’angle des systèmes de numération non unitaires dans la section 3.2, avec la construction d’un nouveau fractal (avec des composantes dans des espaces  $p$ -adiques) qui généralise la construction habituelle. En particulier, dans la section 3.4.1, ce nouveau fractal est utilisé pour étudier la dynamique des systèmes engendrés par une substitution non unitaire (Erg. Th. Dyn. Sys., 2003).

Nous discutons dans la section 3.4.2 comment cette construction pourrait à terme permettre d’obtenir une partition de Markov pour un endomorphisme du tore, c’est-à-dire une application non inversible, en s’appuyant sur un travail en cours avec V. Berthé, P. Surer et J. Thuswaldner sur la représentation de systèmes non unitaires à l’aide de nouveaux systèmes de numération (B., S., S. & T., en cours).

- Une généralisation est aussi de considérer des matrices qui ne sont plus à coefficients seulement positifs. Ici, les objets combinatoires associés ne sont plus des substitutions mais des *automorphismes du groupe libre*, c’est-à-dire des règles de transformations qui autorisent des annulations. Des partitions de Markov pour de tels automorphismes ont été proposées dans [AFHI07], en s’appuyant sur une construction de fractal par approximations successives. Dans la section 3.4.3, nous discutons de la dynamique symbolique pour les laminations attractives de classes d’automorphismes du groupe libre (Ann. Inst. Fourier, 2006), et nous expliquons pourquoi cette approche pourrait permettre d’obtenir des invariants topologiques pour ces classes d’automorphismes.
- Enfin, pour sortir du cadre algébrique, le challenge reste entier pour définir des procédés d’induction généraux pour les additions sur un tore et plus généralement pour comprendre quels sont les codages symboliques que l’on peut obtenir par des procédés d’induction répétés (voir [Fog02, Chapitre 12] et [BFZ05]). Nous abordons cette question à la section 3.3.7, en perspective de travaux sur l’approximation de plans discrets.

### 1.3 La topologie de fractals pour aborder différentes questions interdisciplinaires

En conclusion de cette partie, nous retiendrons qu’une méthode d’étude d’un système dynamique consiste à coder ses orbites dans une partition bien choisie. Cela permet de passer d’objets continus à des objets combinatoires et discrets, et de les étudier avec des points de vue à la fois combinatoires et ergodiques. Les questions relevant de la dynamique symbolique qui sont posées sur les systèmes sont ainsi : quelles sont les bonnes partitions pour coder (éventuellement en faisant appel à des procédés d’induction) ? A-t-on un bon codage (est-il décrit par des mots interdits) ? Quelles sont les propriétés combinatoires et ergodiques des systèmes obtenus ?

Pour choisir une bonne partition pour les automorphismes et les additions sur un tore, nous venons de voir qu’il est naturel de considérer une structure compacte

---

autosimilaire appelée *fractal de Rauzy*, au moins dans le cas où l'application considérée est décrite par un nombre algébrique unitaire dont tous les conjugués sont de module inférieur à 1. Vérifier que cette structure compacte est adaptée pour représenter l'application revient à vérifier une propriété de pavage, question pour laquelle j'ai proposé d'utiliser un *graphe de frontière* qui décrit l'intersection du fractal avec ses voisins dans un recouvrement localement fini.

**Vers des applications autres que dynamiques ?** Il faut insister sur le fait que ces codages symboliques ont des applications *externes* dans différents champs en mathématiques, informatique, ou physique, répondant par là aux ambitions de Poincaré et Hadamard. Nous avons déjà évoqué comment des résultats substantiels en géométrie ont été obtenus en exploitant la combinatoire de points fixes de systèmes d'itération [Mor21] ou l'existence de partitions de Markov [Sma67, Adl98]. De même, en théorie des nombres, les systèmes symboliques, via des aspects ergodiques et combinatoires, ont nourri des conjectures, et des constructions à base de substitutions ont fourni des exemples non-triviaux pour ces conjectures [AS02, ABD06, AB07, Roy03, HM06]. Dans cet esprit, nous allons voir maintenant comment les graphes de frontière peuvent être utilisés dans les domaines d'applications des fractals de Rauzy, ou plus précisément, comment exploiter les propriétés topologiques de fractals pour aborder des questions liées à un processus d'autosimilarité.

**Contribution sur la topologie des fractals** Les propriétés topologiques des fractals s'interprètent souvent en propriétés dans les différents domaines abordés, comme nous le détaillerons dans la section 3.1 : en théorie des nombres, il s'agit de caractériser des développements finis ou purement périodiques [AS05, IR05]. En dynamique, la connexité de la base de la partition est un élément important pour décider si cette partition de Markov est génératrice [Adl98]. En géométrie discrète, la topologie du fractal influe sur la forme d'approximations de plans discrets [ABI02]. En géométrie, l'existence de points de coupure est un bon candidat pour être un invariant par conjugaison par un automorphisme de groupe libre.

Même s'il existe plusieurs résultats disséminés dans la littérature concernant la topologie des fractals (connexité, simple connexité, points intérieurs) [Mes00, Aki02], les méthodes utilisées pour mettre en évidence des phénomènes sont très peu généralisables et concernent des exemples fixés. Avec J. Thuswaldner, dans (Monographie S. & T., 2008), nous montrons que les graphes de frontière permettent non seulement de déterminer si des recouvrements sont des pavages, mais surtout d'étudier la topologie des fractals de Rauzy (voir section 3.1). Les propriétés que nous caractérisons sont les suivantes : mesure de la dimension de la frontière,  $\mathbf{0}$  point intérieur, connexité, homéomorphisme à un disque, simple connexité et groupe fondamental non libre.

### 1.3.1 Quasi-cristaux et espaces de pavages

**Quasi-cristaux** En physique, la dynamique symbolique a fait son apparition dans les années 1980 avec la découverte des quasi-cristaux : en 1984, un cristal, alliage

d'aluminium et de manganèse a été synthétisé et a montré des symétries d'ordre 12. Ceci était théoriquement impossible, dans la mesure où les cristaux sont connus pour avoir des symétries d'ordre 2, 3, 4 ou 6. Le terme de *quasi-cristal* est donc apparu pour désigner ces cristaux avec des symétries interdites. Leur définition n'est pas totalement fixée, cependant, on désigne généralement par ce terme tout solide dont le diagramme de diffraction est essentiellement discret [BT86, LGJJ93, Sen06].

**Modélisation des quasi-cristaux unidimensionnels** D'un point de vue mathématique, la question qui s'est posée (comme pour Hadamard avec ses géodésiques récurrentes) fut de savoir s'il existait effectivement des structures mathématiques pour l'emplacement des atomes qui vérifient ces propriétés. Là encore, la dynamique symbolique (et plus précisément les substitutions) a apporté sa contribution.

En effet, à une substitution on associe un espace de pavages de la droite en considérant des pavages par des intervalles dont la taille et l'ordre sont gouvernés par l'arithmétique et la combinatoire de la substitution. Ces espaces de pavages sont naturellement munis d'une topologie et d'une  $\mathbb{R}$ -action minimale et uniquement ergodique. Les pavages unidimensionnels sont en particulier des candidats naturels pour les cristaux dans la mesure où on peut transformer un pavage en un ensemble discret de points en plaçant un atome à la fin de chaque intervalle. Comprendre si cet ensemble discret modélise correctement un quasi-cristal revient à étudier son spectre de diffraction.

Or, il a été prouvé que le spectre de diffraction d'un ensemble de points positionnés selon un pavage substitutif est purement discret si et seulement si le flot de translation sur l'espace de pavage est lui-même purement discret [LMS02]. On sait aussi que le spectre de cette dernière action est intimement relié au spectre du système dynamique symbolique engendré par la substitution [BK06] (voir section 2.2.3). Savoir si ce spectre est purement discret est donc équivalent à décider si un fractal de Rauzy engendre un pavage autosimilaire, et les graphes de frontière apportent une réponse générale à cette question. Ils permettent de conclure à l'existence de modèles naturels explicites de quasi-cristaux.

**Quasi-cristaux en dimension 2 et plans discrets** Pour généraliser ces résultats en dimension supérieure, le problème est de définir correctement les propriétés de répartition des atomes qui caractérisent un quasi-cristal. Dans le cas des cristaux, les atomes sont répartis selon un réseau. Pour un quasi-cristal, ce réseau est remplacé par une structure non périodique mais qui présente quand même certaines propriétés de répétitivité. Un modèle raisonnable est celui des *ensembles de Meyer*, qui sont obtenus à partir d'un schéma de coupe-et-projection [Moo00]. Les pavages associés à une suite sturmiennes sont des exemples de tels schémas en dimension 1 [GVG04]. En dimension 2, le pavage de Penrose vérifie ces propriétés et il a effectivement un spectre de diffraction purement discret, mais proposer des classes étendues de quasi-cristaux est une tâche ardue [AP98, KP00, KV98, Rob04].

Par analogie avec la dimension 1, de bons candidats pour les schémas de coupe-et-projection sont donnés par les approximations discrètes de plans [ABEI01, BP97]. La notion de discrétisation de surface est formellement apparue dans [Rev91] sous la

forme d'espace arithmétique. Suivant la définition de [ABI02, BV00], on considère un plan  $\mathcal{P}_{(a,b,c)}$  de  $\mathbb{R}^3$  d'équation  $ax + by + cz = 0$ , et on note  $\mathcal{S}$  l'ensemble des cubes à coordonnées entières qui intersectent le demi-espace d'équation  $ax + by + cz < 0$ . Le *plan discret* ou la *surface brisée* associée à  $\mathcal{P}_{(a,b,c)}$  est la frontière de l'ensemble  $\mathcal{S}$ . Ce plan discret est noté  $\mathfrak{P}_{(a,b,c)}$ .

On obtient de bons candidats pour des schémas de coupe-et-projection (et donc de quasi-cristaux), en remplaçant chaque face d'un plan discret par une tuile d'un fractal de Rauzy. Cependant, la généralisation des résultats ergodiques unidimensionnels à ce cadre s'avère difficile à mettre en oeuvre : en effet, les pavages considérés ont des pièces à bord fractal et potentiellement multiples, ce qui n'entre pas dans tous les cadres ergodiques considérés par les chercheurs sur ce domaine (voir par exemple les survols sur les pavages multidimensionnels [Sol06, PF08]).

### 1.3.2 Géométrie discrète et génération de plans discrets

Concernant les plans discrets que nous venons de définir, une question importante est de les engendrer par des itérations de règles de substitutions multidimensionnelles, de manière analogue à ce qui existe pour les approximations de droites du plan qui sont codées par des itérations de substitutions.

**Substitution multidimensionnelle?** Derrière cette question se pose la définition d'une substitution multidimensionnelle. Une première définition *vectorielle* consiste à définir une application qui envoie une face de cube sur une union de faces de cubes, le positionnement de ces faces dépendant du positionnement de la pièce initiale. Le problème ici est de définir une application qui respecte la structure d'un plan discret : les images de deux faces de cubes doivent absolument être totalement disjointes. Dans ce cadre, Arnoux, Ito et Sano ont fait dans [AI01, AIS01] une avancée considérable en définissant une *substitution généralisée*, obtenue comme application duale d'une substitution et qui agit effectivement sur des faces de cubes.

Cependant, cette approche n'est absolument pas efficace pour engendrer un plan discret. Lors de chaque itération, il faut calculer pour chaque pièce du motif un produit de matrice par un vecteur ! Ce n'est pas du tout ce qui se passe dans le cas unidimensionnel, où l'on itère et compose des substitutions en s'appuyant sur la notion d'adjacence d'images de lettres ( $\sigma(UV) = \sigma(U)\sigma(V)$ ), sans calculer explicitement la position de chaque lettre. Si une définition itérative existe pour les substitutions dont les images de lettres sont des rectangles [GA97, AS02, Han00], l'existence de règles d'adjacence dans un cas général est loin d'être évident.

Dans (TCS, 2004), avec P. Arnoux et V. Berthé, nous discutons de l'existence de règles de placements pour la substitution définie sur  $\mathbb{Z}^2$  par  $1 \mapsto \begin{pmatrix} 2 \\ 1 \end{pmatrix}$ ,  $2 \mapsto 3$ ,  $3 \mapsto 1$ .

En exploitant le formalisme de Arnoux-Ito-Sano, nous montrons qu'il est possible de définir des règles de placement pour cette substitution qui, par itération successives, recouvre tout un plan discret fixé. Nous montrons surtout dans (TCS, 2004) que les approches vectorielles et par règles de placements sont globalement équivalentes avec le formalisme de Arnoux-Ito-Sano. On peut donc les considérer



---

simultanément. Des détails sont donnés dans la section 3.3.5.

**Engendrement de plans discrets** Pour généraliser (TCS, 2004), nous avons cherché à savoir s’il était possible d’engendrer des plans discrets en itérant toujours une même substitution généralisée. Dans le cas unidimensionnel, on sait que les droites discrètes engendrées par une unique substitution ont pour pente un nombre quadratique. Engendrer un plan discret avec une seule substitution engendre aussi des restrictions sur le vecteur normal du plan discret : en étudiant les détails du formalisme, on doit supposer que ce vecteur normal est le vecteur dilatant d’une matrice  $\mathbf{M}$  à coefficients positifs avec une unique direction dilatante, on retrouve ainsi le cas Pisot unitaire étudié pour les partitions de Markov dans la section 1. En particulier, on peut associer un fractal de Rauzy à la matrice  $\mathbf{M}$ .

Dans ce cadre, dans (Integers, 2005) nous avons montré qu’il est possible d’engendrer le plan discret par itérations d’une substitution généralisée sur les faces du cube unité si et seulement si  $\mathbf{0}$  est un point intérieur du fractal de Rauzy. On revient ici aux propriétés topologiques des fractals.

Dans un survol sur les conditions de pavage que j’écris actuellement avec V. Berthé et J. Thuswaldner pour un chapitre d’un ouvrage collectif qui sera édité chez Cambridge University Press (CANT, 2009), nous donnons aussi une condition explicite en terme de géométrie discrète pour vérifier cette condition : il faut et il suffit qu’une pièce de taille fixée du plan discret soit recouverte par une itération donnée de la substitution généralisée. Ces travaux sont décrits dans la section 3.3.3.

**Une stratégie d’engendrement globale des plans discrets ?** La question générale est quand même de proposer une stratégie d’engendrement de n’importe quel plan discret. Dans ce cadre, un travail fondateur [IO94] cherche à montrer que tous les plans discrets sont uniformément récurrents. L’idée est d’approximer le vecteur normal d’un plan discret par des vecteurs rationnels, en utilisant un algorithme de fractions continues multidimensionnelles, l’algorithme de Jacobi-Perron (c’est exactement la méthode utilisée dans le cas des approximations de droites). À chaque matrice utilisée dans l’algorithme de fractions continues, ils associent une substitution multidimensionnelle. Ensuite, ils composent adroitement ces substitutions pour engendrer des morceaux du plan discret normal au vecteur considéré au départ.

La question qui reste à déterminer est de vérifier que les morceaux grossissent suffisamment pour engendrer *tout* le plan discret. Pour cela, ils utilisent un lemme qui explique que les morceaux grossissent à partir du moment où ils ont atteint une taille raisonnable. Ils décrivent avec un graphe fini les cas pour lesquels cette taille raisonnable n’est pas atteinte. Même si ce travail est remarquable, il est difficilement généralisable en l’état, et le domaine de validité du lemme utilisé est assez restreint (voir discussion section 3.3.4).

Avec V. Berthé et J. Bourdon, nous travaillons actuellement sur une généralisation du travail de Ito et Ohtsuki pour l’adapter à différents algorithmes de fractions continues (B., B. & S., en cours). Nous sommes partis d’un travail en collaboration avec P. Arnoux (Jour. Montoises, 2006), où nous montrons que pour engendrer une famille de plans discrets, il faut superposer deux propriétés bien distinctes (engen-

drement et extension). Dans la section 3.3.4, nous détaillons comment il est possible de vérifier ces deux propriétés pour plusieurs algorithmes d'engendrement. Le point-clé est de faire appel à un critère topologique sur motifs images de substitutions multidimensionnelles pour vérifier la propriété d'engendrement.

La perspective finale de ce travail rejoint les questions dynamiques discutées dans la section 1.2.3 : il va s'agir de construire une suite de domaines fondamentaux qui permettraient de représenter symboliquement la dynamique n'importe quelle addition sur un tore par la composition de substitutions (conjecture  $S$ -adique).

### 1.3.3 Étude des systèmes de numération en base non entière

**Systèmes de numération** Explicitons maintenant le lien entre les fractals de Rauzy et les systèmes de numération. On considère un nombre  $\beta > 1$ . De manière analogue aux développements binaires ou décimaux, il est possible de décomposer tout  $x \in [0, 1[$  en base  $\beta$ , c'est-à-dire sous la forme  $x = \frac{a_0}{\beta} + \frac{a_1}{\beta^2} + \dots + \frac{a_n}{\beta^n} + \dots$ . Les chiffres  $a_n$  ne sont pas quelconques : par analogie avec le développement décimal, ils sont obtenus par un algorithme glouton relié à la *beta-transformation*  $T_\beta : x \rightarrow \beta x - E(\beta x)$ . Ils appartiennent donc à l'ensemble  $\{0, 1, 2, \dots, E(\beta)\}$ .

Cette écriture est liée aux systèmes dynamiques dans la mesure où le développement d'un réel  $x$  n'est rien d'autre que le codage de l'orbite de  $x$  sous l'action de la beta-transformation, pour la partition donnée par ses intervalles de continuité [Par60].

Les questions naturelles qui se posent lors de l'étude d'un système de numération sont avant tout de caractériser les développements des nombres, puis de déterminer les éléments qui ont un développement fini, périodique, ou ultimement périodique.

Pour les beta-numérations, on a des réponses assez précises aux deux premières questions. En effet, on sait que les développements des réels  $x$  sont entièrement caractérisés par l'écriture de 1 en base  $\beta$ , notée  $d_\beta^*(1)$ . Si  $\beta$  est un nombre de Pisot, on utilise les propriétés algébriques de  $\beta$  pour caractériser aussi les développements ultimement périodiques (il s'agit des éléments de  $x \in \mathbb{Q}(\beta)$ ) (voir section 2.1.5).

**Développements finis et topologie** L'identification des développements finis est par contre bien plus complexe. Par construction, si un nombre a un  $\beta$ -développement fini, alors il appartient à l'ensemble  $\mathbb{Z}[1/\beta] \cap \mathbb{R}^+$ . La question qui se pose est la réciproque : quels sont les nombres  $\beta$  pour lesquels tous les éléments de  $\mathbb{Z}[1/\beta] \cap \mathbb{R}^+$  ont un développement fini ? Si cette propriété est vraie, on dit que  $\beta$  vérifie une propriété de finitude, ou *propriété (F)*. Cette propriété est bien comprise dans le cas Pisot unitaire [FS92, Aki00], où elle est même caractérisée en termes topologiques :  $\mathbf{0}$  doit être un point intérieur du fractal de Rauzy associé à la numération [Aki02].

Dans le cas où  $\beta$  est un nombre non unitaire, cependant, le fractal de Rauzy ne vérifie pas de propriété de pavage : l'action de la matrice associée à  $\beta$  sur son hyperplan contractant n'a pas un ratio suffisant. Ceci empêche d'utiliser les approches

géométriques ou arithmétiques pour relier la propriété (F) au fait que  $\mathbf{0}$  est point intérieur du fractal.

Pour traiter ce cas, j'ai proposé dans la publication (Erg. Th. Dyn. Sys, 2003) de construire un nouveau fractal de Rauzy qui tient compte de l'arithmétique sous-jacente à  $\beta$ . Puisque  $\beta$  est un nombre non unitaire, il existe des corps locaux (extensions finies de corps  $p$ -adiques) dans lesquels  $\beta^n$  converge exponentiellement vite, et dans ces corps on peut faire converger des séries basées sur des systèmes de numération. J'ai donc introduit des fractals de Rauzy complets, incluant des composantes  $p$ -adiques, et j'ai montré en collaboration avec G. Barat, V. Berthé, et S. Akiyama dans (Monas. Math., 2008) que ces fractals induisent un pavage autosimilaire d'un espace constitué de corps euclidiens et locaux dès que la propriété (F) est vérifiée. En fait,  $\mathbf{0}$  est un point intérieur du fractal complet et du fractal de Rauzy original, même si ce dernier ne pave pas l'hyperplan dans lequel il vit. Des détails sont donnés dans la section 3.2.

**Développements purement périodiques** Les points de vue dynamiques et topologiques permettent aussi de mieux comprendre les développements périodiques en beta-numération. Pour le développement en fractions continues, le théorème de Galois dit qu'un nombre a un développement purement périodique si et seulement s'il est quadratique et son conjugué est de module inférieur à 1. La question qui s'est posée a été de caractériser les développements purement périodiques en base  $\beta$ .

Or, les techniques habituelles pour rechercher les développements purement périodiques consistent à s'appuyer sur une *extension naturelle* de la dynamique, qui corrige son défaut d'injectivité [DKS96]. Différents travaux [IR06] ont montré que dans le cas Pisot unitaire, on peut construire une extension naturelle pour  $T_\beta$  comme produit de la  $\beta$ -transformation  $T_\beta$  et d'une contraction diagonale sur le fractal de Rauzy (en fait, cette construction est très proche du domaine de Markov discuté dans la section 1.1). À partir de cette extension naturelle, on déduit un théorème de Galois pour les développements en base  $\beta$  Pisot unitaire : un élément de  $\mathbb{Q}(\beta)$  admet un beta-développement purement périodique si et seulement si le point formé par ses conjugués appartient au fractal de Rauzy associé à  $\beta$  et s'il vérifie des conditions métriques lui-même [IR06]. Par contre, la démonstration de [IR06] n'est valable que dans le cas unitaire.

Comme nous le détaillerons dans la section 3.2, Dans (JNT, 2007), V. Berthé et moi-même avons proposé une extension de ce résultat au cas non unitaire, en utilisant les fractal de Rauzy  $p$ -adiques, et en donnant une nouvelle preuve qui fait bien apparaître le rôle de l'extension naturelle.

**Rationnels avec un développement purement périodique** Cette caractérisation des développements périodiques à l'aide de fractals permet d'exhiber de nouvelles propriétés des développements. Je me suis en particulier intéressée aux rationnels dont le développement purement périodique. Lorsque  $b$  est un entier, on sait depuis longtemps que les nombres rationnels  $p/q$  avec un développement périodique en base  $b$  sont exactement ceux pour lesquels  $q$  est premier avec la base  $b$ . Une question est de généraliser ce résultat en base non entière. Pour le cas qua-

dratique unitaire, on dispose d'une bonne généralisation, liée à la propriété (F) [Sch80, HI97, Aki98].

En particulier, si  $\beta$  est un nombre Pisot unitaire, Akiyama a montré que dès que la propriété (F) est vérifiée, il existe un intervalle non vide de la forme  $[0, \epsilon[$  dont tous les rationnels ont un  $\beta$ -développement est purement périodique [Aki98]. Par contre, tous les rationnels de  $[0, 1[$  ne vérifient pas cette propriété ; on appelle  $\gamma(\beta)$  le plus grand intervalle de la forme  $[0, a[ \cap \mathbb{Q}$  dans lequel les rationnels de dénominateur premier avec la norme de  $\beta$  ont un développements périodiques. Pour le plus petit nombre de Pisot, on sait que  $\gamma(\beta)$  est approximé par 0.66666666608644067488. Tous les rationnels n'ont donc pas un développement périodique.

Ceci est assez inattendu, dans la mesure où on attend que la répartition des développements périodiques en base  $\beta$  soit *aléatoire*, en tout cas indépendante de la position par rapport à 0. Par contre, comme nous le verrons dans la section 3.2), ceci devient moins mystérieux si on prend en compte le fait qu'une caractérisation des développements périodiques est donnée par un fractal de Rauzy, et est donc reliée à sa topologie.

- Dans (Monas. Math., 2008), nous avons caractérisé  $\gamma(\beta)$  en terme d'intersection de la frontière du fractal de Rauzy  $p$ -adique, de droite dans les espaces euclidiens et de bandes dans les espaces  $p$ -adiques. En considérant un graphe de frontière pour le fractal de Rauzy  $p$ -adique, nous sommes parvenus à calculer la valeur exacte de  $\gamma(\beta)$  pour deux nombres quadratiques non unitaires :  $\gamma(2 + \sqrt{7}) = 0$  alors que ce nombre vérifie la propriété (F) (il s'agit de la racine de  $X^2 - 4X - 3$ ) et  $\gamma(5 + 2\sqrt{7}) = (7 - \sqrt{7})/12$ , qui vérifie toujours la propriété (F) (il s'agit de la racine de  $X^2 - 10X - 3$ ).
- Dans un travail en cours avec B. Adamczewski, W. Steiner et C. Frougny (A., F., S.& S., en cours), nous avons utilisé la description de la frontière par le fractal de Rauzy pour montrer que dans le cas cubique unitaire  $\gamma(\beta)$  est un nombre irrationnel. En particulier, nous prouvons ainsi que la quantité 0.66666666608644067488 estimée dans [AS05] est bien l'approximation d'un nombre irrationnel. Il faut bien noter que les preuves d'irrationalité faisant appel à des considérations topologiques sont extrêmement rares.

## 1.4 Utilisation de systèmes dynamiques en sciences expérimentales

Jusqu'à maintenant, nous avons décrit des systèmes dynamiques considérés par des mathématiciens, mais ce faisant nous avons oublié la motivation initiale et mis la charrue avant les bœufs. En effet, les systèmes dynamiques ont été introduits pour permettre de mieux comprendre les observations faites dans différents cadres *expérimentaux*. Le problème se pose alors légèrement différemment : on dispose d'observations sur un système, plus ou moins variées. On fait l'hypothèse que ce système est régi par une loi dynamique et peut être décrit par une dynamique différentielle continue de la forme  $\frac{dX}{dt} = F(X, t)$ . L'objectif est de proposer des lois explicites puis d'exploiter la connaissance des lois pour décrire les trajectoires.

---

Ainsi, la mécanique céleste a consisté à proposer des lois raisonnables pour décrire les trajectoires des planètes [Arn99]. De la même manière, les lois d'action de masse, ou de Michaelis-Mentens ont été proposées pour décrire le comportement de systèmes chimiques [CB04, HS96]. Des modèles spécifiques ont été introduits pour décrire le comportement de systèmes économiques, ou la dynamique des populations [Tur03]. Les chaînes de régulations transcriptionnelles avec une rétro-régulation négative sont bien décrites par un modèle de Goodwin [Goo63, Goo65]. Plus globalement, des sigmoïdes de type “loi de Hill” décrivent bien les observations expérimentales faites sur des régulations transcriptionnelles [Yag75]. Chaque loi fait intervenir des paramètres, que des observations variées permettent d'estimer avec une précision plus ou moins grande.

Dans tous ces cadres, en plus de la question dynamique de l'étude du futur des éléments, se pose la question de la *fiabilité* des prédictions ou des simulations. En effet, même si le nombre d'observations sur un système peut-être important (il suffit de considérer la quantité phénoménale d'observations intégrées dans des modèles météorologiques [Pie01, Hol04]), ces observations sont toujours limitées. Or, ce sont ces observations qui définissent les états du système. Cependant, lorsqu'on a à faire à des dynamiques chaotiques, on sait que, par essence, la moindre erreur dans la définition d'un état générera impérativement une erreur importante dans les prédictions. Il est ainsi fondamental de prendre en compte le niveau de fiabilité à la fois dans l'estimation des paramètres du modèle et dans la qualité des observations, pour évaluer le caractère prédictif des modèles.

### 1.4.1 Spécificités de la biologie moléculaire

**Prédictions basées sur la multiplication du nombre d'observations** Une première approche consiste alors à multiplier le nombre d'observations pour complexifier les modèles et augmenter la fiabilité de leurs prédictions, au moins à moyen terme. C'est la stratégie utilisée lorsque d'une part on a la possibilité technique de produire des observations variées sur le système, et d'autre part on cherche à produire des prédictions quantitatives à court ou moyen terme sur un système, sans en déterminer le comportement à long terme. L'exemple typique est ici donné par les modèles météorologiques [PG03, Kal03].

Sans disposer d'autant de données, la biochimie entre dans ce cadre, dans la mesure où les produits entrant en réaction sont en quantité suffisante pour rendre l'hypothèse déterministe réaliste, et où on dispose de la possibilité matérielle de tester des voies métaboliques et obtenir des valeurs de paramètres, ainsi qu'un indice de leur fiabilité [CB04, CBJS05]. À partir de ces modèles, on procède généralement à des simulations de modèles différentiels [HSG<sup>+</sup>06, THT<sup>+</sup>99] avec des applications concrètes en pharmacologie par exemple [BMOH97, ECB98]. Des analyses plus formelles concernent la recherche d'états stationnaires et leur stabilité [AFS04]. Ensuite, une communauté importante a exploré les propriétés de la dynamique fournie par la topologie et la stœchiométrie des réactions [PSP<sup>+</sup>04] : les *voies élémentaires* sont en fait des espaces affines invariants pour cette dynamique, qui fournissent une analyses des voies qui fonctionnent indépendamment [SFD00]. L'existence d'inva-

riants thermodynamiques (coefficients de contrôles et d'élasticité) [BB73] permet des analyses de sensibilité au voisinage d'un état stationnaire [Red88, Fel97, HS96] qui sont exploitées via des approches de contraintes basées sur de l'analyse convexe [PPP02, LB06] ou d'optimisation [SVC02, KPE03, SKB<sup>+</sup>02].

De même, en biologie végétale ou en dynamique des populations, les modèles de croissance sont bien compris et permettent de mettre en œuvre des simulations numériques qui donnent des résultats réalistes [PL90, Var03].

**Biologie moléculaire** En biologie moléculaire par contre, la situation est totalement différente. Les séquences d'ADN sont analysées avec des méthodes dynamiques, en particulier pour identifier des signatures de génomes. Cependant, comme nous le discuterons dans la section 4.1.1, on ne peut pas aujourd'hui considérer que les séquences d'ADN sont des systèmes dynamiques symboliques, au sens où elles ne codent pas les trajectoires d'un système sur un espace.

Même si les génomes ne sont pas les codages de la dynamique d'une molécule, tous les mécanismes permettant de passer d'un génome à l'expression explicite d'une protéines sont des phénomènes dynamiques qui peuvent être étudiés par discrétisation. Plus précisément, des modèles décrivant les interactions entre molécules existent depuis plus d'un demi siècle ; cependant, les observations sur les *états* de la cellule étaient extrêmement éparpillées jusqu'à la fin des années 1990.

Or, il y a une quinzaine d'années, un saut technologique a eu lieu, sous la forme de puces (à ADN, protéines-protéines, oligo...) qui permettent d'observer *simultanément* la concentration de milliers de molécules au sein d'une cellule. Le spectre de molécules observables est très vaste : ARN messagers [LW00], métabolites [Fie02, Kel04], protéines [PM00, PBZ<sup>+</sup>03], flux métaboliques [Sau04]... Ce saut technologique a marqué la naissance d'un nouveau domaine en biomathématiques et bioinformatique, appelé *biologie systémique*. Le challenge est ici de proposer des méthodes et des concepts pour exploiter ces observations [Kit02a, Pal06, Kan06, Alo06].

Ce domaine englobe d'abord des problématiques statistiques, dans la mesure où les observations sont en quantité limitée (pour des raisons de coût financier) et de qualité médiocre, ce qui nécessite des méthodes avancées pour en extraire une information fiable des données bruitées [Fri04, Kit01]. Le domaine de la biologie systémique concerne surtout le monde des systèmes dynamiques, dans la mesure où il renseigne (enfin !) sur l'état complet des cellules. Si on reprend le formalisme introduit au début de ce chapitre, on dispose maintenant à la fois de lois dynamiques et d'observations des états ; il devient donc légitime de chercher à prédire le comportement des molécules à l'intérieur des cellules [JP06].

**Spécificité** Cependant, le traitement des problèmes posés en biologie moléculaire ne peut pas importer les méthodes développées en physique ou en biochimie et ceci pour deux raisons principales. D'abord, le nombre de composants en jeu dans des processus génétiques est assez faible ; l'effet du bruit est important et la modélisation par une dynamique différentielle déterministe est assez peu réaliste. Tout au moins, les comportements attestés par des approches de dynamique différentielle doivent être suffisamment robustes à un niveau de bruit important. Autre raison, l'obtention

de réplicats pour déterminer les paramètres des lois est très coûteuse. Dans ce contexte, il est extrêmement complexe de construire des modèles quantitatifs et fiables à assez grande échelle. Les modèles quantitatifs restent ainsi exceptionnels (on notera la régulation du cycle cellulaire [TCN01, CCCN<sup>+</sup>04] ou les voies de transduction de signal associées à EGF [SEJGM02] et NF $\kappa$ B [NIE<sup>+</sup>06]).

Partant de ce constat, différentes approches ont été explorées pour simuler et modéliser la dynamique des réseaux de régulation génétique. Certaines restent différentielles (en se focalisant sur des réseaux de petite taille), d'autres discrètes (voir section suivante); d'autres encore adoptent un point de vue probabiliste, voire stochastique [RWA02, KEBC05]. La revue de référence en la matière est [DJ02].

### 1.4.2 Systèmes dynamiques en biologie moléculaire

**Dynamique symbolique pour des systèmes linéaires par morceaux** Une école importante de la modélisation en biologie systémique travaille avec un point de vue qualitatif et a finalement adopté une approche de dynamique symbolique. L'idée est avant tout de discrétiser l'espace des états, pour ramener les analyses à des modèles finis.

La première approximation qui est faite est de remplacer les lois de Hill qui sont utilisées pour modéliser les régulations transcriptionnelles par des applications linéaires par morceaux; différents travaux ont montré que cette approche conserve les propriétés du système biologique en terme d'attracteurs et de comportements périodiques [Sno89]. À partir d'une dynamique définie par des équations linéaires par morceaux, une partition de l'espace est naturellement définie par les intervalles de continuité de l'application. On peut donc chercher à comprendre quelle est la dynamique symbolique obtenue à partir de cette partition. Il a été montré que cette dynamique symbolique est complexe dès que deux ou trois gènes sont impliqués dans le système [Far06b, Far06a, LU06, CFLM06] et nécessite des investigations poussées.

**Sur-approximation markovienne** Pour obtenir des systèmes simulables et exploitables, il a donc été proposé de sur-approximer les dynamiques par des applications markoviennes. Pour cela, on conserve la partition de l'espace selon les intervalles de continuité de l'approximation linéaire par morceaux. On construit une dynamique en supposant que chaque pièce de la partition peut être suivie de tous les éléments qui sont dans la direction de l'attracteur de la dynamique sur la pièce considérée. Ce faisant, on ne différencie plus les points au sein d'une pièce, et on dispose d'une dynamique sur un ensemble fini définie par des contraintes [DJ02]. Des raffinements de cette approche pour mieux identifier les attracteurs s'appuient sur la théorie du contrôle et permettent de rajouter à la dynamique symbolique, en plus des intervalles de continuité, des pièces pour chaque élément de la frontière (solutions de Filipov) [GS02].

La grande force de cette théorie est que la détermination des enchaînements admissibles ne dépend pas de la valeur des paramètres de la dynamique considérée au départ, mais seulement du nombre d'états pour chaque variable ayant un effet

différencié sur la régulation du système, et de la comparaison des seuils de passage à chaque état. Le fait qu’il n’existe qu’un nombre fini d’états avec une influence différente sur le comportement du système est une des spécificités des mécanismes de régulation génétique. En d’autres termes, les informations nécessaires pour déterminer les conditions d’admissibilité dans la dynamique symbolique sont qualitatives et peuvent être obtenues par une lecture approfondie de la littérature couplée avec quelques expérimentations.

Finalement, cette approche permet concrètement d’abstraire un système dynamique portant sur des régulations transcriptionnelles en un système de contraintes locales sur des trajectoires vues comme des systèmes symboliques. Pour ces systèmes, il devient possible d’explorer la dynamique via des approches informatiques qui entrent dans le cadre de la vérification de modèles [dJGHP03]. En codant les contraintes du système, on peut d’abord simuler les trajectoires du système et surtout savoir, via des approches en méthodes formelles, quels sont les états stationnaires, s’il existe des trajectoires périodiques, ou si certains états peuvent être atteints dans certaines conditions. Des illustrations de cette approche ont été données dans le cadre de la régulation du stress nutritionnel chez la bactérie *E. Coli* [BRdJ<sup>+</sup>05], la virulence d’une bactérie pectinolytique [SRN07] ou le comportement de protéines de signalisation [GA04].

**Les modèles de logiques multivaluées et réseaux booléens** Une alternative à cette approche consiste à oublier totalement le point de vue différentiel et abstraire le système en un ensemble de règles logiques. Les éléments du réseau deviennent alors des variables qui peuvent admettre plusieurs états logiques (on parle d’états multivalués) [Tho79, dJRCT05]. Les interactions sont modélisées par des transitions qui à chaque état associent son attracteur ; cela signifie qu’on ne définit pas le successeur d’un élément dans la dynamique mais la position relative de l’attracteur vers laquelle l’élément va se diriger. On dispose ici d’un peu plus de liberté que dans le cas linéaire par morceaux, en particulier pour définir les lois de transitions. Différentes analyses peuvent être faites pour ces modèles, partant de modèles synchrones ou asynchrones [Tho91], en utilisant à nouveau des approches en méthodes formelles : utilisation de logique temporelle pour requêter les trajectoires [BCRG04] ou parcours d’arbres de décisions représentant l’espace des états [NTC07]. Des études poussées ont été menées avec ce formalisme au sujet du développement embryonnaire de la drosophile [GCT08] ou du réseau de régulation sous-jacent au développement de la mucoviscidose [GBC<sup>+</sup>04].

Les modèles logiques multivalués sont apparus dans les années 1990, suivant une évolution assez naturelle depuis les premiers réseaux booléens proposés dans les années 1970 comme modèles aléatoires des réseaux de régulation génétique [Kau69]. Dans les réseaux booléens, on considère que les gènes sont allumés ou éteints, et que des règles logiques permettent de prédire de manière déterministe l’état d’un gène au temps  $n + 1$  en fonction du temps  $n$ .

Les réseaux booléens sont en fait une dynamique symbolique déterministe “simple” dans la mesure où les successeurs au temps suivant sont déterminés uniquement en fonction de la valeur des variables au temps donné. Avec les logiques multiva-



luées (ou les fonctions linéaires par morceaux), l'approche devient non-déterministe puisque plusieurs états peuvent être les successeurs d'un état donné, puisque ces successeurs sont déterminés en fonction des attracteurs de la dynamique. De plus, l'approche booléenne est plus simple que l'approche multivaluée car le nombre de valeurs qu'une variable peut prendre est limité à 2, ce qui n'est pas suffisant pour décrire la complexité des régulations entre de multiples molécules.

Des simulations sur ces réseaux ont montré que les réseaux booléens ne sont pas assez précis pour décrire la dynamique : il peut manquer des trajectoires observées dans la réalité. Ils sont donc peu employés pour décrire la trajectoire des réseaux de régulation. Par contre, ils ont montré leur intérêt pour l'étude des propriétés globales des réseaux grande-échelle, avec des approches probabilistes [Kau93, SL98] illustrées principalement sur le réseau de la levure *S. cerevisiae* [KPBT03, KPST03].

**Limites de l'approche markovienne** Ces approches informatiques sont extrêmement puissantes pour la simulation et l'étude de la dynamique d'un réseau de régulation génétique. Elles sont cependant limitées par la taille des réseaux qui peuvent être étudiés : du fait de l'explosion combinatoire des trajectoires considérées et de la quantité d'information biologique (même qualitative) nécessaire pour effectivement réaliser le modèle, les études sont limitées à deux ou trois dizaines de sommets. Une autre limitation est l'intégration dans le réseau de processus continus ou avec des échelles de temps différentes. Dans ce cas, la discrétisation des états est moins naturelle, et la notion d'attracteur pour la dynamique renseignable à partir de la littérature peu évidente.

**Représentation par des langages et exploration combinatoire** Pour étendre la classe de modèles étudiables, une piste explorée a été de travailler sur la présence des molécules aux différentes étapes des transformations; elle concerne les modèles sous forme de graphes de réactions. Dans ce cadre, il est généralement proposé de modéliser un réseau dans un langage ad-hoc, puis de proposer une sémantique dynamique pour les éléments du langage (booléen, différentielle ou autre), et ensuite d'interroger le système en fonction de cette sémantique, généralement avec des algèbres de processus, des logiques temporelles ou des parcours de diagrammes de décision [DL04, RPS<sup>+</sup>04, CFS06, APUM03].

Les questions posées sont variées : citons entre autres l'atteignabilité, l'existence d'états stationnaires ou d'oscillations (dans le cas différentiel), l'enchaînement de réactions pour produire une molécule [CRCD<sup>+</sup>04, MRM<sup>+</sup>08]. La correction de modèle est abordée par le biais de la recherche de réactions pour expliquer des requêtes. On retrouve ainsi le même genre de questions que celles évoquées dans le cas de la modélisation linéaire par morceaux, si ce n'est que la classe de modèles étudiables est plus large; en contrepartie, on ne peut pas procéder à des simulations de la dynamique.

**Prédictions à partir d'invariants robustes** Proposer des prédictions à partir d'un modèle présenté sous la forme d'un graphe d'interactions ou d'un graphe

de réactions revient à proposer des analyses *robustes* d'un modèle, c'est-à-dire indépendantes de la valeur exacte des paramètres et ne prenant en compte que les informations connues sur ces paramètres. Un exemple couramment utilisé dans ce domaine est la *règle de Thomas*, qui dit que si un système admet plusieurs états stationnaires non dégénérés, alors son graphe d'interactions contient au moins un cycle positif [Tho81]. On a ici une information globale sur le système, ayant un sens biologique important (la bistabilité du système, induisant des phénomènes d'hystérésis), sans une étude détaillée de la dynamique. Cette règle, proposée dans les années 1980, a été prouvée pour les différents formalismes existants (logique, multivalué, différentiel...) dans les années 2000 [Sou06], en faisant appel à un résultat général sur l'univalence de flots différentiel, le théorème de Gale-Nikaido [Par83].

Plus généralement, les travaux mathématiques existant sur les prédictions de systèmes à partir d'informations qualitatives se concentrent pour la plupart sur l'étude de signes de déterminants : les informations au sujet de l'existence ou l'unicité d'attracteurs et leur stabilité sont principalement contenues dans la matrice jacobienne de la dynamique (en particulier ses sous-déterminants), qui est précisément donnée par le graphe d'interaction du réseaux [HKG]. Comme nous le détaillerons dans la section 4.1.3, les approches qui se distinguent à ce sujet sont les réductions par échelles de temps et les réductions de systèmes monotones [Son07].

**Contribution : comparaison d'états stationnaires pour appréhender la dynamique** J'ai axé mes recherches sur des méthodes de prédiction robustes pour un réseau, dans des modèles assez peu renseignés dont la taille est croissante.

Du fait de l'absence de connaissances sur la valeur numérique des paramètres, les méthodes de prédictions vont s'appuyer sur la comparaison de différents états stationnaires (modèle non régulé génétiquement, muté, régulation génétique). Comme nous le détaillerons dans le chapitre 4, cette restriction permet quand même de comprendre le fonctionnement d'un système, en particulier de détecter les zones qui méritent une modélisation fine. Dans ce travail, nous verrons comment coupler différents niveaux d'abstraction permet de comprendre le fonctionnement d'un unique processus biologique : la régulation du métabolisme des acides gras.

D'abord, lorsqu'un système est modélisé sous la forme de graphes de réaction, l'étude des sous-déterminants du graphe de réaction à l'aide du théorème de Gale-Nikaido permet de caractériser les effets de certaines régulations. Nous mettons en évidence que les régulations génétiques permettent d'augmenter l'effet tampon d'un système métabolique sur les variations d'énergie. Nous montrons aussi que le rôle des grandes classes d'acides gras dans leur métabolisme doit être précisé : dans la littérature, on a souvent tendance à distinguer les acides gras saturés ou mono-insaturés des acides gras poly-insaturés ; une de nos conclusions est que cette classification n'est pas assez fine pour rendre compte des observations sur le système (R., S., P., C.& L. en cours).

Ensuite, lorsqu'un système est modélisé par son graphe d'interaction, des résultats sur la réponse d'un système à des paramètres extérieurs permet d'appréhender le rôle de boucles positives dans le comportement du système (Roy. Soc. Inter, 2006). Cela nous amène à concentrer nos analyses sur la voie des désaturations des aci-

des gras dans leur métabolisme. Des raisonnements sur les observations pendant un jeûne suggèrent que cette voie admet un régulateur inconnu. Or, nous montrons que cette voie est en fait contrôlée par une boucle positive qui fausse les raisonnements intuitifs et enlève les incohérences.

Enfin, à un niveau dynamique beaucoup plus précis (quantitatif en particulier), avec une analyse de données temporelles sur les différentes classes d'acides gras pendant un jeûne chez des souris mutées, P. Blavy a illustré pourquoi la boucle positive n'est pas active, et comment la modélisation de procédés dynamiques de transformation inter acides gras est suffisante pour conclure à l'existence de régulations inconnues (JOBIM, 2008).

### 1.4.3 Comment construire des réseaux sur lesquels appliquer des simulations dynamiques ?

**Pourquoi reconstruire des réseaux ?** Il existe donc un certain nombre de méthodes inspirées par les systèmes dynamiques qui permettent d'étudier un système en biologie moléculaire. Cependant, ces travaux sont limités par la quantité d'informations disponibles sur un système. En particulier, les approches d'analyse dynamique ne peuvent porter que sur les interactions connues et assez bien détaillées au sein d'un réseau. La question qui se pose est donc d'aider le biologiste à identifier dans l'ensemble des molécules qui sont actives dans son système, un tout petit nombre de produits sur lesquels une modélisation fine pourra être appliquée. Ce domaine ne fait à priori plus partie de l'étude de systèmes dynamiques. Il s'agit de travailler à partir des observations et des connaissances sur un système pour suggérer l'existence d'interactions que des expérimentations précises permettront de valider. Cependant, puisqu'on recherche des *régulateurs* à partir d'*observations* sur un système, il est légitime d'imaginer qu'une expertise en systèmes dynamiques peut avoir une valeur ajoutée.

Différentes approches existent pour identifier les interactions potentielles qui existent entre différentes molécules. On entre alors dans le domaine de la *reconstruction de réseaux* (voir section 5.1.1) : il s'agit principalement de rassembler les informations dans la littérature ou d'optimiser des modèles probabilistes à partir de données grande-échelle sur un mécanisme (voir section 5.1.1).

**Diagnostic et règle causale** Différentes analyses [FMS<sup>+</sup>07] ont montré qu'aucune méthode n'est convaincante prise isolément, et les réseaux grande-échelle qui apparaissent actuellement sont finalement une compilation d'interactions déduites de différentes méthodes et expérimentations sous différentes conditions. Il a été nécessaire de proposer des méthodes pour concilier différentes données hétérogènes, repérer les incohérences pour raffiner et améliorer les modèles. Pour répondre à ces deux questions, différentes méthodes ont été proposées. Elles diffèrent par le type de règles sur lesquelles s'appuient la notion de cohérence ou les prédictions. Cependant, on retrouve dans la plupart une règle implicite, appelée *règle causale*, qui affirme que la modification de l'expression d'une protéine lorsque un gène est supprimé ou sur-exprimé implique qu'une régulation (au moins indirecte) existe entre le gène

manipulé et la protéine observée [DWFS99].

Au sein du projet Symbiose à l'IRISA, je me suis particulièrement concentrée sur la manière dont les systèmes dynamiques permettent d'analyser les connaissances disponibles (même incomplètes) sur un ensemble large de molécules en fonction des données pour identifier des petits ensembles de molécules sur lesquels les connaissances sont incomplètes, ouvrant la voie, en fonction des cas, à la recherche de régulations additionnelles ou de modélisation fines.

Au sein du projet Symbiose, en collaboration avec M. Le Borgne, O. Radulescu et les doctorants P. Veber et C. Guziolowski, nous avons travaillé sur la règle causale pour développer une méthode pour le diagnostic et l'analyse de réseaux qui tient compte du manque de précision des données et des connaissances.

**Validité de la règle causale ?** Comme nous l'avons déjà mentionné, les modèles causaux avaient d'ailleurs été proposés dans les années 80 [Kau69] pour étudier la dynamique de systèmes ; ils ont été abandonnés au profit de modèles de régulation par seuil quand il a été montré que l'ensemble des trajectoires obtenus pour les modèles causaux n'englobaient pas tous les comportements observés. Leur utilisation pour le diagnostic de réseau est donc à priori plus que sujette à caution. Comme nous le détaillerons dans la section 5.1.2, nous avons d'abord étudié le cadre de validité de la règle causale ((Biosystems, 2006) et (Roy. Soc. Inter, 2006)). Même si elle est largement utilisée, nous avons montré que cette règle n'a un sens que lorsqu'on décrit les variations de produits entre un état de départ (stable) et l'état stationnaire consécutif à une perturbation du système (qui peut être génétique ou un stress environnemental). Ainsi, même si d'un point de vue dynamique les modèles causaux sont trop restrictifs, ils ont toute légitimité pour être utilisés pour étudier des données portant sur les comparaisons d'états stationnaires.

**Contribution : application au diagnostic de modèles** Cette règle causale, dans son cadre restreint d'application, nous a permis de développer un formalisme pour analyser un système biologique modélisé sous la forme d'un graphe d'influence. Dans le chapitre 5, nous discuterons ainsi des différentes questions pertinentes sur un plan biologique qu'il est possible d'aborder sur des réseaux de grande taille à l'aide d'approches de résolution de contraintes très efficaces malgré la complexité (NP-complète) des questions abordées (Complex Us, 2006).

- *Diagnostiquer* un réseau, c'est-à-dire décider si l'ensemble des connaissances intégrées dans un modèle est suffisant pour expliquer un comportement donné.
- *Localiser* des zones contradictoires s'il en existe (CIBB, 2008).
- *Prédire* le comportement du système suite à une perturbation, y compris en présence de bruit dans les données.
- *Mesurer la robustesse* d'un ensemble de données par rapport aux connaissances.
- *Identifier le signe de régulations* dans un réseau pour lequel les interactions sont connues mais les modes de régulations inconnus (BMC bioinfo, 2008).
- *Estimer le nombre de jeux expérimentaux* nécessaires pour faire une analyse.

Ainsi, une “simple” considération sur les déplacements d’états stationnaires permet de répondre à des questions relatives à la construction de réseaux biologiques. Cependant, on constatera que la démarche pour répondre à ces questions a surtout consisté à aider le biologiste à formaliser ses questions sous forme de requêtes résolubles avec des algorithmes adaptés. Le cadre d’application des méthodes présentées au chapitre 5 s’avère donc bien plus large que la question de la reconstruction de réseau, et nous discutons dans la section 5.4 des différentes questions biologiques auxquelles il est envisageable d’apporter une réponse dans le futur, en gardant à l’esprit la notion de systèmes dynamiques (voir aussi (Biofutur, 2007)).

## 1.5 Plan du document

Le but de ce document est donc de détailler quels peuvent-être les apports des méthodes de discrétisations de systèmes dynamiques dans différents domaines. Nous allons maintenant pouvoir détailler ces approches, en commençant par des questions de théorie ergodique (chapitre 2) sur les automorphismes et les additions sur un tore. Leurs applications en informatique théorique (théorie des nombres et géométrie discrète), en relation avec l’étude de propriétés topologiques de fractal seront discutées dans le chapitre 3. Dans le chapitre 4, le point de vue sera complètement modifié puisqu’on discutera de discrétisation des déplacements d’états stationnaires dans des systèmes différentiels plutôt que de codages de trajectoires, en considérant la modélisation de systèmes biologiques. Enfin, dans le chapitre 5, nous adopterons un point de vue informatique d’analyse de systèmes et données de grande échelle, en insistant sur les questions biologiques qui peuvent être formalisées en considérant des discrétisations de déplacements d’états stationnaires, avec un objectif de résolution automatique.

---

## Chapitre 2

# Discrétiser de (simples ?) additions ou multiplications

Dans ce chapitre, nous allons détailler les problématiques encore actuelles posées par le codage symbolique des applications les plus *simples* possibles, c'est-à-dire les additions et les multiplications sur des tores.

La section 2.1 présente une synthèse des principaux résultats de dynamique sur ces applications sur les tores unidimensionnels.

Dans la section 2.2, nous rappellerons les différentes directions étudiées pour passer du cas unidimensionnel au cas multidimensionnel, et les difficultés qui sont apparues pour obtenir une bonne dynamique symbolique. Nous y montrerons que l'obtention d'une bonne dynamique symbolique se ramène finalement à la vérification d'une propriété de pavage.

Enfin, dans la section 2.3, nous proposerons une méthode qui englobe tous les cas traités dans la littérature et répond à la question de la recherche des partitions de Markov de manière décidable : il va s'agir de coder par un graphe l'ensemble des points de recouvrement entre des tuiles, pour vérifier explicitement la condition de pavage. Le résultat principal ici sera le Théorème 2.3.4.

Les résultats présentés dans ce chapitre seront exploités dans le chapitre suivant pour des applications dans différents domaines des mathématiques et de l'informatique théorique.

### 2.1 Les enjeux de la dynamique symbolique

#### 2.1.1 Systèmes dynamiques symboliques

Grossièrement, un *système dynamique* est un système qui évolue au cours du temps de façon à la fois *causale* (son avenir ne dépend que du passé ou du présent) et *déterministe* (une condition à l'instant présent détermine un unique futur possible). L'évolution du système dynamique peut alors être modélisée

- soit par une évolution continue dans le temps (une équation différentielle). Mathématiquement, le système  $T : X \times M \rightarrow X$  est déterminé par une loi  $T$ , définie sur un espace des états  $X$  et un espace de paramètres temporels  $M$ . Cette définition est la plus naturellement d'un point de vue physique ou biologique, puisque la notion de temps  $y$  est continue. Nous utiliserons ces modèles dans les chapitres 4 et 5 pour les applications des systèmes dynamiques en biologie moléculaire.
- soit par une évolution discontinue dans le temps, qui correspond à modéliser l'état du système après un pas de temps donné. Le système est alors donné par une application  $T : X \rightarrow X$ . Ces modèles, moins naturels, sont toutefois d'une grande utilité pour mettre en évidence des résultats importants qui se généralisent ensuite aux systèmes différentiels. Dans les systèmes discontinus, l'application  $T$  peut être un homéomorphisme sur un espace topologique (système dynamique topologique), un difféomorphisme sur une variété différentielle (système dynamique différentiel) ou une application qui préserve la mesure sur un espace mesurable (système dynamique mesuré).

**Systèmes dynamiques mesurés** La dynamique symbolique s'intéresse à des systèmes dynamiques mesurés, tout en s'inspirant de la dynamique topologique. On considère un compact  $X$ , qui est muni d'une topologie et d'une mesure. On suppose que l'application  $T$  est continue sauf sur un ensemble de mesure nulle, qu'elle est surjective sur  $X$  sauf sur un ensemble de mesure nulle, et qu'elle préserve la mesure. Le système  $T : X \rightarrow X$  est donc un système dynamique mesuré. Par contre, on ne pourra pas parler de système dynamique topologique puisque l'inverse de  $T$  soit n'existera pas, soit sera défini à un ensemble de mesure nulle près.

Le système dynamique est *minimal* lorsque l'orbite de tout point  $x$  permet de l'engendrer  $X = \{T^n(x), n \geq 0\}$ .

On considère une *partition mesurée* de  $X$  sous la forme  $X = X_1 \cup X_2 \cup \dots \cup X_n$ , où les  $X_i$  sont des fermés et les intersections  $X_i \cap X_j$  sont de mesure nulle pour la mesure sur  $X$  invariante par  $T$ .

**Système symbolique associé à un système dynamique** Le *codage d'un point*  $x$  pour l'action de  $T$  est le mot infini  $r_T(x) = u_0 \dots u_n \dots$  sur l'alphabet  $\mathcal{A} = \{1, \dots, n\}$  tel que  $T^n(x) \in X_{u_n}$  pour tout entier  $n$  [LM95].

Le système symbolique associé à  $X$  est la fermeture dans  $\mathcal{A}^{\mathbb{N}}$  de l'ensemble des codages des éléments de  $X$ , il s'agit donc de l'ensemble  $X_T = \overline{\{r_T(x), x \in X\}}$ .

Lorsque l'application  $T$  est bijective, on peut considérer l'ensemble  $\widehat{X}_T$  des orbites bi-infinies, c'est-à-dire les codages  $\dots u_{-k} \dots u_0 u_1 \dots u_n \dots \in \mathcal{A}^{\mathbb{Z}}$  tels que  $T^n(x) \in X_{u_n}$  pour tout entier  $n$  dans  $\mathbb{Z}$ .

Le *langage* du système symbolique (qu'il soit infini à droite ou bi-infini) est l'ensemble des mots finis qui apparaissent dans au moins un codage d'un point du système. La définition du système symbolique  $X_T$  (et le fait qu'on considère l'adhérence des codages des points) implique que  $X_T$  contient toutes les mots infinis qui ont le même langage que le codage d'un point.

Dans le cas où le système  $(X, T)$  est minimal, toutes les orbites de points ont le même langage et le système symbolique contient exactement l'ensemble des mots infinis qui admettent ce langage.

**Décalage** Une application naturelle agit sur l'ensemble  $X_T$ , il s'agit du *décalage*  $S$  qui supprime la première lettre du mot lorsqu'on considère des mots infinis à droite.

$$S : u_0 \dots u_n \dots \in X_T \mapsto u_1 \dots u_n \dots \in X_T.$$

Ce décalage s'étend aux mots bi-infinis en déplaçant simplement l'indice du mot :

$$\tilde{S} : \dots u_{-2}u_{-1}[\cdot]u_0 \dots u_n \dots \in \widetilde{X_T} \mapsto \dots u_{-2}u_{-1}u_0[\cdot]u_1 \dots u_n \dots \in \widetilde{X_T}.$$

Par construction, on a la relation  $r_T(Tx) = S(r_T(x))$ , ce qui signifie que considérer les orbites de  $T$  et celles de  $S$  revient globalement au même. On dit que le décalage  $S$  sur le système  $X_T$  commute avec l'application  $T$  sur le système  $X$ .

L'intérêt de  $\tilde{S}$  par rapport à  $S$  est que le premier est inversible tandis que le second ne l'est pas. Quand cela est possible, on privilégie donc  $\tilde{S}$  à  $S$ ; nous discuterons de la différence entre ces deux applications (même dans le cas inversible) sur un exemple à la section 2.2.6.

Lorsque  $T$  n'est pas inversible (même en mesure), on ne peut pas utiliser  $\tilde{S}$ , mais on recherche souvent une extension de  $S$  inversible (nous en discuterons dans la section 3.2). Pour une facilité d'écriture, on assimilera dans l'avenir les deux notations  $\tilde{S}$  et  $S$ , en se référant au contexte pour savoir si on considère des mots infinis ou bi-infinis.

**Partition génératrice** La partition est dite *génératrice* lorsque l'application de codage est un bon codage. Autrement dit, l'application  $x \in X \mapsto r_T(x) \in X_T$  doit être injective en mesure sur l'ensemble  $X$ .

**Dynamique symbolique** La dynamique symbolique consiste à trouver une partition génératrice pour un système. Avec cette opération, la dynamique du système, c'est-à-dire l'étude des trajectoires, est ramenée à l'étude de la combinatoire de certaines suites symboliques.

L'étape suivante de la dynamique symbolique (et son intérêt) consiste à caractériser l'ensemble des mots infinis qui appartiennent au système symbolique  $X_T$ . Les caractérisations usuelles portent sur le langage de ces systèmes, c'est-à-dire l'ensemble des mots finis qui apparaissent dans au moins un codage d'une orbite. Lorsqu'on arrive à caractériser de manière simple le système symbolique, on peut procéder à toutes sortes d'études (recherches de sous-ensembles invariants, propriétés métriques, études des orbites...) qui seront autant d'informations sur le système  $(X, T)$  initialement considéré.



### 2.1.2 Deux situations orthogonales où la dynamique symbolique a montré son utilité

La dynamique symbolique a été introduite par Hadamard et Morse au début du vingtième siècle pour étudier les propriétés de certaines courbes de surfaces. Étant donné un système dynamique, on recherche une partition du système qui permet de remplacer la dynamique par un décalage sur un ensemble de mots infinis ou bi-infinis sur un alphabet fixé. Une bonne partition doit être telle que l'ensemble des mots infinis ainsi obtenus (appelés codage des orbites) a une structure simple à manipuler. Deux classes d'exemples se sont distingués dans l'histoire de la dynamique symbolique : les systèmes markoviens et les systèmes auto-induits.

**Systèmes markoviens (ou de type fini) et systèmes sofiques** Les premiers exemples intéressants sont ceux pour lesquels la dynamique peut-être codée par un ensemble fini de mots interdits. Un système dynamique est dit *markovien* lorsque son système symbolique pour une partition donnée est *sous-décalage de type fini* : son langage est constitué de tous les mots infinis de  $\mathcal{A}^{\mathbb{N}}$  dans lesquels n'apparaissent jamais un nombre fini de motifs finis.

Si un système admet une partition génératrice qui est codée par un sous-décalage de type fini, on dit que la partition est une *partition de Markov* pour le système. Par construction, ces systèmes sont d'entropie non nulle. On sait en particulier que les automorphismes hyperboliques des tores et les difféomorphismes pseudo-Anosov des surfaces [Bow75] admettent tous une partition de Markov.

Un cas un peu moins contraint que les sous-décalage de type fini correspond aux systèmes sofiques. Il s'agit d'ensembles de mots infinis dont le langage est donné par un langage régulier, c'est-à-dire les étiquettes de chemins d'un graphe fini. Les sous-décalage de type fini sont bien entendu sofiques, mais la réciproque n'est pas vraie : les systèmes symboliques sofiques sont les facteurs topologiques des sous-décalages de type fini [LM95].

**Les systèmes auto-induits et les systèmes substitutifs** Une autre classe d'exemples concerne les dynamiques auto-induites, c'est-à-dire celles qui sont conjuguées à leur application de premier retour sur un sous-ensemble bien choisi. Ces systèmes sont de faible complexité, en particulier, d'entropie nulle, et on peut donc les considérer comme orthogonaux aux systèmes markoviens. Pour les décrire de manière symbolique, on utilise des codages provenant de procédés de répétitions, appelés *substitutions*.

Plus précisément, une substitution sur un alphabet fini  $\mathcal{A}$  est un morphisme de  $\mathcal{A}$  vers  $\mathcal{A}^*$ . L'exemple type est la substitution  $\sigma$  définie sur l'alphabet  $\{1, 2\}$  par  $\sigma(1) = 12$ ,  $\sigma(2) = 1$ .

L'intérêt des substitutions est qu'on peut les itérer : l'image de 1 est 12, qui admet 121 pour image, puis 12112, 12112121, 1211212112112, ... On remarquera que ces mots ont des débuts en commun de plus en plus long. La limite de ces mots est un mot infini  $u$  qui est fixé par la substitution : il vérifie  $\sigma(u) = u$ , on parlera de *point fixe* de la substitution.

Toutes les substitutions n'admettent pas un point fixe (il suffit de considérer la substitution  $1 \rightarrow 21, 2 \rightarrow 12$ ). Par contre, on sait que toute substitution admet au moins un point périodique : il existe  $n$  et un mot  $u$  infini tel que  $\sigma^n(u) = u$ .

La question naturelle qui se pose alors est de savoir plusieurs points périodiques peuvent avoir des langages différents. Pour éviter cela, on introduit la condition de primitivité : il s'agit de s'assurer que chaque lettre de  $\mathcal{A}$  apparaît au bout d'un moment dans l'image de toutes les autres lettres. Ceci implique qu'un point périodique, dont le langage contient tous les mots de la forme  $\sigma^{mk}(u_0)$ , contiendra des bouts de longueur croissante  $\sigma^{mk}(u_1)$  de tous les autres points périodiques ; les langages seront donc globalement les mêmes.

Plus formellement, on dit que la substitution est *primitive* lorsqu'elle sa *matrice linéarisée*  $\mathbf{M}_\sigma$  est *primitive* : il existe  $m$  tel que  $\mathbf{M}_\sigma^m$  ne contient que des coefficients strictement positifs. Par exemple, la matrice de  $\sigma$  est  $\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$  (la  $i$ -ème colonne indique la composition en lettres de  $\sigma(i)$ ). Cette substitution est primitive puisque  $\mathbf{M}_\sigma^2$  n'a aucun coefficient nul.

Sous la condition de primitivité, on sait que tous les points périodiques d'une substitution ont même langage [Que87]. On construit un système dynamique symbolique  $X_\sigma$  en considérant tous les mots infinis qui admettent ce langage. L'action naturelle sur  $X_\sigma$  est donnée par le décalage  $S$ , qui préserve bien entendu le langage d'un mot infini. Il s'agit du *système dynamique symbolique engendré par la substitution*. En utilisant l'hypothèse de primitivité, on montre que ce système dynamique est minimal : pour n'importe quel mot infini  $u \in X_\sigma$ , on sait que  $u$  permet d'engendrer entièrement l'ensemble  $X_\sigma$  par décalage (son orbite par  $S$  est dense) :

$$X_\sigma = \overline{\{S^n(u), n \in \mathbb{N}\}}.$$

De plus, l'ensemble  $\mathcal{A}^\mathbb{N}$  dans lequel vit  $X_\sigma$  est muni d'une topologie naturelle, et est donc un espace mesurable. Toujours sous l'hypothèse de primitivité, on montre qu'il existe une unique mesure invariante par  $S$  pour cette topologie, et le système dynamique est donc uniquement ergodique [Que87].

### 2.1.3 Les questions classiques en dynamique symbolique

Lors de l'étude d'une dynamique symbolique, le schéma est assez classique.

- On propose d'abord une partition pour le système dynamique, généralement obtenue en considérant les sous-ensembles de  $X$  sur lesquels  $T$  est continues (*ensembles de continuité*).
- On caractérise ensuite les codages du système dynamique pour cette partition.
- On montre que le codage est un bon codage (autrement dit, la partition est *génératrice*) ; il s'agit de montrer que, sauf sur un ensemble petit, deux points distincts ont des codages distincts.
- Ensuite, on peut travailler sur le système symbolique ainsi obtenu : est-il markovien ? S'il est de faible complexité, est-il substitutif, ou engendré par une suite de substitutions ou une structure de graphes infinis qui généralise les odomètres (diagrammes de Bratelli, voir [DHS99]) ?

- Cette étude préliminaire permet de se concentrer sur les propriétés métriques du système de départ : existence de mesures invariantes, ergodicité, spectre...
- Enfin, sur un plan topologique on cherche souvent à caractériser les orbites finies, périodiques et ultimement périodiques, qui correspondent à autant de points remarquables pour la dynamique de départ.

**Deux classes d'exemples canoniques** Les deux classes d'exemples canoniques pour ce schéma d'études sont les multiplications et les additions sur un tore. Comme mentionné dans le chapitre 1, ces exemples ont été étudiés en détail, et le schéma d'étude fonctionne remarquablement bien (nous allons rapidement illustrer cela par la suite).

Cependant, la question se pose de l'utilisation de ce schéma dans le cas multidimensionnel. Dans ce cadre, il est plus difficile de choisir une partition, et encore plus de prouver qu'elle est bien génératrice. Dans un cas bien précis (le cas Pisot), depuis les travaux de Rauzy [Rau82] et Thurston [Thu89] dans les années 80, on commence à avoir une bonne idée des candidats naturels pour la partition. Néanmoins, la question de la partition génératrice reste un point clé, qui va être le principal sujet de ce chapitre.

#### 2.1.4 Multiplications par des nombres simples

Les exemples canoniques de systèmes markoviens sont les multiplications sur un intervalle. Historiquement, il s'agit des dynamiques les plus étudiées. Nous allons rapidement rappeler les résultats fondamentaux à leur sujet.

**Multiplier par 2 modulo 1** Plutôt que de travailler sur l'intervalle  $[0, 1[$ , on préfère travailler sur le tore de dimension 1  $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ , qui peut-être vu comme l'intervalle fermé  $[0, 1]$  dans lequel 0 et 1 sont identifiés comme égaux. Ces deux ensembles sont en bijection, mais le tore est un compact alors que  $[0, 1[$  ne l'est pas.

Un des systèmes dynamiques les plus simples est l'application sur le tore de dimension 1 :  $T_2 : x \in \mathbb{T} \mapsto 2x \bmod 1 \in \mathbb{T}$ . Pour cette application, le schéma d'étude proposé plus haut s'applique très naturellement :

- Une partition naturelle est donnée par ses intervalles de continuité  $[0, 1/2[$ , qui sera noté par l'indice 0, et  $[1/2, 1[$ , qui sera désigné par l'indice 1.
- On montre facilement que le codage de l'orbite d'un réel  $x \in \mathbb{T}$  vis-à-vis de cette partition est le développement binaire du réel  $x$ . On en déduit que tous les mots infinis de  $\{0, 1\}^{\mathbb{N}}$  sont des codages, à part ceux qui se terminent par un infini de 1. Le système dynamique symbolique associé est donc tout l'ensemble  $\{0, 1\}^{\mathbb{N}}$ . Il est bien de type fini.
- Le développement binaire détermine entièrement un nombre réel de  $[0, 1[$ , donc de  $\mathbb{T}$ . Cette partition est donc génératrice.
- Les propriétés métriques du système ont été fortement exploitées dans divers travaux pour proposer différentes conjectures sur les développements des nombres. On montre par exemple que dans le développement binaire de presque tous les nombres réels, la fréquence de 0 et de 1 est égale à  $1/2$ ; en

fait, pour n'importe quelle base entière, la fréquence de chaque chiffre est uniforme dans le développement de presque tous les réels [Bor09, Lan94]. Notons cependant qu'un piège se trouve dans le terme *presque tous* : déterminer si la fréquence de 1 dans le développement binaire de  $\sqrt{2}$  est égale à  $1/2$  est une question toujours ouverte [Bor50, Wal06].

- Les informations sur les fréquences se traduisent naturellement en termes probabilistes : la probabilité de revenir dans un intervalle de longueur donnée est déterminée uniquement par la longueur de l'intervalle. On en déduit aussi que  $T_2$  est d'entropie  $\log 2$ .
- Enfin, les points périodiques pour ce système sont les rationnels dont le dénominateur est impair. Les points ultimement périodiques sont tous les nombres rationnels.
- En fait, le système dynamique est fondamentalement imprévisible puisque des réels proches (dont le développement binaires commence par les mêmes chiffres) peuvent avoir des futurs sans aucun rapport après un certain temps (puisque les futurs sont déterminés par les fins du développement binaire, qui peuvent être très variées). L'existence d'un ensemble dense de points périodiques et l'imprévisibilité sont des caractéristiques habituelles de *chaos*.

Des résultats identiques s'appliquent en fait pour les multiplications par un rationnel modulo 1. Par contre, si on considère un nombre irrationnel, les choses se compliquent un peu.

**Multiplier par le nombre d'or modulo 1** On considère maintenant l'application  $T_\phi : x \in \mathbb{T} \mapsto \phi x \pmod{1}$ , où  $\phi$  désigne le nombre d'or, caractérisé par son équation  $\phi^2 = \phi + 1$ . Alors le schéma décrit plus haut s'applique encore :

- Une partition naturelle existe, celle qui correspond aux intervalles de continuité de  $T_\phi$ , c'est-à-dire  $[0, 1/\phi[$  et  $[1/\phi, 1[$ .
- Les codages de l'application sont un peu plus complexes que pour le développement binaire mais toujours bien compris : un mot infini est le codage de l'orbite d'un point si et seulement si d'une part il ne contient pas deux 1 consécutifs et d'autre part il ne termine pas par une succession infinie de 10. De ce fait, le système dynamique symbolique associé est l'ensemble des mots infinis de  $\{0, 1\}^{\mathbb{N}}$  qui ne contiennent pas le motif 11. Il s'agit donc d'un codage de type fini.
- L'entropie de ce système est  $\log \phi$ . La probabilité de revenir dans un intervalle de longueur donnée est bien caractérisée, mais elle ne dépend plus seulement de la longueur de l'intervalle.
- Les réels dont l'orbite est ultimement périodique sont tous les éléments de  $\mathbb{Q}(\phi)$  [Sch80] ; tous les rationnels de  $[0, 1[$  ont une orbite purement périodique.
- Puisque l'entropie est non nulle, comme pour  $T_2$ , le système est imprévisible et on peut parler de système *chaotique*.

En particulier, intimement lié à cette application existe un système de numération en base  $\phi$  : tout réel  $x \in \mathbb{R}$  se décompose de manière unique sous la forme

$$x = \sum_{i \geq i_0} a_i \phi^{-i},$$

où la suite des chiffres  $a_i$  ne contient pas deux 1 consécutifs. Ces chiffres sont obtenus par algorithme glouton à partir de  $T_\phi$  ; il s'agit précisément du codage de l'orbite de  $\phi^{-i_0}x$  choisi tel que  $\phi^{-i_0}x \in [1/\phi, 1]$ .

Là encore, les propriétés métriques (sur les probabilités de retour dans un intervalle) ont des incidences sur les développements. On montre par exemple que pour presque tout réel, la fréquence du chiffre 0 dans le développement en base  $\phi$  est égale à  $1/\phi$  tandis que la fréquence de 1 est égale à  $1/\phi^2$ . On remarquera qu'ici, les fréquences existent encore, mais, contrairement aux bases entières, elles ne sont plus uniformes.

### 2.1.5 Multiplier par un réel $\beta$ ; beta-numérations

Le cas général consiste à étudier l'application  $T_\beta : x \in \mathbb{T} \mapsto \beta x \pmod{1}$ , où  $\beta$  est un réel quelconque supérieur à 1. Cette application est appelée *beta-transformation*. Le schéma d'étude s'applique alors partiellement.

**Partition et caractérisation des codages : beta-shift** Une partition naturelle est encore donnée par les intervalles de continuité de  $T_\beta$ , qui dépendent de la partie entière de  $\beta$ , notée  $b = \lfloor \beta \rfloor$ . Il s'agit donc des  $b$  intervalles  $[0, 1/\beta[$ ,  $[1/\beta, 2/\beta[$ ,  $\dots$ ,  $[(b-1)/\beta, 1]$ .

Les codages obtenus avec cette partition sont bien déterminés. Pour tout  $x \in \mathbb{T}$ , on note  $d_\beta(x)$  le codage de son orbite par  $T_\beta$ . Même si l'application  $T_\beta$  n'est pas définie au point 1 (pour faire de  $T_\beta$  une application bijective), il n'y a aucune restriction à étendre  $T_\beta$  au point 1 et à définir le codage de l'orbite de 1, que l'on note  $d_\beta(1)$ . De plus, si le codage de l'orbite de 1 est de la forme  $a_1 \dots a_n 0000000$ , on considère le nouveau mot infini  $d_\beta(1)^* = (a_1 \dots a_{n-1}(a_n - 1))^\infty$ . On sait alors depuis Parry [Par60] qu'un mot infini  $u_1 \dots u_n \dots$  sur l'alphabet  $\{0, \dots, b-1\}$  est le codage d'une orbite de  $T_\beta$  si et seulement si la condition d'admissibilité suivante est vérifiée pour l'ordre lexicographique [Bla89, Par60]

$$\forall p \geq 1, \quad u_p u_{p+1} u_{p+2} \dots <_{lex} d_\beta(1)^*.$$

L'ensemble des mots infinis qui vérifie cette condition étendue au cas fermé  $\leq_{lex}$  est appelé *beta-shift*. Il s'agit du système symbolique qui code  $T_\beta$  pour la partition précisée plus haut.

Notons que dans le cas du nombre d'or, on a  $d_\phi(1) = 110000000\dots$ , et donc  $d_\phi^* = (10)^\infty$ . On retrouve ici la condition selon laquelle le motif 11 est interdit dans le codage.

**Les cas markoviens et sofiqes** À partir de cette caractérisation, on en déduit que la partition produit un codage de type fini si et seulement si le mot  $d_\beta(1)^*$  est périodique [BM86]. Le nombre  $\beta$  est alors appelé un *nombre de Parry simple*.

On peut aller plus loin. Le codage symbolique de  $T_\beta$  est un système sofique si et seulement si  $d_\beta(1)^*$  est ultimement périodique ; on parle de *nombre de Parry*. Le graphe qui décrit le langage du système symbolique est alors entièrement déterminé par  $d_\beta(1)^*$ , comme illustré Fig. 2.1 [BM86].

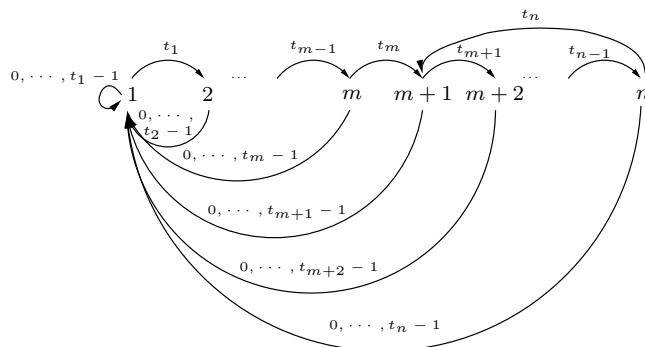


FIG. 2.1 – Graphe qui décrit les codages des orbites de la multiplication par  $\beta$  modulo 1, notée  $T_\beta$ , dans le cas où  $\beta$  est un nombre de Parry. Pour construire ce graphe, on considère le mot infini obtenu à partir du codage de 1  $d_\beta^*(1) = t_1 \cdots t_m (t_{m+1} \cdots t_n)^\infty$ .

Les nombres de Parry font l'objet de nombreuses études et restent mystérieux [Boy96, Sol94]. Une condition suffisante existe pour montrer qu'un nombre est de Parry : si  $\beta$  est un *nombre de Pisot*, c'est-à-dire si le module de tous ses conjugués de Galois appartient à  $]0, 1[$ , alors  $\beta$  est un nombre de Parry [Ber77].

**Partition génératrice et beta-numération** Si les beta-transformations sont aussi étudiées, c'est non seulement parce qu'il s'agit d'exemples canoniques de systèmes dynamiques, mais aussi parce qu'il s'agit de transformations naturelles d'un point de vue numérique. En effet, on montre facilement que si le codage de l'orbite d'un point  $x$  est noté  $d_\beta(x) = u_1 \dots u_n \dots$ , alors  $x$  s'écrit en base  $\beta$  à partir de son développement :

$$x = \frac{u_1}{\beta} + \dots + \frac{u_n}{\beta^n} + \dots,$$

et cette écriture est unique dès que la condition d'admissibilité est vérifiée. On parle de *développement de Renyi* de  $x$ . L'unicité de cette écriture signifie que la partition est génératrice pour  $T_\beta$  : le codage d'un point est unique.

**Propriétés métriques des beta-transformations** À partir du moment où une partition markovienne ou sofique a été considérée, il devient possible d'étudier les propriétés métriques de la beta-transformation  $T_\beta$ , et ces propriétés ont fait l'objet d'une abondante littérature, dans le cas Pisot en particulier [Bro96, BKS91].

**Points remarquables pour la dynamique** Naturellement, on cherche à caractériser les réels qui ont un codage remarquable pour une application  $T_\beta$ , au moins dans le cas des nombres de Parry. On montre ainsi que les orbites ultimement périodiques sont données par les éléments de  $\mathbb{Q}(\beta)$  [BM86, Sch80]. Les orbites finies

(qui se terminent sur 0) et celles qui sont purement périodiques font aussi l'objet de nombreuses études, que nous détaillerons dans la section 3.2. Nous proposerons des extensions, en particulier quand  $\beta$  n'est pas un nombre unitaire.

**Points remarquables pour la numération** Du point de vue des systèmes de numération, certains réels sont d'un grand intérêt : il s'agit des réels qui n'admettent pas de partie fractionnaire dans leur développement en base  $\beta$ . Pour les définir proprement, on doit comme mentionné dans le cas du nombre d'or étendre le développement en base  $\beta$  à l'ensemble des réels positifs. Si  $x > 1$ , il existe un unique  $n_0$  tel que  $x\beta^{-n_0} \in [1/\beta, 1[$ . En considérant le développement de ce nombre, on obtient un développement pour  $x$  sous la forme d'une partie entière et d'une partie fractionnaire qui dépendent de  $\beta$  :

$$x = \underbrace{u_{-n_0}\beta^{n_0} + u_{-n_0+1}\beta^{n_0-1} + \dots + u_0}_{\text{partie entière}} + \underbrace{\frac{u_1}{\beta} + \dots + \frac{u_n}{\beta^n} + \dots}_{\text{partie fractionnaire}}$$

Les *nombre entiers pour la beta-numération* sont les réels qui n'admettent pas de partie fractionnaire.

$$\text{Int}(\beta) = \left\{ z = \sum_{0 \leq i \leq K} u_i \beta^i \in \mathbb{Z}[\beta]; u_K \dots u_0 = d_\beta(\beta^{-K-1}z) \right\}. \quad (2.1)$$

Ces nombres forment un ensemble discret de  $\mathbb{R}^+$ . Dans le cas Parry, ces nombres présentent une régularité puisque deux éléments consécutifs ne diffèrent que par un nombre fini de valeurs, en fait les nombres  $T_\beta^{a-1}(1)$ ,  $a \in \{1, \dots, n\}$  [Aki07, Thu89]. Cependant, les régularités de cet ensemble ne sont pas évidentes. On montre que les intervalles sont ordonnés suivant l'ordre des lettres dans le point fixe d'une substitution naturellement associée à  $\beta$ , qui est définie à partir de  $d_\beta(1)^*$ , aussi appelée beta-substitution [Thu89] (voir aussi (Integers, 2005)).

### 2.1.6 Additions et systèmes auto-induits

Un autre classe d'applications très étudiées est constituée des additions sur le tore. Etant donné  $\alpha < 1$ , il s'agit de comprendre la dynamique de l'application

$$R_\alpha : x \in \mathbb{T} \mapsto x + \alpha \pmod{1}.$$

**Rigidité** Ces applications sont fondamentalement différentes des multiplications  $T_\beta$ . D'abord, elles sont bijectives, alors que les multiplications ne l'étaient pas. Sur-tout, dans le cas des multiplications, il existait une infinité d'orbites ultimement périodiques (donc finies). De plus, le système dynamique était fondamentalement imprévisible (puisque deux points proches peuvent avoir des orbites sans aucun rapport après un certain temps). Pour les additions  $R_\alpha$ , au moins dans le cas où  $\alpha$  est irrationnel, la situation est inverse : il n'y a aucune orbite périodique ; tous les points

sont d'orbite dense dans  $\mathbb{T}$  (autrement dit, le système dynamique est minimal). De plus, les transformations  $R_\alpha$  préservent la métrique circulaire. La distance sur le cercle entre les éléments des orbites de deux points est donc constante. Il y a donc prévisibilité à long terme.

**Codage symbolique et caractérisation combinatoire** Pour ces applications, une partition naturelle est encore donnée par les intervalles de continuité  $[0, 1 - \alpha[$  et  $[1 - \alpha, 1[$ . Le schéma d'étude se passe alors relativement bien, comme l'ont montré Hedlund et Morse dans des articles qui ont fondé la dynamique symbolique [MH38, MH40]. Les mots infinis obtenus par codage des  $R_\alpha$  sont très spécifiques d'un point de vue combinatoire : un mot  $u_0 u_1 \dots u_k \dots$  est le codage de la trajectoire de  $x$  pour une addition  $R_\alpha$  si et seulement si ce mot contient exactement  $n + 1$  motifs de longueur  $n$  pour tout  $n$ .

**Une meilleure caractérisation des codages symboliques : application de premier retour** Cependant, la caractérisation des codages via leur fonction de complexité n'est pas suffisante pour une addition donnée : comment distinguer deux mots infinis qui sont tous deux de complexité  $n + 1$  ? Pour cela, Morse et Hedlund ont utilisé des applications de premier retour et de substitution, en s'appuyant sur les substitutions :  $\sigma_0(0) = 01$ ,  $\sigma_0(1) = 1$  et  $\sigma_1(0) = 0$ ,  $\sigma_1(1) = 10$ .

Les *application de premier retour* sont des concepts classiques en systèmes dynamiques : on considère un système  $(X, R)$  et un sous-ensemble  $X_1$  de  $X$ . L'*application de premier retour*  $R^*$  de  $R$  sur  $X_1$  est définie pour tout  $x$  de  $X_1$  par  $R^*(x) = R^{n(x)}(x)$ , où  $n(x) \geq 1$  est le plus petit entier tel que  $R^{n(x)}(x) \in X_1$ .

Supposons que  $\alpha > 1/2$ . Pour les additions  $R_\alpha$ , on montre que l'application de premier retour de  $R_\alpha$  sur l'intervalle  $[0, \alpha]$  est conjuguée par une homothétie  $h$  à une nouvelle addition  $R_{\alpha_1}$ , où  $\alpha_1 = \frac{1}{\alpha} - [\frac{1}{\alpha}]$  est obtenu par l'algorithme des fractions continues :  $R_\alpha^* = h R_{\alpha_1} h^{-1}$ . On en déduit que le codage  $r_\alpha(1 - \alpha)$  de la trajectoire de  $1 - \alpha$  par  $R_\alpha$  est lié au codage  $r_{\alpha_1}(1 - \alpha_1)$  de  $1 - \alpha_1$  par  $R_{\alpha_1}$  par  $\sigma_0$  :

$$r_\alpha(1 - \alpha) = \sigma_0^{[1/\alpha]-1} r_{\alpha_1}(1 - \alpha_1).$$

En itérant ce processus, on obtient une écriture de  $r_\alpha(1 - \alpha)$  comme limite des mots  $\sigma_0^{a_1-1} \sigma_1^{a_2} \sigma_0^{a_3}(0)$  où  $[a_1, a_2, \dots]$  est le développement en fraction continue de  $\alpha$ .

Ceci permet de caractériser entièrement les codages des additions  $R_\alpha$  : lorsque  $\alpha > 1/2$ , une suite  $u_0 \dots u_n \dots$  est le codage d'une addition  $R_\alpha$  pour la partition donnée par ses intervalles de discontinuité si et seulement si son langage est engendré par les mots finis  $\sigma_0^{a_1-1} \sigma_1^{a_2} \sigma_0^{a_3} \dots \sigma_0 a_{2n+1}(0)$ , où  $\alpha = [0, a_1, a_2, \dots]$  désigne le développement en fractions continues de  $\alpha$ . Le cas où  $\alpha < 1/2$  se traite de manière similaire, ainsi que le cas où les intervalles de continuité dans la partition sont considérés avec des parties entières supérieures [MH38, MH40, Fog02].

Avec cette caractérisation, on montre que le codage de l'orbite d'un point pour  $R_\alpha$  est unique sauf pour un nombre dénombrable de points. La partition considérée au départ est donc génératrice en mesure. Les propriétés que l'on peut déduire de cette approche ne sont pas nouvelles : les additions irrationnelles sur un tore sont



minimales, d'entropie nulle, sans orbite périodique ; mais ceci se montre directement, sans besoin de codages...

L'intérêt de cette approche réside dans la réciproque : tout système dynamique qui est codé par des suites de complexité minimale (c'est-à-dire  $n+1$ ) sera finalement isomorphe à une addition sur un tore, et du fait de cette caractérisation, la dynamique symbolique sera engendrée par la composition de substitutions. Ainsi, une importante littérature a été consacrée à l'identification de systèmes de complexité  $n+1$ , et à exploiter les relations entre complexité et dynamique. Ces propriétés ont été utilisées par exemple dans le cadre de l'optimisation de l'allocation de ressources dans les réseaux de télécom [GHVdL07].

**Le cas substitutif; caractérisations des cas auto-induits** Un cas en particulier a fait l'objet d'investigations poussées : il s'agit du cas où l'addition est auto-induite, c'est-à-dire que son application de premier retour sur son plus grand intervalle de continuité est conjugué à l'addition considérée au départ. La question est alors : quelle sont les additions dont le codage symbolique est substitutif ? Avec les approches précédentes, on arrive à un résultat très simple à exprimer : l'application  $R_\alpha$  est auto-induite si et seulement si la pente  $\alpha$  admet un développement en fractions continues purement périodique. Selon le théorème de Galois, il s'agit d'un nombre quadratique irrationnel avec un conjugué de Galois en dehors de l'intervalle  $]0, 1[$  [All98, CMPS93]. Dans ce cas là, l'orbite du point  $\alpha$  est un point fixe de substitution, en particulier il est très simple de l'engendrer. Le cas le plus connu est celui de la substitution de Fibonacci  $1 \rightarrow 12, 2 \rightarrow 1$ , dont le point fixe code l'orbite du nombre d'or sous l'addition ayant pour pente le nombre d'or [All98, CMPS93]. La description du codage de  $R_\alpha(x)$ , pour un  $x$  quelconque et  $\alpha$  quadratique, a été donnée dans [Yas99, BEIR07].

Les applications de ces approches de codage ont principalement porté sur l'engendrement des droites discrètes. En effet, imaginons que l'on dessine un quadrillage du plan par des droites verticales et horizontales en suivant le réseau  $\mathbb{Z}^2$ . La question qui se pose est de discrétiser une droite  $Y = \gamma X + \rho$ , c'est-à-dire, partant de  $\rho$ , de suivre la direction de la droite  $Y = \gamma X + \rho$  en restant sur le quadrillage. Lorsqu'on traverse une direction verticale, on retient un  $a$ , lorsqu'on traverse une direction horizontale, on retient la lettre  $b$  [Fog02, Chapitre 6]. Un simple calcul montre qu'une suite de coupure pour une droite de pente  $\gamma$  est le codage d'un point donné pour l'addition d'angle  $\alpha = \gamma/(\gamma + 1)$ . ainsi, en identifiant exactement le codage des orbites de  $R_\alpha$ , on obtient des procédés efficaces pour engendrer les débuts des approximations discrètes de droites par itérations successives à partir du calcul du développement en fractions continues de  $\alpha$  [Fog02, Chapitre 6].

## 2.2 Multidimensionnel : quel candidat pour une partition ?

Les cas canoniques que sont les additions et les multiplications sur l'intervalle sont donc très bien compris d'un point de vue symbolique et abondamment ex-

plottés. Ces travaux ont ouvert la voie à des recherches très actuelles dans le cas multidimensionnel.

### 2.2.1 Multiplier par une matrice

#### Partitions de Markov pour les automorphismes hyperboliques du tore

La généralisation de la multiplication par  $\beta$  sur l'intervalle est la multiplication par une matrice  $\mathbf{M}$  sur un tore  $\mathbb{T}^n$ . On parle d'endomorphismes du tore, et d'automorphismes lorsque le déterminant de  $\mathbf{M}$  est égal à 1. Lorsque  $\mathbf{M}$  n'a pas de valeurs propres de module 1, on dit que l'automorphisme est hyperbolique.

On dispose sur ces applications d'une information importante due à Bowen : tous les automorphismes hyperboliques des tores sont des systèmes markoviens [Bow75] ; autrement dit, pour chaque multiplication sur le tore  $\mathbb{T}^n$  par une matrice unitaire et sans valeur propre de module 1, il existe une partition de  $\mathbb{T}^n$  qui permet de coder l'action de la matrice par des mots interdits.

**Construction effective ?** Précisons ici ce que signifie ici une partition de Markov. On sait qu'un codage de type fini est associé à un ensemble de mots interdits. En raffinant la partition (et en augmentant donc le nombre  $N$  de pièces dans cette partition), on montre qu'un codage markovien est en fait caractérisé par un ensemble  $\mathcal{M}$  de paires de lettres  $(i, j)$ , où,  $i, j \leq N$  qui sont autorisées dans le codage d'une trajectoire, tandis que les paires  $(i, j)$  qui n'appartiennent pas à  $\mathcal{M}$  ne doivent jamais apparaître dans un codage [LM95]. Notons bien que ici, les paires  $(i, j)$  sont ordonnées : la lettre  $i$  peut ou non être *suivie* de  $j$  dans le codage. On définit ainsi un graphe orienté sur les sommets  $\{1, \dots, N\}$  en posant une arête de  $i$  vers  $j$  si  $(i, j) \in \mathcal{M}$ .

Rechercher une partition de Markov pour un automorphisme  $\mathbf{M}$  consiste finalement à trouver un domaine fondamental  $\mathcal{X}$  du réseau  $\mathbb{Z}^n$  et une partition de ce domaine en  $N$  pièces  $\mathcal{X} = \cup_{i=1}^n \mathcal{X}_i$  tels que l'image par  $\mathbf{M}$  des points d'une pièce  $\mathcal{X}_i$  respectent les contraintes données par l'ensemble  $\mathcal{M}$ . Ceci revient à dire que  $\mathbf{M}\mathcal{X}_i \bmod \mathbb{Z}^n$  est exactement égal à l'union des pièces  $\mathcal{X}_j$ , pour les  $j$  tels que  $(i, j) \in \mathcal{M}$ . La condition pour que la partition soit markovienne est donc

$$\mathbf{M}\mathcal{X}_i = \bigcup_{i \rightarrow j} \mathcal{X}_j \quad \bmod \mathbb{Z}^n.$$

Dans le cas des automorphismes du tore de dimension 2, une partition de Markov s'obtient à l'aide de rectangles de manière explicite [Bow78]. Cependant, dès qu'on dépasse la dimension 3, on sait que la partie contractante du bord des éléments d'une partition de Markov ne peut pas être lisse [Bow78, Caw91]. Il est donc vain d'espérer construire une partition par des moyens simples.

**Le cas des automorphismes avec une seule valeur propre dilatante** Après les résultats de Bowen sur l'existence de partitions de Markov, un certain nombre de travaux ont cherché à construire explicitement ces partitions. La *dynamique arithmétique* a été développée par différents auteurs en suivant un article fondateur de

Vershik [Ver92], qui a étudié le cas du nombre d'or puis étendu ses travaux au cas quadratique Pisot [SV98]. Suivant la définition de [Sid03], le but de cette approche est de construire des codages explicites pour des automorphismes hyperboliques de tores ou de solénoïdes, de sorte que les propriétés géométriques de l'automorphisme sont clairement transposées dans le cadre dynamique. L'approche de Vershik [Ver92, SV98] et poursuivie dans [Sch00a, KV98, Sid01, Sid02, BK05, LS05] est de considérer des automorphismes du tore avec une seule valeur propre dilatante, et de décomposer les points du tore sous la forme de séries entières dont la base est donnée par un point homocline, c'est-à-dire un point qui se trouve à l'intersection des variétés stable et instable de la dynamique.

Avec ces constructions, on obtient un codage symbolique de l'automorphisme du tore par un système markovien. Cependant, ce codage n'est pas générateur : on dit qu'il s'agit d'une extension *finie-à-un* de la transformation, ce qui signifie que le nombre de points de l'espace de départ admettant le même codage est constant : on ne sait pas s'il est égal à un, ce qui est nécessaire pour obtenir une vraie partition de Markov. Plus précisément, on sait que le domaine  $\mathcal{X}$  qui est construit contient un nombre entier de copies d'un domaine fondamental du tore, mais on ne sait pas montrer qu'il s'agit réellement d'un domaine fondamental ; il faudrait pour cela montrer qu'il est de volume 1. La question de la genericité du codage n'est donc pas résolue dans ces travaux.

## 2.2.2 Des multiplications par une matrice vers des additions sur un tore

**Représentation géométrique** La question de la genericité du codage peut-être abordée avec un point de vue géométrique, en proposant explicitement une construction pour une partition de Markov. Les premiers travaux dans cette direction [IO93, KV98, Pra99] puis [Ber99, BK05] se sont appuyés sur un ensemble compact introduit par Rauzy [Rau82] pour étudier une addition sur un tore, et repris par Thurston [Thu89]. Il s'agit de remplacer les séries entières de [Sch00a] par des séries entières géométriques faisant intervenir les vecteurs propres dilatants à droite et à gauche de l'automorphisme du tore considéré [BBLT06].

**Sections des partitions de Markov dans le cas Pisot : une intuition géométrique** La dynamique arithmétique étudie une classe spécifique d'automorphismes hyperboliques : il s'agit des automorphismes du tore hyperbolique admettant une seule valeur propre dilatante. La valeur propre dominante est alors un nombre de Pisot et on montre que le polynôme caractéristique est irréductible (Trans. AMS, 2001). On appelle ces automorphismes *Pisot irréductibles*. Géométriquement, ces automorphismes se dilatent selon une unique direction, ce qui simplifie amplement la dynamique.

On en déduit des propriétés géométriques intéressantes sur le pavage donné par une partition de Markov du tore. Ces propriétés sont bien visibles dans le cas bi-dimensionnel (voir Fig. 2.2 pour la matrice  $\begin{pmatrix} 1 & 1 \\ 2 & 1 \end{pmatrix}$ ), mais elle sont vérifiées

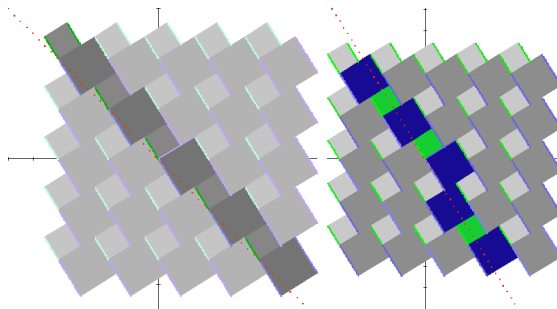


FIG. 2.2 – Une partition de Markov pour un automorphisme du tore  $\mathbb{T}^2$ . Les sections du domaine fondamental induisent un pavage périodique de la droite  $x + y = 0$  et un pavage stable par la contraction par  $-1/\phi$  de la direction dilatante de la matrice

pour tous les automorphismes Pisot irréductibles. On a un domaine de Markov, décomposé en pièces, qui est aussi un domaine fondamental de  $\mathbb{Z}^n$ . Ce domaine remplit donc tout l'espace  $\mathbb{R}^n$  de manière régulière. On découpe ce recouvrement de  $\mathbb{R}^n$  par différents sous-espaces.

- Lorsqu'on découpe le recouvrement périodique de  $\mathbb{R}^n$  selon l'espace contractant de la matrice, on obtient une partition de l'hyperplan contractant. Puisque  $\mathbf{M}$  stabilise à la fois l'hyperplan contractant et  $\mathbb{Z}^n$ , l'action de  $\mathbf{M}$  sur sa partition (une contraction par définition) permet de décomposer chaque domaine de l'hyperplan en domaines de la même forme.
- La trace du domaine fondamental sur l'espace dilatant de la matrice (qui est une droite) forme une partition de la droite stable par la multiplication par la valeur propre dominante de la matrice.
- Puisque la partition représente un domaine fondamental du réseau  $\mathbb{Z}^n$ , les sections du domaine de Markov selon n'importe quel hyperplan à coordonnées entières, (en particulier l'hyperplan orthogonal au vecteur  $(1, \dots, 1)$ ), forment un pavage périodique du plan.

**Facteur de translation** À partir de la figure 2.2, on comprend mieux le fait que le domaine fondamental du tore donné par la partition de Markov se sectionne à la fois en un ensemble autosimilaire et en un domaine fondamental du plan. Pour cela, on appelle  $\mathcal{X}$  l'intersection entre le domaine de Markov et l'espace contractant de l'automorphisme. Sur la figure 2.2, il s'agit d'un intervalle, dans le cas général, on obtiendra un compact dans l'espace contractant. Puisque nous avons considéré un domaine de Markov,  $\mathcal{X}$  est un domaine fondamental de l'espace contractant pour un réseau engendré par les projections des vecteurs canoniques sur le plan contractant.

Considérons maintenant cet ensemble  $\mathcal{X}$  dans le tore : cela revient à placer dans  $\mathbb{R}^n$  tous les ensembles  $\mathbf{z} + \mathcal{X}$ , avec  $\mathbf{z} \in \mathbb{Z}^n$ . L'application qui nous intéresse est l'application de premier retour sur  $\mathcal{X}$  (plongé dans le tore  $\mathbb{T}^n$  en suivant le flot le long de la direction dilatante. Nous l'appellerons  $\Phi$ . Remonté à  $\mathbb{R}^n$ , il s'agit donc

de partir d'un point de  $\mathcal{X}$ , de suivre la direction dilatante et de s'arrêter lorsqu'on croise une nouvelle copie de la forme  $\mathbf{z} + \mathcal{X}$  (que nous appellerons feuille).

Le point important, que l'on constatera sur la figure 2.2, est que le nombre de feuilles  $\mathbf{z} + \mathcal{X}$  sur lesquelles on peut arriver par  $\Phi$  est fini. Ce résultat est en fait général : dans un voisinage borné de  $\mathcal{X}$ , il n'y a qu'un nombre fini de feuilles  $\mathbf{z} + \mathcal{X}$ , et pour des raisons de densité, il suffit d'un nombre fini de feuilles  $\mathbf{z} + \mathcal{X}$  pour recouvrir  $\mathcal{X}$ , après projection dans le plan contractant.

Ceci signifie qu'on peut décomposer  $\mathcal{X}$  en fonction de l'application  $\Phi$  : on va rassembler dans un ensemble  $\mathcal{X}_{\mathbf{z}}$  les points  $\mathbf{x}$  dont l'image par  $\Phi$  arrive dans une même feuille  $\mathbf{z} + \mathcal{X}$ . Dans le tore, appliquer  $\Phi$  revient alors à déplacer  $\mathcal{X}_{\mathbf{z}}$  par la projection du vecteur  $\mathbf{z}$  sur le plan contractant. On définit ainsi un échange de morceaux dans  $\mathcal{X}$ , les vecteurs de translations étant de la forme  $\pi(\mathbf{z})$ , avec  $\mathbf{z} \in \mathbb{Z}^n$ , et  $\pi$  la projection sur l'espace contractant parallèlement à la direction dilatante. Ainsi, les vecteurs de translation sont fortement liés aux directions du réseau dont  $\mathcal{X}$  est un domaine fondamental.

De plus, les propriétés des partitions de Markov impliquent que l'application de premier retour le long de la feuille dilatante "respecte" l'action de l'automorphisme considéré, qui est décrite par l'action d'une substitution. Symboliquement, l'échange de morceaux devrait donc respecter une structure substitutive.

Tout ceci est bien entendu intuitif, mais ces intuitions ont incité à considérer géométriquement le lien entre les échanges de morceaux sur des tores, les additions sur des tores et les systèmes symboliques engendrés par une substitution.

### 2.2.3 Additions sur un tore

La généralisation naturelle de l'addition sur un intervalle est l'addition sur un tore de dimension supérieure à 1. On considère donc maintenant un tore de dimension  $n$ , noté  $\mathbb{T}^n = \mathbb{R}^n / \mathbb{Z}^n$ , et une addition sur ce tore par un vecteur de translation  $\mathbf{x}_0 : \mathbf{x} \in \mathbb{T}^n \mapsto \mathbf{x} + \mathbf{x}_0 \in \mathbb{T}^n$ .

**À la recherche de tous les systèmes engendrés par une substitution qui codent une addition** Pour mieux comprendre ces additions sur un tore, la communauté scientifique s'est inspirée des connaissances sur les additions de l'intervalle détaillées dans la section 2.1.6. Il a donc fallu mieux comprendre le lien entre les substitutions et ces additions dans différents contextes. Inspirées par le cas unidimensionnel, les questions principales ont porté sur la relation entre le codage des additions et les compositions de systèmes substitutifs.

**Étude spectrale** Avec le cas unidimensionnel, on a appris que certains systèmes substitutifs sur deux lettres, mais pas tous, sont le codage d'additions sur l'intervalle. Or, sur un plan ergodique, on sait avec le théorème de Von Neumann qu'un système dynamique est isomorphe en mesure à une addition sur un groupe compact si et seulement si son spectre est purement discret ; autrement dit, l'ensemble  $L^2(X_\sigma, S)$  est engendré par les fonctions propres de l'opérateur unitaire  $U : f \in L^2(X, S) \mapsto f \circ S \in L^2(X_\sigma, S)$  [Wal82]. Identifier les substitutions qui codent une addition sur un

tore revient donc à comprendre les propriétés spectrales des systèmes symboliques engendrés par une substitution. Cette question a occupé une large partie de la littérature du domaine dans les années 1970 et 1980 : j'ai écrit une synthèse sur ce sujet dans l'ouvrage collectif Pytheas-Fogg dans la collection Lectures Notes in Mathematics (Pytheas Fogg, LNM, 2002). En tout état de cause, il s'avère que même avec un alphabet à deux lettres, caractériser les systèmes substitutifs à spectre purement discret, et plus globalement déterminer le spectre d'un système dynamique substitutif est une tâche ardue.

**Le cas des substitutions Pisot sur deux lettres** Il ressort de l'analyse spectrale des années 1980 que de bons candidats pour les systèmes à spectre purement discret sont ceux qui sont engendrés par une substitution dont la matrice d'incidence a une valeur propre Pisot dont le degré algébrique est égal à la taille de l'alphabet (comme pour les automorphismes du tore, on parlera de substitution *Pisot irréductible*).

Pour cette classe de substitutions, on connaît explicitement le spectre [BK06]. On sait aussi maintenant que, si l'alphabet ne contient que deux lettres, le système engendré par une substitution Pisot unitaire est systématiquement le codage d'une addition sur un tore [BD02]. Le cas des substitutions sur deux lettres est donc bien maîtrisé.

**Espace de pavage associé à une substitution Pisot** Pour ce qui est du cas général, l'école américaine a publié ces dernières années un certain nombre de travaux concernant l'étude spectrale des systèmes substitutifs. Leur approche est la suivante : à partir d'une substitution (réductible ou irréductible) Pisot, on peut définir un flot de translation sur un espace de pavage engendré par n'importe quel mot infini qui est un point fixe (ou éventuellement périodique) pour la substitution [BBK06, BK05]. Les valeurs propres de ce flot de translation ne dépendent que de la matrice de la substitution [BBK06]. Une première question ouverte est de savoir si, dans le cas Pisot unitaire, ce flot de translation dont on connaît le spectre est effectivement à spectre purement discret.

Dans cette construction d'espaces de pavages, deux dynamiques coexistent : la première est celle du flot de translation sur l'espace de pavage (dont on connaît explicitement le spectre), la seconde est l'application de premier retour sur une section adéquate. Un autre résultat de Barge et Kwaplicz montre que symboliquement, une substitution code la deuxième dynamique [BBK06].

En travaillant sur la coexistence de ces deux dynamiques, il devient possible de montrer que le système symbolique engendré par la substitution  $\sigma(1) = 162111$ ,  $\sigma(2) = 435211$ ,  $\sigma(3) = 35211$ ,  $\sigma(4) = 435444$ ,  $\sigma(5) = 162544$ ,  $\sigma(6) = 62544$  n'est pas à spectre purement discret [BBK06]. Par contre, le flot de translation qui lui est associé est à spectre purement discret. Ainsi, les systèmes substitutifs Pisot unitaires *ne sont pas tous* à spectre purement discret.

On pourra cependant remarquer que si la valeur propre dominante de  $\sigma$  est bien un nombre de Pisot de degré 3, elle n'est pas irréductible puisqu'elle admet des valeurs propres de degré 2. La question reste donc entière de savoir si le système

associé à une substitution *Pisot irréductible unitaire* est à spectre discret. On sait simplement que dans ce cas irréductible, cette propriété est équivalente au spectre discret de la première dynamique par flot de translation.

Dans le contre-exemple de la substitution  $\sigma$ , même si le système substitutif n'est pas explicitement le codage d'une addition, Barge et Kwapicz ont quand obtenu un résultat dynamique intéressant, qui formalise et généralise des observations sur des substitutions provenant de beta-numérations [EI05]. En effet, ils ont montré que dès qu'un flot de translation est à spectre purement discret, le système symbolique associé peut ne pas être une addition sur un tore, mais il sera toujours le codage de l'application de premier retour d'une addition sur un tore. Autrement dit, il existe toujours un réseau lié à ce système et une addition qui permet de le décrire. Dans ce contexte, plusieurs questions ressortent :

- Vérifier que le flot de translation sur un espace de pavage engendré par une substitution Pisot unitaire est à spectre purement discret.
- Vérifier que le système dynamique symbolique associé à la substitution est à spectre purement discret, ou, si ça n'est pas le cas, est conjugué à l'application de premier retour d'une translation. Cela revient à trouver un bon domaine fondamental pour un réseau pour représenter le système.
- Étant donnée une translation sur un tore, trouver (s'il en existe) une substitution qui décrit les orbites de cette translation dans une partition adéquate.

Répondre à ces questions revient aussi à répondre à la question de la construction d'une partition de Markov. Les approches arithmétiques discutées à la section 2.2 ayant montré leurs limites, nous allons voir comment il est possible de répondre partiellement à ces questions en adoptant un point de vue algorithmique sur les constructions géométriques inspirées par Rauzy.

## 2.2.4 Les fractals de Rauzy : différentes approches

Les fractals de Rauzy ont d'abord été introduits dans [Rau82] pour étudier l'addition du nombre de Tribonacci (qui vérifie  $\alpha^3 = \alpha^2 + \alpha + 1$ , avec  $|\alpha| < 1$ ) sur le tore de dimension 2 défini par  $\mathbb{C}/(\mathbb{Z} + \mathbb{Z}\alpha)$ . Rauzy a montré que cette addition est codée symboliquement par le système dynamique substitutif engendré par la substitution de Tribonacci définie par  $\sigma(1) = 12$ ,  $\sigma(2) = 13$ ,  $\sigma(3) = 1$ . Il s'agit du premier exemple d'échange de morceaux dans  $\mathbb{R}^2$  qui était symboliquement codé par une substitution.

Ces fractals ont été repris par Thurston dans un cadre de beta-numération, qui les a alors nommé *tuiles centrales* [Thu89].

Ensuite, différents travaux ont généralisé leur définition pour des substitutions Pisot ou des beta-numérations. Il y a globalement trois méthodes de construction pour les fractals de Rauzy.

- La première est inspirée par la construction initiale de Rauzy [Rau82] et considère des séries entières. Elle a été étudiée en particulier dans [Mes98, Mes00, SW02]. C'est cette approche qui a inspirées la dynamique arithmétique de [Ver92, Sch00a]. Il s'agit alors de considérer une substitution Pisot irréductible unitaire (matrice d'incidence avec une unique valeur propre supérieure

à 1).

- La seconde approche s’appuie sur les propriétés d’autosimilarité du fractal et des substitutions généralisées pour le construire par approximations successives ; elle a été introduite par l’école japonaise [IK91, AI01, AIS01, IR06]. Elle a d’abord été définie pour les substitutions Pisot unitaires irréductibles, avant d’être étendue à toutes les substitutions Pisot unitaires. Même si le point de vue est plus dynamique, la construction d’un fractal de Rauzy par espace de pavage développée par l’école américaine [BK06, BBK06] entre dans ce cadre.
- La troisième approche suivie à partir des beta-numérations dans [Thu89] consiste à représenter de manière compacte l’ensemble des entiers pour une beta-numération (voir Eq. (2.1)). Cette approche a particulièrement été étudiée dans [Aki98, AS98, Aki99, BBLT06] On peut associer à ce nombre de Pisot une substitution, mais sa matrice d’incidence peut admettre des valeurs propres dilatantes autres que  $\beta$ , on s’éloigne donc des deux premiers cas (on dit qu’on se trouve alors dans le cas Pisot *réductible*, éventuellement non unitaire).

**Contribution** Il s’avère que le lien entre ces trois approches se repose que sur une seule propriété : les ensembles obtenus vérifient les mêmes relations d’autosimilarité, ce qui permet de montrer qu’ils sont égaux (nous reviendrons sur ce point en détail dans la section suivante).

Pendant et après ma thèse, j’ai travaillé sur l’unification de ces approches, en étendant la première construction de Rauzy à différentes classes de substitutions Pisot, ce qui a permis d’englober dans un même formalisme les classes traitées par les autres méthodes de construction.

- La classe des substitutions irréductibles unitaires est étudiée dans (Trans. AMS, 2001) et (JTN Bordeaux, 2001), en collaboration avec V. Canterini.
- La classe des substitutions réductibles est étudiée dans (Integers, 2005) et (JNT, 2007), en collaboration avec Valérie Berthé, créant un pont avec les approches de numérations.
- La relation entre la construction par beta-numération et celle par approximation successive est étudiée dans (Integers, 2005) : nous y montrons que les nombres remarquables du point de vue de la beta-numération (beta-entiers) ne sont rien d’autre que les coordonnées des sommets de plans discrets, qui sont remarquables du point de vue des approximations successives.
- Le cas non unitaire est introduit dans (Erg. Th. Dyn. Sys, 2004) et repris dans (JNT, 2007). Il s’agit de la seule représentation connues pour des substitutions non unitaires.

Nous allons maintenant détailler la première construction, et rapidement mentionner pourquoi les autres constructions permettent d’obtenir le même ensemble.

## 2.2.5 Construction du fractal de Rauzy

Nous présentons ici une construction synthétique du fractal de Rauzy inspirée par la construction arithmétique de Rauzy, qui intègre cependant le point de vue



géométrique développé par l'école japonais. Cette construction est détaillée dans une monographie écrite en collaboration avec J. Thuswaldner en cours de soumission.

**Hypothèses sur la substitution** On considère une substitution sur un alphabet  $\mathcal{A}$  qui contient  $n$  éléments. Sa matrice d'incidence (obtenue par linéarisation) est notée  $\mathbf{M}_\sigma$ . On suppose que cette matrice est primitive. En particulier, cette matrice admet une unique valeur propre dominante, que l'on note  $\beta$ .

On suppose d'abord que  $\beta$  est un nombre unitaire ; nous discuterons du relâchement de cette hypothèse dans le chapitre suivant.

On note  $d$  le degré algébrique de  $\beta$ . On sait ainsi que  $d \leq n$ . Cependant, il n'y a pas nécessairement égalité, puisque le polynôme caractéristique de  $\mathbf{M}_\sigma$  peut être réductible. Il s'agit d'une des différences majeures entre l'approche des fractals de Rauzy venant des systèmes dynamiques (où on a très longtemps supposé que le polynôme caractéristique était irréductible) et l'approche venant des systèmes de numération (où l'hypothèse d'irréductibilité est rarement vérifiée).

On suppose que  $\beta$  est un nombre de Pisot : ses conjugués de Galois sont de module inférieur à 1. On note  $\beta^{(2)}, \dots, \beta^{(r)}$  les conjugués réels et  $\beta^{(r+1)}, \overline{\beta^{(r+1)}}, \dots, \beta^{(r+s)}, \overline{\beta^{(r+s)}}$  les conjugués complexes.

**Vecteurs propres** On note  $\mathbf{e}_i$  les vecteurs propres canoniques de  $\mathbb{R}^n$ . On note  $\mathbf{v}_\beta$  le vecteur propre dominant de  ${}^t\mathbf{M}_\sigma$  tel que  $\langle \mathbf{v}_\beta, \mathbf{e}_1 \rangle = 1$  et  $\mathbf{u}_\beta$  le vecteur propre dominant de  $\mathbf{M}_\sigma$  tel que  $\langle \mathbf{v}_\beta, \mathbf{u}_\beta \rangle = 1$ . Plusieurs normalisations ont en fait été considérées dans la littérature. Celle-ci permet de retrouver les beta-numération.

Les deux vecteurs ont leurs coordonnées dans  $\mathbb{Q}(\beta)$ , et ces coordonnées sont toutes positives. On construit des vecteurs propres pour les autres valeurs propres en remplaçant  $\beta$  par chaque  $\beta^{(i)}$  dans les coordonnées des vecteurs propres.

**Décomposition de l'espace** À partir de ces vecteurs, on décompose l'espace  $\mathbb{R}^n$  en différents sous-espaces :

- L'espace *beta-contractant* de  $\mathbf{M}_\sigma$  est le sous-espace  $\mathbb{H}_c$  engendré par les vecteurs propres pour les valeurs propres conjuguées à  $\beta$ . Il est de dimension  $d-1$ . On note  $\mathbf{h} : \mathbb{H}_c \rightarrow \mathbb{H}_c$  la restriction de  $\mathbf{M}_\sigma$  à  $\mathbb{H}_c$  ; c'est donc une contraction.
- On note  $\mathbb{H}_e$  la *droite dilatante* de  $\mathbf{M}_\sigma$ , dirigée par  $\mathbf{u}_\beta$ .
- L'espace *beta-supplémentaire* est l'espace invariant  $\mathbb{H}_s$  à coordonnées entières tel que  $\mathbb{R}^n = \mathbb{H}_c \oplus \mathbb{H}_e \oplus \mathbb{H}_s$ . Il est engendré par les espaces propres pour les valeurs propres de  $\mathbf{M}_\sigma$  qui ne sont pas conjuguées à  $\beta$  [BBK06, EIR06].

On note  $\pi : \mathbb{R}^n \rightarrow \mathbb{H}_c$  la projection canonique sur  $\mathbb{H}_c$  parallèlement à  $\mathbb{H}_e \oplus \mathbb{H}_s$ . Par construction,  $\pi$  vérifie une relation de commutation avec la contraction  $\mathbf{h}$  : pour tout  $w \in A^*$ , on a  $\pi(\mathbf{l}(\sigma(w))) = \mathbf{h}\pi(\mathbf{l}(w))$ , où  $\mathbf{l}$  désigne la linéarisation de  $A^*$  dans  $\mathbb{R}^n$ .

**Ligne brisée et fractal** Puisque la substitution est supposée primitive, elle admet au moins un mot périodique bi-infini, noté  $u = \dots u_{-k} \dots u_{-1} u_0 u_1 u_2 \dots$  [Que87]. On plonge ce mot bi-infini en une ligne brisée de  $\mathbb{R}^n$  en remplaçant chaque lettre de  $u$  par le vecteur canonique correspondant.

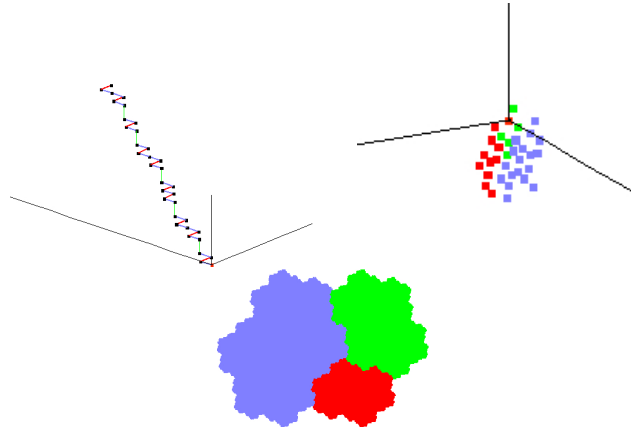


FIG. 2.3 – L’escalier représentant le point fixe de la substitution de Tribonacci définie par  $\sigma(1) = 12$ ,  $\sigma(2) = 13$ ,  $\sigma(3) = 1$ . En haut, le même escalier est vu depuis différents points focaux de l’espace : l’escalier s’enroule autour d’une droite ; vu depuis cette droite, les sommets de l’escalier forment un ensemble borné. En bas, l’escalier est projeté dans le plan contractant de la matrice d’incidence, ce qui donne un fractal de Rauzy.

Dans le cas irréductible, selon l’hypothèse Pisot, la ligne brisée reste à une distance bornée de la droite dilatante [AI01]. Dans le cas réductible, la projection sur le beta-espace contractant reste toujours un ensemble borné. À partir de cette projection, on définit le *fractal de Rauzy* (Trans. AMS, 2001).

$$\mathcal{T} := \overline{\{\pi(\mathbf{I}(u_0 \dots u_{k-1})) \mid k \in \mathbb{N}\}}.$$

Les sous-tuiles du fractal sont naturellement définies par la direction de l’escalier qui est projetée pour chaque point :  $\mathcal{T}(i) := \overline{\{\pi(\mathbf{I}(u_0 \dots u_{k-1})) \mid k \in \mathbb{N}, u_k = i\}}$ . On a donc  $\mathcal{T} = \bigcup_{i \in A} \mathcal{T}(i)$ .

La primitivité de la substitution implique aussi que la définition du fractal et de ses sous-tuiles est indépendante du point périodique choisi (voir [AI01, Que87] et (Trans. AMS, 2001)).

### 2.2.6 Propriété topologique fondamentale : autosimilarité

**Autosimilarité dirigée par un graphe orienté** La principale propriété des fractals de Rauzy est leur propriété d’autosimilarité. Selon [MW88], on dit qu’un compact  $K$  est *autosimilaire dirigé par un graphe orienté* (GIFS) lorsqu’il existe une décomposition du compact  $K = \bigcup K_i$  (les  $i$  désignant les sommets du graphe) et des contractions  $\tau_e$  (les  $e$  désignant les arêtes du graphe) telles que chaque sous-ensemble vérifie

$$K_i = \bigcup_{i \rightarrow j} \tau_e(K_j).$$

Autrement dit, chaque ensemble  $K_i$  se décompose en copies de différents autres ensembles  $K_j$ , préalablement contractés par des applications  $\tau_e$ . Cette opération de décomposition est décrite par un graphe orienté dont les sommets désignent les ensembles  $K_i$  eux-mêmes et pour lesquels chaque copie de  $K_j$  au sein de  $K_i$  correspond à une arête du graphe. Nous donnerons un exemple explicite à la Figure 2.4.

La propriété principale de ces ensembles est que le compact et sa décomposition vérifiant l'équation associée à un GIFS est unique, dès que toutes les applications  $\tau_e$  sont des contractions. Nous allons utiliser cette unicité pour montrer que différentes définitions de compacts produisent le même ensemble.

**Désubstitution et décomposition du système symbolique** Les fractals de Rauzy permettent de représenter le système dynamique  $X_\sigma$  engendré par le point périodique d'une substitution, qui est formé des mots bi-infinis ayant le même langage que le point périodique. Sur cet ensemble de mots, la substitution agit comme une contraction. Après diverses tentatives, il a été prouvé qu'il est possible d'utiliser cette contraction pour simuler une division sur  $X_\sigma$  : tout mot bi-infini  $v$  est l'image d'un mot  $w$  (dont la lettre centrale est notée  $w_0$ ) du même système par la substitution, mis à part un petit décalage dans la séquence :

$$v = S^k(\sigma(w)), \quad \text{avec } w \in X_\sigma \text{ et } k < |\sigma(w_0)|.$$

Dans cette écriture, on sait que le couple  $(k, w)$  est unique [Mos96, DHS99]. Notons que cette unicité est loin d'être évidente, il s'agit de la principale raison pour laquelle le système symbolique considéré contient des mots *bi-infinis* et pas des mots infinis à droite seulement (en fait, pour certaines substitutions, cette unicité est fautive si on ne considère que des mots infinis à droite).

Dans cette écriture, on peut considérer  $w$  comme le quotient de  $v$  par  $\sigma$  et  $k$  comme le reste de cette division. Plutôt que  $k$ , on préférera souvent le préfixe de longueur  $k$  de  $\sigma(w_0)$ . Cette écriture permet de décomposer les éléments d'une sous-tuile en fonction de leur origine : pour tout mot bi-infini  $v$  qui commence par la lettre  $i$ , il existe, un unique mot  $w$  et une décomposition de l'image de la lettre centrale  $\sigma(w_0)$  sous la forme  $\sigma(w_0) = pis$ , avec  $p$  de longueur  $k$ , tel que  $v$  coïncide avec  $\sigma(w)$  modulo un décalage de la longueur de  $p$ .

**Graphe directeur des fractals de Rauzy** Ce raisonnement s'applique aux mots bi-infinis dans  $X_\sigma$  mais il peut être transposé sur les débuts du point périodique. Hors, lors de la projection par  $\pi \circ \mathbf{l}$ , la substitution  $\sigma$  commute avec la contraction  $\mathbf{h}$ , et le décalage  $S$  commute avec une translation. On en déduit la relation suivante (voir [AI01] et (Trans. AMS, 2001)) :

$$\mathcal{T}(i) = \bigcup_{\sigma(j)=pis} \mathbf{h}\mathcal{T}(j) + \pi\mathbf{l}(p).$$

Ainsi, le fractal de Rauzy d'une substitution primitive est la solution d'un GIFS dirigé par un graphe orienté appelé *graphe des préfixes-suffixes*.

---

**Définition 2.2.1** *Les sommets du graphe des préfixes-suffixes sont les éléments de l'alphabet  $\mathcal{A}$ .*

*Il y a une arête de  $i$  vers  $j$  dans ce graphe, étiquetée par la décomposition  $(p, i, s)$  si et seulement si  $\sigma(j) = pis$ .*

Vincent Cantérini et moi-même avons proposé ce graphe dans (JTN Bordeaux, 2001). D'autres versions avaient été introduites antérieurement. Ainsi, dans ses travaux sur la substitution de Tribonacci, Rauzy travaillait avec un graphe des suffixes (les arêtes ne contiennent que le suffixe  $s$ ) [Rau82]. Dans le cadre des beta-numérations, un graphe similaire apparaît, étiqueté seulement par des préfixes [DT93]. Or, ces deux formalismes (préfixes ou suffixes uniquement) ne conviennent pas pour un système substitutif d'un point de vue dynamique : la principale raison est que les étiquettes de chemins du graphe ne sont pas en bijection avec les chemins eux-mêmes ; il peut exister deux chemins étiqueté de manière semblable qui partent de différents sommets. Avec Vincent Canterini, nous avons donc proposé le graphe des préfixes-suffixes, et nous avons ainsi pu généraliser les résultats de Rauzy sur la description du système engendré par la substitution de Tribonacci.

**Théorème 2.2.2 (JTN Bordeaux, 2001)** *Soit  $\sigma$  une substitution primitive Pisot. Le système symbolique constitué de mots bi-infinis engendré par  $\sigma$  est isomorphe en mesure à un odomètre sur l'ensemble des étiquettes de chemins infinis reconnus par le graphe des préfixes-suffixes.*

**Tuiles disjointes** Dans [Rau82], Rauzy montrait facilement que l'aire des intersections entre les sous-tuiles est nulle : ces tuiles ne s'intersectent que sur leur frontière. Il s'appuyait fortement sur cette propriété pour obtenir ses résultats, comme nous le détaillerons plus loin.

Cependant, lors de la construction d'autres fractals, il est vite apparu que cette propriété de sous-tuiles disjointes en mesure n'était pas prouvée. Après différents tâtonnements, une condition combinatoire a été introduite dans [AI01] pour assurer cette propriété (voir le survol (Pytheas Fogg, LNM, 2002)). Grossièrement, on dit qu'une substitution vérifie la *condition de fortes coïncidences* dès que toutes les paires de mots bi-infini périodiques pour la substitution partagent un segment en commun lorsqu'elles sont représentées sous la forme d'une ligne brisée dans  $\mathbb{R}^n$ .

Cette condition combinatoire, simple a priori, est en fait assez frustrante : on sait que toutes les substitutions Pisot sur deux lettres la vérifient [BD02], on ne connaît aucun contre-exemple sur des alphabets plus gros, mais on ne sait pas pour l'instant démontrer que toutes les substitutions Pisot la vérifient.

Cette condition est cependant fondamentale, dans la mesure où elle assure que les fractals pour les substitutions qui la vérifient sont bien la solution unique du GIFS donné par le graphe des préfixes-suffixes.

**Théorème 2.2.3 ([AI01, SW02], (Trans. AMS, 2001), (JNT, 2007))** *Soit  $\sigma$  une substitution primitive, unitaire et Pisot. Si  $\sigma$  satisfait la condition de fortes coïncidences, alors les sous-tuiles de son fractal de Rauzy ont leurs intérieurs disjointes. Ces sous-tuiles sont l'unique solution du GIFS donné par le graphe des préfixes-suffixes.*

Ce théorème a une histoire assez longue. Il a d'abord été démontré par Rauzy pour la substitution de Tribonacci [Rau82]. Ensuite, ce résultat a été prouvé par Arnoux et Ito dans [AI01] pour le cas des substitutions irréductibles, avec une construction du fractal basée sur des approximations successives. Sirvent et Wang ont insisté sur l'aspect GIFS, toujours pour le cas des substitutions irréductibles [SW02].

La preuve de ce résultat pour les fractals obtenus par la méthode arithmétique de Rauzy [Rau82] a été donnée par V. Canterini et moi-même dans (Trans. AMS, 2001), en s'appuyant sur la notion de désubstitution étudiée dans (JTN Bordeaux, 2001). Nous en déduisons des résultats sur les facteurs topologiques des systèmes substitutifs. L'intérêt de cette construction par rapport à celle de [AI01] était que nous pouvions envisager de travailler sur des substitutions non unitaires, comme nous aurons l'occasion d'en discuter plus tard.

Surtout, la généralisation de ces résultats au cas réductible (où le polynôme caractéristique de la matrice n'est plus irréductible) a été proposée dans (JNT, 2007), par Valérie Berthé et moi-même, pour considérer les problèmes liés aux beta-substitutions. Cela a en particulier d'explicitier le lien entre les fractals de Rauzy et les systèmes de numération. Le cas des substitutions non unitaires a aussi été traité dans (JNT, 2007), nous en rediscutons au chapitre 3.

**Exemple** Considérons la substitution  $\sigma_0(1) = 112$ ,  $\sigma_0(2) = 113$ ,  $\sigma_0(3) = 4$ ,  $\sigma_0(4) = 1$ . Il s'agit d'une substitution primitive unitaire, donc la valeur propre dominante est un nombre de Pisot de degré 3 ( $\beta^3 - 3\beta^2 + \beta - 1 = 0$ ).

Pour obtenir le dessin de la figure 2.4, la procédure consiste à

- Calculer les  $2^{17}$  premiers termes du point fixe (par itération successive).
- Calculer la projection de chaque sommet de la ligne brisée sur le plan contractant de la matrice. Pour cela, on remarque que la différence entre un sommet de la ligne brisée et son successeur est égale à la projection  $\pi(\mathbf{e}_i)$  (sur le plan contractant), d'un vecteur de base de l'espace. Il "suffit" donc de calculer les  $\pi(\mathbf{e}_i)$  et de les ajouter une à une en suivant l'ordre des lettres du point fixe.
- Pour éviter des erreurs d'approximation, les coordonnées des projections des sommets de la ligne brisée doivent d'abord être calculées dans les corps  $\mathbb{Q}(\alpha_i)$  (où les  $\alpha_i$  sont les valeurs propres contractantes) avant de les plonger dans  $\mathbb{C}$  pour tracer la figure.

Par construction, le fractal a 4 sous-tuiles, ce qui correspond au nombre de lettres de l'alphabet. La première  $\mathcal{T}(1)$  se décompose en cinq pièces (qui correspondent au nombre de 1 apparaissant dans une image de lettres), c'est-à-dire

$$\mathcal{T}(1) = \mathbf{h}\mathcal{T}(1) \cup (\mathbf{h}\mathcal{T}(1) + \pi(e_1)) \cup \mathbf{h}\mathcal{T}(2) \cup (\mathbf{h}\mathcal{T}(2) + \pi(e_1)) \cup \mathbf{h}\mathcal{T}(4).$$

Les autres pièces ne sont que des copies des précédentes :

$$\mathcal{T}(2) = \mathbf{h}(\mathcal{T}(1) + 2\pi(e_1)) \quad \mathcal{T}(3) = \mathbf{h}(\mathcal{T}(2) + 2\pi(e_1)) \quad \mathcal{T}(4) = \mathbf{h}(\mathcal{T}(3)).$$

Pour retrouver cette décomposition, il suffit de considérer le graphe des préfixes-suffixes de la substitution, montré dans la figure 2.4 : par définition du graphe, pour

chaque tuile  $\mathcal{T}(i)$ , il y a autant de sous-tuiles que de flèches qui partent du sommet  $i$  dans le graphe. De plus, le vecteur de translation associé à la sous-tuile  $\mathcal{T}(j)$  dans la décomposition de  $\mathcal{T}(i)$  est donné par le préfixe apparaissant dans l'étiquette. Ainsi, puisqu'une seule arête part des sommets 2, 3 et 4, on sait que les tuiles  $\mathcal{T}(2)$ ,  $\mathcal{T}(3)$  et  $\mathcal{T}(4)$  ne se décomposeront qu'en une seule sous-tuile.

Le fractal et sa décomposition sont montrés dans la figure 2.4. Dans cette figure, la contraction  $\mathbf{h}$  est la composée d'une homothétie contractante et d'une rotation d'environ  $\pi/2$  : l'image des tuiles (qui ont la forme de rectangles horizontaux) est alors des rectangles pratiquement verticaux.

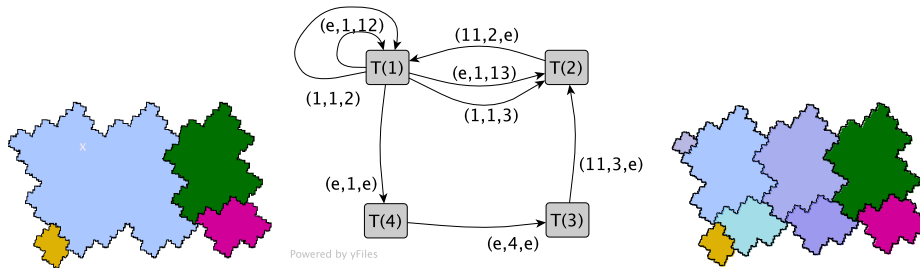


FIG. 2.4 – Le fractal de Rauzy pour la substitution  $\sigma_0(1) = 112$ ,  $\sigma_0(2) = 113$ ,  $\sigma_0(3) = 4$ ,  $\sigma_0(4) = 1$ . À gauche, chaque sous-tuile est décomposée en fonction des autres sous-tuiles (sur lesquelles ont préalablement appliqué  $\mathbf{h}$ ). L'application  $\mathbf{h}$ , qui contracte les éléments et les fait tourner de  $\pi/2$ . Ainsi, les images par  $\mathbf{h}$  des sous-tuiles (qui sont presque horizontales dans la figure de gauche) sont presque verticales dans la figure du droite.

Le graphe des préfixes-suffixes, qui représente la décomposition, est montré au centre. Le symbole  $e$  désigne le mot vide. Pour chaque tuile  $\mathcal{T}(i)$ , il y a autant de sous-tuiles que de flèches qui partent du sommet  $i$  dans le graphe.

## 2.2.7 Retour à la dynamique symbolique

**Échange de morceaux dans le fractal** Pour montrer que l'addition par le nombre de Tribonacci sur un tore de dimension 2 est symboliquement décrite par un système substitutif, Rauzy s'appuie sur deux points

- D'abord, montrer que le système symbolique engendré par la substitution commute avec une addition sur un tore. Concrètement, il montre que le système symbolique est isomorphe en mesure à un échange de morceaux sur un compact de  $\mathbb{R}^2$  puis il montre que cet échange de morceaux se quotiente en une addition sur un tore.
- Ensuite, montrer que le quotient de  $\mathbb{R}^2$  dans le tore  $\mathbb{T}^2$  est injectif. Cela revient à montrer que le fractal de Rauzy est un domaine fondamental de  $\mathbb{R}^2$  pour un réseau bien choisi.

Lors de la généralisation de la définition des fractals aux substitutions Pisot, ces deux points ont été traités séparément. Nous allons voir que la condition combinatoire de forte-coïncidence déjà introduite permet de généraliser le premier point. Nous verrons dans la section suivante qu'une autre condition combinatoire (dite de super-coïncidences) permet de généraliser le second point, ainsi que d'autres conditions plus géométriques ou arithmétiques.

**Échange de morceaux** La condition de fortes coïncidences assure que les pièces du fractal sont disjointes en mesure.

Il existe une autre partition naturelle du fractal, qui consiste, dans la définition des sous-tuiles  $\mathcal{T}(i)$ , à remplacer la condition  $u_k = i$  par la condition  $u_{k-1} = i$ . On obtient une autre partition du fractal en pièces disjointes, et on passe de la première partition à la seconde en translatant chaque pièce  $\mathcal{T}(i)$  par le vecteur  $-\pi \mathbf{1}(i)$ .

On définit ainsi une transformation sur le fractal de Rauzy : coder cette transformation selon la partition en sous-tuiles permet de retrouver exactement le système engendré par la substitution. La partition est génératrice puisque chaque point du fractal admet un unique codage sauf sur un ensemble de mesure nulle.

**Théorème 2.2.4 ([AI01])** *Soit  $\sigma$  une substitution primitive unitaire et Pisot. On définit un échange de morceaux sur le fractal de Rauzy par :  $\mathbf{f} : \mathbf{x} \in \mathcal{T}(i) \mapsto \mathbf{x} + \pi \mathbf{e}_i \in \mathcal{T}$ . Si  $\sigma$  vérifie la condition de fortes coïncidences, le codage symbolique de la transformation  $\mathbf{f}$  selon la partition donnée par les sous-tuiles est exactement le système symbolique engendré par la substitution  $\sigma$ . De plus, ce codage est injectif en mesure, ce qui signifie que la partition est génératrice pour l'échange de morceaux.*

Là encore, ce résultat a été montré par étapes. D'abord dans le cas irréductible dans [AI01]. Avec Vincent Canterini, nous avons redémontré ce résultat avec les approches de séries entières dans (Trans. AMS, 2001). Cette nouvelle approche m'a permis d'introduire le cas non unitaire dans (Erg. Th. Dyn. Sys., 2003). Le cas réductible a été traité par Valérie Berthé et moi-même dans (JNT, 2007).

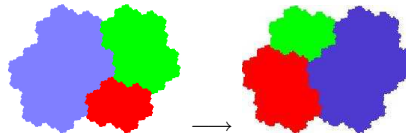


FIG. 2.5 – Le fractal de Rauzy pour la substitution  $\sigma(1) = 12, \sigma(2) = 13, \sigma(3) = 1$  et l'échange de morceaux défini à partir de ses trois sous-tuiles.

**Quotient sur le tore : une condition suffisante** On cherche toujours à comprendre des additions sur un tore. Pour cela, il faut remplacer l'échange de morceaux par une simple addition. Une manière simple est de quotienter le plan beta-contractant dans lequel vit le fractal de Rauzy par un réseau qui assure que tous les

vecteurs de translations  $\pi(\mathbf{e}_i)$  sont égaux dans le quotient. L'existence de ce passage au quotient est assuré dans le cas où la substitution est irréductible (la taille de l'alphabet est égale au degré de la valeur propre dominante).

Dans le cas général, Joerg Thuswaldner et moi-même avons montré qu'une condition supplémentaire est nécessaire pour pouvoir procéder à un passage au quotient.

**Définition 2.2.5 (Monographie S. & T., 2008)** *On dit qu'une substitution primitive Pisot vérifie la condition de quotient map, s'il existe  $d$  lettres distinctes  $B(1), \dots, B(d)$  dans  $\mathcal{A}$  telles que*

$$\forall i \in \{1, \dots, n\} \quad \langle \mathbf{e}_i - \mathbf{e}_{B(1)}, \mathbf{v}_\beta \rangle \in \sum_{k \in \{2, \dots, d\}} \mathbb{Z} \langle \mathbf{e}_{B(k)} - \mathbf{e}_{B(1)}, \mathbf{v}_\beta \rangle. \quad (2.2)$$

On vérifie par exemple que la substitution  $\sigma_0$  vérifie cette condition.

Par contre, une des substitutions canoniques issues de la beta-numération ne vérifie pas cette condition  $\sigma_1(1) = 12, \sigma_1(2) = 3, \sigma_1(3) = 4, \sigma_1(4) = 5, \sigma_1(5) = 1$  (voir la monographie). Cette substitution entre dans le cadre des contre-exemples pour la propriété de systèmes substitutifs à spectre purement discret discuté dans la section [BBK06]. Le système substitutif n'est pas isomorphe à une translation sur un tore : puisqu'il ne vérifie pas la condition de quotient map, son spectre est trop large d'un point de vue algébrique pour être associé à un réseau. Par contre, comme cela a été montré dans [EI05], le système substitutif est isomorphe à l'application de premier retour d'une translation sur un tore, qui est obtenu comme le quotient de  $\mathbb{R}^2$  par un réseau dont le domaine fondamental consiste à adjoindre au fractal de Rauzy une copie de sa plus grande pièce.

**Systèmes isomorphes à une translation sur un tore** Pour les substitutions  $\sigma$  qui vérifient à la fois la condition de fortes coïncidences et la condition de quotient, nous avons montré que le système substitutif admet pour facteur topologique une translation sur un tore de dimension  $d - 1$ , où  $d$  est le degré algébrique de la valeur propre dominante de la matrice (voir (Trans. AMS, 2001) et (Monographie S. & T., 2008)). De plus, la projection du système substitutif  $X_\sigma$  sur le tore admet des fibres qui sont constantes presque partout (voir une ébauche de la preuve dans (Pytheas Fogg, LNM, 2002), ainsi que dans [IR06]).

En particulier, du point de vue des codages symboliques, on s'approche d'une condition pour que le système substitutif soit une translation sur un tore. En collaboration avec Vincent Cantérini (pour le cas irréductible) et Joerg Thuswaldner (pour le cas réductible), nous avons obtenu la condition suivante.

**Proposition 2.2.6 (Trans. AMS, 2001) et (Monographie S. & T., 2008)** *Soit  $\sigma$  une substitution primitive Pisot unitaire qui vérifie la condition de fortes coïncidences et la condition de quotient. Soit  $d$  le degré algébrique de la valeur propre dominante de  $\sigma$ .*

*Si le fractal de Rauzy est un domaine fondamental du réseau du plan beta-contractant défini par  $\sum_{k \in \{2, \dots, d\}} \mathbb{Z} \langle \pi \mathbf{e}_{B(k)} - \pi \mathbf{e}_{B(1)}, \cdot \rangle$ , alors il existe une translation sur un tore de dimension  $d - 1$  et une partition de ce tore qui est codée symboliquement par le système symbolique engendré par la substitution.*



Ainsi, savoir si un système substitutif est un codage générateur pour une addition sur un tore est ramené à une question de pavage.

**Une condition suffisante pour obtenir une partition de Markov** Ce résultat et les intuitions concernant les sections de partitions de Markov décrites à la section 2.2.2 permettent aussi de proposer un procédé pour construire des partitions de Markov pour les automorphismes hyperboliques du tore unitaires, Pisot et irréductibles. Il s'agit de construire un ensemble de  $\mathbb{R}^n$  qui a une forme pre-cylindrique : la base est constituée des sous-tuiles du fractal de Rauzy dans l'hyperplan contractant, le cylindre s'élevant le long de la direction dilatante avec une hauteur qui dépend du volume de la sous-tuile considérée :

$$\mathcal{X} = \cup \mathcal{T}(i) \times [0, \langle \mathbf{u}_\beta, \mathbf{e}_i \rangle].$$

**Proposition 2.2.7 (Thèse de troisième cycle)** *Soit  $\mathbf{M}$  un automorphisme hyperbolique du tore  $\mathbb{T}^n$  à coefficients tous positifs qui admet une unique valeur propre dilatante. Soit  $\sigma$  une substitution qui admet  $\mathbf{M}$  pour matrice d'incidence. Si  $\sigma$  vérifie la condition de fortes coïncidences et si le fractal de Rauzy est un domaine fondamental du réseau engendré par les vecteurs  $\pi(\mathbf{e}_k) - \pi(\mathbf{e}_1)$  pour  $k = 2 \dots n$ , alors l'ensemble  $\mathcal{X}$  est une partition de Markov pour l'action de  $\mathbf{M}$  sur le tore.*

Ce résultat a été démontré dans [IR06] avec des approches inspirées par la géométrie discrète. Auparavant, j'en avais donné une démonstration dans ma thèse de troisième cycle.

Dans le cas plus général où l'automorphisme admet plusieurs valeurs propres dilatante mais où la plus grande est un nombre de Pisot, alors le fractal de Rauzy permet de construire une partition de Markov pour l'action de la matrice sur l'espace engendré par les espaces propres associés à la valeur propre dominante et à ses conjugués algébriques [IR06].

Comme dans le cas des additions, on a ainsi ramené l'existence d'une partition de Markov pour un automorphisme du tore à une question de pavage. Nous allons pouvoir maintenant nous attaquer concrètement à la vérification de cette propriété de pavage.

### 2.3 Obtenir (enfin) une bonne dynamique symbolique

Finalement, que ce soit pour comprendre la dynamique symbolique de certaines additions ou multiplications sur un tore, nous sommes ramenés à considérer des substitutions primitives Pisot et unitaires et nous cherchons à savoir si le système symbolique engendré par n'importe quel point périodique de la substitution est le codage d'une translation sur un tore. Pour cela, nous avons proposé un candidat pour la translation sur le tore, ainsi que la partition du tore qui produirait un codage convenable. La seule hypothèse à vérifier maintenant est de savoir si le fractal de Rauzy est un domaine fondamental pour le réseau considéré.

### 2.3.1 Des pavages périodiques aux pavages autosimilaires

**Ébauche de la preuve pour la substitution de Tribonacci** Pour vérifier que le fractal de Rauzy est un domaine fondamental pour un réseau dans le cas de la substitution de Tribonacci, Rauzy a montré que l'aire du fractal était exactement égale à l'aire d'une maille du réseau. Plus précisément, il a identifié les points du réseau qui sont inclus dans le fractal [Rau82]. Or, cette démonstration n'a pas pu être généralisée, elle est vraiment spécifique au système de numération associé au nombre de Tribonacci. Une autre direction de recherche a donc été engagée par la communauté : travailler sur différentes conditions de pavages.

**Recouvrement multiple et pavages** Précisons maintenant ce qu'on entend par un pavage. Un recouvrement multiple du plan beta-contractant  $\mathbb{H}_c$  par les sous-tuiles  $\mathcal{A}$  est donné par un ensemble de translations  $\Gamma \subset \mathbb{H}_c \times A$  tel que : (1)  $\mathbb{H}_c = \bigcup_{(\gamma, i) \in \Gamma} \mathcal{T}(i) + \gamma$ , (2) Tout compact de  $\mathbb{H}_c$  intersecte un nombre fini de tuiles, (3) il existe un entier  $p$  tel que presque tous les points  $\mathbb{H}_c$  appartiennent à exactement  $p$  tuiles.

Ainsi, les éléments de  $\Gamma$  sont de la forme  $[\gamma, i]$  et donnent un nom et un emplacement aux pièces du recouvrement :  $\gamma$  dit qu'il faut positionner une tuile à l'emplacement  $\gamma$  dans l'espace  $\mathbb{H}_c$ ,  $i$  spécifie quel genre de tuile doit être positionné à cet endroit.

Quand les éléments du recouvrement ont des intérieurs disjoints deux-à-deux (autrement dit  $p = 1$ ), on dit que le recouvrement multiple est un *pavage*.

Ainsi, la proposition 2.2.6 nous dit que si la substitution vérifie la condition de quotient map (2.2) (en particulier dans le cadre irréductible), le plan  $\mathbb{H}_c$  admet un recouvrement multiple par l'ensemble de translation

$$\Gamma_{lat} = \left\{ [\gamma, i] \in \pi(\mathbb{Z}^n) \times A \mid \gamma \in \sum_{k=2}^n \mathbb{Z}(\pi(\mathbf{e}_{B(k)}) - \pi(\mathbf{e}_{B(1)})) \right\}. \quad (2.3)$$

Le recouvrement est ici périodique parce que la condition sur les tuiles est indépendante du nom des tuiles : pour tout point  $\gamma$  dans la grille  $\sum_{k=2}^n \mathbb{Z}(\pi(\mathbf{e}_{B(k)}) - \pi(\mathbf{e}_{B(1)}))$ , on place une copie de toutes les tuiles. Un exemple est illustré à la Figure 2.6 pour la substitution  $\sigma_0$  qui vérifie bien la condition de quotient.

**Recouvrement autosimilaire** On obtient un nouvel ensemble de translations pour un recouvrement multiple en considérant l'ensemble

$$\Gamma_{srs} = \{[\pi(\mathbf{x}), i] \in \pi(\mathbb{Z}^n) \times \mathcal{A}; \quad 0 \leq \langle \mathbf{x}, \mathbf{v}_\beta \rangle < \langle \mathbf{e}_i, \mathbf{v}_\beta \rangle\}. \quad (2.4)$$

D'un point de vue géométrique, cet ensemble de points et de noms de pièces correspond à une discrétisation de plan dans l'esprit de Reveilles [Rev91] : il s'agit de sélectionner les points à coordonnées entières qui sont au dessus de  $\mathbb{H}_c$  et qui, translatés par un vecteur canonique  $\mathbf{e}_i$ , sont au dessous de  $\mathbb{H}_c$  [ABI02].

On montre que cet ensemble est un ensemble de translation pour un recouvrement multiple de  $\mathbb{H}_c$  : il est uniformément discret, relativement dense et localement fini [ABI02, BK06, IR06, EIR06].  $\mathbb{H}_c = \bigcup_{[\gamma, i] \in \Gamma_{srs}} (\mathcal{T}(i) + \gamma)$ .

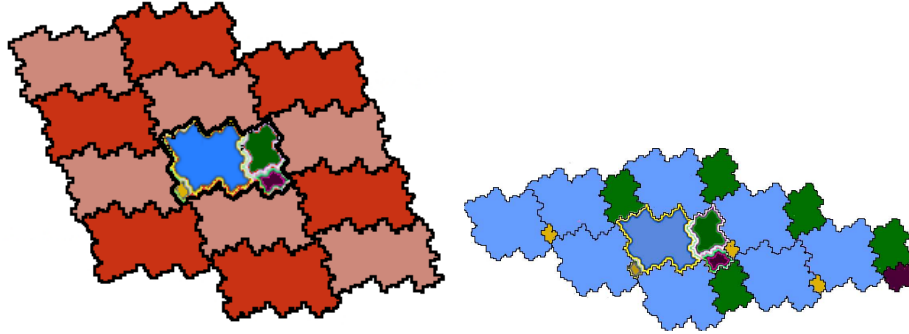


FIG. 2.6 – Recouvrements périodiques (à gauche) et autosimilaire (à droite) associé au fractal de Rauzy de la substitution  $\sigma_0$ .

La spécificité de ce recouvrement est son autosimilarité [LW96] : il est stable lorsqu'on remplace chaque tuile par sa décomposition autosimilaire puis procède à une renormalisation. Autrement dit, on vérifie que  $\Gamma_{sr,s}$  est stable par la transformation  $[\pi(\mathbf{x}), i] \rightarrow \cup_{\sigma(j)=pis} [\mathbf{x} + \mathbf{h}^{-1}\pi\mathbf{l}(p), j]$ .

**Partition de Markov** Dans la section précédente, nous avons mis en évidence une relation entre l'existence d'une partition de Markov basée sur les fractals de Rauzy et la vérification d'une propriété de pavage périodique. Avec les approches d'espaces de pavage, on montre que ces conditions sont équivalentes dans le cas irréductible.

**Théorème 2.3.1 ([IR06, EIR06, BK06])** *Soit  $\sigma$  une substitution primitive Pisot unitaire. Alors les conditions suivantes sont équivalentes :*

- Les sous-tuiles du fractal de Rauzy pavent l'espace contractant  $\mathbb{H}_c$  de manière autosimilaire.
- Le fractal de Rauzy est la base d'une partition de Markov pour l'action de la matrice d'incidence de la substitution sur un tore obtenu en quotientant l'espace engendré par les espaces propres associés à sa valeur propre dominante et à ses conjugués algébriques.

Si, de plus, la substitution est irréductible, alors ces conditions sont équivalentes au fait que le système symbolique engendré par la substitution est un codage d'une translation sur un tore de dimension  $d - 1$ , pour une partition génératrice donnée par les sous-tuiles du fractal. La partition de Markov représente alors l'action de la matrice sur le tore  $\mathbb{T}^n$ .

Ainsi, pour la matrice  $\mathbf{M} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$ , qui correspond à la substitution

de Tribonacci, une partition de Markov est donnée à partir du fractal de Rauzy, comme montré Fig. 2.7. Cette pièce pave l'espace selon un  $\mathbb{Z}^3$ , et elle montre que

l'orbite de tout point de l'espace est caractérisé par des chemins dans un graphe à trois états. Les arêtes de ce graphe ayant des étiquettes toutes distinctes, le graphe est déterministe et la représentation symbolique est markovienne.

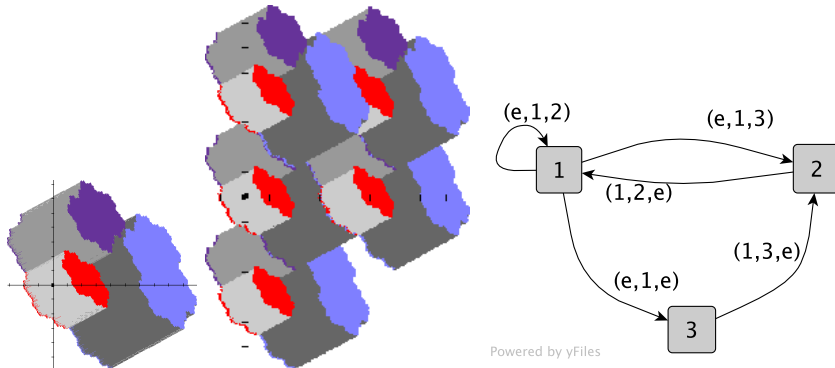


FIG. 2.7 – Partition de Markov pour l'action de matrice de la substitution de Tribonacci sur le tore de dimension 3. Cette pièce est un domaine fondamental du réseau  $\mathbb{Z}^3$ , comme illustré sur la figure centrale. À partir de cette représentation, on en déduit que l'action de la matrice sur le tore est exactement décrite par les chemins du graphe des préfixes-suffixes présenté à droite.

**Conjecture Pisot** Dans ce contexte, la conjecture Pisot exprime le fait que le recouvrement autosimilaire est un pavage. Elle a d'abord été énoncée par les physiciens [BT86] avant d'être formalisée en terme de substitutions [Fog02] dans le cadre irréductible puis étendue au cadre réductible [BBK06]. Elle se dérive différemment en fonction du type de substitution considérée.

- Pour toute substitution Pisot unitaire irréductible, les recouvrements périodiques et autosimilaires par le fractal de Rauzy sont tous les deux des pavages.
- Pour toute substitution Pisot unitaire (réductible), le recouvrement autosimilaire par le fractal de Rauzy est un pavage.

Cette conjecture est prouvée pour les substitutions irréductibles sur deux lettres.

À ce jour, tous les exemples testés par les mathématiciens n'ont pas permis d'exhiber un contre-exemple à cette conjecture. Cependant, comme nous allons le voir, les méthodes qui ont été utilisées ne sont pas réellement automatiques, et les exemples considérés concernent principalement des fractals de Rauzy dans  $\mathbb{R}^2$  ou  $\mathbb{R}^3$ . Une étude systématique de cette conjecture reste à faire, et demande pour cela des algorithmes efficaces pour vérifier la condition de pavage.

### 2.3.2 Différentes conditions de pavage

Les différentes conditions de pavage existant sont liées aux modes de construction du fractal évoqués plus haut ; avec Valérie Berthé et Joerg Thuswaldner, j'écris

actuellement une synthèse sur ce sujet (CANT, 2009) à paraître chez Cambridge University Press.

**Construction par projection d’escalier : super-coïncidences** La construction du fractal de Rauzy comme projection de lignes brisées est intimement liée à l’approche par flot de translation. De cette approche ressort une condition combinatoire sur la substitution pour que le recouvrement autosimilaire soit un pavage. Il s’agit de la condition de super-coïncidences. Cette condition est parue simultanément (issues de points de vues combinatoires et ergodiques) dans [IR06, BK06] pour le cas Pisot unitaire irréductible. Elle a ensuite été étendue au cas réductible dans [EIR06, BBK06].

Grossièrement, cette condition demande que pour toute paire de lignes brisées partant d’une face du plan discret  $\Gamma_{sr,s}$  associé à la substitution, les escaliers partagent au moins un segment en commun. Cette définition est à mettre en balance avec la condition de fortes coïncidences qui assure que les sous-tuiles du fractal sont disjointes : il fallait alors que les escaliers partant des faces du cube unité partagent un segment. Si on veut un pavage, il faut étendre cette condition à toutes les faces du plan discret.

En utilisant cette condition, il a été montré qu’une famille importante de substitutions associées à des beta-numérations vérifie la condition de pavage [BBK06]. La preuve est de nature ergodique.

La vérification de cette condition peut aussi se faire à l’aide d’un algorithme de paires balancées, dans la lignée des idées introduites dans les années 1980 pour l’étude du spectre des systèmes substitutifs ([Liv87], voir aussi mon survol (Pytheas Fogg, LNM, 2002)). Cet algorithme n’a cependant jamais été implémenté, ni sa complexité étudiée. Par construction, cet algorithme termine si et seulement si la condition de super-coïncidences est vérifiée.

**Construction arithmétique : condition (W)** Comme déjà évoqué, les fractals de Rauzy apparaissent naturellement en théorie des nombres. Par construction, le plan beta-contractant  $\mathbb{H}_c$  est isomorphe à l’ensemble  $\mathbb{R}^{r-1} \times \mathbb{C}^s$ , où  $r-1$  est le nombre de conjugués réels de  $\beta$  et  $2s$  le nombre de conjugués complexes. Or, il existe un plongement naturel de  $\mathbb{Q}(\beta)$  dans cet ensemble :

$$\varphi : x = P(\beta) \in \mathbb{Q}(\beta) \rightarrow (P(\beta^{(2)}), \dots, P(\beta^{(r+s)})) \in \mathbb{R}^{r-1} \times \mathbb{C}^s \quad (2.5)$$

Selon l’approche de Thurston [Thu89], on définit la tuile centrale associée à  $\beta$  comme la représentation par  $\phi$  de l’ensemble des entiers pour la numération défini en (2.1) :  $\tilde{T} := \overline{\varphi(\text{Int}(\beta))}$ . Avec mes collaborateurs, nous avons montré, avec une décomposition spécifique, que cette tuile centrale vérifie exactement le GIFS associé à une substitution qui est définie à partir du développement de 1 (voir (Monas. Math., 2008), (JNT, 2007), (Integers, 2005)). Par unicité des solutions d’un GIFS, on en déduit que la tuile centrale pour une beta-numération est exactement le fractal de Rauzy d’une substitution spécifique.

Dans (Integers, 2005), nous avons montré la réciproque de cette propriété : tout fractal de Rauzy d’une substitution Pisot unitaire est représentation de l’ensemble

des entiers d'une numération. Il suffit pour cela d'étendre la définition des beta-numérations aux numérations de Dumont-Thomas. En effet, selon [DT93], pour toute substitution sur un alphabet  $\mathcal{A}$  et pour chaque lettre  $a$  de  $\mathcal{A}$ , il est possible de développer les réels de l'intervalle  $[0, \langle \mathbf{v}_\beta, \mathbf{e}_a \rangle[$  sous la forme  $x = \sum_{i \geq 1} \langle \mathbf{v}_\beta, \mathbf{l}(q_i) \rangle \beta^{-i}$ , où les  $q_i$  sont des suites de préfixes lus dans le graphe des préfixes-suffixes à partir du sommet  $a$ . On note  $d_{\sigma,a}(x)$  la suite des préfixes  $q_i$ . Dans (Integers, 2005), Valérie Berthé et moi-même montrons que les chiffres  $q_i$  de ce nouveau développement peuvent être produits de manière gloutonne à partir d'un système dynamique sur une suspension d'intervalles.

**Proposition 2.3.2 (Integers, 2005)** *Soit  $\sigma$  une substitution primitive Pisot unitaire. Soit  $a$  une lettre de l'alphabet. Soit  $\mathbf{v}_\beta$  un vecteur propre dominant de la transposée de la matrice d'incidence de  $\sigma$ .*

*Pour tout  $x \in [0, \langle \mathbf{v}_\beta, \mathbf{e}_a \rangle[$ , on note  $q_1, q_2, \dots$  les préfixes d'images de lettres qui constituent le développement de Dumont-Thomas de  $x$  :  $x$  peut s'écrire sous la forme  $x = \sum_{i \geq 1} \langle \mathbf{v}_\beta, \mathbf{l}(q_i) \rangle \beta^{-i}$  où les  $q_i$  sont les étiquettes d'un chemin du graphe des préfixes-suffixes qui part de  $a$ .*

*Alors les  $q_i$  sont obtenus par codage de la dynamique suivante, qui étend les beta-transformations.*

$$T_\sigma : \bigcup_{a \in \mathcal{A}} [0, \langle \mathbf{v}_\beta, \mathbf{e}_a \rangle) \times \{a\} \rightarrow \bigcup_{a \in \mathcal{A}} [0, \langle \mathbf{v}_\beta, \mathbf{e}_a \rangle) \times \{a\}$$

$$(x, b) \mapsto (\beta x - \langle \mathbf{v}_\beta, \mathbf{l}(q) \rangle, c) \quad (2.6)$$

$$\text{avec } \begin{cases} \sigma(b) = qcs \\ \beta x - \delta_\sigma(q) \in [0, \langle \mathbf{v}_\beta, \mathbf{e}_c \rangle). \end{cases}$$

Nous en avons déduit que pour toute substitution primitive, le fractal de Rauzy est la représentation des réels qui sont des entiers pour cette nouvelle numération (Integers, 2005). Ceci a permis d'explicitier définitivement pourquoi les approches par représentation des entiers et par représentation des points fixes sont strictement équivalentes.

Du point de vue des beta-numération, la condition de pavage autosimilaire s'appelle *propriété (W)* pour *weak finiteness condition* [Hol96, Aki02]. Elle spécifie que

$$\forall z \in \mathbb{Z}[\beta^{-1}] \cap [0, 1), \forall \varepsilon > 0, \exists x, y \in \text{Fin}(\beta) \text{ tel que } z = x - y \text{ et } y < \varepsilon.$$

Pour vérifier cette condition, un algorithme est donné par [ARS04]; Il a été utilisé pour caractériser tous les nombres cubiques qui la vérifient. Dans (CANT, 2009), nous donnons une généralisation de cette condition à toutes les substitutions, à l'aide de géométrie discrète. On ne sait pas encore si un algorithme permet de la vérifier.

### Construction par approximations successives : mesure de la frontière

Une autre approche pour obtenir le fractal de Rauzy consiste à construire des approximations polygonales successives et à les renormaliser. On montre que la limite de Hausdorff de ces approximations vérifie le GIFS donné par le graphe des préfixes-suffixes, et est donc égale à au fractal de Rauzy [AI01].

Dans ce cadre, pour vérifier que la condition de pavage autosimilaire est satisfaite, il suffit de montrer que le bord des approximations successives converge vers un ensemble de mesure nulle [IR06]. Pour vérifier cela, l'école japonaise propose d'utiliser une description des approximations de la frontière par un morphisme de groupe libre, et lorsqu'il y a peu d'annulation dans cette description, en déduit qu'il y a effectivement pavage [IK91, EI05].

Cette méthode est intéressante pour certaines substitutions, mais aucun travail à ce jour n'en a déduit des critères généraux de pavage.

**Construction par autosimilarité : description de la frontière** Enfin, une dernière approche pour montrer qu'un recouvrement est un pavage consiste à calculer les intersections du fractal avec tous ses voisins dans le recouvrement considéré, puis à montrer que ces intersections sont de mesure nulle.

La propriété fondamentale utilisée ici est la relation d'autosimilarité du fractal, couplée avec le fait que les recouvrements sont définis par des ensembles localement finis permet de définir un graphe fini qui décrit les intersections du fractal avec ses voisins.

Cette approche a été d'abord utilisée dans le cas de la substitution de Tribonacci [Mes98, Mes00], puis pour des substitutions Pisot irréductibles, dans [FFIW06] pour les recouvrements autosimilaires. Je l'ai moi-même généralisée dans (Ann. Inst. Fourier, 2004) pour étudier des recouvrements périodiques.

**Avantages et limitations** Après avoir synthétisé toutes ces approches, il faut noter qu'elles ont finalement été développées de manière ad-hoc sans prendre en compte les différents points de vue existant sur les fractals de Rauzy.

- Aucune ne traite complètement de manière décidable le cas des recouvrements autosimilaires : la condition de super-coïncidences, via l'algorithme des paires balancées, n'est que semi-décidable, puisqu'il ne termine que si la condition est vérifiée. La condition (W) n'est vérifiable que pour les beta-substitutions ; la construction des graphes de frontière n'est valable que dans le cas irréductible.
- Le cas du recouvrement périodique n'est traité que dans le cas irréductible, la plupart du temps via l'équivalence avec le recouvrement autosimilaire (valable uniquement dans cette situation). Or, dans le cas réductible, il n'y a plus équivalence entre pavage autosimilaire et pavage périodique. Les méthodes décrites ne sont donc plus valables. Par exemple, aucune de ces méthodes ne permet de déterminer si le recouvrement périodique associé à la substitution  $\sigma_0$  et montré Figure 2.6 est bien un pavage.

Les seules approches qui intègrent directement l'aspect périodique sont la construction par approximations successives, qui hélas ne propose pas de critère explicite [FFIW06], et la construction par graphe de frontière, qui dans (Ann. Inst. Fourier, 2004) ne concerne que les substitution irréductibles.

Ces constats m'ont mené, avec Joerg Thuswaldner, à proposer une méthode globale pour répondre globalement à la question des pavages de manière décidable.

### 2.3.3 Graphes de frontière et condition décidable de pavage

Dans (Monographie S. & T., 2008), avec J. Thuswaldner, nous faisons une synthèse des approches par graphes de frontière de (Ann. Inst. Fourier, 2004) et [FFIW06], en incluant une nouvelle famille de graphes (dits de contacts) introduits dans [Thu06], pour généraliser leur construction au cas des substitutions réductibles et aux conditions de pavages périodiques.

**Idée générale** On considère l'intersection entre une sous-tuile  $\mathcal{T}(i)$  et une autre tuile du recouvrement, notée  $\mathcal{T}(j) + \gamma$ . Il est possible de décomposer les deux sous-tuiles selon le GIFS qui gouverne le fractal, on obtient alors

$$\begin{aligned} \mathcal{T}(i) \cap (\mathcal{T}(j) + \gamma) &= \bigcup_{\substack{\sigma(i_1)=p_1 i s_1 \\ \sigma(j_1)=p_2 j s_2}} (\mathbf{h}\mathcal{T}(i_1) + \pi\mathbf{P}(p_1)) \cap (\mathbf{h}\mathcal{T}(j_1) + \pi\mathbf{P}(p_2) + \gamma). \\ &= \mathbf{h} \bigcup_{\substack{\sigma(i_1)=p_1 i s_1 \\ \sigma(j_1)=p_2 j s_2}} (\mathcal{T}(i_1) + \mathbf{h}^{-1}\pi\mathbf{P}(p_1)) \cap (\mathcal{T}(j_1) + \mathbf{h}^{-1}\pi\mathbf{P}(p_2) + \mathbf{h}^{-1}\gamma). \end{aligned}$$

Le jeu dans les publications considérées consiste à exhiber un vecteur  $\mathbf{e}$  et un autre vecteur  $\gamma'$  dans l'ensemble de translation associé au recouvrement tel que les intersections dans la partie droite des équations ci-dessus se réécrivent sous une forme convenable à translation près :

$$(\mathcal{T}(i_1) + \mathbf{h}^{-1}\pi\mathbf{P}(p_1)) \cap (\mathcal{T}(j_1) + \mathbf{h}^{-1}\pi\mathbf{P}(p_2) + \mathbf{h}^{-1}\gamma) = \mathbf{e} + \mathcal{T}(i') \cap (\mathcal{T}(j') + \gamma')$$

on obtient finalement

$$\mathcal{T}(i) \cap (\mathcal{T}(j) + \gamma) = \mathbf{h} \bigcup_{\substack{\sigma(i_1)=p_1 i s_1 \\ \sigma(j_1)=p_2 j s_2}} \mathbf{e} + \mathcal{T}(i') \cap (\mathcal{T}(j') + \gamma') \quad (2.7)$$

Ainsi, l'intersection entre  $\mathcal{T}(i)$  et  $\mathcal{T}(j) + \gamma$  se décompose en intersections du même type, modulo des translations de vecteurs  $\mathbf{e}$ .

Or, les recouvrements considérés sont tous localement finis, il n'existe donc qu'un ensemble fini d'intersections de la forme  $\mathcal{T}(i) \cap (\mathcal{T}(j) + \gamma)$  qui sont non vides. On peut donc construire un graphe fini, qui contient toutes les intersections a priori non vides de la forme  $\mathcal{T}(i) \cap (\mathcal{T}(j) + \gamma)$ , et où les arêtes représentent des inclusions décrites dans (2.7).

#### Positionnement des différentes approches par rapport à l'idée générale

Même si cette approche semble naturelle, elle n'était pas tout-à-fait présentée de cette manière dans [FFIW06, Thu06] ou (Ann. Inst. Fourier, 2004). En particulier, deux points clé ont posé problème :

- Les tuiles  $\mathcal{T}(j) + \gamma$  étaient construites sous la forme  $\mathcal{T}(j) + \mathbf{x}$ , avec  $\mathbf{x} \in \mathbb{Z}^n$ . Autrement dit, ces publications assimilaient le plan discret  $\Gamma_{srs}$  ou le réseau  $\Gamma_{lat}$  avec  $\mathbb{Z}^n$ . Lorsque l'espace beta-contractant est un hyperplan, cela ne pose



aucun problème puisqu'il y a une correspondance une-à-une entre ces deux ensembles. Dans le cas réductible, cette correspondance n'existe plus.

Dans (Monographie S. & T., 2008), nous avons donc proposé de définir les tuiles du recouvrement sous la forme  $\mathcal{T}(j) + \gamma$ , avec  $\gamma \in \pi(\mathbb{Z}^n)$ , ce qui assure l'unicité de l'écriture et surtout le fait que les vecteurs  $\gamma$  forment un ensemble localement fini. En particulier, cela a mené à modifier la définition des ensembles de translation  $\Gamma_{srs}$  et  $\Gamma_{lat}$  par rapport à celles de [Thu06, FFIW06] en introduisant la projection  $\pi$  *au sein même de la définition des vecteurs de translation*.

- Dans le cas des pavages périodiques, les vecteurs  $\gamma'$  obtenus dans (2.7) n'appartiennent plus à l'ensemble de translation  $\Gamma_{lat}$  considéré initialement. Cela pose un problème de stabilité.

Nous avons résolu ce problème en introduisant un nouvel ensemble de translation, qui contient les vecteurs  $\gamma_0 + \gamma^*$ , où  $\gamma^* \in \Gamma_{srs}$  et  $\gamma_0$  appartient à sous-ensemble fini de  $\Gamma_{lat}$  (qui contient toutes les tuiles du pavage périodique ayant une chance d'intersecter le fractal de Rauzy). Ces nouveaux vecteurs forment un ensemble localement fini (même s'il n'a plus d'interprétation géométrique simple), et il s'agit de la seule hypothèse nécessaire pour définir un graphe de frontière.

- Une intersection de la forme  $\mathcal{T}(i) \cap (\mathcal{T}(j) + \gamma)$  est non vide si et seulement si  $\mathcal{T}(j) \cap (\mathcal{T}(i) - \gamma)$  est elle aussi non vide. Il y a donc redondance dans la construction du graphe, qui pose des problèmes d'explosion combinatoire lors des calculs, ce qui peut déjà se voir dans [Mes98] et (Ann. Inst. Fourier, 2004). Suivant la proposition de [FFIW06], nous avons proposé de définir un unique représentant pour chaque couple d'intersections, ce qui revient à ordonner les variables pour définir un bon parcours des contacts.

### Définition des graphes de frontière

1. *Nommer les noeuds du graphe : éviter la redondance.* Les noeuds des graphes de frontière représentent les intersections de la forme  $\mathcal{T}(i) \cap \mathcal{T}(j) + \gamma$ . On désigne cette intersection par un triplet  $[i, \gamma, j] \in A \times \pi(\mathbb{Z}^n) \times A$ . Pour éviter des redondances, on limite les noeuds du graphe à l'ensemble

$$\mathfrak{D} = \{[i, \gamma, j] \in A \times \pi(\mathbb{Z}^n) \times A \mid \gamma = \pi(\mathbf{x}), (\langle \mathbf{x}, \mathbf{v}_\beta \rangle > 0) \text{ ou } (\gamma = \mathbf{0} \text{ et } i \leq j)\}.$$

2. *Limiter les noeuds du graphe : utiliser la propriété de finitude locale.* Nous nous intéressons seulement aux intersections non vides. Or, les sous-tuiles sont bornées. Cela fournit une borne pour  $\gamma$  dans le triplet  $[i, \gamma, j]$  :

$$\|\gamma\| \leq \frac{2 \max\{\|\pi \mathbf{P}(p)\|; (p, a, s) \in \mathcal{P}\}}{1 - \max\{|\beta^{(j)}|; j = 2, \dots, d\}} \quad (2.8)$$

où la norme est définie sur  $\mathbb{H}_c$  par  $\|\mathbf{x}\| = \max\{|\langle \mathbf{x}, \mathbf{v}_{\beta^{(i)}} \rangle|; i = 2, \dots, r + s\}$ .

Cette condition, ajoutée à une contrainte donnée par la définition de l'ensemble de translation considéré pour le recouvrement, va assurer que les  $\gamma$  sont en nombre fini.

3. *Définir les arêtes du graphe : inclusion dans la décomposition par GIFS.* Les arêtes du graphes sont définies par la relation (2.7). Un travail technique est nécessaire pour identifier les représentants dans  $\mathcal{D}$  des intersections qui apparaissent dans la partie droite de (2.7) (voir (Monographie S. & T., 2008)).
4. *S'adapter au pavage considéré : considérer des chemins infinis.* C'est seulement à cette étape que le graphe est spécifié en fonction du recouvrement considéré (périodique ou autosimilaire). Pour un ensemble de translation donné  $\Gamma$ , le *graphe de frontière* est l'ensemble des chemins infinis dans le graphe précédent issus des sommets  $[i, \gamma, j]$  tels que  $[\gamma, j] \in \Gamma$ . On montre dans (Monographie S. & T., 2008) que les conditions qui définissent ce graphe en font un graphe fini.

**Propriété fondamentale du graphe de frontière : description des intersections** L'intérêt du graphe de frontière réside dans le résultat suivant : un graphe de frontière permet de décrire l'ensemble des tuiles qui s'intersectent dans un recouvrement, que la condition de pavage soit vérifiée ou non.

**Théorème 2.3.3 (Monographie S. & T., 2008)** *Soit  $\sigma$  une substitution primitive Pisot unitaire. Soit  $\Gamma \subset \pi(\mathbb{Z}^n) \times \mathcal{A}$  un ensemble de translations localement fini qui permet de recouvrir l'espace beta-contractant à l'aide des sous-tuiles du fractal de Rauzy.*

*L'intersection d'une sous-tuile  $\mathcal{T}(i)$  avec une autre tuile  $\mathcal{T}(j) + \gamma$  du recouvrement est non vide si et seulement si  $[i, \gamma, j]$  ou  $[j, -\gamma, i]$  est un noeud du graphe de frontière associé à  $\Gamma$ .*

De plus, le graphe de frontière fournit une description autosimilaire des intersections de la forme  $\mathcal{T}(i) \cap (\mathcal{T}(j) + \gamma)$ , notée  $B[i, \gamma, j]$ . En effet, on déduit du théorème 2.3.3 que ces intersections vérifient le GIFS

$$B[i, \gamma, j] = \bigcup_{[i, \gamma, j] \xrightarrow{\mathbf{e}} [i_1, \gamma_1, j_1] \text{ in } \mathcal{G}_\Gamma^{(B)}} \mathbf{h}B[i_1, \gamma_1, j_1] + \mathbf{e}.$$

Ceci fournit un critère explicite pour vérifier la condition de pavage : il suffit de vérifier que la mesure de Hausdorff (ou la *box-dimension*) des intersections est strictement inférieure à la dimension de l'espace contractant. Hors, on sait calculer la box-dimension d'un GIFS en fonction aux valeurs propres de la matrice graphe qui dirigé le GIFS.

On obtient donc un critère explicite, décidable et adapté à toutes les substitutions Pisot, pour vérifier qu'un recouvrement autosimilaire ou périodique est effectivement un pavage, qui est le résultat principal de ce chapitre.

**Théorème 2.3.4 (Monographie S. & T., 2008)** *Soit  $\sigma$  une substitution primitive Pisot et  $\beta$  la valeur propre dominante de sa matrice d'incidence.*

- *Le recouvrement autosimilaire associé à  $\sigma$  est un pavage si et seulement si la valeur propre dominante de la matrice du graphe de frontière pour l'ensemble de translation autosimilaire  $\Gamma_{sr,s}$  est strictement inférieure à  $\beta$ .*

- Supposons que la substitution vérifie la condition de quotient. Le recouvrement périodique associé à  $\sigma$  est un pavage si et seulement si la valeur propre dominante de la matrice du graphe de frontière pour l'ensemble de translation périodique  $\Gamma_{lat}$  est strictement inférieure à  $\beta$ .

### 2.3.4 Description de la frontière

**GIFS pour la frontière** Lorsque les conditions de pavage sont vérifiées, on en déduit en particulier une équation pour la frontière des sous-tuiles  $\mathcal{T}(i)$  :

$$\partial\mathcal{T}(i) := \bigcup_{\substack{[i,\gamma,j] \in \mathcal{G}_{srs}^{(B)} \\ [i,\gamma,j] \neq [i,0,i]}} B[i,\gamma,j] \cup \bigcup_{\substack{[j,\mathbf{0},i] \in \mathcal{G}_{srs}^{(B)} \\ j < i}} B[j,\mathbf{0},i].$$

Ces équations seront systématiquement utilisées dans le chapitre 3 pour obtenir des propriétés topologiques des fractals ainsi que des propriétés des beta-numérations.

**Exemple** Dans la figure 2.8, nous donnons le graphe de frontière pour le recouvrement autosimilaire de  $\sigma_0$ . En considérant tous les sommets qui sont de la forme  $[1,\gamma,j]$ , on peut ainsi retrouver toutes les tuiles du recouvrement qui sont adjacentes à la tuile centrale  $\mathcal{T}(1)$ . Ainsi, on montre que  $\mathcal{T}(1)$  intersecte les trois autres sous-tuiles  $\mathcal{T}(2)$ ,  $\mathcal{T}(3)$  and  $\mathcal{T}(4)$  (puisque les sommets  $[1,\mathbf{0},2]$ ,  $[1,\mathbf{0},3]$  and  $[1,\mathbf{0},4]$  apparaissent dans le graphe).

Cette tuile intersecte aussi cinq autres tuiles du recouvrement :  $\mathcal{T}(1) + \pi(0,0,1,0)$ ,  $\mathcal{T}(2) + \pi(0,0,1,0)$ ,  $\mathcal{T}(1) + \pi(0,1,-1,0)$ ,  $\mathcal{T}(1) + \pi(0,1,0,0)$  et  $\mathcal{T}(1) + \pi(1,-1,1,0)$ , qui correspondent aux sommets du graphe  $[1,\pi(0,0,1,0),1]$ ,  $[1,\pi(0,0,1,0),2]$ ,  $[1,\pi(0,1,-1,0),1]$ ,  $[1,\pi(0,1,0,0),1]$ ,  $[1,\pi(1,-1,1,0),1]$ .

En étudiant la valeur propre maximale de la matrice d'incidence de ce graphe, on montre que le recouvrement autosimilaire est bien un pavage, comme illustré Figs. 2.6 et 2.8.

**Graps des points triples, quadruples et autres** Une remarque importante est à faire pour ce formalisme : rien n'empêche de remplacer les intersections entre deux tuiles par des intersections entre trois tuiles, voire quatre tuiles. On obtient simplement des graphes de taille plus importante, mais leur définition théorique ne pose aucun problème (si ce n'est des points techniques). Ces graphes sont précisément définis dans (Monographie S. & T., 2008). Ils permettent de décrire explicitement l'ensemble des points triples et quadruples dans le pavage autosimilaire.

Par exemple, on montre formellement que, comme illustré Fig. 2.8, le pavage autosimilaire pour la substitution  $\sigma_0$  induit exactement 5 points quadruples, c'est-à-dire qu'il y a exactement 5 points du plan qui se trouvent à l'intersection d'au moins 4 tuiles du pavage, dont une appartient à la tuile centrale.

L'identification des points triples et quadruple est nécessaire pour comprendre la topologie des fractals, comme cela est discuté dans le chapitre 3.

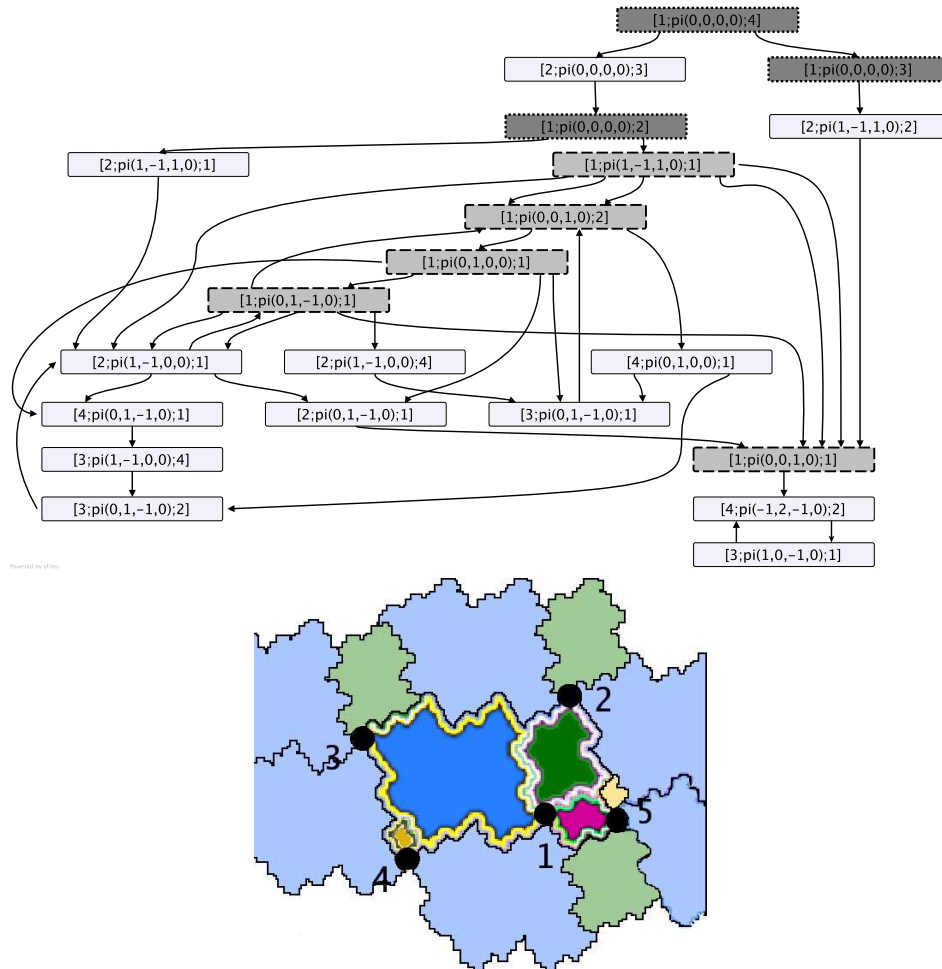


FIG. 2.8 – Graphe de frontière pour le recouvrement autosimilaire de la substitution  $\sigma_0(1) = 112$ ,  $\sigma_0(2) = 113$ ,  $\sigma_0(3) = 4$ ,  $\sigma_0(4) = 1$ . Les noms des arêtes sont omis. “pi” désigne la projection  $\pi$ . Par construction, un sommet  $[i, \gamma, j]$  apparaît dans ce graphe si et seulement si la sous-tuile  $\mathcal{T}(i)$  la sous-tuile  $\mathcal{T}(j) + \gamma$  dans le pavage autosimilaire. On en déduit par exemple que la sous-tuile  $\mathcal{T}(1)$  (en bleu foncé) a exactement 8 voisins dans le pavage autosimilaire, comme illustré dans la figure du bas.

## 2.4 Perspectives de travail : le cas non Pisot

En conclusion, la représentation symbolique des additions et des multiplications sur l'intervalle est bien connue et très exploitée. La généralisation au cas multidimensionnel est plus complexe. Un cas moins difficile que les autres se distingue, il s'agit du codage des automorphismes hyperboliques du tore à coefficients positifs qui admettent une unique valeur propre dilatante; dans ce cas, on dispose de conditions combinatoires qui assurent que l'automorphisme est représenté par un système symbolique markovien. De plus, la translation sur un tore définie à partir du vecteur propre de la transposée de la matrice est représentée par un système symbolique engendré par une substitution.

Les perspectives dans ce domaine sont principalement des questions de tests automatiques et de généralisation à des cas non Pisot.

**Complexité et expérimentations sur des alphabets grande échelle** La première question qui n'a pas été encore traitée est de déterminer la complexité des algorithmes de vérification des conditions de pavage. Ensuite, il serait intéressant d'appliquer cet algorithme à des jeux de matrices aléatoires Pisot de grande taille pour savoir si la conjecture Pisot est spécifique à des dimensions de petite taille ou si elle est encore vraie expérimentalement pour de grands alphabets.

**Propriétés des partitions de Markov** Dans le cas Pisot unitaire, on dispose maintenant de critères pour obtenir des partitions de Markov. Cependant, on ne maîtrise ni la forme (propriétés topologiques) de ces partitions, ni la complexité du langage associé. Or, pour chaque automorphisme du tore Pisot, il existe plusieurs substitutions qui admettent cette matrice pour matrice d'incidence. Un domaine important de recherche consiste donc à proposer des critères de sélection d'une substitution pour une matrice donnée, et vérifier que pour toute matrice il existe une substitution qui vérifie ces critères. Dans cette direction, nous proposons au chapitre 3 (section 3.1) des critères pour contrôler la topologie des fractals.

**Partitions de Markov pour les endomorphismes du tore** Les résultats décrits ici concernent le cas Pisot unitaire. Cependant, les cas non unitaire et surtout non Pisot sont tout aussi intéressants. Cependant, dans ces cas là, le modèle d'une partition de Markov définie comme une suspension d'un fractal n'est plus valable.

Le cas *non unitaire Pisot* peut-être traité en ajoutant à la représentation des éléments  $p$ -adiques, nous discutons cette question dans le chapitre 3 (section 3.2). On peut alors définir un fractal et des critères pour que le système substitutif soit isomorphe à une translation sur un groupe compact. Obtenir une partition de Markov pour un endomorphisme du tore Pisot (et non inversible) est plus difficile, mais l'approche  $p$ -adique est prometteuse pour construire la base de la partition, comme discuté à la section 3.4.

**Partitions de Markov pour des automorphismes non positifs du tore** Les résultats obtenus ici concernent des automorphismes Pisot dont les coefficients sont

---

tous positifs. Pour étudier des automorphismes avec des coefficients négatifs, une piste consiste à considérer non plus des substitutions mais des *automorphismes du groupe libre*, c'est-à-dire des règles de transformations qui autorisent des annulations.

Par exemple, dans [AFHI07], les auteurs exploitent les constructions de fractals à l'aide d'approximations polygonales ainsi que les formalismes d'itérations définis dans [SAI01]. Cela permet de construire une partition de Markov pour un automorphisme du tore  $\mathbb{T}^4$  admettant deux directions dilatantes et deux directions contractantes, avec certains coefficients négatifs. La généralisation de cette approche et la définition de graphes de frontière pour les pièces de la partition de Markov obtenue dans [AFHI07] est le prochain travail à effectuer dans cette direction.

L'étude des automorphismes du groupe libre présente des intérêts plus généraux que la construction de partition de Markov. Dans cette direction, dans la section 3.4.3, nous discuterons de la dynamique symbolique pour les laminations attractives de classes d'automorphismes du groupe libre et ses applications pour la définition d'invariants à conjugaison près.

**Représentation des additions sur un tore** Dans le cas non algébrique, pour les additions, la principale question concerne la description de la dynamique symbolique des additions sur un tore, quelles qu'elles soient. Cela nécessite de trouver une partition du tore convenable vis-à-vis de l'addition. S'inspirant du cas unidimensionnel, on cherche à construire ce domaine fondamental par itération et renormalisations successives de différentes substitutions, reliées à un développement en fractions continues du vecteur de translation. On entre alors dans le monde des fractions continues multidimensionnelles et des substitutions généralisées, que nous aborderons dans le chapitre 3 (section 3.3).

Si la construction d'un tel domaine fondamental est réalisée, on peut espérer pouvoir en déduire des propriétés pour construire des partitions de Markov pour des automorphismes du tore moins spécifiques.



---

## Chapitre 3

# Utiliser les graphes de frontière

Dans ce chapitre, nous présentons différentes applications des systèmes dynamiques substitutifs dans différents domaines des mathématiques. Nous insistons en particulier sur le lien entre la topologie des fractals, qui peut-être déterminée à partir de dérivés des graphes de frontières, et des propriétés en théorie des nombres ou en géométrie. Nous discutons aussi des perspectives de recherche dans les différents domaines considérés.

La section 3.1 permet d'illustrer comment les graphes de frontière, introduits dans le chapitre 2 pour montrer des propriétés de partition de Markov, permettent en fait de vérifier un grand nombre de propriétés topologiques des fractals de Rauzy.

Dans la section 3.2, nous utiliserons ces propriétés topologiques de fractals pour mieux comprendre les développements en base beta des réels, en particulier les réels qui ont des développements fini ou purement périodiques). Dans le cas où la base  $\beta$  est un nombre non unitaire, nous verrons pourquoi il faut pour cela décrire explicitement la frontière de fractals avec des composantes  $p$ -adiques.

Dans la section 3.3, nous discuterons de la question de l'engendrement de plans discrets par des méthodes substitutives. Nous verrons que les propriétés topologiques des fractals ont alors des conséquences sur la forme des pièces engendrées. Nous discuterons aussi pourquoi cette question est importante pour répondre à la question de la représentation des additions sur un tore mentionnées dans les perspectives du chapitre 1.

Enfin, dans la section 3.4, nous discutons de l'intérêt des propriétés topologiques de fractals (avec de nouvelles constructions) en géométrie et en théorie ergodique, pour sortir du cadre Pisot dans la représentations des additions et des multiplications sur des tores.



## 3.1 Topologie des fractals de Rauzy

### 3.1.1 Où la topologie des fractals s'interprète de manières variées

Comme mentionné dans le premier chapitre d'introduction, le succès des fractals de Rauzy vient de leur utilisation dans une variété de domaines. En fait, dès qu'un objet mathématique est auto-induit (au sens où il se contient lui-même), et que le "ratio" d'induction est un nombre de Pisot, ce phénomène d'induction peut se traduire par une substitution et se représenter par un fractal de Rauzy. De manière moins évidente, les propriétés de l'objet mathématique considéré au départ se traduisent souvent en des propriétés topologiques des fractals.

- En théorie des nombres, des propriétés diophantiennes (recherche des meilleures approximations simultanées de vecteurs pour certains nombres) se déduisent de la largeur de la plus grande boule centrée en zéro et incluse dans le fractal.  
Par ailleurs, le fait que 0 est point intérieur du fractal signifie que les écritures de nombres dans une numération en base  $\beta$  sont toutes finies pour les éléments dans l'anneau  $\mathbb{R}[1/\beta]$  [AS05]. De plus, le fractal de Rauzy permet de caractériser les rationnels dont le développement est purement périodique [IR05]. Nous reviendrons longuement sur ces deux derniers points dans la section 3.2, en discutant des travaux (publié ou en cours de rédaction) en collaboration avec V. Berthé, S. Akiyama, G. Barat, B. Adamczewski, C. Frougny et W. Steiner ((Monas. Math., 2008),(JNT, 2007)).
- Comme nous l'avons aussi déjà mentionné, le fractal de Rauzy permet de construire une partition de Markov pour certains automorphismes d'un tore [IO93, Pra99]. La connexité du fractal est un élément important pour décider si cette partition de Markov est génératrice [Adl98].
- En géométrie discrète, les fractals de Rauzy sont très liés à des procédés d'engendrement de plans discrets [ABI02]. La forme des pièces générés lors de ce procédé est contrôlée par la topologie du fractal. Nous discuterons cette question en détails dans la section 3.3, en discutant des travaux en collaboration avec P. Arnoux, V. Berthé et J. Bourdon ((TCS,2004), (Jour. Montoises, 2006), (Integers, 2005)).
- En géométrie, les fractals de Rauzy permettent de représenter des flots pour des espaces de pavages. Certaines propriétés topologiques (comme l'existence de points de coupure) sont de bons candidats pour des invariants par conjugaison par un automorphisme de groupe libre : nous discuterons ce point dans la section 3.4, en s'appuyant sur des travaux (publiés ou en cours de rédaction) en collaboration avec P. Arnoux, V. Berthé et A. Hilion (Ann. Inst. Fourier, 2006).

Il existe plusieurs résultats disséminés dans la littérature concernant la topologie des fractals. Par exemple, le fractal de Rauzy pour la substitution de Tribonacci est connu pour être homéomorphe à un disque [Mes00], mais des choses assez surprenantes arrivent pour d'autres fractals : ils peuvent ne pas être connexes, ne pas

être simplement connexes, 0 peut ne pas être point intérieur [Aki02]. Les méthodes utilisées pour mettre en évidence des phénomènes sont très peu généralisables et concernent des exemples fixés.

### 3.1.2 Graphes de pavages et propriétés topologiques

Comme nous l'avons déjà mentionné, les travaux sur les graphes de frontière décrits dans la section 2.3.3 permettent, en plus de donner une condition de pavage pour les fractals de Rauzy, d'identifier explicitement les triplets et quadruplets de tuiles qui s'intersectent dans un pavage. Dans la même monographie en cours de soumission (Monographie S. & T., 2008), nous exploitons cette connaissance sur les points doubles, triples et quadruple pour donner des conditions algorithmiques pour de nombreuses propriétés de fractals. Les démonstrations sont assez variées, nous encourageons donc le lecteur à lire les détails dans (Monographie S. & T., 2008). Nous allons ici simplement rappeler les principales propriétés topologiques que nous caractérisons et discuter des apports des graphes de frontière par rapport aux résultats existant.

1. **Mesure de la dimension de la frontière** Des exemples de calculs de dimension de la frontière apparaissent dans [FFIW06, IK91, Mes00, Thu06]. Puisque les graphes de pavages fournissent une description par GIFS de la frontière du fractal, on en déduit directement une mesure de la dimension par boîtes (ou *box-dimension*) de cette frontière, qui dépend des valeurs propres de la matrice d'incidence du graphe de la frontière. La principale limitation de ces méthodes est qu'elles ne sont valables que dans le cas irréductible, puisqu'elles s'appuient sur un procédé d'approximation de la frontière par des polygones. Une autre limitation (sauf pour [Thu06]) est aussi l'absence d'algorithme pour calculer la dimension. Au contraire, nous proposons dans (Monographie S. & T., 2008) une méthode pour calculer la dimension par boîtes, qui est entièrement automatique puisqu'elle s'appuie sur un algorithme pour calculer les graphes de frontière. Cette méthode est utilisable pour toutes les substitutions Pisot unitaire, cas réductible compris.

Dans le cas où les conjugués de Galois de la valeur propre dominante de la substitution sont de même norme, on obtient un calcul de la dimension de Hausdorff de la frontière.

2. **Le point 0 est un point intérieur du fractal** Dans le cas des systèmes de numération, nous savons avec [Aki02] que 0 est un point intérieur du fractal de Rauzy si et seulement si les écritures des éléments de  $\mathbb{Z}[1/\beta]$  sont toutes finies. Avec une approche géométrique, nous prouvons dans (Monographie S. & T., 2008) que ce résultat est vrai pour toute substitution Pisot unitaire si on considère le système de numération de Dumont-Thomas associé ([DT93], voir aussi (Integers, 2005)).
3. **Connexité** En généralisant les critères donnés dans [Can03, Rau82], nous donnons dans (Monographie S. & T., 2008) une condition nécessaire et suffisante pour que les pièces du fractals soient connexes. Cette condition étant liée aux graphes de frontières, elle est vérifiable par l'application d'un algorithme.

4. **Homéomorphisme à un disque** Des exemples de fractals homéomorphes à un disque ont été étudiés dans [Luo02, Mes00, Mes06] : il s’agissait d’exemples où une seule tuile permet de décomposer le fractal. Dans (Monographie S. & T., 2008), nous proposons une condition en terme de graphes de frontière et de points triples qui est suffisante pour qu’un fractal de Rauzy soit homéomorphe à un disque, lorsque le coefficient de dilatation de la substitution est un nombre de Pisot de degré 3. Comme pour les précédentes propriétés topologiques, on vérifie cette condition de manière algorithmique.

Notre approche est générale : en fait, nous proposons un critère pour que la solution d’un GIFS soit une courbe de Jordan. Nous nous sommes inspirés d’une approche utilisée dans [LZ04] pour étudier des systèmes de numération dans l’anneau des entiers gaussiens. Cependant, dans notre cas, la présence de plusieurs pièces dans le GIFS empêche de déduire l’homéomorphisme à un disque à partir de la connexité de l’intérieur [LRT02]. Nous avons donc utilisé plusieurs résultats sur la topologie du plan pour aboutir à nos critères.

En guise d’exemple, nous détaillerons la condition d’homéomorphisme à un disque dans la section 3.1.3.

5. **Simple connexité et groupe fondamental non libre** Dans le cas où le fractal n’est pas homéomorphe à un disque, nous proposons dans (Monographie S. & T., 2008) un critère pour montrer que ce fractal contient un grand nombre de trous. Plus précisément, notre critère permet de montrer que le groupe fondamental de certains fractals non seulement n’est pas libre, mais n’est même pas dénombrable. Ainsi, topologiquement, certains fractals sont de la même nature que le tapis de Sierpinski. En particulier, puisque le groupe fondamental n’est pas dénombrable, l’intérieur de ces fractals n’est pas simplement connexe.

Tous les critères proposés s’appuient sur les graphes de frontière, et sur la combinatoire de la substitution. Ils sont donnés par des algorithmes qui ont été implémentés dans le programme de calcul formel MuPad, accessible sur demande.

En illustration, on pourra considérer la figure 3.1 qui propose différents exemples de fractals ; la croix blanche y désigne le point  $\mathbf{0}$ .

- $\sigma_1$  Cette substitution est définie par  $\sigma_1(1) = 12$ ,  $\sigma_1(2) = 3$ ,  $\sigma_1(3) = 4$ ,  $\sigma_1(4) = 5$ ,  $\sigma_1(5) = 1$ . Elle est réductible, sa valeur propre dominante vérifie  $\beta^3 = \beta - 1$ . Le fractal et toutes ses sous-tuiles sont homéomorphes à un disque et  $\mathbf{0}$  est un point intérieur.
- $\sigma_2$  Cette substitution est définie par  $\sigma_2(1) = 2$ ,  $\sigma_2(2) = 3$ ,  $\sigma_2(3) = 12$ . Elle est irréductible, son polynôme minimal est le même que celui de  $\sigma_1$ . Le fractal et toutes ses sous-tuiles sont connexes, mais leur groupe fondamental est non-dénombrable. Le point  $\mathbf{0}$  se trouve sur la frontière du fractal.
- $\sigma_3$  Cette substitution est définie par  $\sigma_3(1) = 3$ ,  $\sigma_3(2) = 23$ ,  $\sigma_3(3) = 31223$ . Son fractal n’est pas connexe et  $\mathbf{0}$  est sur la frontière du fractal.
- $\sigma_4$  Cette substitution est définie par  $\sigma_4(1) = 11112$ ,  $\sigma_4(2) = 11113$ ,  $\sigma_4(3) = 1$ . Elle est irréductible et le groupe fondamental de son fractal est non trivial (la

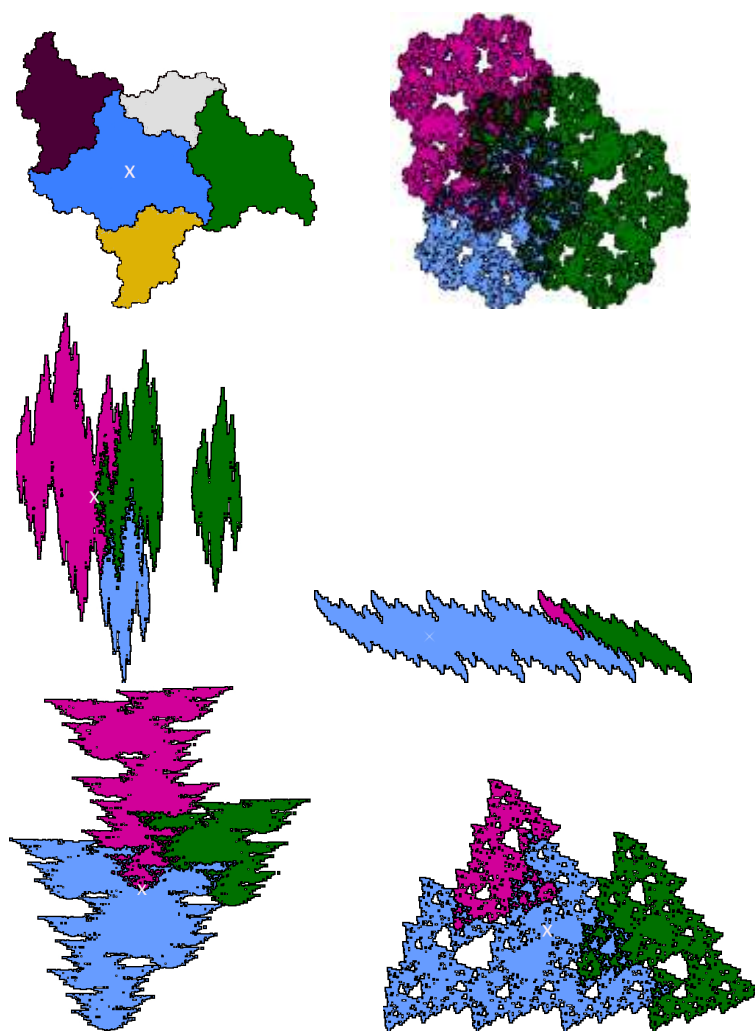


FIG. 3.1 – Différents fractals de Rauzy pour les substitutions  $\sigma_1$ ,  $\sigma_2$ ,  $\sigma_3$ ,  $\sigma_4$ ,  $\sigma_5$  et  $\sigma_6$  définies dans le texte. Les sous-tuiles sont indiquées par des couleurs distinctes. La croix blanche indique la position de l'origine  $\mathbf{0}$ . Les propriétés topologiques de ces fractals sont obtenues à partir des graphes de frontières, elles sont décrites dans le texte.

frontière n'est pas une courbe de Jordan). Par contre,  $\mathbf{0}$  est point intérieur du fractal.

$\sigma_5$  Cette substitution est définie par  $\sigma_5(1) = 123$ ,  $\sigma_5(2) = 1$ ,  $\sigma_5(3) = 31$ . Le groupe fondamental du fractal est non dénombrable.  $\mathbf{0}$  se trouve sur la frontière du fractal : même si cela est peu visible sur le dessin, il existe en fait un isthme très fin qui part de l'extérieur du fractal et se termine en  $\mathbf{0}$ .

$\sigma_6$  Cette substitution est définie par  $\sigma_6(1) = 12$ ,  $\sigma_6(2) = 31$ ,  $\sigma_6(3) = 1$ . Le groupe fondamental du fractal est non dénombrable.  $\mathbf{0}$  se trouve sur la frontière du fractal.

### 3.1.3 Exemple : homéomorphisme à un disque

Tous les détails concernant les résultats sur la topologie des fractals se trouvent dans (Monographie S. & T., 2008). Pour illustrer ces approches, nous donnons ici quelques détails sur un critère pour déterminer que la frontière du fractal de Rauzy est une courbe de Jordan. Cette propriété a été montrée pour le fractal associé à la substitution de Tribonacci dans [Mes00]; le critère que nous proposons ici est général.

Dire que la frontière d'un compact est une courbe de Jordan est équivalent à dire que le compact lui-même est homéomorphe à un disque. Pour vérifier cette propriété sur un fractal de Rauzy, on utilise les propriétés d'autosimilarité du bord : on sait que le bord se décompose de manière autosimilaire, en considérant le graphe de la frontière par exemple. De plus, chaque morceau du bord admet lui-même une décomposition. En considérant les points triples du fractal, on peut vérifier si, topologiquement, le bord peut se décomposer sous forme de cercle.

**Grphe de connexité** On définit ainsi le *graphe de connexité de la frontière* d'une sous-tuile  $\mathcal{T}(i)$  : ses sommets sont les ensembles de la forme  $\mathcal{T}(i) \cap (\mathcal{T}(j_2) + \gamma)$  qui contiennent au moins deux points et apparaissent dans le graphe de la frontière du fractal. On met une arête entre deux sommets (qui correspondent chacun à un morceau de frontière) si et seulement si l'intersection entre ces bouts de frontière est non vide.

Plus précisément, chaque sommet du graphe de connexité  $\mathcal{T}(i) \cap (\mathcal{T}(j_2) + \gamma)$  correspond au sommet  $[i, \gamma, j_2]$  du graphe de frontière. Pour calculer ces graphes, il faut d'abord calculer le graphe de la frontière du fractal. Il faut ensuite étudier ses chemins pour déterminer les sommets desquels partent au moins deux chemins distincts. On obtient ainsi les sommets du graphe de connexité. Ensuite, pour placer des arêtes, on doit considérer toutes les intersections de la forme  $(\mathcal{T}(i) \cap (\mathcal{T}(j_2) + \gamma)) \cap (\mathcal{T}(i') \cap (\mathcal{T}(j'_2) + \gamma'))$  et trouver celles qui sont non vides. Ceci correspond à calculer les graphes des points triples et des points quadruples du fractal, sur le même modèle que le graphe de frontière (voir (Monographie S. & T., 2008) pour plus détails).

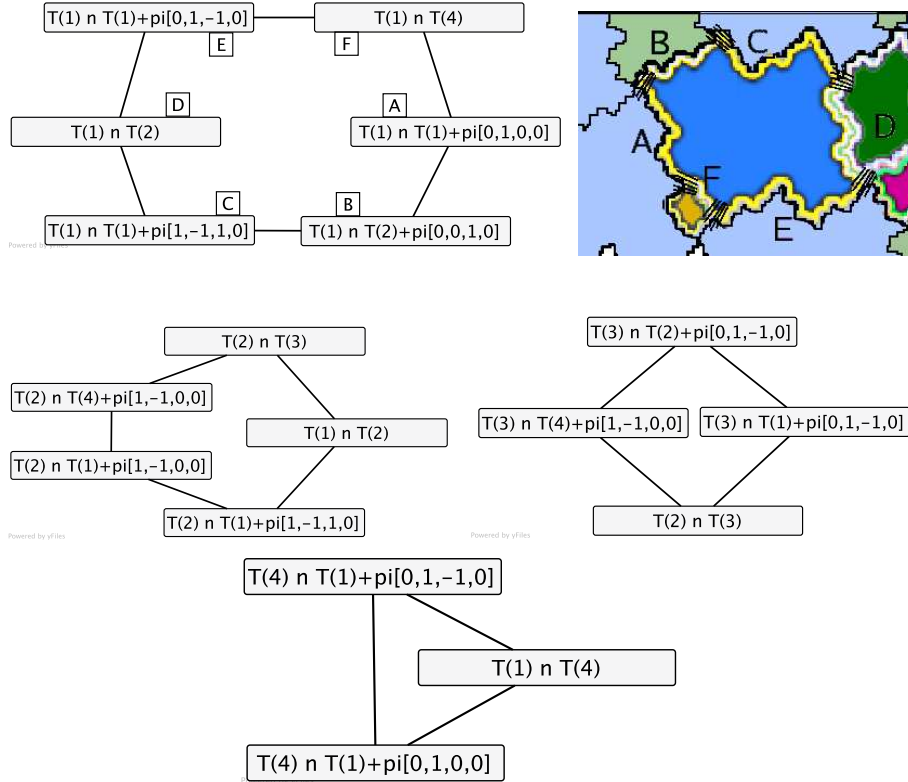


FIG. 3.2 – Les graphes de connectivité pour les sous-tuiles du fractal de la substitution  $\sigma_0$ . Chaque sommet des graphes correspond à une intersection de la forme  $\mathcal{T}(i) \cap \mathcal{T}(j) + \gamma$  qui apparaît dans la décomposition de la frontière de  $\mathcal{T}(i)$ . Un sommet entre deux arêtes signifie que les morceaux de la frontière correspondant s'intersectent.

**Exemple** Pour la substitution  $\sigma_0$ , on obtient quatre graphes de connectivité (un pour chaque sous-tuile) décrits dans la figure 3.2. Dans le haut de la figure, on voit que le graphe de la frontière de  $\mathcal{T}(1)$  se décompose en six morceaux, chacun étant adjacent à deux autres morceaux. Cela s'illustre bien sur la représentation de  $\mathcal{T}(1)$  en haut à droite.

**Homéomorphisme à un disque** On peut construire un graphe de connectivité non seulement pour les sous-tuiles  $\mathcal{T}(i)$  mais aussi pour tous les éléments qui apparaissent dans la décomposition autosimilaire de  $\mathcal{T}(i)$ . Une condition suffisante pour que le fractal soit un homéomorphe à un disque est alors que ces graphes de connectivité aient une structure topologique simple : on attend que, en première approximation, le bord soit un polygone (c'est-à-dire que son graphe de connectivité soit une boucle). Ensuite, on impose que tous les morceaux du polygone se dé-

composent sous forme de lignes brisées (leurs graphes de connexité doivent être des lignes). En appliquant le processus d'autosimilarité, on va ainsi construire des lignes polygonales qui convergent vers le bord du fractal en maîtrisant le processus de décomposition, et obtenir finalement le bord d'un disque.

**Théorème 3.1.1 (Monographie S. & T., 2008)** *Soit  $\sigma$  une substitution primitive Pisot unitaire dont la valeur propre dominante est de degré 3. On suppose que le recouvrement auto-similaire est un pavage.*

*Alors chaque sous-tuile  $\mathcal{T}(i)$  ( $i \in \mathcal{A}$ ) est homéomorphe à un disque si les conditions suivantes sont vérifiées.*

1. *Tous les graphes de connexité des sous-tuiles  $\mathcal{T}(i)$  sont des boucles.*
2. *Tous les graphes de connexité des éléments qui apparaissent dans la décomposition des sous-tuiles  $\mathcal{T}(i)$  sont des lignes ou réduits à un noeud.*
3. *Les intersections entre trois tuiles du pavage autosimilaires sont vides ou réduites à un point.*

En appliquant ce théorème, on en déduit ainsi que toutes les sous-tuiles du fractal de Rauzy de  $\sigma_0$  sont homéomorphes à un disque. On en déduit que le fractal lui-même est homéomorphe à un disque en considérant la topologie des intersections entre les sous-tuiles.

### 3.1.4 Perspectives d'étude

En conclusion de cette partie, nous pouvons dire que nous disposons maintenant d'un outil puissant (les graphes de frontière et leurs dérivés) pour caractériser les propriétés topologiques de fractals de Rauzy. Avant de passer aux applications de ces propriétés topologiques, il faut mentionner que de nombreux travaux restent à développer sur la topologie des fractals et la compréhension des graphes de frontière.

**Intérieur du fractal de Rauzy** La principale question concerne l'intérieur du fractal de Rauzy. Dans les exemples mentionnés, il apparaît que cet est non connexe. Une question naturelle porte sur la structure des composantes connexes de cet intérieur, ou plutôt leurs adhérences : sont-elles autosimilaires ? Sont-elles homéomorphes à un disque ? Des questions semblables ont été traitées dans le cas autosimilaire avec une seule tuile dans la décomposition [LT06, NN03b, NT04]. Le fait qu'il y ait ici plusieurs tuiles rend les choses bien plus difficiles. Une approche peut-être tentée par le théorème de Torhorst [Kur68, §61, II, Theorem 4], qui relie la question de l'homéomorphisme entre un disque et l'intérieur des composantes connexes à l'existence de points de coupure pour la tuile. Une première étape consiste donc à rechercher des critères pour l'existence de points de coupure, qui est reliée à l'étude d'invariants pour les automorphismes de groupes libres (voir section 3.4).

**Groupe fondamental** Nous avons exhibé des compacts du plan euclidien dont le groupe fondamental est non libre et non dénombrable, ce qui est pathologique d'un point de vue topologique. Il serait intéressant de décrire la structure du groupe fondamental.

Comme déjà mentionné, le fait que le groupe fondamental est non dénombrable implique que l'intérieur du fractal n'est pas simplement connexe. Or, des exemples de structures qui ne sont pas simplement connexes existent (tapis de Sierpinski, Hawaiian Earring) et leur groupe fondamental est décrit par des mots [CL08, EK98]. Cependant, les méthodes ne sont pas tout à fait transposables ici dans la mesure où les exemples qui ont été considérés jusqu'à maintenant étaient de dimension topologique 1 alors que nos fractals sont de dimension topologique 2.

**Dimension supérieure à 3** Une dernière direction est la généralisation des critères sur les homéomorphismes à un disque au cas où la constante de dilatation est de degré algébrique supérieur à trois. Nous nous sommes restreints à la dimension 3 dans la mesure où nos arguments pour étudier la simple connexité sont basés sur des propriétés de séparation du plan, tel que le théorème de Jordan. Ces résultats ne sont plus valables en dimension supérieure. Par contre, les graphes de frontière et les propriétés de pavage sont toujours valables. Pour obtenir des critères de simple connexité, on pourrait s'appuyer par exemple sur le théorème de Moore-van Kampen-Zippin [vK35], qui donne une caractérisation de la sphère  $\mathbb{S}^2$  en fonction d'ensembles de coupure. L'objectif serait de trouver un exemple de fractal dont la frontière serait homéomorphe à une sphère, par exemple en étudiant la substitution qui généralise Tribonacci  $\sigma(1) = 12, \sigma(2) = 13, \sigma(3) = 14, \sigma(4) = 1$ .

## 3.2 Théorie des nombres. Développement en bases non-entières : le cas non unitaire

**Les substitutions en théorie des nombres** Depuis leurs premières utilisations, les substitutions ont toujours été exploitées en théorie des nombres. Par exemple, la suite de Thue-Morse, point fixe de substitution, permet de classer les entiers en fonction de la parité de la somme de chiffres dans leur développement binaire. La suite de Baum-Sweet caractérise les entiers dont le développement binaire contient une plage impaire de zéro ; il s'agit d'une projection lettre à lettre du point fixe d'une substitution [AS02].

Un autre pont entre la théorie des nombres et les substitutions est le théorème de Cobham [Cob72], qui établit qu'une suite est la projection lettre à lettre du point fixe d'une substitution de longueur constante  $k$  si et seulement si les éléments  $u_n$  de cette suite sont obtenus en alimentant un  $k$ -automate fini avec le développement en base  $k$  de  $n$ .

Les substitutions sont aussi utilisées pour proposer des résultats de transcendance : les nombres réels dont le développement en fraction continue est donné par la suite de Thue-Morse, la suite de Baum-Sweet ou la suite de Rudin-Shapiro sont tous des nombres transcendants, la preuve utilisant la structure substitutive de ces



suites [ABD06]. Dans le même esprit, les nombres irrationnels dont le développement binaire est le point fixe d'une substitution sont transcendants [AB07].

En approximation diophantienne enfin, les substitutions ont montré leur utilité : leurs points fixes produisent des nombres transcendants qui sont très mal approximables par des entiers algébriques cubiques [Roy03, HM06].

**Beta-numérations et beta-substitutions** Enfin, comme nous l'avons vu, les substitutions sont intimement liées aux numérations en base non-entière puisque, dans le cas Pisot, les points dont le beta-développement est entier sont décrits par le point fixe d'une substitution.

Cette substitution, appelée *beta-substitution*, est définie entièrement par le développement  $d_\beta(1)$  de 1 en base  $\beta$ . Si  $d_\beta(1) = t_1 \dots t_{m-1} t_m 0^\infty$  est fini, on construit un développement infini sous la forme  $d_\beta^*(1) = (t_1 \dots t_{m-1} (t_m - 1))^\infty$  ; s'il est infini, on pose  $d_\beta^*(1) = d_\beta(1)$ . Dans tous les cas, on sait que  $d_\beta^*(1)$  est ultimement périodique, on le note alors  $d_\beta^*(1) = (a_1 \dots a_m)(a_{m+1} \dots a_n)^\infty$ . La beta-substitution est alors définie sur l'alphabet  $\{1, \dots, n\}$  par  $\sigma(k) = 1^{a_k}(k+1)$  for each  $k < n$  et  $\sigma(n) = 1^{a_n}(m+1)$ .

Par exemple, pour le nombre  $\beta = 5 + 2\sqrt{7}$ , le polynôme annulateur est  $X^2 - 10X - 3 = 0$ , ce qui implique que 1 s'écrit  $1 = 10/\beta + 3/\beta^2$ . Son développement est donc  $d_\beta(1) = (10)30^\infty$ . En utilisant des réécritures, on a aussi  $1 = 10/\beta + 2/\beta^2 + 10/\beta^3 + 2/\beta^4 + \dots$ . Le développement impropre de 1 est donc  $d_\beta^*(1) = ((10)2)^\infty$  et la substitution qui décrit l'arrangement des points entiers en beta-numération est définie par  $\sigma(1) = 1^{10}2$ ,  $\sigma(2) = 11$ .

Nous allons maintenant voir comment l'étude du fractal de Rauzy de cette substitution, en particulier de sa topologie, permet de prouver des propriétés substantielles des développements en base  $\beta$ , notamment dans le cas non unitaire.

### 3.2.1 Fractals dans le cas non unitaire

**Beta-numérations et dynamique** Nous avons déjà longuement discuté de l'intérêt des fractals de Rauzy pour construire une partition de Markov pour la dynamique correspondant à la multiplication par une matrice sur un tore. Du point de vue des beta-numérations, une autre dynamique est tout aussi importante : il s'agit de la beta-transformation  $T_\beta$ . Par définition, on part de cette application pour en coder les orbites et définir le développement en base  $\beta$  des réels positifs. Un des défauts de cette application est qu'elle n'est pas injective, puisque l'image de l'intervalle  $[0, 1[$  est une copie de  $[\beta]$  intervalles. Pour mieux l'étudier, on recherche donc une *extension naturelle* de cette application, c'est-à-dire une application sur un espace plus large qui est bijective et qui admet  $T_\beta$  pour facteur topologique.

Un procédé habituel et général pour construire des extensions naturelles consiste à construire des tours de Rokhlin, et à en étudier la structure [DKS96]. Cependant, ces extensions naturelles ne préservent pas les propriétés algébriques de  $\mathbb{Q}(\beta)$ . Dans le cas Pisot unitaire, cependant, on peut construire une extension naturelle pour  $T_\beta$  comme produit de la  $\beta$ -transformation  $T_\beta$  et d'une contraction diagonale sur le fractal de Rauzy (en fait, cette construction est très proche de la partition de Markov

discutée au chapitre 2) [IR06]. En particulier, l'image de l'anneau  $\mathbb{Z}[\beta]$  dans cette extension naturelle est un réseau. Cela permet par exemple de caractériser les réels dont le  $\beta$ -développement est purement périodique (voir section 3.2.3).

### Pourquoi le fractal usuel n'est pas convenable dans le cas non-unitaire

Dans la situation où  $\beta$  est un nombre de Pisot *non unitaire*, cependant, ces propriétés d'extension naturelle ne sont plus valables. En effet, l'extension basée sur le fractal de Rauzy n'est plus une application qui préserve la mesure : cette extension est le produit de la  $\beta$ -transformation  $T_\beta$  et d'une contraction diagonale sur le fractal de Rauzy ; la  $\beta$ -transformation  $T_\beta$  est dilatante, de rapport  $\beta$ , et la contraction sur le fractal de Rauzy a pour rapport le produit des normes des conjugués de  $\beta$ , c'est-à-dire  $N(\beta)/\beta$ . Ainsi, l'extension est dilatante, de rapport  $N(\beta)$ , et il ne peut s'agir d'une extension naturelle.

Ces arguments de mesure impliquent en particulier que les propriétés de pavages autosimilaire ou périodique ne peuvent pas être vérifiées par un fractal de Rauzy dans le cas non unitaire.

**Espace complet de représentation** Dans la publication (Erg. Th. Dyn. Sys, 2003), j'ai proposé de rajouter au fractal de Rauzy des composantes basées sur des corps locaux. La question était alors de comprendre le spectre des systèmes engendrés par une substitution non unitaire (nous reviendrons sur ce sujet dans la section 3.4), mais j'ai ensuite utilisé le formalisme décrit dans cette publication pour étudier les systèmes de numération en base non entière.

La principale idée que j'ai proposée est d'étendre l'espace de représentation pour les fractals. On considère l'anneau des entiers  $\mathfrak{O}$  de  $\mathbb{Q}(\beta)$ , et on décompose l'idéal  $(\beta)$  en idéaux premiers dans cet anneau :  $(\beta) = \beta\mathfrak{O} = \prod_{i=1}^{\nu} \mathfrak{P}_i^{n_i}$ . Chaque idéal premier  $\mathfrak{P}$  ainsi obtenu définit une topologie sur le corps  $\mathbb{Q}(\beta)$  ; sa complétion produit un corps noté  $\mathbb{K}_{\mathfrak{P}}$ . Ce corps supporte une valeur absolue  $|x|_{\mathfrak{P}} = |N_{\mathbb{K}_{\mathfrak{P}}/\mathbb{Q}_p}(y)|_p^{1/e(\mathfrak{P})f(\mathfrak{P})} = p^{-f(\mathfrak{P})v_{\mathfrak{P}}(y)}$ , où  $p$  est défini par  $\mathfrak{P} \cap \mathbb{Z} = p\mathbb{Z}$ . Ce corps supporte aussi une mesure de Haar définie par  $\mu_{\mathfrak{P}}(a + \mathfrak{p}_{\mathfrak{P}}^m) = p^{mf(\mathfrak{P})}$ .

L'espace complet de représentation  $\mathbb{K}_{\beta}$  est obtenu en ajoutant à l'espace de représentation euclidien  $\mathbb{R}^{r-1} \times \mathbb{C}^s$  le produit des corps locaux  $\mathbb{K}_{\mathfrak{P}_i}$ . Le corps  $\mathbb{Q}(\beta)$  s'injecte naturelle dans cet espace en étendant l'application  $\phi$  définie dans le cas unitaire en (2.5) :

$$\phi_{\beta}: x = P(\beta) \in \mathbb{Q}(\beta) \longrightarrow \phi_{\beta}(x) = (P(\beta_2), \dots, P(\beta_{r+s}), x, \dots, x) \in \mathbb{K}_{\infty} \times \prod_{i=1}^{\nu} \mathbb{K}_{\mathfrak{P}_i}. \quad (3.1)$$

L'espace  $\mathbb{K}_{\beta}$  est naturellement muni de la topologie produit, et des additions et multiplications sur chaque coordonnée. Il s'agit donc d'un anneau abélien localement compact. De plus, une mesure de Haar normalisée  $\mu_{\beta}$  est obtenue comme le produit des mesures sur chaque composante. On en déduit en particulier que pour tout borélien  $B$  de  $\mathbb{K}_{\beta}$ , on a

$$\mu_{\beta}(\beta \cdot B) = \beta^{-1} \mu_{\beta}(B).$$

**Fractal complet** À partir de l'injection  $\phi_\beta$ , on définit naturellement le fractal de Rauzy complet, intégrant des composantes sur des corps locaux.

$$\tilde{\mathcal{T}} = \overline{\phi_\beta(\text{Int}(\beta))} \subset \mathbb{K}_\beta$$

Pour différencier ce fractal du fractal construit au chapitre 2, le fractal de Rauzy usuel est appelé *fractal euclidien*.

**Justification a posteriori** Il est nécessaire de s'attarder sur les raisons de ce choix de représentation. Pour obtenir des propriétés de GIFS, de séparation des sous-tuiles ou de pavage, il faut que l'espace de représentation vérifie plusieurs propriétés

- les tuiles doivent être suffisamment larges pour couvrir l'ensemble de l'espace considéré, lorsqu'elles sont déplacées selon l'ensemble de points autosimilaire produit par l'approximation de plan.
- les tuiles doivent être suffisamment petites pour ne pas produire de recouvrement, à la fois combinatoire (dans le recouvrement), topologique (intérieur disjoint), ou, de manière plus forte, métrique (tuiles disjointes en mesure).

Lorsque  $\beta$  n'est pas unitaire, considérer seulement les représentation archimédiennes de  $\mathbb{Q}(\beta)$  produit des tuiles trop grosses pour respecter ces propriétés. Considérer seulement les places locales aurait produit des propriétés topologiques et algébriques intéressantes, mais cette fois-ci les tuiles auraient été trop petites, puisque les adèles principales forment un ensemble discret du groupe des adèles. Avec le choix proposé, nous nous trouvons entre ces deux extrêmes.

**Propriétés de la tuile complète** En collaboration avec Valérie Berthé, j'ai montré dans (JNT, 2007) que cet espace de représentation est effectivement le bon espace pour étudier les propriétés de la  $\beta$ -numération.

- Le fractal de Rauzy complet vérifie une relation d'autosimilarité dirigée par le graphe des préfixes-suffixes

$$\tilde{\mathcal{T}}(a) = \bigcup_{a \xrightarrow{(p,a,s)} b} \left( \phi_\beta(\beta)\tilde{\mathcal{T}}(b) + \phi_\beta(p) \right).$$

- Les sous-tuiles  $\tilde{\mathcal{T}}^{(a)}$  sont disjointes en mesure dans l'espace complet de représentation, de mesure non nulle et elles sont l'adhérence de leurs intérieurs.
- On obtient un recouvrement de l'espace en considérant les translatés des sous-tuiles selon un ensemble autosimilaire et localement fini (ici, l'ensemble des paires  $(x, a)$  telles que  $x \in \mathbb{Z}[1/\beta] \cap [0, T_\beta^{(a-1)}(1))$ ) :

$$\mathbb{K}_\beta = \bigcup_{x \in \mathbb{Z}[1/\beta] \cap [0, T_\beta^{(a-1)}(1))} \tilde{\mathcal{T}}(a) + \phi_\beta(x). \quad (3.2)$$

Une des questions qui vont nous intéresser va être de donner des conditions explicites pour que ce recouvrement de  $\mathbb{K}_\beta$  soit un pavage.

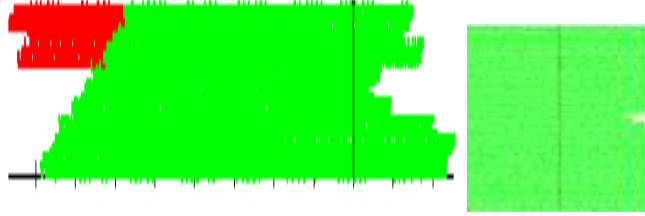


FIG. 3.3 – Représentation du fractal de Rauzy complet (incluant les composantes  $p$ -adiques) pour la substitution associée au nombre  $\beta = 5 + 2\sqrt{7}$ . le polynôme minimal de  $\beta$  est alors  $X^2 - 10X - 3$ . L'espace de représentation est alors  $\mathbb{R} \times \mathbb{Z}_3$ , où  $\mathbb{Z}_3$  désigne l'ensemble des entiers 3-adiques. L'axe vertical contient une représentation de  $\mathbb{Z}_3$  plongé dans l'intervalle  $[0, 1)$ . L'axe horizontal correspond à l'axe réel. Puisque le développement de 1 en base  $\beta$  est fini de longueur 2, il y a deux sous-tuiles pour le fractal.

À droite, on a procédé à un zoom autour de l'axe vertical  $x = 0$ . Il semble ainsi que le fractal contiennent une bande de la forme  $[0, \varepsilon] \times \mathbb{Z}_3$ . Nous vérifions cela formellement à l'aide de graphes de frontière et en déduisons des propriétés des développements purement périodiques.

**Exemple** Pour représenter ces fractals, on utilise un plongement de l'anneau des entiers  $\mathbb{Z}_p$  dans  $\mathbb{R}$  défini comme suit :  $x = \sum a_i p^i \in \mathbb{Z}_p \mapsto \sum a_i p^{-i} \in [0, 1]$ . Ce plongement est continu, il respecte la mesure mais pas la topologie de  $\mathbb{Z}_p$ . Deux exemples sont donnés dans les figures 3.3 et 3.4.

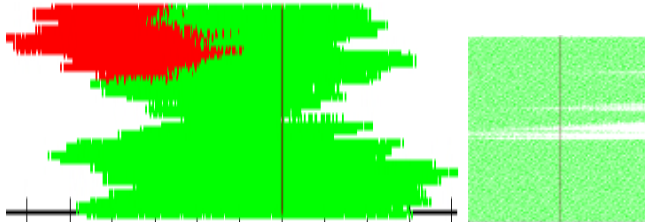


FIG. 3.4 – Fractal de Rauzy pour la substitution associée au nombre  $\beta^2 = 4\beta + 3$ . Comme dans la figure précédente, le fractal se représente dans  $\mathbb{R} \times \mathbb{Z}_3$ . Par contre, le zoom autour de l'axe vertical suggère que le fractal ne contient aucune bande de la forme  $[0, \varepsilon] \times \mathbb{Z}_3$ .

### 3.2.2 Nombres dont le $\beta$ -développement est fini

**La propriété (F)** Dans le cadre de l'étude des développements en  $\beta$ -numération, une question naturelle a été d'identifier les éléments dont le  $\beta$ -développement est

fini. Par construction, si un nombre  $a$  a un  $\beta$ -développement fini, alors il appartient à l'ensemble  $\mathbb{Z}[1/\beta] \cap \mathbb{R}^+$ . La question qui se pose est la réciproque : quels sont les nombres  $\beta$  pour lesquels tous les éléments de  $\mathbb{Z}[1/\beta] \cap \mathbb{R}^+$  ont un développement fini ? Si cette propriété est vraie, on dit que  $\beta$  vérifie une propriété de finitude, ou *propriété (F)*.

**Familles qui vérifient cette propriété** Dans [FS92], les auteurs montrent que dès que  $\beta$  est une racine dominante d'un polynôme du type  $X^n = a_{n-1}X^{n-1} + \dots + a_0$ , avec  $a_{n-1} \geq a_{n-2} \geq \dots \geq a_0 \geq 1$ , alors  $\beta$  vérifie la propriété (F). Si on considère la  $\beta$ -substitution engendrée par  $\beta$ , lorsque le polynôme  $X^n - a_{n-1}X^{n-1} + \dots - a_0$  est irréductible, le système dynamique symbolique est à spectre purement discret : il est mesurablement isomorphe à une addition sur un groupe compact [Sol92].

Une caractérisation des nombres de Pisot de degré 2 ou 3 qui vérifient (F) est explicitement donnée dans [Aki00].

**Interprétation géométrique** Dans le cas unitaire, la propriété (F) a des conséquences en termes de pavages par les tuiles du fractal de Rauzy. Selon [Aki02], cette propriété est équivalente au fait que  $\mathbf{0}$  est un point intérieur du fractal de Rauzy. Cela implique que le recouvrement autosimilaire associé au fractal est un pavage ; et on retrouve ainsi le fait que dans le cas où le polynôme à coefficients décroissants est le polynôme minimal de  $\beta$ , le système substitutif associé est à spectre purement discret.

**Extension au cas substitutif unitaire ; preuve géométrique** Dans (Integers, 2005), Valérie Berthé et moi-même proposons une généralisation de la propriété (F) dans le cas substitutif. Considérons une substitution  $\sigma$  Pisot. Comme à la section 2.2.5, on désigne par  $\mathbf{v}_\beta$  un vecteur propre dominant de la transposée de la matrice d'incidence de  $\sigma$ . Cependant, on normalise cette fois-ci le vecteur  $\mathbf{v}_\beta$  pour que son premier coefficient soit égal à 1. Les coefficients de ce vecteur propre sont donnés par les produits scalaires  $\langle \mathbf{v}_\beta, \mathbf{e}_k \rangle$  avec les éléments de la base canonique.

**Définition 3.2.1 ((Integers, 2005))** *On dit que la substitution vérifie la propriété (F) si et seulement pour tout  $x$  engendré par les coordonnées  $\langle \mathbf{v}_\beta, \mathbf{e}_k \rangle$ , le développement de Dumont-Thomas  $d_{(\sigma,a)}(x)$  de  $x$  est fini (voir Théorème 2.3.2).*

$$\begin{aligned} \forall a \in \mathcal{A} = \{1 \dots n\}, \\ \forall z \in \mathbb{Z}[\langle \mathbf{v}_\beta, \mathbf{e}_k \rangle, \dots, \langle \mathbf{v}_\beta, \mathbf{e}_k \rangle] \cap [0, \langle \mathbf{v}_\beta, \mathbf{e}_a \rangle), \quad d_{(\sigma,a)}(z) \text{ est fini.} \end{aligned} \quad (3.3)$$

Dans le cas des  $\beta$ -substitutions, on a  $\delta_\sigma(k) = 1/\beta^k$ , permettant de retrouver la propriété (F).

Dans (Integers, 2005) nous montrons que si cette propriété est vérifiée avec  $\beta$  unitaire, le recouvrement autosimilaire associé à la substitution est bien un pavage, généralisant ainsi les résultats de [Aki02].

De plus, dans (Monographie S. & T., 2008), J. Thuswaldner et moi-même montrons que dans le cas unitaire, cette propriété est équivalente au fait que  $\mathbf{0}$  est un

point intérieur du fractal de Rauzy. On obtient ainsi un résultat plus général que [Aki02] puisque notre preuve, basée sur la géométrie des plans discrets, englobe toutes les substitutions unitaires Pisot (y compris les substitutions réductibles). De plus, on en déduit un critère algorithmique pour vérifier cette propriété, inspiré par les graphes de frontière.

**Théorème 3.2.2 (Monographie S. & T., 2008)** *Une substitution  $\sigma$  primitive Pisot unitaire vérifie la condition de finitude définie en (3.3) si et seulement si  $\mathbf{0}$  est un point intérieur du fractal de Rauzy. Cette propriété est vérifiable au moyen de graphes de frontière.*

**Le cas non unitaire** Les preuves de (Monographie S. & T., 2008) et [Aki02] sont valables seulement dans le cas unitaire : dans le cas non-unitaire, le fractal de Rauzy euclidien ne vérifie pas de propriété de pavage, puisque l'ensemble produit par approximation de plan n'est pas un ensemble localement fini ; pour la même raison, ses sous-tuiles ne sont pas disjointes. Ceci empêche d'utiliser les approches géométriques ou arithmétiques pour relier la propriété (F) au fait que  $\mathbf{0}$  est point intérieur du fractal.

Dans ce cadre, le fractal de Rauzy complet (intégrant des places non-archimédiennes) est d'une large utilité. En effet, nous prouvons que sous la condition (F), les sous-tuiles sont disjointes et des propriétés de pavages sont vérifiées. Dans (Monas. Math., 2008), en collaboration avec Shigeki Akiyama, Valérie Berthé et Guy Barat, nous exploitons ces propriétés de pavages dans l'espace complet de représentation pour en déduire que si  $\beta$  vérifie la propriété (F), alors  $\mathbf{0}$  est point intérieur du fractal de Rauzy complet et du fractal de Rauzy euclidien.

**Théorème 3.2.3 (Monas. Math., 2008)** *Soit  $\beta$  un nombre de Pisot. Si  $\beta$  vérifie la propriété (F), alors toutes les tuiles de l'espace complet (incluant les places locales) décrites dans (3.2) sont d'intérieurs disjoints et leur frontière est de mesure nulle. De plus,  $\mathbf{0}$  est un point intérieur du fractal de Rauzy complet.*

On projette ce résultat sur l'espace de représentation euclidien pour obtenir une généralisation de [Aki02] dans le cas non unitaire.

**Corollaire 3.2.4** *Soit  $\beta$  un nombre de Pisot. Si  $\beta$  vérifie la propriété (F), alors  $\mathbf{0}$  est un point intérieur du fractal de Rauzy euclidien  $\mathcal{T}$ .*

On ne sait pas si la réciproque du corollaire 3.2.4 est vraie :  $\mathbf{0}$  point intérieur du fractal euclidien implique-t-il que la propriété (F) est vérifiée ?

### 3.2.3 Nombres dont le développement est périodique

**Développements ultimement périodiques** Depuis l'introduction des  $\beta$ -développements par Parry [Par60] une question aussi naturelle que la finitude des développements a été d'identifier les nombres admettant des propriétés de périodicité. Pour les bases entières, on sait depuis le lycée que les nombres dont le développement

en base  $b$  est ultimement périodique sont les nombres rationnels. Plus généralement, lorsque  $\beta$  est un nombre de Pisot, les nombres dont le  $\beta$ -développement est ultimement périodique sont tous les éléments de  $\mathbb{Q}(\beta)$  [Ber77, Sch80].

**Rationnels dont le développement est purement périodique : bases entières ou quadratiques** Le cas des développements purement périodiques est plus complexe. Lorsque  $b$  est un entier, on sait depuis longtemps que les nombres rationnels  $p/q$  avec un développement périodique en base  $b$  sont exactement ceux pour lesquels  $q$  est premier avec la base  $b$ .

Avec des méthodes d'approximation et de renormalisation, Schmidt a prouvé dans [Sch80] que lorsque  $\beta^2 = n\beta + 1$  avec  $n \in \mathbb{N}^*$ , tous les nombres rationnels inférieurs à 1 ont un  $\beta$ -développement purement périodique. Dans [HI97], il a été montré que dans le cas où  $\beta^2 = n\beta - 1$ ,  $n \geq 3$ , aucun nombre rationnel n'admet un développement purement périodique. Pour cela, les auteurs construisent l'extension naturelle de la  $\beta$ -transformation à partir du fractal du Rauzy et l'utilisent pour comprendre les développements purement périodiques.

Ce résultat "négatif" sur les développements purement périodiques se généralise assez bien aux cas cubiques et autres : dès que  $\beta$  admet un conjugué de Galois dans  $[0, 1[$ , aucun rationnel ne peut avoir un développement purement périodique [Aki98].

**Base non-entière unitaire** Pour comprendre le cas non quadratique (au moins cubique), Akiyama a montré que dès que la propriété (F) est vérifiée, il existe un intervalle non vide de la forme  $[0, \epsilon[$  dont tous les rationnels ont un  $\beta$ -développement purement périodique [Aki98]. Ceci est assez inattendu, dans la mesure où on attend que la répartition des développements périodiques en base  $\beta$  soit *aléatoire*, en tout cas indépendante de la position par rapport à 0.

Ce résultat est moins surprenant lorsqu'on considère le point de vue du fractal de Rauzy. En effet, Ito et Rao ont proposé d'utiliser le fractal de Rauzy et ses propriétés d'extension naturelle [IR05]. Puisque l'extension naturelle respecte la structure algébrique de  $\mathbb{Z}[\beta]$  dans le cas unitaire, on en déduit qu'un élément  $x$  de  $\mathbb{Q}(\beta)$  a un développement purement périodique si et seulement s'il existe une lettre  $a$  telle que

- $x < T^{(a)}(1)$ ,
- La tuile  $\mathcal{T}(a)$  du fractal euclidien contient le point du plan contractant dont les coordonnées sont données par les conjugués de Galois de  $-x$ .

Si on suppose que la propriété (F) est vérifiée, on sait que  $\mathbf{0}$  est point intérieur du fractal, en particulier d'une de ses tuiles  $\mathcal{T}(a)$ . Cette tuile  $\mathcal{T}(a)$  contient donc un voisinage de  $\mathbf{0}$ . Si on considère maintenant un rationnel  $x$  suffisamment petit, tous ses conjugués de Galois sont égaux à  $x$  et sont donc petits. Tout point ayant  $(-x, \dots, -x)$  pour coordonnées dans l'espace contractant sera donc près de  $\mathbf{0}$  et appartiendra ainsi à  $\mathcal{T}(a)$ , ce qui implique que le développement de  $x$  est purement périodique.

**Bases non-unitaires** Dans le cas non-unitaire, comme nous l'avons déjà discuté, les propriétés d'extension naturelle de la tuile euclidienne ne sont plus vérifiées. Dans (JNT, 2007), avec Valérie Berthé, nous utilisons le fractal complet pour caractériser les points au développement purement périodique dans le cas non-unitaire. Nous généralisons le résultat de [IR05] au cas non-unitaire, en remplaçant le fractal de Rauzy euclidien par le fractal complet. Nous donnons aussi une preuve plus simple que celle de [IR05] et qui fait ressortir la structure d'extension naturelle proposée par le fractal de Rauzy.

**Théorème 3.2.5 (JNT, 2007)** *Soit  $\beta$  un nombre de Pisot et  $\mathcal{T}$  son fractal de Rauzy complet (incluant les places locales lorsque  $\beta$  est non-unitaire). Soit  $\phi_\beta$  le plongement de  $\mathbb{Q}(\beta)$  dans  $\mathbb{K}_\beta$  défini en (3.1). Soit  $x \in [0, 1] \cap \mathbb{Q}(\beta)$ . Alors  $x$  a un  $\beta$ -développement purement périodique si et seulement si*

$$(-\phi_\beta(x), x) \in \bigcup_{a \in \mathcal{A}} (\tilde{\mathcal{T}}^{(a)} \times [0, T_\beta^{a-1}(1))). \quad (3.4)$$

**Plus petit rationnel qui admet un développement non périodique** On peut utiliser ce résultat pour rechercher précisément le plus grand intervalle de rationnels de la forme  $[0, \epsilon[ \cap \mathbb{Q}$  qui ne contient que des points périodiques. Ce nombre est appelé  $\gamma(\beta)$ .

$$\gamma(\beta) = \sup\{\varepsilon, \forall \frac{p}{q} < \varepsilon, d_\beta(\frac{p}{q}) \text{ est purement périodique}\}.$$

On obtient un encadrement de cette quantité en considérant l'intersection du fractal de Rauzy avec une droite ; il s'agit d'un résultat que j'ai montré dans (Monas. Math., 2008) en collaboration avec Shigeki Akiyama, Valérie Berthé et Guy Barat. Le résultat exact que nous obtenons s'écrit comme suit dans le cas unitaire. Sa généralisation au cas non unitaire demande d'introduire un certain nombre de notations liées aux corps locaux. Nous renvoyons le lecteur à l'article pour plus de détail. La figure 3.5 résume cette situation.

**Proposition 3.2.6 (Monas. Math., 2008)** *Soit  $\beta$  un nombre de Pisot unitaire. Soient  $a, b \in \mathcal{A}$  tels que  $T_\beta^{(b-1)}(1) \leq T_\beta^{(a-1)}(1)$  et  $x \in \mathbb{Z}[1/\beta]$ . Soit  $\Delta_\infty$  l'application diagonale :  $\Delta : x \in \mathbb{R} \rightarrow (x, \dots, x) \in \mathbb{R}^{r-1} \times \mathbb{C}^s$ . On définit les ensembles suivants pour décrire l'intersection entre différentes tuiles :*

- $A_{a,b} = -\mathcal{T}^{(a)} \cap -\mathcal{T}^{(b)} \cap \Delta_\infty([T_\beta^{(b-1)}(1), T_\beta^{(a-1)}(1)])$
- $B_{a,x} = -\mathcal{T}^{(a)} \cap -\mathcal{T}(x) \cap \Delta_\infty((0, T_\beta^{(a-1)}(1)))$
- $A = \bigcup_{(a,b) \in \mathcal{A}^2, T_\beta^{(b-1)}(1) \leq T_\beta^{(a-1)}(1)} A_{a,b}$
- $B = \bigcup_{a \in \mathcal{A}, x \in \mathbb{Z}[1/\beta] \cap (0,1)} B_{a,x}$ .

Alors une borne inférieure pour  $\gamma(\beta)$  est donnée par

$$\gamma(\beta) \geq \min\left( \min_{\substack{(a,b) \in \mathcal{A}^2 \\ T_\beta^{(b-1)}(1) \leq T_\beta^{(a-1)}(1) \\ A_{a,b} \neq \emptyset}} \min_{x \in A_{a,b}} \|x\|_\infty, \min_{\substack{a \in \mathcal{A} \\ x \in \mathbb{Z}[1/\beta] \cap (0,1) \\ B_{a,x} \neq \emptyset}} \inf_{x \in B_{a,x}} \|x\|_\infty \right).$$



Une borne supérieure est donnée par :

$$\gamma(\beta) \leq \max \{ \eta; [0, \eta] \subset A \cup B \}.$$

**Irrationalité de  $\gamma(\beta)$**  Dans un travail en cours de rédaction avec Boris Adamczewski, Christiane Frougny et Wolfgang Steiner (A., F., S. & S., en cours), nous exploitons ce résultat et la présence de spirale sur la frontière du fractal pour montrer que dans le cas cubique totalement complexe, la quantité  $\gamma(\beta)$  est un nombre irrationnel. Ceci confirme une observation de [AS05], où, lorsque  $\beta$  est le plus petit nombre de Pisot, les auteurs calculent une approximation de  $\gamma(\beta)$  par 0.66666666608644067488.

**Théorème 3.2.7 (A., F., S. & S., en cours)** *Soit  $\beta$  un nombre de Pisot cubique et unitaire. Soit  $\mathcal{T}$  le fractal de Rauzy associé. Alors tous les points de  $\partial\mathcal{T} \cap \mathbb{R}$  sont des nombres irrationnels. En particulier,  $\gamma(\beta) \notin \mathbb{Q}$ .*

En plus du résultat en lui-même, l'intérêt de ce théorème est qu'il est démontré en utilisant des propriétés topologiques de pavages, ce qui est rare dans les preuve d'irrationalité.

**Le cas non unitaire** Dans le cas non unitaire, un calcul assez simple montre que tous les rationnels de la forme  $1/N(\beta)^n$  n'ont pas un développement périodique. On a donc  $\gamma(\beta) = 0$ .

On introduit alors la quantité suivante, qui évite les nombres dont le dénominateur est certain de générer un développement non périodique.

$$\tilde{\gamma}(\beta) = \sup \{ \varepsilon, \forall p/q < \varepsilon, \text{pgcd}(q, N(\beta)) = 1, d_\beta(p/q) \text{ est purement périodique} \}.$$

Nous montrons dans (Monas. Math., 2008) que contrairement au cas unitaire, cette quantité n'est pas liée à la propriété (F). En effet, dans le cas quadratique non-unitaire, la proposition 3.2.6 se généralise pour montrer que  $\tilde{\gamma}(\beta)$  est intimement lié à la description de la frontière du fractal de Rauzy complet (incluant les places locales). En décrivant cette frontière par un GIFS, on parvient à calculer cette quantité dans des cas particuliers.

**Théorème 3.2.8 (Monas. Math., 2008)** *Les rationnels ont un comportement inattendu vis à vis des développements purement périodiques en base  $\beta$  non unitaire et de la propriété de finitude.*

- $\tilde{\gamma}(2 + \sqrt{7}) = 0$  alors que ce nombre vérifie la propriété (F) (il s'agit de la racine de  $X^2 - 4X - 3$ ).
- $\tilde{\gamma}(5 + 2\sqrt{7}) = (7 - \sqrt{7})/12$ , qui vérifie toujours la propriété (F) (il s'agit de la racine de  $X^2 - 10X - 3$ ).

Dans (A., F., S.& S., en cours), nous montrons aussi que dans le cas quadratique non ramifié,  $\tilde{\gamma}(\beta)$  appartient toujours à  $\mathbb{Q}(\beta)$ , et nous proposons un algorithme pour calculer cette valeur à partir de la notion de graphe de frontière.

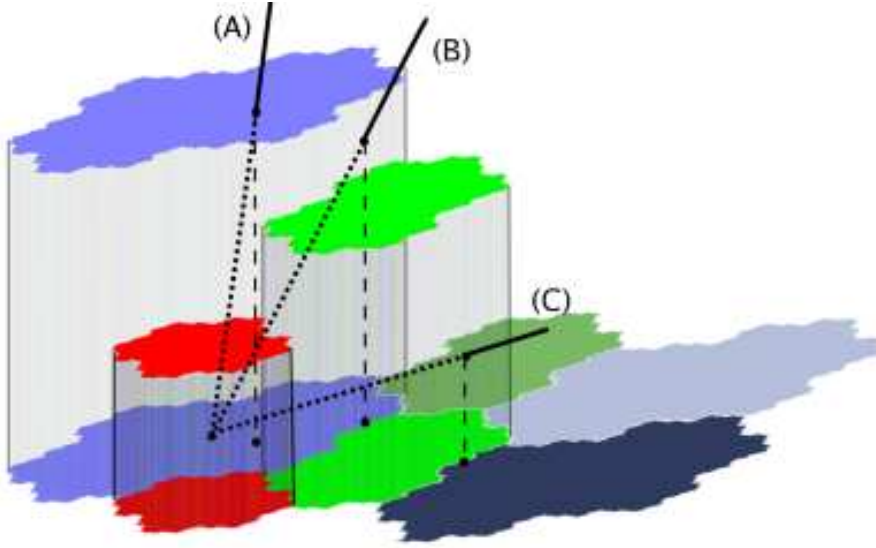


FIG. 3.5 – Illustration des trois cas de calcul des quantités  $\gamma(\beta)$  et  $\tilde{\gamma}(\beta)$ . L'illustration est faite dans un cas unitaire, mais la démonstration reste valable dans le cas non unitaire. La quantité  $\tilde{\gamma}(\beta)$  est donné par la partie la plus large de la diagonale qui est entièrement inclus dans l'extension naturelle  $\bigcup_{a \in \mathcal{A}} (-\tilde{\mathcal{T}}(a)) \times [0, T_\beta^{(a-1)}(1))$ . Cette extension naturelle est représentée par les sous-tuiles  $-\tilde{\mathcal{T}}(a)$  dans la direction horizontale et par l'intervalle  $[0, 1)$  sur l'axe vertical. Ainsi, l'extension naturelle est une union de cylindres à base fractale et de hauteur verticale. La hauteur du cylindre dépend de la sous-tuile considérée.

En fonction du premier endroit où la diagonale sort de cette extension naturelle, on obtient différentes situations pour  $\tilde{\gamma}(\beta)$ , décrites dans (Monas. Math., 2008) et dans la proposition 3.2.6.

La première situation correspond au cas où la diagonale part de 0 et sort de l'extension naturelle sur un plateau de hauteur  $T_\beta^{(a-1)}(1)$ . Alors,  $\tilde{\gamma}(\beta)$  appartient à l'orbite de 1 sous l'action de  $T_\beta$ .

La deuxième situation signifie que la diagonale sort de l'extension naturelle en traversant une verticale entre deux plateaux. Le point d'intersection se situe alors au dessus de l'intersection entre deux sous-tuiles  $(-\tilde{\mathcal{T}}(a)) \cap (-\tilde{\mathcal{T}}(b))$ .

La dernière situation signifie que la diagonale traverse complètement l'extension naturelle et sort au dessus d'une nouvelle tuile du pavage autosimilaire.

À partir de cette caractérisation et d'une description par graphe des intersections entre les tuiles du pavage, on obtient un calcul exact de  $\tilde{\gamma}(\beta)$  pour certains  $\beta$  et des informations sur son irrationalité.

---

**Théorème 3.2.9 (A., F., S.& S., en cours)** *Supposons que  $\beta$  est un nombre quadratique non unitaire, solution de  $\beta^2 = a\beta + b$ . Soit  $d$  l'entier sans facteur carré tel que  $\mathbb{Q}(\beta) = \mathbb{Q}(\sqrt{d})$ . Si*

- *$b$  est sans facteur carré,*
- *$b$  est premier avec le discriminant de  $\mathbb{Q}(\beta)$ ,*
- *$d$  est un résidu quadratique pour tout les diviseurs premiers impairs de  $b$ ,*
- *$d \equiv 1 \pmod{8}$  si  $b$  est pair.*

*Alors  $\tilde{\gamma}(\beta)$  est un élément de  $\mathbb{Q}(\beta)$ . De plus, il existe un algorithme pour le calculer à partir du graphe de frontière du fractal de Rauzy associé à la  $\beta$ -numération.*

### 3.2.4 Perspectives de travail

L'objectif principal sur cette question reste de mieux utiliser les propriétés topologiques de fractals pour comprendre les propriétés des beta-numérations. Cela se décline en plusieurs points.

**Interprétation des propriétés topologiques du fractal en terme de numération** Nous avons vu que dans le cas unitaire, le fait que 0 est point intérieur du fractal de Rauzy signifie que tous les éléments de  $\mathbb{Z}[\beta] \cap \mathbb{R}^+$  ont un développement fini. Une question naturelle est alors de comprendre la signification de la connexité locale du fractal, de sa connexité ou de sa simple connexité.

Dans le cas non-unitaire, ces questions sont délicates dans la mesure où l'espace de représentation est totalement déconnecté. Par contre, quel est l'équivalent de la propriété de finitude dans ce cas ? Peut-on caractériser topologiquement l'ensemble des points dont le développement est fini ?

**Transcendance de  $\gamma(\beta)$**  Une question difficile est d'aller plus loin dans l'étude de  $\gamma(\beta)$ , au moins dans le cas unitaire cubique.

En effet, nous savons montrer que  $\gamma(\beta)$  n'est alors pas un nombre rationnel (A., F., S.& S., en cours). Nous utilisons pour cela le fait que cette quantité est décrite par l'intersection d'une courbe fractale et d'une droite. Cependant, on n'utilise pas la description explicite du fractal par son graphe de frontière. La connaissance de ce graphe peut-elle permettre de déterminer la valeur de  $\gamma(\beta)$  (comme c'est le cas dans le cas quadratique), ou au moins montrer qu'il ne s'agit pas d'un élément de  $\mathbb{Q}(\beta)$  ?

Dans le cas quadratique non-unitaire, on obtient l'algébraïcité de  $\tilde{\gamma}(\beta)$  en étudiant la frontière du fractal par une description autosimilaire. Dans le cas cubique, il faut se restreindre à l'intersection du fractal avec une droite, qui n'est pas décrite par un IFS a priori. Pour avancer dans cette direction, il faudra décrire explicitement l'intersection du fractal et la droite, sans doute avec un arbre infini du type diagramme de Bratelli.

**Vers des approximations diophantiennes** Une autre question d'intérêt est relative au calcul des meilleures approximations simultanées d'un couple de réels.

En effet, dans [HM06, IFHY03a], en suivant les idées de [Rau82], les auteurs considèrent des nombres de Pisot  $\beta$  cubiques unitaires avec un conjugué de Galois non réel. Ils montrent que les meilleures approximations du  $(1/\beta, 1/\beta^2)$  pour une norme euclidienne décrite par  $\beta$  sont données par la suite récurrence associée à  $\beta$ , et, en tout état de cause, que ces nombres sont très mal approximatés par des nombres rationnels (puisque  $\sqrt{N}\|N(1/\beta, 1/\beta^2)\|$  la quantité est minorée par une quantité indépendante de  $N$ ).

La preuve s'appuie sur le fractal de Rauzy, en particulier ses propriétés de pavages, sur le fait que  $\mathbf{0}$  est point intérieur, et l'étude des ellipses incluses dans ce fractal.

Pour généraliser cette approche, il faut être capable de déterminer le rayon du fractal de Rauzy, c'est-à-dire la taille maximale de la boule incluse dans le fractal, éventuellement pour différentes topologies. Ensuite, il faudra déterminer la valeur du rayon de cette boule ; sans doute en la décrivant par un arbre pondéré.

**Numération de Dumont-Thomas** Une autre direction d'étude consiste à généraliser les résultats sur la description des éléments purement périodiques aux numérations de Dumont-Thomas [DT93].

Dans (Integers, 2005) , nous avons proposé une dynamique relative aux numérations de Dumont-Thomas. Il est raisonnable de penser que les fractals de Rauzy décrivent eux aussi les développements purement périodiques dans ce système de numération.

De plus, pour les systèmes de numération en base entière ou certains systèmes en base  $\beta$ , des liens ont été mis en évidence entre les graphes de frontière et des automates pour décrire le développement de l'addition de deux nombres [Berar, ST02]. Il sera intéressant de généraliser ces approches pour décrire le développement de Dumont-Thomas de la somme de deux nombres à partir du graphe des frontières.

### 3.3 Géométrie discrète : engendrement de plans discrets

Comme mentionné dans le chapitre 1, les codages de droites discrètes dans le plan sont parfaitement identifiés et peuvent être générés au moyens de substitutions en considérant le développement en fractions continues de la pente de la droite approximée [Fog02, Chapitre 6].

Une question qui occupe les informaticiens théoriciens et les mathématiciens est de généraliser ce résultat au cas multidimensionnel : par exemple, est-il possible de caractériser les approximations de plans discrets dans  $\mathbb{R}^3$  ? Peut-on les engendrer de manière efficace ? D'un point de vue mathématique, répondre à cette question permettrait d'identifier pour tout vecteur de  $\mathbb{R}^2$  une suite de réseaux de  $\mathbb{R}^2$  et une partition de leurs domaines fondamentaux qui décriraient l'action de la translation par le vecteur.

Pour cela, plusieurs aspects doivent être considérés : il faut d'abord caractériser les approximations de plans de manière combinatoire, et si possible dynamique, de

même que les suites sturmiennes sont caractérisées comme les codages d'additions sur le tore de dimension 1. Il faut ensuite être capable de définir un développement en fraction continue multidimensionnel d'un vecteur (qui sera le vecteur orthogonal au plan approximé). Il faut enfin définir des substitutions multidimensionnelles qui permettront d'engendrer le codage de l'approximation de plan.

Dans cette direction, un procédé d'extension des substitutions en grande dimension a été développé par Arnoux et Ito [AI01], et il semble très prometteur. Nous allons dans cette section détailler comment il est possible de l'utiliser pour répondre partiellement à la question de l'engendrement des plans discrets, et quelles sont les perspectives dans ce domaine.

### 3.3.1 Plans discrets

**Définitions** La notion de discrétisation de surface est formellement apparue dans [Rev91] sous la forme d'espace arithmétique. Suivant la définition de [ABI02, BV00], on considère un plan  $\mathcal{P}_{(a,b,c)}$  de  $\mathbb{R}^3$  d'équation  $ax + by + cz = 0$ , et on note  $\mathcal{S}$  l'ensemble des cubes à coordonnées entières qui intersectent le demi-espace d'équation  $ax + by + cz < 0$ . Le *plan discret* ou la *surface brisée* associée à  $\mathcal{P}_{(a,b,c)}$  est la frontière de l'ensemble  $\mathcal{S}$ . Ce plan discret est noté  $\mathfrak{P}_{(a,b,c)}$ .

**Faces du plan discret** Le plan discret  $\mathfrak{P}_{(a,b,c)}$  est naturellement recouvert par différentes faces de cubes placées en des points à coordonnées entières. Pour décrire un plan discret, on recherche donc les couples de la forme

[ point à coordonnées entières , orientation de la face de cube ]

qui permettent de recouvrir  $\mathfrak{P}_{(a,b,c)}$ . On introduit donc trois faces de cubes qui vont permettre de reconstruire  $\mathfrak{P}_{(a,b,c)}$ .

$$\begin{aligned} F_1 &= \{ \lambda \mathbf{e}_2 + \mu \mathbf{e}_3, (\lambda, \mu) \in [0, 1[^2 \} \\ F_2 &= \{ \lambda \mathbf{e}_1 + \mu \mathbf{e}_3, (\lambda, \mu) \in [0, 1[^2 \} \\ F_3 &= \{ \lambda \mathbf{e}_1 + \mu \mathbf{e}_2, (\lambda, \mu) \in [0, 1[^2 \}. \end{aligned}$$

Selon [ABI02], une face  $\mathbf{x} + F_i$  de  $\mathfrak{P}_{(a,b,c)}$  sera alors désignée par la notation  $[\mathbf{x}, i^*]$ . L'ensemble des faces de  $\mathfrak{P}_{(a,b,c)}$  est appelé  $\Gamma_{(a,b,c)}$ ; c'est cet ensemble qu'il faut déterminer.

$$\mathfrak{P}_{(a,b,c)} = \bigcup_{[\mathbf{x}, i^*] \in \Gamma_{a,b,c}} \mathbf{x} + F_i.$$

**Caractérisation arithmétique** Par définition de la frontière dans ce cas bien particulier, l'ensemble  $\mathbf{x} + F_i$  est sur la frontière du demi-espace  $\mathcal{S}$  si et seulement si le vecteur  $\mathbf{x} - \mathbf{e}_i$  appartient à  $\mathcal{S}$  tandis que  $\mathbf{x}$  n'appartient pas à  $\mathcal{S}$ . Puisque  $\mathcal{S}$  est défini par l'équation  $ax + by + cz < 0$ , on obtient une caractérisation des vecteurs  $\mathbf{x}$  à coordonnées entières  $(a, b, c)$  :

- L'ensemble  $\mathbf{x} + F_1$  est une face du plan discret si et seulement  $0 \leq ax + by + cz < a$ .
- L'ensemble  $\mathbf{x} + F_2$  est une face du plan discret si et seulement  $0 \leq ax + by + cz < b$ .

- L'ensemble  $\mathbf{x} + F_3$  est une face du plan discret si et seulement  $0 \leq ax + by + cz < c$ .

On obtient donc une description arithmétique de  $\mathfrak{F}_{(a,b,c)}$  :

$$\Gamma_{a,b,c} = \{[\mathbf{x}, i^*], \langle \mathbf{x}, (a, b, c) \rangle \geq 0 \text{ et } \langle \mathbf{x}, (a, b, c) \rangle < \langle \mathbf{e}_i, (a, b, c) \rangle\}. \quad (3.5)$$

### 3.3.2 Substitutions généralisées

L'objectif ici est de pouvoir engendrer par un formalisme répétitif un plan discret. En particulier, il faut être capable d'agir sur les faces  $[\mathbf{x}, i^*]$ . Dans ce cadre, Arnoux et Ito ont fait dans [AI01, AIS01] une avancée considérable en définissant une *substitution généralisée*, obtenue comme application duale d'une substitution et qui agit effectivement sur des faces de cubes.

**Définition** Soit  $\sigma$  une substitution unimodulaire sur trois lettres, de matrice  $\mathbf{M}$ . On appelle *substitution généralisée* l'application  $\mathbf{E}_1(\sigma)^*$  définie comme suit :

$$\mathbf{E}_1(\sigma)^*[\mathbf{x}, i^*] = \bigcup_{k \in \{1,2,3\}} \bigcup_{S, \sigma(k)=PiS} [\mathbf{M}^{-1}\mathbf{x} + \mathbf{M}^{-1}\mathbf{1}(S), k^*]. \quad (3.6)$$

$$\mathbf{E}_1(\sigma)^*(A \cup B) = \mathbf{E}_1(\sigma)^*(A) \cup \mathbf{E}_1(\sigma)^*(B).$$

On considère par exemple la substitution définie par  $\sigma(1) = 3, \sigma(2) = 13, \sigma(3) = 233$ . Sa matrice est  $\mathbf{M} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 2 \end{pmatrix}$ . Son inverse est  $\mathbf{M}^{-1} = \begin{pmatrix} -1 & -2 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$ .

Pour calculer l'image d'une face de type 1 par  $\mathbf{E}_1(\sigma)^*$ , on recherche les 1 qui apparaissent dans l'image d'une lettre. Il s'avère que 1 n'apparaît que dans l'image de la lettre 2, suivie d'un suffixe égal 3. L'image par  $\mathbf{E}_1(\sigma)^*$  d'une face de type 1 est donc une face de type 1, déplacée par le vecteur qui correspond au suffixe suivant 1, c'est-à-dire  $\mathbf{M}^{-1}\mathbf{1}(3)$ . L'image d'une face de type 2 ne contient elle aussi qu'une face. Par contre, puisque 3 apparaît 4 fois dans les images de lettres, l'image d'une face de type 3 contiendra 3 faces différentes. Finalement, la substitution généralisée est définie comme suit. De manière plus visuelle, on peut dessiner les images de faces, comme illustré à la figure 3.6.

$$\begin{aligned} \mathbf{E}_1(\sigma)^*[\mathbf{x}, 1^*] &= \mathbf{M}^{-1}\mathbf{x} + [\mathbf{M}^{-1}\mathbf{1}(3), 2^*] = \mathbf{M}^{-1}\mathbf{x} + [(1, 0, 0), 2^*] \\ \mathbf{E}_1(\sigma)^*[\mathbf{x}, 2^*] &= \mathbf{M}^{-1}\mathbf{x} + [\mathbf{M}^{-1}\mathbf{1}(33), 3^*] = \mathbf{M}^{-1}\mathbf{x} + [(2, 0, 0), 3^*] \\ \mathbf{E}_1(\sigma)^*[\mathbf{x}, 3^*] &= \mathbf{M}^{-1}\mathbf{x} + [\mathbf{0}, 1^*] \cup [\mathbf{0}, 2^*] \cup [\mathbf{0}, 3^*] \cup [(1, 0, 0), 3^*] \end{aligned}$$

**Propriétés fondamentales** Les substitutions duales ont plusieurs propriétés fondamentales :

- La substitution duale de la composition  $\sigma_1\sigma_2$  est la composée de  $\mathbf{E}_1(\sigma_2)^*$  par  $\mathbf{E}_1(\sigma_1)^*$ .

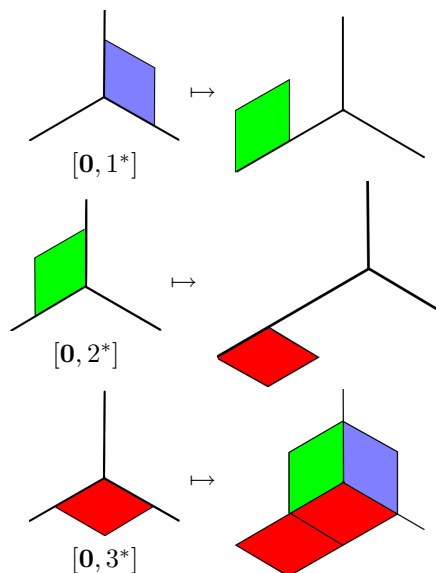


FIG. 3.6 – Images des faces du cube unité par la substitution généralisée de la substitution  $\sigma(1) = 3, \sigma(2) = 13, \sigma(3) = 233$ . La face  $[0, 1^*]$  est représentée par la face du cube orthogonale à l'axe des  $x$  (similairement pour les autres faces).

- Si  $A$  et  $B$  sont deux sous-ensembles disjoints d'un plan discret  $\Gamma_{\mathbf{x}}$  de vecteur normal  $\mathbf{x} = (a, b, c)$ , alors  $\mathbf{E}_1(\sigma)^*(A)$  et  $\mathbf{E}_1(\sigma)^*(B)$  sont constitués de pièces disjointes du plan  $[\text{AI01}]$  discret de vecteur normal  $\mathbf{M}\mathbf{x}$  [Fer05] :

$$\mathbf{E}_1(\sigma)^*\Gamma_{\mathbf{x}} \subset \Gamma_{\mathbf{M}\mathbf{x}}.$$

- Si on note  $\mathcal{U} = [0, 1^*] \cup [0, 2^*] \cup [0, 3^*]$  l'ensemble des faces du cube unité, ce cube unité est contenu dans son image par  $\mathbf{E}_1(\sigma)^*$  :

$$\mathcal{U} \subset \mathbf{E}_1(\sigma)^*\mathcal{U}.$$

**Développement en fractions continues** Même si une définition de fractions continues fait encore débat [Sch00b, Lag94, Bre81, Gra92, Nog95], globalement, tous les algorithmes de développements en fraction continue (unidimensionnels comme multidimensionnels) consistent à produire à partir d'un vecteur  $\mathbf{x}$  des suites de matrices  $\mathbf{M}_i$  à coefficients entiers, et qui appartiennent à une famille finie de  $SL(\mathbb{N})$ , de telle sorte que le produit des matrices  $\mathbf{M}_1\mathbf{M}_2 \dots \mathbf{M}_n \mathbb{R}^n$  converge vers une unique direction  $\mathbb{R}\mathbf{x}$ .

En dimension 2, il est prouvé qu'il n'existe globalement qu'une seule manière d'arriver à ce résultat, via l'algorithme d'Euclide, à cause de la rigidité de  $SL_2(\mathbb{N})$ , qui est un monoïde libre engendré par deux matrices. De cette unicité dérivent

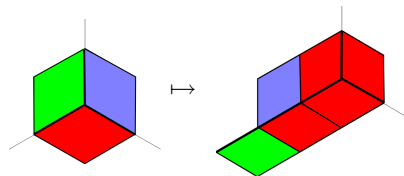


FIG. 3.7 – Image du cube unité par la substitution généralisée de  $\sigma(1) = 3$ ,  $\sigma(2) = 13$ ,  $\sigma(3) = 233$ . Chaque couleur dans la figure de droite correspond à l'image de la même de la dite couleur dans le cube unité à gauche. On remarquera que les images des différentes faces du cube unité ne se superposent pas. De plus, l'image du cube unité contient le cube unité lui-même. Ces deux propriétés sont vérifiées par toutes les substitutions généralisées.

de nombreuses propriétés remarquables pour les fractions continues (en terme de meilleures approximations rationnelles de nombres, de caractérisation des nombres quadratiques, ou de vitesses de convergences d'algorithmes) [Khi63, IK02].

En dimension supérieure, il n'existe plus de manière canonique de choisir la suite des matrices  $\mathbf{M}_i$ . À partir de l'algorithme d'Euclide, de nombreux auteurs ont dérivé différents algorithmes ; les plus célèbres sont sans doute les algorithmes de Jacobi-Perron [Jac68, FP54], Brun [Pod77], Selmer [Sel78] ou Poincaré [Sch00b], mais d'autres sont encore proposés actuellement [Yas05]. Même si ce domaine est extrêmement vaste, une conclusion s'impose : l'ensemble des propriétés remarquables obtenues pour l'algorithme d'Euclide en dimension 2 ne pourra pas se retrouver simultanément en dimension supérieure et le choix d'un algorithme de fractions continues multidimensionnel doit se faire principalement en fonction des applications recherchées. Par exemple, les algorithmes de Brun et Jacobi-Perron sont de bons candidats pour obtenir de bonnes approximations rationnelles de vecteurs [BAG01, FIKO96], mais ne semblent pas totalement appropriés pour caractériser les corps cubiques [DPLR72, IFHY03b, IY07]. Les vecteurs admettant un développement en fractions continues d'Arnoux-Rauzy [AR91] périodique semblent de bons candidats pour être tous conjugués à une rotation.

**Engendrement de bouts de plans discrets à partir d'un algorithme de fractions continues** Quel que soit l'algorithme, cependant, on parvient donc à écrire un vecteur  $\mathbf{x}$  sous la forme

$$\mathbf{x} = \mathbf{M}_1 \mathbf{M}_2 \dots \mathbf{M}_n \mathbf{x}_n. \quad (3.7)$$

Imaginons maintenant que les matrices  ${}^t\mathbf{M}_i$  sont les matrices de substitutions  $\sigma_i$ . Dans les faits, il n'y a aucun problème pour choisir une substitution pour chaque matrice, dès qu'elle est à coefficients positifs. Selon les propriétés des substitutions généralisées, on en déduit que le codage  $\Gamma_{\mathbf{x}}$  du plan discret normal à  $\mathbf{x}$  se construit en partie à partir du plan discret normal à  $\mathbf{x}_n$ , puisqu'il contient  $\mathbf{E}_1(\sigma_1)^* \dots \mathbf{E}_1(\sigma_n)^*(\Gamma_{\mathbf{x}_n})$ . En particulier, puisque tous les plans discrets contiennent le cube unité  $\mathcal{U}$ , on peut



déduire que le plan discret  $\Gamma_{\mathbf{x}}$  contient un motif généré par des substitutions généralisées

$$\mathbf{E}_1(\sigma_1)^* \dots \mathbf{E}_1(\sigma_n)^*(\mathcal{U}) \subset \Gamma_{\mathbf{x}}. \quad (3.8)$$

**Stratégie d'engendrement** Différents travaux ont exhibé des conditions suffisantes sur le choix des substitutions ou des vecteurs pour s'assurer que les motifs qui apparaissent dans la formule (3.8) permettent de recouvrir tout le plan discret  $\Gamma_{\mathbf{x}}$ . Nous allons détailler ces différentes approches et nos contributions dans les lignes qui suivent.

### 3.3.3 Le cas purement substitutif

**Restriction sur les plans considérés** Une première question qui a été largement étudiée concerne le cas où les matrices  $\mathbf{M}$  dans le développement (3.7) sont toutes identiques. Le vecteur  $\mathbf{x}$  est alors un vecteur propre de la matrice  $\mathbf{M}$ , pour sa valeur propre dominante.

Dans ce cas, les motifs  $\mathbf{E}_1(\sigma_1)^* \dots \mathbf{E}_1(\sigma_n)^*(\Gamma_{\mathbf{x}_n})$  s'étalent dans le plan discret normal à  $\mathbf{x}$ . Cependant, pour espérer recouvrir tout ce plan discret, il faut que les motifs s'étalent de manière relativement uniforme.

Or, s'il existe une valeur propre dilatante pour  $\mathbf{M}$  (autre que la valeur propre dominante), on montre que les motifs vont se rassembler autour de l'espace propre pour cette deuxième valeur propre; ils ne pourront donc pas recouvrir tout le plan discret.

Ainsi, on doit supposer que la matrice  $\mathbf{M}$  admet une seule valeur propre dilatante : selon le formalisme introduit dans le chapitre 2, elle est Pisot irréductible. Le plan approximé est alors orthogonal à  $\mathbf{x}$ , il s'agit du plan contractant de la matrice  ${}^t\mathbf{M}$ .

Ce cas est la généralisation du cas purement substitutif dans le cadre de l'approximation des droites de  $\mathbb{R}^2$ . En effet, on sait que le codage d'une droite discrète est engendré par une substitution si et seulement si la pente de la droite est un nombre quadratique dont le conjugué de Galois est en dehors de l'intervalle  $(0, 1)$  [Yas99]. La généralisation de cette restriction en dimension 3 correspond à considérer des matrices Pisot.

Notons que même si ces restrictions paraissent farfelues et peu naturelles, elles concernent une famille suffisamment grande pour permettre d'avancer dans la direction de l'engendrement de tous les plans discrets. Nous discuterons de cela dans la section 3.3.6.

**Engendrer un plan discret par itération à partir du cube unité** Dans le cas Pisot, on est ramené à rechercher quels sont les vecteurs  $\mathbf{x}$  pour lesquels le plan discret est recouvert par les motifs  $(\mathbf{E}_1(\sigma)^*)^n(\mathcal{U})$ . En réalisant que les substitutions généralisées ne font que mimer les développements en base non-entières existant sur la droite réelle, on montre dans (Integers, 2005) et (Monographie S. & T., 2008) la propriété suivante (la condition est exprimé sous forme suffisante dans (Integers, 2005), le fait qu'elle soit nécessaire est prouvé dans (Monographie S. & T., 2008)).

---

**Proposition 3.3.1** ((Integers, 2005),(Monographie S. & T., 2008)) *Soit  $\mathbf{x}$  le vecteur propre dilatant à gauche de la matrice  $\mathbf{M}$  d'une substitution unitaire et irréductible Pisot sur trois lettres. Les ensembles  $(\mathbf{E}_1(\sigma)^*)^n(\mathcal{U})$  recouvrent la discrétisation  $\Gamma_{\mathbf{x}}$  du plan contractant de  $\mathbf{M}$  si et seulement si la substitution  $\sigma$  vérifie la condition de finitude étendue (3.3).*

En particulier, selon la section 3.2.2, ceci est équivalent au fait que  $\mathbf{0}$  est un point intérieur du fractal de Rauzy associé à  $\sigma$ . Cette propriété peut donc se vérifier en construisant un graphe de frontière. Cependant, une telle vérification sort du cadre de la géométrie discrète et est difficilement généralisable au cas où on composera des substitutions différentes. Nous allons donc introduire un nouveau critère pour l'engendrement du plan  $\Gamma_{\mathbf{x}}$ .

**Remarque : préfixes, suffixes et versions analogues** La formule (3.6) qui définit  $\mathbf{E}_1(\sigma)^*$  fait intervenir les *suffixes* d'images de lettres, alors que la formule (3.3) fait intervenir les *préfixes* d'images de lettres. Ceci s'explique pour des raisons historiques et d'applications : la formule (3.6) est obtenue par dualité de l'action d'une substitution sur des escaliers ; pour avoir une expression simple, ce sont les inverses de préfixes  $-\mathbf{l}(P)$  qui apparaissent, avec des jeux de faces de cube  $F_1, F_2, F_3$  qui ne contiennent pas  $\mathbf{0}$  (Arnoux et Ito considère les faces supérieures du cube unité). Recentrer les faces en zéro dans la formule de Arnoux et Ito remplace les préfixes  $-\mathbf{l}(P)$  par des suffixes  $\mathbf{l}(S)$  ; on obtient ainsi la formule (3.6) qui fait intervenir des suffixes.

Par contre, la formule (3.3) est inspirée des approches en beta-numération, où les préfixes d'images de lettres ont été déterminants à cause de la définition des beta-substitutions.

En fait, on peut parfaitement remplacer les suffixes  $\mathbf{l}(S)$  dans (3.6) par des préfixes  $\mathbf{l}(P)$ , cela revient à considérer la substitution miroir à la substitution initiale. Dans ce cas, les propriétés des substitutions généralisées ne changent pas, puisqu'elles ne font intervenir que la linéarisation de la matrice. De même, le fractal de Rauzy ainsi obtenu est le symétrique  $-\mathcal{T}$  du fractal obtenu pour la substitution initiale ; le fait que  $\mathbf{0}$  est point intérieur est donc toujours vérifié.

Pour respecter les raisons historiques, nous garderons donc dans la suite la définition (3.6), mais il faut avoir à l'esprit que les applications de ces résultats aux beta-substitutions et en numération doivent être faites en considérant les images miroir des substitutions.

**Engendrer un plan discret à partir d'un motif** Il existe un grand nombre de substitutions pour lesquels on sait que les itérations à partir du cube unité ne permettront pas de recouvrir tout le plan discret  $\Gamma_{\mathbf{x}}$  ; il suffit de considérer toutes les substitutions pour lesquelles  $\mathbf{0}$  n'est pas point intérieur du fractal de Rauzy, que nous avons illustré au début du chapitre. Notons cependant qu'on ne sait pas si pour une matrice donnée il existe au moins une substitution ayant cette matrice pour laquelle la condition est vérifiée.

Dans ce cas, il est possible de considérer un motif de taille plus importante à partir duquel on peut engendrer tout le plan discret : il existe un motif  $\mathcal{V}$  de  $\Gamma_{\mathbf{x}}$  tel que

$$\Gamma_{\mathbf{x}} = \cup_{n \geq 0} (\mathbf{E}_1(\sigma)^*)^n(\mathcal{V}).$$

L'existence de ce motif peut être démontrée de plusieurs manières. Dans le cas des beta-substitutions, l'existence d'un tel motif est prouvée dans [Aki00] en s'appuyant sur la convergence de séries. Dans le survol (CANT, 2009), V. Berthé, J. Thuswaldner et moi-même introduisons une approche purement géométrique qui s'appuie sur la notion de réseau. On obtient alors une formule explicite pour le motif  $\mathcal{V}$ , qui dépend du vecteur propre dilatant à droite et des vecteurs propres contractants à gauche de la matrice d'incidence de la substitution (voir proposition suivante).

### Une nouvelle condition équivalente pour la propriété de finitude (F)

L'existence de ce motif permet de donner une nouvelle condition nécessaire et suffisante pour l'engendrement du plan discret  $\Gamma_{\mathbf{x}}$ . La preuve sera publiée dans (CANT, 2009).

**Proposition 3.3.2 (CANT, 2009)** *Soit  $\sigma$  une substitution unitaire de type Pisot. Soit  $\mathbf{x}$  le vecteur propre dilatant à gauche de la matrice d'incidence  $\mathbf{M}$  de  $\sigma$ . La valeur propre dilatante est notée  $\beta$ . On note aussi  $\mathbf{v}_{\beta^{(2)}}, \dots, \mathbf{v}_{\beta^{(r+s)}}$  les vecteurs propres contractants de  $\mathbf{M}$  pour les conjugués de  $\beta$ . On introduit les constantes  $\mathcal{K}_k = \max\{|\langle \mathbf{1}(S), \mathbf{v}_{\beta^{(k)}} \rangle|\}$  pour  $s$  suffixe propre de l'image d'une lettre par  $\sigma$  et  $\mathcal{K}'_k = \min\{|\langle \mathbf{1}(S), \mathbf{v}_{\beta^{(k)}} \rangle|\}$  pour  $s$  suffixe propre de l'image d'une lettre par  $\sigma$ .*

*La constante d'itération associée à la substitution est*

$$N = \min \left\{ \frac{\log(2 \frac{\mathcal{K}_k}{1-|\beta^{(k)}|}) - \log(\mathcal{K}'_k)}{-\log(|\beta^{(k)}|)}, 2 \leq k \leq r+s \right\}.$$

*Alors  $\sigma$  satisfait la propriété de finitude étendue, c'est-à-dire que le plan discret  $\Gamma_{\mathbf{x}}$  est engendré à partir des itérations du cube unité si et seulement si le motif  $\mathbf{E}_1(\sigma)^{*N}(\mathcal{U})$  recouvre le motif  $\mathcal{V}$  défini par :*

$$\mathcal{V} = \left\{ [\mathbf{x}, i^*] \in \mathbb{Z}^n \times \mathcal{A}, \left\{ \begin{array}{l} 0 \leq \langle \mathbf{x}, \mathbf{u}_\beta \rangle < \langle \mathbf{e}_i, \mathbf{u}_\beta \rangle \\ |\langle \mathbf{x}, \mathbf{v}_{\beta^{(k)}} \rangle| < \frac{\mathcal{K}_k}{1-|\beta^{(k)}|}, \forall 1 < k \leq r+s \end{array} \right. \right\}.$$

### 3.3.4 Approximations de plans discrets tridimensionnels : l'école japonaise

Après le cas purement substitutif se pose naturellement la question de l'engendrement de n'importe quel plan discret. Pour cela, comme dans le cas des droites discrètes, on remplace l'itération d'un même substitution par la composition de substitutions différentes. La propriété de finitude (F) n'a alors plus aucun sens, et il faut trouver un lemme qui permet de faire grossir les différentes itérations du cube unité.

Cette approche a été développée dans les années 1990 par l'école japonaise portée par S. Ito [IO94] à partir des outils définis plus haut, en se basant sur l'algorithme de fractions continues de Jacobi-Perron. Avec P. Arnoux et V. Berthé (Jour. Montoises, 2006), et dans un travail en cours avec V. Berthé et J. Bourdon (B., B. & S., en cours), nous travaillons sur la généralisation l'automatisation de l'approche de [IO94] pour l'appliquer à d'autres algorithmes de fractions continues et étudier plus précisément les différentes approximations de plans discrets. Avant de discuter des apports, il est nécessaire de détailler un peu les points principaux de l'approche de Ito et Ohtsuki.

**Une stratégie d'engendrement des plans discrets** Ito et Ohtsuki cherchent à montrer que toutes les approximations discrètes de plans sont uniformément récurrentes. Ceci a été redémontré avec des approches arithmétiques plus tard [BV00], mais l'approche de [IO94] demeure fort intéressante pour étudier la structure des plans discrets. Leur idée est la suivante :

1. Tous les plans discrets sont caractérisés par leur vecteur normal  $\mathbf{x}$  ; ils sont ainsi de la forme  $\Gamma_{\mathbf{x}}$ .
2. Le vecteur  $\mathbf{x}$  peut être approximé par un algorithme de fractions continues. En particulier, l'algorithme de Jacobi-Perron permet de décrire tout vecteur  $\mathbf{x}$  par une alternance de deux paramètres. Plus précisément, on considère les familles de matrices

$$\mathbf{M}(B, C) = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & B & C \end{pmatrix}$$

L'algorithme de Jacobi-Perron considère un vecteur  $\mathbf{x} = (a, b, c)$  tel que  $0 \leq a, b < c$  et construit un vecteur  $\mathbf{x}_1 = (a_1, b_1, c_1)$  tel que  $0 \leq a_1, b_1 < c_1$  avec

$$\mathbf{x} = {}^t\mathbf{M}(B_0, C_0)\mathbf{x}_1, \quad B_0 = [b/a] \quad C_0 = [c/a].$$

Cette opération se répète infiniment, et permet de décomposer le vecteur  $\mathbf{x}$  sous la forme

$$\mathbf{x} = {}^t\mathbf{M}(B_0, C_0) {}^t\mathbf{M}(B_1, C_1) \dots {}^t\mathbf{M}(B_n, C_n) \mathbf{x}_n.$$

Les suites  $B_n, C_n$  vérifient alors la condition d'admissibilité  $0 \leq B_n \leq C_n$ ,  $C_n \geq 1$  et  $C_{n+1} \neq 0$  dès que  $B_n = C_n$ .

3. Pour chaque matrice  $\mathbf{M}(B, C)$  apparaissant dans l'algorithme de Jacobi-Perron, on introduit une substitution qui a cette matrice pour matrice d'incidence.

$$\sigma_{B,C}(1) = 3 \quad \sigma_{B,C}(2) = 13^B \quad \sigma_{B,C}(3) = 23^C.$$

4. Chaque substitution admet une substitution généralisée, qui se représente comme suit :

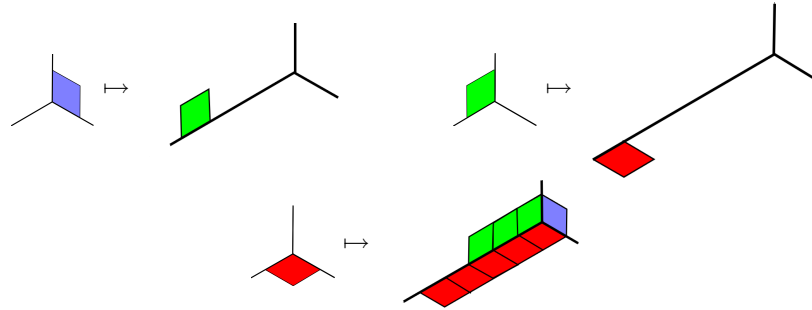


FIG. 3.8 – Images des faces du cube unité par la substitution généralisée d’une substitution de Jacobi-Perron  $\sigma_{3,5}$ . Il y a  $B$  faces de type 2 dans l’image des faces de type 3 par  $\sigma_{B,C}$ , et  $C$  faces de type 3.

5. En construisant un graphe fini, on peut montrer que dès que le développement en fractions continues de  $\mathbf{x}$  est tel qu’il existe une suite infinie d’entiers  $k$  tels que

$$B_n \neq C_n \text{ ou } C_{n+1} = B_{n+1} \text{ ou } B_{n+2} \neq 0, \quad (3.9)$$

alors pour ces entiers, le motif  $\mathbf{E}_1(\sigma(B_0, C_0))^* \dots \mathbf{E}_1(\sigma(B_n, C_n))^*(\mathcal{U})$  contient le cube  $\mathcal{U}$  et l’entoure avec un anneau.

Autrement dit,  $\mathbf{E}_1(\sigma(B_0, C_0))^* \dots \mathbf{E}_1(\sigma(B_n, C_n))^*(\mathcal{U}) \setminus \mathcal{U}$  admet exactement deux composantes connexes, dont une est bornée.

6. Un lemme (dit de l’anneau) remplace la condition de finitude (F). Ce lemme établit que l’image d’un anneau par une substitution généralisée est encore un anneau.
7. On conclut que pour tous les vecteurs  $\mathbf{x}$  qui vérifient la condition du point 3, les motifs  $\mathbf{E}_1(\sigma(B_0, C_0))^* \dots \mathbf{E}_1(\sigma(B_n, C_n))^*(\mathcal{U})$  recouvrent le plan  $\Gamma_{\mathbf{x}}$ . Si cette condition n’est pas vérifiée, une approche identique permet de montrer que le plan est engendré à partir d’un motif  $\mathcal{V}$  légèrement plus gros.

**Limites de l’approche** Il faut faire plusieurs remarques au sujet de ces résultats.

- Le choix de l’algorithme de Jacobi-Perron est arbitraire. Il existe d’autres algorithmes pour approximer des vecteurs, dont les applications sont différentes. En particulier l’algorithme de Brun, utilisé pour la reconnaissance de plans discrets [Pod77, BF08, Fer08], et l’algorithme d’Arnoux-Rauzy, qui permet d’engendrer des systèmes substitutifs dont on conjecture qu’ils sont isomorphes à des additions sur un tore [AR91, AI01]. Pour chaque algorithme, la même question se pose de reconnaître les développements de vecteurs pour lesquels le plan discret est engendré à partir d’un motif simple (si possible le cube unité  $\mathcal{U}$ ).
- La construction du graphe fini a été faite manuellement dans [IO94]. Dans le cas de l’algorithme de Brun, ce graphe devient de taille trop importante et il n’a été ni construit explicitement, ni étudié.

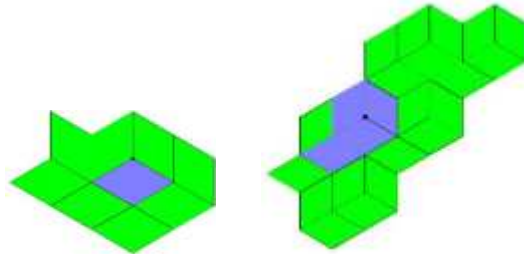


FIG. 3.9 – Un anneau et son image par une substitution généralisée, qui n'est pas un anneau

- L'énoncé du lemme de l'anneau est hélas incomplet. En effet, si on considère la figure 3.9 on constate que l'image de l'anneau de gauche, qui est présentée à droite, n'est pas elle-même un anneau.

Même si cette approche est prometteuse, il y a donc un certains de points à régler avant de l'appliquer à différents algorithmes et l'exploiter à fond. En particulier, le lemme de l'anneau tel qu'énoncé n'est pas exact. Avec P. Arnoux, V. Berthé et J. Bourdon, nous avons retravaillé le lemme de l'anneau en introduisant des conditions combinatoires locales, appelées *règles locales*, qui sont plus proches de la notion de substitutions usuelles. Ces règles locales sont mentionnées dans [IO94] mais elles ne sont pas réellement exploitées.

À partir de ces règles locales, on espère montrer comment il est possible d'engendrer efficacement des plans discrets avec différents algorithmes. Nous allons détailler ces questions dans les deux sous-sections qui suivent.

### 3.3.5 Un nouveau mode d'engendrement : règles locales pour des substitutions bidimensionnelles

**Itérer une formule ou construire par règles d'adjacence ?** Il faut ici faire preuve d'honnêteté : la formule (3.6) n'est absolument pas efficace pour engendrer un plan discret. Lors de chaque itération, il faut calculer pour chaque pièce du motif un produit de matrices par un vecteur ! Or, si on revient au cas unidimensionnel, ce n'est pas du tout de cette manière que les itérations sont calculées pour obtenir des suites sturmiennes : lorsqu'on itère une substitution, on s'appuie plutôt sur la notion d'adjacence d'images de lettres. Plus précisément, la formule de base qui permet de construire les mots sturmiens de manière efficace est la suivante :  $\sigma(UV) = \sigma(U)\sigma(V)$ . Il n'est donc pas nécessaire de calculer systématiquement la position des lettres dans les itérations  $\sigma^n(a)$  ; ce mot est construit petit à petit.

**Motifs** La communauté des informaticiens théoriciens a fait un travail important pour généraliser cette approche au cas multidimensionnel [AS02, Han00]. Est-il possible de définir des règles de positionnement qui permettent d'itérer une substitution

généralisée? Le premier cas non trivial dans ce cadre est de considérer des suites infinies bidimensionnelles, c'est-à-dire des suites  $U_{i,j}$  indicées par  $\mathbb{Z}^2$ . Dans ce cadre, la notion de motifs bi-dimensionnel est naturelle : il s'agit d'un motif purement géométrique, qui ne contient aucune information de localisation dans  $\mathbb{Z}^2$ . Par exemple, un motif peut être constitué d'un 1, placé à gauche d'un 3 et en dessous d'un 2. On ne précise par à quel endroit de  $\mathbb{Z}^2$  ce motif se trouve.

$$\begin{array}{|c|c|} \hline & 2 \\ \hline 3 & 1 \\ \hline \end{array}$$

Cette notion de motif (positionné ou non positionné) est discutée dans la publication (TCS, 2004).

**Une substitution définie par règle de placement ?** Dans ce cadre, une substitution bidimensionnelle le devrait être une application qui associe à chaque lettre un motif bi-dimensionnel non vide. Considérons par exemple :

$$\Sigma_0 : \quad 1 \mapsto \begin{array}{c} 2 \\ 1 \end{array} \quad 2 \mapsto 3 \quad 3 \mapsto 1.$$

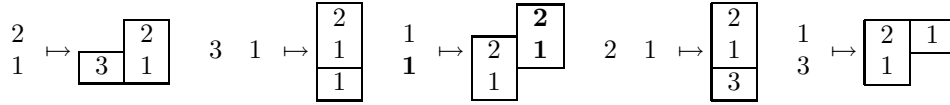
Pour être un morphisme, il faut que cette application puisse s'appliquer sur les motifs et non pas sur les lettres. Pour cela, il est nécessaire de savoir placer les images de lettres les unes par rapport aux autres. Le problème ici est qu'il n'existe pas de structure de monoïde sur  $\mathbb{Z}^2$  : il n'y a pas de manière naturelle pour positionner deux motifs l'un à côté de l'autre. Ainsi, l'image de 12 doit-elle être  $\begin{array}{c} 2 \ 3 \\ 1 \end{array}$  ou

$$\begin{array}{c} 2 \\ 3 \ 1 \end{array}, \text{ or autre chose ?}$$

En fait, on ne sait même pas si des motifs peuvent se positionner correctement pour recouvrir tout  $\mathbb{Z}^2$ , sauf dans le cas particulier des motifs rectangulaires de même taille ou des opérations canoniques existent (mettre deux motifs l'un au dessus de l'autre ou sur leur côté) [GA97].

Même si cette question du placement est résolue, il faut aussi vérifier que ces règles de placement sont compatibles, c'est-à-dire qu'en appliquant différentes règles de placement les éléments ne se recouvrent pas. Là encore, ces questions se résolvent bien dans le cas où les images de lettres sont des rectangles de même taille (la substitution se découpe dans ce cas en deux substitutions unidimensionnelles).

**Un exemple non trivial** Dans (TCS, 2004), P. Arnoux, V. Berthé et moi-même utilisons les substitutions généralisées définies plus haut ainsi que la caractérisation arithmétique des plans discrets et la proposition 3.3.2 pour montrer que des règles non triviales de substitution existent dans  $\mathbb{Z}^2$  pour la substitution  $\Sigma_0$ . En fait, cette substitution est un codage bi-dimensionnel de la substitution généralisée associée par la formule 3.6 à la substitution  $1 \rightarrow 13, 2 \rightarrow 1, 3 \rightarrow 2$ . Les règles locales suivantes sont obtenues en travaillant à partir de la formule (3.6).



En utilisant les propriétés des substitutions généralisées, nous parvenons donc à montrer que ces règles ne se superposent pas et permettent de recouvrir tout  $\mathbb{Z}^2$  en s'appliquant itérativement. Le mot bidimensionnel obtenu est un point fixe de la substitution  $\Sigma_0$  (TCS, 2004).

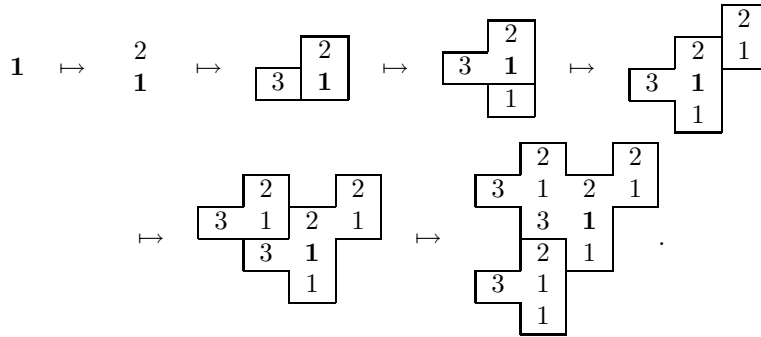


FIG. 3.10 – Itération des règles locales associées à  $\Sigma_0$ .

**Une classe d'exemples non-triviaux** En particulier, toutes les substitutions unidimensionnelles qui vérifient la condition de finitude (F) vont permettre de générer un jeu de règles locales pour des substitutions bidimensionnelles non triviales dont les itérations engendrent des points fixes indicés par  $\mathbb{Z}^2$ . L'étude de ces points fixes et des différents motifs qui y apparaissent est une question encore à l'étude.

Ainsi, dans [Jam04], il est montré que les motifs  $\begin{pmatrix} 3 & 2 \\ 1 & 1 \end{pmatrix}$  et  $\begin{pmatrix} 2 & 2 \\ 3 & 2 \end{pmatrix}$  sont interdits dans ces mots bidimensionnels.

Savoir si les substitutions généralisées peuvent toutes être intégralement définies à partir de telles règles locales est une question d'actualité. Un travail de Mozès va dans ce sens [Moz89], s'appuyant sur des résultats de Goodman-Strauss [GS98]. Tout récemment, J.-M. Gambaudo aurait démontré que tout pavage linéairement répétitif (dont la fonction de récurrence est sous-linéaire) serait à règles locales, c'est-à-dire qu'il existe une coloration des pièces du pavage telle que l'espace de pavage associé est un sous-shift de type fini. Or, les plans discrets entrent dans le cadre des pavages. Ce résultat impliquerait que tous les plans discrets linéairement répétitifs sont à règles locales. Il reste alors à caractériser les plans discrets linéairement répétitifs, en particulier, comprendre le lien entre cette propriété et le fait que les coefficients du développement en fractions continues (de Jacobi-Perron par exemple)



sont bornés. Il reste aussi à proposer des jeux de règles locales explicites dans ce cas.

### 3.3.6 Engendrer des plans discrets en s'appuyant sur des règles locales

Les règles locales ont été définies précédemment dans le cas des mots bidimensionnels. Cependant, rien n'empêche de les définir plus globalement dans un espace  $\mathbb{Z}^n$ , en particulier dans l'ensemble des pièces de la forme  $[\mathbf{x}, i^*]$  de  $\mathbb{R}^3$  qui permettent de décrire les plans discrets de  $\mathbb{R}^3$ .

**Admissibilité d'un motif** La question sous-jacente à l'approche par règles locales est la notion d'admissibilité de motif. Plus précisément, étant donné un motif discret fini de la forme  $\mathfrak{U} = \cup[\mathbf{y}, i^*]$ , ce motif peut-il apparaître dans un plan discret  $\Gamma_{\mathbf{x}}$ , à translation près ?

On peut répondre à cette question en utilisant la caractérisation arithmétique donnée par la formule (3.5) : selon cette caractérisation, on recherche un vecteur de translation  $\mathbf{y}_0$  tel que pour chaque pièce  $[\mathbf{y}, i^*]$  de  $\mathfrak{U}$ , l'intervalle  $[0, \langle \mathbf{x}, \mathbf{e}_i \rangle]$  contient la quantité  $\langle \mathbf{y} + \mathbf{y}_0, \mathbf{x} \rangle$ . On est ainsi ramené à vérifier qu'une intersection d'intervalles est non vide, ce qui se vérifie avec des approches de contraintes usuelles, par exemple en Prolog.

**Calcul des règles locales pour un jeu de substitutions généralisées** La notion d'admissibilité de motif permet d'identifier tous les motifs connexes contenant deux pièces qui apparaissent dans un plan discret (on généralise ainsi les travaux de [Jam04]). Par connexe on entend des pièces qui partagent au moins une face. Ces motifs sont appelés *primaires*.

On considère maintenant un jeu de substitutions généralisées.

On appelle *jeu de règles locales* un ensemble de motifs primaires minimal pour l'inclusion tel que

- Les images de pièces de base  $[\mathbf{0}, i^*]$  par toutes les substitutions considérées sont recouvertes par des motifs du jeu de règles locales.
- Les images des éléments du jeu sont elles-mêmes recouvertes par des règles locales du jeu.

À partir de ces deux propriétés, on est certain d'engendrer les itérés des pièces de base par toutes les substitutions. Si on veut itérer à partir d'un motif de base différent, dont on est certain qu'il apparaît dans au moins un plan discret par la condition d'admissibilité précédente, il suffit de rechercher un jeu de règles locales qui recouvre ce motif initialement.

Notons bien qu'il n'est pas nécessaire de vérifier dans ce cadre que les règles locales sont compatibles parce que les propriétés des substitutions généralisées nous assurent que dans un plan discret, les positionnements d'images de lettres sont uniques.

Il existe bien entendu plusieurs jeu de règles locales. Pour en trouver un, une approche gloutonne est suffisante (recouvrir l'image de  $[\mathbf{0}, 1^*]$  par des règles locales,

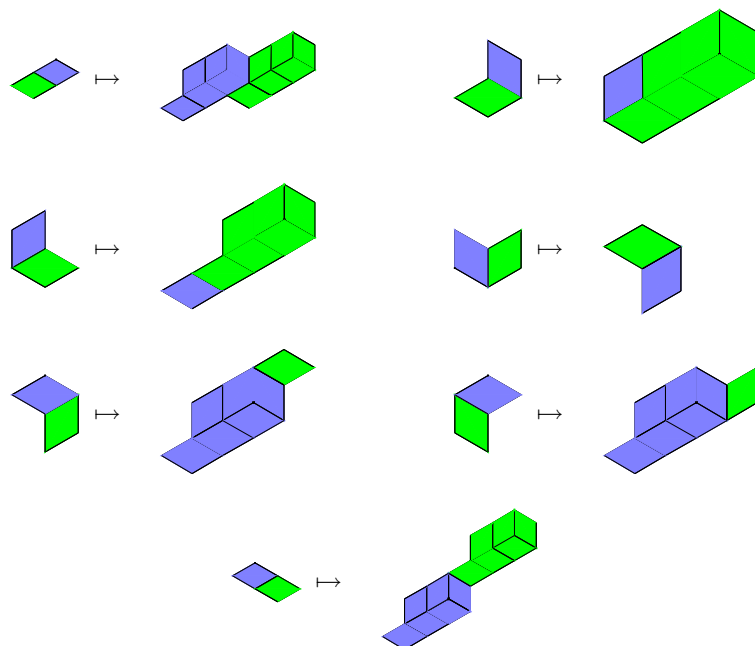


FIG. 3.11 – Règles locales pour la substitution généralisée de la substitution de Jacobi-Perron  $\sigma_{(2,3)}$ . Ces règles donnent les positions relatives des images de face à l'intérieur d'une paire. Elles sont suffisantes pour itérer la substitution généralisée  $\Sigma_{1,2}$ .

puis les images de ces règles locales par de nouvelles règles, jusqu'à ne plus produire de nouvelles règles).

Ainsi, pour la substitution de Jacobi-Perron  $\sigma_{(2,3)}$ , on obtient le jeu de règles locales montré Fig. 3.11.

**La stratégie d'engendrement** Dans (Jour. Montoises, 2006), P. Arnoux, V. Berthé et moi-même proposons une approche pour généraliser l'approche japonaise de [IO94]. Nous montrons que pour engendrer une famille de plans discrets, il faut superposer deux propriétés bien distinctes :

1. Une propriété qui caractérise les développements tels que la structure topologique de certains anneaux est respectée (propriété d'encadrement), un peu plus contrainte que le lemme de l'anneau énoncé dans [IO94]. Cette propriété remplace la propriété de finitude (F).
2. Une propriété qui caractérise les développements qui produisent un anneau sur lequel la propriété d'engendrement s'applique (propriété de génération). Dans [IO94], la propriété d'engendrement étant donnée par le lemme de l'anneau,

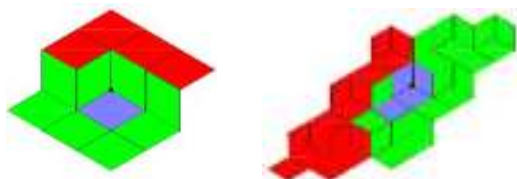


FIG. 3.12 – L’anneau présenté à la figure 3.9 n’était pas recouvert par règles locales pour les substitutions de Jacobi-Perron. Par contre il est inclus dans un anneau recouvert par règles locales, dont l’image est elle-même recouverte par règles locales.

il suffisait de caractériser les développements pour lesquels le cube unité était totalement entouré.

Nous pouvons maintenant détailler deux résultats qui permet de concrétiser cette approche.

**Partie (1). Le lemme de l’anneau couvert par des règles locales** Dans notre travail en cours, V. Berthé, J. Bourdon et moi-même introduisons une nouvelle définition pour les anneaux, qui intègrent la combinatoire de la substitution :

**Définition 3.3.3 ((B., B. & S., en cours))** *un sous-ensemble  $\mathfrak{A}$  d’un plan discret est un anneau couvert par un jeu de règles locales s’il existe une suite de motifs primaires dans le jeu de règles locales qui recouvrent l’anneau et tels que chaque motif intersecte exactement deux autres motifs du recouvrement.*

Avec cette notion qui prend en compte la combinatoire des substitutions, nous montrons dans (B., B. & S., en cours) que les substitutions généralisées respectent les anneaux. Cette proposition est illustrée à la figure 3.12.

**Proposition 3.3.4 (Lemme de l’anneau couvert par règles locales, (B., B. & S., en cours))** *Soit  $\mathfrak{A}$  anneau recouvert par les règles locales associées aux substitutions généralisées des  $\sigma_{B,C}$ . Alors les images de  $\mathfrak{A}$  par les substitutions généralisées des  $\sigma_{B,C}$  sont des anneaux recouverts par des règles locales.*

Il faut noter que la preuve de ce résultat n’est pas purement topologique mais fait intervenir la *combinatoire* des substitutions. Il s’agit de considérer les préimages de chaque règle locale du jeu et vérifier qu’elles ne se déconnectent pas, par une étude de cas. Le point principal est que cette vérification est automatique et se fait en un temps fini.

**Graphe de génération** À partir du nouveau lemme de l’anneau, on peut construire un graphe qui caractérise tous les développements qui produisent effectivement une configuration entourant le cube unité par un anneau recouvert par règles locales.

**Définition 3.3.5** *On considère un ensemble de substitutions  $\sigma_i$  et un jeu de règles locales pour leurs substitutions généralisées qui vérifie le lemme de l'anneau couvert par règles locales.*

*Le graphe d'engendrement de ces substitutions est défini comme suit :*

- *Les configurations terminales du graphe sont tous les anneaux couverts par règles locales qui contiennent le cube unité  $[0, 1^*] \cup [0, 2^*] \cup [0, 3^*]$  et ne contiennent pas d'autre anneau de ce type.*
- *Les sommets sont obtenus itérativement en considérant les images successives du cube unité par les substitutions généralisées des  $\sigma_i$  et en les intersectant avec l'ensemble des configurations terminales.*
- *Il existe un arc entre deux sommets  $\mathfrak{B}$  et  $\mathfrak{C}$  si et seulement s'il existe une substitution généralisée  $\mathbf{E}_1(\sigma_i)^*$  et une configuration terminale  $\mathfrak{A}$  telle que  $\mathbf{E}_1(\sigma_i)^*(\mathfrak{C}) \cap \mathfrak{A} \subset \mathfrak{B}$ . Cet arc est étiqueté par  $i$ .*

Il faut noter ici que le sens des arcs est inversé par rapport à la notion de substitution : les flèches représentent les préimages des motifs par une substitution et non leurs images.

Ce graphe est fini puisqu'il s'agit de sous-motifs d'un nombre fini de motifs (les configurations terminales). D'un point de vue algorithmique, la difficulté consiste à identifier l'ensemble des configurations terminales pour le jeu de règles locales. Pour cela, on utilise la notion de motif admissible défini plus haut et un critère topologique pour s'assurer que les configurations considérées sont bien minimales. Il s'agit du point le plus long en terme de calcul dans cette approche.

**Engendrement de plans discrets** À partir du graphe de génération et du lemme de l'anneau couvert par règles locales, on détermine les développements qui permettent d'engendrer effectivement un plan discret.

**Théorème 3.3.6 (B., B. & S., en cours)** *Soit  $\sigma_i$  des substitutions associées à un algorithme de fractions continues en dimension 3. On considère un jeu de règles locales pour lesquelles les substitutions généralisées vérifient le lemme de l'anneau par règles locales.*

*On considère un vecteur  $\mathbf{x}$  dont le développement en fractions continues est décrit par les indices  $i_0, \dots, i_n, \dots$ . Si le développement est tel que pour tout  $n$  il existe  $n'$  tel que  $i_n i_{n+1} \dots i_{n+n'}$  indice un chemin qui part d'une configuration terminale vers le cube unité, alors le plan discret  $\Gamma_{\mathbf{x}}$  est engendré à partir du cube unité :*

$$\Gamma_{\mathbf{x}} = \cup_{n \geq 1} \mathbf{E}(\sigma_{i_1})^* \dots \mathbf{E}(\sigma_{i_n})^* ([0, 1^*] \cup [0, 2^*] \cup [0, 3^*]).$$

**Premières constructions** Nous avons vérifié que cette approche généralise exactement le résultat de [IO94] dans la mesure où le graphe d'engendrement relatif aux configurations terminales pour le lemme de l'anneau donné dans [IO94] est exactement celui qui est décrit dans [IO94]. De plus, la condition (3.9) décrit exactement les chemins du graphe qui partent d'une configuration terminale mais ne reviennent jamais vers le cube unité.

### 3.3.7 Perspectives

Nous avons finalement proposé une approche qui permet de décrire les développements en fractions continues de vecteurs en dimension 3 pour lesquels le plan discret qui leur est orthogonal est engendré à partir du cube unité. Cette approche combine des vérifications de contraintes, des constructions de graphes et des vérifications topologiques. À partir de cette approche théorique, plusieurs perspectives sont à l'étude.

**Comparaison de différents algorithmes de fractions continues** La première application de cette méthode va être de comparer les graphes pour différents algorithmes de fractions continues : Jacobi-Perron, Brun et Arnoux-Rauzy. En particulier, nous souhaitons vérifier si, pour ces différents algorithmes, les plans discrets dont le vecteur normal a un développement en fractions continues avec des quotients bornés peuvent être engendrés à partir du cube unité.

Pour les plans discrets qui ne sont pas engendrés à partir du cube unité, la question qui se posera sera de trouver un motif plus gros que le cube unité qui permet de les engendrer.

**Topologie des motifs qui engendrent les plans** Un objectif sera de choisir le meilleur algorithme de génération de plan discret en fonction de la topologie des morceaux engendrés. Dans le cas substitutif Pisot, la non-connexité du fractal de Rauzy implique que les itérations successives ne sont pas connexes à partir d'un certain rang. Ainsi, il sera intéressant de déterminer quels sont les meilleurs algorithmes de fractions continues qui engendrent des plans discrets à partir de pièces au moins connexes, si possible simplement connexes, et avec un volume équitablement réparti.

En relation avec cette question, des approches basées sur les substitutions généralisées et l'algorithme de fractions continues de Brun fournissent des bases pour la reconnaissance de plans discrets [ABFJ07, BF08, Fer08]. La reconnaissance de la topologie de ces motifs est maintenant une question qu'on peut espérer comprendre en combinant les approches topologiques présentées au début de ce chapitre avec les questions d'engendrement de plans discrets décrites ici.

#### **Familles de systèmes substitutifs isomorphes à une translation sur un tore**

Avec cette approche, nous pourrons vérifier la conjecture d'Arnoux-Ito [AI01] selon laquelle toutes les substitutions obtenues par trois substitutions de bases dites d'Arnoux-Rauzy engendrent un système symbolique isomorphe en mesure à une translation sur un tore ; on sait avec [AI01] que ces substitutions sont Pisot irréductibles et unitaires, il reste à montrer que leur fractal de Rauzy engendre un pavage. On considérera aussi les développements qui engendrent des suites unidimensionnelles qui ne sont pas uniformément équilibrées [CFZ00], pour mieux comprendre les propriétés des plans discrets associés.

### Construction de fractals de Rauzy pour des paramètres non rationnels

Dans le cas Pisot substitutif, les pièces qui engendrent un plan discret peuvent être renormalisées pour engendrer le fractal de Rauzy de la substitution, c'est-à-dire un candidat pour un domaine fondamental d'un réseau qui code l'action d'une translation par le système substitutif. Dans le cas général, les fractions continues devraient permettre de construire une suites de domaines fondamentaux qui permettraient de représenter symboliquement la dynamique n'importe quelle addition sur un tore par la composition de substitutions (conjecture  $S$ -adique).

## 3.4 Invariants pour des systèmes dynamiques

En dernière application des fractals de Rauzy se trouvent la continuation du travail sur le codage des systèmes dynamiques et la définitions d'invariants.

### 3.4.1 Dynamique des systèmes substitutifs non unitaires

Comme mentionné dans les perspectives du chapitre 1, si le cas Pisot irréductible unitaire commence à être correctement appréhendé, les cas non unitaire et non Pisot sont pour l'instant simplement ébauchés. Les approches d'engendrement de plans discrets décrites dans la section 3.3 sont une piste pour traiter le cas unitaire non Pisot. Au sujet des substitutions non-unitaires, on s'inspire des approches développées pour la beta-numération non unitaire (section 3.2) et à des graphes de pavages pour démontrer des résultats non triviaux.

**Fractal de Rauzy  $p$ -adique pour toutes les substitution** La construction du fractal de Rauzy complet décrites pour les beta-numérations se généralise à toutes les substitutions non unitaires. En effet, on considère une substitution dont la matrice d'incidence n'est pas de déterminant 1. Pour ces substitutions, on peut construire un fractal de Rauzy euclidien ; pour les raisons détaillées dans la section 3.2.1, les sous-tuiles de ce fractal s'intersectent sur un ensemble de mesure non nulle.

En exploitant le lien entre le système symbolique et les chemins de l'automate des préfixes-suffixes donné dans le théorème 2.2.2 ainsi que la représentation des fractals de Rauzy par numération de Dumont-Thomas explicitée dans le théorème 2.3.2, on étend la définition du fractal de Rauzy euclidien à un fractal de Rauzy complet qui vit dans un espace  $p$ -adique. Les formalismes sont identiques à ceux décrits pour les beta-numérations non unitaires (section 3.2) ; je les ai détaillés dans la publication (Erg. Th. Dyn. Sys, 2003).

**Echange de morceaux dans des espaces  $p$ -adiques** J'ai montré que les résultats dynamiques connus dans le cas unitaire se généralisent au cas non unitaire : en particulier, un système substitutif code les trajectoires d'un échange de morceaux dans un espace un peu surprenant.

**Théorème 3.4.1 (Erg. Th. Dyn. Sys, 2003)** *Soit  $\sigma$  une substitution irréductible Pisot sur  $d$  lettres, qui vérifie la condition de fortes coïncidences. Soit  $\mathbf{M}_\sigma$  la matrice*

d'incidence de  $\sigma$  et  $\beta$  la valeur propre dominante de  $\mathbf{M}_\sigma$ . Pour  $p \in \mathbb{N}$  premier, on note  $D_p = \sum_{i=1}^k e(\mathcal{J}_i)$ , où  $e(\mathcal{J}_i)$  désigne l'indice de ramification de chaque diviseur premier  $\mathcal{J}_i$  de l'idéal  $\beta \mathcal{O}_{\mathbb{Q}(\alpha)} + p \mathcal{O}_{\mathbb{Q}(\beta)} \subset \mathbb{Q}(\beta)$ .

Le système dynamique symbolique  $(X_\sigma, S)$  associé à  $\sigma$  est isomorphe en mesure à un échange de morceaux sur une partie autosimilaire compacte de mesure non nulle dans  $\mathbb{R}^{d-1} \times \mathbb{Q}_{p_1}^{D_{p_1}} \times \dots \times \mathbb{Q}_{p_k}^{D_{p_k}}$ , où les  $p_i$  sont les diviseurs premiers de  $\det \mathbf{M}_\sigma$ .

Avec des méthodes identiques au cas unitaire, on peut aller plus loin dans l'étude de la dynamique du système substitutif : l'échange de morceaux peut se quotienter sur un groupe abélien compact en une translation.

**Proposition 3.4.2 (Erg. Th. Dyn. Sys, 2003)** *Tout système substitutif de type Pisot est une extension finie presque partout de son facteur équicontinu maximal, qui est une translation sur un groupe compact.*

**Graphes de pavage et systèmes substitutifs à spectres discrets** La question importante d'un point de vue dynamique est de déterminer si cette opération de quotient est injective, c'est-à-dire si l'extension est en fait un isomorphisme en mesure. Pour cela, j'ai généralisé dans (Ann. Inst. Fourier, 2004) la définition des graphes de pavages au cas non unitaire, et j'ai produit des graphes de pavage associé à un réseau périodique de  $\mathbb{K}_\sigma$ . Cette méthode, algorithmique, permet de montrer que des systèmes substitutifs non unitaire sont effectivement isomorphes à une translation donnée sur un groupe compact, et donc à spectre purement discret.

**Théorème 3.4.3 (Ann. Inst. Fourier, 2004)** *Le système dynamique associé à la substitution non unitaire  $\sigma(1) = 2$ ,  $\sigma(2) = 3$ ,  $\sigma(3) = 11233$  a un spectre purement discret.*

Avant cette approche, les seuls résultats connus sur le spectre discret des systèmes substitutifs non unitaires concernaient les substitutions sur deux lettres [BD02]. On voit ici que les graphes de pavages peuvent être exploités dans des contextes très généraux.

### 3.4.2 Vers des partitions de Markov pour des endomorphismes du tore

Cette section est plus prospective. Nous y discutons de pistes pour construire des partitions de Markov pour des endomorphismes du tore.

#### Partition de Markov pour l'action d'une matrice sur des corps locaux

On considère donc une substitution Pisot irréductible *non unitaire* sur  $n$  lettres. Comme nous venons de le rappeler, nous pouvons définir un fractal de Rauzy, qui est un sous-ensemble de l'espace complet de représentation  $\mathbb{K}_\beta$ . La matrice de la substitution  $\mathbf{M}$  agit naturellement sur cet espace (via la multiplication par la représentation de  $\beta$  dans chaque composante de l'espace  $\mathbb{K}_\beta$ ). Sous des conditions

naturelles (qui généralisent la condition Pisot) données dans (Erg. Th. Dyn. Sys, 2003), la matrice agit aussi naturellement sur un groupe compact obtenu comment quotient de  $\mathbb{K}_\beta$ . Comme pour le cas unitaire, l'extension naturelle introduite à la section 3.2 est alors un bon candidat pour une partition de Markov pour cette action. Il faut pour cela qu'une extension de la condition de pavage exprimée par des graphes de frontière soit vérifiée, comme illustré au théorème 3.4.3 (voir (Ann. Inst. Fourier, 2004) et (Erg. Th. Dyn. Sys, 2003)).

La différence fondamentale entre les cas unitaire et non-unitaire est que cette partition de Markov décrit l'action de la matrice sur l'espace de représentation complet  $\mathbb{K}_\beta$ , qui inclut des parties non-archimédiennes. Dans cet espace, l'action de  $\mathbf{M}$  est inversible et peut donc être décrite par un sous-shift de type fini inversible : il va s'agir du décalage sur l'ensemble des étiquettes de chemins bi-infinis du graphe des préfixes-suffixes (notons bien que sur un ensemble de mots bi-infinis, ce décalage est bien inversible). Rajouter les composantes non-archimédiennes nous ramène ainsi à un cas équivalent à l'action d'une matrice inversible sur un tore.

**Partition de Markov pour un endomorphisme du tore ?** Dans la réalité, on recherche une partition de Markov pour l'action de la matrice  $\mathbf{M}$  sur le *tore*  $\mathbb{T}^n$ . Cette deuxième action n'est plus inversible. Un domaine de Markov doit permettre de représenter cette action par un décalage sur un sous-shift de type fini unilatéral : puisque l'action de  $\mathbf{M}$  n'est plus inversible, sa représentation symbolique ne le sera plus, ce qui est le cas d'un décalage sur des mots infinis à droite seulement. Nous recherchons donc un système symbolique de type fini unilatéral qui est isomorphe en mesure à l'action de la matrice.

Or, l'action de la matrice  $\mathbf{M}$  sur le tore est un facteur topologique de son action sur l'espace  $\mathbb{K}_\beta$ . Cependant, la projection de  $\mathbb{K}_\beta$  sur  $\mathbb{T}^n$  ne respecte pas la représentation du fractal de Rauzy complet par le graphe des préfixes-suffixes ; la question de trouver une bonne partition de Markov reste donc entière.

Géométriquement, on cherche à construire un domaine fondamental de l'espace contractant  $\mathbf{H}_c$  de  $\mathbb{R}^n$  qui respecte l'action de  $\mathbf{M}$ . Et on cherche à obtenir ce domaine fondamental à partir du fractal de Rauzy complet  $\tilde{\mathcal{T}}$ . Comme nous l'avons déjà vu, le fractal de Rauzy dans  $\mathbf{H}_c$  existe bien, mais ses sous-tuiles ne sont pas disjointes : il faut les séparer par des corps locaux pour arriver à les rendre disjointes.

On peut voir cette question autrement : le fractal de Rauzy euclidien recouvre plusieurs domaines fondamentaux du réseau de  $\mathbf{H}_c$  qui représente la section de l'action de  $\mathbf{M}$ , et il faut parvenir à en sélectionner un sous-ensemble représente exactement un domaine fondamental. Autrement dit, la question est de sélectionner pour chaque élément  $\mathbf{x}$  de  $\mathbf{H}_c$  un unique point de  $\mathbb{K}_\beta$  dont la partie euclidienne coïncide avec  $\mathbf{x}$ .

**Recherche d'un candidat** Un bon candidat pour ces points consiste à représenter la partie non-archimédienne de l'espace de représentation comme une limite inverse d'anneaux d'entiers. Dans ce cadre, l'élément  $z = \langle \mathbf{x}, \mathbf{v}_\beta \rangle$  considéré au départ admet plusieurs préimages dans le quotient  $\mathcal{O}(\mathbb{Q}(\beta))/\beta\mathcal{O}(\mathbb{Q}(\beta))$ . On peut sélectionner la seule préimage  $z_1$  qui appartient  $\mathbb{Z}[\beta]$ , puis recommencer cette opération. La



suite des représentations dans l'espace complet des  $z_n$ , notée  $\phi_\beta(z_n)$  converge alors dans le fractal de Rauzy complet vers un point  $z_\infty$  qui admet  $z$  pour représentation euclidienne.

Intuitivement, cela revient à sélectionner dans le fractal de Rauzy représenté Fig. 3.3 la ligne horizontale d'ordonnée  $y = 1$ . Formellement, on peut définir ce nouvel ensemble en considérant une généralisation des beta-numérations.

**Une généralisation des beta-numérations : shift radix systems** Les SRS (shift radix systems) sont définis comme suit dans [ABB<sup>+</sup>05]. On considère  $\mathbf{r} \in \mathbb{R}^d$  et on définit la fonction  $\tau_{\mathbf{r}} : \mathbf{z} = (z_0, \dots, z_{d-1}) \in \mathbb{Z}^d \rightarrow (z_1, \dots, z_{d-1}, -\lfloor \langle \mathbf{r}, \mathbf{z} \rangle \rfloor) \in \mathbb{Z}^d$ . L'application  $\tau_{\mathbf{r}}$  est un *SRS* si pour tout  $\mathbf{z} \in \mathbb{Z}^d$  il existe un entier  $k \in \mathbb{N}$  tel que  $\tau_{\mathbf{r}}^k(\mathbf{z}) = \mathbf{0}$ .

Ces systèmes de numération sont des généralisations des beta-numérations dans le cas unitaire [ABB<sup>+</sup>05]. En effet, si on considère un nombre  $\beta$  et si on développe son polynôme minimal sous la forme  $(X - \beta)(X^{d+r_{d-1}}X^{d-1} + r_{d-2}X^{d-2} + \dots + r_0)$ , le shift radix system associé au vecteur  $\mathbf{r} = (r_0, \dots, r_{d-1})$  coïncide avec la dynamique de la beta-transformation dans le cas unitaire.

Les propriétés arithmétiques des SRS ont été intensivement étudiées [ABB<sup>+</sup>05, ABPT06, ABPT08]. En particulier, on leur associe une tuile centrale, qui correspond à rechercher les préimages dans  $\mathbb{Z}^n$  de  $\mathbf{0}$  et à les renormaliser par la matrice compagnon du vecteur  $\mathbf{r}$ . Là encore, on montre que, pour les SRS associé au vecteur  $\mathbf{r}$  qui correspond à un  $\beta$  unitaire, la tuile ainsi obtenue coïncide avec un fractal de Rauzy [ABB<sup>+</sup>05].

**Un cas plus spécifique : le cas non unitaire** Cependant, dans le cas non unitaire, les applications  $\tau_{\mathbf{r}}$  apparaissent comme une restriction injective de la dynamique de la beta-transformation ; en fait, cette dynamique procède exactement à la sélection de points dans  $\mathcal{O}(\mathbb{Q}(\beta))/\beta\mathcal{O}(\mathbb{Q}(\beta))$  : nous démontrons ceci dans un travail en cours de rédaction avec Paul Surer, Joerg Thuswaldner et Valérie Berthé (B., S., S. & T., en cours).

L'intérêt de ce processus de sélection est que, avec les SRS, on peut, y compris dans le cas non unitaire, définir un ensemble compact qui représente les entiers pour ces systèmes de numération. Par contre, contrairement au cas unitaire, l'ensemble compact ainsi obtenu ne coïncide pas avec le fractal de Rauzy euclidien : il s'agit d'un sous-ensemble de ce fractal euclidien.

Une manière de visualiser cet ensemble est la suivante. On considère le fractal de Rauzy complet (incluant les parties  $p$ -adique associé au nombre  $\beta$ . Chaque composante  $p$ -adique de ce fractal se plonge dans  $[0, 1[$  par l'application  $\sum a_i p^i \in \mathbb{Z}_p \rightarrow \sum a_i p^{-i} \in [0, 1[$ . Comme nous l'avons déjà mentionné, ce plongement respecte la mesure du fractal, ainsi que ses parties compactes. L'image du fractal complet dans le fractal est alors un sous ensemble de  $\mathbb{R}^{d-1}$  croisé avec un produits d'intervalles  $[0, 1]^N$ . Le fractal obtenu avec les SRS correspond aux éléments  $\mathbf{x}$  de  $\mathbb{R}^{d-1}$  tels que  $(\mathbf{x}, 1, \dots, 1)$  appartient au plongement du fractal. Autrement dit, ce sont les préimages dans le fractal euclidien des éléments qui se trouvent au bout de  $\mathbb{Z}_p$ .

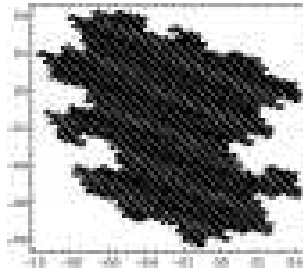


FIG. 3.13 – Tuile centrale pour le *shift radix system* associée à la beta-substitution du nombre de Pisot non unitaire  $\beta^3 = 7\beta^2 + 3\beta + 2$ .

Les propriétés de ce nouveau compact sont complexes. Nous avons montré que  $\mathbf{0}$  est un point intérieur de ce fractal. Par contre, il n'est plus autosimilaire mais décrit par un graphe infini. La dimension de Hausdorff de son bord n'a pas encore été calculée.

**Une nouvelle propriété de pavage** La propriété intéressante de cette tuile concerne à nouveau des aspects de pavages. En effet, graphiquement, pour les exemples considérés, il semble que ce fractal induise un pavage du plan par un nombre de tuiles infinies liées au système de numération. En collaboration avec Paul Surer, Joerg Thuswaldner et Valérie Berthé, nous avons pu montrer que le fractal induit effectivement un pavage (B., S., S. & T., en cours). La méthode consiste à estimer numériquement la taille d'approximations polygonales de la frontière. Or, il s'avère que ces deux propriétés sont disjointes. Pour vérifier la propriété de pavage, nous avons introduit un nouvel algorithme basé sur le recouvrement de la tuile par des réseaux.

**Théorème 3.4.4 (B., S., S. & T., en cours)** Soit  $\beta$  défini par  $\beta^3 = 7\beta^2 - 3\beta + 2$ . Les tuiles associées au *shift radix system* défini à partir de  $\beta$  forment un pavage du plan  $\mathbb{R}^2$ .

**Perspective : base d'une partition de Markov ?** La tuile centrale pour les SRS est donc un sous-ensemble compact du fractal de Rauzy euclidien qui induit des propriétés de pavages intéressantes dans le plan et est obtenu comme limite inverse de la dynamique. Selon moi, il s'agit d'un candidat naturel pour la base d'une partition de Markov. Les travaux futurs dans cette direction consistent à vérifier si ce nouveau fractal est le domaine fondamental d'un réseau. Si tel est le cas, il y a de bons espoirs pour que l'action de la matrice compagnon du vecteur  $\mathbf{r}$  initialement considéré admette cette structure pour partition de Markov.

**Perspective : étude de la topologie des nouveaux fractals** Les nouveaux ensembles obtenus ne sont en fait pas autosimilaires mais décrits par un graphe infini.

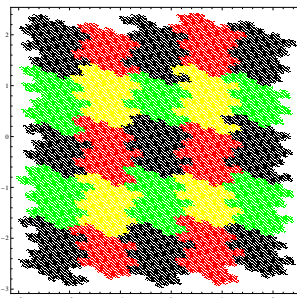


FIG. 3.14 – Pavage associé au *shift radix system* de la beta-substitution de  $\beta = 3$ . Contrairement au cas des fractals de Rauzy, les tuiles de ce pavage ne sont pas en nombre fini : elles ont une infinité de formes.

La question maintenant est de décrire les propriétés topologiques de cet ensemble, en particulier sa frontière, sans doute à l'aide de graphes infinis.

### 3.4.3 Dynamique symbolique des homéomorphismes de surfaces. Automorphismes de groupes libres

**Des substitutions aux automorphismes du groupe libre** Une dernière application de la dynamique symbolique concerne les automorphismes du groupe libre. En effet, il est assez naturel d'étendre la notion de substitution à la notion d'endomorphisme du groupe libre. Parmi ces endomorphismes se trouvent ceux qui sont inversibles ; on parle alors d'*automorphisme* du groupe libre. Les substitutions qui se prolongent en un automorphisme du groupe libre sont appelées inversibles, et leurs spécificités sont bien étudiées, en particulier sur les alphabets à deux lettres.

**De la géométrie vers la symbolique** L'intérêt porté aux automorphismes du groupe libre vient de l'étude des difféomorphismes des surfaces par Nielsen et ensuite Thurston [Thu88, CB88]. En effet, tout difféomorphisme d'une surface  $f : M_g \rightarrow M_g$  engendre un automorphisme du groupe fondamental de la surface, qui est un groupe libre dès que la surface est orientable et a un bord. Notons que les automorphismes qui sont ainsi obtenus sont assez spécifiques ; ils sont appelés *géométriques*. Si tous les automorphismes du groupe  $\mathbb{F}_2$  sont géométriques, aucun automorphisme d'un groupe libre de rang impair ne peut coder un homéomorphisme de surface (Ann. Inst. Fourier, 2006).

Pour étudier la dynamique de l'automorphisme à isotopie près, Thurston a introduit la notion de graphe plongé dans la surface (*train track*). Ces objets géométriques sont structurés autour de chaque sommet par un ordre cyclique qui provient du fait que la surface est localement plane. On peut alors considérer un feuilletage mesurable à partir de poids qui respectent la structure du graphe sur les arêtes.

**Automorphismes versus substitutions** La dynamique de ces trains-track et plus généralement des automorphismes du groupe libre est difficile à appréhender. Pour mieux la comprendre, on oublie le point de vue géométrique et on considère un automorphisme du groupe libre d'un point de vue purement symbolique. Or, même du point de vue symbolique, l'étude des automorphismes est délicate. En effet, l'étude des systèmes symboliques engendrés par une substitution est basée sur l'existence d'un mot infini périodique pour la substitution ; ce point périodique existe pour la seule raison que la longueur de toutes les images de lettres croît vers l'infini. Or, cette propriété n'est plus vérifiée par les automorphismes du groupe libre puisque des annulations apparaissent (par exemple, la conjugaison fixe le mot fini  $i_w : a \mapsto waw^{-1}$  et ses itérations successives ne font donc pas croître sa longueur).

**Une condition suffisante pour limiter les annulations** Pour régler ce problème, Bestvina et Handel ont proposé de coder un automorphisme  $\Phi$  du groupe libre par un représentant topologique, plus formellement une application sur un graphe marqué  $f : G \rightarrow G$  qui induit l'automorphisme  $\Phi$  sur le groupe fondamental de  $G$ , en l'occurrence un groupe libre  $F_n$ . Un tel représentant est appelé un *train-track* si les itérations de  $f$  ne contiennent aucune annulation. Elles se comportent donc comme une substitution.

Bestvina et Handel se concentrent sur les équivalents algébriques des homéomorphismes pseudo-Anosov des surfaces, appelés *iwip* (irreducible with irreducible power). Formellement, ces automorphismes sont tels qu'aucun facteur libre n'est envoyé par l'automorphisme et ses puissances sur un conjugué de ce facteur.

Le résultat fondamental dit que tout automorphisme iwip, à conjugaison par un automorphisme intérieur près, admet un *train-track* pour représentant topologique [BFH97]. La construction de ce *train-track* est en plus proposée sous une forme proche d'un algorithme.

**Dynamique d'un automorphisme du groupe libre** De ce point de vue géométrique, on obtient que les points intéressants à étudier pour la dynamique d'un automorphisme du groupe libre sont les points fixes qui apparaissent sur le bord du groupe libre.

En effet, pour les automorphismes iwip, il devient possible de construire un mot bi-infini sur lequel l'automorphisme agit sans annulation et qui est fixé par une puissance de l'automorphisme. À partir de ce mot bi-infini, on construit un système dynamique [BFH00], appelé la *lamination attractive symbolique*, obtenue comme la fermeture de l'orbite du point bi-infini sous l'action du décalage.

Ainsi, une lamination attractive symbolique peut être considérée de deux points de vue [CHL06] : algébriquement, il s'agit d'un ensemble de géodésiques qui est fermé pour la topologie induite sur le bord du groupe libre, qui est aussi invariante pour l'action du groupe libre et invariante par renversement d'orientation. Symboliquement, il s'agit d'un système dynamique symbolique contenant des mots bi-infinis et invariant sous l'action du décalage et de la symétrie par rapport à l'indice 0.

**Lien avec les substitutions** Dans (Ann. Inst. Fourier, 2006), avec V. Berthé, P. Arnoux et A. Hilion, nous nous intéressons aux laminations attractives d'un point de vue purement symbolique. Nous étudions d'abord le lien entre la propriété iwip et la propriété de primitivité. Il s'avère que, même si elles sont reliées, ces deux propriétés ne sont pas équivalentes.

Nous montrons ensuite que la lamination attractive d'un automorphisme extérieur iwip dans les coordonnées données par un représentant train-track  $f$  est un *système dynamique symbolique associé à une substitution double*  $\sigma_f$ , obtenue en doublant l'alphabet constitué par les arêtes sur lequel le train-track est défini.

En particulier, nous montrons que ce système est *minimal* et invariant par le flip si et seulement si la substitution double  $\sigma_f$  est non-orientable, c'est-à-dire qu'il n'existe pas de sous-ensemble de l'alphabet qui le partitionne en deux et est stable ou disjoint de son image par la substitution.

Ainsi, tout automorphisme extérieur  $\Phi$  iwip non-orientable contient une dynamique substitutive minimale, qui représente la lamination attractive de  $\Phi$  du bord du groupe libre.

**Représentation par une addition** Comme nous l'avons expliqué au chapitre 2, les systèmes symboliques engendrés par une substitution sont de bons candidats pour être à spectre purement discret, c'est-à-dire isomorphes à une addition sur un tore, en particulier lorsque leur valeur propre dominante est un nombre de Pisot. Pour les substitutions doubles associées à un train-track, la valeur propre dominante est le coefficient de dilatation  $\lambda$  de  $\Phi$ . Lorsque celui-ci est Pisot, on peut donc construire un fractal de Rauzy, qui admet des symétries très particulières et est stable par un échange de morceaux (voir (Ann. Inst. Fourier, 2006) pour les détails).

Dans ce cas là, on sait surtout que la substitution associée vérifie la condition de quotient. Nous avons donc utilisé les graphes de pavages pour vérifier si le système symbolique considéré est conjugué à une addition sur un tore.

**Théorème 3.4.5 (Ann. Inst. Fourier, 2006)** *Les laminations attractives de train-tracks fixés pour  $\phi_1 : a \mapsto c, b \mapsto c^{-1}a, c \mapsto b$  et  $\phi_2 : a \mapsto c, b \mapsto c^{-1}a, c \mapsto bc^{-1}$  sont à spectre discret. Elles sont isomorphes en mesure à une translation sur le tore de dimension 2.*

On montre ainsi une propriété métrique de certaines laminations attractives associées à un automorphisme du groupe libre. Dans la lignée de [CHL06], la question reste entière pour comprendre ce que ces propriétés signifient sur la topologie et les propriétés des arbres réels engendrés par un automorphisme du groupe libre.

**Recherche d'invariants topologique à une conjugaison près** Un des défauts de l'approche proposée est que nous construisons un fractal de Rauzy à partir d'un représentant train-track. Or, il existe plusieurs représentants train-track pour un même automorphisme du groupe libre. On ne sait pas encore si le fait d'être conjuguées à une addition sur un tore est une propriété commune aux laminations attractives engendrées par tous les représentants train-tracks d'un automorphisme. L'intuition dit que les différentes représentations seront des sections d'une même

suspension décrite par un domaine fondamental du tore, et auront donc globalement les mêmes propriétés, mais tout reste à faire pour formaliser cette intuition.

Parallèlement, une direction de recherche consiste à comprendre les propriétés des fractals de Rauzy qui sont invariantes lorsqu'on change de représentant train-track. En particulier, nous avons observé que les fractals de Rauzy pour des automorphismes du groupe libre conjugués semblent avoir des formes semblables (voir Fig. 3.15). Une première étape dans cette direction serait donc d'utiliser les critères topologiques proposés dans la section 3.1 pour comprendre quels sont les invariants topologiques à une conjugaison près.

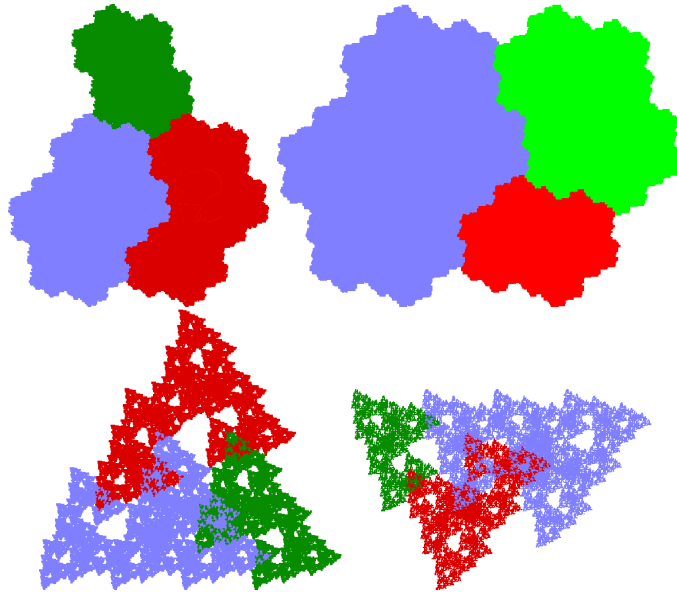


FIG. 3.15 – Fractals de Rauzy pour deux paires  $(\tau_4, \sigma_4)$ , et  $(\tau_2, \sigma_2)$  de substitutions Pisot unitaires irréductibles. Dans chaque paire, les substitutions sont conjuguées par un automorphisme du groupe libre. Ces fractals semblent avoir des propriétés topologiques communes



---

## Chapitre 4

# Appréhender le fonctionnement d'un système dynamique en biologie moléculaire

Dans les deux chapitres précédents, nous avons vu comment on pouvait tirer parti de la connaissance d'une application pour considérer ses propriétés dynamiques et en extraire des informations sur les codages de cette application. Cependant, cela nécessite d'une part de connaître explicitement l'application, d'autre part d'étudier une application dont les régularités sont assez importantes. Dorénavant, nous allons discuter des approches de systèmes dynamiques dans un autre domaine, la biologie moléculaire.

La situation est dans ce cas là vraiment orthogonale à ce que nous avons considéré jusqu'alors : on dispose d'observations sur un système ; on suppose simplement que ce système suit une loi dynamique, au sujet de laquelle on dispose d'informations très partielles. Le but est d'exploiter les observations disponibles sur le système de manière optimale.

Un des points qui frappe dans ce nouveau domaine est son manque de contours ; il s'agit là d'un contraste frappant avec la théorie des systèmes dynamiques symboliques où les enjeux sont bien déterminés. La raison principale de ce flou en biologie des systèmes est la jeunesse de ce domaine : mis à part quelques articles précurseurs et fondateurs des années 1960 (le travail de Jacob et Monod en particulier), l'importance des concepts de dynamique est réellement discutée en biologie moléculaire depuis seulement une quinzaine d'années : l'intérêt des biologistes croît avec l'amélioration des techniques d'observation à grande échelle dans une cellule. Mais finalement, ce domaine n'en est qu'à ses balbutiements. Il est donc caractérisé par une profusion de propositions de méthodes d'analyses de réseaux et de données. A long terme, la sélection entre les différentes approches se fera sans doute par leur



utilisation réelle en biologie.

La jeunesse de ce domaine (et l'esprit du projet Symbiose à l'Irisa) m'a incitée à travailler avant tout sur l'identification de questions qui ont un sens d'un point de vue biologique et peuvent aussi être abordées avec les données mises à disposition. Pour cela, j'ai travaillé en étroite collaboration avec des biologistes moléculaires spécialistes des régulations du métabolisme des acides gras (Agrocampus Rennes), ainsi qu'avec des informaticiens (Michel Le Borgne, Philippe Veber, Carito Guziolowski) et mathématiciens (Ovidiu Radulescu) de l'université.

Du fait de cette approche *du terrain vers la formalisation*, les concepts mathématiques qui vont être introduits dans la suite de ce document sont beaucoup plus simples que dans les chapitres précédents (on considère des déplacements d'états stationnaires pour des modèles différentiels). Par contre, les aspects de résolution informatique, de complexité ont été plus poussés que dans les parties précédentes, puisque les applications envisagées vont jusqu'à l'analyse de données réelles. Ainsi, l'état d'esprit des prochains chapitres est complètement différent des chapitres précédents.

Nous allons d'abord discuter de l'intérêt des systèmes dynamiques en biologie moléculaire, pour l'analyse de séquences ou la modélisation du comportement au sein d'une cellule. Ceci nous amènera à présenter une approche qui consiste encore à discrétiser les trajectoires de systèmes dynamiques, non plus dans l'esprit de la dynamique symbolique (c'est-à-dire partitionner l'espace pour discrétiser les trajectoires), mais plutôt dans un esprit statique : on supposera que dynamiquement, le système étudié est simple (il passe d'un état stationnaire à un autre état stationnaire), et nous nous concentrerons sur la comparaison entre les différents états stationnaires pour identifier des interactions manquant dans le modèle de départ ou orienter de nouvelles expérimentations.

## 4.1 Dynamique en biologie moléculaire

### 4.1.1 Systèmes dynamiques et séquences d'ADN ?

**Dynamique et génomes** Un premier cadre d'application des méthodes de dynamique symbolique en biologie moléculaire pourrait concerner l'étude des séquences d'ADN. En effet, les séquences d'ADN sont des suites de très grande taille sur un alphabet à quatre lettres, et une grande part des recherches à leur sujet se focalise sur la recherche de régularités ou d'irrégularités en leur sein. On peut naturellement voir ces séquences comme des mots infinis, et considérer que ces mots ont été produits par des phénomènes de dynamiques sur les génomes (duplication, insertion, délétion, substitution), ce qui est exploité pour identifier des signatures de génomes [NAT<sup>+</sup>04, OG04, DGV<sup>+</sup>99, JVEGS02].

**Visualiser les régularités** Une méthode pour visualiser des régularités consiste à représenter une séquence par une image de dimension 2, avec des techniques de jeu de chaos. Il s'agit là de représenter une séquence  $u_1 \dots u_n \dots$  par l'ensemble de points  $\{\sum_{i=1}^n \mathbf{l}(u_i)/2^i\}$ , où les vecteurs  $\mathbf{l}(A)$ ,  $\mathbf{l}(C)$ ,  $\mathbf{l}(G)$  et  $\mathbf{l}(T)$  sont quatre points

---

du plan (typiquement,  $(1, 1)$ ,  $(1, -1)$ ,  $(-1, 1)$ ,  $(-1, -1)$  [Jef90]. En considérant de cette manière des génomes d'espèces différentes, on observe à l'œil nu que les représentations obtenues sont bien distinctes, et qu'elles présentent des régularités, qui font penser à des fractals. Lors d'un stage de Master 2, T. Hénin a montré que ces régularités sont expliquées par la présence de mots interdits dans les séquences d'ADN, ces mots interdits étant liés à des faits bien connus en biologie des séquences. Ainsi, la représentation par jeu du chaos est d'autant plus fractale que le langage du génome considéré est proche d'un langage régulier [Hén06].

**Dynamique ou algorithmique ?** Les régularités dans les génomes proviennent ainsi de mécanismes temporels qui ont fait évoluer les génomes jusqu'à leur forme actuelle [DGV<sup>+</sup>99]. Cependant, il faut bien noter que la dynamique se lit non pas sur *un* génome, mais plutôt l'ensemble des génomes des espèces liées par l'évolution. En particulier, l'ordre des bases dans une séquence d'ADN ne semble pas être lui-même le codage de la trajectoire d'une molécule dans le temps. Il faut cependant noter qu'un lien semble exister entre la séquence d'ADN d'une part, la structure et la dynamique de la chromatine d'autre part (par exemple, le placement des histones sur l'ADN n'est pas du tout uniforme), mais on est encore loin de penser que la séquence d'ADN serait un codage de la dynamique associée [NAT<sup>+</sup>04].

Cela ne signifie absolument pas que les séquences d'ADN ne méritent pas une étude attentive. En effet, en tant que produit de l'évolution de génomes, chaque génome contient un grand nombre de régularités et de contraintes cachées qui doivent être explorées via l'étude de son langage. Cependant, cette étude entre dans le domaine de l'algorithmique des séquences et a priori les méthodes de discrétisation de systèmes dynamiques ne semblent pas de grande utilité. Il s'agit là de la spécialité d'une bonne partie du projet Symbiose [Nic08], et j'ai ainsi eu l'occasion de participer à la définition de concepts pour rechercher efficacement des éléments répétés dans les séquences via l'exploitation d'une représentation astucieuse des langages sous la forme d'arbre des préfixes (Repeat, submitted, 2008). Mais on s'éloigne ici du sujet principal de ce document : force est de constater que les séquences d'ADN ne sont pas dynamiques, au sens où elles ne codent pas les trajectoires d'un système sur un espace.

### 4.1.2 Dynamique au niveau cellulaire

**Expression d'une protéine** Les systèmes dynamiques sont cependant loin d'être absents de la biologie moléculaire. Autant les génomes ne codent pas une dynamique, autant tous les mécanismes permettant de passer d'un génome à l'expression explicite d'une protéine sont des phénomènes dynamiques. Rappelons rapidement que les protéines, composées d'une suite d'acides aminés dont la composition est déterminée par le code génétique, assurent l'intégrité et le fonctionnement des cellules. Pour aboutir à la production d'une protéine, une partie d'une séquence d'ADN doit d'abord être transcrite en un ARN, qui lui-même est traduit en protéine (voir Fig. 4.1). La protéine obtenue va ensuite agir à différents endroits de la cellule : certaines protéines jouent le rôle d'enzyme (catalyseur de réactions biochimiques),

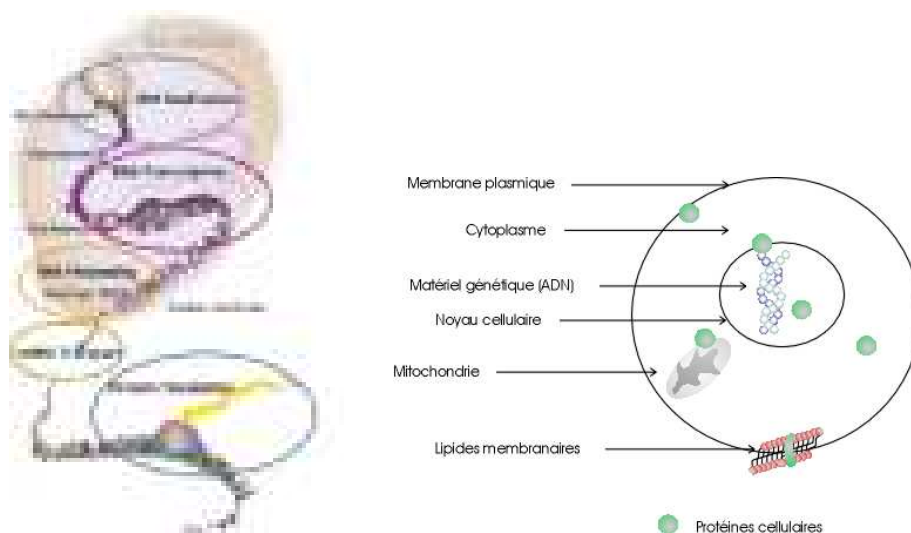


FIG. 4.1 – Gauche : schéma des mécanismes transformant un gène en une protéine. Droite : les différents rôles des protéines (Sources : <http://www.hbcprotocols.com/>, <http://www.hon.ch/Dossier/Ageing/>)

d'autres de composante membranaire, d'autres encore ont un rôle structural extra ou intracellulaire, et les dernières constituent la machinerie indispensable à différents mécanismes tels que la réplication du matériel génétique, la transcription, la traduction des protéines ou la mort cellulaire (sous la forme de protéases ou d'inhibiteurs) (voir Fig. 4.1).

**Régulation et dynamique** Dans ce schéma, le point fondamental qui rend ces phénomènes dynamiques est que la cellule s'est munie de la capacité de *décider* d'activer une transcription, une traduction, ou même la fonction d'une protéine. Et cette décision est principalement due à la présence ou l'absence de différentes autres protéines. Ainsi, la transcription est activée si différents *facteurs de transcription* (qui sont des protéines), sont présents pour se fixer à l'avant de la partie du génome qui est traduite. De même, une protéine va jouer son rôle dans la cellule à partir du moment où elle est *activée*, souvent par un phénomène physique tel que l'adjonction d'un groupement phosphate (phosphorylation), un clivage, ou la formation d'un complexe avec une autre protéine de la cellule.

Ce schéma est bien entendu très incomplet, de nombreux mécanismes de régulations existent et méritent chacun d'être décrits, étudiés et approfondis. Ce qu'il faut retenir ici, c'est que la cellule décide de produire ou non une molécule en fonction de la quantité des autres molécules dans la cellule. Et en ce sens, la concentration d'une molécule est fonction des concentrations de l'ensemble des molécules de la cellule, et on peut considérer que l'ensemble de ces concentrations forment l'état

d'un système dynamique.

**Spécificité des systèmes** Comparés aux systèmes dynamiques expérimentaux déjà connus et étudiés, en particulier en physique et en chimie, voire en écologie, les spécificités de la biologie moléculaire sont :

- la difficulté pour observer des molécules au sein de la cellule ;
- la difficulté pour perturber le système sans le détruire ;
- la difficulté pour avoir beaucoup de réplicats : les expériences sont qualitativement reproductibles. D'un point de vue quantitatif, cependant, la variabilité est telle qu'il faudrait quelques dizaines de réplicats pour obtenir des informations fiables ; en pratique, les coûts financiers et les difficultés expérimentales font que le nombre de réplicats est de 3 à 5.

Des techniques d'observations haut-débit se sont cependant développées ces 15 dernières années, faisant entrer la biologie dans l'ère *omique* : elles permettent d'observer plus de variables, mais avec une précision moindre. De plus, le coût limite le nombre de réplicats. Autrement dit, nous sommes en présence de systèmes pour lesquels :

- On dispose de lois assez générales pour décrire les phénomènes principaux de régulation.
- On dispose d'observations qualitatives sur les concentrations des molécules pendant des expérimentations, qui peuvent être des déplacements d'états stationnaires en réponse à un stress (privation de nourriture, privation d'oxygène, introduction dans un milieu azoté...) ou des réponses à une perturbation génétique (délétion ou activation d'un gène). Lorsqu'elles sont dites temporelles, ces observations consistent au maximum en une vingtaine de points temporels, répartis souvent sur plusieurs heures voire plusieurs jours.
- On dispose de connaissances sur les phénomènes de régulation d'un système, dans l'immense production bibliographique publiée dans le domaine de la biologie. Cependant, ces informations sont actuellement difficilement exploitables, parce que insuffisamment structurées sous la forme d'interactions.

### 4.1.3 Plusieurs stratégies pour modéliser la dynamique d'un réseau à un niveau qualitatif

Il devient donc légitime de chercher à prédire le comportement des molécules à l'intérieur des cellules, mais les problématiques posées en biologie moléculaire ne peuvent pas importer les méthodes développées en physique ou en biochimie, à cause de la difficulté pour construire des modèles quantitatifs et fiables à assez grande échelle. Il faut noter quelques exceptions obtenues grâce à des travaux expérimentaux considérables associés à de gros efforts de modélisation : on dispose maintenant de modèles différentiels réalistes pour la régulation du cycle cellulaire [CCCN<sup>+</sup>04] ou les voies de transduction de signal associées à EGF [SEJGM02] et NF $\kappa$ B [NIE<sup>+</sup>06].

Différentes approches ont été explorées pour simuler et modéliser la dynamique des réseaux de régulation génétique. Certaines restent différentielles (en se focalisant

sur des réseaux de petite taille), d'autre discrètes, d'autres encore adoptent un point de vue probabiliste, voire stochastique [DJ02]. Les questions posées sont variées : citons entre autres l'atteignabilité, l'existence d'états stationnaires ou d'oscillations, l'enchaînement de réactions pour produire une molécule [MRM<sup>+</sup>08].

**Approches discrètes** Les approches discrètes en particulier, dont nous avons discuté dans la section 1.4.2, sont extrêmement puissantes pour la simulation et l'étude de la dynamique d'un réseau ; elles ont montré des résultats très encourageants ces dix dernières années [BRdJ<sup>+</sup>05, GCT08, GBC<sup>+</sup>04]. Elles sont cependant limitées par la taille des réseaux qui peuvent être étudiés : du fait de l'explosion combinatoire des trajectoires considérées et de la quantité d'information biologique (même qualitative) nécessaire pour effectivement spécifier le modèle, les études sont limitées à deux ou trois dizaines de sommets.

Pour élargir la classe de modèles étudiables, la communauté des informaticiens s'est attaché à formuler des requêtes liées à la présence de molécules aux différentes étapes des transformations. Les modèles sont alors décrits par des réactions élémentaires et utilisent des lois d'action de masse ; la valeur des paramètres n'y est souvent pas spécifiée. Ces approches proposent de modéliser un réseau dans un langage ad-hoc, puis d'utiliser une sémantique dynamique pour les éléments du langage (booléen, différentielle ou autre). Notons bien que la précision de la sémantique a des conséquences sur la précision des informations demandées à l'utilisateur. Ceci permet d'interroger le système en fonction de cette sémantique, généralement avec des algèbres de processus, des requêtes en logique temporelle ou des analyses de diagrammes de décision [DL04, RPS<sup>+</sup>04, CFS06].

**Prédiction à partir d'invariants robustes** Adoptant un autre point de vue, une communauté de mathématiciens s'est attaché à définir des méthodes de prédictions *robustes* à partir d'invariants globaux, partiellement indépendantes de la valeur exacte des paramètres du système. Il s'agit de proposer des prédictions à partir d'un modèle présenté sous la forme d'un graphe d'interaction ou d'un modèle de réactions élémentaires. La principale illustration de cette approche est la *règle de Thomas*, qui dit que si un système admet plusieurs états stationnaires non dégénérés, alors son graphe d'interaction contient au moins un cycle positif [Tho81]. On a ici une information globale sur le système, ayant un sens biologique important (la bistabilité du système, induisant des phénomènes d'hystéresis), sans une étude détaillée de la dynamique. Cette règle, proposée dans les années 1980, a été prouvée pour les différents formalismes existants (logique, multivalué, différentiel...) dans les années 2000 [Sou06]. Elle est fréquemment exploitée dans les études de réseaux déjà mentionnées : la première étape de l'étude d'un réseau consiste à identifier les cycles positifs, qui peuvent induire une multistationnarité, et ensuite à proposer des analyses (simulation, prédiction, abduction) pour identifier les cycles qui sont effectifs, c'est-à-dire qui provoquent effectivement une multistationnarité. Cependant, pour les réseaux qui intègrent des aspects métaboliques, la règle de Thomas s'applique mal : les réactions métaboliques réversibles introduisent des cycles positifs dans les graphes d'interaction, qui souvent ne sont pas effectifs mais empêchent une analyse

simple de la stationnarité.

**Signes de déterminants et univalence** Plus globalement, les travaux existant sur les prédictions de systèmes à partir d'informations globales se concentrent pour la plupart sur l'étude de signes de déterminants, ce qui est naturel puisque les questions posées concernant l'existence ou l'unicité d'attracteurs, ainsi que leur stabilité. Or, ces informations sont principalement contenues dans la matrice jacobienne de la dynamique (en particulier ses sous-déterminants), qui est précisément donnée par le graphe d'interaction du réseaux [HKG]. Ce domaine est en plein essor actuellement et évolue encore rapidement. Une méthode usuelle de réduction utilise les échelles de temps ; elle oblige cependant à privilégier les variables lentes par rapport aux variables rapides. On doit aussi distinguer une théorie plus mature que les autres, inspirée de la théorie du contrôle, qui obtient pour des systèmes monotones (pour lesquels une condition nécessaire est la présence de cycles positifs) des résultats importants au sujet de l'existence d'états stationnaires et de leur stabilité. Ceci offre un critère pour décomposer le système en modules qui présente un comportement entrée-sortie : chaque module pris isolément converge vers un état stationnaire pour toute condition initiale [Son07].

Lorsque seuls des cycles négatifs sont présents, nous verrons comment le théorème de Gale-Nikaido [Par83] permet de montrer que le système possède un unique état stationnaire.

**Contributions** J'ai axé mes recherches sur des méthodes permettant des prédictions robustes pour un réseau, dans des modèles assez peu renseignés et de grande taille. Du fait de l'absence de connaissances sur la valeur numérique des paramètres, les méthodes de prédiction vont s'appuyer sur la comparaison de différents états stationnaires. Nous allons cependant voir que cette restriction permet quand même de mieux comprendre le fonctionnement d'un système, en particulier de détecter les zones qui méritent une modélisation fine.

- Les premiers résultats portent sur des réseaux formalisés sous la forme de modèles de réactions élémentaires. Nous verrons que l'étude des sous-déterminants de la jacobienne du système issu du modèle de réactions permet de caractériser les effets de certaines régulations.
- La seconde classe de résultats étudie la réponse d'un système à des paramètres extérieurs. Les systèmes sont alors décrits seulement par leur graphe d'interaction, aussi appelé graphe d'influence. Lorsqu'ils sont taille moyenne, on arrive à comprendre le rôle de boucles positives dans le comportement du système.
- La troisième classe de système comprend des réseaux de grande échelle couplés avec des données de puces à ADN. On entre alors dans une problématique de correction de modèles avec des méthodes de diagnostic. Cette troisième classe de modèle sera considérée dans le chapitre suivant.

Pour toutes ces questions, une question biologique reviendra régulièrement : il s'agit de la modélisation des régulations du métabolisme des acides gras. Nous allons en particulier présenter différents modèles pour ce système, avec un niveau d'abstraction plus ou moins important. Avec ce fil conducteur, nous allons illus-

trer pourquoi la biologie systémique doit adopter une démarche progressive pour comprendre un système dans ses différentes composantes : considérer d'abord des données qualitatives avec un graphe d'influence, puis un modèle défini par des réactions mais sans spécifier ses coefficients, pour enfin se concentrer sur l'analyse de données numériques sur une toute petite partie du modèle. Chaque type d'abstraction correspond à une classe de questions et engendre des méthodes spécifiques de résolution. Pour revenir à ce qui a été écrit au début de ce chapitre, on pourra ainsi illustrer l'une des principales causes de la difficulté à poser des contours pour la biologie systémique, à savoir, la complémentarité entre les différents degrés d'abstraction pour l'étude d'un même système et l'impossibilité d'en préférer une par rapport aux autres.

## 4.2 Prédications pour des réseaux décrits par des réactions

### 4.2.1 Existence des états stationnaires dans un modèle dynamique

**Modèle de réactions** Nous nous plaçons maintenant dans un contexte général de dynamique différentielle : on suppose que le flux d'un produit, de concentration  $X_i$ , est le résultat de différentes réactions  $R_j$  pour lesquels  $X_i$  est soit consommé soit produit. Si  $X_i$  est consommé par la réaction  $R_j$ , cette réaction intervient avec un coefficient négatif dans le flux de  $X_i$ . Si  $X_i$  est un produit de  $R_j$ , cette réaction apparaît avec un coefficient positif.

La *dynamique d'un système* est donc décrite par le système suivant :

$$\frac{d\mathbf{X}}{dt} = \mathbf{F}(\mathbf{X}, \mathbf{p}) \quad \mathbf{F}(\mathbf{X}, \mathbf{p}) = \sum_i^r \nu_i \mathbf{R}_i(\mathbf{X}, \mathbf{p}), \quad (4.1)$$

où  $\mathbf{X}$  désigne la concentration des produits  $X_i$ ;  $\mathbf{p} \in \mathbb{R}^q$  désigne les paramètres du système,  $\mathbf{F} : \mathbb{R}^n \times \Delta \rightarrow \mathbb{R}^n$  est un champ de vecteur différentiel avec  $\Delta$  est un sous-ensemble compact de  $\mathbb{R}^q$ ;  $\nu_i$  est le vecteur stoechiométrique de la réaction élémentaire  $i$  et  $R_i(\mathbf{X}, \mathbf{p})$  son taux.

Un état stationnaire est alors défini par  $\mathbf{F}(\mathbf{X}, \mathbf{p}) = \mathbf{0}$ .

**Une condition suffisante d'existence d'un état stationnaire** Comme mentionné dans la section précédente, la recherche des états stationnaires pour les systèmes biologiques a fait l'objet d'une littérature importante, en particulier leur unicité et leur stabilité. Les résultats d'existence sont par contre plus épars.

Avec Ovidiu Radulescu et Elisabeth Pécou, nous avons utilisé la formule de Poincaré-Hopf pour montrer le résultat suivant sur des champs de vecteurs.

**Proposition 4.2.1** ((Roy. Soc. Inter, 2006) et (R., S., P., C.& L. en cours))

Soit  $\Phi(\mathbf{X}) = \mathbf{G}(\mathbf{X}) - \Lambda(\mathbf{X})$  un champ de vecteurs différentiable sur  $\mathbb{R}_+^n$  tel que

-  $\mathbf{G}$  est borné,

- Pour tout  $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_n)$  tel que  $X_i = 0$  et  $X_j \neq 0$  pour  $j \neq i$ ,  $\mathbf{G}$  vérifie  $G_i(\mathbf{X}) > 0$ ,
- $\mathbf{\Lambda} = (\mathbf{\Lambda}_1(\mathbf{X}_1), \dots, \mathbf{\Lambda}_n(\mathbf{X}_n)) : \mathbb{R}_+^n \rightarrow \mathbb{R}_+^n$ , et  $\Lambda_i$  sont différentiables  $\Lambda_i(0) = 0$  et vérifient  $\lim_{\|\mathbf{X}\| \rightarrow +\infty} \Lambda_i(\mathbf{X}) = +\infty$ , pour tout  $1 \leq i \leq n$ .

Alors l'équation  $\Phi(\mathbf{X}) = \mathbf{0}$  admet au moins une solution dans  $\mathbb{R}_+^n$ .

Même si ce résultat est peu avenant, il est en fait assez naturel d'un point de vue biologique :  $\mathbf{G}(\mathbf{X})$  représente alors les régulations entre les différents composants  $\mathbf{X}$  d'un modèle et  $\mathbf{\Lambda}(\mathbf{X})$  représente la dégradation naturelle de ces composants. La proposition dit que si les réactions élémentaires dans le modèle vérifient les hypothèses (raisonnables biologiquement) suivantes, alors la production et la dégradation des composants vont s'équilibrer et exhiber un état stationnaire :

1. les dégradations doivent être des fonctions strictement croissantes de la concentration des produits ;
2. l'absence d'un substrat pour un flux le bloque ;
3. tous les flux qui ne sont pas des dégradations saturent à haute concentration ;
4. le système active la production d'un produit dès que sa concentration devient nulle.

On obtient ainsi une condition générale pour l'existence d'un état d'équilibre. L'intérêt de ce résultat est qu'il est totalement indépendant des paramètres des fonctions  $R_i$  ; on peut donc le considérer comme une propriété robuste du système.

**Exemple : métabolisme des acides gras et ses régulations** Dès le début de mes travaux en bioinformatique, j'ai travaillé avec S. Lagarrigue, du laboratoire de génétique animale de l'Agrocampus Rennes. L'ambition de ce laboratoire est de trouver des régulateurs génétiques du caractère d'engraissement chez le poulet. La question de la régulation des acides gras reviendra donc régulièrement tout au long de ce document.

Pour commencer à étudier ce problème, nous avons proposé un modèle abstrait du métabolisme des acides gras. Pour cela, nous avons volontairement abstrait chaque voie du métabolisme (glycolyse, synthèse, dégradation des acides gras) en une seule réaction. Nous avons aussi abstrait les phénomènes de régulation en deux voies de régulation : les enzymes de la synthèse des acides gras sont globalement connues pour être régulées par les facteurs de transcription LXR $\alpha$ -SREBP1, tandis que l'oxydation des acides gras est régulée par le facteur de transcription PPAR. De plus, une classe d'acides gras, dits *polyinsaturés* ou PUFA, agissent comme des ligands pour l'activation du facteur de transcription connu pour réguler l'oxydation des acides gras (PPAR) et l'inactivation du facteur de transcription qui régule la synthèse des acides gras (LXR $\alpha$ ) [Jum04].

A partir de ces connaissances, nous avons proposé le modèle décrit dans la Figure 4.1. Ses spécificités sont les suivantes :

- Toutes les voies sont abstraites.
- Les métabolites agissent comme des régulateurs des voies de régulation transcriptionnelle, induisant une boucle de régulation transcription/métabolisme/transcription peu étudiée du point de vue de la modélisation.



- Les transformations de l'ATP sont prises en compte dans le modèle, via la variable  $T$ .

Les variables de ce modèle sont données dans la table suivante. Pour faciliter la comparaison avec les modèles de niveau d'abstraction inférieur que nous présenterons dans la suite du document, nous mettons aussi les noms des enzymes correspondantes.

Type	Name	Concentration symbol		$\frac{d \text{ product}}{dt}$
		abstract node	corresponding enzyme	
Metabolic parameter	Glucose	$G$		
Metabolic variable	Acetyl Co-A	$\mathcal{A}$		$\Phi_{\mathcal{A}}$
	Saturated and monounsaturated fatty acids	$F_1$	S/MU-FA	$\Phi_{F_1}$
	Poly-unsaturated fatty acids	$F_2$	PUFA	$\Phi_{F_2}$
Energetic variable	Energy ATP	$T$		$\Phi_T$
Genetic variable	Active form of PPAR	PP	PPAR-a	$\Psi_1$
	Active form of the regulation path LXR-SREBP	L	LXR-a, SREBP-a, SCAP	$\Psi_2$
	Enzymes of S/MU-FA synthesis	$E_1$	ACL, ACC, FAS	$\Psi_3$
	Enzymes of S/MU-FA oxidation	$E_2$	SCD1	$\Psi_4$
	Enzymes of PUFA oxidation	$E_3$	FADS1, FADS2	$\Psi_5$
	Enzymes of Ketone body exit	$E_4$		$\Psi_6$

**Modèle abstrait de réaction du métabolisme des lipides** Un modèle différentiel pour formaliser la représentation du métabolisme donnée Fig. 4.2 est donné dans la Table 4.1. Volontairement, les fonctions de flux et de régulations Gly, Oxi1, Oxi2, Krebs, ... y sont laissées à l'état symbolique. Ce modèle prend simplement en compte la stoechiométrie des transformations en introduisant des paramètres  $n_1$ ,  $n_2$ ,  $\alpha_G$ , ...

$$\left\{ \begin{array}{l}
 \frac{dA}{dt} = \text{Gly}(G, T) + n_1 \text{Oxi1}(F_1, T, E_2) + n_2 \text{Oxi2}(F_2, T, E_3) \\
 \quad - \text{Krebs}(A, T) - \text{Kout}(A, E_4) - n_1 \text{Syn}(A, T, E_1) - \delta_A A \\
 \frac{dF_1}{dt} = \text{Syn}(A, T, E_1) - \text{Oxi1}(F_1, T, E_2) + \text{Fin1}(F_1, T) - \delta_{F_1} F_1 \\
 \frac{dF_2}{dt} = -\text{Oxi2}(F_2, T, E_3) + \text{Fin2}(F_2, T) - \delta_{F_2} F_2 \\
 \frac{dT}{dt} = \alpha_G \text{Gly}(G, T) + \alpha_K \text{Krebs}(A, T) + \alpha_{O1} \text{Oxi1}(F_1, T, E_2) \\
 \quad + \alpha_{O2} \text{Oxi2}(F_2, T, E_3) - \alpha_S \text{Syn}(A, T, E_1) - \text{DegT}(T) \\
 \frac{dPP}{dt} = \tilde{\Psi}_1(F_2) - \delta_{PP} PP \\
 \frac{dL}{dt} = \tilde{\Psi}_2(F_2) - \delta_L L \\
 \frac{dE_1}{dt} = \tilde{\Psi}_3(L) - \delta_{E_1} E_1 \\
 \frac{dE_2}{dt} = \tilde{\Psi}_4(PP) - \delta_{E_2} E_2 \\
 \frac{dE_3}{dt} = \tilde{\Psi}_5(PP) - \delta_{E_3} E_3 \\
 \frac{dE_4}{dt} = \tilde{\Psi}_6(PP) - \delta_{E_4} E_4
 \end{array} \right.$$

TAB. 4.1 – Un modèle différentiel abstrait pour la synthèse des acides gras et ses régulations.

On obtient un résultat d'existence d'un état stationnaire pour ce modèle dès que les hypothèses biologiques associées à la Proposition 4.2.1 sont vérifiées.

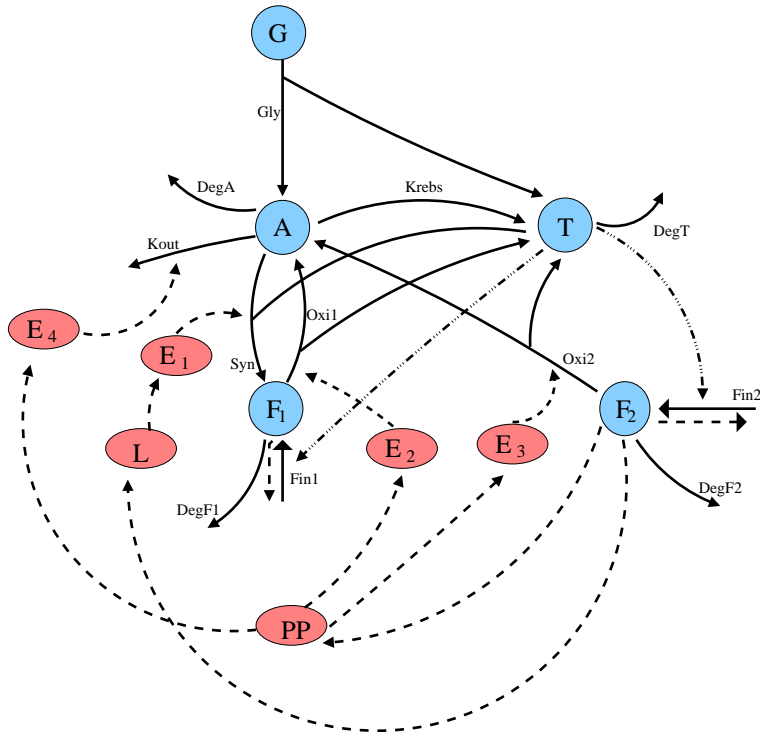


FIG. 4.2 – Un modèle du métabolisme des acides gras et ses régulations, sous la forme d'un graphe qui schématise un modèle différentiel défini par des réactions élémentaires (voir 4.1). Les flèches en pointillés désignent des régulations génétiques. Les flèches pleines désignent des flux métaboliques. Les flèches pointillées/traités désignent des régulations énergétiques impliquant l'ATP ( $T$ ).

**Corollaire 4.2.2** *Le modèle de la régulation du métabolisme des acides gras admet au moins un état stationnaire et un état quasi-stationnaire (où les concentrations des enzymes sont considérées comme constantes) dès que tous les flux sont bornés, qu'ils sont nuls lorsque la concentration d'un de leurs substrats est égale à zéro et qu'ils sont strictement positifs lorsque la concentration d'un de leurs produits est nulle.*

**Deux états stationnaires mis en évidence expérimentalement** Les observations sur le métabolisme font apparaître deux états stationnaires distincts pendant les protocoles de mise à jeûn [LCL<sup>+</sup>04] : lors des phases de nutrition (où on considère que les apports de glucose sont importants), la quantité d'acide gras dans le foie est petite ; lors des phases de jeûne, la quantité de glucose diminue et la quantité d'acide gras augmente. Savoir si le passage d'un état à l'autre est continu ou si les deux états peuvent coexister est alors une question naturelle du point de

vue de la modélisation. Pour cela, on doit se poser la question de l'unicité de l'état stationnaire.

## 4.2.2 Condition d'unicité

**Table de signes des matrices de différentiation** Dans ce modèle, nous avons donc volontairement choisi de garder des informations abstraites sur les différents flux. Par contre, nous avons constaté que les signes de matrice de dérivation  $\frac{\partial \mathbf{R}_i}{\partial \mathbf{X}_j}$  sont spécifiés à partir des connaissances sur le système. Cette table de signes est présentée dans la Table 4.2.

$\frac{\partial \text{flux}}{\partial \text{var}}$	Gly	Krebs	Kout	Syn	Oxi1	Oxi2	Fin1	Fin2	DegT	$\Psi_1$	$\Psi_2$	$\Psi_3$	$\Psi_4$	$\Psi_5$	$\Psi_6$
A	0	+	+	+	0	0	0	0	0	0	0	0	0	0	0
F <sub>1</sub>	0	0	0	0	+	0	-	0	0	0	0	0	0	0	0
F <sub>2</sub>	0	0	0	0	0	+	0	-	0	+	-	0	0	0	0
T	-	-	0	+	-	-	-	+	0	0	0	0	0	0	0
PP	0	0	0	0	0	0	0	0	0	0	0	0	+	+	+
L	0	0	0	0	0	0	0	0	0	0	0	+	0	0	0
E <sub>1</sub>	0	0	0	+	0	0	0	0	0	0	0	0	0	0	0
E <sub>2</sub>	0	0	0	0	+	0	0	0	0	0	0	0	0	0	0
E <sub>3</sub>	0	0	0	0	0	+	0	0	0	0	0	0	0	0	0
E <sub>4</sub>	0	0	+	0	0	0	0	0	0	0	0	0	0	0	0
G	+	0	0	0	0	0	0	0	0	0	0	0	0	0	0

TAB. 4.2 – Table des signes des dérivées  $\frac{\partial \mathbf{R}_i}{\partial \mathbf{X}_j}$  pour le modèle de la régulation du métabolisme des acides gras.

Ceci est en fait un résultat assez général : dans le cas des réseaux transcriptionnels, on est capable de déterminer les signes de la matrice jacobienne de  $\Phi$ , et on peut s'appuyer sur cette connaissance pour appliquer la règle de Thomas et en déduire des conditions d'unicité d'états d'équilibre. Pour les réseaux qui incluent des flux métaboliques, ce processus se passe généralement mal : les signes de la matrice jacobienne ne sont plus constants, ou, lorsqu'ils le sont, ils introduisent des cycles positifs superflus liés à la présence de réactions réversibles. Par contre, les signes des dérivées des réactions élémentaires  $\frac{\partial \mathbf{R}_i}{\partial \mathbf{X}_j}$  peuvent être identifiés à partir de la littérature. La question que nous allons aborder maintenant est de savoir comment on peut s'appuyer sur cette connaissance pour proposer des études d'unicité d'états stationnaires et de prédictions sur les modèles.

**Extension de la règle de Thomas ?** Pour tester l'unicité des états stationnaires d'un système, nous proposons dans (Roy. Soc. Inter, 2006) et (R., S., P., C.& L. en cours) de nous appuyer sur le théorème de Gale-Nikaido-Inada [Par83], qui généralise le fait qu'une fonction décroissante sur  $\mathbb{R}$  s'annule au plus une fois.

**Théorème 4.2.3 (Gale-Nikaido-Inada)** *Soit  $\mathbf{X} \rightarrow \Phi(\mathbf{X})$  une application différentiable de  $\mathbb{R}_+^n$  dans  $\mathbb{R}^n$  et de jacobien  $J$ . Si tous les mineurs principaux de  $-J$  sont positifs, alors le système  $\Phi = \mathbf{0}$  a au plus une solution.*

Comme l'a fait remarqué C. Soulé, ce théorème est en fait une généralisation de la règle de Thomas qui dit qu'une condition suffisante pour qu'un système ait un unique état stationnaire est que son graphe d'interaction ne contienne pas de

boucle positive [Tho81, Sno98, Gou98, Sou03]. En effet, le graphe d'interaction d'un système biologique est intimement lié à la matrice jacobienne du champ de vecteur associé : ses arêtes correspondent aux coefficients non nuls de la matrice jacobienne ; les signes du graphe sont donnés par les signes des coefficients de la matrice. Or, les mineurs principaux de la matrice peuvent être exprimés comme sommes d'éléments dont les signes sont déterminés par les cycles dans le graphe d'interaction. Si ces cycles sont tous négatifs, on peut en déduire que tous les mineurs sont positifs et appliquer le théorème de Gale-Nikaido (voir (R., S., P., C.& L. en cours)).

Cependant, ce résultat est trop restrictif pour la majeure partie des applications, en particulier dans le cas des réseaux métaboliques, du fait de la présence de réactions réversibles, qui induisent des cycles positifs inactifs dans les réseaux.

Dans (R., S., P., C.& L. en cours), nous montrons comment il est possible d'appliquer itérativement le théorème de Gale Nikaido pour obtenir une condition pour avoir un unique état stationnaire pour le système du métabolisme des lipides. Nous allons détailler maintenant cet aspect.

**Coefficients de contrôle** Puisque les réactions dans le modèle sont des éléments symboliques, au sujet desquels on connaît seulement les signes des variations par rapport aux variables du modèle, on introduit des *coefficients de contrôle* pour interpréter le rôle des réactions dans la dynamique du système.

Usuellement, les coefficients de contrôle et d'élasticité comparent l'importance des variations de flux. Nous appelons ainsi *coefficients de contrôle* les dérivées  $|\frac{\partial R_i}{\partial X_j}|$  qui sont non nulles dans la Table 4.2. Pour les réseaux métaboliques, ces coefficients sont utilisés pour quantifier et étudier le rôle des enzymes régulatrices des flux [CB04]. Ici, vu le rôle particulier joué par les métabolites que sont les PUFA et l'ATP, on considère des coefficients de contrôles pour les variables  $F_2$  et  $T$ .

### Réductions successives pour une unicité d'états stationnaires

- Le théorème de Gale-Nikaido s'applique d'abord au sous-groupe de variables génétiques. Ceci signifie que la valeur des concentrations de  $E_1, E_2, E_3, E_4, L$  et  $PP$  à un état stationnaire est une fonction de la valeur des concentrations des autres variables en cet état stationnaire. On peut donc construire un nouveau système formel en réduisant les variables génétiques, que ne fait plus intervenir que les variables métaboliques, et qui admet les mêmes états stationnaires que le premier modèle.
- Dans ce nouveau modèle de réaction, les signes des coefficients de contrôle sont encore connus : ils sont déterminés en utilisant le théorème des fonctions implicites. Pour ce nouveau modèle, on montre que les variables  $F_1, F_2, A$  vérifient le théorème de Gale-Nikaido et peuvent donc être réduites en un nouveau modèle qui admet  $T$  pour seule variable et  $G$  pour seul paramètre. Là encore, par construction, ce modèle admet les mêmes états stationnaires que les précédents.
- Pour ce dernier modèle, une condition naturelle pour avoir un unique état stationnaire est que la dérivée du flux de  $T$  ait un signe constant.

L'intérêt de cette approche est qu'elle est explicite. Finalement, on obtient une condition d'unicité d'état stationnaire qui est donnée par une fonction polynomiale des coefficients de contrôle.

**Proposition 4.2.4 (R., S., P., C.& L. en cours)** *Il existe quatre fonctions  $f, g, h, l$ , dont les variables sont les coefficients de contrôle  $|\frac{\partial \mathbf{R}_i}{\partial \mathbf{X}_j}|$ , et qui sont toutes positives telles que le modèle abstrait du métabolisme des lipides admet un unique état stationnaire dès que les conditions suivantes sont vérifiées pour tout  $\mathbf{X}$  :*

$$f(\mathbf{X})\left(\frac{\partial \text{Fin1}}{\partial \mathbf{T}}(\mathbf{X}) - \frac{\partial \text{Oxi1}}{\partial \mathbf{T}(\mathbf{X})}\right) + h(\mathbf{X}) > l(\mathbf{X}), \quad (4.2)$$

$$|g(\mathbf{X})\left(\frac{\partial \text{Fin2}}{\partial \mathbf{T}}(\mathbf{X}) - \frac{\partial \text{Oxi2}}{\partial \mathbf{T}}(\mathbf{X})\right)| \ll \ll f(\mathbf{X})\left|\frac{\partial \text{Fin1}}{\partial \mathbf{T}}(\mathbf{X}) - \frac{\partial \text{Oxi1}}{\partial \mathbf{T}}(\mathbf{X})\right| \quad (4.3)$$

$$\alpha_S < \alpha_{O1} < n_1 \alpha_G, \quad n_2 \alpha_{O1} < n_1 \alpha_{O2}. \quad (4.4)$$

**Interprétation biologique** L'intérêt de ce résultat s'évalue par l'interprétation biologique des conditions algébriques qui sont introduites.

- On s'assure que la condition stoechiométrique (4.4) est vérifiée en considérant un bilan des molécules d'Acétyl-coA et d'ATP produites et consommées pendant la dégradation et la synthèse des différentes classes d'acides gras.
- La condition sur la lipolyse (4.2) signifie que les variations d'énergie ont un effet suffisamment fort sur l'entrée des acides gras dans la cellule. Ceci est en accord avec des observations dans [LCL<sup>+</sup>04] pendant des protocoles de mise à jeûn chez des souris.
- la condition sur les proportions d'acides gras (4.3) signifie que les quantités d'acides gras polyinsaturés sont minoritaires dans les quantités d'acides gras totaux, ce qui est confirmé par des observations de [LCL<sup>+</sup>04].

Ainsi, les conditions de la proposition 4.2.4 sont raisonnables d'un point de vue biologique, et on peut conclure que cette analyse suggère que le métabolisme des lipides incluant ses régulations admet un unique état stationnaire.

**Perspective : effectivité de la méthode** Dans (R., S., P., C.& L. en cours), les calculs pour obtenir la condition d'équilibre sont fait manuellement. Une question naturelle est d'identifier automatiquement les boîtes sur lesquelles des réductions sont possibles, puis d'opérer ces réductions avec des méthodes de calcul algébrique. Nos premiers essais ont abouti à une explosion combinatoire lors du calcul des signes de mineurs, même en utilisant des propriétés de stabilité de signes.

Dans une autre direction, Ovidiu Radulescu a proposé avec Vincent Noël de faire appel à la notion de réaction pendante [GR07] pour classer les variables en fonction de leurs effets, ce qui permet de vérifier la condition de Gale-Nikaido sur un plus petit sous-ensemble de déterminant. L'applicabilité de cette approche est en cours d'étude actuellement.

### 4.2.3 Prédications et comparaison de modèles

**Faire de la dynamique en comparant des états stationnaires ?** Comme nous l'avons expliqué, le théorème de Gale-Nikaido peut-être utilisé pour réduire

astucieusement un système de réactions et en déduire des conditions générales pour avoir un unique état stationnaire. La question qui se pose naturellement alors est d'utiliser ces méthodes pour comprendre le fonctionnement du système. Bien entendu, comme les réactions ne sont pas paramétrées, il est impossible de procéder à des simulations de la dynamique.

Pour obtenir des prédictions dynamiques, nous avons proposé de tirer parti du fait que l'ordre de réduction des variables n'est absolument pas contraint par la biologie. En particulier, comme détaillé ci-dessus, la première étape de réduction consiste à réduire les variables génétiques, c'est-à-dire les variables lentes. Ceci est complètement contraire aux méthodes de réduction par échelles de temps qui réduisent les variables rapides pour obtenir une bonne approximation d'une dynamique sur sa variété lente. Ici, la stratégie est inverse : on a le droit de réduire les variables lentes et obtenir une nouvelle dynamique. Cette nouvelle dynamique n'a aucun sens physique, au sens où il ne s'agit pas d'une approximation du comportement réel du système). Par contre, la nouvelle dynamique partage avec le premier modèle les mêmes états stationnaires.

L'intérêt de cet approche est de pouvoir mieux comprendre le comportement des métabolites sous l'effet de différentes conditions. Nous avons illustré cela sur deux types de prédictions.

**Rôle des régulations génétiques** Pour appréhender le rôle des régulations génétiques dans un modèle, on compare la valeur de son état stationnaire à celle de l'état stationnaire du même modèle dans lequel les régulations génétiques sont supposées constantes. L'état stationnaire de ce deuxième modèle est en fait l'état quasi-stationnaire du modèle général. La réduction des variables génétiques permet d'intégrer au niveau métabolique l'effet des régulations génétiques. Les deux modèles réduits sont montrés dans la figure 4.3.

On peut aller plus loin que cette étude graphique. En considérant les expressions des flux après réduction, on montre dans (R., S., P., C.& L. en cours) que la valeur à l'état quasi-stationnaire de  $(\frac{dT}{dG})$  est strictement supérieure à sa valeur à l'état stationnaire.

Biologiquement, on interprète cette relation algébrique comme suit : un des rôles des régulations génétiques est de renforcer l'effet tampon sur la quantité d'ATP : les variations d'ATP pour une variation fixée de nutriment sont moins importantes quand les régulations génétiques sont présentes.

Avec des comparaisons similaires entre les états quasi-stationnaires et stationnaires, on prédit aussi que les courbes qui représentent la concentration des PUFA pendant le jeûne devraient montrer un pic : l'augmentation de la concentration des PUFA est plus importante à court terme qu'à long terme.

Ainsi, avec un modèle extrêmement abstrait et non paramétré, une étude fine des signes de la matrice de dérivation permet d'obtenir des informations non triviales sur le comportement dynamique du modèle, en particulier de comparer les états quasi-stationnaires et stationnaires.

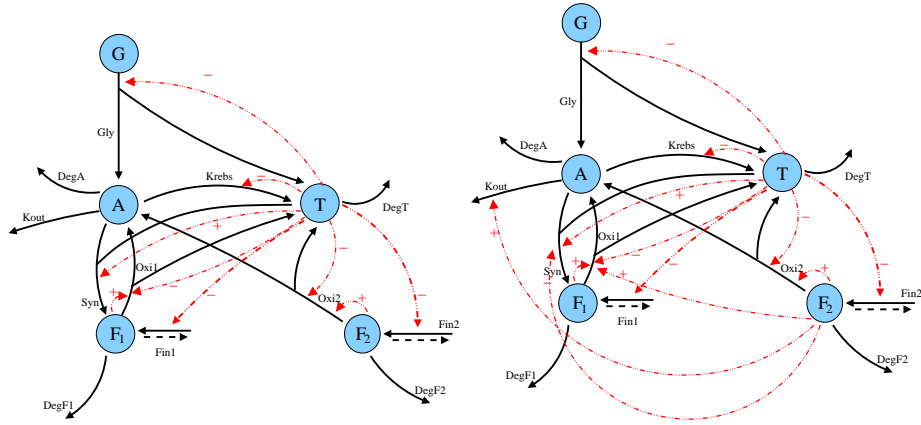


FIG. 4.3 – Gauche : régulations entre métabolites pendant la partie de la dynamique où les régulations génétiques ne sont pas mises en place (jusqu'à l'état quasi-stationnaire). Droite : régulations implicites entre les métabolites lorsque les régulations génétiques sont mises en place. On constate que les régulations génétiques apparaissent implicitement sous la forme de régulations additionnelles venant des PUFA.

**Prédictions expérimentales** Des méthodes semblables permettent de prédire l'effet de mutations sur des sujets. En effet, muter le gène PPAR chez un animal s'interprète dans le modèle par l'annulation de certaines régulations. On peut à nouveau comparer la valeur des états stationnaires et quasi-stationnaires entre les animaux mutés et les sauvages, ce qui donne la relation suivante :

$$\left(\frac{dT}{dG}\right)_{eq, PPAR-/-} > \left(\frac{dT}{dG}\right)_{eq, sauvage} \quad \text{et} \quad \left|\frac{dF_2}{dG}\right|_{eq, PPAR-/-} > \left|\frac{dF_2}{dG}\right|_{eq, sauvage}$$

Cette relation signifie qu'une mutation des PPAR diminue l'effet tampon sur la quantité de PUFA : l'augmentation de la concentration des PUFA pendant le jeûne est plus important dans les cellules mutée en PPAR que dans les cellules sauvages. Ceci est en accord avec des observations de [BLC<sup>+</sup>04, LCL<sup>+</sup>04].

#### 4.2.4 Discussion

Avec ces travaux, nous avons ainsi pu montrer qu'un modèle extrêmement abstrait reproduit le principal comportement connu du métabolisme, c'est-à-dire le passage d'un régime de synthèse à un régime d'oxydation qui stabilise la quantité d'énergie en remplaçant les apports de nourriture par des réserves. A des échelles de temps courtes, ce passage est assuré par des contrôles métaboliques. A des échelles de temps longues, l'effet de régulation des PUFA renforce le contrôle par les facteurs de transcription LXR et PPAR.

---

**Apport méthodologique** Sur un plan méthodologique, l'intérêt de cette approche est que ces conclusions ont été faites à partir de la seule connaissance des signes des dérivées des réactions par rapport aux variables du système. En particulier, les prédictions sont relativement indépendantes de la valeur numérique de constantes cinétiques.

De plus, l'absence de paramètres numériques, et l'impossibilité de procéder à des réductions dynamiques du système peuvent finalement être vues comme un avantage puisqu'il a permis de mettre en avant le comportement des variables métaboliques alors que des approches de réduction usuelles auraient favorisé les variables génétiques. Cependant, ce zoom sur l'aspect métabolique a un prix : on n'apporte des informations sur la dynamique que via la comparaison entre les états quasi-stationnaires et stationnaires.

**Généricité de l'approche ?** L'enjeu est maintenant de généraliser les critères de réduction pour les rendre utilisables sur des modèles de plus grande taille. Pour la partie *réduction* de cette approche, les travaux de thèse de Vincent Noël (discutés plus haut) sont prometteurs. L'obtention de résultats fins sur les prédictions passera sans doute par des études fines des équations obtenues à la fin du processus de réduction. On touche sans doute ici aux limites de l'automatisation.

**Limites** Les limites du modèle sont bien entendu données par son degré d'abstraction. En particulier, les données de [LCL<sup>+</sup>04] montrent que les PUFA ne se comportent pas tous de la même manière pendant un jeûne chez des souris dans lesquelles PPAR a été muté : en particulier, la concentration de l'acide linoléique (PUFA avec un petit nombre de carbones) est amplifiée, tandis que la concentration des acides gras avec un grand nombre de carbones devient quasiment nulle. Cette observation sous-entend que la désaturation qui permet de passer des PUFA à courtes chaînes aux PUFA à longues chaînes doit être mieux comprise, ce que nous allons faire dans la section suivante en nous plaçant à un autre niveau d'abstraction.

## 4.3 Étude de la réponse d'un système à une influence

Pour mieux comprendre les effets des régulations sur la désaturation dans le métabolisme des acides gras, nous avons cherché à analyser le comportement des enzymes de ces voies et à les mettre en relation avec les principaux régulateurs transcriptionnels du métabolisme. Pour cela, on adopte un nouveau niveau d'abstraction : les modèles sont décrits par un graphe d'interaction, qui met en retrait les transformations entre produits.

### 4.3.1 Modèles différentiels décrits par un graphe d'influence

**Modèles différentiels** On suppose maintenant qu'un système biologique est modélisé par un système différentiel de forme générale, qui intègre toutes les réactions



au sein d'une même fonction  $\mathbf{F}$  :

$$\frac{d\mathbf{X}}{dt} = \mathbf{F}(\mathbf{X}, \mathbf{P})$$

- $\mathbf{X}$  est l'ensemble des *concentrations des molécules* considérées dans le système. La concentration de la molécule  $A$  est noté  $X_{(A)}$ .
- $\mathbf{P}$  est l'ensemble des paramètres de contrôle du système.
- $\mathbf{F}$  est une application continue. La composante correspondant à la molécule  $B$  est notée  $\mathbf{F}_{\mathbf{X}_{(B)}}$ .

**Graphe d'influence** Les informations concernant les régulations transcriptionnelles sur un système biologique sont usuellement représentées sous la forme d'un *graphe d'influence* ou *graphe d'interaction*. Un exemple est donné dans la figure 4.4.

- les sommets du graphe représentent la concentration de molécules au sein du système. Les molécules peuvent être des ARN, des métabolites, des protéines (actives ou non) ou des complexes protéiques. Ces sommets peuvent être étiquetés par +, - ou 0 lorsque les variations du produit considéré pendant une expérimentation sont connues.
- Les arcs représentent les *influences entre ces molécules*. Ces arcs peuvent être étiquetés par + ou - lorsque l'effet de l'influence de la source sur la cible est connu. Plus précisément, on met une flèche d'un sommet  $A$  vers une sommet  $B$  lorsque une variation dans la *concentration* de  $A$  induit un changement dans la *vitesse de production* de  $B$ . Des exemples d'influences positives sont l'activation de la transcription d'un gène ou la formation d'un complexe protéique actif. Des exemples d'influences négatives sont par exemple l'inhibition de la transcription d'un gène ou l'inactivation d'une protéine par un clivage.

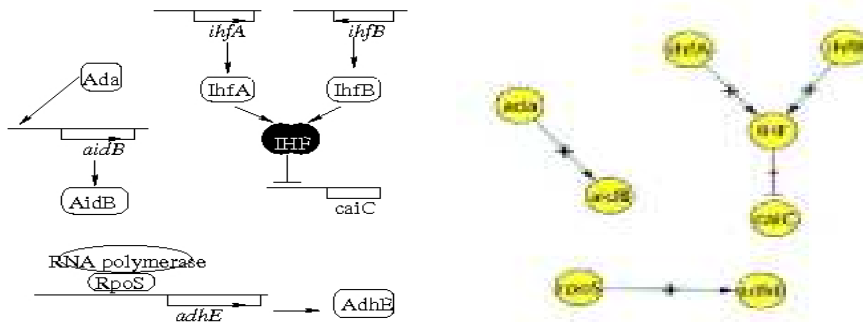


FIG. 4.4 – Gauche : Quelques interactions chez la bactérie *Echerischia coli* : les sommets sous forme de cercles représentent des protéines, les sommets sous forme de cercle plein sont des complexes protéiques, les rectangles représentent des gènes, c'est-à-dire des portions d'une séquence d'ADN.

Droite : les relations entre les molécules permettent de créer un graphe d'influence dont les flèches représentent des actions positives (+) ou négatives (-).

**Des modèles différentiels aux graphes d'influence** Le lien avec la question de l'analyse des données biomoléculaires réside dans les remarques suivantes :

- Le graphe d'influence est relié à la matrice jacobienne de  $\mathbf{F}$ , dans la mesure où il existe une flèche du sommet  $A$  vers le sommet  $B$  si et seulement si  $\frac{\partial \mathbf{F}_{\mathbf{x}(B)}}{\partial X_{(A)}} \neq 0$ . Des études concernant le passage d'un modèle différentiel à un graphe d'influence ont été précisément faites dans [RZL07, FS08].
- Les états stationnaires  $\mathbf{X}$  pour un ensemble de paramètres  $\mathbf{P}$  vérifient la relation  $\mathbf{F}(\mathbf{X}, \mathbf{P}) = \mathbf{0}$ .

**Exemple de la régulation du métabolisme des acides gras** Nous revenons ici à notre exemple référence. La Figure 4.5 présente un graphe d'influence pour les principales enzymes et régulateurs impliqués dans le métabolisme des acides gras. ACL, ACC et FAS sont des enzymes impliquées dans la synthèse des acides gras. La delta-9-désaturase (SCD1) catalyse la synthèse des acides gras mono-insaturés. La delta-5-désaturase (FADS1) et la delta-6-désaturase (FADS2) sont les enzymes clés de la synthèse des acides gras polyinsaturés (notés PUFA).

Comme mentionné plus haut, PPAR $\alpha$  (noté PPAR) est un facteur de transcription qui agit sur l'oxydation des acides gras, en particulier les PUFA. Les facteurs LXR $\alpha$  (noté LXR) et SREBP1 (noté SREBP), par contre, activent la transcription des enzymes de la synthèse. Tous les trois ont une forme active (on adjoint "-a" à leur nom) : PPAR-a est activée par les PUFA tandis que LXR-a est inhibée par les PUFA.

Le point frappant dans ce graphe d'influence est que les désaturases SCD1, FADS1 et FADS2 sont activées à la fois par les protéines SREBP-a et PPAR-a. Ceci est paradoxal dans la mesure où SREBP régule la synthèse des acides gras tandis que PPAR induit leur oxydation ; ces enzymes ont ainsi la capacité théorique de jouer un rôle dans les deux processus que sont la synthèse et l'oxydation.

De plus, les observations pendant les protocoles de mise-à-jeûn chez des poulets et des rats indiquent que, même si ces trois éléments possèdent les mêmes régulateurs, la concentration des ARN de SCD1 augmente alors que celles de FADS1 et FADS2 restent stables. On revient là à des observations paradoxales du même type que celles mentionnées dans la section précédente sur le comportement des acides gras à longue chaîne pendant le jeûne.

La différence entre SCD1 et les désaturases FADS1 et FADS2 réside dans leur actions sur les PUFA : SCD1 est impliqué dans la production d'acides gras mono-insaturés, qui ne sont pas des PUFA, tandis que que FADS1 et FADS2 sont des enzymes de production de PUFA. En particulier, ces régulations induisent deux boucles positives dans le réseau : FADS1-PUFA-PPAR et FADS2-PUFA-PPAR.

Pour mieux comprendre les relations entre ces différents produits, j'ai suggéré de vérifier si les comportements observés pendant un protocole expérimental sur différents produits peuvent être expliqués par les influences indirectes données dans un graphe d'influence. Suivant les idées d'Ovidiu Radulescu, Michel Le Borgne et Philippe Veber, nous avons formalisé cette question en proposant une théorie de la réponse linéaire d'un système à un influence. Ces travaux ont été en particulier

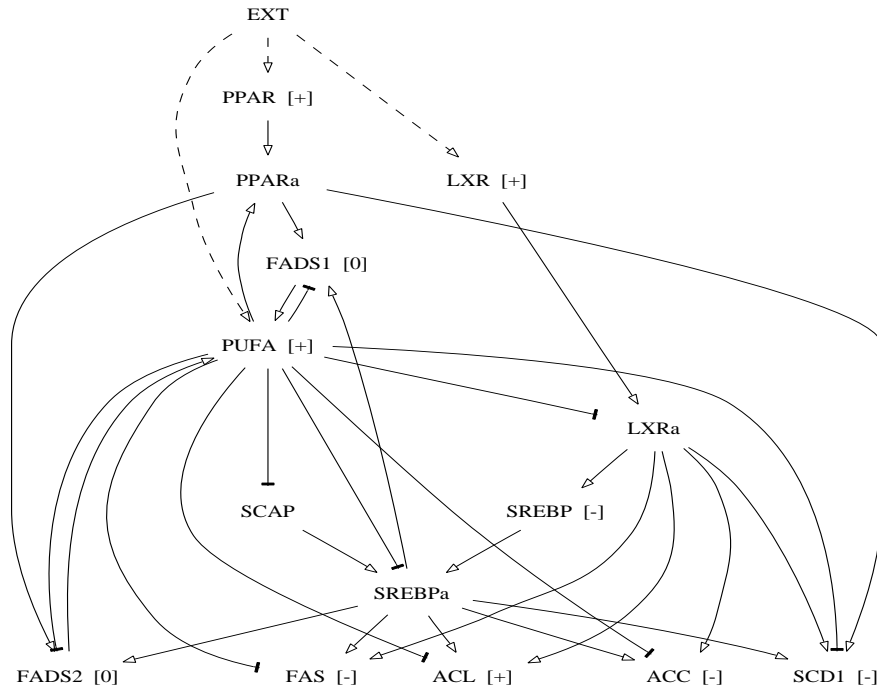


FIG. 4.5 – Graphe d'influence réduit pour les régulations du métabolisme des acides gras. Les boucles d'auto-régulation négatives sont omises. Les flèches qui se terminent par  $->$  signifient que l'interaction est positive tandis que les flèches qui se terminent par  $-|$  correspondent à des interactions négatives. Le symbole *Ext* signifie que les produits cibles sont régulés par des éléments qui ne sont pas intégrés au graphe. Les signes placés entre crochets correspondent aux variations d'ARN observées pendant un jeûne.

publiés dans (Biosystems, 2006) et (Roy. Soc. Inter, 2006), et nous allons maintenant en détailler les principaux résultats.

### 4.3.2 Réponse linéaire d'un système à un influence extérieure

Dans le cas des régulations du métabolisme des acides gras, la question que nous nous sommes posés était de savoir si les comportements différents observés sur SCD1 et FADS1/FADS2 pouvaient être expliqués par les relations connues dans le réseau, ou si, comme certains biologistes le suggèrent, il existe une régulation additionnelle sur FADS1 et FADS2. On doit donc comprendre si les déplacements d'états stationnaires observés suite à un stress (ici, le jeûne) peuvent être globalement expliqués par les entrées du système.

**Réponse linéaire d'un système à un influence extérieure** Une telle problématique n'est rien d'autre qu'un problème de réponse statique d'un système à une influence extérieure. La réponse est considérée comme linéaire lorsque les déplacements sont petits et non linéaire lorsque les déplacements sont importants. Des théories de la réponse linéaire de systèmes dynamiques ont été développées dans différents domaines, qui couvrent la physique [KTH98, Lar88], la mécanique [Mac70], la chimie [OPK73, PO74], ou la biologie [KB73, HR74, KKB<sup>+</sup>02, VAR04].

Dans toutes ces théories, les quantités considérées sont des *susceptibilités* qui représentent la dérivée des sorties par rapport aux entrées. Ces coefficients de susceptibilité sont déduits d'équations qui doivent être compatibles avec les lois de la thermodynamique, le cas non linéaire étant bien plus difficile à manipuler [OPK73, Lar88, Grm01].

Différents travaux ont mis en évidence que les susceptibilités dépendent de la topologie du réseau. Ainsi, les déplacements d'états stationnaires qui correspondent à une petite concentration permettent de calculer la matrice jacobienne d'équations d'évolution d'une cinétique chimique [CSR93, VAR04]. En relation avec cela, des travaux sur le contrôle des réseaux métaboliques [KKB<sup>+</sup>02, CB04] montrent que la matrice des élasticités, dans ce cas là le jacobien du système de réactions, n'est rien d'autre que l'inverse des coefficients de contrôle (dont nous avons déjà parlé dans la section précédente), c'est-à-dire les susceptibilités.

**Formule de Mason-Coates pour les réseaux biologiques** Le problème de Dirichlet consiste à rechercher comment il est possible de déterminer la conductivité d'un système électrique à partir des mesures de courants et de tensions sur ses bords [Cal80, CM00]. Dans (Roy. Soc. Inter, 2006), nous montrons comment des approches de discrétisation permettent de répondre au problème de Dirichlet dans le cadre des systèmes biologiques. Nous y étudions en particulier la réponse de sous-graphes à des modifications sur la frontière d'un système, qui est principalement définie par les informations manquant sur le système. Nous généralisons ainsi la formule de Mason-Coates pour les réseaux électriques [Mas53, Coa59].

**Théorème 4.3.1 ( réponse locale à un influence extérieure (Roy. Soc. Inter, 2006))** Soit  $\mathcal{G}$  un sous-graphe du graphe d'influence d'un réseau  $\mathcal{I}$  qui suit une dynamique différentielle décrite par une fonction  $\mathbf{F}$ . Soit  $\overset{\circ}{G}$  l'intérieur de  $\mathcal{G}$ , c'est-à-dire l'ensemble des sommets qui ne sont pas connectés à un sommet extérieur. On considère  $\overset{\circ}{J}$  la restriction du jacobien de  $\mathbf{F}$  à  $\overset{\circ}{G}$ , calculée en un état stationnaire non dégénéré :  $\overset{\circ}{J}_{i,j} = \frac{\partial F_i}{\partial X_j}$ ,  $i, j \in \overset{\circ}{G}$ .

On suppose que  $\overset{\circ}{\Delta} = \det(\overset{\circ}{J}) \neq 0$  et qu'il n'existe aucune influence directe des paramètres du système sur l'intérieur de  $\mathcal{G}$ .

Alors la réponse de tout élément de  $\mathcal{G}$  à des petites variations sur la frontière est décrite par

$$\delta X_i = \sum_{j \in \overset{\circ}{\mathcal{T}}^{in} G} \sum_{j \rightsquigarrow i \in \mathcal{P}_{\overset{\circ}{G}}} \frac{a_{j \rightsquigarrow i}}{\overset{\circ}{C}_{j \rightsquigarrow i}} \delta X_j. \quad (4.5)$$

– La notation  $\overset{\circ}{\mathcal{T}}^{in} G$  désigne l'ensemble des sommets qui se trouvent sur la frontière de  $G$  (supposés connectés à un sommet extérieur). L'ensemble  $\mathcal{P}_{\overset{\circ}{G}}$  ras-

semble les chemins inclus dans  $\mathcal{G}$  qui partent de la frontière et n'y reviennent pas.

- $a_{j \rightsquigarrow i}$  désigne le produit des coefficients d'interaction  $\frac{\partial F_k}{\partial X_i}$  parcourus en suivant le chemin de  $j$  vers  $i$ .  $l(j \rightarrow i)$  désigne la longueur du chemin de  $j$  vers  $i$ .
- $\check{C}_{j \rightsquigarrow i} = (-1)^{l_{k(j) \rightsquigarrow i} + 1} \frac{\hat{\Delta}}{\Delta_{k(j) \rightsquigarrow i}}$  désigne le module du chemin  $k(j) \rightsquigarrow i$  où  $k(j) \in \mathring{G}$  est le successeur de  $j$  sur le chemin considéré  $a_{j \rightsquigarrow i}$ . Les quantités  $\hat{\Delta}_{k(j) \rightsquigarrow i}$  sont les mineurs du jacobien obtenu en supprimant les lignes et les colonnes qui correspondent aux sommets parcourus sur le chemin  $j \rightsquigarrow i$ .
- Si  $k(j) = i$ , on pose alors  $\check{C}_{j \rightsquigarrow i} = \check{C}_i = -\frac{\hat{\Delta}}{\Delta_i}$

L'intérêt de cette approche réside dans la détermination des signes des modules  $C_i$  et  $\check{C}_{j \rightsquigarrow i}$ . En effet, comme nous l'avons mentionné et utilisé en appliquant le théorème de Gale-Nikaido, les mineurs d'une matrice se décomposent en produits dont les signes sont liés au nombre d'éléments dans les partitions en cycles disjoints du graphe d'influence :

$$C_{j \rightsquigarrow i} = \frac{\sum_{L \in \mathcal{L}(G)} (-1)^{|L|} lp(L)}{\sum_{L \in \mathcal{L}(G_{j \rightsquigarrow i})} (-1)^{|L|} lp(L)} \quad (4.6)$$

$$C_i = \frac{\sum_{L \in \mathcal{L}(G)} (-1)^{|L|} lp(L)}{\sum_{L \in \mathcal{L}(G_i)} (-1)^{|L|} lp(L)}, \quad (4.7)$$

où  $G_{j \rightsquigarrow i}$  désigne le sous-graphe de  $G$  obtenu en supprimant les sommets apparaissant sur le chemin  $j \rightsquigarrow i$  et  $\mathcal{L}(G)$  désigne l'ensemble des partitions de  $G$  en boucles disjointes,  $|L|$  désigne le nombre de boucles dans une partition donnée et  $lp(L)$  désigne le produit des coefficients de la matrice jacobienne qui correspondent aux arêtes considérées dans les boucles.

On en déduit en particulier que si le sous graphe  $\mathcal{G}$  ne contient aucune boucle positive pour un état stationnaire non dégénéré, chaque partition  $L$  contient  $|L|$  cycles négatifs, impliquant que le signe de  $lp(L)$  est lui aussi  $(-1)^{|L|}$ . Ainsi, tous les modules  $C_{j \rightsquigarrow i}$  et  $C_i$  sont positifs (Roy. Soc. Inter, 2006).

**Grandes variations ?** Concrètement, nous pouvons maintenant considérer la situation où nous modifions l'ensemble de paramètres pour que le système se déplace d'un état stationnaire non dégénéré à un autre. En utilisant des conditions reliées au théorème de Gale-Nikaido discuté dans la section précédente, on montre que le système vérifie un résultat d'unicité qui permet d'intégrer les petites variations du Théorème 4.3.1 sur n'importe quel chemin reliant les deux états stationnaires et ne passant pas par des sommets dégénérés.

$$\Delta X_i = \int_{\mathcal{D}} \sum_{j \in \mathbb{T}^{\text{in}} G} \sum_{j \rightsquigarrow i \in \mathcal{P}_{\check{C}}} \frac{a_{j \rightsquigarrow i}}{\check{C}_{j \rightsquigarrow i}} dX_j. \quad (4.8)$$

Cependant, cette relation devient fautive lorsque le système rencontre un point singulier, où il perd sa stabilité et le déplacement d'état d'équilibre n'est pas continu.

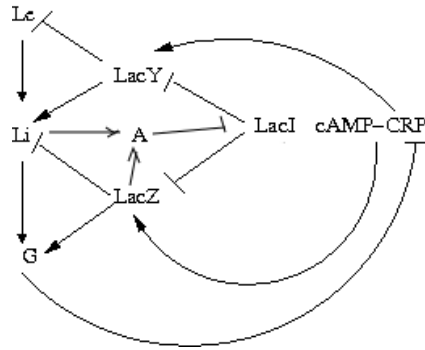


FIG. 4.6 – Graphe d'influence pour les interactions concernant le métabolisme du lactose chez *E. coli* (opéron lactose). Les arêtes qui se terminent par ">" indiquent une activation tandis que les arêtes qui se terminent par "-|" indiquent une inhibition. Les boucles de retro-régulation négatives sur chaque sommet ont été omises.

Il faut alors modifier l'équation 4.8 en y ajoutant un saut fini (voir discussion dans (Roy. Soc. Inter, 2006) pour éclairer la question de la robustesse et des conditions de changement d'états.).

### 4.3.3 Transport et métabolisme du lactose chez *E. coli*

L'étude des influences dans un graphe d'influence peut aussi être utilisée à des fins de prédiction. Illustrons cela sur le modèle du métabolisme du lactose chez la bactérie *E. coli*.

**Opéron lactose** Le modèle de l'opéron-lactose a été étudié par Jacob et Monod dans les années 1960 [JM61]. Il s'agit d'un travail fondateur qui a mis en évidence l'importance des régulations génétiques pour expliquer le comportement d'un système biologique.

Le métabolisme du lactose chez *E. coli* nécessite la présence de 3 enzymes : la  $\beta$ -galactosidase *LacZ*, la perméase *LacY* et la transacétylase *LacA*. La perméase *LacY* transporte le lactose de l'extérieur de la cellule (*Le*) à l'intérieur du milieu cellulaire (*Li*). *LacZ* hydrolyse le lactose *Li* en glucose (*G*) et en allolactose (*A*). Les protéines *LacY* et *LacZ* sont sous contrôle direct d'un inhibiteur appelé *LacI*. Inversement, un complexe formé d'AMP cyclique et de la protéine CRP (*cAMP - CRP*), agit sur l'activité des ARN-polymérase et permet d'activer fortement la transcription de *LacY* et *LacZ*. Enfin, *A* a une action inhibitrice sur *LacI* via la formation d'un complexe. Ces connaissances permettent de construire un graphe d'influence montré dans la figure 4.6. Il faut noter que *Le* est une entrée du système, contrôlé par l'expérimentateur (il s'agit noeud externe).

**Influences respectives des boucles** Considérons le cas où le lactose extérieur  $Le$  augmente. On sait alors que le système passe d'un état pauvre en lactose interne  $Li$  où la transcription est bloquée à un état riche en lactose interne où la transcription est activée. Cela revient à dire que la variation  $\frac{\partial X_{Li}}{\partial Le}$  est positive.

Or, le théorème 4.3.1 implique que

$$\delta Li = \frac{a_{Le \rightsquigarrow Li}}{\dot{C}_{Le \rightsquigarrow Li}} \delta Le,$$

avec  $a_{Le \rightsquigarrow Li} > 0$ .

On explique  $\dot{C}_{Le \rightsquigarrow Li}$  en considérant les boucles du système qui ne passent pas par  $Le$  :

$$\begin{aligned} l_1^+ &= \{Li, A, LacI, LacY, Li\}, \\ l_2^+ &= \{Li, G, cAMP - CRP, LacZ, Li\}, \\ l_3^+ &= \{LacZ, A, LacI\}, \\ l_1^- &= \{Li, A, LacI, LacZ, G, cAMP - CRP, LacY, Li\}, \\ l_2^- &= \{Li, G, cAMP - CRP, LacY, Li\}, \\ l_3^- &= \{G, cAMP - CRP, LacZ\}, \\ l_4^- &= \{Li, A, LacI, LacZ, Li\}, \end{aligned}$$

Pour calculer  $\dot{C}_{Le \rightsquigarrow Li}$ , on recherche toutes les partitions en boucles disjointes du graphe d'influence privé de  $Le$ , puis du graphe d'influence privé de  $Li$  et  $Le$ . Dans l'équation suivante, pour chaque boucle  $l$ ,  $\tilde{l}p(l)$  désigne le produit des coefficients d'interactions ( $\frac{\partial X_a}{\partial b}$  sur les arcs  $a \rightarrow b$  contenus dans  $l$ , divisé par les coefficients d'auto-régulation ( $\frac{\partial X_n}{\partial n}$  des nœuds  $n$  de la boucle. Le signe de  $\tilde{l}p(l)$  est égal au produit des signes des arcs contenus dans  $l$ .

$$\begin{aligned} \dot{C}_{Le \rightsquigarrow Li} &= \left( \frac{\partial X_{Li}}{\partial Le} \right)^{-1} \left( 1 - \tilde{l}p(l_4^+) - \tilde{l}p(l_3^-) \right)^{-1} \left( 1 - \tilde{l}p(l_1^+) - \tilde{l}p(l_2^+) - \tilde{l}p(l_3^+) \right. \\ &\quad \left. - \tilde{l}p(l_1^-) - \tilde{l}p(l_2^-) - \tilde{l}p(l_3^-) - \tilde{l}p(l_4^-) + \tilde{l}p(l_4^+) \tilde{l}p(l_2^-) + \tilde{l}p(l_1^+) \tilde{l}p(l_3^-) \right). \end{aligned} \quad (4.9)$$

Si  $Li$  croît en fonction de  $Le$ , on obtient une contrainte non triviale sur les différentes contribution des boucles du système à son fonctionnement :

$$\frac{\partial X_{Li}}{\partial Le} > 0 \text{ ssi } \sum_{i=1}^3 \tilde{l}p(l_i^+) < 1 - \sum_{i=1}^3 \tilde{l}p(l_i^-) - \tilde{l}p(l_2^+) \tilde{l}p(l_4^-) + \tilde{l}p(l_1^+) \tilde{l}p(l_3^-).$$

**Effets de mutations** En considérant la formule 4.9, on peut aussi prédire les effets de mutations sur le système. par exemple, une mutation sur  $CRP$  ou sur l'AMP cyclique implique une inactivation des boucles  $l_2^+, l_1^-, l_2^-, l_3^-$ . en particulier,  $\dot{C}_{Le \rightsquigarrow Li}$  va diminuer dès que le système sauvage vérifie

$$\tilde{l}p(l_2^+) - \tilde{l}p(l_1^+) \tilde{l}p(l_3^-) - \tilde{l}p(l_4^+) \tilde{l}p(l_2^-) > -(\tilde{l}p(l_1^-) + \tilde{l}p(l_2^-) + \tilde{l}p(l_3^-)).$$

L'intensité du flux de lactose entrant sera alors moins importante que chez la bactérie sauvage. Si l'inégalité inverse est vraie, le flux de lactose augmentera.

### 4.3.4 Régulation du métabolisme des acides gras

Cette approche permet aussi de mieux comprendre les effets des différents éléments du système sur la production des désaturases  $FADS1$  et  $FADS2$ .

**Des boucles et des chemins...** Pour cela, on considère les boucles du modèle donné dans la Figure 4.5.

$$\begin{aligned} l_1^- &= \{PUFA, SCAP, SREBP_a, FADS2, PUFA\}, \\ l_2^- &= \{PUFA, SREBP_a, FADS2, PUFA\}, \\ l_3^- &= \{PUFA, LXRa, SREBP, SREBP_a, FADS2, PUFA\}, \\ l_4^- &= \{PUFA, FADS2, PUFA\}, \\ l^+ &= \{PUFA, PPAR_a, FADS2, PUFA\}. \end{aligned}$$

On considère aussi des chemins qui passent par les produits  $PUFA$ ,  $FADS2$  et  $LXR$  :

$$\begin{aligned} p &= \{PUFA, LXR - a, SREBP\}, \\ p' &= \{LXR, LXR - a, SREBP\}, \\ p_1 &= \{PUFA, FADS2\}, \\ p_2 &= \{PUFA, SCAP, SREBP_a, FADS2\}, \\ p_3 &= \{FADS2, PUFA\}, \\ p_4 &= \{PUFA, LXRa, SREBP, SREBP_a, FADS2\}, \\ p_5 &= \{LXR, LXRa, SREBP, SREBP_a, FADS2\}, \\ p_6 &= \{PPAR, PPAR_a, FADS2\}, \\ p_7 &= \{PUFA, PPAR_a, FADS2\}, \\ p_8 &= \{PUFA, SREBP_a, FADS2\}, \\ p_9 &= \{PPAR, PPAR_a, FADS2, PUFA\}, \\ p_{10} &= \{LXR, LXRa, SREBP, SREBP_a, FADS2, PUFA\}. \end{aligned}$$

**Rôle de la double régulation des désaturases par SREBP et PPAR** On applique le théorème 4.3.1 au graphe

$$G_3 = \{PUFA, PPAR, PPAR_a, FADS2, SREBP_a, SREBP, LXRa, LXR, SCAP\}.$$

Selon les informations principales de la littérature, on suppose que toutes les interactions sur les éléments de ce système passent par  $PPAR$ ,  $LXR$  et  $PUFA$ . Ce sont donc les seuls éléments du graphe qui ont une variation forcée  $\delta X^f$ . Avec le théorème 4.3.1, on peut exprimer les variations réelles du système en fonction des variations forcées sur  $PPAR$ ,  $LXR$  et  $PUFA$ .

$$\delta PUFA = \frac{1}{\tilde{C}} \left[ \frac{\delta PUFA^f}{\chi_X} + \tilde{a}_{p_3} \delta FADS2^f + \tilde{a}_{p_9} \delta PPAR^f + \tilde{a}_{p_{10}} \delta LXR^f \right] \quad (4.10)$$

$$\tilde{C} = 1 - \tilde{l}_p(l^+) - \tilde{l}_p(l_1^-) - \tilde{l}_p(l_2^-) - \tilde{l}_p(l_3^-) - \tilde{l}_p(l_4^-). \quad (4.11)$$

Dans ces équations, les quantités  $\tilde{a}_p$  sont obtenues comme des produits des éléments de la matrice jacobienne qui correspondent aux arêtes parcourues par le chemin  $p$ .

L'équation (4.10) permet de mieux comprendre le rôle de la régulation duale de  $FADS1$  et  $FADS2$  par  $PPAR$  et  $SREBP$ . En effet, la boucle positive  $l^+$  qui



se passe par *PUFA*, *FADS2* et *PPAR* diminue la valeur de  $\tilde{C}$ , et selon (4.10), va augmenter les variations de  $\delta PUFA$ . L'interprétation biologique est la suivante : la double régulation des désaturases par *PPAR* et *SREBP* induit un effet de réaction compensatoire à la demande croissante de PUFA causées par l'oxydation pendant le jeûne. Cette interprétation est discutée en termes biologiques dans [NN03a].

**Régulation des désaturases** Le théorème 4.3.1 permet aussi d'obtenir la relation suivante sur les variations de *FADS2* :

$$\delta FADS2 = \frac{1}{\tilde{C}} [(\tilde{a}_{p_1} + \tilde{a}_{p_2} + \tilde{a}_{p_4} + \tilde{a}_{p_7} + \tilde{a}_{p_8})\delta PUFA^f + \tilde{a}_{p_5}\delta LXR^f + \tilde{a}_{p_6}\delta PPAR^f]$$

Dans cette équation, les chemins  $p_1, p_2, p_4, p_8$  ont des influences négatives tandis que les chemins  $p_7, p_5, p_6$  ont des influences positives. La conclusion biologique est que, intrinsèquement, l'existence de boucles positives dans le système implique qu'il a la *capacité théorique* pour que les désaturases *FADS1* et *FADS2* ne varient pas pendant le jeûne.

En particulier, contrairement à la conclusion qui est faite habituellement, *on ne peut pas conclure a priori à partir de l'absence de variation de FADS1 et FADS2 pendant le jeûne qu'une régulation autre que celle de SREBP et PPAR existe sur ces protéines*. La présence de boucles positives fausse tout ce raisonnement intuitif.

### 4.3.5 Ajustement numérique pour analyser l'influence des voies

Avec cette dernière approche, le rôle des PUFA et des désaturases commence à être précisé mais des inconnues demeurent. Dans le cadre de sa thèse (co-encadrement INRA-INRIA), Pierre Blavy a travaillé sur la régulation des désaturations, en s'inspirant du modèle abstrait présenté Fig. 4.2. Son travail a en particulier consisté à exploiter des données temporelles sur le jeûne de souris sauvages et mutées en *PPAR*, pour mieux comprendre les mécanismes de transformations entre acides gras, et leurs régulations.

**Modèle abstrait des entrées des acides gras** L'approche adoptée a consisté à partir d'un modèle volontairement abstrait pour les mécanismes actifs pendant le jeûne. Pour les souris mutées pendant le jeûne, seule la bêta-oxydation est active. De plus, les désaturations pour l'obtention des PUFA sont inactives, puisqu'elles sont régulées par *FADS1* et *FADS2* qui sont elles-mêmes régulées par *PPAR $\alpha$*  et *SREBP* (or *PPAR $\alpha$*  n'est pas présent chez les souris mutées et l'activité de *SREBP* décroît pendant le jeûne). Le modèle abstrait qui a été considéré contient donc simplement

- une entrée des acides gras dans le foie à partir du tissu adipeux,
- des bêta-oxydations,
- différentes transformations métaboliques entre les acides gras (sauf les désaturations qui sont inactivées).

La question principale était de savoir si ce modèle différentiel extrêmement générique pouvait se coupler avec les données, les paramètres pour le taux d'entrée

---

et la bêta-oxydation étant indépendants de la classe d'acide gras. Les conclusions, détaillées dans les actes de la conférence JOBIM (JOBIM, 2008), sont les suivantes.

**Régulation des désaturases** Les données pour les souris mutées montrent que la production de l'acide gras C22 :6 $\omega$ 3 par désaturation du C18 :3 $\omega$ 3 est active chez ces espèces, avec un taux d'accumulation constant. Or, si les désaturations étaient inactives dans le modèle, on devrait observer une diminution de la production de C22 :6 $\omega$ 3 chez les souris mutées. On aboutit ici à une contradiction puisque les données indiquent que l'activité de FADS1 et FADS2 est non nulle chez les souris mutées pendant le jeûne, ce qui suggère fortement l'existence d'un régulateur additionnel pour ces deux protéines.

Pour vérifier cette hypothèse, nos collaborateurs à l'INRA (Pascal Martin et Hervé Guillou, Toulouse) ont procédé à des mesures additionnelles (PCR quantitative) sur les ARNs de FADS1 et FADS2. Comme nous l'avions prédit, les concentrations d'ARN sont non nulles, et il existe donc sans doute un régulateur inconnu aux désaturases.

D'un point de vue méthodologique, ce résultat illustre l'importance du niveau de modélisation et du choix de l'expérimentation pour obtenir une conclusion sur le comportement d'un système. Dans la section précédente, nous avons expliqué que des données de jeûne chez des souris sauvages ne suffisaient pas pour conclure à l'existence de ce régulateur inconnu pour FADS1 et FADS2, à cause de la présence de boucles de régulations positives dans le système qui faussent les analyses intuitives du comportement du système. Ici, l'expérimentation s'est faite chez des souris mutées, et la construction d'un modèle qui prend en compte toutes les transformations entre les acides gras a permis de conclure positivement.

**Rôle de PPAR dans la régulation de la bêta-oxydation** Les analyses de P. Blavy ont aussi mis en évidence une différence dans les taux d'oxydation entre les souris mutées et sauvages. Cela suggère que PPAR- $\alpha$  a un effet sur l'oxydation indépendant de l'état nutritionnel de l'individu. Le résultat important est que l'ajustement peut-être fait avec un taux constant en fonction du temps, chez les souris mutées comme chez les souris sauvages. On en déduit que le rôle de PPAR $\alpha$  dans la régulation de l'oxydation pendant le jeûne (via les PUFA), que nous avons discuté dans la section 4.2.3, est en fait négligeable par rapport à son rôle constitutif comme régulateur de l'expression des enzymes de l'oxydation elle-même.

## 4.4 Conclusion

Dans ce chapitre, nous avons donc illustré certaines spécificités de la modélisation de systèmes biologiques au sujet desquels l'obtention de paramètres ne peut être réalisée que sur une très petite portion.

Les questions principales qui se sont posées ont été

- Une étude globale de l'unicité des états stationnaires du système.

- La recherche de prédictions globales sur le rôle des régulations génétiques sur le système métabolique, en particulier l'intérêt pour le système d'avoir des contrôles à des échelles de temps courtes et longues.
- L'effet de grosses perturbations sur le système, via des inhibitions de voies de régulation.
- Le rôle de voies métaboliques qui a priori ne se distinguaient pas des autres, à savoir la désaturation des acides gras polyinsaturés.

Pour toutes ces questions, nous avons considéré des modèles volontairement abstraits et non paramétrés, à *des niveaux d'abstraction adaptés à la question posée*. C'est une lapalissade de dire qu'un modèle doit être construit en fonction des questions posées sur le système considéré ; cette évidence doit cependant rester systématiquement à l'esprit en biologie moléculaire, à cause de la difficulté inhérente à ce domaine pour obtenir des observations fiables sur le système.

Puisque, dans notre exemple récurrent, les mécanismes couplaient des réactions métaboliques et des régulations génétiques, nous avons travaillé à la fois sur la matrice des signes des dérivées du modèle de réactions, et sur le graphe d'influence du système. En se concentrant sur l'étude des déplacements d'états stationnaires avec des méthodes algébriques, nous avons pu apporter une réponse biologique aux questions mentionnées plus haut.

Cette approche abstraite a aussi permis d'identifier les zones du modèle qui étaient trop abstraites dans la littérature et méritaient une étude attentive, sur laquelle des modélisations plus fines ont été réalisées.

L'intérêt de cette démarche se trouve donc dans la complémentarité des niveaux d'abstraction considérés, et dans l'apport de méthodes algébriques pour élucider le comportement d'un système biologique.

Dans le chapitre suivant, nous allons nous concentrer sur les types de raisonnements qui permettent de définir une prédiction ou d'aboutir à une contradiction. Pour cela, nous allons considérer une version simplifiée du théorème 4.3.1 et l'utiliser comme outil de diagnostic systématique sur des réseaux de grande taille.

---

## Chapitre 5

# De la dynamique à l'aide au raisonnement par l'analyse de contraintes

Dans le chapitre précédent, nous avons montré comment des approches algébriques associées à différents niveaux d'abstraction de modèles permettent de mieux comprendre le rôle de certains éléments d'un réseau. Dans ce chapitre, nous allons nous intéresser à l'automatisation des raisonnements pour étudier un réseau. Il va donc s'agir d'identifier différentes questions fondamentales du point de vue de l'expérimentateur, et d'apporter un cadre de réponse générique à ces questions en s'appuyant le théorème 4.3.1 qui décrit la réponse linéaire d'un système.

### 5.1 Première question : comment construire des réseaux sur lesquels appliquer des simulations dynamiques ?

Comme nous l'avons vu au chapitre 4, il existe un certain nombre de méthodes inspirées par les systèmes dynamiques qui permettent d'étudier un système engendré par la biologie. Cependant, et cela est apparu lors de nos études du métabolisme des acides gras, ces travaux sont limités par la quantité d'information disponible sur un système. En particulier, les approches d'analyse dynamique (même qualitative) ne peuvent porter que sur les interactions connues et assez bien détaillées au sein d'un réseau. Or, un des objectifs majeurs de la biologie moléculaire est de découvrir de nouvelles interactions, et de comprendre leurs effets. Par exemple, dans le cas du métabolisme des acides gras, comment identifier le régulateur inconnu des désaturases ?

Par ailleurs, en particulier sur les systèmes eucaryotes, vu la complexité des organismes, il n'est possible d'acquérir des données pour une dynamique précise

que pour une très petite partie du système. La question qui se pose est donc d'aider le biologiste à identifier, au sein de l'ensemble des molécules qui sont actives dans son système, un tout petit nombre de produits sur lesquels une modélisation fine pourra être appliquée.

Ce domaine ne fait a priori plus partie de l'étude de systèmes dynamiques. Il s'agit de travailler à partir des observations et des connaissances sur un système pour suggérer l'existence d'interactions que des expérimentations précises permettront de valider. Cependant, puisqu'on recherche des *régulateurs* à partir d'*observations* sur un système, les systèmes dynamiques offrent un cadre utile à la production d'hypothèses.

### 5.1.1 Différentes approches pour la construction de modèles

**Construction de modèles** La première question qui se pose ici est d'exploiter les données pour identifier les interactions dans un système. On entre dans le domaine de la *construction de modèles*. Pour cela, la stratégie la plus naturelle consiste à rassembler les connaissances dans la littérature, ce qu'on fait naturellement pour des modèles de petite taille. À très grande échelle, cette démarche a été faite pour l'organisme modèle qu'est la bactérie *E. coli* [SGCPG<sup>+</sup>06] ; il a été récemment estimé que même pour cet organisme, on connaît au plus 30% des interactions existant. Pour les autres organismes, des travaux similaires ont lieu portés par des applications en médecine en particulier : citons entre autres KEGG, ou les bases proposées par les sociétés Ingenuity ou Biobase [IS98, SCG<sup>+</sup>01, HBC<sup>+</sup>01, BBH03, ZMPQ<sup>+</sup>02, KGK<sup>+</sup>04, HMPL<sup>+</sup>04] ... Cependant, ces travaux sont moins avancés que ceux sur *E. coli* et aucune banque de données n'est à ce jour suffisamment renseignée pour permettre la construction d'un modèle de régulation de manière rapide, fiable et efficace.

Pour reconstruire des modèles, des approches statistiques ont aussi été proposées, exploitant les données grande échelle disponibles, et les couplant avec les informations stockées dans les banques de données [Fri04, STV04, JP06, BBAIdB07]. Le premier exemple a porté sur le réseau d'interaction de la levure, pour laquelle des données d'immuno-précipitation couplées avec des algorithmes de comparaison de séquences ont permis de proposer à la communauté des réseaux transcriptionnels pour la levure [LRR<sup>+</sup>02, HGL<sup>+</sup>04, MWG<sup>+</sup>06]. Or, ces réseaux "benchmark" n'ont pas encore été validés expérimentalement, et de nouvelles approches apparaissent régulièrement, qui font intervenir des méthodes variées (corrélations, information mutuelle, études de causalité, réseaux Bayésiens, analyses de chemins, équations différentielles...) (voir [BBAIdB07] et la discussion dans (BMC bioinfo, 2008)). Force est de constater que les résultats de ces multiples travaux divergent considérablement [FMS<sup>+</sup>07] ; d'autant plus que peu de sources permettent effectivement de valider les réseaux à partir de publications (mis à part [RBB<sup>+</sup>06, GBBK02, HBC<sup>+</sup>01]).

Enfin, une dernière stratégie pour reconstruire des réseaux (principalement biochimiques) a consisté à isoler des interactions à partir de raisonnements logiques sur les transformations entre molécules [JDSZ05, BV03, KGC05a]. Elle concerne principalement des réseaux de petite taille.

**Diagnostic** Les réseaux grande échelle qui apparaissent actuellement sont finalement une compilation d'interactions déduites de différentes méthodes et expérimentations sous différentes conditions. Ainsi, les différentes sources utilisées pour construire ces réseaux introduisent des erreurs, ce qui rend les biologistes dubitatifs sur les études faites sur ces systèmes. Depuis quelques années, différents auteurs mettent ainsi en avant le besoin de méthodes pour concilier différentes données hétérogènes, repérer les incohérences pour raffiner et améliorer les modèles [IGH01, Pal00, Kit02b]. L'objectif est alors double :

- Vérifier que le modèle est cohérent avec les observations disponibles, et sinon, proposer des corrections.
- Proposer des prédictions pour le comportement du système. Ces prédictions peuvent ensuite être validées par la littérature. Elles peuvent aussi mener à la proposition d'expérimentation, dans une boucle qui alimente le modèle avec de nouvelles données pouvant potentiellement l'invalider à nouveau.

Pour répondre à ces deux questions, différentes méthodes ont été proposées. Elles diffèrent par le type de règles sur lesquelles s'appuient la notion de cohérence ou les prédictions. Cependant, on retrouve dans la plupart une règle implicite, appelée *règle causale*, qui affirme que la modification de l'expression d'une protéine lorsque un gène est supprimé ou sur-exprimé implique qu'une régulation (au moins indirecte) existe entre le gène manipulé et la protéine observée [DWFS99]. Cette règle a ainsi été utilisée pour procéder à des test globaux de compatibilité [GRRL<sup>+</sup>03], pour vérifier la cohérence de systèmes avec des raisonnements logiques [BSPL03, JDSZ05], pour suggérer des voies de régulations actives en s'appuyant sur les connaissances d'interaction protéines-protéines [YIJ04, YMM<sup>+</sup>05, Ide04] ou pour analyser la cohérence de données d'activités des voies métaboliques [HLPP06].

**Contribution générale : utilisation d'une règle causale dans un cadre dynamique réduit pour le diagnostic de modèles** J'ai travaillé à l'IRISA avec Michel Le Borgne, Ovidiu Radulescu et les doctorants Philippe Veber et Carito Guziolowski sur l'élaboration d'une méthode pour le diagnostic et l'analyse de réseaux qui tient compte du manque de précision des données et des connaissances, basées sur une règle causale.

Nous avons ainsi proposé une approche pour isoler les modules incohérents dans un réseau. La première étape consiste à construire un système de contraintes qui clarifient et formalisent le raisonnement intuitif des biologistes. Pour cela, nous avons étudié le cadre de validité exact de la règle causale mentionnée ci-dessus : même si d'un point de vue dynamique les modèles causaux sont trop restrictifs, ils ont toute légitimité pour être utilisés pour étudier des données portant sur les comparaisons d'états stationnaires.

S'appuyant sur cette règle causale, nous avons interprété les objectifs principaux du biologiste (cohérence globale, aide à la correction, prédiction) comme des propriétés de l'ensemble des solutions du système. Il a fallu proposer des méthodes efficaces de résolution du système de contraintes, adaptées à l'étude de réseaux de régulation de grande taille tels que celui de *E. coli* : nous avons ainsi montré que nos méthodes sont aussi fiables que les autres méthodes disponibles pour le diagnostic

et la prédiction [HLPP06] mais peuvent de plus être appliquées sur des réseaux de taille bien plus grande.

Nous avons ensuite cherché à étendre le cadre d'application de ces systèmes de contraintes, en particulier à la recherche du mode de régulation (activation ou inhibition) des facteurs de transcription. Nous travaillons maintenant sur l'élaboration de plans expérimentaux.

Ces résultats sont vraiment le fruit de la complémentarité entre divers chercheurs de cultures différentes. Ma propre contribution a porté sur l'identification des questions qui pouvaient se poser en fonction des connaissances et des données disponibles sur un réseau et son comportement. Nous allons donc maintenant détailler les résultats rapidement mentionnés ici en insistant sur le processus de formalisation des questions sur un système et des approches de résolution.

### 5.1.2 Formalisation : deux questions et une règle intuitive

**Des questions en biologie...** Si nous reprenons les résultats décrits au chapitre 4, les questions biologiques qui se sont posées entraînent dans plusieurs catégories :

- Nous avons cherché à savoir si les connaissances sont suffisantes pour expliquer des observations. Cette question s'est particulièrement posée pour les désaturases dont le comportement, intuitivement, n'était pas celui qui était attendu mais pour lequel nous avons vu que les conclusions intuitives n'étaient pas nécessairement justifiées.
- Nous avons cherché à savoir à quels comportements du système il fallait s'attendre après avoir perturbé le système, par exemple en inactivant (par mutation) le gène de PPAR dans le métabolisme des acides gras, ou en inactivant les régulateurs de l'opéron lactose.
- Nous avons cherché à comprendre le fonctionnement du système, c'est-à-dire à identifier l'effet de variations de certains produits sur le comportement global du système.

**... aux requêtes en informatique ?** D'un point de vue bioinformatique, pour apporter une réponse plus précise à ces questions, on se doit d'isoler des concepts sous-jacents à ces objectifs et de les formaliser en requêtes avec une résolution efficace. Nous avons identifié deux concepts principaux derrière les questions biologiques :

- Le *diagnostic* de réseau, c'est-à-dire, décider si l'ensemble des connaissances intégrées dans un modèle est suffisant pour expliquer un comportement donné.
- La *prédiction*, qui revient à trouver le comportement du système suite à une perturbation.

Ces questions apparaissent dans de nombreux domaines des sciences expérimentales. La spécificité ici est de leur apporter une réponse qui est adaptée au type de connaissances sur un système (représenté sous la forme de graphe d'influence) et au type d'observations (souvent des données de transcriptome, qualitatives, avec très peu de points temporels quand il y en a plusieurs).

**Règle intuitive** Nous avons travaillé sur les raisonnements intuitifs des biologistes et cherché à les formaliser. Après une lecture approfondie de la littérature et de nombreuses discussions avec des biologistes, il nous est apparu que les raisonnements utilisés implicitement par les biologistes pour conclure à l'invalidité d'un modèle (et donc à l'existence de régulations additionnelles) ou faire des prédictions étaient basées sur une simple règle causale très intuitive : *la variation d'un élément doit être qualitativement expliquée par la variation d'au moins un prédécesseur dans le graphe d'influence.*

Un exemple est donné dans la figure 5.1, où  $A$  et  $B$  sont deux protéines qui activent la transcription de  $C$ . Plusieurs cas peuvent se produire lors d'un stress environnemental :

- Si la concentration de  $A$  et  $B$  augmentent lors du stress, alors  $C$  doit nécessairement augmenter. Ou plutôt il *ne peut pas diminuer*...
- En particulier, si  $A$  et  $B$  augmentent et qu'une observation établit que  $C$  diminue simultanément, la règle de cohérence n'est pas respectée et on doit conclure que soit les observations sont erronées, soit il manque une interaction dans le réseau. Dans tous les cas, nous détectons un *conflit* dans le réseau.
- Si la concentration de  $A$  augmente et celle de  $B$  diminue, alors  $C$  peut aussi bien augmenter que diminuer, et on ne peut rien conclure.

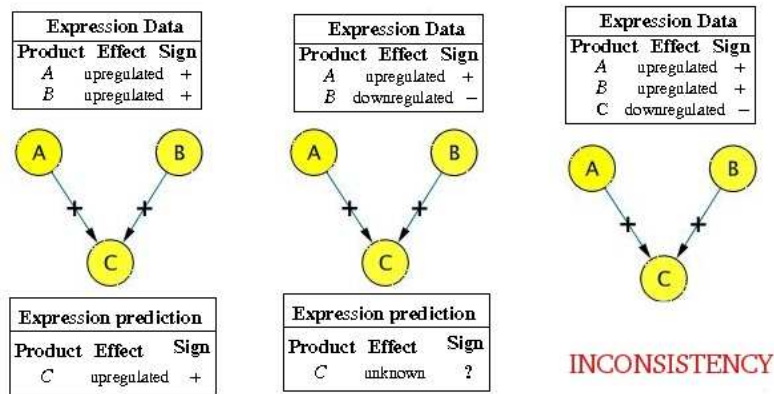


FIG. 5.1 – Illustration du processus de vérification de cohérence. Gauche : lorsque les influences de  $A$  et  $B$  sont de même signe, on en déduit que  $C$  augmente. Milieu : Dans certains cas, aucune prédiction n'est possible. Droite : Les données fournies ne sont pas cohérentes avec le réseau.

### 5.1.3 De l'intuition à une contrainte satisfaite par une régulation

**Domaine de validité de la règle intuitive ?** Le problème est que cette règle est intuitive et, mal utilisée, elle peut mener à des erreurs de raisonnement : nous l'avons vu dans le cadre du métabolisme des acides gras, où le fait que FADS1 et FADS2



ne varient pas pendant un jeûne peut être expliqué par des influences indirectes dans le réseau, alors qu'on s'attend intuitivement à l'existence d'une régulation additionnelle. Dans ce cas là, d'autres expérimentations appuient la thèse de la régulation additionnelle, mais cet exemple est intéressant pour illustrer les erreurs de raisonnement qu'on peut faire si on ne formalise pas les raisonnements utilisés et les hypothèses sous-jacentes.

Ainsi, les modèles causaux ont été utilisés puis abandonnés dans les années 1990 pour modéliser la dynamique de réseaux de régulation, parce que non adaptés à un cadre dynamique. Nous avons souhaité reprendre ces modèles causaux dans un cadre plus restreint. En effet, même si les modèles causaux ne sont pas adaptés pour étudier la *dynamique* des réseaux biologiques, nous avons cherché à savoir si ces modèles peuvent être utilisés dans un cadre de *déplacements d'états stationnaires*. On ne cherche plus alors à décrire l'état d'une molécule au moment  $n + 1$  en fonction de son état au temps  $n$ , mais à poser une contrainte globale sur les variations entre deux états stationnaires. Nous allons expliquer maintenant pourquoi que cette règle est vraie sous des conditions raisonnables d'un point de vue biologique.

**Déplacements d'états stationnaires** La restriction à des considérations sur des états stationnaires est naturelle dès lors qu'on considère le type de données généralement utilisées pour analyser des systèmes : la majeure partie est constituée d'observation du transcriptome à grande échelle, sous la forme de puces à ADN qui comparent la quantités d'ARN entre le début et la fin d'une expérimentation. Hormis les cas où les systèmes considérés ont des comportements périodiques, le début et la fin d'une expérimentation sont des états stationnaires d'un système dynamique.

Ainsi, un *stress environnemental* désigne l'effet d'une perturbation des conditions initiales. Il se modélise par le passage continu d'un état stationnaire pour un ensemble de paramètres  $\mathbf{P}_1$  à un état stationnaire pour un ensemble de paramètres  $\mathbf{P}_2$ .

Finalement, on cherche donc à comprendre comment exploiter des informations relatives à un déplacement d'états stationnaires sachant des informations partielles sur un système.

Dans (Biosystems, 2006), nous considérons le théorème 4.3.1 sur la réponse linéaire d'un système pour comprendre l'effet des prédécesseurs d'un sommet d'un graphe d'influence sur ce sommet. Le théorème 4.3.1 porte sur des variations locales, nous avons cherché à savoir quelles conditions étaient suffisantes pour en déduire une information sur de grandes variations.

**Théorème 5.1.1** *Si on suppose que pour une molécule  $i$  :*

1. *Il existe une constante  $C$  telle que  $\frac{\partial \mathbf{F}_{X(i)}}{\partial X(i)} < -C$ ,  $C > 0$ .*
2. *Il n'y a pas d'influence directe de  $\mathbf{P}$  sur  $X(i)$ .*
3.  $\mathbf{F}_{X(i)}(\{\mathbf{X}, X(i) = 0\}) > 0$ .
4. *Pour toute flèche  $k \rightarrow i$  dans le graphe d'influence, le signe de  $\frac{\partial \mathbf{F}_{X(i)}}{\partial X(k)}$  est constant sur une plage d'expérimentation.*

Alors les signes des variations entre deux états stationnaires consécutifs à un stress environnemental vérifient la contrainte suivante :

$$\text{signe}(\Delta X_{(i)}) \simeq \sum_{k \neq i, k \rightarrow i} \text{signe} \left( \frac{\partial \mathbf{F}_{X_{(i)}}}{\partial X_{(k)}} \right) \times \text{signe}(\Delta X_{(k)}).$$

Par exemple, si on considère le graphe d'influence de l'opéron-lactose, on en déduit que l'équation qui régit la variation de *LacY* est donné par l'équation :

$$\text{signe}(\Delta \text{LacY}) \simeq \text{signe}(\Delta \text{cAMP} - \text{CRP}) - \text{signe}(\Delta \text{LacI}).$$

**Spécificité du théorème** Le lecteur pourra se demander quelle est la difficulté dans l'énoncé du théorème 5.1.1, en regard surtout du théorème 4.3.1. La différence majeure entre ces deux résultats réside dans la taille des variations : le théorème 4.3.1 proposait des contraintes sur des petites variations (approximation linéaire), alors que le Théorème 5.1.1 porte sur le signe de grands déplacements d'états stationnaires. Il s'agit donc de montrer qu'on a le droit de conserver les relations de signes valables sur les petites variations lorsqu'on compare deux états stationnaires qui ne sont plus proches. On montre ceci en utilisant des méthodes usuelles de différentiation.

**Discussion biologique** Ce résultat donne un domaine de validité au raisonnement implicitement utilisé dans les publications de biologie. Les hypothèses sont raisonnables biologiquement puisque

- L'hypothèse (1) est vérifiée dès que la molécule ne rétro-agit pas sur elle-même comme un activateur direct de sa production : en dehors de ce cas, la dégradation de la molécule assure que l'influence de la molécule sur elle-même est négative. Il faut bien noter qu'il existe un certain nombre de molécules  $X_{(i)}$  (facteurs de transcription en particulier) qui sont leurs propres rétro-activateurs. Dans ces cas là, la contrainte n'est pas vérifiée pour  $X_{(i)}$ . On ne prendra tout simplement pas en compte cette contrainte. On peut voir cela comme un premier apport du théorème 5.1.1 : identifier les sommets d'un graphe d'influence pour lesquels les comportements sont *contraires à l'intuition* détaillée dans la figure 5.1 et au sujet desquels il faut prendre beaucoup de précautions.
- L'hypothèse (2) signifie que la molécule considérée est un sommet interne du système.
- L'hypothèse (4) signifie que les influences entre les molécules restent constantes pendant les expérimentations. Cette hypothèse est la plus forte, dans la mesure où des phénomènes biomoléculaires ont montré que le signe d'une interaction peut changer en fonction de la quantité d'éléments agissant sur leur cible [TS01].

L'intérêt principal du théorème 5.1 est qu'il est modulaire. Si une des hypothèses n'est pas vérifiée pour une molécule, il suffit de ne pas prendre en compte l'équation associée à ce sommet ; mais cela n'empêche pas de considérer les équations pour les autres sommets du réseau qui de leur côté vérifient la relation.

**Algèbre de signes** Pour comprendre ces équations, il faut réaliser que les éléments  $signe(\Delta X_{(i)})$  ne sont pas des réels mais doivent plutôt être lus dans l'algèbre de signes. Cette algèbre est définie par l'ensemble  $\{+, -, ?, \mathbf{0}\}$ . Le symbole  $+$  désigne les variations positives,  $-$  désigne les variations négatives et  $\mathbf{0}$  désigne les variations nulles. Le symbole  $?$  désigne les variations sur lesquelles on n'a pas d'information précise.

Sur cette algèbre de signes, on définit des opérations d'addition et de multiplication ainsi qu'une relation de compatibilité  $\simeq$ . On peut voir cette algèbre comme un codage des relations ensemblistes sur  $\mathbb{R}$ ,  $\mathbb{R}^+$ ,  $\mathbb{R}^-$  et  $\{0\}$ . Cette algèbre de signe a été introduite dans les années 80 pour étudier des questions de raisonnements qualitatifs à partir de règles causales; le point de vue était plutôt dynamique que statique [Kui84, dKB84, dKB82].

+	+	-	?	$\mathbf{0}$	$\times$	+	-	?	$\mathbf{0}$	$\approx$	+	-	?	$\mathbf{0}$
+	+	?	?	+	+	+	-	?	$\mathbf{0}$	+	V	F	V	V
-	?	-	?	-	-	-	+	?	$\mathbf{0}$	-	F	V	V	V
?	?	?	?	?	?	?	?	?	$\mathbf{0}$	?	V	V	V	V
$\mathbf{0}$	+	-	?	$\mathbf{0}$	$\mathbf{0}$	$\mathbf{0}$	$\mathbf{0}$	$\mathbf{0}$	$\mathbf{0}$	$\mathbf{0}$	V	V	V	V

**Codage de la réponse à un stress dans l'algèbre des signes** Finalement, la situation que nous considérons est la suivante :

- Un système est modélisé par son graphe d'influence, dont les arcs sont signés par un élément de l'ensemble  $\{+, -, \mathbf{0}\}$ . Le signe de l'arc du sommet  $A$  vers le sommet  $B$  est noté  $S_{A,B}$ .
- On dispose d'observations sur les variations pendant une expérimentation sur une partie des composants du système. Ces variations sont décrites par un élément de  $\{+, -, \mathbf{0}\}$ .
- Pour chaque élément  $A$  du système sur lequel  $A$  ne rétroagit pas positivement, on peut poser une contrainte sur la variation de la concentration de  $A$ , notée  $x_A$  pendant un déplacement d'états stationnaires consécutif à un stress environnemental :

$$x_A \approx \sum_{B \rightarrow A, B \neq A} S_{B,A} x_B \quad (5.1)$$

**Validité pour des modèles non différentiels ?** Dans (Biosystems, 2006), nous avons étudié la validité de ces contraintes en supposant que les systèmes étaient régis par des modèles différentiels. Cependant, cette hypothèse peut s'étendre. Ainsi, en collaboration avec A. Richard, nous avons travaillé sur la validité des contraintes dans le cadre des modèles logiques multivalués. Là encore, cette règle s'avère assez générale [Veb07].

#### 5.1.4 Extension de la contrainte à d'autres cadres

Nous avons étudié la validité de la règle causale pour les déplacements d'états stationnaires lors de stress environnementaux (Biosystems, 2006). Cependant, les

expérimentations opérées en biologie moléculaire ne sont pas toujours des stress environnementaux ; au contraire, avec les progrès techniques en biologie moléculaire, de nombreuses perturbations de systèmes consistent à forcer la sur-expression ou la sous-expression d'un gène, par création de mutants. Dans ce cadre, des contraintes qualitatives existent encore, mais elles doivent prendre en compte les modifications au sein de la cellule, comme cela est détaillé ci-dessous.

De plus, lors de nos expérimentations sur différents modèles, nous avons constaté que la règle causale est toujours valide, mais parfois trop générale pour contraindre efficacement le système. Nous avons donc proposé de nouvelles règles causales en fonction des types d'interaction qui apparaissent dans le système.

**Perturbations génétiques** Dans (BMC bioinfo, 2008), nous considérons des données de perturbations génétiques : après croissance comparable des cellules, on compare la quantité d'ARN dans une cellule mutée à la quantité des ARN dans une cellule non mutée. Dans ces cas là, avec une preuve semblable au cas des stress environnementaux, on montre que les contraintes qualitatives qui régissent le système sont les suivantes.

$$\left\{ \begin{array}{ll} x_A \approx \sum_{B \rightarrow A, B \neq A} S_{B,A} x_B & \text{si tous les prédécesseurs de } A \\ & \text{ne sont pas perturbés} \\ & \text{génétiquement} \\ x_A \approx -S_{M,A} + \sum_{\substack{B \rightarrow A \\ A \neq M, A \neq B}} S_{B,A} x_B & \text{si un prédécesseur de } A, \text{ noté } M \\ & \text{est inactivé} \\ x_A \approx S_{M,A} + \sum_{\substack{B \rightarrow A \\ A \neq M, A \neq B}} S_{B,A} x_B & \text{si un prédécesseur de } A, \text{ noté } M \\ & \text{est sur-exprimé} \end{array} \right. \quad (5.2)$$

**Formation de complexes** Dans le cadre de l'association de deux protéines pour former un complexe, la règle booléenne causale est vérifiée, mais elle est en fait trop peu précise. Dans le cadre du travail de thèse de C. Guziolowski, en étudiant plus précisément les phénomènes entrant en jeu et leurs échelles de temps, nous montrons dans (CIBB, 2008) qu'une règle booléenne qui décrit mieux les déplacements d'états stationnaires lors de la formation des complexes est *Le plus faible gagne*. Autrement dit, la variation de la concentration du complexe est guidée par la variation du composant en plus faible quantité.

**Schémas plus généraux : phosphorylation** On peut plus globalement modéliser la plupart des phénomènes biochimiques en proposant des règles booléennes pour décrire les déplacements d'états stationnaires dans les systèmes biologiques. Dans le cadre de la modélisation des régulations induites par la molécule chimère *EWf/FLI1* (collaboration avec l'institut Curie, projet ANR SITCON), Carito Guziolowski et Sylvain Blachon travaillent actuellement sur l'élaboration de règles pour la phosphorylation de la protéine *P53*. Là encore, il faut bien noter que ces règles ne décrivent absolument pas la dynamique des interactions mais des conditions globales sur les variations après un stress.

## 5.2 Une approche pour le diagnostic d'un réseau

Au début de la section 5.1.2, nous avons détaillé comment les questions principales que se posent un biologiste se ramènent, d'un point de vue informatique, à rechercher des méthodes efficaces pour poser un diagnostic sur un réseau et, lorsque le diagnostic est bon, procéder à des prédictions. Nous allons voir maintenant comment nous réalisons ces tâches concrètement : l'idée principale est de partir des contraintes définies par le théorème 5.1.1 comme formalisation de l'intuition du biologiste, et de traduire les notions de diagnostic et de prédiction comme des objets liés à ces contraintes.

### 5.2.1 Construire le système de contraintes décrivant un réseau et les observations disponibles

Nous considérons donc maintenant un ensemble d'observations qui comparent l'état initial de molécules pendant des expérimentations à leur état final. Nous pouvons construire un système de contraintes qui décrivent les états du système.

**Variables** Les variables correspondent soit au signe des interactions, soit à la variation des concentrations des molécules du système. Elles sont à valeur dans l'algèbre des signes  $\{+, -, \mathbf{0}\}$ .

- On considère une variable  $s_{i,j}$  pour décrire le signe de l'interaction allant de la molécule  $i$  vers la molécule  $j$ , lorsque le modélisateur veut intégrer cette interaction dans le modèle (appuyé par une publication scientifique ou des données d'immuno-précipitation (chip-chip) par exemple).
- S'il y a  $k$  expérimentations concernant des molécules  $M_1, \dots, M_N$ , on considère une variable  $x_j^{(l)}$  pour décrire la variation de la concentration de la molécule  $M_j$  pendant l'expérimentation numérotée  $l$  pour  $j \in \{1, \dots, N\}$  et  $l \in \{1, \dots, k\}$ . La comparaison se fait par rapport à un état référence commun à toutes les expérimentations.

**Affectations de valeurs** Les observations et connaissances permettent de fixer la valeur de certaines variables du système.

Si la quantité d'une molécule  $M_j$  varie positivement, négativement ou ne varie pas pendant l'expérimentation  $l$ , on fixe la valeur de  $x_j^{(l)}$  correspondante.

Si la littérature ou des expérimentations permettent de déterminer le sens de l'interaction de  $i$  vers  $j$ , on fixe le signe de la variable  $s_{i,j}$ . Sinon, la variable  $s_{i,j}$  est une inconnue du système.

**Contraintes entre les variables** Jusqu'à maintenant, nous n'avons fait que transcrire la connaissance et les observations, mais nous n'avons pas relié les variables entre elles. Pour cela, nous utilisons les règles décrites dans la section précédente. En fonction du type d'interaction sur la molécule  $M_j$  et du type d'expérimentation, on introduit une contrainte sur chaque variable  $x_j^{(l)}$ , qui fait intervenir

les variables  $x_i^{(l)}$  pour lesquels  $s_{i,j}$  est non nul ou inconnu.

Dans le cas de l'opéron-lactose, on obtient le système de contraintes suivant.

$$\left\{ \begin{array}{l} LacI \approx -A \quad (1) \\ A \approx LacZ \quad (2) \\ LacZ \approx cAMP - LacI \quad (3) \\ Li \approx Le + LacY - LacZ \quad (4) \\ G \approx Li + LacZ \quad (5) \\ cAMP \approx -G \quad (6) \\ LacY \approx cAMP - LacI \quad (7) \end{array} \right.$$

### 5.2.2 Exploiter le système de contraintes : invariants et modules spécifiques

Les deux objectifs fondamentaux des biologistes discutés dans la section 5.1.2 s'interprètent alors facilement dans le cadre des systèmes d'équations (Complex Us, 2005) :

- la notion de *cohérence* revient à décider s'il existe au moins une solution au système ;
- la notion de prédiction revient à rechercher les invariants dans l'ensemble de toutes les solutions.

**Solution** Une *solution* dans  $\{+, -\}$  d'un système de contraintes associé à un modèle biologique est une affectation de toutes les variables du système avec les valeurs  $+$  et  $-$  qui satisfait simultanément toutes les équations du système.

Une solution à valeur dans  $\{+, -, \mathbf{0}\}$  introduit la valeur  $\mathbf{0}$  pour les variables et permet de distinguer des sous-ensembles de variables pertinents.

Ces solutions correspondent aux différents observations qu'il est *possible* d'obtenir avec un stress.

Pour le cas de l'opéron-lactose (voir Fig. 4.6), il est possible d'énumérer l'ensemble des solutions du système dans  $\{+, -\}$ , décrites dans la Table 5.1. Il y a une symétrie dans l'ensemble des solutions puisque le système est linéaire. On voit ici que les contraintes associées au système restreignent fortement l'espace des possibilités, puisqu'on passe de 256 jeux d'observations possibles à priori à 18 jeux réellement admissibles.

**Cohérence** On dit que le système de contraintes est *cohérent* lorsqu'il admet au moins une solution.

Ainsi, le système de l'opéron lactose est cohérent puisqu'il admet au moins une solution. Par contre, le système présenté dans la figure 5.1 est incohérent lorsqu'on introduit les observations, puisqu'il s'écrit :

$$\left\{ \begin{array}{l} C \approx A + B \\ C = + \\ A = - \\ B = - \end{array} \right.$$

<i>Le</i>	<i>LacI</i>	<i>A</i>	<i>LacZ</i>	<i>Li</i>	<i>G</i>	<i>cAMP</i>	<i>LacY</i>	<i>Le</i>	<i>LacI</i>	<i>A</i>	<i>LacZ</i>	<i>Li</i>	<i>G</i>	<i>cAMP</i>	<i>LacY</i>
-	-	+	+	+	+	-	+	+	-	+	+	-	-	+	+
-	-	+	+	-	+	-	+	+	-	+	+	+	+	-	+
-	-	+	+	-	+	-	-	+	-	+	+	+	+	-	-
-	-	+	+	-	-	+	+	+	-	+	+	+	+	-	-
-	+	-	-	+	-	+	-	+	+	+	+	-	+	-	+
-	+	-	-	+	-	+	+	+	+	-	+	+	-	+	+
-	+	-	-	-	-	+	-	+	+	-	-	+	-	+	+
-	+	-	-	-	-	+	+	+	+	-	-	-	-	+	-
-	+	-	-	+	+	-	-	+	+	-	-	+	+	-	-

TAB. 5.1 – Liste des 18 jeux expérimentaux (parmi les  $2^8=256$  possibles) compatibles avec le réseau modélisant l’opéron lactose.

De même, les observations sur la mise à jeûn des poulets décrites à Fig. 4.5 sont cohérentes avec le réseau. On retrouve ainsi de manière automatique que la variation nulle des désaturases FADS1 et FADS2 n’est absolument pas une contradiction du système, comme nous l’avons déjà discuté à la section 4.3.4.

**Prédictions** On dit que la variable  $s_{i,j}$  ou  $x_j^{(l)}$  est une *prédiction* du système lorsque la valeur de cette variable est *invariante* dans toutes les solutions du système de contraintes.

Par exemple, si nous revenons sur le système de l’opéron-lactose, et si on observe que *Le* est sous-exprimé et *LacZ* est sur-exprimé, on peut déduire de la liste des solutions que, nécessairement, *LacI* sera sous-exprimé et *A* sera sur-exprimé, puisque toutes les solutions telles que  $Le = -$  et  $LacZ = +$  sont aussi telles que  $LacI = -$  et  $A = +$ .

**Diagnostic (ou module) incompatible** Un *diagnostic* est un sous-ensemble de contraintes du système, minimal pour l’inclusion, qui n’admet pas de solution. Nous reviendrons plus précisément sur cette notion dans une section ultérieure.

### 5.2.3 Cohérence dans la littérature : vers une contrainte plus restreinte mais tout aussi intuitive ?

**Contraintes locales vs solution globale** La notion de cohérence apparaît pour l’étude de réseaux génétiques [GRRL<sup>+</sup>03] ou à la fois métaboliques et génétiques [HCP03, CKR<sup>+</sup>04]. Les auteurs calculent une mesure de cohérence pour différents types de réseaux en exploitant des données de transcriptome. Des notions d’influences indirectes apparaissent aussi dans [JDSZ05, BMO<sup>+</sup>01] pour contraindre des variations .

La première différence entre ces méthodes et le schéma de vérification que nous proposons réside dans des considérations locales plutôt que globales : nous appliquons une règle de cohérence à l’ensemble des nœuds du système et nous considérons une solution *globale* pour ce système. Au contraire, dans les approches de [GRRL<sup>+</sup>03, HCP03, CKR<sup>+</sup>04], la cohérence est étudiée localement (elle ne prend

pas en compte les contraintes successives apportées par les sommets éloignés d'un nœud). L'apport du point de vue global est d'abord de contraindre plus fortement le système, et surtout de localiser les modules à corriger (quelle que soit leur taille) pour orienter le modélisateur dans sa construction de réseau.

**Changement d'échelle** La deuxième différence entre les approches exposées dans [GRRL<sup>+</sup>03, HCP03, CKR<sup>+</sup>04] et notre schéma de vérification est le type de données qui peuvent être analysées : nous pouvons écrire un système de contraintes dès qu'un jeu de données est disponible, quelle que soit sa taille. Guteriez et al [GRRL<sup>+</sup>03] comme l'école allemande [HCP03, CKR<sup>+</sup>04] ont besoin d'une grande variété de jeux expérimentaux pour explorer tout l'espace des dynamiques et poser un diagnostic de compatibilité ou d'incompatibilité.

Enfin, la dernière différence porte sur la taille des jeux de données considérés : comme nous allons le voir plus loin, Michel Le Borgne, Philippe Veber et Carito Guziolowski ont développé des méthodes très efficace pour traiter des systèmes de contraintes qui contiennent plusieurs milliers de variables, un ordre de grandeur innatignable avec les approches de [GRRL<sup>+</sup>03, HCP03, CKR<sup>+</sup>04].

**Une autre approche de la cohérence** On retrouve aussi la notion d'influence de chemins pour étudier un système dans les travaux de l'équipe de T. Ideker pour proposer des plans expérimentaux en tant qu'optimum global d'influences diverses [YIJ04, YMM<sup>+</sup>05]. Derrière ces approches se trouve une nouvelle notion de cohérence, que nous allons appeler *cohérence par chemins* : une affectation des variables est cohérente par chemins par rapport à un jeu d'observations si pour tous les sommets du système, il existe un chemin partant d'une entrée du système dont le signe est compatible avec l'affectation du sommet.

P. Veber a montré que ces notions ne sont pas équivalentes : dans la Figure 5.2, le graphe n'est pas cohérent au sens des équations qualitatives, puisque la variation de  $E$  doit être égale à celle de  $D$ , qui est elle-même égale à celle de  $F$ . Par contre, le graphe est cohérent au sens des chemins puisque la variation de  $E$  peut-être expliquée par le chemin partant de  $A$  et passant par  $C$ , tandis que celle de  $F$  peut-être expliquée par le chemine partant de  $A$  et passant par  $B$ .

Inversement, le réseau de droite est cohérent pour les équations qualitatives mais n'est pas cohérent par chemin si  $C$  est la seule entrée puisqu'aucun chemin partant de  $C$  ne peut expliquer la variation de  $B$  et  $A$  ; en fait, les variations de  $A$  et  $B$  s'autojustifient.

Ces deux approches sont fondamentalement différentes :

- la cohérence par chemins force toute variation à être expliquée par une cause extérieure, mais ne vérifie pas localement que les variations sont cohérentes, en particulier les variations introduites sur les sommets non observés.
- Inversement, les équations qualitatives vérifient que les variations sont compatibles localement sur tous les sommets, mais elles n'imposent pas que les variations aient une cause. Dans la Figure 5.2, les équations qualitatives admettent très bien que  $A$  et  $B$  augmentent alors que  $C$  diminue puisque  $A$  et  $B$  se justifient l'un l'autre. Mais quel produit justifie leur accroissement ?



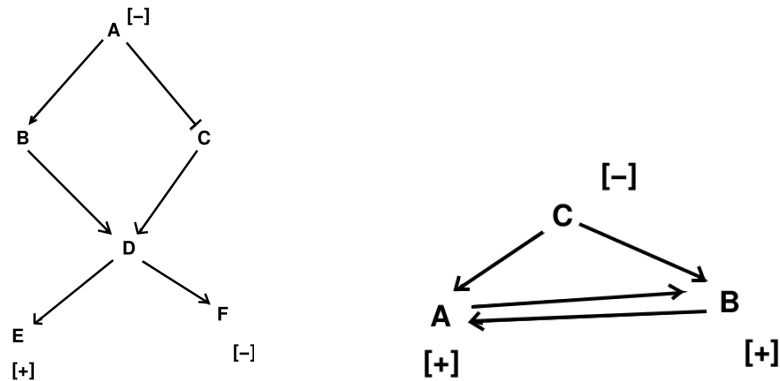


FIG. 5.2 – Deux réseaux qui sont cohérents respectivement seulement par chemin et pour les équations qualitatives.

**Une nouvelle contrainte plus stringente et unificatrice** On peut ainsi introduire une nouvelle notion de cohérence plus stringente, que nous appellerons *cohérence biologique* : un modèle est cohérent avec des données s'il existe une valuation pour toutes les variables du système telle que toute variation non nulle sur un sommet doit être prouvée comme une suite finie et non circulaire de déductions.

Cependant, cette notion est plus difficile à vérifier que la cohérence par équations qualitatives définies plus haut (en particulier avec les arbres de décision dont nous allons discuter tout de suite). Nous proposons donc de l'utiliser seulement lors de la recherche de cause pour l'état d'un système, comme nous en rediscuterons à la section 5.4.4.

#### 5.2.4 Calcul des invariants et des diagnostics

Les concepts étant posés, la question est de résoudre effectivement et efficacement la question de la cohérence, des prédictions et de la localisation des zones incohérentes. Une première approche aurait pu être d'utiliser les algorithmes usuels de résolutions d'équations qualitatives [TMD03, Dor88]. Ces algorithmes sont basés globalement sur une adaptation de l'algorithme de Gauss au cas qualitatif. La difficulté réside dans le fait que la relation de compatibilité  $\simeq$  n'est pas transitive et induit des problèmes algorithmiques complexes pour le choix de l'ordre de réduction des équations.

Or, les systèmes que nous considérons peuvent contenir plusieurs milliers de variables. De plus, répondre au problème de la satisfiabilité d'un système de contraintes associées à un graphe d'influence est NP-complet (voir (Complex Us, 2005)); autrement dit, le fait de considérer seulement les systèmes associés à une question biologique ne permet pas de diminuer la complexité des systèmes étudiés [Dor88]. M. Le Borgne, P. Veber et C. Guziolowski ont proposé de faire appel à différentes méthodes de codage des solutions pour explorer de la manière la plus efficace possible

l'espace des solutions des contraintes.

**Énumérer les solutions** Une première stratégie pour identifier des diagnostics et répondre à la question des prédictions consiste à explorer l'ensemble des solutions. Pour cela, il faut résoudre deux questions distinctes.

1. *Réduire le système de contraintes* à un système de taille plus petite mais équivalent en terme de cohérence. Plus précisément, un certain nombre d'équations ne contraignent en fait pas le système. Il s'agit typiquement des successions de transcriptions de gènes non observés. On obtient alors des équations de la forme  $x_{i+1} = x_i$  qui n'apportent rien à la cohérence du système.

Ainsi, nous montrons dans (Complex Us, 2005) que la question de la cohérence globale du système est équivalente à la cohérence du système construit pour le graphe d'influence qui correspond à l'image inverse des états observés et des cycles du système (voir aussi la discussion dans la thèse de P. Veber [Veb07]). Sur les réseaux transcriptionnels, cette opération de réduction diminue considérablement la taille du système de contraintes, et permet d'envisager une étude de l'ensemble des solutions.

2. *Représenter efficacement l'espace des solutions.* Le système étant réduit à un système de taille plus raisonnable équivalent pour la question de la cohérence, il devient possible de représenter l'espace des solutions sous la forme d'un arbre de décision binaire (BDD pour *Binary Decision Diagram*). Il s'agit d'une structure de données dans laquelle toute formule logique (booléenne) peut être représentée par un graphe orienté acyclique avec une racine consistant en nœuds de décisions, et deux nœuds terminaux pour les valeurs 0 et 1. Un chemin de la racine au nœud 1-terminal représente un modèle admissible pour la formule considérée. Ces BDD sont très utilisés par les programmes de conception assistée par ordinateur, pour générer des circuits (synthèse logique) et en vérification de modèles [Bry92].

Concrètement, Michel Le Borgne, Philippe Veber et Carito Guziolowski ont ainsi proposé une implémentation sous la forme de BDD et des algorithmes efficaces pour les parcourir, qui permettent de rechercher des diagnostics et procéder à des prédictions pour les graphes d'influence. Une suite logicielle (Pyquali et Bioquali) a été développée à ce propos [LBV07], et une interface web a été développée (<http://www.irisa.fr/symbiose/bioquali>) ainsi qu'un plugin pour la suite logicielle Cytoscape dédiée à la représentation de réseaux biologiques [SMO<sup>+</sup>03].

**Rechercher au moins une solution** Une autre stratégie consiste non pas à étudier *l'ensemble des solutions* mais à rechercher s'il existe *au moins une solution* à différents problèmes. La question de la cohérence d'un système entre naturellement dans ce cadre. La recherche des prédictions est aussi abordable en considérant la recherche des éléments qui ne sont pas invariants (pour chaque sommet, existe-t-il deux solutions pour le système qui prennent des valeurs opposées?). On entre ici dans le cadre de la résolution de contraintes.

Pour aborder cette question, P. Veber a proposé de se concentrer sur la programmation par ensembles réponses (ASP), en collaboration avec l'équipe de T. Schaub à Postdam avec laquelle nous entretenons des relations scientifiques et qui développe un solveur efficace Clasp [GKNS07]. Indiquons rapidement que la programmation par ensembles réponses est une forme de programmation déclarative dédiée à l'exploration de problèmes réputés difficiles (NP-complets) [Nie99, Bar03]. Dans ce domaine, en pleine expansion, l'objectif est d'allier l'expressivité de la programmation logique (type Prolog) avec la puissance des algorithmes SAT. Par construction, les algorithmes implémentés terminent toujours en principe ; il s'agit de la principale différence avec les solveurs de type Prolog, qui peuvent amener à des boucles infinies.

À l'aide de Clasp, il été possible de montrer que la question de la cohérence, celle de la recherche de diagnostic, et la recherche des prédictions sont abordables sans passer par une étape de réduction préalable [GST<sup>+</sup>08].

**Complémentarité** Nous avons ainsi à notre disposition deux approches bien différentes mais tout aussi efficaces pour résoudre les questions biologiques qui ont été identifiées en début de chapitre. On pourra se demander pourquoi les conserver toutes les deux : en fait, ces approches sont complémentaires, et doivent être étudiées en parallèle pour pouvoir répondre à des questions plus précises que celle du diagnostic et des prédictions.

Ces questions seront abordées plus en détail à la fin de ce chapitre, mais nous pouvons déjà insister sur les apports des deux méthodes :

- La spécificité du parcours de l'ensemble des solutions via les BDD est qu'il permet d'*énumérer* les solutions à des questions variées, même s'il ne s'agit pas d'invariants. Ceci nous permettra en particulier d'assouplir les contraintes sur les prédictions en recherchant non pas des variables qui sont invariantes pour l'ensemble des solutions du système, mais des variables dont la valeur est identique pour un pourcentage assez élevé de solutions. Ceci est impraticable par une approche directe la programmation par ensembles réponses.
- La spécificité de la programmation par ensembles réponses est de pouvoir interroger des systèmes de contraintes de très grande taille sans les réduire préalablement, et d'agrandir le spectre de questions posées. En particulier, la notion de détermination de plans expérimentaux et la recherche de causes communes à un évènement semblent abordables en ASP mais pas avec des algorithmes de parcours de BDD. Nous en discuterons plus précisément dans la dernière section du chapitre.

### 5.2.5 Schéma d'utilisation et validation

Les concepts informatiques étant posés, nous pouvons proposer un schéma d'utilisation pour l'étude d'un réseau.

**Schéma général** Le processus que nous mettons en place pour analyser un réseau est illustré dans la figure 5.3 :

- Construire un graphe d’influence à partir des connaissances disponibles, incluant avant tout les régulations transcriptionnelles avec une règle booléenne.
- Appliquer un test de cohérence.
- Si le test est positif, rechercher les prédictions du système, les valider via une étude bibliographique ou les valider expérimentalement.
- Si le test de cohérence est négatif, localiser l’emplacement des incompatibilités. Corriger le modèle dans les zones incompatibles, éventuellement en introduisant des régulations post-transcriptionnelles modélisées avec des règles spécifiques. Retourner alors à l’étape 2.

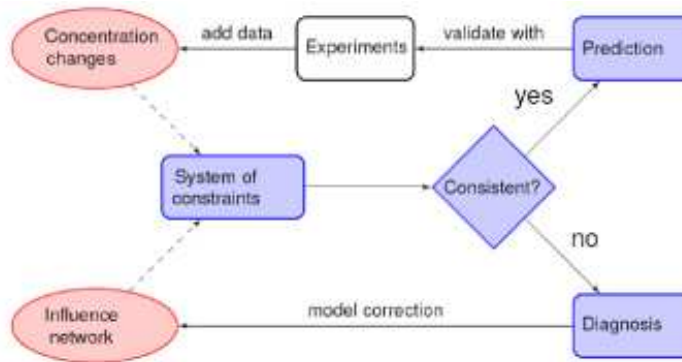


FIG. 5.3 – Illustration du processus de vérification de cohérence pour un système : on alterne des phases de construction (“manuelles”) de modèles avec des phases automatiques de localisation de zones sur lesquelles une incompatibilité existe ou une prédiction est possible.

**Réseau de régulation transcriptionnel de *E. coli*** De manière à valider nos méthodes sur un réseau réel, nous avons considéré le réseau transcriptionnel de la bactérie modèle *E. coli*. Pour cette seule bactérie, un travail a été entrepris, qui rassemble toutes les interactions entre protéines au sein de cette cellule. Cette information est publiquement accessible via la base de donnée RegulonDB [SGCPG<sup>+</sup>06]. À partir de cette information, C. Guziolowski, pendant son master puis sa thèse, a obtenu un réseau contenant 1763 produits et 4491 interactions.

Après réduction à l’image inverse des cycles du système (voir Sec. 5.2.4), on obtient un graphe de 68 produits et 198 influences (voir Fig. 5.4). Par construction, ce graphe réduit, aussi appelé *cœur* du réseau, est l’ensemble des gènes qui régulent le reste du réseau, sans a priori sur les observations.

À ce réseau nous avons adjoint des données correspondant à un stress nutritionnel. Là encore, la base RegulonDB classe les variations des ARN pour différents stress environnementaux, avec une référence précise pour chaque variation. Nous avons ainsi collecté 45 variations concernant des produits du réseau.

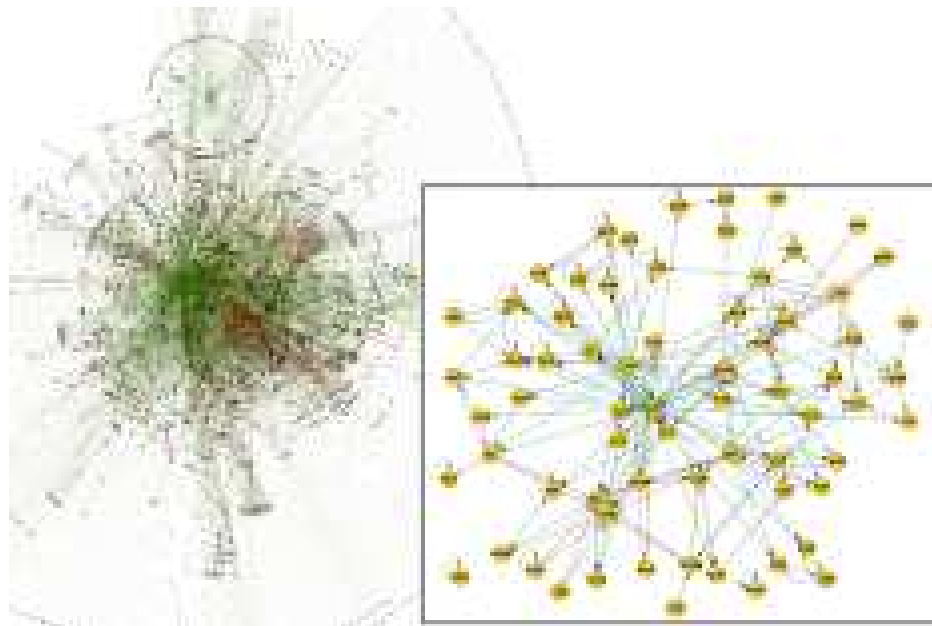


FIG. 5.4 – Le graphe d'influence pour les interactions transcriptionnelles de la bactérie *E. coli* et son graphe réduit

**Application du schéma** Nous avons appliqué le schéma proposé ci-dessus au réseau transcriptionnel de *E. coli* et le jeu de 45 observations. La boucle diagnostic-correction a dû être utilisée 3 fois avant d'obtenir un réseau cohérent.

- Dans le réseau, certaines protéines ne font pas directement partie de la machinerie de transcription, il s'agit des facteurs-sigma. Nous avons montré que sans ces facteurs sigma, le réseau transcriptionnel n'est pas cohérent avec les variations connues au sujet du stress nutritionnel. La phase de diagnostic a mis en évidence que seules trois équations posaient un problème. C. Guziolowski a alors étudié la bibliographie et introduit les facteurs-sigma dans le réseau, ainsi qu'un complexe IHF associé à une règle booléenne spécifique (voir Fig. 5.5). Ces résultats sont explicités dans (JPBC, 2006) et discutés dans (CIBB, 2008).
- Après cette première correction, une nouvelle incohérence est apparue, qui a été localisée autour de la molécule *appY*. Là encore, une étude poussée de la littérature a montré que cette molécule reçoit un signal post-transcriptionnel pendant le stress, qui n'avait pas été inclus initialement dans le réseau. Les détails sont donnés dans la publication (CIBB, 2008).

Nous avons ensuite produit un ensemble de prédictions pour ce graphe compatible. À partir des 45 observations initiales, nous avons prédit la variation de 526 autres éléments du réseau. Pour valider ces prédictions, nous avons considéré des

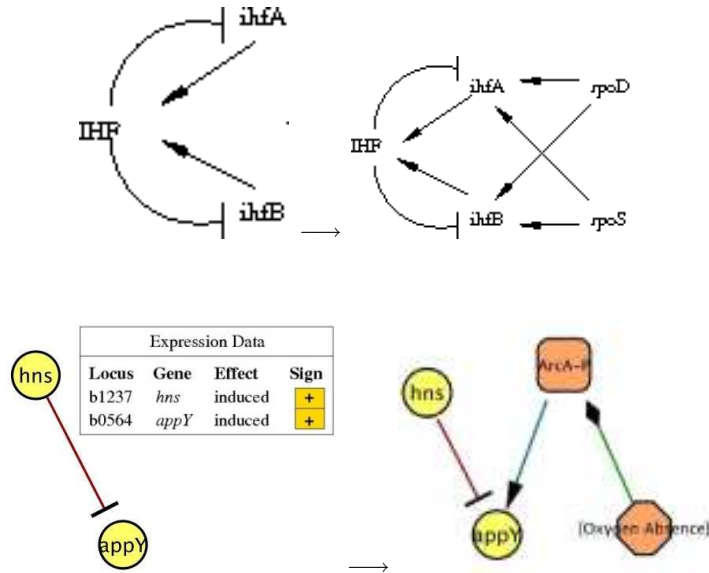


FIG. 5.5 – Les deux sous-modules du réseau transcriptionnel de *E. coli* auxquels il a fallu adjoindre des interactions spécifiques pour obtenir une compatibilité.

données de transcriptome [FHT<sup>+</sup>07]. Le consensus entre les observations de transcriptome et nos prédictions est de 80% (CIBB, 2008). A posteriori, ce taux élevé de consensus est surprenant dans la mesure où le modèle que nous considérons n'intègre que peu de phénomènes post-transcriptionnels ; ainsi, les modules de réception des signaux extérieurs sont peu pris en compte. Il faut aussi avoir à l'esprit que le réseau est a priori incomplet puisqu'il est construit seulement à partir des connaissances [GW02]. Malgré tout, cela est suffisant pour contraindre une grande partie du système.

Notons enfin que le niveau de consensus est équivalent à celui obtenu avec les autres approches du domaine [CKR<sup>+</sup>04, CP02, EP00], si ce n'est que nous travaillons avec des réseaux de bien plus grande taille et avec beaucoup moins de jeux de données.

### Réseau de grande taille : restriction des données à leur partie cohérente

Nous avons ainsi illustré que le schéma de diagnostic et prédictions est efficace sur des réseaux de grande taille, à partir d'un jeu de données de petite taille. Par contre, à partir d'un jeu de données de grande taille (typiquement des données de transcriptome), si le réseau n'est pas assez détaillé, le nombre d'incohérences augmente. Pour illustrer ce point, nous avons testé systématiquement la cohérence entre le réseau transcriptionnel de *E. coli* et différents jeux de données concernant des stress publiés dans [FHT<sup>+</sup>07], plus précisément 226 jeux de données correspondant à 61 conditions expérimentales différentes. Seuls 31 conditions expérimentales se sont

avérés cohérentes avec le réseau transcriptionnel (BMC bioinfo, 2008).

Dans ces cas là, corriger le modèle devient un travail de très longue haleine. Pour avoir une première estimation du travail à faire, nous avons considéré chaque jeu de données indépendamment et procédé itérativement à des diagnostics : pour chaque nœud du réseau sur lequel une incompatibilité est détectée, les contraintes impliquant la variable sur ce nœud dans le réseau ont été enlevées du réseau. Cela revient à ne pas prendre en compte l'ensemble des informations (régulation et observation) qui portent sur le nœud problématique. À la fin du processus, on a isolé un sous-ensemble du réseau qui porte les incohérences, et on se concentre sur la partie (restante) qui est compatible avec les données observées. Cette opération de *nettoyage* des données sera appelée *restriction cohérente des données*.

Nous avons alors observé que, après cette restriction des jeux de données, les observations du système portaient en moyenne sur 12,62% du réseau (BMC bioinfo, 2008), ce qui restreignait très fortement les jeux de données initiaux.

### Étude préalable pour un jeu de grande taille : mesure de la robustesse

Pour estimer la nécessité de procéder à la correction du modèle, nous proposons une phase préalable, qui consiste à utiliser les jeux de données restreints pour vérifier que le réseau est *globalement cohérent*. Nous proposons ainsi un protocole pour tester la compatibilité d'un graphe d'influence et d'un jeu de données de transcriptome (voir (CIBB, 2008)).

- On considère un réseau formalisé sous la forme d'un graphe d'influence. D'un jeu de données transcriptome, on extrait la liste des ARN qui varient significativement de manière positive et de manière négative.
- On fixe un taux de données  $x$  (de 5 à 20%) et on procède à une sélection aléatoire de  $x\%$  des variables ayant une variation significative. Ces variables sélectionnées sont considérées comme des observations pour le système.
- On restreint le jeu d'observations ainsi obtenu à sa partie cohérente, selon la procédure décrite ci-dessus.
- On compare les prédictions des observations restantes avec les données transcriptome qui n'avaient pas été sélectionné comme observation, pour calculer un taux de consensus entre les prédictions et les observations.
- On itère ce processus avec d'autres tirages aléatoires de  $x\%$  parmi les nœuds qui varient significativement, on en déduit un *taux de consensus moyen* pour le ratio  $x$ .
- Tout le processus est itéré avec de nouvelles valeurs de  $x$ .

Cette analyse permet d'estimer si les observations sont globalement cohérentes avec le réseau analysé. Comme nous le montrons sur le réseau transcriptionnel de *E. coli* dans (CIBB, 2008), le taux de consensus ne varie pas significativement pour les différentes valeurs de  $x$ . Il est de 90% pour les données concernant un stress nutritionnel. Notre conclusion est que, malgré les erreurs et son aspect incomplet, le réseau transcriptionnel de *E. coli* donné par RegulonDB est une bonne description pour un stress nutritionnel (90% de consensus en moyenne). Par contre, les stress anaérobie et de température sont moins bien décrits, puisque le taux de consensus est de 80% en moyenne. Les détails se trouvent dans (CIBB, 2008).

### 5.2.6 Etude (en cours) du rôle de la molécule EWS/FLI1 dans le sarcome d'Ewing

Depuis 2007, dans le cadre du projet SITCON soutenu par le programme ANR BIOSYS, nous avons collaboré avec des équipes de l'institut Curie dans le cadre d'une modélisation des régulations de la prolifération des tumeurs d'Ewing.

**Contexte expérimental** Le sarcome d'Ewing est par ordre d'importance la deuxième tumeur osseuse maligne chez les enfants et les jeunes adultes. Dans 80% des cas, elle est associée à une translocation des chromosomes  $t(11;22)(q24;q12)$  qui engendre une fusion du gène EWS avec un membre de la famille ETS, appelé FLI. Le gène chimère qui en résulte, EWS/FLI est considéré comme un activateur transcriptionnel de la tumeur d'Ewing, par altération de l'expression de ses cibles dans un environnement cellulaire permissif.

Bien que la tumeur d'Ewing soit l'une des tumeurs les plus étudiées, ses cellules d'origine sont encore inconnues. Un des objectifs du projet SITCON est ainsi d'identifier les gènes cibles de EWS/FLI ainsi que les voies de régulation. Dans ce but, G. Stoll (Institut Curie) a produit un réseau de régulation décrivant les effets connus de la protéine chimère EWS/FLI1 (145 nœuds, 269 arcs). Ce réseau est particulièrement intéressant dans la mesure où, contrairement aux graphes d'influence chez *E. coli* et la levure, les régulations sont extrêmement interconnectées et une abondante combinatoire est présente : 76% des nœuds ont au moins deux successeurs. De plus, la réduction du réseau à l'image inverse des cycles du système contient 83 nœuds et 155 : il s'agit d'une différence majeure avec le réseau transcriptionnel de *E. coli*, pour lequel la réduction permettait de passer d'un réseau contenant plusieurs milliers de nœuds à un réseau de 68 nœuds seulement, soit une réduction de la taille de 96%, alors que le procédé de réduction permet simplement de réduire le système de 42 %.

Par ailleurs, l'équipe d'Olivier Delattre à l'institut Curie a obtenu des séries temporelles pendant une expérience qui a consisté à inhiber artificiellement la protéine EWS/FLI1 puis à l'activer à nouveau. On dispose ainsi de données permettant de déterminer les gènes de la cellule qui sont corrélés ou anti-corrélés à EWS/FLI1.

**Analyse de cohérence** Les données de corrélation et d'anti-corrélation suivent exactement les mêmes règles causales que les stress environnementaux. T. Baumratova a donc procédé à un test de cohérence entre le graphe d'influence et les données de corrélation. Il s'est avéré que le réseau était cohérent partout sauf en un sommet.

Après analyse de la situation biologique, l'incohérence est apparue comme résultant de la non-distinction des effets transcriptionnels et post-transcriptionnels de deux régulations sur une protéine. Plus précisément, dans la version du graphe d'influence considéré, les ARN et les protéines actives ont été assimilées dans un même sommet du graphe. Or, il est apparu que plus de 50% des interactions portait sur des régulations post-transcriptionnelles. Pour cette raison, nous avons construit un nouveau graphe d'influence à partir du modèle de régulation, ce deuxième graphe



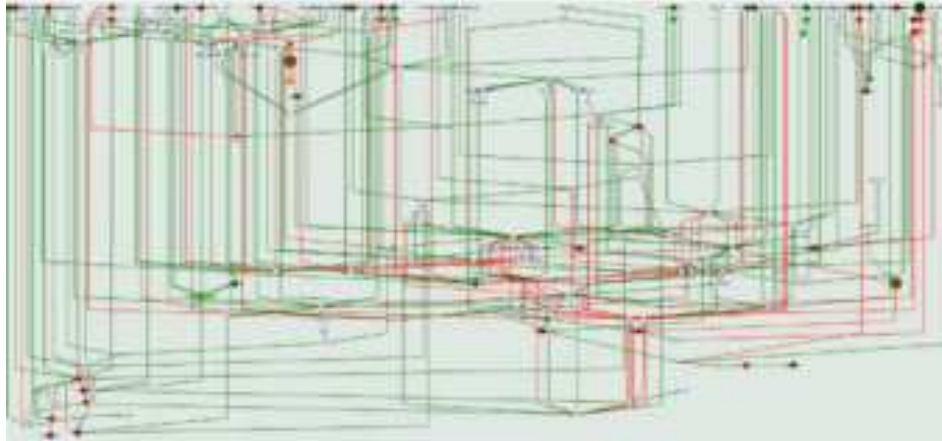


FIG. 5.6 – Le réseau de régulation de la protéine chimérique EWS/FLI1, dans lequel les ARN sont distingués de leur protéine (produit par G. Stoll et T. Baumuratova, Institut Curie). Ce réseau comporte 329 nœuds et 583 arcs. La combinatoire y est importante et la réduction à l'image inverse des cycles du système réduit le système de seulement 42%.

distingue les ARN des protéines actives, et est présenté Fig. 5.6.

Avec ce deuxième graphe, les incohérences ont été résolues.

**Prédictions** Nous avons ensuite procédé à des calculs de prédictions sur le réseau. Nous avons ainsi obtenu 20 prédictions sur des sommets non-observés. En particulier, les prédictions ont porté principalement sur des protéines actives. À partir d'une nouvelle analyse des données, nous avons orienté nos recherches sur des protéines actives dont la variation est prédite mais pour lesquelles les ARN ne varient pas significativement. Selon notre interprétation, cela signifie qu'un phénomène post-transcriptionnel renverse l'effet naturel de la transcription sur cette protéine. Des expérimentations sont programmées à l'automne 2008 pour rechercher la pertinence des prédictions.

**Perspective : mieux contraindre les phénomènes post-transcriptionnels pour capter plus de comportements**

L'analyse du réseau de régulation de EWS/FLI1 a mis en évidence que lorsque de nombreuses interactions post-transcriptionnelles sont actives, la règle de causalité est valable mais peu contraignante pour le système. Dans ce cadre, C. Guziolowski et S. Blachon ont proposé de construire des règles spécifiques pour les différents modules d'interaction dans le réseau. Comme cela a été déjà expliqué dans la section 5.1.4, les déplacements d'états stationnaires lors de la formation de complexe peuvent être modélisés par une règle booléenne du type *le plus faible gagne*. Pour d'autres processus post-transcriptionnels tels que les signaux, il s'agit, dans le même esprit, de proposer des

---

contraintes booléennes qui décrivent mieux les déplacements d'états stationnaires et de contraindre plus le système.

Ces travaux sont en cours, et nous espérons obtenir à la fin un meilleur taux de prédiction, et surtout, une meilleure description pour l'élaboration de plans expérimentaux que nous aborderons en fin de chapitre.

## 5.3 Une nouvelle question : identifier le signe des régulations

Nous avons ainsi proposé une approche pour isoler les modules incohérents dans un réseau. L'idée principale est de construire un système de contraintes qui clarifient et formalisent le raisonnement intuitif des biologistes, puis d'interpréter les objectifs principaux du biologiste (cohérence globale, aide à la correction, prédiction) comme des propriétés de l'ensemble des solutions du système.

L'idée de cette démarche était au départ d'aider le modélisateur à corriger son modèle, mais le cadre d'application de cette approche s'avère finalement bien plus général. En fait, dès qu'un expérimentateur procède à des raisonnements intuitifs sur des données qui comparent différents états d'un système, il est possible de formaliser les étapes du raisonnement sous la forme de contraintes : cela permet de mieux comprendre le domaine de validité du raisonnement, et de limiter les erreurs de raisonnements qui peuvent apparaître en particulier à cause de la présence de boucles positives dans un réseau.

La première difficulté qui se pose est d'interpréter la question initiale comme une propriété de l'ensemble des solutions du système. La seconde difficulté est ensuite de proposer un algorithme de vérification de la propriété, qui est généralement de nature difficile. Dans cette section, nous allons illustrer cela pour la question de l'inférence du rôle des facteurs de transcription dans les réseaux de régulation, qui s'avère finalement assez bien circonscrite. Dans la dernière section, nous étendrons cette discussion à l'élaboration de plans expérimentaux.

### 5.3.1 Inférence de signes

**Différentes stratégies de nature probabiliste** La problématique est de savoir comment, à partir de données diverses sur les interactions (en particulier les mesures d'immuno-précipitation et les connaissances sur les interactions protéine-protéine) on peut retrouver quelles sont les interactions qui existent réellement entre les protéines et éventuellement leur rôle (activation ou inhibition) par rapport à un réseau de régulations génétiques.

De nombreuses méthodes ont été proposées à ce propos [BBAIdB07], usuellement testées sur le réseau de la levure *S. cerevisiae* (voir la discussion dans (BMC bioinfo, 2008)). La plupart de ces méthodes sont de nature probabiliste : elles consistent à considérer différents modèles pour les données informant sur les régulations, et ensuite à calculer un optimum qui est considéré comme le réseau inféré. Il s'agit de problèmes d'optimisation non convexes, et, en pratique, rien n'assure que les solu-

tions trouvées sont des optimums globaux plutôt que locaux. Dans les faits, comme nous l'avons déjà mentionné, les différentes approches probabilistes de reconstruction de réseaux ne coïncident pas sur les organismes modèles ; elles coïncident encore moins pour la question de l'inférence des signes.

La difficulté est accrue pour l'inférence de signes en comparaison avec la problématique de la reconstruction de modèle dans la mesure où les modèles prédits sont difficilement validables. À cause des jeux de données disponibles, l'organisme modèle habituellement étudié dans ce cadre est la levure *S. cerevisiae*, mais il n'existe pas, comme pour *E. coli*, de base qui référence les différentes interactions entre les molécules de cet organisme dans la littérature. Un travail considérable a été fait pour élaborer la base *SGD* intégrant les données génétiques sur la levure [HBC<sup>+</sup>01]. Cependant les relations protéine-protéine y sont fournies sans détailler les influences transcriptionnelles et en particulier leur sens. Le seul travail de référencement des publications concernant les interactions chez la levure que nous connaissons actuellement a été fait dans [GBBK02]. Malheureusement, les informations proposées alors ne sont pas disponibles sous la forme de base de données ni actualisées.

On se retrouve donc dans une situation où la communauté propose des méthodes pour inférer les réseaux de régulation mais n'a pas la possibilité de valider ses travaux par une référence commune.

**Apports d'une approche par contraintes ?** La problématique générale de l'inférence de signes entre parfaitement dans le cadre des contraintes que nous avons explicitées dans la section 5.2.1 : contrairement à la question de l'analyse d'un graphe d'influence dans lequel les signes des arcs sont connus, on suppose qu'on a à notre disposition un graphe d'influence dont les signes des arcs sont inconnus (ce qui est exactement l'information fournie par les données d'immuno-précipitation) et différents jeux de données correspondant à des stress environnementaux ou des perturbations génétiques. On peut alors écrire une contrainte pour chaque sommet et chaque jeu de données disponible, ce qui amène à un nouveau système de contraintes. Les concepts de diagnostic et de prédiction trouvent alors tout leur sens ici :

- Les phases de diagnostic vont permettre de détecter les modules du réseau qui présentent un comportement non trivial. En effet, vue la forme du système de contraintes, les incohérences sont apportées par des jeux de données différents qui prédisent des signes différents pour certaines interactions.
- Les phases de prédiction permettent d'identifier des signes de certaines régulations.

L'apport de cette approche par rapport à celles existantes est double. D'abord, la partie diagnostic n'est généralement pas abordée, puisque les méthodes existantes proposent toujours une solution considérée comme optimale, qui n'est pas confrontée a posteriori aux données observées. Plus fondamentalement, l'approche par contrainte consiste à sur-approximer l'ensemble des possibilités admissibles pour les signes des interactions, puis à rechercher des invariants en leur sein ; on ne s'oblige plus à proposer un signe pour l'ensemble des interactions du système comme le font les approches probabilistes, qui courent le risque de calculer un optimum local et

non pas global. L'approche par contrainte permet donc d'identifier la partie qui devrait faire consensus entre les différentes approches probabilistes.

### 5.3.2 Diagnostic : identifier des interactions avec des effets multiples

#### Interprétation des incohérences : interactions avec de multiples effets

Comme mentionné plus haut, les phases de diagnostic permettent d'identifier les sous-modules pour lesquels différents jeux de données indiquent des modes de régulations différents pour une même interaction. Selon nous, cela signifie que soit une interaction est manquante dans le module concerné, soit qu'une protéine du module peut à la fois agir en tant que répresseur et activateur sur sa cible en fonction de l'état du système. On parle alors d'*interaction avec de multiples effets*. Nous avons analysé les types d'incohérences qui se retrouvent dans le réseau de la levure et de *E. coli*. Globalement, les incohérences se classifient comme illustré dans la figure 5.7.

- les modules de Type 1 sont composés d'une seule régulation. Plusieurs jeux de données différents indiquent que la source inhibe et active sa cible en fonction du contexte expérimental.
- les modules de Type 2 sont composés de gènes régulés par le même prédécesseur direct. Dans l'exemple, l'interaction entre *Sum1* et *YFL040W* est inférée pour être une activation, tandis que celle entre *Sum1* et *DIT2* est inférée comme une inhibition. Cependant, un dernier jeu expérimental indique que (*YFL040W* et *DIT2*) sont tous les deux sur-exprimés, ce qui génère une interaction avec un effet multiple.
- Les modules de Type 3 correspondent à des gènes dont la variation est prédite mais qui partagent un même prédécesseur indirect, induisant une interaction avec un effet multiple.
- Les modules de type IV concernent des gènes dont la variation est prédite, induisant des effets multiples entre leurs prédécesseurs et leurs successeurs.

**Réseau transcriptionnel de *E. coli*** Une illustration de l'intérêt des approches de diagnostic pour l'inférence du rôle des facteurs de transcription peut-être faite avec le réseau déjà étudié de *E. coli*. Nous avons repris les jeux de données correspondant aux 61 stress environnementaux différents donnés dans [FHT<sup>+</sup>07], en oubliant les signes donnés par RegulonDB pour les régulations dans ce réseau. Nous avons alors constaté que pour dix modules, présentés dans la Figure 5.8, différents jeux de données indiquent que des régulations sont à la fois des activations et des inhibitions. Dans ces cas là, il est probable que des mécanismes régulateurs des molécules concernés manquent dans le réseau pour décrire correctement les stress considérés.

**Application au réseau de la levure** Nous avons aussi appliqué nos algorithmes sur l'inférence des signes au réseau transcriptionnel de la levure *S. cerevisiae*. Pour procéder à cette étude, nous avons considéré différents réseaux publiés

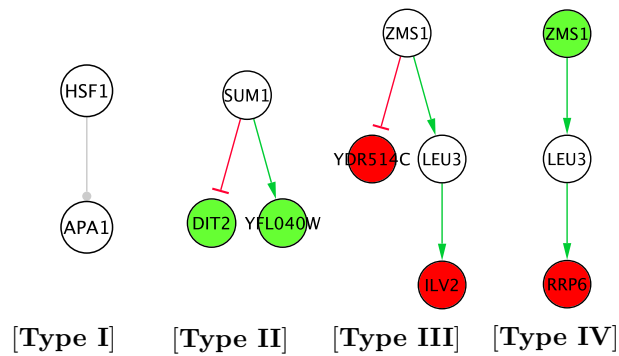


FIG. 5.7 – Les différents types de modules incohérents quand au mode de régulation retrouvés chez des réseaux transcriptionnels

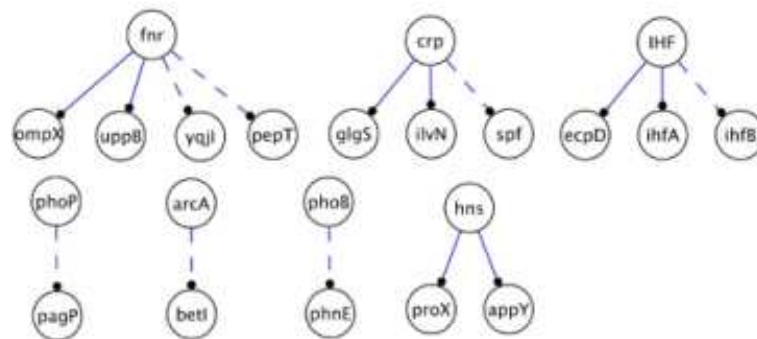


FIG. 5.8 – Les modules du réseau transcriptionnel de *E. coli* pour lesquels soit une interaction est manquante, soit une protéine peut agir à la fois comme activateur et inhibiteur de sa cible en fonction du contexte expérimental. Les flèches pleines représentent des modules de Type 2, les flèches pointillées des modules de Type 1.

sur cette espèce. La plupart de ces réseaux ont été construits à partir d'un travail expérimental phénoménal qui a consisté à rechercher par immuno-précipitation l'ensemble des protéines qui se fixent sur la plupart des facteurs de transcriptions connus chez la levure [LRR<sup>+</sup>02]. Ces données ont permis à différents auteurs de proposer des réseaux transcriptionnels orientés mais non-signés pour la levure [HGL<sup>+</sup>04, MWG<sup>+</sup>06, SSR<sup>+</sup>03, NTIM05, BBAIdB07]. À partir de ces sources, nous avons sélectionné différents réseaux :

- Le premier est le cœur du réseau directement déduit des données chIP-chip de [LRR<sup>+</sup>02], et déjà introduit dans [KPST03]. Il contient 31 nœuds et 52 interactions.
- Le deuxième réseau contient toutes les interactions entre les facteurs de transcription testés dans l'étude de [LRR<sup>+</sup>02]. Dans ce réseau, les interactions sur les protéines qui ne sont pas des facteurs de transcription sont omises du réseau.
- Le troisième est le réseau étendu à l'ensemble des facteurs de transcription de la levure, publié dans [MWG<sup>+</sup>06].
- Le dernier réseau considéré, grande-échelle, consiste en l'ensemble des régulations obtenues par chIP-chip dans [LRR<sup>+</sup>02]. Il contient 2419 sommets et 4344 nœuds.

Concernant la phase de diagnostic, il faut constater que le nombre de modules pour lesquels des interactions avec de multiples effets apparaissent est très important pour l'ensemble des quatre réseaux considérés. Ces modules ainsi que nos prédictions ont été listés et proposés à la communauté (voir le matériel supplémentaire de (BMC bioinfo, 2008)). Nous espérons pouvoir travailler avec des expérimentateurs pour prouver en particulier les phénomènes de régulations dépendante du contexte.

### 5.3.3 Prédiction : utiliser la redondance pour tenir compte des incertitudes ?

**Utilisation de la redondance** La présence de nombreuses incohérences entre les différents jeux de données, à la fois pour *E. coli* et *S. cerevisiae* rend les phases de prédiction délicates. Pour procéder à des prédictions, nous avons d'abord réduit les jeux de données à leur partie cohérente, comme expliqué dans la partie 5.2.5, pour identifier la partie des données et du réseau pour laquelle nous étions certains qu'il y avait compatibilité.

Nous avons ensuite appliqué la procédure d'inférence de signes sur le réseau de *E. coli* dont nous avons préalablement oublié les signes, puis comparé les résultats inférés aux signes initiaux. Le taux de faux-positifs était alors de 28,3%. La raison de ce nombre important de faux positifs réside dans le fait que certains signes d'interactions sont inférés à partir d'un seul jeu de données, ce qui n'est pas suffisant pour détecter des interactions manquantes ou des comportements complexes.

Pour répondre à ce problème, nous avons proposé de tirer profit de la présence de plusieurs jeux de données pour considérer qu'une prédiction est valable si elle est obtenue de manière redondante pour plusieurs jeux de données indépendants. Avec cette opération de *filtrage des prédictions par redondance*, nous avons obtenu

un taux de faux-positif de 2,5% (voir (BMC bioinfo, 2008) pour plus de précisions). Ainsi, la redondance permet de faire des prédictions sur les signes des régulations malgré l'incertitude des données et les incohérences contenues dans le réseau.

**Données de perturbations génétiques** Nous avons d'abord appliqué nos algorithmes à ces réseaux avec des données correspondant à des des inhibitions de gènes (Knock-Out) et des activations de gènes [HMJ<sup>+</sup>00]. Bien qu'un grand nombre de mutants (plus de 200) aient été produits dans cette étude, nous avons constaté que les variations observées entre les espèces normales et mutants concernaient peu de sommets (seulement 1.6 % est en moyenne différentiellement exprimé, les expériences couvrant au total 12% du réseau). À partir de ces données, nous avons inféré 162 signes du réseau grande-échelle ; parmi celles référencées dans la littérature, 75% des signes inférés ont été validés.

Avec cette étude, nous avons surtout pu comparer notre approche à la seule approche que nous connaissons et permettant d'inférer des signes à partir de données d'expression [YMM<sup>+</sup>05]. Les taux d'inférence des deux méthodes sont comparables, et les prédictions sont identiques sur leur ensemble d'intersection. Par contre, nous avons constaté que les deux méthodes sont complémentaires concernant leurs prédictions : nos méthodes permettent d'inférer des signes à la limite externe du réseau réduit tandis que l'approche de [YMM<sup>+</sup>05] infère des signes à l'intérieur du réseau réduit.

Cela provient du type d'analyse utilisé : dans [YMM<sup>+</sup>05], les auteurs utilisent une notion de cohérence par chemin (voir Section 5.2.3). Ils cherchent à identifier les chemins les plus probables qui connectent un gène muté à des protéines dont la concentration a varié suite à la mutation. La méthode d'optimisation recherche donc le plus petit nombre d'interactions qui supportent la plus grande information possible dans le réseau, et se concentre donc dans le cœur du réseau (réseau réduit qui rassemble ses cycles). Or, c'est au sein de ce cœur que se concentre la dynamique comme résultant de la combinatoire des interactions. Les considérations de déplacements d'états stationnaires qui fondent l'approche par contraintes ne sont donc pas assez fines pour capter des informations dans cette zone du réseau.

Inversement, les approches par localisation des chemins physiques probables n'empruntent pas les voies extérieures au cœur, qui sont des arbres acycliques relativement indépendants. Les prédictions de [YMM<sup>+</sup>05] ne peuvent donc pas porter sur ces interactions. Par contre, ces voies sont bien adaptées pour une approche par contraintes sur les déplacements d'états stationnaires.

**Données de stress environnementaux** Nous avons ensuite exploité nos algorithmes à partir de 15 jeux de données sur la levure correspondant à des stress environnementaux, disponibles sur la base SGD [HBC<sup>+</sup>01]. Nos conclusions ont porté sur plusieurs points :

- Les données de stress expérimentaux sont beaucoup plus informatives que les données de perturbations génétiques au sujet des interactions existant dans un réseau. En ce sens, il serait intéressant d'adapter les différents algorithmes

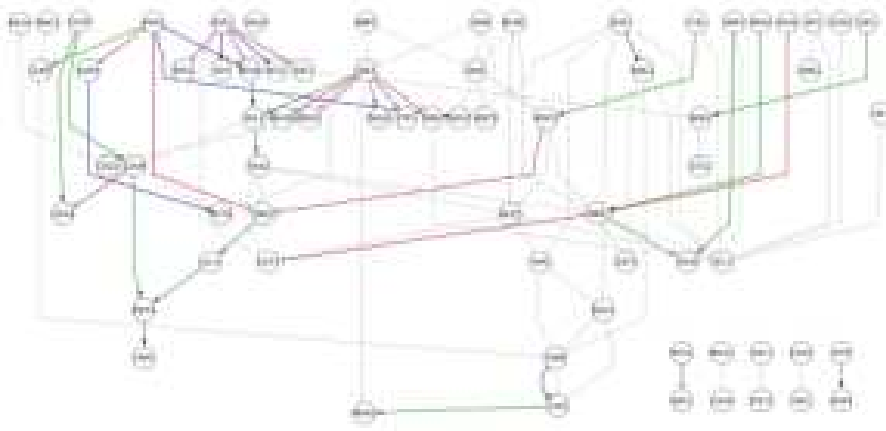


FIG. 5.9 – Signes inférés dans le réseau transcriptionnel de la levure proposé dans [MWG+06]

d'optimisation [YMM<sup>+</sup>05, SSR<sup>+</sup>03] pour qu'ils prennent en compte ce type de données plutôt que des données de perturbations génétiques.

- En fonction des réseaux, le taux d'inférence est compris entre 19% et 37%. Ce taux est similaire au taux théorique pour une vingtaine d'expériences calculé dans le cas de *E. coli*.
- Comme dans le cas de *E. coli*, l'opération de filtrage (prédictions obtenues à partir de jeux expérimentaux différents) permet de diminuer significativement le nombre de prédictions invalidées par les références bibliographiques de [GBBK02], sans doute dues à des interactions manquantes.

### 5.3.4 Combien de jeux expérimentaux sont-ils nécessaires pour procéder à une inférence ?

Les applications réelles des méthodes de [YMM<sup>+</sup>05] ou de nos méthodes ont montré que les prédictions de signes portaient finalement sur une petite partie du réseau (au mieux 20%), malgré la variété des jeux de données considérés. Nous nous sommes alors demandé quelle est l'influence du nombre de jeux de données sur le résultat du processus d'inférence : est-il nécessaire de multiplier les observations pour obtenir un meilleur taux d'inférence de signes ?

**Maximum d'inférence de signes d'interactions par règles causales** Nous avons considéré le réseau de *E. coli* en y intégrant les signes des arcs donnés par RegulonDB pour générer aléatoirement des solutions au système de contraintes. Chaque solution du système modélise ainsi un jeu d'observation théorique en réponse à un stress environnemental. À partir de ces observations aléatoires, nous



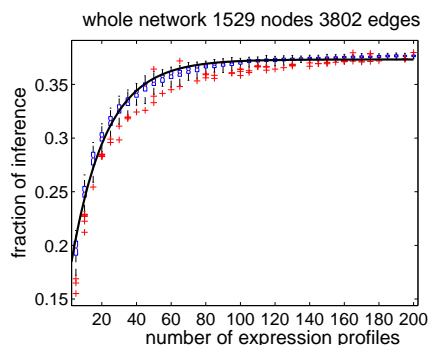


FIG. 5.10 – Pourcentage de signes inférables dans le réseau de *E. coli* en fonction du nombre de jeux de données d’observations disponibles

avons supprimé les signes des régulations et calculé la proportion de signes déductibles. Cette proportion dépend du nombre de jeux d’observations disponibles pour procéder à l’inférence des signes. Les courbes obtenues montrent qu’il semble exister un taux maximal de signes de régulation inférables, qui semble proche de 40%. Les trois cinquièmes restant du réseau ne sont pas inférables à partir de méthodes causales (voir Fig 5.10).

Cependant, rien n’indique que le palier visible sur la courbe est effectivement un maximum. En mêlant les diagrammes de décision, des notions d’élimination de variables et une partition des contraintes, (voir l’algorithme 2 dans (BMC bioinfo, 2008)), P. Veber a montré que le pourcentage maximal de signes inférables dans le réseau de *E. coli* à partir de données de stress expérimentaux est effectivement de 40.8%.

Nous pouvons également conclure que 30 expériences distinctes sont en moyenne suffisantes pour inférer la moitié des signes inférables ; ceci sous-entend que multiplier les expérimentations au delà de quelques dizaines ne permet pas d’obtenir un gain significatif dans la problématique de l’inférence des modes de régulation avec notre approche.

**Rôle de la dynamique ?** Nous avons aussi procédé au calcul du maximum théorique pour le réseau réduit de *E. coli*. Si le maximum théorique est assez proche de celui du réseau global (autour de 40%) (BMC bioinfo, 2008), il est notable de constater que le nombre moyen d’expérimentations nécessaires pour retrouver la moitié des signes inférables est de 1000 expérimentations. Notre interprétation est assez proche de celle que nous avons eu au sujet des localisations différentes pour les interactions dont le signe est inférable avec une approche par contraintes ou par recherche de chemin le plus probable [YMM<sup>+</sup>05]. La dynamique du système se concentre au cœur du réseau, où le degré entrant moyen est plus important qu’en périphérie, et des observations de stress expérimentaux ne permettent pas de la déterminer uniquement. Par contre, il existe une grande partie du réseau de *E. coli*

---

pour laquelle les aspects dynamiques jouent peu : le graphe d'influence y ressemble à un graphe acyclique. Sur cette partie du réseau, des analyses causales avec peu d'expérimentations sont largement suffisantes pour comprendre la structure et le fonctionnement du système ; la dynamique du système entre peu en jeu.

## 5.4 De nombreuses questions à explorer : vers une aide au raisonnement pour le biologiste ?

Au début du chapitre, nous sommes partis d'une contrainte qui formalise l'intuition des biologistes sur les comportements relatifs de molécules au sein d'un système suite à un stress. Nous avons ainsi montré comment l'étude de solutions de systèmes de contraintes permet :

- de localiser des incohérences dans un réseau présenté sous la forme d'un graphe d'influence, à partir d'un ou plusieurs jeux d'observations de taille variable ;
- de prédire les variations de nœuds non observés, y compris lorsque des incohérences sont présentes dans certaines zones du réseau ;
- de mesurer la robustesse d'un réseau par rapport à un jeu de données de grande taille ;
- d'identifier le rôle de régulateur d'un composant (activateur ou répresseur),
- d'estimer le nombre de jeux expérimentaux suffisant pour procéder à des inférences de signes.

On doit constater que la question (dynamique) de la prédiction de comportement a été quelque peu laissée de côté au profit d'une aide au raisonnement des biologistes pour exploiter leurs données et leurs connaissances. En terme de perspective, nous allons maintenant décrire dans quelles directions il est important de pousser cette aide au raisonnement.

Les différentes questions qui se posent, et qui semblent abordables avec les concepts décrits dans ce chapitre concernent la notion de plan expérimental. Il va s'agir de proposer les actions sur un système qui sont optimales pour obtenir une réponse attendue par le biologiste.

Cependant, comme nous allons le voir, ce concept de plan expérimental est très vague : dans la réalité, les objectifs des expérimentateurs sont très variés, et la difficulté réside avant tout à identifier les différentes attentes des expérimentateurs pour les formaliser sous forme de problèmes résolubles. Nous allons donc proposer ici une classification (subjective) des différentes ambitions des biologistes qui semblent abordables avec des approches de contraintes. Il s'agit de perspectives dans la mesure où aucune implémentation n'a encore été effectuée.

### 5.4.1 Différentes questions autour des plans expérimentaux

**Qu'est ce qu'un plan expérimental ?** C'est un lieu commun de dire que tous les éléments d'un système sont loin d'être observables, ou à un prix déraisonnable. Une expérimentation est donc loin d'être faite au hasard, et nous allons discuter

---

dans ce chapitre de la fonction que peut avoir un informaticien ou un mathématicien dans le processus d'élaboration d'une expérimentation.

Par *plan expérimental* nous voulons dire le fait de choisir de manière raisonnée différentes actions sur différents produits du système de manière à produire des observations nouvelles sur le système éclairant son fonctionnement. L'aspect *plan* réside dans le processus de sélection des expérimentations.

**Différentes approches dans la littérature** Différents groupes de scientifiques se sont intéressés à cette question. Notons en particulier les travaux du groupe d'Ideker sur la levure [YIJ04, YMM<sup>+</sup>05], dont la stratégie a été d'inférer des réseaux de régulations transcriptionnelles pour *S. cerevisiae* avec des méthodes d'inférence par maximum de vraisemblance. Le résultat de ces analyses a produit différents réseaux d'interactions ayant statistiquement les mêmes probabilités d'être actifs. Pour évaluer la pertinence de ces différents réseaux, le groupe a sélectionné plusieurs expériences de mutation dont les effets sur les réseaux théoriques étaient différents, de manière à isoler le réseau d'interaction réellement actif [YMM<sup>+</sup>05]. Il faut noter que dans ce cas, la sélection des expérimentations via l'analyse de leurs effets théoriques a pu se faire *à la main* puisque l'analyse probabiliste avait permis de réduire le choix à un petit nombre de réseaux de petite taille.

Une approche d'analyse *automatique* des résultats d'analyse pour proposer de nouvelles expérimentations a été mise en place par l'école anglaise de logique inductive autour de R. King, dans le programme *Robot Scientist* [BMO<sup>+</sup>01, RKK<sup>+</sup>01, KGC05b]. Le projet a consisté à produire une boucle d'expérimentations couplée avec du raisonnement automatique pour sélectionner parmi différentes voies de synthèse des acides aminés aromatiques possibles, celles qui étaient réellement actives. La partie expérimentale était seulement semi-automatique, dans la mesure où des expérimentateurs venaient procéder aux manipulations des cellules, mais le choix des différentes molécules à insérer dans les expérimentations était par contre opéré par des analyses logiques.

Ce genre d'approche est promis à un avenir certain, avec l'essor des différentes techniques *omiques* et des progrès dans l'automatisation de la manipulation des produits génétiques et de l'observation des cellules. Concrètement, différentes plateformes grande-échelle sont en train de se mettre en place au sein des laboratoires de biologie. La fonction de ces plateformes est de procéder à des manipulations (KO, injection de siRNA ou d'anticorps, combinaisons de différents produits chimiques) de manière automatique et d'observer les effets de ces manipulations. Les observations sont souvent assez grossières (survie ou non de la cellule, éventuellement quelques mesures physiques).

Dans ce cadre, un besoin important va rapidement se faire sentir pour l'analyse des résultats et surtout le choix des combinaisons de produits à injecter dans les cellules en fonction de leurs effets. En précurseur à ces analyses, différents travaux portent sur l'analyse des connaissances sur le caractère *essentiel* ou non des gènes chez *E. coli* ou la levure [HGH01].

Un dernier domaine où la notion de plan expérimental est primordiale est la

---

biologie synthétique, là encore en plein essor. Il s'agit de combiner des constructions biochimiques pour *forcer* un comportement attendu. Le choix des différents mutants qui sont produits est alors guidé par la modélisation des effets attendus. Dans ce cadre, des approches de vérification de modèles avec les logiques temporelles ont donné des résultats prometteurs [BYBW07]

**Différentes attentes des expérimentateurs** Une remarque naïve mais importante est à faire avant de discuter de l'intérêt des approches qualitatives pour l'élaboration de plans expérimentaux : les expérimentateurs peuvent avoir des attentes de nature complètement différentes concernant le résultat de leurs expérimentations. Et les analyses qui doivent leur être proposées doivent être adaptées à ces différentes attentes.

- Le premier objectif des expérimentateurs est de *valider* le modèle qui est étudié. En effet, les modèles considérés sont usuellement constitués à partir d'indications dans la littérature déduites d'expérimentations dans divers milieux, ou de données expérimentales sujettes au bruit. Avant de faire confiance à un tel modèle et en particulier exploiter ses prédictions, il est nécessaire de valider expérimentalement le fait que le modèle soit un bon niveau de description des phénomènes considérés lors de la construction du modèle.
- Le deuxième objectif est d'exploiter le modèle pour *contrôler le système*. Dans ce cas là, l'objectif est de trouver les expérimentations qui permettent de perturber le système de manière la plus pertinente et convaincante possible.
- Le troisième objectif est d'exploiter le modèle pour *comprendre le fonctionnement du système*. Dans ce cadre, on cherchera en particulier à faire des expérimentations qui permettent de démêler la combinatoire des voies, et d'identifier le cœur du fonctionnement.

**Des attentes à des questions concrètes et formalisables?** Les lignes qui vont suivre vont détailler comment ces trois objectifs (validation, contrôle, fonctionnement) peuvent être dérivés dans des questions qui sont formalisables en termes de contraintes associées à un graphe d'influence. Si nous avons procédé à quelques tests pour évaluer la pertinence des différents concepts présentés ici, les implémentations réelles des approches proposées ne sont pas encore réalisées, et il s'agira d'une des principales perspectives de ce travail.

#### 5.4.2 Comment sélectionner des sommets pour valider un système ?

La première attente d'un biologiste est de valider son modèle, c'est-à-dire de montrer que les interactions intégrées dans le modèle sont suffisantes pour expliquer les variations observées. Nous discutons ici de l'apport des approches de contraintes pour cette notion, en fonction du rôle donné aux observations préalables sur le système.

**Approche usuelle pour valider un modèle vis-à-vis d'un protocole expérimental donné : bloquer les influences sur un sommet** Dans les publications en biologie, la validité d'un modèle est prouvée en considérant un certain nombre de sommets du modèle pour lesquels on observe une variation pendant un protocole expérimental. Successivement, pour chaque sommet, la validation consistera à inactiver tous les composants qui agissent directement sur le sommet, à renouveler le protocole expérimental et à vérifier que la concentration du produit ne bouge plus pendant l'expérimentation.

Dans ce cas, on admettra que, pendant l'expérimentation, les voies d'influence du système sur le sommet considéré initialement sont à chercher parmi ses prédécesseurs dans le modèle. Pour l'exemple de l'opéron Lactose, une validation possible consiste ainsi à bloquer les effets de Le, LacY et LacZ et de vérifier que la concentration de Li ne varie alors plus.

Le fait de bloquer les influences entrant sur un composant plutôt que le composant lui-même est rendu nécessaire pour valider les interactions du modèle : en invalidant simplement le composant, on ne peut pas savoir si toutes les interactions intégrées dans le modèle qui concernent ce composant sont suffisantes pour expliquer ses variations.

Idéalement, il faudrait procéder ainsi pour tous les sommets du réseau. Dans les faits, les expérimentateurs se concentrent sur quelques sommets du réseau, souvent sélectionnés parce que les inactivations des prédécesseurs sont techniquement possibles et non léthales.

Cette approche mérite quelques commentaires :

- Avec ce type d'expérimentations, on montre que le modèle est valide pour décrire le fonctionnement du système *pendant l'expérimentation* considérée. Il est sous-entendu que dans un autre cadre expérimental, d'autres influences peuvent apparaître et la démarche est à renouveler.
- Avec cette approche, on obtient un modèle *valide*, mais qui peut contenir des influences *superflues*. A priori, selon notre expérience, avoir des influences en trop n'est pas considéré comme gênant d'un point de vue expérimental. Les analyses concernant le fonctionnement du système auront justement pour objet dans l'avenir d'identifier le superflu et obtenir un modèle minimal.

**Quelle perturbation est la plus intéressante ? Rechercher une perturbation maximale du système** A priori, l'expérimentateur n'a pas besoin de l'informaticien ou du mathématicien pour procéder à cette démarche de validation. Il lui *suffit* de choisir les sommets qu'il considère comme clés et de procéder à des inhibitions de voies. Le terme *suffit* est largement provocateur dans la mesure où il peut-être extrêmement difficile d'inhiber simultanément l'effet de plusieurs produits sans provoquer la mort des cellules.

Ici, l'informaticien peut cependant être un appui pour aider à identifier les sommets qui permettent de valider au mieux le modèle : il va s'agir de choisir le sommet pour lequel l'inactivation des influences des prédécesseurs induira le plus d'effets possibles dans le système, en comparaison avec les observations du système non perturbé pour le même protocole expérimental. Il sera de plus possible de vérifier

ces effets théoriques pendant l'expérimentation.

Pour formaliser cette question, il faut modéliser précisément les effets de l'inactivation des prédécesseurs d'un composant  $\mathcal{S}$  : puisque nous sommes dans un cadre systémique, l'invalidation des prédécesseurs peut avoir des effets sur de multiples autres composants. Pour mesurer les effets des inactivations des prédécesseurs de  $\mathcal{S}$ , on va comptabiliser le nombre de prédictions obtenues pour un nouveau graphe d'influence (déduit du premier en supprimant tous les prédécesseurs de  $\mathcal{S}$  et leurs influences) et un jeu d'observations obtenu en supprimant du jeu d'observations initiales celles qui partagent avec  $\mathcal{S}$  un prédécesseur indirect. Il s'agit donc de résoudre un système de contraintes avec un nombre polynomial d'appels à un oracle NP-complet.

**Proposer un plan expérimental pour la validation du modèle *indépendant* d'un protocole initial ?** Comme nous l'avons déjà noté, dans cette démarche, le modèle est validé par rapport à un protocole expérimental donné, qui fournit des données sur lesquelles on s'appuie pour optimiser le choix des sommets à perturber. On peut cependant introduire un degré de complexité supplémentaire et chercher quel est le moyen de valider le modèle indépendamment de protocoles expérimentaux.

Cette question se pose en particulier avec la mise en place dans différents laboratoires de plate-formes automatiques de manipulations génétiques : il va bientôt être possible d'inactiver automatiquement de grands nombres de combinaisons de quelques composants différents, et d'observer les effets de ces inactivations.

Pour répondre à cette question, nous avons proposé dans (Complex Us, 2005) et (Biofutur, 2007) d'associer à chaque groupe de  $p$  sommets d'un graphe d'influences un *pouvoir de validation*, qui est le nombre de valeurs pour ce groupe qui sont compatibles avec au moins une solution du système de contraintes :

$$\tau(X_1, \dots, X_p) = 1 - \frac{\text{val}(X_1, \dots, X_p)}{2^p},$$

où  $\text{val}(X_1, \dots, X_p)$  désigne le nombre de valeurs de  $(X_1, \dots, X_p)$  dans  $\{+, -\}^p$  qui s'étendent en une solution globale du système de contraintes associé au réseau considéré.

Nous nous sommes appuyés sur une observation là encore naïve : certains jeux de données n'apportent en fait aucune information sur un système, dans la mesure où toute variation de leurs composants peut s'étendre en une solution globale du système qualitatif. Par exemple, en étudiant la Table 5.1, on constate que pour chaque valeur du triplet  $(Le, G, A)$ , il existe une valuation de  $(Li, LacY, LacZ, LacI, cAMP)$  qui est solution du système associé à l'opéron lactose. Une observation du système limitée aux nœuds  $(Le, G, A)$  ne peut donc pas mettre en défaut le modèle proposé. Notons par ailleurs qu'il ne s'agit pas d'un cas isolé : il y a 56 possibilités d'observer 3 nœuds dans le graphe, et 22 d'entre elles ont un pouvoir de validation nul.

Inversement, parmi les huit valeurs possibles du triplet  $(LacI, A, LacZ)$ , seuls les jeux  $(+, -, -)$  et  $(-, +, +)$  s'étendent en une solution du système. Ainsi, ce jeu

de données s'avère très contraignant pour la validité du modèle. Parmi les groupes de trois composants, il s'agit du jeu qui est le plus astreignant.

Observation	Pouvoir de validation	Observation	Pouvoir de validation
$(LacZ, LacI, G, A)$	0.75	$(Li, Le, LacY, G)$	0
$(cAMP - CRP, LacI, G, A)$	0.75	$(cAMP - CRP, Li, Le, LacY)$	0
$(cAMP - CRP, LacZ, G, A)$	0.75	$(Li, Le, LacY, A)$	0.125
$(LacZ, LacY, LacI, A)$	0.75	$(Li, Le, LacY, LacI)$	0.125
$(Le, LacZ, LacI, A)$	0.75	$(Li, Le, LacZ, LacY)$	0.125
$(cAMP - CRP, LacZ, LacI, G)$	0.75	$(cAMP - CRP, Le, LacY, A)$	0.25

TAB. 5.2 – Pouvoir de validation en fonction du groupe de sommets observés, sur le réseau de l'opéron lactose. Le pouvoir de validation a été calculé pour tous les groupes de 4 sommets ; dans le tableau figurent les six meilleurs scores et les six moins bons. Si le taux d'un groupe est proche de 1, une observation des molécules compatible avec les équations qualitatives du réseau valide fortement le réseau. Notons qu'il ne semble pas y avoir de règle simple pour deviner, à partir du graphe d'influence, quels groupes de sommets sont les plus importants à observer.

Un ensemble de sommets est d'autant plus à même de valider un modèle que  $\tau(X_1, \dots, X_p)$  est proche de 1. On peut donc utiliser ce taux de deux manières différentes :

- si on dispose d'observations sur des composants  $X_1, \dots, X_p$  qui sont compatibles avec le modèle,  $\tau(X_1, \dots, X_p)$  permet de mesurer la pertinence de la validation.
- Si on vient de construire un modèle que l'on doit valider par des expérimentations, on peut rechercher, pour des petites valeurs de  $p$ , quelles sont les  $p$  variables les plus contraignantes et pertinentes pour cette validation.

### 5.4.3 Plan expérimentaux autour du contrôle d'un système

La notion de contrôle de système est bien ancienne en informatique, il s'agit de tout un pan de l'automatique. D'un point de vue biologique, deux notions liées au contrôle me semblent pertinentes et étudiables à l'aide de contraintes : il s'agit d'abord de renverser une compétition, et ensuite de rechercher comment induire un phénotype attendu.

**Comment inverser le fonctionnement d'un système ? Identifier les lieux de compétition** Une première approche concrète pour contrôler le système repose sur la notion de compétition. A priori, on peut considérer qu'une compétition existe dès qu'un sommet  $\mathcal{S}$  subit des influences de signes contraires, ce qui l'oblige à faire un choix. Autrement dit, la variation de  $\mathcal{S}$  ne peut pas être déduite des variations de ses prédécesseurs.

Pour identifier les sommets sur lesquels un choix réel est fait, on propose dans (Biosystems, 2006) de définir comme suit un sommet  $\mathcal{S}$  *porteur de compétition*

pendant une expérimentation : en  $\mathcal{S}$  arrivent deux chemins (a) qui ne partagent aucun sommet en commun (b) dont la source fait l'objet d'une prédiction à partir du jeu d'observation initial dans lequel l'observation sur  $\mathcal{S}$  (s'il y en a une) a été retirée.

Le plan expérimental naturel associé à ce concept est de renverser le comportement du système. En particulier, puisque  $\mathcal{S}$  subit des influences de signes contraires (sans quoi il serait prédit), il est possible de bloquer toutes les influences qui induisent la variation observée. Le modèle prédit alors que la variation de  $\mathcal{S}$  dans ces nouvelles conditions sera inversée.

Dans (Biosystems, 2006), nous discutons de cette notion de compétition, même si les termes utilisés ne sont pas exactement les mêmes. Nous avons ainsi étudié un graphe d'influence étendu pour les régulations autour du métabolisme des acides gras. Nous en déduisons en particulier que LXR, SCD1 ; FADS1 et FADS2, largement mentionnés au chapitre 4, sont porteurs de compétitions.

**Comment induire un phénotype ?** Un autre manière d'envisager la notion de contrôle consiste à rechercher comment provoquer un phénotype donné. Contrairement à ce qui a été discuté jusqu'à maintenant, on ne dispose plus d'expérimentations préalables sur le système. On suppose simplement qu'on dispose d'un certain nombre de produits (si-RNA) ou d'expérimentations (knock-out, sur-expressions grâce à des produits pharmacologiques) permettant d'inhiber ou activer certaines molécules du système, rassemblées dans un ensemble  $\mathcal{C}$ . On souhaite savoir si une combinaison de ces produits permet de produire un phénotype, ce qui, d'un point de vue théorique, correspond à fixer la variation d'un sommet  $\mathcal{S}$  donné.

D'un point de vue contrainte, cela revient à trouver une combinaison d'affectations  $\{+, 0, -\}$  pour chaque variable appartenant à l'ensemble des produits contrôlables  $\mathcal{C}$  telle que la prédiction de ces affectations est l'observation souhaitée sur  $\mathcal{S}$ .

Obtenir une solution pertinente à ce type de contraintes n'est possible que lorsqu'il est possible d'agir sur toutes les entrées du système, et sur des sommets qui admettent un nombre limité de prédécesseurs. En particulier, ce type de question vise à être utilisé dans des laboratoires sur puces, où des expérimentations en masse pourront être produites automatiquement.

#### 5.4.4 Identifier des groupes d'intérêt dans un réseau

Si on met de côté la notion de plan expérimental, on peut interpréter la dernière question *comment induire un phénotype* en terme fonctionnel : résoudre cette question revient à identifier un groupe de sommets fondamental pour comprendre le fonctionnement du réseau. Cette notion se décline plus généralement vers l'identification de sous-groupes du réseau qui ont un sens en terme de comportement. Là encore, aucun concept unique n'émerge pour l'importance et le sens donné à un sous-groupe. Nous avons quand même distingué trois questions fondamentales qui sont formalisables en terme de contraintes et ont un sens du point de vue biologique.



**Comment expliquer un phénotype ?** Autour de cette notion de phénotype on trouve aussi la notion d'explication : si on dispose d'observations grande-échelle sur un système (type transcriptome), on peut se demander naturellement quels sont les éléments clé qui induisent les observations données.

Cette question se traduit bien en terme de contraintes : on considère un jeu d'observations sur un système, et on y ajoute ses prédictions. Il est alors possible de rechercher parmi les sous-ensembles de sommets observés ou prédits un *système générateur*, qui induit l'ensemble des prédictions. On obtient ainsi un certain nombre de jeux minimaux qui expliquent une expérimentation donnée et dont la perturbation permettra par la suite de changer le comportement du système.

Cette notion permet aussi de mieux comprendre le fonctionnement du système : un groupe minimal rassemble des éléments qui signent cette expérimentation. Imaginons ensuite qu'on associe à ces éléments une fonction via une annotation par Gene Ontology, on pourrait parvenir à mieux comprendre en quoi une expérimentation est la combinaison de différentes influences.

**Comment trouver l'origine commune de différents profils ?** On peut aller plus loin dans cette approche et rechercher une explication parmi une liste de causes possibles. Nous travaillons actuellement par exemple sur la question de l'analyse de données e-QTL [GDZ<sup>+</sup>06].

- Les *quantitative trait loci* (QTL) sont des portions d'ADN qu'on a pu associer à la présence d'un phénotype donné. Dans notre cas, le facteur d'intérêt est l'état d'engraissement du poulet. Ainsi, ces études de QTL permettent d'identifier une liste de gènes dans laquelle se trouvent les causes génétiques du phénotype considéré [LPCa06, APL<sup>+</sup>06].
- En parallèle, une approche e-QTL permet de différencier pour de multiples gènes du génome leur différence d'état d'expression entre les individus dont le phénotype est différent. Dans cette approche, on dispose de données de transcriptome : grossièrement, on sait observer la variation des ARN de tous les gènes du génome.

La question qui se pose est alors de démêler, parmi les gènes causes, quelle combinaison permet de produire le phénotype observé. Pour cela, nous proposons de chercher pour chaque produit dans l'ensemble des causes l'ensemble des gènes sur lesquels il y a une influence, et de rechercher les combinaisons qui sont compatibles avec les données e-QTL.

Plus formellement, il va s'agir de rechercher une valuation des sommets considérés comme des causes possibles (issues des données QTL) qui est cohérente biologiquement avec les observations issues des données e-QTL. Dans ce cas, les perturbations (solutions sans signification biologique) introduites par les boucles positives dans les solutions des systèmes qualitatifs ne peuvent plus être admises. Nous proposons de remplacer la notion de cohérence que nous utilisons habituellement par la notion plus restrictive de *cohérence biologique* discutée à la section 5.2.3.

Cela nécessite avant tout de disposer de relations entre les gènes causes et les cibles observées dans les e-QTL. Dans le cadre de sa thèse, Pierre Blavy, soutenu par François Moreews, construit actuellement un réseau grande-échelle qui rassemble les

---

connaissances sur les régulations entre les gènes liés au métabolisme des acides gras. Il s'agit de rassembler à la fois les connaissances des experts sur des protéines identifiées comme importantes (*LXR*, *PPAR*, *SCD1*, ...), et de les comparer aux informations qui existent au sujet des gènes variant significativement dans les e-QTL, disponibles soit dans la base PathwayCommons soit dans les services des sociétés Ingenuity et Biobase. À la fin de ce travail, nous aurons un ensemble de connaissances dans lequel nous pourrions rechercher les origines du facteur d'engraissement.

**Comment contraindre le comportement du système ?** Enfin, d'un point de vue théorique, de même que nous avons défini le pouvoir de validation d'un sous-ensemble du modèle, on peut définir le pouvoir de prédiction d'un ensemble de sommets comme le ratio entre le nombre de prédictions que produisent des observations sur cet ensemble et le nombre total de jeux compatibles. En choisissant un ensemble de sommets optimal pour ce ratio, on obtient des jeux de molécules dont les observations sont optimales en terme de couverture dans les prédictions induites.

Surtout, avec ce dernier concept, on obtient une définition des sommets clés du modèle comme ceux qui influent le plus sur ce modèle. On obtient une alternative aux approches statistiques de la théorie des réseaux biologiques, où l'importance d'un sommet est en rapport avec sa valence et sa conservation au cours de l'évolution [BA99]. Nous avons en perspective de comparer ces deux notions sur le réseau de *E. coli*, pour vérifier que les sommets conservés au cours de l'évolution sont effectivement ceux qui contraignent le plus le réseau.

## 5.5 Synthèse des perspectives

**Étude de la dynamique des réseaux** Finalement, les différents travaux présentés dans ce chapitre permettent d'illustrer l'idée selon laquelle l'étude d'un réseau doit se faire de manière progressive, à plusieurs échelles.

- Une première approche large échelle permettra d'analyser avec un niveau grossier de détail les différentes connaissances et données à la disposition du modélisateur. Dans ce cadre, les méthodes qualitatives de ce chapitre semblent adaptées.
- Une deuxième étude concernera la dynamique de sous-groupes identifiés comme complexes dans la première approche. Il peut s'agir de modules porteurs d'incohérences, ou du cœur du réseau, qui rassemble les aspects dynamiques. Pour ces réseaux, des approches mêlant la description de réseaux métaboliques et des approches par modèles de réactions (abordés dans la section 4.2) sont à généraliser, en particulier pour fournir des règles simples de recherche et d'analyse des états stationnaires, ainsi que des méthodes de prédiction de différentes modifications du systèmes (mutations...).
- Enfin, en dernière étape, des méthodes numériques peuvent être utilisées sur des éléments très restreints du modèle, couplées avec des jeux de données spécifiques. Dans ce cadre, l'élaboration de schémas de raisonnement pour

décider quels sont les jeux d'expérimentations nécessaires peuvent être intéressants.

**Contraintes pour la mise en place de plans expérimentaux** Plus concrètement, les perspectives concernant l'étude des réseaux à leur plus haut niveau d'abstraction (graphe d'influence) sont avant tout de réaliser concrètement une démarche d'aide automatique au raisonnement des biologistes.

- Il va s'agir d'abord d'apporter une réponse aux différentes questions qui ont été identifiées et discutées dans la section précédente, cette réponse devant être formalisée sous la forme d'une propriété de l'espace des solutions de systèmes de contraintes.
- Il faudra ensuite être capable de vérifier ces propriétés de manière efficace en exploitant les possibilités et la complémentarité des solveurs de contraintes comme des diagrammes de décision.

**Une contrainte supplémentaire : schéma d'explication simple** En plus de résoudre ces problèmes, hautement non triviaux, une question fondamentale va être d'intégrer dans le processus de résolution une notion de simplicité de preuve. En effet, du point de vue de l'expérimentateur, un modèle n'est obtenu que comme la juxtaposition de faits relevés dans la littérature ou suggérés par des données. Il s'agit d'un modèle, et avant de se baser sur ses prédictions pour lancer des expérimentations, l'expérimentateur se doit de comprendre l'origine des prédictions, ne serait-ce que pour vérifier une dernière fois que les informations utilisées pour obtenir une prédiction sont valides et que rien n'a été oublié.

En ce sens, il semble important, lors des phases de cohérence, diagnostic, prédiction ou proposition expérimentale, d'associer à toute proposition une *preuve* de celle-ci. Ceci est bien entendu antinomique avec l'approche des solveurs de contraintes SAT, qui procèdent par recherche d'un modèle et pas par déduction comme peut le faire PROLOG.

Cependant, il serait intéressant de proposer, à partir du résultat des solveurs de contraintes, une dernière phase d'analyse avec représentation graphique pour illustrer une preuve simple de l'analyse recherchée, en particulier des arguments utilisés pour aboutir à une prédiction.

Cela serait un outil efficace pour achever de convaincre la communauté en biologique de l'intérêt des approches automatiques en modélisation. Il s'agit cependant d'une question difficile en informatique, qui nécessite des investigations poussées et illustre encore la richesse des questions que pose la biologie aux mathématiques et à l'informatique.

---

# Bibliographie

- [AB07] B Adamczewski and Y Bugeaud. On the complexity of algebraic numbers. I. Expansions in integer bases. *Ann. of Math. (2)*, 165(2) :547–565, 2007.
- [ABB<sup>+</sup>05] S Akiyama, T Borbély, H Brunotte, A Pethő, and JM Thuswaldner. Generalized radix representations and dynamical systems. I. *Acta Math. Hungar.*, 108(3) :207–238, 2005.
- [ABD06] B Adamczewski, Y Bugeaud, and L Davison. Continued fractions and transcendental numbers. *Ann. Inst. Fourier (Grenoble)*, 56(7) :2093–2113, 2006. Numération, pavages, substitutions.
- [ABEI01] P Arnoux, V Berthé, H Ei, and S Ito. Tilings, quasicrystals, discrete planes, generalized substitutions, and multidimensional continued fractions. In *Discrete models : combinatorics, computation, and geometry (Paris, 2001)*, Discrete Math. Theor. Comput. Sci. Proc., AA, pages 059–078 (electronic). Maison Inform. Math. Discrèt. (MIMD), Paris, 2001.
- [ABFJ07] P Arnoux, V Berthé, T Fernique, and D Jamet. Functional stepped surfaces, flips, and generalized substitutions. *Theoret. Comput. Sci.*, 380(3) :251–265, 2007.
- [ABI02] P Arnoux, V Berthé, and S Ito. Discrete planes,  $\mathbb{Z}^2$ -actions, Jacobi-Perron algorithm and substitutions. *Ann. Inst. Fourier (Grenoble)*, 52(2) :305–349, 2002.
- [ABPT06] S Akiyama, H Brunotte, A Pethő, and JM Thuswaldner. Generalized radix representations and dynamical systems. II. *Acta Arith.*, 121(1) :21–61, 2006.
- [ABPT08] S Akiyama, H Brunotte, A Pethő, and JM Thuswaldner. Generalized radix representations and dynamical systems. III. *Osaka J. Math.*, to appear, 2008.
- [Adl98] RL Adler. Symbolic dynamics and Markov partitions. *Bull. Amer. Math. Soc. (N.S.)*, 35(1) :1–56, 1998.
- [AFHI07] P Arnoux, M Furukado, E Harriss, and S Ito. Algebraic numbers and automorphisms of free groups. submitted, 2007.
- [AFS04] JD Angeli, JE Ferrell, and ED Sontag. Detection of multi-stability, bifurcations, and hysteresis in a large class of biological positive-feedback systems. *PNAS*, pages 1822–1827, 2004.
- [AI01] P Arnoux and S Ito. Pisot substitutions and Rauzy fractals. *Bull. Belg. Math. Soc. Simon Stevin*, 8(2) :181–207, 2001.
- [AIS01] P Arnoux, S Ito, and Y Sano. Higher dimensional extensions of substitutions and their dual maps. *J. Anal. Math.*, 83 :183–206, 2001.
- [Aki98] S Akiyama. Pisot numbers and greedy algorithm. In *Number theory (Eger, 1996)*, pages 9–21. de Gruyter, Berlin, 1998.
- [Aki99] S Akiyama. Self affine tiling and Pisot numeration system. In *Number theory and its applications (Kyoto, 1997)*, volume 2 of *Dev. Math.*, pages 7–17. Kluwer Acad. Publ., Dordrecht, 1999.
- [Aki00] S Akiyama. Cubic Pisot units with finite beta expansions. In *Algebraic number theory and Diophantine analysis (Graz, 1998)*, pages 11–26. de Gruyter, Berlin, 2000.
- [Aki02] S Akiyama. On the boundary of self affine tilings generated by Pisot numbers. *J. Math. Soc. Japan*, 54 :283–308, 2002.
- [Aki07] S Akiyama. Pisot number system and its dual tiling. In *Physics and Theoretical Computer Science (Cargese, 2006)*, pages 133–154. IOS Press, 2007.
- [AKM65] RL Adler, AG Konheim, and MH McAndrew. Topological entropy. *Trans. Amer. Math. Soc.*, 114 :309–319, 1965.
- [All98] C Allauzen. Une caractérisation simple des nombres de Sturm. *J. Théor. Nombres Bordeaux*, 10(2) :237–241, 1998.

- 
- [Alo06] U Alon. *An Introduction to Systems Biology : Design Principles of Biological Circuits*. CRC Press, 2006.
- [AP98] J Anderson and I Putnam. Topological invariants for substitution tilings and their associated  $C^*$ -algebras. *Ergodic Theory Dynam. Systems*, 18 :509–537, 1998.
- [APL<sup>+</sup>06] B Abasht, F Pitel, S Lagarrigue, P Le Bihan-Duval, P Le Roy, and al., Fatness qtl on chicken chromosome 5 and interaction with sex. *Genet. Sel. Evol.*, 38 :297–311, 2006.
- [APUM03] M Antoniotto, A Policriti, N Ugel, and B Mishra. Model building and model checking for biochemical processes. *Cell Biochemistry and Biophysics*, 38 :271–286, 2003.
- [AR91] P Arnoux and G Rauzy. Geometric representation of sequences of complexity  $2n + 1$ . (Représentation géométrique de suites de complexité  $2n + 1$ ). *Bull. Soc. Math. Fr.*, 119(2) :199–215, 1991.
- [Arn99] VI Arnold. *Mathematical methods of classical mechanics*, volume 60 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 199?. Translated from the 1974 Russian original by K. Vogtmann and A. Weinstein, Corrected reprint of the second (1989) edition.
- [ARS04] S Akiyama, H Rao, and W Steiner. A certain finiteness property of Pisot number systems. *J. Number Theory*, 107 :135–160, 2004.
- [AS98] S Akiyama and T Sadahiro. A self-similar tiling generated by the minimal Pisot number. *Acta Math. Info. Univ. Ostraviensis*, 6 :9–26, 1998.
- [AS02] JP Allouche and JO Shallit. *Automatic sequences : Theory and Applications*. Cambridge University Press, 2002.
- [AS05] S Akiyama and K Scheicher. Intersecting two-dimensional fractals with lines. *Acta Sci. Math. (Szeged)*, 71(3-4) :555–580, 2005.
- [ASY97] KT Alligood, TD Sauer, and JA Yorke. *Chaos*. Textbooks in Mathematical Sciences. Springer-Verlag, New York, 1997. An introduction to dynamical systems.
- [AW70] RL Adler and B Weiss. *Similarity of automorphisms of the torus*. Memoirs of the American Mathematical Society, No. 98. American Mathematical Society, Providence, R.I., 1970.
- [BA99] AL Barabási and R Albert. Emergence of scaling in random networks. *Science*, 286 :509–512, 1999.
- [BAG01] A Broise-Alamichel and Y Guivarc’h. Exposants caractéristiques de l’algorithme de Jacobi-Perron et de la transformation associée. *Ann. Inst. Fourier*, 51(3) :565–686, 2001.
- [Bar03] C Baral. *Knowledge Representation, Reasoning and Declarative Problem Solving*. Cambridge University Press, 2003.
- [BB73] K Burns and J Burns. The control of flux. *Symp Soc Exp Biol.*, pages 65–104, 1973.
- [BBAIdB07] M Bansal, V Belcastro, A Ambesi-Impiombato, and D di Bernardo. How to infer gene networks from expression profiles. *Mol Syst Biol.*, 3(78), 2007.
- [BBH03] GD Bader, D Betel, and CW Hogue. BIND : the Biomolecular Interaction Network Database. *Nucleic Acids Res*, 31(1) :248–50, 2003.
- [BBK06] V Baker, M Barge, and J Kwapisz. Geometric realization and coincidence for reducible non-unimodular pisot tiling spaces with an application to beta-shifts. *Ann. Inst. Fourier*, 56(7) :2213–2248, 2006.
- [BBLT06] G Barat, V Berthé, P Liardet, and J Thuswaldner. Dynamical directions in numeration. *Ann. Inst. Fourier (Grenoble)*, 56(7) :1987–2092, 2006. Numération, pavages, substitutions.
- [BCRG04] G Bernot, J-P Comet, A Richard, and J Guespin. A fruitful application of formal methods to biological regulatory networks : Extending thomas’ asynchronous logical approach with temporal logic. *Journal of Theoretical Biology*, 229(3) :339–347, 2004.
- [BD02] M Barge and B Diamond. Coincidence for substitutions of Pisot type. *Bull. Soc. Math. Fr.*, 130(4) :619–626, 2002.
- [BDV05] C Bonatti, LJ Díaz, and M Viana. *Dynamics beyond uniform hyperbolicity*, volume 102 of *Encyclopaedia of Mathematical Sciences*. Springer-Verlag, Berlin, 2005. A global geometric and probabilistic perspective, Mathematical Physics, III.
- [BEIR07] V Berthé, H Ei, S Ito, and H Rao. On substitution invariant Sturmian words : an application of Rauzy fractals. *Theor. Inform. Appl.*, 41(3) :329–349, 2007.
- [Ber77] A Bertrand. Développements en base de Pisot et répartition modulo 1. *C. R. Acad. Sci. Paris Sér. A-B*, 285(6) :A419–A421, 1977.
- [Ber99] A Bertrand. Codage des endomorphismes de Pisot du tore  $[0, 1]^r$  et mesures simultanément invariantes pour deux homomorphismes du tore. *Math. Z.*, 231 :369–381, 1999.
- [Berar] J Bernat. Arithmetic automaton for perron numbers. *Discrete Mathematics and Theoretical Computer Science*, to appear.

- 
- [BF08] V Berthé and T Fernique. Brun expansions of stepped surfaces. preprint, 2008.
- [BFH97] M Bestvina, M Feighn, and M Handel. Laminations, trees, and irreducible automorphisms of free groups. *GAF*, 7 :215–244, 1997.
- [BFH00] M Bestvina, M Feighn, and M Handel. The Tits alternative for  $\text{Out}(F_n)$ , I : Dynamics of exponentially growing automorphisms. *Ann. Math.*, 151 :517–623, 2000.
- [BFZ05] V Berthé, S Ferenczi, and LQ Zamboni. Interactions between dynamics, arithmetics and combinatorics : the good, the bad, and the ugly. In *Algebraic and topological dynamics*, volume 385 of *Contemp. Math.*, pages 333–364. Amer. Math. Soc., Providence, RI, 2005.
- [BG97] J Barrow-Green. *Poincaré and the three body problem*, volume 11 of *History of Mathematics*. American Mathematical Society, Providence, RI, 1997.
- [BK05] M Barge and J Kwapisz. Elements of the theory of unimodular Pisot substitutions with an application to  $\beta$ -shifts. In *Algebraic and topological dynamics*, volume 385 of *Contemp. Math.*, pages 89–99. Amer. Math. Soc., 2005.
- [BK06] M Barge and J Kwapisz. Geometric theory of unimodular Pisot substitutions. *Amer. J. Math.*, 128(5) :1219–1282, 2006.
- [BKS91] T (ed.) Bedford, M (ed.) Keane, and C (ed.) Series. *Ergodic theory, symbolic dynamics, and hyperbolic spaces. Lectures given at the workshop "Hyperbolic geometry and ergodic theory", held at the International Centre for Theoretical Physics in Trieste, Italy, 17-28 April, 1989*. Oxford University Press, 1991.
- [Bla] F Blanchard. Topological chaos : what may this mean? arXiv :0805.0232v1.
- [Bla89] F Blanchard.  $\beta$ -expansions and symbolic dynamics. *Theoret. Comput. Sci.*, 65 :131–141, 1989.
- [BLC<sup>+</sup>04] S Barnouin, F Lassere, M Cantiello, H Guillou, T Pineau, and P Martin. A kinetic view of coordinate modulations of gene expression and metabolism in wild-type and ppara<sup>-/-</sup> mice during fasting. In *Conference of the Paul Hamel Institute, Monaco*, 2004.
- [BM86] A Bertrand-Mathis. Développement en base  $\theta$  ; répartition modulo un de la suite  $(x\theta^n)_{n \geq 0}$  ; langages codés et  $\theta$ -shift. *Bull. Soc. Math. France*, 114(3) :271–323, 1986.
- [BMO<sup>+</sup>01] CH Bryant, SH Muggleton, SG Olivier, DB Kell, P Reiser, and RD King. Combining inductive logic programming, active learning and robotics to discover the function of genes. *Electronic Transaction in Artificial Intelligence*, 5 :1–36, 2001.
- [BMOH97] BM Bakker, PAM Michels, FR Opperdoes, and Westerhoof HV. Glycolysis in bloodstream from *trypanosoma brucei* can be understood in terms of the kinetics of the glycolytic enzymes. *J. Biol. Chem.*, 272 :3207–3215, 1997.
- [Bor09] E Borel. Les probabilités d’énombrables et leurs applications arithmétiques. *Rend. Circ. Mat. Palermo*, 1909.
- [Bor50] E Borel. Sur les chiffres décimaux de  $\sqrt{2}$  et divers problèmes de probabilités en chaîne. *C. R. Acad. Sci. Paris*, 230 :591–593, 1950.
- [Bow75] R Bowen. *Equilibrium states and the ergodic theory of Anosov diffeomorphisms*. Springer-Verlag, Berlin, 1975. Lecture Notes in Mathematics, Vol. 470.
- [Bow78] R Bowen. Markov partitions are not smooth. *Proc. Amer. Math. Soc.*, 71(1) :130–132, 1978.
- [Boy96] DW Boyd. On beta expansions for Pisot numbers. *Math. Comput.*, 65(214) :841–860, 1996.
- [BP97] MP Béal and D Perrin. Symbolic dynamics and finite automata. In Grzegorz Rozenberg and Arto Salomaa, editors, *Handbook of Formal Languages*, volume 2, pages 463–503. Springer, 1997.
- [BRdJ<sup>+</sup>05] G Batt, D Ropers, H de Jong, J Geiselman, R Mateescu, M Page, and D Schneider. Validation of qualitative models of genetic regulatory networks by model checking : Analysis of the nutritional stress response in escherichia coli. *Bioinformatics*, 21(Suppl 1) :i19–i28, 2005.
- [Bre81] AJ Brentjes. *Multidimensional continued fraction algorithms*. Mathematisch Centrum, Amsterdam, 1981.
- [Bro96] A Broise. Transformations dilatantes de l’intervalle et théorèmes limites. *Astérisque*, 238 :1–109, 1996.
- [Bry92] RE Bryant. Symbolic boolean manipulation with ordered binary-decision diagrams. *ACM Comput. Surv.*, 24(3) :293–318, 1992.
- [BSPL03] SD Bay, J ; Shrager, A Pohorille, and P Langle. Revising regulatory networks : from expression data to linear causal models. *Journal of Biomedical Informatics*, 35(289-297), 2003.

- 
- [BT86] E Bombieri and JE Taylor. Which distributions of matter diffract? An initial investigation. *J. Physique*, 47(7, Suppl. Colloq. C3) :C3–19–C3–28, 1986. International workshop on aperiodic crystals (Les Houches, 1986).
- [BV00] V Berthé and L Vuillon. Tilings and rotations on the torus : a two-dimensional generalization of sturmian sequences. *Discrete Math.*, 223 :27–53, 2000.
- [BV03] F Boyer and A Viari. Ab initio reconstruction of metabolic pathways. *Bioinformatics*, 19(supp. 2), 2003.
- [BYBW07] G Batt, B Yordanov, C Belta, and R Weiss. Robustness analysis and tuning of synthetic gene networks. *Bioinformatics*, 23(18) :2415–2422, 2007.
- [Cal80] AP Calderon. On an inverse boundary value problem. In Soc. Brasileira de Mathematica, editor, *Seminar on Numerical Analysis and its applications to Continuum Physics*, pages 65–73, Rio de Janeiro, 1980.
- [Can03] V Canterini. Connectedness of geometric representation of substitutions of Pisot type. *Bull. Belg. Math. Soc. Simon Stevin*, 10(1) :77–89, 2003.
- [Caw91] E Cawley. Smooth Markov partitions and toral automorphisms. *Ergodic Theory Dynam. Systems*, 11(4) :633–651, 1991.
- [CB88] A Casson and S Bleiler. *Automorphisms of surfaces after Nielsen and Thurston*, volume 9 of *London Mathematical Society Student Texts*. Cambridge University Press, Cambridge, 1988.
- [CB04] A Cornish-Bowden. *Fundamentals of Enzyme Kinetics (third edition)*. Portland Press, Londres, 2004.
- [CBJS05] A Cornish-Bowden, M Jamin, and V Saks. *Cinétique enzymatique*. EDP Sciences, 2005.
- [CCCN<sup>+</sup>04] K Chen, L Calzone, A Csikasz-Nagy, FR Cross, B Novak, and JJ Tyson. Integrative analysis of cell cycle control in budding yeast. *Molecular Biology of the Cell*, 15(8) :3841–3862, 2004.
- [CFLM06] R Coutinho, B Fernandez, R Lima, and A Meyroneinc. Discrete time piecewise affine models of genetic regulatory networks. *J. Math. Biol.*, 52(4) :524–570, 2006.
- [CFS06] L Calzone, F Fages, and S Soliman. Biocham : An environment for modeling biological systems and formalizing experimental knowledge. *Bioinformatics*, 22 :1805–1807, 2006.
- [CFZ00] J Cassaigne, S Ferenczi, and LQ Zamboni. Imbalances in Arnoux-Rauzy sequences. *Ann. Inst. Fourier*, 50(4) :1265–1276, 2000.
- [CHL06] T Coulbois, A Hilion, and M Lustig.  $\mathbb{R}$ -trees and laminations for free groups. Preprint, 2006.
- [CKR<sup>+</sup>04] MW Covert, EM Knight, JL Reed, MJ Herrgard, and BO Palsson. Integrating high-throughput and computational data elucidates bacterial networks. *Nature*, 429(6987) :92–6, 2004.
- [CL08] GR Conner and JW Lamoreaux. On the existence of universal covering spaces for metric spaces and subsets of the euclidean plane. *preprint*, 2008.
- [CM00] EB Curtis, , and JA Morrow. *Inverse problems for electrical networks*. World Scientific, 2000.
- [CMPS93] D Crisp, W Moran, A Pollington, and P Shiue. Substitution invariant cutting sequences. *J. Théor. Nombres Bordeaux*, 5(1) :123–137, 1993.
- [Coa59] CL Coates. Flow-graph solutions of linear algebraic equations. *IRE Trans. Circuit Theory*, CT-6 :170–187, 1959.
- [Cob72] A Cobham. Uniform tag sequences. *Math. Systems Theory*, 6 :164–192, 1972.
- [CP02] MW Covert and BO Palsson. Transcriptional regulation in constraints-based metabolic models of escherichia coli. *J. biol. chem.*, 277(31) :28058–28064, 2002.
- [CRCD<sup>+</sup>04] N Chabrier-Rivier, M Chiaverini, V Danos, F Fages, and V Schächter. Modeling and querying biomolecular interaction networks. *Theor. Comput. Sci.*, 325(1) :25–44, 2004.
- [CSR93] T Chevalier, T Schreiber, and J Ross. Toward a systematic determination of complex reaction mechanisms. *J. Phys. Chem.*, 97 :6776 – 6787, 1993.
- [Dev86] RL Devaney. *An introduction to chaotic dynamical systems*. The Benjamin/Cummings Publishing Co. Inc., Menlo Park, CA, 1986.
- [DGV<sup>+</sup>99] PJ Deschavanne, A Giron, J Vilain, G Fagot, and B Fertil. Genomic signature : characterization and classification of species assessed by chaos game representation of sequences. *Molecular Biology and Evolution*, 16 :1391–1399, 1999.
- [DHS99] F Durand, B Host, and C Skau. Substitutional dynamical systems, Bratteli diagrams and dimension groups. *Ergodic Theory Dynam. Systems*, 19(4) :953–993, 1999.
- [DJ02] H De Jong. Modeling and simulation of genetic regulatory systems : A literature review. *Journal of Computational Biology*, 9(1) :67–103, 2002.

- 
- [dJGHP03] H de Jong, J Geiselmann, Céline Hernandez, and Michel Page. Genetic network analyzer : qualitative simulation of genetic regulatory networks. *Bioinformatics*, 19(3) :336–344, 2003.
- [dJRCT05] H de Jong, D Ropers, C Chaouiya, and D Thieffry. Modélisation, analyse et simulation de réseaux de régulation génique. *Biofutur*, 252(36-40), 2005.
- [dKB82] J de Kleer and JS Brown. Foundations of envisioning. In *Proceedings of the National Conference of the American Association for Artificial Intelligence*, 1982.
- [dKB84] J de Kleer and JS Brown. A qualitative physics based on confluences. In *Qualitative Reasoning about Physical Systems*. Elsevier Science Publishers, 1984.
- [DKS96] K Dajani, C Kraaikamp, and B Solomyak. The natural extension of the  $\beta$ -transformation. *Acta Math. Hungar.*, 73(1-2) :97–109, 1996.
- [DL04] V Danos and C Laneve. Formal molecular biology. *Theoretical Computer Science*, 325(1) :69 – 110, 2004.
- [Dor88] JL Dormoy. Controlling qualitative resolution. In *Proceedings of the seventh National Conference on Artificial Intelligence, AAAI88', Saint-Paul, Minn.*, 1988.
- [DPLR72] E Dubois and R Paysan-Le Roux. Approximations simultanées dans un corps de séries formelles. (Simultaneous approximation in a formal power series field). *C. R. Acad. Sci., Paris, Sér. A*, 274 : 437–440, 1972.
- [DR04] O Delgrange and E Rivals. Star : an algorithm to search for tandem approximate repeats. *Bioinformatics*, 20(16) :2812–20, 2004.
- [DT93] JM Dumont and A Thomas. Digital sum moments and substitutions. *Acta Arith.*, 64 :205–225, 1993.
- [DWFS99] P D’haeseleer, X Wen, S Fuhrman, and R Somogyi. Linear modeling of mrna expression levels during cns development and injury. In *Pacific Symposium on Biocomputing*, pages 41–52, 1999.
- [ECB98] R Eisenthal and A Cornish-Bowden. Propsects for antiparasitic drugs : the case of *trypanosoma brucei*, the causative agent of african sleeping sickness. *J. Biol. Chem.*, 272 :5500–5505, 1998.
- [EI05] H Ei and S Ito. Tilings from some non-irreducible, Pisot substitutions. *Discrete Math. Theor. Comput. Sci.*, 7(1) :81–121, 2005.
- [EIR06] H Ei, S Ito, and H Rao. Atomic surfaces, tilings and coincidences II. reducible case. *Ann. Inst. Fourier*, 56 :2285–2313, 2006.
- [EK98] K Eda and K Kawamura. The fundamental groups of one-dimensional spaces. *Topology Appl.*, 87(3) :163–172, 1998.
- [EP00] JS Edwards and BO Palsson. The escherichia coli mg1655 in silico metabolic genotype : its definition, characteristics, and capabilities. *Proc Natl Acad Sci U S A*, 97(10) :5528–33, May 2000.
- [ES97] M Einsiedler and K Schmidt. Markov partitions and homoclinic points of algebraic  $\mathbf{Z}^d$ -actions. *Tr. Mat. Inst. Steklova*, 216(Din. Sist. i Smezhnye Vopr.) :265–284, 1997.
- [Far06a] E Farcot. Geometric properties of a class of piecewise affine biological network models. *J. Math. Biol.*, 52(3) :373–418, 2006.
- [Far06b] E Farcot. Symbolic numeric analysis of attractors in randomly generated piecewise affine models of gene networks. In *ISSAC 2006*, pages 79–86, New York, 2006. ACM.
- [Fel97] D Fell. *Understanding the Control of Metabolism*. Portland Press, 1997.
- [Fer05] T Fernique. Bidimensional sturmian sequences and substitutions. In *DLT’05*, volume Lecture Notes in Computer Science 3572, pages 236–247, 2005.
- [Fer08] T Fernique. Generation and recognition of digital planes using multi-dimensional continued fractions. In *DGCI’08 : International Conference on Discrete Geometry for Computer Imagery*, 2008.
- [FFIW06] DJ Feng, M Furukado, S Ito, and J Wu. Pisot substitutions and the Hausdorff dimension of boundaries of atomic surfaces. *Tsukuba J. Math.*, 30(1) :195–223, 2006.
- [FHT<sup>+</sup>07] JJ Faith, B Hayete, JT Thaden, I Mogno, J Wierzbowski, and al.’ Large-scale mapping and validation of Escherichia coli transcriptional regulation from a compendium of expression profiles. *PLoS Biol.*, 5(1) :e8, 2007.
- [Fie02] O Fiehn. Metabolomics : the link between genotypes and phenotypes. *Plant Molecular Biology*, 48(1-2) :155–171, 2002.
- [FIKO96] T Fujita, S Ito, M Keane, and M Ohtsuki. On almost everywhere exponential convergence of the modified Jacobi-Perron algorithm : A corrected proof. *Ergodic Theory Dyn. Syst.*, 16(6) :1345–1352, 1996.



- 
- [FMS<sup>+</sup>07] F Ferrazzi, P Magni, L Sacchi, A Nuzzo, U Petrovic, and R Bellazzi. Inferring gene regulatory networks by integrating static and dynamic data. *Int J Med Inform*, Epub 2007 Sept 6, 2007.
- [Fog02] N. Pytheas Fogg. *Substitutions in dynamics, arithmetics and combinatorics*, volume 1794 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2002. Edited by V. Berthé, S. Ferenczi, C. Mauduit and A. Siegel.
- [FP54] E Frank and O Perron. Remark on a certain class of continued fractions. *Proc. Am. Math. Soc.*, 5 :270–283, 1954.
- [Fri04] N Friedman. Inferring cellular networks using probabilistic graphical models. *Science*, 303 :799–805, 2004.
- [FS92] C Frougny and B Solomyak. Finite beta-expansions. *Ergodic Theory Dynamical Systems*, 12 :45–82, 1992.
- [FS08] F Fages and S Soliman. From reaction models to influence graphs and back : a theorem. In *Formal Methods in Systems Biology*, Lectures Notes in Bio-Informatics, Springer-Verlag, 2008.
- [GA97] D Giammaressi and Restivo A. *Two-dimensional languages*. Springer-Verlag, Berlin, 1997.
- [GA04] R Ghoshn and C Andomlin. Symbolic reachable set computation of piecewise affine hybrid automata and its application to biological modelling : Delta-notch protein signalling. *Systems Biology*, 1(1) :170–183, 2004.
- [GBBK02] N Guelzim, S Bottani, P Bourguine, and F Képès. Topological and causal structure of the yeast transcriptional regulatory network. *Nature Genetics*, 31 :60–63, 2002.
- [GBC<sup>+</sup>04] J Guespin, G Bernot, J-P Comet, A Mérieau, A Richard, C Hulen, and B Polack. Epigenesis and dynamic similarity in two regulatory networks in pseudomonas aeruginosa . *Acta Biotheoretica*, 52(4) :379–390, 2004.
- [GCT08] A Gonzalez, C Chaouiya, and D Thieffry. Logical modelling of the role of the hh pathway in the patterning of the drosophila wing disc. *Bioinformatics*, to appear, 2008.
- [GDZ<sup>+</sup>06] A Ghazalpour, S Doss, B Zhang, S Wang, C Plaisier, Ruth Castellanos, Alec Brozell, Eric E Schadt, Thomas A Drake, Aldons J Lusic, and Steve Horvath. Integrating genetic and network analysis to characterize genes related to mouse weight. *PLoS Genet*, 2(8) :e130, 2006.
- [GHVdL07] B Gaujal, A Hordijk, and D Van der Laan. On the optimal open-loop control policy for deterministic and exponential polling systems. *Probability in Engineering and Informational Sciences*, 21 :157–187, 2007.
- [GKNS07] M Gebser, B Kaufmann, A Neumann, and T Schaub. clasp : A conflict-driven answer set solver. In *Proceedings of the Ninth International Conference on Logic Programming and Nonmonotonic Reasoning (LPNMR'07)*, pages 260–265. Springer, 2007.
- [Goo63] BC Goodwin. *Temporal organization in cells*. Academic Press, 1963.
- [Goo65] BC Goodwin. Oscillatory behavior in enzymatic control processes. In G. Weber, editor, *Advances in Enzyme Regulation*, pages 425–438. Pregamon Press, Oxford, 1965.
- [Gou98] JL Gouzé. Positive and negative circuits in dynamical systems. *J.Biol.Syst.*, 6 :11–15, 1998.
- [GR07] A Gorban and O Radulescu. Dynamical robustness of biological networks with hierarchical distribution of time scales. *IET Systems Biology*, 1(238-246), 2007.
- [Gra92] DJ Grabiner. Farey nets and multidimensional continued fractions. *Monatsh. Math.*, 114(1) :35–61, 1992.
- [Grm01] M Grmela. Complex fluids subjected to external influences. *J. Non-Newtonian Fl. Mech.*, 96 :221–254, 2001.
- [GRRL<sup>+</sup>03] RM Gutierrez-Rios, DA Rosenblueth, JA Loza, AM Huerta, JD Glasner, FR Blattner, and J Collado-Vides. Regulatory network of Escherichia coli : consistency between literature knowledge and microarray profiles. *Genome Res*, 13(11) :2435–2443, 2003.
- [GS98] C Goodman-Strauss. Matching rules and substitution tilings. *Ann. of Math. (2)*, 147(1) :181–223, 1998.
- [GS02] JL Gouzé and T Sari. A class of piecewise linear differential equations arising in biological models. *Dynamical Systems*, 17(4), 2002.
- [GST<sup>+</sup>08] M Gebser, T Schaub, S Thiele, B Usadel, and P Veber. Detecting inconsistencies in large influence networks with answer set programming. In *Workshop on Constraint Based Methods for Bioinformatics*, 2008.
- [GVG04] JP Gazeau and JL Verger-Gaugry. Geometric study of the beta-integers for a Perron number and mathematical quasicrystals. *J. Théor. Nombres Bordeaux*, 16(1) :125–149, 2004.

- 
- [GW02] B Grunfelder and EA Winzeler. Treasures and traps in genome-wide data sets : case examples from yeast. *Nat Rev Genet*, 3(9) :653–661, Sep 2002.
- [Had98] J Hadamard. Les surfaces à courbures opposées et leurs lignes géodésiques. *Journal de Mathématiques Pures & Appliquées*, 4, 1898.
- [Han00] CW Hansen. *Dynamics of multi-dimensional substitutions*. PhD thesis, George Washington University, 2000.
- [HBC<sup>+</sup>01] EL Hong, R Balakrishnan, KR Christie, MC Costanzo, SS Dwight, and *al.* Saccharomyces genome database. <http://www.yeastgenome.org/>, 2001.
- [HCP03] MJ Herrgard, MW Covert, and BØ Palsson. Reconciling gene expression data with known genome-scale regulatory network structures. *Genome Res*, 13(11) :2423–34, Nov 2003.
- [HGH01] J Hartman, G Garvik, , and L Hartwell. Principles for the buffering of genetic variation. *Science*, 291 :1001–1004, 2001.
- [HGL<sup>+</sup>04] CT Harbison, DB Gordon, TI Lee, NJ Rinaldi, K Macisaac, and *al.* Transcriptional regulatory code of a eukaryotic genome. *Nature*, 431(7004) :99–104, 2004.
- [HI97] M Hama and T Imahashi. Periodic  $\beta$ -expansions for certain classes of Pisot numbers. *Comment. Math. Univ. St. Paul.*, 46(2) :103–116, 1997.
- [HKG] JW Helton, I Klep, and R Gomez. Determinant expansions of signed matrices and of certain jacobians. arXiv :0802.4319v1.
- [HLPP06] MJ Herrgard, BS Lee, V Portnoy, and BO Palsson. Integrated analysis of regulatory and metabolic networks reveals novel regulatory mechanisms in Saccharomyces cerevisiae. *Genome Res*, 16(5) :627–35, 2006.
- [HM06] P Hubert and A Messaoudi. Best simultaneous Diophantine approximations of Pisot numbers and Rauzy fractals. *Acta Arith.*, 124(1) :1–15, 2006.
- [HMJ<sup>+</sup>00] T Hughes, M Marton, A Jones, C Roberts, R Stoughton, and *al.* Functional discovery via a compendium of expression profiles. *Cell*, 102(1) :109–126, 2000.
- [HMPL<sup>+</sup>04] H Hermjakob, L Montecchi-Palazzi, C Lewington, S Mudali, S Kerrien, and *al.* IntAct : an open source molecular interaction database. *Nucleic Acids Res*, 32(Database issue) :D452–5, 2004.
- [Hol96] M Hollander. *Linear numeration systems, finite beta expansions, and discrete spectrum of substitution dynamical systems*. PhD thesis, Washington University, 1996.
- [Hol04] JR Holton. *An Introduction to Dynamic Meteorology*. Academic Press, 2004.
- [HP05] D Hinrichsen and AJ Pritchard. *Mathematical systems theory. I*, volume 48 of *Texts in Applied Mathematics*. Springer-Verlag, Berlin, 2005. Modelling, state space analysis, stability and robustness.
- [HR74] R Heinrich and TA Rapoport. A linear steady-state treatment of enzymatic chains. general properties, and effector strength. *Eur. J. Biochem.*, 42 :89–95, 1974.
- [HS96] R Heinrich and S Schuster. *The Regulation of Cellular Systems*. Chapman and Hall, 1996.
- [HSG<sup>+</sup>06] S Hoops, S Sahle, R Gauges, C Lee, J Pahle, N Simus, M Singhal, L Xu, P Mendes, and U Kummer. Copasi—a complex pathway simulator. *Bioinformatics*, 22(24) :3067–74, 2006.
- [Hén06] T Hénin. Représentation par jeux du chaos de séquences d’adn. Master’s thesis, IRISA, Université de Rennes 1, 2006. présenté à "Algorithmique, combinatoire du texte et applications en bio-informatique", Chessy France.
- [Ide04] T Ideker. A systems approach to discovering signaling and regulatory pathways—or, how to digest large interaction networks into relevant pieces. *Adv Exp Med Biol*, 547(NIL) :21–30, 2004.
- [IFHY03a] S Ito, J Fujii, H Higashino, and S Yasutomi. On simultaneous approximation to  $(\alpha, \alpha^2)$  with  $\alpha^3 + k\alpha - 1 = 0$ . *J. Number Theory*, 99(2) :255–283, 2003.
- [IFHY03b] S Ito, J Fujii, H Higashino, and SI Yasutomi. On simultaneous approximation to  $(\alpha, \alpha^2)$  with  $\alpha^3 + k\alpha - 1 = 0$ . *J. Number Theory*, 99(2) :255–283, 2003.
- [IGH01] T Ideker, T Galitski, and L Hood. A new approach to decoding life : systems biology. *Annu Rev Genomics Hum Genet.*, 2 :343–72, 2001.
- [IK91] S Ito and M Kimura. On Rauzy fractal. *Japan J. Indust. Appl. Math.*, 8(3) :461–486, 1991.
- [IK02] M Iosifescu and C Kraaikamp. *Metrical theory of continued fractions*. Mathematics and its Applications, Kluwer Academic Publishers, 2002.
- [IO93] S Ito and M Ohtsuki. Modified Jacobi-Perron algorithm and generating Markov partitions for special hyperbolic toral automorphisms. *Tokyo J. Math.*, 16(2) :441–472, 1993.

- [IO94] S Ito and M Ohtsuki. Parallelogram tilings and Jacobi-Perron algorithm. *Tokyo J. Math.*, 17(1) :33–58, 1994.
- [IR05] S Ito and H Rao. Purely periodic  $\beta$ -expansion with Pisot base. *Proc. Amer. Math. Soc.*, 133 :953–964, 2005.
- [IR06] S Ito and H Rao. Atomic surfaces, tilings and coincidences I. Irreducible case. *Israel J. Math.*, 153 :129–155, 2006.
- [IS98] Ingenuity-Systems. Ingenuity pathways knowledge base. Available : [www.ingenuity.com](http://www.ingenuity.com), 1998.
- [IY07] S Ito and SI Yasutomi. On simultaneous diophantine approximation to periodic points related to modified Jacobi-Perron algorithm. *Advanced Studies in Pure Mathematics*, 49 :171–184, 2007.
- [Jac68] CGJ Jacobi. General theory of continued-fraction-like algorithms in which each number is formed from three previous ones. (From the posthumous papers communicated by E. Heine.). (Allgemeine Theorie der kettenbruchähnlichen Algorithmen, in welche jede Zahl aus drei vorhergehenden gebildet wird.). *Borchardt J.*, LXIX : 29–64, 1868.
- [Jam04] D Jamet. On the language of discrete planes and surfaces. In *Proceedings of the Tenth International Workshop on Combinatorial Image Analysis*, pages 227–241. Springer-Verlag, 2004.
- [JDSZ05] P Juvan, J Demsar, G Shaunlsky, and B Zupan. Genepath : from mutations to genetic networks and back. *Nucleic Acids Research*, 33(Web Server issue) :W749–W752, 2005.
- [Jef90] J Jeffrey. Chaos game representation of gene structure. *Nucleic Acids Research*, 18(8) :2163–2170, 1990.
- [JM61] F Jacob and J Monod. Genetic regulatory mechanisms in the synthesis of proteins. *J Mol Biol.*, 3 :318–56, 1961.
- [JP06] Andrew R Joyce and Bernhard O Palsson. The model organism as a system : integrating 'omics' data sets. *Nat Rev Mol Cell Biol*, 7(3) :198–210, 2006.
- [Jum04] DB Jump. Fatty acid regulation of gene transcription. *Crit. Rev. Clin. Lab. Sci.*, 41(1) :41–78, 2004.
- [JVEGS02] R Jansen, JDA Van Embden, W Gaastra, and LM Schouls. Identification of genes that are associated with dna repeats in prokaryotes. *Mol Microbiol*, 43(6) :1565–1575, 2002.
- [K06] Davar K. Normal numbers are normal. *Clay Mathematics Institute Annual Report 2006*, pages 15 ;27–31, 2006.
- [Kal03] E Kalnay. *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge University Press, 2003.
- [Kan06] K Kaneko. *An Introduction to Complex Systems Biology*. Springer-Verlag, 2006.
- [Kau69] SA Kauffman. Metabolic stability and epigenesis in randomly constructed genetic nets. *Journal of Theoretical Biology*, 22 :437–467, 1969.
- [Kau93] SA Kauffman. *The origins of Order : Self-Organization and Selection in Evolution*. Oxford University Press, 1993.
- [KB73] H Kacser and JA Burns. The control of flux. *Symp. Soc. Exp. Biol.*, 27 :65–104, 1973.
- [Kea75] M Keane. Interval exchange transformations. *Math. Z.*, 141 :25–31, 1975.
- [KEBC05] M Kaern, TA Elston, WJ Blake, and JJ Collins. Stochasticity in gene expression : from theories to phenotypes. *Nature Rev.Genet.*, 6 :451–464, 2005.
- [Kel04] DB Kell. Metabolomics and systems biology : making sense of the soup. *Current Opinion in Microbiology*, 7(3) :296–307, 2004.
- [KGC05a] RD King, SM Garrett, and GM Coghill. On the use of qualitative reasoning to simulate and identify metabolic pathways. *Bioinformatics*, 21(9) :2017–26, 2005.
- [KGC05b] RD King, SM Garrett, and GM Coghill. On the use of qualitative reasoning to simulate and identify metabolic pathways. *Bioinformatics*, 21(9) :2017–2026, 2005.
- [K GK<sup>+</sup>04] M Kanehisa, S Goto, S Kawashima, Y Okuno, and M Hattori. The KEGG resource for deciphering the genome. *Nucleic Acids Res*, 32(Database issue) :D277–80, 2004.
- [Khi63] AY Khintchine. *Continued fractions*. P. Noordhoff, Ltd., Groningen, 1963.
- [Kit01] H Kitano, editor. *Foundations of Systems Biology*. MIT Press, 2001.
- [Kit02a] H Kitano. Systems biology : A brief overview. *Science*, 295(5560) :1662 – 1664, 2002.
- [Kit02b] H Kitano. Systems biology : A brief overview. *Science*, 295(5560) :1662–1664, Mar 2002.
- [KKB<sup>+</sup>02] BN Kholodenko, A Kiyatkin, FJ Bruggeman, E Sontag, HV Westerhoff, and JB Hoek. Untangling the wires : a strategy to trace functional interactions in signaling and gene networks. *PNAS*, 99(20) :12841–12846, 2002.

- 
- [KP00] J Kellendonk and I Putnam. Tilings,  $C^*$ -algebras, and  $K$ -theory. In M. Baake et al., editor, *Directions in mathematical quasicrystals*, volume 13 of *AMS CRM Monogr. Ser.*, pages 177–206, Providence, RI, 2000.
- [KPBT03] S Kauffman, C Peterson, Samuelsson B, and C Troein. Random boolean network models and the yeast transcriptional network. *PNAS*, 100 :14796–14799, 2003.
- [KPE03] K Kauffman, P Prakash, and J Edwards. Advances in flux balance analysis. *Current Opinion in Biotechnology*, 2003.
- [KPST03] S Kauffman, C Peterson, B Samuelsson, and C Troein. Random boolean network models and the yeast transcriptional network. *PNAS*, 100(25) :14796–14799, 2003.
- [KTH98] R Kubo, M Toda, and N Hashitsume. *Statistical Physics II. Nonequilibrium Statistical Mechanics*. Springer, 1998.
- [Kui84] BJ Kuipers. Commonsense reasoning about causality : Deriving behaviour from structure. In *Qualitative Reasoning about Physical Systems*. Elsevier Science Publishers, 1984.
- [Kur68] K Kuratowski. *Topology. Vol. II*. New edition, revised and augmented. Translated from the French by A. Kirkor. Academic Press, New York, 1968.
- [KV98] R Kenyon and A Vershik. Arithmetic construction of sofic partitions of hyperbolic toral automorphisms. *Ergodic Theory Dynam. Systems*, 18(2) :357–372, 1998.
- [Lag94] JC Lagarias. Geodesic multidimensional continued fractions. *Proc. Lond. Math. Soc., III. Ser.*, 69(3) :464–488, 1994.
- [Lan94] S Lang. *Algebraic number theory*, volume 110 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1994.
- [Lap86] PS Laplace. *Essai philosophique sur les probabilités*. Collection Epistémè. [Epistémè Collection]. Christian Bourgeois Éditeur, Paris, fifth edition, 1986.
- [Lar88] RG Larson. *Constitutive Equations for Polymer Melts and Solutions*. Butterworths : Boston, 1988.
- [LB06] Abdelhalim Larhlimi and Alexander Bockmayr. A new approach to flux coupling analysis of metabolic networks. In *Computational Life Sciences II, CompLife'06*, LNBI, number = 4216, pages 205–215. Springer-Verlag, 2006.
- [LBV07] M Le Borgne and P Veber. Decision diagrams for qualitative biological models. Technical Report RR-6182, INRIA, 2007.
- [LCL<sup>+</sup>04] SS Lee, WY Chan, CK Lo, DC Wan, DS Tsang, and WT Cheung. Requirement of ppar $\alpha$  in maintaining phospholipid and triacylglycerol homeostasis during energy deprivation. *J Lipid Res.*, 45(11) :2025–37, 2004.
- [LGJJ93] JM Luck, C Godrèche, A Janner, and T Janssen. The nature of the atomic surfaces of quasiperiodic self-similar structures. *J. Phys. A*, 26(8) :1951–1999, 1993.
- [Liv87] AN Livshits. On the spectra of adic transformations of Markov compact sets. *Uspekhi Mat. Nauk*, 42(3(255)) :189–190, 1987. In Russian. English translation in *Russian Math. Surveys* 42 (1987), 222–223.
- [LM95] D Lind and B Marcus. *An introduction to symbolic dynamics and coding*. Cambridge University Press, Cambridge, 1995.
- [LMS02] JY Lee, RV Moody, and B Solomyak. Pure point dynamical and diffraction spectra. *Ann. Henri Poincaré*, 3(5) :1003–1018, 2002.
- [Lot05] M Lothaire. *Applied combinatorics on words*, volume 105 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, 2005.
- [LPCal06] S Lagarrigue, F Pitel, W Carré, and al., Mapping quantitative trait loci affecting fatness and breast muscle weight in meat-type chicken lines divergently selected on abdominal fatness. *Genet. Sel. Evol.*, 38 :85–97, 2006.
- [LRR<sup>+</sup>02] TI Lee, NJ Rinaldi, F Robert, DT Odom, Z Bar-Joseph, and al., Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science*, 298(5594) :799–804, 2002.
- [LRT02] J Luo, H Rao, and B Tan. Topological structure of self-similar sets. *Fractals*, 10 :223–227, 2002.
- [LS05] E Lindenstrauss and K Schmidt. Symbolic representations of nonexpansive group automorphisms. *Israel J. Math.*, 149 :227–266, 2005.
- [LT06] J Luo and JM Thuswaldner. On the fundamental group of self-affine plane tiles. *Ann. Inst. Fourier (Grenoble)*, 56(7) :2493–2524, 2006. Numération, pavages, substitutions.
- [LU06] R Lima and E Ugalde. Dynamical complexity of discrete-time regulatory networks. *Non-linearity*, 19(1) :237–259, 2006.
- [Luo02] J Luo. A note on a self-similar tiling generated by the minimal Pisot number. *Fractals*, 10(3) :335–339, 2002.
- [LW96] JC Lagarias and Y Wang. Self affine tiles in  $\mathbb{R}^n$ . *Adv. Math.*, 121 :21–49, 1996.

- 
- [LW00] D Lockhart and E Winzeler. Genomics, gene expression and dna arrays. *Nature*, 405 :827–836, 2000.
- [LZ04] J Luo and ZL Zhou. Disk-like tiles derived from complex bases. *Acta Math. Sin. (Engl. Ser.)*, 20(4) :731–738, 2004.
- [Mac70] AGJ MacFarlane. *Dynamical Systems Models*. Harap : London, 1970.
- [Mas53] SJ Mason. Feedback theory - some properties of signal flow graphs. *Proc. IRE*, 41 :1144–1156, 1953.
- [Mes98] A Messaoudi. Propriétés arithmétiques et dynamiques du fractal de Rauzy. *J. Théor. Nombres Bordeaux*, 10(1) :135–162, 1998.
- [Mes00] A Messaoudi. Frontière du fractal de Rauzy et système de numération complexe. *Acta Arith.*, 95(3) :195–224, 2000.
- [Mes06] A Messaoudi. Propriétés arithmétiques et topologiques d’une classe d’ensembles fractales. *Acta Arith.*, 121(4) :341–366, 2006.
- [MH38] M Morse and GA Hedlund. Symbolic Dynamics. *Amer. J. Math.*, 60(4) :815–866, 1938.
- [MH40] M Morse and GA Hedlund. Symbolic dynamics II. Sturmian trajectories. *Amer. J. Math.*, 62 :1–42, 1940.
- [Moo00] RV Moody. Model sets : a survey. In F. Axel and eds J.-P. Gazeau, editors, *From Quasicrystals to More Complex Systems*, pages 145–166. Les Editions de Physique, Springer-Verlag, Berlin, 2000.
- [Mor21] HM Morse. Recurrent geodesics on a surface of negative curvature. *Trans. Amer. Math. Soc.*, 22(1) :84–100, 1921.
- [Mos96] B Mossé. Reconnaissabilité des substitutions et complexité des suites automatiques. *Bull. Soc. Math. France*, 124(2) :329–346, 1996.
- [Moz89] S Mozes. Tilings, substitution systems and dynamical systems generated by them. *J. Anal. Math.*, 53 :139–186, 1989.
- [MRM<sup>+</sup>08] PT Monteiro, D Ropers, R Mateescu, AT Freitas, and H de Jong. Temporal logic patterns for querying dynamic models of cellular interaction networks. *Bioinformatics*, to appear, 2008.
- [MW88] RD Mauldin and SC Williams. Hausdorff dimension in graph directed constructions. *Trans. Amer. Math. Soc.*, 309(2) :811–829, 1988.
- [MWG<sup>+</sup>06] KD MacIsaac, T Wang, DB Gordon, DK Gifford, G Stormo, and E Fraenkel. An improved map of conserved regulatory sites for *Saccharomyces cerevisiae*. *BMC Bioinformatics*, 7(NIL) :113, 2006.
- [NAT<sup>+</sup>04] S Nicolay, F Argoul, M Touchon, Y d’Aubenton Carafa, C Thermes, and A Arneodo. Low frequency rhythms in human dna sequences : a key to the organization of gene location and orientation? *Phys. Rev. Lett.*, 93 :108101, 2004.
- [Nic08] J Nicolas. *Habitation à Diriger des Recherches : Syntaxe et Raisonnement sur les génomes*. PhD thesis, Université de Rennes 1, 2008.
- [Nie99] I Niemela. Logic programs with stable model semantics as a constraint programming paradigm. *Annals of Mathematics and Artificial Intelligence*, 25 :241–273, 1999.
- [NIE<sup>+</sup>06] DE Nelson, AE Ihekweba, M Elliot, JR Johnson, CA Gibnet, Foreman BE, and al. Oscillations in nf- $\kappa$ b signaling control the dynamics of gene expression. *Science*, 22(5696) :704–708, 306.
- [NN03a] MT Nakamura and TY Nara. Essential fatty acid synthesis and its regulation. *Prostaglandins, Leukotrienes and Essential Fatty Acids*, 68 :145–150, 2003.
- [NN03b] SM Ngai and N Nguyen. The Highway dragon revisited. *Discrete Comput. Geom.*, 29(4) :603–623, 2003.
- [Nog95] A Nogueira. The three-dimensional Poincaré continued fraction algorithm. *Isr. J. Math.*, 90(1-3) :373–401, 1995.
- [NT04] SM Ngai and TM Tang. A technique in the topology of connected self-similar tiles. *Fractals*, 12(4) :389–403, 2004.
- [NTC07] A Naldi, D Thieffry, and C Chaouiya. Decision diagrams for the representation and analysis of logical models of genetic networks. In *Conference on Molecular Systems Biology 2007*, number 4695 in LNCS/LNBI, pages 233–247. Springer-Verlag, 2007.
- [NTIM05] N Nariai, Y Tamada, S Imoto, and S Miyano. Estimating gene regulatory networks and protein-protein interactions of *Saccharomyces cerevisiae* from multiple genome-wide data. *Bioinformatics*, 21 Suppl 2(NIL) :ii206–ii212, 2005.
- [OG04] N Oiwa and J Glazier. Self-similar mitochondrial dna. *Cell Biochemistry and Biophysics*, 41 :41–62, (2004).

- 
- [OPK73] GF Oster, AS Perelson, and A Katchalsky. Network thermodynamics : dynamic modelling of biophysical systems. *Quat. Rev. Biophys.*, 6 :1–134, 1973.
- [Pal00] B Palsson. The challenges of in silico biology. *Nature Biotechnology*, 18 :1147–1150, 2000.
- [Pal06] B Palsson. *Systems Biology - Properties of Reconstructed Networks*. Cambridge University Press, 2006.
- [Par60] W Parry. On the  $\beta$ -expansions of real numbers. *Acta Math. Acad. Sci. Hungar.*, 11 :401–416, 1960.
- [Par83] T Parthasarathy. *On Global Univalence Theorems*, volume 977 of *Lecture Notes in Mathematics*. Springer, 1983.
- [PBZ<sup>+</sup>03] E Phizicky, P Bastiaens, H Zhu, M Snyder, and S Fields. Protein analysis on a proteomic scale. *Nature*, pages 208–215, 2003.
- [PF08] N Priebe-Franck. A primer on substitution tilings of euclidean space. To appear in *Expositiones Mathematicae*, 2008.
- [PG03] J Pailleux and JF Geleyn. La modélisation de l’atmosphère et la prévision numérique du temps. *Annales des Ponts et Chaussées*, 107-108 :56–62, 2003.
- [Pie01] RA Pielke. *Mesoscale Meteorological Modeling*. Academic Press, 2001.
- [PL90] P Prusinkiewicz and A Lindenmayer. *The algorithmic beauty of plants*. Springer-Verlag, 1990.
- [PM00] A Pandey and M Mann. Proteomics to study genes and genomes. *Nature*, 405 :835–846, 2000.
- [PO74] AS Perelson and GF Oster. Chemical reaction dynamics. part ii : Reaction networks. *Arch. Rat. Mech. Anal.*, 57 :31–98, 1974.
- [Pod77] EV Podsypanin. A generalization of the continued fraction algorithm that is related to the Viggo Brun algorithm. *Zap. Naučn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI)*, 67 :184–194, 1977.
- [Poi87] H Poincaré. *Les méthodes nouvelles de la mécanique céleste. Tome I. Les Grands Classiques Gauthier-Villars*. [Gauthier-Villars Great Classics]. Librairie Scientifique et Technique Albert Blanchard, Paris, 1987.
- [PPP02] JA Papin, ND Price, , and BO Palsson. Extreme pathway lengths and reaction participation in genome-scale metabolic networks. *Genome Research*, 12 :1889–1900, 2002.
- [Pra99] B Praggastis. Numeration systems and Markov partitions from self-similar tilings. *Trans. Amer. Math. Soc.*, 351(8) :3315–3349, 1999.
- [PSP<sup>+</sup>04] A Papin, J Stelling, ND Price, S Klamt, S Schuster, and BO Palsson. Comparison of network-based pathway analysis methods. *Trends Biotechnol.*, 22(8) :400–405, 2004.
- [Que87] M Queffélec. *Substitution dynamical systems—spectral analysis*. Lecture Notes in Mathematics, 1294. Springer-Verlag, Berlin, 1987.
- [Rau82] G Rauzy. Nombres algébriques et substitutions. *Bull. Soc. Math. France*, 110(2) :147–178, 1982.
- [RBB<sup>+</sup>06] T Reguly, A Breikreutz, L Boucher, BJ Breikreutz, GC Hon, CL Myers, and *al.*. Comprehensive curation and analysis of global interaction networks in *Saccharomyces cerevisiae*. *J Biol*, 5(4) :11, 2006.
- [Red88] C Reeder. Metabolism control theory : A structural approach. *J Theor Biol*, 2(135) :175–201, 1988.
- [Rev91] JP Reveillès. *Géométrie discrète, calcul en nombres entiers et algorithmique*. Thèse de Doctorat, Université Louis Pasteur, Strasbourg, 1991.
- [RKK<sup>+</sup>01] PGK Reiser, RD King, DB Kell, SH Muggleton, CH Bryant, and SG Oliver. Developing a logical model of yeast metabolism. *Electronic Transaction in Artificial Intelligence*, 5 :223–244, 2001.
- [Rob04] EA Robinson. Symbolic dynamics and tilings of  $\mathbb{R}^d$ . In *Symbolic dynamics and its applications*, volume 60 of *Proc. Sympos. Appl. Math.*, Amer. Math. Soc. Providence, RI, pages 81–119, 2004.
- [Roy03] D Roy. Approximation to real numbers by cubic algebraic integers. II. *Ann. of Math. (2)*, 158(3) :1081–1087, 2003.
- [RPS<sup>+</sup>04] A Regev, AM Panina, W Silverman, L Cardelli, and E Shapiro. Bioambients : an abstraction for biological compartments. *Theor. Comput. Sci.*, 325(1) :141–167, 2004.
- [RWA02] CV Rao, DM Wolf, and AP Arkin. Control exploitation and tolerance of intracellular noise. *Nature*, 420 :231–237, 2002.
- [RZL07] O Radulescu, A Zinovyev, and A Lilienbaum. Model reduction and model comparison for nfkb signaling. In *Foundations of Systems Biology in Engineering, FOSBE’07*, Stuttgart, Germany, 2007.

- 
- [SAI01] Y Sano, P Arnoux, and S Ito. Higher dimensional extensions of substitutions and their dual maps. *J. Anal. Math.*, 83 :183–206, 2001.
- [Sau04] U Sauer. High-throughput phenomics : experimental methods for mapping fluxomes. *Current Opinion in Biotechnology*, 15(1) :58–63, 2004.
- [SCG<sup>+</sup>01] F Schacherer, C Choi, U Gotze, M Krull, S Pistor, and E Wingender. The TRANS-PATH signal transduction database : a knowledge base on signal transduction networks. *Bioinformatics*, 17(11) :1053–7, 2001.
- [Sch80] K Schmidt. On periodic expansions of Pisot numbers and Salem numbers. *Bull. London Math. Soc.*, 12(4) :269–278, 1980.
- [Sch00a] K Schmidt. Algebraic coding of expansive group automorphisms and two-sided beta-shifts. *Monatsh. Math.*, 129(1) :37–61, 2000.
- [Sch00b] F Schweiger. *Multidimensional Continued Fraction*. Oxford Univ. Press, New York, 2000.
- [SEJGM02] B Schoeberl, C Eichler-Jonsson, ED Gilles, and G Müller. Computational modeling of the dynamics of the map kinase cascade activated by surface and internalized egf receptors. *Nature Biotechnology*, pages 370 – 375, 2002.
- [Sel78] ES Selmer. Unit fraction expansions and a multiplicative analog. *Nordisk Mat. Tidskr.*, 25-26 :91–109, 1978.
- [Sen06] M Senechal. What is... a quasicrystal? *Notices Amer. Math. Soc.*, 53(8) :886–887, 2006.
- [SFD00] S Schuster, DA Fell, and T Dandekar. A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nature Biotechnology*, 18 :326–332, 2000.
- [SGCPG<sup>+</sup>06] H Salgado, S Gama-Castro, M Peralta-Gil, E Diaz-Peredo, F Sanchez-Solano, A Santos-Zavaleta, and *al.*, 'RegulonDB (version 5.0) : Escherichia coli K-12 transcriptional regulatory network, operon organization, and growth conditions. *Nucleic Acids Res*, 34(Database issue) :D394–7, 2006.
- [Sid01] N Sidorov. Bijective and general arithmetic codings for Pisot toral automorphisms. *J. Dynam. Control Systems*, 7(4) :447–472, 2001.
- [Sid02] N Sidorov. An arithmetic group associated with a Pisot unit, and its symbolic-dynamical representation. *Acta Arith.*, 101(3) :199–213, 2002.
- [Sid03] N Sidorov. Arithmetic dynamics. In S. Bezuglyi et al., editor, *Topics in dynamics and ergodic theory*, volume 310 of *Lond. Math. Soc. Lect. Note Ser.*, pages 145–189. Cambridge University Press, 2003.
- [SKB<sup>+</sup>02] J Stelling, S Klamt, K Bettenbrock, S Schuster, , and ED Gilles. Metabolic network structure determines key aspects of functionality and regulation. *Nature*, 420 :190–193, 2002.
- [SL98] Z Szallasi and S Liang. Modeling the normal and neoplastic cell cycle with 'realistic boolean genetic networks' : their application for understanding carcinogenesis and assessing therapeutic strategies. In *Proc. Pacific Symposium on Biocomputing*, volume 3, pages 66–76, 1998.
- [Sma67] S Smale. Differentiable dynamical systems. *Bull. Amer. Math. Soc.*, 73 :747–817, 1967.
- [SMO<sup>+</sup>03] P Shannon, A Markiel, O Ozier, NS Baliga, JT Wang, and *al.*, Cytoscape : a software environment for integrated models of biomolecular interaction networks. *Genome Res.*, 13(11) :2498–504, 2003.
- [Sno89] EH Snoussi. Qualitative dynamics of piecewise-linear differential equations : a discrete mapping approach. *Dyn. Stab. Syst.*, 4(3-4) :189–207, 1989.
- [Sno98] EH Snoussi. Necessary conditions for multistationarity and stable periodicity. *J. Biol. Syst.*, 6(3-9), 1998.
- [Sol92] B Solomyak. Substitutions, adic transformations, and beta-expansions. In *Symbolic dynamics and its applications (New Haven, CT, 1991)*, volume 135 of *Contemp. Math.*, pages 361–372. Amer. Math. Soc., Providence, RI, 1992.
- [Sol94] B Solomyak. Conjugates of beta-numbers and the zero-free domain for a class of analytic functions. *Proc. Lond. Math. Soc.*, III. Ser., 68(3) :477–498, 1994.
- [Sol06] B Solomyak. Tilings and dynamics. In *EMS Summer School on Combinatorics, Automata and Number Theory*, Liege, 2006.
- [Son07] ED Sontag. Monotone and near-monotone biochemical networks. *Systems and Synthetic Biology*, 1 :59–87, 2007.
- [Sou03] C Soulé. Graphic requirements for multistationarity. *ComplexUs*, 1 :123–133, 2003.
- [Sou06] C Soulé. Mathematical approaches to differentiation and gene regulation. *C.R. Paris Biologies*, 329, 2006.
- [SRN07] JA Sepulchre, S Reverchon, and W Nasser. Modeling the onset of virulence in a pectinolytic bacterium. *J. of Theor. Biol.*, 244 :239–257, 2007.

- 
- [SSR<sup>+</sup>03] E Segal, M Shapira, A Regev, D Pe'er, D Botstein, and *al.* Module networks : identifying regulatory modules and their condition-specific regulators from gene expression data. *Nat Genet.* 34(2) :166–76, 2003.
- [ST02] K Scheicher and JM Thuswaldner. Canonical number systems, counting automata and fractals. *Math. Proc. Cambridge Philos. Soc.*, 133(1) :163–182, 2002.
- [STV04] B Scholkopf, K Tsuda, and J-P Vert, editors. *Kernel Methods in Computational Biology*. MIT Press, 2004.
- [SV98] N Sidorov and A Vershik. Bijective arithmetic codings of hyperbolic automorphisms of the 2-torus, and binary quadratic forms. *J. Dynam. Control Systems*, 4(3) :365–399, 1998.
- [SVC02] D Segrè, D Vitkup, and GM Church. Analysis of optimality in natural and perturbed metabolic networks. *PNAS*, 99(23) :15112–15117, 2002.
- [SW02] VF Sirvent and Y Wang. Self-affine tiling via substitution dynamical systems and Rauzy fractals. *Pacific J. Math.*, 206(2) :465–485, 2002.
- [TCN01] JJ Tyson, K Chen, and B Novak. Network dynamics and cell physiology. *Nat Rev Mol Cell Biol*, 2 :908–916, 2001.
- [Tho79] R Thomas, editor. *Kinetic logic : a boolean approach to the analysis of complex regulatory systems*, volume 29 of *ecture Notes in Biomathematics*. Springer-Verlag, 1979.
- [Tho81] R Thomas. On the relation between the logical structure of systems and their ability to generate multiple steady states or sustained oscillations. *Springer Ser. Synergetics*, 9 :180–193, 1981.
- [Tho91] R Thomas. Regulatory networks seen as asynchronous automata : a logical description. *J. Theor. Biol.*, 153 :1–23, 1991.
- [THT<sup>+</sup>99] M Tomita, K Hashimoto, K Takahashi, TS Shimizu, Y Matsuzaki, and *al.* E-cell :software environment of whole-cell simulation. *Bioinformatics*, 15 :72–84, 1999.
- [Thu88] W Thurston. On the geometry and dynamics of diffeomorphisms of surfaces. *Bull. Amer. Math. Soc. (N.S.)*, 19(2) :417–431, 1988.
- [Thu89] WP Thurston. *Groups, tilings and finite state automata*. Lectures notes distributed in conjunction with the Colloquium Series, in *AMS Colloquium lectures*, 1989.
- [Thu06] JM Thuswaldner. Unimodular Pisot substitutions and their associated tiles. *J. Théor. Nombres Bordeaux*, 18(2) :487–536, 2006.
- [TMD03] L Travé-Massuyès and P Dague, editors. *Modèles et raisonnements qualitatifs*. Hermes sciences, 2003.
- [TS01] D Thieffry and L Sanchez. A logical analysis of the drosophila gap-gene system. *Journal of Theoretical Biology*, 211(2) :115–141, 2001.
- [Tur03] P Turchin. *Complex Population Dynamics : a Theoretical/Empirical Synthesis*. NJ : Princeton University Press, 2003.
- [Val03] B Vallée. Dynamical analysis of a class of Euclidean algorithms. *Theoret. Comput. Sci.*, 297(1-3) :447–486, 2003. Latin American theoretical informatics (Punta del Este, 2000).
- [Var03] F Varenne. La simulation informatique face à la méthode des modèles. le cas de la croissance des plantes. *Nature Sciences Sociétés*, 11 :16–28, 2003.
- [VAR04] MO Vlad, A Arkin, and J Ross. Response experiments for nonlinear systems with application to reaction kinetics and genetics. *PNAS*, 101(19) :7223–7228, 2004.
- [Veb07] P Veber. *Modélisation grande échelle de réseaux biologiques : vérification par contraintes booléennes de la cohérence des données*. PhD thesis, Université de Rennes 1, 2007.
- [Vee78] WA Veech. Interval exchange transformations. *J. Anal. Math.*, 33 :222–272, 1978.
- [Ver92] AM Vershik. Arithmetic isomorphism of hyperbolic automorphisms of a torus and of sofic shifts. *Funktsional. Anal. i Prilozhen.*, 26(3) :22–27, 1992.
- [vK35] ER van Kampen. On some characterizations of 2-dimensional manifolds. *Duke Math. J.*, 1(1) :74–93, 1935.
- [Wal82] P Walters. *An introduction to ergodic theory*. Springer-Verlag, New York, 1982.
- [Wal06] M Waldschmidt. Diophantine analysis and words. In *Diophantine analysis and related fields 2006*, volume 35 of *Sem. Math. Sci.*, pages 203–221. Keio Univ., Yokohama, 2006.
- [Yag75] G Yagil. Quantitative aspects of protein induction. In BL Horecker and ER Stadtman, editors, *Current topics in Cell regulation*, pages 183–237. Academic Press, 1975.
- [Yas99] SI Yasutomi. On Sturmian sequences which are invariant under some substitutions. In *Kanemitsu, Shigeru (ed.) et al., Number theory and its applications. Proceedings of the conference held at the RIMS, Kyoto, Japan*, pages 347–373. Kluwer Academic Publishers, 1999.



- [Yas05] SI Yasutomi. On a new algorithm for inhomogeneous Diophantine approximation. *Tsukuba J. Math.*, 29(1) :173–195, 2005.
- [YIJ04] CH Yeang, T Ideker, and T Jaakkola. Physical network models. *J Comput Biol*, 11(2-3) :243–62, 2004.
- [YMM<sup>+</sup>05] CH Yeang, HC Mak, S McCuine, C Workman, T Jaakkola, and T Ideker. Validation and refinement of gene-regulatory pathways on a network of physical interactions. *Genome Biol*, 6(7) :R62, 2005.
- [ZMPQ<sup>+</sup>02] A Zanzoni, L Montecchi-Palazzi, M Quondam, G Ausiello, M Helmer-Citterich, and G Cesareni. MINT : a Molecular INTeraction database. *FEBS Lett*, 513(1) :135–40, 2002.

---

# Activités scientifiques

## 5.6 Publications

### Livres et ouvrages

- N. Pytheas-Fogg. *Substitutions in Dynamics, Arithmetics and Combinatorics*. Lectures Notes in Mathematics 1794, Springer-Verlag, 2002. V. Berthé, S. Ferenczi, C. Mauduit and A. Siegel, editors.
- Journées Montoises d'Informatique Théorique (Rennes, 2006), Theoretical Informatics and Applications, Vol 42 (3), 2008. D. Caucal and A. Siegel, guest editors.

### Revue à comité de lecture

- (Monas. Math., 2008) S. Akiyama, G. Barat, V. Berthé and A. Siegel, Boundary of central tiles associated with Pisot beta-numeration and purely periodic expansions, to appear in *Monashefte für Mathematik*
- (BMC bioinfo, 2008) P. Veber, C. Guziolowski, M. Le Borgne, O. Radulescu, and A. Siegel, Inferring the role of transcription factors in regulatory networks, *BMC Bio-Informatics* 9, 2008. .
- (JNT, 2007) V. Berthé and A. Siegel, Purely Periodic beta-Expansions in the Pisot Non-unit Case *Journal of Number Theory* 153 (2) 2007, 153-172.
- (JPBC, 2006) C. Guziolowski, P. Veber, M. Le Borgne, O. Radulescu, and A. Siegel Checking Consistency Between Expression Data and Large Scale Regulatory Networks : A Case Study *Journal of Biological Physics and Chemistry* (7) 2007, 37-43 (Presented at Réseaux d'interaction : analyse, modélisation et simulation. RIAMS'06, Lyon, France.)
- (Ann. Inst. Fourier, 2006) P. Arnoux, V. Berthé, A. Hilion and A. Siegel Fractal representation of the attractive lamination of an automorphism of the free group, *Annales de l'Institut Fourier* 56(7), 2006, 2161-2212.
- (Biosystems, 2006) A. Siegel, O. Radulescu, M. Le Borgne, P. Veber, J. Ouy, S. Lagarrigue Qualitative analysis of the relation between DNA microarray data and behavioral models of regulation networks *BioSystems* 84, 2006, 153-174
- (Roy. Soc. Inter, 2006) O. Radulescu, S. Lagarrigue, A. Siegel, M. Le Borgne, P. Veber Topology and linear response of interaction networks in molecular biology, *Journal of The Royal Society Interface* 3(6), 2006, pp. 185 - 196
- (Complex Us, 2005) P. Veber, M. Le Borgne, A. Siegel, S. Lagarrigue, O. Radulescu Complex Qualitative Models in Biology : a new approach *Complexus* 2, 2004/2005, pp. 140-151 (Presented at of ECCS, Paris, 2005)

- (Integers, 2005) V. Berthé and A. Siegel, Tilings associated with beta-numeration and substitutions, *INTEGERS*, 5(3), pp. A2, 2005
- (Ann. Inst. Fourier, 2004) A. Siegel, Pure discrete spectrum dynamical system and periodic tiling associated with a substitution. *Annales de l'Institut Fourier* 2(54), 2004, p. 288-299.
- (TCS, 2004) P. Arnoux, V. Berthé and A. Siegel. Two-dimensional iterated morphisms and discrete planes . *Theoretical Computer Science*, 319, 2004, p. 145–176.
- (Erg. Th. Dyn. Sys., 2003) A. Siegel. Représentation des systèmes dynamiques substitutifs non unimodulaires. *Ergodic Theory and Dynamical Systems*, 23, 2003, 1247-1273
- (Trans. AMS, 2001) V. Canterini and A. Siegel. Geometric Representation of primitive substitution of Pisot type. *Transactions of the AMS*, 353(12), 2001, pp. 5121–5144.
- (JTN Bord., 2001) V. Canterini and A. Siegel. Automate des préfixes-suffixes associé à une substitution primitive. *Journal de théorie des nombres de Bordeaux*, 13(2), 2001, pp. 353–369.
- (Acta Arith, 2001) A. Siegel. Théorème des trois longueurs et suites sturmiennes : mots d'agencement des longueurs. *Acta Arithmetica*, 97 (1-3), 2001, pp. 195–210.

## Chapitre dans les ouvrages

- (Pytheas Fogg Intro, LNM, 2002) V. Berthé and A. Siegel. Basic notions on substitutions. In N. Pytheas-Fogg, *Substitutions in Dynamics, Arithmetics and Combinatorics*. Lectures Notes in Mathematics 1794, Springer-Verlag, 2002, pp. 1–34 (Chapter 1).
- (Pytheas Fogg, LNM, 2002) A. Siegel, Spectral theory and geometric representation of substitutions. In N. Pytheas-Fogg, *Substitutions in Dynamics, Arithmetics and Combinatorics*. Lectures Notes in Mathematics 1794, Springer-Verlag, 2002, pp. 199–252 (Chapter 7).

## Actes de colloques à comité de lecture

- (CIBB, 2008) C. Guziolowski, J. Gruel, O. Radulescu and A. Siegel, *Curating a large-scale regulatory network by evaluating its consistency with expression datasets*, CIBB 2008 : Computational Intelligence Methods for Bioinformatics and Biostatistics, To appear in Lecture Notes in Bioinformatics LNBI/LNCS, Springer-Verlag.
- (JOBIM, 2008) P. Blavy, F. Gondret, H. Guillou, S. Lagarrigue, P. Martin, O. Radulescu, A. Siegel and J. Van Milgen, *A minimal and dynamic model for fatty acid metabolism in mouse liver*, JOBIM, Lille, 2008.
- (Jour. Montoises, 2006) P Arnoux, V Berthé, and A Siegel, *Finiteness properties for Pisot S-adic tilings* Journées Montoises d'Informatique théorique. JM'06, Rennes, 2006
- (EWT, 2003) A. Siegel, *Spectral theory for dynamical system arising from substitutions*, EWM, K. Dajani, J. Von Reis (eds.), CWITract, Marseille, 2003.

## Soumissions

- (Repeat, submitted, 2008) J. Nicolas ; C. Rousseau ; A. Siegel ; P. Peterlongo ; F. Coste ; P. Durand ; S. Tempel ; A.-S. Valin and F. Mahé, *Local and Maximal Repeats*, 2008.

---

(Monographie S. & T., 2008) A. Siegel and J. Thuswaldner, *Topological properties of Rauzy fractals*, 2008.

## Travaux en cours

- (R., S., P., C.& L. en cours) O. Radulescu, A. Siegel, E. Pécou, C. Chatelain and S. Lagarrigue, *Algebraic mddules and qualitative constraints for the study of equilibria of biochemical models*
- (A., F., S.& S., en cours) B. Adamczewski, C. Frougny, A. Siegel, W. Steiner, *Algebraicity of the length of the largest interval  $[0, \varepsilon[ \cap \mathbb{Q}$  containing only purely periodic beta-expansions.*
- (B., B. & S., en cours) V. Berthé, J.Bourdon, A. Siegel *Génération de plan discret par des algorithmes de fractions continues.*
- (B., S., S. & T., en cours) V. Berthé, A. Siegel, P. Surer, J. Thuswaldner *Fractal tiles associated to generalized radix representations and shift radix systems.*
- (CANT, 2009) V. Berthé, A. Siegel, J. Thuswaldner *Tilings, substitutions and Rauzy fractals : effectivity and complexity*, Chapter of *Combinatorics, Automata and Number Theory*, Cambridge University Press.
- (A., G. & S., en cours) P. Arnoux, S. Giabicani, A. Siegel *Dynamique du nombre d'or.*

## Reuves sans comité de lecture

- (Biofutur, 2007) A. Siegel, C. Guziolowski, P. Veber, O. Radulescu, M. Le Borgne, *Optimiser un plan d'expérience à partir de modèles qualitatifs ?*, BioFutur (275), 2007, 27-31
- P. Durand, D. Lavenier, M. Leborgne, A. Siegel, P. Veber and J. Nicolas, *Applying Complex Models on Genomic Data*. ERCIM News (60), 2005.
- A. Siegel. Répétitions dans les figures géométriques. *Quadrature* 34, 2002, pp. 40-47.
- A. Siegel, Fractals a la carte, *Tangente*, Hors Série 18 Les fractales, 2004, pp. 80-84.

## Divers

- A. Siegel, *Représentations géométrique, combinatoire et arithmétique des systèmes substitutifs de type Pisot*, Thèse de troisième cycle, Université de la Méditerranée, 2000.

## 5.7 Communications

### Congrès, symposium

- Workshop “Number Theory and Ergodic Theory 2008 April”, Kanazawa, Japan (04/2008) *Topological properties of central tiles and boundary graphs*
- Journée annuelle du Groupe de Travail « Systèmes Dynamiques, Automates et Algorithmes » (GT SDA2) du GDR Informatique Mathématique, Paris (10/2007) *Propriétés topologiques des fractals de Rauzy*
- ICIAM'07 : International Conference on Applied and Industrial Mathematics, Zurich (07/2007). Minisymposium : new reasearch in bioinformatics *Qualitative response of interaction networks : application to the validation of biological models.*

- 
- Journée de Numération Graz 2007 (04/2007) *Topological properties of central tiles for substitutions.*
  - Troisième carrefour OUEST-genopole, Brest, janvier 2006 *Modélisation in silico de la régulation génétique du métabolisme des lipides : Problèmes et méthodes* (with S. Lagarrigue)
  - International Conference on Probability and Number Theory 2005, Japan (P&NT 05), juin 2005. *Beta-numeration and Rauzy fractals for non unit Pisot numbers*
  - Numeration Tilings and Substitutions, Grenoble, March 2005. *Finiteness properties for substitution dynamical systems,*
  - Number Theoretic Algorithms and Related Topics, Strobl, Austria, October 2004. *Covering associated to a beta-shift and different conditions for tiling.*
  - Université d'été Sciences Mathématiques et Modélisation, Bordeaux, 2004, France. *Dynamique du nombre d'or* (with P. Arnoux).
  - Aperiodic Order : Dynamical Systems, Combinatorics and Operators, Banff, Canada, mai 2004. *Spectral theory of substitutive systems : combinatorial conditions for pure discrete spectrum and tilings.*
  - EWM (European Women in Mathematics), Marseille, novembre 2003. *Spectral theory for dynamical system arising from substitutions,*
  - Workshop on Beta-numeration, generalized substitutions and tilings, mars 2003. *Purely periodic orbits of beta expansion for Pisot non unit,*
  - Forum des jeunes mathématiciennes et des jeunes informaticiennes, pp. 77-80, Paris, March 2002. *Autour des fractals de Rauzy.*
  - Systèmes dynamiques–l'odyssée dynamique, Session résidentielle du CIRM, troisième semaine, février 2001, Marseille. *Substitutions, Rauzy fractals and tilings,*
  - Systèmes dynamiques–l'odyssée dynamique, Session résidentielle du CIRM, cinquième semaine, février 2001, Marseille. *A  $P$ -adic representation of substitutions,*
  - Fourth international joint meeting of the AMS and SSM, Session on Geometric and Symbolic Dynamical Systems, Denton, Texas, mai 1999. *Geometric realizations of Pisot substitutions*
  - Deuxième Ecole d'automne doctorale AMIcale, Marseille, décembre 1997. *Théorème des trois longueurs et suites sturmiennes*

## Groupes de travail et Ateliers

- Journées CMM-INRIA, Santiago du Chili (07/2008) *from questions in systemic biology to constraints resolution*
- Atelier Substitutions et automorphismes de groupes libres. Porq'roll 2008 (06/2008) *Propriétés topologiques des fractals de Rauzy.*
- VicAnne Meeting, Paris, IHP, janvier 2006. *Etude des déplacements d'équilibre d'un modèle différentiel par décomposition en modules.*
- Atelier Substitutions et automorphismes de groupes libres, Marseille CIRM (2005)
- Groupe de travail Sur les exemples de Lattes, IRMAR, Rennes (03/2005)
- Groupe de travail Spectre des systèmes adiques, Marseille, CIRM, January 2005. *Systèmes substitutifs spectre discret : conditions explicites,*
- Research in team , Montpellier, July 2004 *representation of Non Pisot Unit substitutions*
- Workshop on Beta-numeration, generalized substitutions and tilings, March 2003. *Purely periodic orbits of beta expansion for Pisot non unit,*
- Atelier Substitutions généralisées, pavages et numération, Marseille, CIRM (2002).

- 
- Groupe d'étude sur la numération, CIRM, Marseille, mars 1999. *Automate des préfixes-suffixes et représentation des substitutions*

## Séminaires

- Station biologique de Roscoff, séminaire (10/2008)
- Université de Rennes I, Séminaire de systèmes dynamiques (02/2008)
- Laboratoire Systèmes Navals Complexes, Toulon (12/2007)
- Université de Rennes, groupe de travail application des mathématiques en biologique (12/2007)
- Université de Nice (03/2007)
- Université Paris XI, groupe de travail Systèmes Dynamiques (05/2006)
- ENS Lyon, séminaire de mathématiques, (04/2006)
- Université de Montpellier, LIRMM, groupe de travail d'info. théorique (02/2006)
- University of Neuchâtel, colloquium de mathématiques, (01/2006)
- University of Vienna, séminaire de mathématiques (11/2005)
- Université de Rennes I, séminaire de systèmes dynamiques (11/01, 02/05)
- Université de Brest, séminaire de systèmes dynamiques , (11/04)
- ENS-Cachan, antenne Bretagne, séminaire des élèves (05/04)
- Université Paris XI, séminaire de systèmes dynamiques (06/01)
- Dijon, séminaire de systèmes dynamiques (03/01)
- Université de Montpellier, séminaire de mathématiques (03/01)
- Université Paris VII, séminaire d'informatique du LIAFA (02/01)
- Université de Marne-la-Vallée, séminaire d'informatique (12/00)
- Université Paris VI, séminaire de théorie ergodique (11/00)
- Ecole Polytechnique, séminaire de géométrie ergodique (11/00)
- Université de Bordeaux, séminaire de théorie des nombres (10/00)
- Université de Turku (Finlande), séminaire de mathématiques (09/98)
- Université de la méditerranée, séminaire de théorie ergodique (04/98, 04/00)
- Université d'Amiens, séminaire d'informatique (03/98)
- Université de Provence, séminaire de dynamique symbolique (10/97, 03/99, 04/00)

## 5.8 Encadrement et enseignements

### Doctorat

- Directrice de la thèse de Carito Guziolowski, 2006-2009. *Méthodes d'analyse qualitative de réseaux biologiques à grande échelle.*
- Co-directrice de la thèse de Pierre Blavy, 2006-2009. *Modèles multi-organes et multi-espèces du métabolisme des lipides : nouvelles approches mathématiques et validation biologique.*
- Co-directrice de la thèse de Philippe Veber, 2004-2007. *Modélisation grande échelle de réseaux biologiques : vérification par contraintes booléennes de la cohérence des données.*

### Stages de master

- Laurent Chamoin, Master 2 bio-informatique, 2006-2007. *Recherche de modules dans le réseau MAPK.*

- Carito Guziolowski, Master 2 bio-informatique, 2005-2006. *Testing a new approach of qualitative modelling in E. Coli transcriptional regulatory network*
- Mathieu Morvan, Master 2 informatique, 2004-2005. *diagnostic pour la dynamique de réseaux biologiques.*
- Julien Ouy, Master 2 informatique, 2003-2004. *Modélisation de réseaux biologiques et application pour l'analyse de données expérimentales.*

## Enseignements

- Master 2 Modélisation de Systèmes Biologiques, Rennes. *Construction et exploitation de réseaux biomoléculaires* (2008) [24h]
- Formation interne Inria, Liffré. *Modélisation de réseaux biologiques* (2005) [2h]
- Master 1 de bio-informatique, Rennes. (2005) [4h]
- Université d'été Sciences Mathématiques et Modélisation, Bordeaux. *Dynamique du nombre d'or* (2004) [6h]
- *Niveau Licence* : Premier degré universitaire, Préparation concours (Mathématiques, Marseille et Rennes) (1998, 1999, 2000, 2001) [202h]

## 5.9 Coopérations et contrats

### Participation à des projets scientifiques

- Projet LAREDA, contrat ANR blanc (STIC) 2007-2010. *Lattice Reduction Algorithms : Dynamics, Probabilities, Experiments, Applications.* Responsable : Brigitte Vallée (CNRS, univ. Caen).
- Projet SITCON, contrat ANR Biosys 2006-2009. *Modeling Signal Transduction Induced By A Chimeric Oncogene.* Responsable : Andrei Zynoviev (Institut Curie)
- Projet DYCONUM. contrat ANR jeunes chercheurs (mathématiques) 2006-2009. *Études diophantiennes, dynamiques et combinatoires de différentes numérations.* Responsable : Wolfgang Steiner (LIAFA, univ. paris VII)
- Projet MOCA, contrat ARC INRIA 2005-2007 *MODularité, Compositionnalité et Abstraction dans les réseaux géniques et protéiques.* Responsable : Sylvain Soliman (projet contraintes, INRIA Rocquencourt)
- Projet VICANNE, contrat ACI IMPBio 2004-2007. *Modélisation dynamique et simulation des systèmes biologiques.* Responsable : Jean-Pierre Mazat (Univ. Bordeaux)
- Projet NUMERATION, contrat ACI Interfaces des mathématiques, 2003-2006. Responsable : Pierre Arnoux (Univ. Méditerranée)
- Projet MATHRESOGEN, contrat ACI IMPBio 2003-2006. *Méthodes mathématiques pour l'identification d'acteurs critiques dans des processus biochimiques régulés par un réseau de gènes.* Responsable : Ovidiu Radulescu (Univ. Rennes 1)

### Relations internationales

- programme Egide Sakura (Franco-japonais), 2007-2008. *Number theory and discrete dynamical systems*
- programme Egide Amadeus (franco-autrichien), 2008-2009. *Fractals and topological structures arising from dynamics*
- relation bilatérales avec le CMM, université du Chili (visite d'un mois en 2005).

## Séjours à l'étranger

- Laboratoire de mathématiques de l'université de Kanazawa, Japon. Visite d'une semaine, 2008.
- Laboratoire de mathématiques de Keio university, Japon. Visite de 3 jours, 2008.
- Laboratoire de mathématiques de l'université de Leoben, Autriche. Visite d'une semaine, 2005.
- Centre de modélisation mathématiques, Université du Chili, Santiago du Chili. Visite de 5 semaines, 2005.
- Laboratoire de Mathématiques de l'université du North Texas, Denton, USA. Visite de 2 semaines, 2000.
- Laboratoire de mathématiques de l'Université de Turku, Finlande. Visite de 2 semaines, 1999.

## 5.10 Animation de la recherche

### Administration

- Membre élue de la commission d'évaluation de l'INRIA, depuis 2008.
- Membre élue du conseil de laboratoire de l'IRISA (2003-2008)

### Participation à des jurys de thèse

- Jury de thèse de M. Manceny (Evry, 2006). *Réseaux de jeux. Une extension de la théorie des jeux pour la modélisation des interactions locales. Application aux réseaux de régulation génétique*. Directeur de thèse : Franck Delaplace.

### Participation à des comités de programmes

- Journées montoises d'informatique théorique, Rennes (2006), Mons (2008).
- ECAI 2008.
- Conférence de bioinformatique JOBIM : Bordeaux (2006), Marseille (2007), Lille (2008)
- RIAMS 2006, Réseaux d'Interactions : Analyse, Modélisation et Simulation, Workshop satellite à IPG Lyon (2006) ; Lyon (2007)

## Animation de la recherche

### Mathématiques et informatique

- Organisation de la conférence internationale Numeration : Mathematics and Computer Science, March 23-27, 2009, CIRM, Marseille. (70 participants attendus)
- Organisation de la conférence internationale Journées montoises d'informatique théorique, Rennes, août 2006. (70 participants)
- J'organise régulièrement des ateliers de travail sur le thème *Generalized substitution and tilings and numeration*, avec V. Berthé. Ces ateliers rassemblent une quarantaine de chercheurs ; la moitié sont étrangers. Des sessions ont eu lieu en Juin 2007 (porquerolles), Mars 2006 (CIRM, Marseille), Avril 2005 (CIRM, Marseille), Mars 2004 (CIRM, Marseille), Mars 2003 (CIRM, Marseille),



### **Bioinformatique**

- Organisation du séminaire de bioinformatique de l'IRISA depuis 2002.
- Organisation de 14 journées thématiques de bioinformatique depuis 2002, dans le cadre de l'animation de Ouest-Génope.
- Supervision de l'animation sur les réseaux biologiques dans le cadre du projet ACI VICANNE (programme IMPBio, 2004-2007).
- Organisation de deux journées satellites de la conférence de bioinformatique française JOBIM :
  - *Modélisation dynamique et simulation des réseaux biologiques* Lille, 2008
  - *Modélisation dynamique et simulation des réseaux biologiques Où en est-on ?*, Marseille, 2007.
- Organisation d'ateliers sur les réseaux biologiques. Ces réunions ont rassemblé de 20 à 50 participants français.
  - *Réseaux booléens et automates cellulaires pour la modélisation de systèmes biologiques*, Rennes, 2007 (organisé avec P. Veber).
  - *Aspects stochastiques de la modélisation des réseaux de régulation*, Nice 2005 (organisé avec J.A. Sepulchre).
  - *Modélisation de réseaux de gènes*, Dijon 2004 (organisé avec E. Pécou).

---

# Résumé

Ce travail présente des contributions théoriques et pratiques à la théorie des codages symboliques de systèmes dynamiques. Les applications concernent différents champs mathématiques et la modélisation en biologie moléculaire. Le but est d'illustrer comment des méthodes de discrétisation de systèmes dynamiques et une approche algorithmique permettent d'exploiter au mieux les connaissances disponibles sur le système, même partielles. Un premier objectif est d'exhiber des informations au sujet d'une dynamique que l'on connaît explicitement et les traduire en propriétés concrètes. Un deuxième objectif est de produire de la connaissance sur une dynamique ou un modèle lorsqu'on ne le connaît pas explicitement. Dans ce document, ces deux questions sont abordées sur deux grandes classes de systèmes dynamiques.

Les premiers systèmes considérés sont des automorphismes et des translations sur un tore. Inspirés par les cas unidimensionnels (beta-numération, étude des suites sturmiennes), la question principale qui se pose est de trouver un domaine fondamental pour le tore dans lequel les trajectoires de la dynamique considérée se codent par des systèmes symboliques simples. Dans le cas où l'automorphisme du tore considéré admet une unique direction dilatante (le cas Pisot), un bon candidat pour ces partitions est donné par un domaine dont la base est fractale, introduit par G. Rauzy dans les années 1980. Nous décrivons comment une approche décidable pour décrire le bord fractal du domaine et ses propriétés de pavage, permet de s'assurer qu'il s'agit d'un domaine adéquat pour un codage de l'automorphisme. La description du bord du domaine permet de décrire ses propriétés topologiques, et de les exploiter dans les différents domaines d'informatique théorique où les automorphismes et les additions sur un tore apparaissent. Ainsi, en théorie des nombres, nous nous appuyons sur la topologie du domaine pour caractériser les propriétés des développements finis ou purement périodiques de rationnels en base non entière. En géométrie discrète, ces propriétés s'interprètent en termes de conditions pour l'engendrement de plans discrets par des méthodes itératives.

La deuxième classe de systèmes concerne les systèmes dynamiques de grande échelle en biologie moléculaire. Il s'avère que les données et les connaissances sur les modèles de régulations transcriptionnelles dans une cellule sont souvent trop partielles pour leur appliquer les méthodes usuellement utilisées pour la modélisation de systèmes expérimentaux. Dans ce document, nous discutons d'un formalisme (inspiré par la dynamique) qui permet d'interpréter les observations en biologie mo-

léculaire, pour aider à la correction de modèles, et, dans le futur, à la mise en place de plans expérimentaux. Au vu de la qualité des données, les aspects dynamiques sont alors remplacés par des considérations sur les déplacements d'états stationnaires, et analyser les données revient à formaliser puis résoudre des contraintes portant sur des ensembles discrets. Nous montrons ainsi comment aborder les notions de corrections de modèles et de diagnostic de réseaux grande échelle.

**Mots clés** : systèmes dynamiques discrets. Fractal. Numération. Biologie des systèmes. Diagnostic. Contraintes.