Analyses de l'algorithme de Gauss. Applications à l'analyse de l'algorithme LLL.

Antonio Vera

UNIVERSITÉ de CAEN/BASSE-NORMANDIE

U.F.R. : Sciences

ÉCOLE DOCTORALE : SIMEM

THÈSE

présentée par

Antonio VERA

et soutenue

le 15 juillet 2009

en vue de l'obtention du

DOCTORAT de l'UNIVERSITÉ de CAEN

spécialité : Informatique

(Arrêté du 7 août 2006)

Analyses de l'algorithme de Gauss. Applications à l'analyse de l'algorithme LLL

MEMBRES du JURY

Valérie BERTHÉ	Directrice de Recherche au CNRS	LIRMM (Montpellier)	
Hervé DAUDÉ	Maître de Conférences	LATP (Aix-Marseille)	
Philippe FLAJOLET	Directeur de Recherche à l'INRIA	INRIA Rocquencourt	
Guillaume HANROT	Directeur de Recherche à l'INRIA	INRIA Nancy Grand-Est	(rapporteur)
Alejandro MAASS	Profesor Titular	Universidad de Chile (Chili)	(rapporteur)
Brigitte VALLÉE	Directrice de Recherche au CNRS	GREYC (Caen)	(directrice)
Alfredo VIOLA	Profesor Titular	U. de la República (Uruguay)	(rapporteur)

Mis en page avec la classe thloria.

Table des matières

Introduction

Le travail de cette thèse se situe au carrefour de deux domaines algorithmiques, le domaine de *l'analyse d'algorithmes* et celui de la *réduction des réseaux euclidiens*.

L'analyse d'algorithmes. C'est une branche de l'informatique mathématique fondée par Don Knuth dans les années 60, qui étudie mathématiquement le comportement des algorithmes, non pas dans le pire des cas comme c'est l'habitude, mais plutôt sur des instances "génériques". Ces analyses permettent de prédire le comportement "pratique" des algorithmes, mais aussi, et c'est souvent le plus important, de mieux comprendre leur structure, et d'isoler les noeuds de difficulté algorithmique. C'est donc aussi un puissant moteur d'amélioration algorithmique. L'ouvrage fondateur du sujet est l'ensemble des trois livres qui forment *The Art of Computer Programming* [35, 36, 37], tous parus entre la fin des années 60 et le début des années 70. Depuis, toute une communauté s'est créée sur cette thématique, et le livre de Flajolet et Sedgewick, [23] tout récemment publié, peut être considéré comme l'ouvrage qui fonde le domaine de la combinatoire analytique, qui est historiquement l'outil mathématique principal de l'analyse d'algorithmes jusqu'à l'heure actuelle.

La combinatoire analytique traite les problèmes combinatoires en utilisant l'objet central des séries génératrices, en utilisant des méthodes à la fois formelles et analytiques. La série génératrice est d'abord considérée comme une série formelle. Sa structure permet alors de refléter la combinatoire du problème et ce sont ses coefficients, via leur analyse asymptotique, qui vont permettre de revenir au problème de départ. Le comportement asymptotique de ces coefficients va dépendre fortement des singularités de la série génératrice, désormais vue comme une fonction de variable complexe.

Cette méthode générale a permis d'analyser très précisément le comportement générique de beaucoup d'algorithmes célèbres, et dans des domaines très variés de l'algorithmique. Il faut cependant remarquer que peu d'algorithmes arithmétiques ont été étudiés par ces techniques. Si l'algorithmique des polynômes sur un corps fini a été largement analysée, via cette méthodologie de combinatoire analytique, les algorithmes sur les nombres semblent se prêter plus difficilement à ces méthodes, car la présence des retenues trouble le paysage en introduisant des corrélations, que les outils de combinatoire analytique ne savent pas bien gérer. C'est pourquoi il a fallu introduire d'autres outils, comme les syustèmes dynamiques, en complément des outils généraux de combinatoire analytique ou de probabilités. Nous en reparlerons plus loin.

La réduction des réseaux euclidiens. On y étudie un objet très simple, le réseau euclidien, sous-groupe additif discret de l'espace euclidien $\mathbb{R}^n$, décrit aussi comme l'ensemble de combinaisons linéaires à coefficients entiers d'un ensemble, appelé *base*, de vecteurs indépendants. L'objet paraît au départ trop simple pour être réellement intéressant, mais cette simplicité est trompeuse, car c'est la coexistence des deux points de vue –algébrique et métrique– qui lui donne toute sa richesse et sa complexité. Réduire un réseau, c'est en trouver une base qui a de bonnes propriétés

euclidiennes, avec des vecteurs assez courts et assez orthogonaux. Là encore, le problème apparaît technique, et on a un peu de mal à lui trouver un intérêt général. Et, là aussi, on se trompe, car ce problème est essentiel, comme nous allons le voir.

La réduction des réseaux est à la fois un problème de mathématiques pures, un problème algorithmique, et c'est aussi maintenant un domaine complet de l'algorithmique qui regroupe toute une thématique, avec ses applications et les algorithmes associés. En tant que problème mathématique, il est né avec la *géométrie des nombres*, créée par Minkowski au dix-neuvième siècle. On envisageait alors le problème de la réduction de manière plutôt contemplative : quelles sont les bonnes notions de réduction ? Existe-t-il des bases réduites ? Parfois, en mathématiques, les preuves d'existence peuvent se révéler constructives, et donner lieu, une fois que la notion s'avère historiquement mûre, à des algorithmes. Mais, ce n'était pas le cas dans ce domaine. Évidemment, on savait bien que le problème de la réduction se ramenait en dimension 1 au calcul du pgcd ; il existait aussi un algorithme en dimension 2 proposé par Lagrange [42] et explicité par Gauss [26] qui résolvait complètement le problème. Mais, le versant algorithmique du problème n'était pas envisagé dans son ensemble.

Il y a eu un véritable tournant dans le domaine, quand, en 1982, Lenstra, Lenstra et Lovász [46] ont créé un algorithme, l'algorithme LLL, qui résout le problème de la réduction, dans un compromis très fructueux entre la qualité (euclidienne) de la base obtenue et le temps mis à l'obtenir. C'est cet algorithme qui a créé le domaine algorithmique de la réduction des réseaux. Mais il a fait plus : il a eu, dès sa création, un énorme impact, car il a servi de boîte à outils pour un grand nombre de problèmes variés, dépassant la première application qui avait motivé sa création (la factorisation des polynômes).

En cryptologie, et plus précisement en cryptanalyse, il s'est révélé essentiel pour casser des cryptosystèmes du type sac-à-dos [40], ou encore des générateurs pseudo-aléatoires [72]. Plus généralement, il a permis de casser presque tous les cryptosystèmes fondés sur des problèmes linéaires ou linéarisables. Par exemple, une méthode générale, due à Coppersmith, et utilisant la réduction des réseaux, permet de trouver les petites racines modulaires d'un polynôme modulaire, dès qu'on en connaît une approximation suffisamment bonne, et sans qu'on ait besoin de factoriser le module. En théorie de nombres, l'algorithme LLL permet de calculer des approximations diophantiennes simultanées [46], et il a été fondamental dans la réfutation par le calcul de la conjecture de Mertens [62].

L'algorithme LLL est ainsi devenu incontournable. Il est implanté dans la plupart des systèmes de calcul formel et de théorie de nombres : Sage, Pari/GP, Maxima, Magma, Maple, Mathematica, bibliothèque NTL, et aussi, de façon autonome, par exemple par Stehlé [71]. L'algorithme LLL est ainsi devenu une opération de base de l'informatique mathématique.

L'analyse de la réduction des réseaux. Les deux domaines que nous venons de présenter –analyse d'algorithmes et réduction des réseaux euclidiens– ne se sont pas encore (véritablement) rencontrés : même si l'algorithme LLL est très utilisé, son comportement n'est pas bien compris. On peut citer Shoup à ce propos, qui, dans la documentation de la librairie NTL [67], écrit :

> *I think it is safe to say that nobody really understands how the LLL algorithm works. The theoretical analyses are a long way from describing what "really" happens in practice. Choosing the best variant for a certain application ultimately is a matter of trial and error.*

Les performances observées apparaissent parfois bien meilleures que les bornes que l'on sait prouver dans le pire des cas. Est-ce une illusion ? Est-ce vrai "presque toujours" ? Est-ce vrai "en moyenne" ? De plus, la multiplicité des applications de l'algorithme, sur des réseaux qui

apparaissent très particuliers et structurés, ne facilite pas la définition d'un cadre général où l'on pourrait mener l'analyse. En conclusion, l'algorithme est à la fois très utilisé et bien mal compris.

Un petit historique. Les premières tentatives d'analyse de l'algorithme LLL ont commencé vers 1990. On peut distinguer trois lignes d'étude : l'étude de l'algorithme LLL lui-même, l'étude précise de la dimension 2 (algorithme de Gauss), l'étude précise de la dimension 1 (algorithme d'Euclide). Et, curieusement, la chronologie des résultats n'a pas suivi les dimensions étudiées.

L'algorithme LLL. L'analyse de l'algorithme LLL ne fait que commencer. La plupart des analyses existantes se placent dans des modèles simples –modèle uniforme de la boule unité de $\mathbb{R}^n$, ou modèles sphériques, un peu plus généraux– qui ne sont malheureusement pas ceux que l'on trouve dans les utilisations les plus habituelles de l'algorithme. Dans ce cadre, Daudé et Vallée [19] ont étudié le nombre d'itérations de LLL, en exhibant une borne supérieure sur le nombre moyen d'itérations, et en estimant la distribution de ce nombre d'itérations. Dans le même esprit, Akhavi [3] a étudié la probabilité qu'une base aléatoire d'entrée soit déjà LLL-réduite. Ces travaux ont été étendus par la suite par Akhavi, Marckert et Rouault [5] à des distributions plus générales, mais les modèles étudiés sont toujours bien loin des modèles "réalistes". De manière complémentaire, Nguyen et Stehlé [60] ont étudié expérimentalement le comportement de l'algorithme, cette fois-ci dans des modèles plus réalistes. Ils ont ainsi contribué à améliorer la compréhension de l'algorithme, énoncé des conjectures très intéressantes, ...mais n'ont rien prouvé. Enfin, il existe aussi un résultat important de complexité, dû à Ajtai, qui montre qu'il est possible d'échantillonner facilement des bases qui sont difficiles à réduire, avec une notion de réduction plus forte que celle de LLL. Mais on ne sait pas montrer que ces bases sont aussi difficiles à réduire dans le sens de LLL.

> *Aucun des deux versants des résultats actuels n'est donc satisfaisant, entre des preuves dans des modèles non réalistes et des conjectures...sans preuves dans des modèles plus réalistes.*

L'algorithme de Gauss. Vallée et Flajolet [73] puis Daudé, Flajolet et Vallée [18] ont effectué la première analyse probabiliste de l'algorithme de Gauss. Ils ont travaillé dans le modèle le plus simple possible, un modèle, dans le plan complexe, à la fois continu et "uniforme". L'algorithme de Gauss y est vu comme l'itération d'une transformation complexe, et on retrouve l'algorithme des fractions continues quand le complexe est réel. Ce travail laisse donc entrevoir que l'algorithme de Gauss a une dynamique reliée, mais différente, à celle de l'algorithme des fractions continues. Cette observation, due à Daudé, a incité Vallée à entreprendre l'étude fine et systématique des algorithmes de la dimension 1, dans toutes leurs versions. Dans le même temps, en 1995, Laville et Vallée [45] ont étudié les caractéristiques probabilistes de la configuration de sortie de l'algorithme de Gauss, notamment le premier minimum et le défaut d'Hermite de la base de sortie.

L'algorithme d'Euclide et l'algorithme des fractions continues. Contrairement à ce qui se passe en dimension $n \geq 2$, il y a, en dimension 1, deux algorithmes, qui ont un comportement bien différent, selon qu'on s'intéresse à un nombre rationnel (l'algorithme termine, s'appelle l'algorithme d'Euclide, et calcule le pgcd entre le numérateur et le dénominateur) ou à un nombre irrationnel (l'algorithme ne termine pas, et calcule le développement en fraction continue du nombre). L'algorithme des fractions continues peut donc se voir comme une version continue de l'algorithme d'Euclide. Et, comme le continu est souvent plus facile à appréhender que le discret, les premiers résultats ont concerné l'algorithme des fractions continues.

L'algorithme des fractions continues. Babenko [8] et Wirsing [82], ont analysé l'algorithme des fractions continues, avec des méthodes d'analyse fonctionnelle. En 1997, dans [77], Vallée uti-

lise les travaux de Mayer [53], et introduit des méthodes fonctionnelles qui fournissent un cadre suffisamment général pour étudier à la fois l'algorithme de Gauss et l'algorithme des fractions continues. Elle fournit aussi un modèle naturel qui permet d'expliquer la transition de l'algorithme de Gauss vers l'algorithme des fractions continues. Elle introduit notamment la notion de valuation qui permet de quantifier cette transition.

L'algorithme d'Euclide. L'analyse en moyenne de l'algorithme d'Euclide a commencé dans les années 70 avec les travaux de Heilbronn [31] et Dixon [21], qui ont utilisé respectivement des méthodes arithmétiques et probabilistes assez spécifiques. Brent a étudié l'algorithme du pgcd binaire [12], en faisant l'hypothèse heuristique qu'il se comportait comme son extension continue. Il a fallu attendre jusqu'en 1994 pour obtenir le premier résultat sur la distribution (limite) de l'algorithme d'Euclide, avec Hensley [32], qui démontre que le nombre d'itérations de l'algorithme d'Euclide suit une loi asymptotiquement gaussienne. Ce résultat frappant s'appuie sur des méthodes d'analyse fonctionnelle et utilise les méthodes initiées par Babenko et Wirsing, tout comme l'opérateur de Mayer.

La méthode d'analyse dynamique. C'est à la suite de ce travail d'Hensley que Vallée introduit la méthode d'analyse dynamique : voyant un algorithme comme un système dynamique, elle considère les opérateurs d'analyse fonctionnelle qui ont servi à l'analyse des fractions continues, comme des opérateurs générateurs qui servent à engendrer des séries génératrices nécessaires à l'analyse de l'algorithme d'Euclide. Elle peut ainsi gérer les corrélations liées aux retenues, et utiliser alors tout l'environnement de la combinatoire analytique. C'est de la combinatoire dynamique–analytique, qui se révèle aussi très fructueuse en théorie de l'information, pour étudier des sources complexes [79, 15].

Cette méthode a fait ses preuves et a permis d'analyser en moyenne tous les algorithmes d'Euclide, et de les classifier (voir [80] pour un article de synthèse). Cette analyse en moyenne inclut l'analyse très technique du pgcd binaire [78] conjecturée par Brent, tout comme l'analyse précise de la complexité en bits (appelée aussi complexité binaire) prouvée par Akhavi et Vallée [6], et aussi son extension à une classe plus vaste d'algorithmes d'Euclide, avec division généralisée, par Bourdon, Daireaux et Vallée [11]. Les travaux les plus récents étudient les versions "Diviser pour Régner" de l'algorithme [14]. Mais cette méthode a aussi permis l'analyse en distribution de ces algorithmes, en généralisant et simplifiant le résultat d'Hensley : Baladi et Vallée [9] démontrent la nature asymptotiquement gaussienne pour toute une classe de coûts, et une classe d'algorithmes. Enfin, Lhote et Vallée [49] montrent le caractère gaussien de la complexité en bits.

Les contributions de cette thèse. Notre but ultime, en particulier dans le cadre du projet Lareda de l'ANR Blanche, est de progresser dans les analyses précises et réalistes de l'algorithme LLL. A cette fin, nous voulons :

(*a*) Déterminer un modèle suffisamment général qui puisse décrire de manière réaliste les entrées de l'algorithme LLL et quantifier en particulier les paramètres géométriques qui les rendent a priori faciles ou difficiles à réduire.

(*b*) Analyser complètemernt et définitivement l'algorithme de Gauss, aussi complètement que ce qui a été fait pour l'algorithme d'Euclide, et, ce, dans un modèle réaliste, aussi bien pour les paramètres d'execution que pour les paramètres liés à la configuration de sortie.

(*c*) Expliquer les liens précis (mais aussi les différences) qui existent entre l'algorithme de Gauss et l'algorithme d'Euclide. En quoi l'algorithme d'Euclide peut-il être vu comme la limite de l'algorithme de Gauss ?

(*d*) Expliquer comment la compréhension très fine de la dimension 2 peut être exploitée dans l'analyse de l'algorithme LLL, et ce, dans des modèles réalistes.

Les principaux résultats. Ce sont les suivants. Entre parenthèses nous indiquons les théorèmes associés.

La notion de valuation. Les densités à valuation sont sous-jacentes à toutes les analyses de cette thèse. Vallée les a déjà introduites dans [77], afin de construire un cadre unificateur qui contienne à la fois l'algorithme de Gauss et l'algorithme des fractions continues. Nous les utilisons dans un cadre beaucoup plus général. Nous pensons en effet qu'elles permettent de construire une échelle simple de modèles de difficulté variable vis-à-vis de la réduction. C'est déjà vrai en dimension 2, car ce modèle à valuation nous permet d'obtenir une analyse paramétrée de l'algorithme de Gauss, à la fois réaliste d'un point de vue algorithmique, et trés satisfaisante d'un point de vue mathématique, puisque des objets classiques, comme les séries d'Eisenstein, y apparaissent naturellement . Ce modèle à valuation nous permet aussi de quantifier très précisément la transition de l'algorithme de Gauss vers l'algorithme d'Euclide. Enfin, nous pensons qu'il est aussi appelé à jouer un rôle important en dimension quelconque, dans les analyses de l'algorithme LLL. En effet, ce modèle à valuation peut englober des modèles aussi différents que les bases d'Ajtai (qui modélisent des instances difficiles vis-à-vis de la réduction) que des modèles faciles vis-à-vis de la réduction (comme le modèle de la boule aléatoire). En conclusion, ce modèle à valuation semble bénéficier d'un grand degré de généralité tout en restant suffisamment maniable.

La géométrie de sortie de l'algorithme de Gauss. Ces résultats sont énoncés dans la Partie III de cette thèse, dans les théorèmes **E**, **F**, **G**, **H**. Nous étudions trois paramètres qui permettent de décrire les propriétés géométriques de la base de sortie de l'algorithme : le premier minimum, le défaut d'Hermite et ce que nous appelons le deuxième minimum orthogonalisé, et ce, dans le modèle à valuation. Les deux premiers paramètres ont déjà été analysés par Laville et Vallée, mais seulement dans un modèle uniforme (correspondant à une valuation nulle). Nous les étudions pour une valuation quelconque. Par ailleurs, le troisième paramètre, celui que nous appelons le "deuxième minimum orthogonalisé" n'avait jamais été étudié précédemment ; après avoir expliqué pourquoi il est amené à jouer un rôle important dans l'analyse ultérieure de l'algorithme LLL, nous l'analysons précisément, dans le modèle à valuation.

Nous nous intéressons aussi à une question plus globale, avec un point de vue dynamique : quelle est la densité de sortie de l'algorithme, quand il a reçu en entrée une densité de valuation r ? Nous montrons que cette densité est liée de très près aux séries d'Eisenstein de poids $2 + r$.

La complexité de l'algorithme de Gauss. Ces résultats sont énoncés dans la Partie II de cette thèse, dans les théorèmes **A**, **B**, **C**, **D**. La complexité (en nombre d'itérations) de l'algorithme de Gauss a déjà été étudiée largement, d'abord dans le modèle uniforme par Daudé, Flajolet et Vallée, puis généralisée (en partie) par Vallée au modèle à valuation. Une telle étude procède en deux temps : on effectue d'abord l'analyse dans un modèle continu, puis on revient au modèle discret par des arguments de comptage de points entiers dans des domaines continus. Ce passage du continu au discret a été effectué dans le cas de la densité uniforme. Mais, plus délicat dans le cas d'une densité à valuation, il n'a pas été abordé par Vallée.

Nous faisons ici ces analyses de complexité, dans le modèle général à valuation, en effectuant aussi l'étape de passage du continu au discret. Et nous étudions des mesures de complexité plus générales, correspondant à ce qu'on appelle des coûts additifs, qui nous permettent d'étudier finalement la complexité en bits de l'algorithme de Gauss. Nous montrons que la complexité moyenne binaire est, pour toute densité à valuation fixée, linéaire en la taille des entiers d'entrée

–contrairement à l'algorithme d'Euclide qui a une complexité quadratique en la taille des entiers d'entrée.

La transition entre l'algorithme de Gauss et l'algorithme d'Euclide. (th. **D**) Le résultat précédent pose de nouveau la question de la transition entre les deux algorithmes, quand ils travaillent tous deux sur des données entières (ou rationnelles). Comment l'algorithme de Gauss se transforme-t-il en l'algorithme d'Euclide quand la valuation se rapproche de sa valeur limite $r \to -1$? Nous répondons très précisément à cette question.

Retour à l'algorithme LLL. Nous faisons donc une analyse précise –et assez exhaustive– de l'algorithme de Gauss, dans une classe de modèles qui permet de paramétriser la difficulté de l'algorithme. En retour, nous proposons deux pistes d'applications possibles de ces résultats à des dimensions supérieures.

Nous expliquons d'abord comment on peut exploiter l'analyse de la configuration de sortie de l'algorithme de Gauss dans l'étude d'une variante de l'algorithme LLL, l'algorithme Pair-Impair. Nous expliquons aussi en quoi notre étude permet de justifier (en partie) l'hypothèse majeure faite dans un travail très récent de Madritsch et Vallée [51]. Ces auteurs proposent une modélisation simplificatrice de l'algorithme LLL par des tas de sable, fondée sur une certaine régularité des étapes de l'algorithme LLL . Notre étude montre que cette hypothèse est justifiée en dimension 2.

Les méthodes. Dans cette thèse, nous utilisons des méthodes assez diverses, relevant de domaines variés ; géométrie élémentaire fine – systèmes dynamiques – analyse fonctionnelle, et théorie spectrale. Même si beaucoup de nos résultats relèvent de la méthodologie générale d'analyse dynamique, nous n'avons pas pu utiliser cette méthode clés en main, car le cadre double que nous avons choisi –valuation r quelconque et retour au modèle discret – nous oblige à raffiner ces méthodes, et à y apporter des contributions originales.

Publications. Les travaux de cette thèse ont fait l'objet de deux publications, en collaboration avec Brigitte Vallée. La première [74], intitulée "Lattice reduction in two dimensions : analysis under realistic probabilistic models", se concentre sur les analyses de l'algorithme de Gauss. Elle est parue dans les actes de la conférence internationale d'analyse d'algorithmes de 2007 (AofA07). La seconde [75], intitulée "Probabilistic Analyses of Lattice Reduction Algorithms", est un long article d'une soixantaine de pages. Il détaille les analyses de l'algorithme de Gauss, en particulier l'étude de l'exécution de l'algorithme, et la replace dans la problématique générale de la réduction des réseaux. Cet article est paru dans les actes de la conférence LLL+25, qui a rassemblé les diverses communautés qui travaillent dans le domaine (cryptologues, algorithmiciens, mathématiciens) autour des trois créateurs de l'algorithme LLL, pour fêter les 25 ans de leur algorithme. Cet article va paraître à la fin de 2009 dans le livre "The LLL algorithm – survey and applications" de la collection "Information Security and Cryptography" (Springer-Verlag). Ce livre regroupera les textes des quinze exposés donnés à cette conférence.

Ces deux articles, même s'ils sont longs, ne nous ont pas laissé encore la possibilité de décrire les preuves détaillées de nos résultats. C'est donc dans cette thèse qu'elles paraissent pour la première fois, et c'est notre projet d'écrire ultérieurement deux articles longs, correspondants à chacune des Parties II et III de cette thèse.

Première partie

Algorithmique des réseaux euclidiens.

Chapitre 1

Les réseaux euclidiens

Sommaire

La géométrie des nombres est une branche de la théorie de nombres introduite par Hermann Minkowski en 1896 [57]. Sa motivation première est l'étude des formes quadratiques définies sur $\mathbb{Z}^n$, et elle adopte un point de vue géométrique. Lorqu'on effectue un changement de base, la base canonique de $\mathbb{Z}^n$ se transforme en une base de $\mathbb{R}^n$, et l'ensemble $\mathbb{Z}^n$ en l'lensemble des combinaisons linéaires entières de vecteurs de cette base, ce qu'on appelle un réseau euclidien, tandis

que les formes quadratiques définies positives se relient à la norme euclidienne. Le minimum de la norme euclidienne, appelé premier minimum, est alors la longueur d'un vecteur le plus court non nul du réseau. Il joue un rôle central dans le domaine.

Ces objets mathématiques que sont les réseaux se révèlent être un outil de modélisation incontournable. Beaucoup de problèmes, de nature a priori très diverse, comme l'approximation diophantienne simultanée, la factorisation de polynômes, la factorisation d'entiers, la programmation linéaire entière, et, plus récemment des systèmes cryptographiques, s'expriment dans le vocabulaire des réseaux ; Leur résolution se ramène à des questions de base sur un réseau, et très souvent, à la détermination du premier minimum du réseau. C'est parce que la modélisation via les réseaux est à la fois universelle et puissante que les problèmes de base des réseaux deviennent eux aussi essentiels à résoudre.

Ce chapitre introduit, dans la section 1.1, les réseaux euclidiens, avec leurs objets de base (déterminant, minima successifs, défaut d'Hermite, parallélotope fondamental). Il décrit les différents points de vue qu'on peut adopter sur les réseaux –sous-groupe discret de $\mathbb{R}^n$, ou ensemble des combinaisons linéaires entières d'un système libre– et les relie. Il envisage ensuite les principaux problèmes qu'on peut se poser sur un réseau, avec des problèmes qui apparaissent plus liés à la structure algébrique, et d'autres plus dépendant de la structure euclidienne de l'espace ambiant. Il décrit la dichotomie qui existent entre ces problèmes du point de vue de la théorie de la complexité, certains étant "faciles" et d'autres s'avérant plus "difficiles" (section 1.2). Le chapitre se termine en décrivant la puissance modélisatrice des réseaux : il parcourt un certain nombre des problèmes algorithmiques naturels, et explique comment leur résolution peut s'effectuer "dans les réseaux" (section 1.3).

1.1 Réseaux euclidiens

Il s'agit d'abord de décrire les principaux objets reliés à une base. On définit ensuite un réseau euclidien, et les pricipaux paramètres qui décrivent sa géométrie (minima successifs, déterminant, défaut d'Hermite) reliés par le théorème de Minkowski.

1.1.1 Orthogonalisée de Gram-Schmidt, Matrice de Gram, Parallélotope fondamental.

L'espace vectoriel réel $\mathbb{R}^n$, $n \geq 1$ est muni de sa structure euclidienne et de la mesure de Lebesgue, notée μ. La base canonique de $\mathbb{R}^n$ est désignée par $(e_1, e_2, \ldots, e_n)$. Le produit scalaire de $v, u \in \mathbb{R}^n$, et la norme euclidienne de u sont respectivement désignés par $v \cdot u$, et $||u|| = (u \cdot u)^{1/2}$. A une partie $E \subseteq \mathbb{R}^n$, nous associons l'espace vectoriel réel engendré par E, que nous désignons par $\langle E \rangle$.

L'ensemble des matrices à $n \geq 1$ lignes et $m \geq 1$ colonnes, avec des coefficients dans un ensemble $\mathbb{S}$ (en pratique $\mathbb{R}$,$\mathbb{Q}$ ou $\mathbb{Z}$) est noté $\mathbb{S}^{n \times m}$. Pour une matrice M, nous désignons sa transposée par ${}^tM$, son déterminant par $\det M$ et sa matrice inverse, (quand elle existe) par M^{-1}.

A un système $B = (b_1, \ldots, b_p)$ de p vecteurs de $\mathbb{R}^n$, on associe la matrice dont les lignes sont les vecteurs b_i exprimés dans la base canonique $(e_1, e_2, \ldots, e_n)$ de $\mathbb{R}^n$. Cette matrice sera appelée la matrice ligne de $(b_1, \ldots, b_p)$ et sera aussi, avec un léger abus de langage, désignée par B.

Les ensembles $[\![a, b]\!]$ définis par $[\![a, b]\!] := [a, b] \cap \mathbb{Z}$ seront appellés *intervalles entiers*. La boule ouverte (resp. fermée) de rayon ρ centrée en a est désignée par $B(a, \rho)$ (resp. $\overline{B}(a, \rho)$) et définie,

comme à l'habitude par

$$B(a,\rho) = \{x \in \mathbb{R}^n \mid ||x-a|| < \rho\}, \qquad \overline{B}(a,\rho) = \{x \in \mathbb{R}^n \mid ||x-a|| \leq \rho\}.$$

Définition 1.1 (Orthogonalisée de Gram-Schmidt, OGS)**.** *Soit une famille $B = (b_1, \ldots, b_p)$ formée de vecteurs linéairement indépendants de $\mathbb{R}^n$. On désigne par B_i la famille commençante, $B_i := (b_1, \ldots, b_i)$ et par H_i le $\mathbb{R}$-espace vectoriel engendré par B_i. La famille* orthogonalisée de Gram-Schmidt *est la famille orthogonale $B^\star = (b_1^\star, \ldots, b_p^\star)$ formée des vecteurs $b_i^\star$, où $b_i^\star$ est la projection orthogonale de b_i sur l'orthogonal de H_{i-1} . Plus précisément*

$$\begin{aligned} b_1^\star &= b_1 \\ b_i^\star &= b_i - \sum_{k=1}^{i-1} m_{i,k} b_k^\star, \quad \text{avec} \quad m_{i,j} = \frac{b_i \cdot b_j^\star}{||b_j^\star||^2} \quad \text{pour} \ \ 1 \leq j < i \leq p \end{aligned}$$

On pose de plus $m_{i,i} = 1$ pour $1 \leq i \leq p$ et $m_{i,j} = 0$ pour $1 \leq i < j \leq p$. Ce procédé d'orthogonalisation de Gram-Schmidt construit aussi la matrice $\mathcal{P} \in \mathbb{R}^{p \times p}$ dont l'entrée $m_{i,j}$ est définie ci-dessus.

Si on désigne aussi par B la matrice de $\mathbb{R}^{p \times n}$ dont la ligne d'indice i est le vecteur b_i (dans la base canonique), et si $B^\star$ est la matrice de $\mathbb{R}^{p \times n}$ dont la ligne i est le vecteur $b_i^\star$ dans la base canonique, le processus *OGS* construit l'égalité matricielle $B = \mathcal{P} B^\star$.

Définition 1.2 (Longueurs et rapports de Siegel)**.** *Soit une famille $B = (b_1, \ldots, b_p)$ et soit $B^\star = (b_1^\star, \ldots, b_p^\star)$ son orthogonalisée. La norme du vecteur $b_i^\star$, désignée par ℓ_i, est appelée la i-ème longueur de Siegel, et le rapport $r_i := \ell_{i+1}/\ell_i$ est appelé i-ème rapport de Siegel.*

Définition 1.3 (Matrice de Gram)**.** *La matrice de Gram d'un système $B = (b_1, \ldots, b_p)$ de p vecteurs de $\mathbb{R}^n$, désignée par $G(b_1, \ldots, b_p)$ ou $G(B)$, est la matrice de $\mathbb{R}^{p \times p}$ définie par*

$$G_{ij} = b_i \cdot b_j \qquad \forall\ i, j \in [\![1, p]\!]\,.$$

Si la matrice B a pour lignes les vecteurs du système $(b_1, \ldots, b_p)$, alors la matrice de Gram s'écrit $G(B) = B \cdot {}^t B$.

Voici quelques propriétés importantes de la matrice de Gram.

Lemme 1.1. *La matrice de Gram $G(B)$ associée à un système $B = (b_1, \ldots, b_p)$ de p vecteurs de $\mathbb{R}^n$ d'orthogonalisé $B^\star = (b_1^\star, b_2^\star, \ldots, b_p^\star)$ vérifie les propriétés suivantes*

(i) Soit C un système de p vecteurs tel que $B = UC$ pour une matrice (carrée) U. Alors $\det G(B) = (\det U)^2 \cdot \det G(C)$.

(ii) Soit U une transformation orthogonale de $\mathbb{R}^n$ et C le système transformé de B en appliquant à chaque vecteur b_i de B la transformation U. Alors $G(C) = (B^t U) \cdot (U^t B) = G(B)$.

(iii) $\det G(B) = \prod_{i=1}^{p} ||b_i^\star||^2 = \prod_{i=1}^{p} \ell_i^2$

(iv) $G(B)$ est inversible si et seulement si le système $B = (b_1, \ldots, b_p)$ est linéairement indépendant.

Démonstration. Le point (i) découle directement de

$$\det G(B) = \det(B\,{}^tB) = \det(U(C\cdot{}^tC){}^tU) = \det U \cdot \det G(C) \cdot \det({}^tU) = (\det U)^2 \det G(C),$$

et (ii) est montrée dans l'énoncé.
Montrons (iii). L'orthogonalisation de Gram-Schmidt écrit la matrice-ligne B sous la forme $B = \mathcal{P}B^\star$, où $\mathcal{P}$ et $B^\star$ sont donnés dans la définition 1.1. Comme $\mathcal{P}$ est une matrice carrée de déterminant 1, il suffit d'appliquer (i) pour conclure que $\det G(B) = \det G(B^\star)$. Par ailleurs comme $B^\star$ est la matrice d'un système orthogonal, la matrice $G(B^\star)$ est diagonale, avec les éléments $||b_i^\star||^2$ sur la diagonale, d'où le résultat. Pour (iv), on observe que si le système $B = (b_1, \ldots, b_p)$ n'est pas inversible, alors l'un des vecteurs $b_i^\star$ est nécessairement nul. Le déterminant de $G(B)$ est alors nul grâce à (iii). □

Définition 1.4 (Parallélotope fondamental)**.** *Le parallélotope construit sur un système indépendant $B = \{b_1, \ldots, b_p\}$ est l'ensemble convexe défini par*

$$\mathcal{Q}(B) = \{\sum_{i=1}^{p} x_i\, b_i \mid x_i \in [0,1[\}\,.$$

On désigne aussi par $\overline{\mathcal{Q}}(B)$ l'adhérence de $\mathcal{Q}(B)$. Les mesures de Lebesgue p-dimensionnelles de $\mathcal{Q}(B)$ ou de $\overline{\mathcal{Q}}(B)$ vérifient

$$\mu(\mathcal{Q}(B)) = \mu(\overline{\mathcal{Q}}(B)) = [\det G(B)]^{1/2} = \prod_{i=1}^{p} ||b_i^\star|| = \prod_{i=1}^{p} \ell_i. \tag{1.1}$$

Remarquons que la formule (1.1) est découle simplement d'un changement de variables.

1.1.2 Réseaux

Un réseau euclidien de $\mathbb{R}^n$ est l'ensemble des combinaisons linéaires à coefficients entiers d'une famille $\{b_1, \ldots, b_p\}$ de p vecteurs linéairement indépendants de $\mathbb{R}^n$, appelée *base* du réseau. Par ailleurs, un réseau peut être défini comme un sous-groupe additif discret de $\mathbb{R}^n$. La proposition suivante montre que les deux définitions sont équivalentes.

Proposition 1.1 (Preuve de Siegel, [69])**.** *Les assertions suivantes sont équivalentes :*

(i) $\mathcal{L}$ est un sous-groupe additif discret de $\mathbb{R}^n$, qui engendre un sous-espace vectoriel de dimension p.

(ii) Il existe un système $B = \{b_1, \ldots, b_p\}$ de p vecteurs linéairement indépendants de $\mathbb{R}^n$, pour lequel

$$\mathcal{L} = \{\sum_{i=1}^{d} x_i b_i \mid \quad x_i \in \mathbb{Z} \quad \forall\, i \in [\![1,p]\!]\}\,.$$

Démonstration. (i)⇒ (ii). Supposons que $\mathcal{L}$ est un sous-groupe additif discret de $\mathbb{R}^n$ engendrant un espace vectoriel de dimension p. Choisissons dans $\mathcal{L}$ un ensemble de p vecteurs linéairement indépendants, qu'on désigne par $C = \{c_1, c_2, \ldots, c_p\}$. À partir de cet ensemble, nous allons construire une famille indépendante $\{b_1, \ldots, b_p\}$, qui engendre $\mathcal{L}$ par combinaisons linéaires à coefficients entiers. Considérons tous les systèmes $\{b_1, \ldots, b_p\} \subset \mathcal{L}$, où, pour tout $i \in [\![1,p]\!]$ le vecteur b_i est un élément de $\overline{\mathcal{Q}}(C)$ qui s'écrit sous la forme

$$b_i = \sum_{j=1}^{i-1} x_{ij}\, c_j + x_i\, c_i \tag{1.2}$$

où les x_{ij} vérifient $0 \leq x_{ij} < 1$ et les x_i vérifient $0 < x_i \leq 1$. De tels systèmes existent, puisque l'on peut toujours choisir $b_i = c_i$ (avec $x_{ij} = 0$, $x_i = 1$) pour $1 \leq j < i \leq p$. Comme $c_i \neq 0$, elles sont formées de vecteurs linéairement indépendants. Nous choissisons un système $\{b_1, \ldots, b_p\} \subset \mathcal{L}$ particulier de la manière suivante : en suivant les indices $i \in [\![1, p]\!]$ dans l'ordre croissant, on choisit à chaque fois le b_i dont le x_i est le plus petit possible. Ce choix s'effectue sur un ensemble non vide, qui est fini car $\mathcal{L}$ est discret et $\overline{\mathcal{Q}}(C)$ compact.

Considérons un vecteur $w \in \mathcal{L}$ qui s'écrit $w = \sum_{i=1}^{p} y_i b_i$, pour certains réels y_i, $i \in [\![1, p]\!]$. Nous allons montrer que les coefficients y_i sont entiers. Supposons, par l'absurde, que ce n'est pas le cas, et considérons le premier indice k, en parcourant les indices depuis la fin vers le début, pour lequel $y_k \notin \mathbb{Z}$. Alors $y'_k := y_k - \lfloor y_k \rfloor$ est non nul, et le choix de k entraîne que

$$\sum_{i=1}^{k} y_i b_i \in \mathcal{L}, \quad \text{et donc que} \quad v := \sum_{i=1}^{k-1} y_i b_i + y'_k b_k \in \mathcal{L}.$$

Maintenant, en remplaçant b_i par son écriture dans la base c_i donnée par (1.2), le vecteur v s'écrit

$$v = \sum_{i=1}^{k-1} \nu_i \, c_i + y'_k \cdot x_k \, c_k,$$

pour des coefficients ν_i adéquats. Posons $\nu'_i := \nu_i - \lfloor \nu_i \rfloor$ pour tout $i \in [\![1, k-1]\!]$. Alors, le vecteur

$$v' = \sum_{i=1}^{k-1} \nu'_i \, c_i + y'_k x_k \, c_k$$

appartient à $\mathcal{L}$, et ses coefficients dans la base c_i vérifient

$$0 \leq \nu'_i < 1 \quad \forall \; i \in [\![1, k-1]\!], \qquad 0 < y'_k \, x_k < x_k.$$

Ceci contredit le choix de x_k, et conclut la première partie de la preuve.

(ii) $\Rightarrow$ (i). Soit $\mathcal{L}$ l'ensemble de combinaisons linéaires entières d'un système indépendant $B = \{b_1, \ldots, b_p\}$. Il est claire que $\mathcal{L}$ est un sous-groupe additif de $\mathbb{R}^n$. Nous prouvons ici que $\mathcal{L}$ ne contient pas des vecteurs arbitrairement petits, hormis le vecteur nul. Nous allons borner inférieurement la longueur d'un vecteur non nul de $\mathcal{L}$ en fonction de la longueur des vecteurs de la base $B^\star$ orthogonalisée de Gram-Schmidt de la base B. Nous énonçons ce résultat en une proposition qui sera utile dans la suite.

Proposition 1.2. *Soit un réseau $\mathcal{L}$ engendré par une base $B = (b_1, \ldots, b_p)$. On désigne par $B^\star$ la base orthogonalisée de B. Alors, pour tout $w \in \mathcal{L} \setminus \{0\}$ on a*

$$||w|| \geq \min \{||b_i^\star||; \quad i \in [\![1, p]\!]\}.$$

Démonstration. Tout vecteur $w \in \mathcal{L} \setminus \{0\}$ s'écrit

$$w = x_1 b_1 + \cdots + x_p b_p \,, \qquad \text{avec} \quad x_i \in \mathbb{Z} \text{ non tous nuls.}$$

Associons au vecteur w le plus petit indice r pour lequel w appartient au sous-espace H_r engendré par le système $(b_1, \ldots, b_r)$: l'indice r vérifie $x_r \neq 0$ et pour tout $i > r$, $x_i = 0$. Dans la base $B^\star$, le vecteur w s'écrit

$$w = \sum_{i=1}^{r} x_i b_i = \sum_{i=1}^{r} x_i \left(\sum_{j=1}^{i} m_{i,j} b_j^\star \right) = x_r b_r^\star + \left[\sum_{j=1}^{r-1} \left(\sum_{i=j}^{r} x_i m_{i,j} \right) b_j^\star \right]. \tag{1.3}$$

On en conclut que $||w||^2 \geq |x_r|^2||b_r^\star||^2$. Mais, comme x_r est un entier non nul, on en déduit que $||w||^2 \geq ||b_r^\star||^2$, ce qui achève la preuve. □

Ainsi, la proposition 1.2 prouve bien que $\mathcal{L}$ est discret, et la proposition 1.1 est achevée. □

En conclusion, un réseau euclidien $\mathcal{L}$ possède toujours une base qui l'engendre par combinaisons linéaires entières, et cette base a un cardinal égal à la dimension de l'espace vectoriel engendré par $\mathcal{L}$.

Proposition 1.3 (Équivalence de bases)**.** *Soient $B, C \in \mathbb{R}^{p\times n}$ deux bases du même réseau. Alors*
(i) il existe une matrice $U \in \mathbb{Z}^{p\times p}$ avec $\det U = \pm 1$ telle que $B = UC$.
(ii) Les deux déterminants $\det G(B)$ et $\det G(C)$ sont égaux.

Démonstration. Comme B et C sont bases du même réseau, il existe des matrices $U, V \in \mathbb{Z}^{p\times p}$ telles que

$$C = UB \qquad \text{et} \qquad B = VC, \qquad \text{et donc} \quad B = (VU)B$$

En multipliant à droite par ${}^tB$, on obtient la relation $G(B) = (VU)G(B)$. Comme la matrice $G(B)$ est inversible, on en déduit l'égalité $I = UV$, et comme les matrices U et V sont carrées, l'égalité $\det U \cdot \det V = 1$. Comme les deux matrices U et V ont des coefficients entiers, leurs déterminants sont entiers, ce qui entraîne l'égalité $|\det U| = |\det V| = 1$ et achève la preuve. □

Pour une base B d'un réseau $\mathcal{L}$, le déterminant de $G(B)$ est indépendant de la base B. C'est par définition de déterminant du réseau. Il est égal au volume p-dimensionnel de n'importe quel parallélotope fondamental.

Définition 1.5 (Déterminant d'un réseau)**.** *Le déterminant d'un réseau $\mathcal{L}$ est un réel positif défini par la relation*

$$(\det \mathcal{L})^2 = \det G(B) = \det(B \cdot {}^tB),$$

qui fait intervenir une base quelconque B du réseau $\mathcal{L}$ et sa matrice-ligne B.

Proposition 1.4. *Le déterminant d'un réseau $\mathcal{L}$ vérifie*

$$\det \mathcal{L} = \prod_{i=1}^{p} ||b_i^\star||$$

pour tout système $B^\star$ associé à une base B de $\mathcal{L}$. Si le réseau est de dimension pleine (i.e., $p = n$), alors, pour toute base b de $\mathcal{L}$, la matrice-ligne B est une matrice carrée, et $\det \mathcal{L} = |\det B|$

Preuve. Direct d'après les propriétés de la matrice de Gram, vues dans le lemme 1.1. □

On note que le déterminant peut se définir d'un point de vue purement algébrique mais aussi d'un point de vue à la fois algébrique et topologique.

1.1.3 Minima successifs

Définition 1.6. *Le premier minimum du réseau $\mathcal{L}$, désigné par $\lambda_1(\mathcal{L})$, est la norme d'un plus court vecteur non nul de $\mathcal{L}$. Plus généralement, le i-ème minimum du réseau $\mathcal{L}$, désigné par $\lambda_i(\mathcal{L})$, est le plus petit nombre réel positif ρ pour lequel la boule fermée de rayon ρ centrée à l'origine contienne au moins i vecteurs linéairement du réseau $\mathcal{L}$,*

$$\lambda_i(\mathcal{L}) := \min\left\{\rho > 0 \mid \dim(\overline{B}(0,\rho) \cap \mathcal{L}) \geq i\right\}.$$

Le résultat suivant établit une minoration du $i$ème minimum en fonction des longueurs des vecteurs de la base $B^\star$ associée à une base quelconque du réseau $\mathcal{L}$. Il généralise la Proposition qui a déjà établi cette minoration dans le cas du premier minimum.

Proposition 1.5. *Le i-ème minimum du réseau $\mathcal{L}$ vérifie*

$$\lambda_i(\mathcal{L}) \geq \min\{L_J; \quad J \subset [\![1,p]\!], |J| = i, J \cap [\![i,p]\!] \neq \emptyset\}, \qquad \textit{avec} \quad L_J := \max\{||b_k^\star||; \quad k \in J\}$$

pour tout système $B^\star$ associé à une base B de $\mathcal{L}$.

Démonstration. Considérons un système $(w_1, w_2, \dots w_j)$ de i vecteurs linéirement indépendants de $\mathcal{L}$ dont la norme est au plus ρ. Associons à chaque vecteur w_k le plus petit indice $r = r(k)$ pour lequel w appartient au sous-espace H_r engendré par le système $(b_1, \dots, b_r)$. Remarquons que tous les indices $r(k)$ ne peuvent être tous strictement inférieurs à i, car cela contredirait l'indépendance des vecteurs w_k. Par ailleurs, le vecteur w_k satisfait $||w_k|| \geq ||b^\star_{r(k)}||$. On en déduit le résultat cherché. □

1.1.4 Théorème de Minkowski

Ce théorème important relie le premier minimum d'un réseau et son déterminant. Il est fondé sur un résultat, dû à Blichfeldt.

Théorème 1.1 (Blichfeldt)**.** *On considère un réseau $\mathcal{L}$ de dimension p. On désigne par μ la mesure de lLebesgue p-dimensionnelle, et on considère un sous-ensemble C du sous-espace vectoriel engendré par $\mathcal{L}$, mu-mesurable. Si C satisfait $\mu(C) > \det(\mathcal{L})$, alors, il existe deux vecteurs distincts $s, t \in C$ pour lesquels $s - t \in \mathcal{L}$.*

Preuve. La preuve est fondée sur le principe des tiroirs. On considère une base B de $\mathcal{L}$ et on désigne par $\mathcal{Q}(B)$ son parallélotope fondamental. Pour chaque $x \in \mathcal{L}$ on pose

$$C_x = (x + \mathcal{Q}(B)) \cap C\,,$$

où, pour un ensemble A, la notation $A + x$ désigne l'ensemble $x + A := \{x + y \mid y \in A\}$. Ces ensembles C_x sont des parties disjointes de C, dont la réunion égale C :

$$C = \bigcup_{x \in \mathcal{L}} C_x, \qquad \text{et donc} \qquad \mu(C) = \sum_{x \in \mathcal{L}} \mu(C_x)\,.$$

En translatant tous ces ensembles dans le parallélotope fondamental, on observe qu'au moins deux d'entre eux ont une intersection non vide : On pose

$$C'_x = C_x - x \subseteq \mathcal{Q}(B),$$

et on raisonne par l'absurde : supposons, au contraire, que ces ensembles sont disjoints deux à deux. Alors,

$$\mu\left(\bigcup_{x \in \mathcal{L}} C'_x\right) = \sum_{x \in \mathcal{L}} \mu\left(C'_x\right) = \sum_{x \in \mathcal{L}} \mu\left(C_x\right) = \mu(C) > \det \mathcal{L}$$

mais en même temps,

$$\bigcup_{x \in \mathcal{L}} C'_x \subseteq \mathcal{P}(b), \quad \text{et donc} \quad \mu(C) = \mu\left(\bigcup_{x \in \mathcal{L}} C'_x\right) \leq \mu(\mathcal{Q}(B)) = \det \mathcal{L},$$

ce qui apporte la contradiction cherchée. Donc il existe deux vecteurs x et y distincts de $\mathcal{L}$ pour lesquels les ensembles C'_x et C'_y sont non disjoints : il existe donc $z \in C'_x \cap C'_y$. Alors les deux points s et t définis par $s := z+x \in C$, $t := z+y \in C$ vérifient $s-t = (z+x)-(z+y) = x-y \in \mathcal{L}$, puisque $x, y \in \mathcal{L}$. La preuve est ainsi achevée. □

Théorème 1.2 (Minkowski)**.** *On considère un réseau euclidien $\mathcal{L}$ de dimension p. On désigne par μ la mesure de Lebesgue p-dimensionnelle, et on considère un sous-ensemble C du sous-espace vectoriel engendré par $\mathcal{L}$, mu-mesurable. qui est convexe, symétrique par rapport à l'origine, et vérifie $\mu(C) > 2^p \det \mathcal{L}$. Alors, C contient au moins un point de $\mathcal{L}$.*

Preuve. L'ensemble $\frac{1}{2}C = \{y/2 \mid y \in C\}$ vérifie

$$\mu(\frac{1}{2}C) = 2^{-d}\mu(C) > \det \mathcal{L}.$$

Donc, le théorème de Blichfeldt prouve qu'il existe deux points distincts $x, y \in \frac{1}{2}C$ dont la différence $x - y$ appartient à $\mathcal{L}$. Alors $2x, 2y \in C$ et la symétrie de C par rapport à l'origine montre que $-2y \in C$. La convexité de C entraîne alors que

$$\frac{1}{2}(2x - 2y) = x - y \in C,$$

et, comme $x - y \in \mathcal{L}$, le théorème est prouvé. □

Théorème 1.3 (Minkowski)**.** *Soit $\mathcal{L}$ un réseau de dimension p. Alors, le premier minimum $\lambda_1(\mathcal{L})$ du réseau et le déterminant $\det \mathcal{L}$ du réseau sont reliés par l'inégalité*

$$\lambda_1(\mathcal{L})| \leq \sqrt{p} \cdot (\det \mathcal{L})^{1/p}.$$

Preuve. La boule fermée centrée à l'origine et de rayon $\sqrt{p}\,(\det \mathcal{L})^{1/p})$ contient strictement l'hypercube de côté $2(\det \mathcal{L})^{1/p}$ centré à l'origine. Elle a donc un volume strictement supérieur à $2^p \det \mathcal{L}$. La boule étant convexe et symétrique autour de l'origine, le théorème 1.2 affirme qu'elle contient un point $w \in \mathcal{L}$, qui évidemment vérifie $||w|| \leq \sqrt{p} \cdot (\det \mathcal{L})^{1/p}$. □

1.1.5 Défaut d'Hermite et constante d'Hermite

Définition 1.7 (Défaut d'Hermite)**.** *Le* défaut d'Hermite *d'un réseau $\mathcal{L}$ de dimension p, désigné par $\gamma(\mathcal{L})$, est défini par la relation*

$$\gamma(\mathcal{L}) = \left(\frac{\lambda_1(\mathcal{L})}{(\det \mathcal{L})^{1/p}}\right)^2.$$

Le carré dans la définition est là pour des raisons historiques, afin de garantir que $\gamma(\mathcal{L})$ est un nombre rationnel quand les réseaux sont entiers. Le théorème de Minkowski (théorème 1.3) établit un majorant pour les défauts d'Hermite $\gamma(\mathcal{L})$ associés à des réseaux $\mathcal{L}$ de dimension p et montre l'inégalité

$$\gamma(\mathcal{L}) \leq p, \qquad \text{pour tout réseau } \mathcal{L} \text{ de dimension } p.$$

Il existe deux interprétations du défaut d'Hermite. D'abord, la quantité $\sqrt{\gamma(\mathcal{L})}$ est égale au rapport entre le le plus petit côté possible d'un parallélotope fondamental $\mathcal{Q}$, et le côté de l'hypercube $\mathcal{H}$ de dimension p et de volume $\det \mathcal{L}$. Les réseaux qui ont un grand défaut d'Hermite sont des réseaux qui sont suffisamment "réguliers", en ce sens qu'ils admettent un parallélotope

fondamental dont la forme "ressemble" à celle d'un hypercube. C'est aussi pourquoi $\gamma(\mathcal{L})$ joue un rôle important dans le problème des empilements réguliers de sphères : comment disposer des sphères identiques dans l'espace de dimension p afin que la densité du remplissage soit maximale et que les centres des sphères sont disposés sur un réseau ? L'empilement le plus dense est achevé par un réseau maximisant le défaut d'Hermite. La borne supérieure du défaut d'Hermite pour une dimension p fixée reçoit le nom *constante d'Hermite* et est définie par

$$\gamma_p = \sup\{\gamma(\mathcal{L}); \quad \dim(\mathcal{L}) = p\}.$$

Les réseaux de dimension p pour lesquels $\gamma(\mathcal{L}) = \gamma_p$ sont appelés des *réseaux critiques*. Pour plus d'information à cet égard, on pourra consulter [16] et [52].

1.1.6 Forme normale d'Hermite.

Puisqu'un réseau admet une infinité de bases, qui n'ont pas du tout la même forme, il peut être utile, pour comparer les réseaux entre eux, par exemple, de déterminer une forme normale pour les bases possibles de ces réseaux. La forme normale la plus employée est la forme normale d'Hermite, que nous décrivons maintenant dans le cas où les réseaux sont de dimension pleine $n = p$.

Définition 1.8. *Une base $B := (b_{i,j}) \in \mathbb{R}^{n\times n}$ d'un réseau de dimension n dans $\mathbb{R}^n$ est* sous forme normale d'Hermite *lorsqu'elle s'exprime sous forme triangulaire dans la base canonique de $\mathbb{R}^n$. Plus précisément,*

(*i*) *B est triangulaire inférieure*

(*ii*) *Les éléments de la diagonale $b_{i,i}$ sont tous strictement positifs.*

(*iii*) *Pour tout $j < i$, on a les inégalités suivantes : $0 \leq b_{i,j} < b_{i,i}$.*

Proposition 1.6. *Tout réseau $\mathcal{L}$ de $\mathbb{R}^n$ de dimension pleine admet une unique base sous forme normale d'Hermite.*

La forme normale d'Hermite permet donc de représenter un réseau de manière canonique, dans un sens "algébrique". En revanche, il n'est pas du tout clair (et de fait c'est rarement le cas) qu'une telle base possède des propriétés euclidiennes intéressantes.

1.2 Problèmes algorithmiques de base.

Dans cette section, nous présentons les problèmes algorithmiques qui se posent naturellement dans l'étude des réseaux, quand ils sont considérés "pour eux-mêmes", indépendamment de leurs applications potentielles. Ces problèmes admettent des énoncés de nature diverse : ensembliste, algébrique, ou euclidien.

1.2.1 Représentation des réseaux

La plupart des problèmes étudiés dans ce chapitre ont un sens quand les réseaux sont des réseaux quelconques de $\mathbb{R}^n$. Les algorithmes qui les résolvent sont aussi le plus souvent bien définis sur des réseaux quelconques de $\mathbb{R}^n$, à condition de définir un modèle de calcul sur les nombres réels.

Mais, si on veut faire de l'algorithmique et étudier la complexité de ces problèmes, il faut considérer une notion de taille d'entrée. Un réseau est donné le plus souvent par une base, parfois

seulement par un système générateur, formé d'éléments de $\mathbb{Q}^n$, qu'on peut toujours ramener dans $\mathbb{Z}^n$ en multipliant tous les rationnels par le ppcm de leurs dénominateurs. On peut alors définir une notion de taille d'entrée. La taille d'un tel système $B = (b_1, \ldots, b_p)$ de p vecteurs de $\mathbb{Z}^n$ est choisie comme étant

$$\tau(B) = \Theta(pn) \cdot \log M \qquad \text{où} \quad M = \max\{b_{i,j}; \quad i \in [\![1, p]\!],\ j \in [\![1, n]\!]\}. \tag{1.4}$$

Remarquons qu'il y a deux composantes dans cette taille d'entrée : la composante qui dépend des deux dimensions n, p, dont le produit mesure le nombre des coefficients de la matrice d'entrée B et la composante $\log M$ qui dépend de la taille des coefficients de cette matrice. Un algorithme polynomial devra être polynomial en chacune de ces deux "parties".

Pour l'instant, nous étudions la complexité des problèmes et les données d'entrée sont donc a priori entières (ou, comme nous l'avons dit, rationnelles.)

1.2.2 Problèmes ensemblistes

Ce sont ceux qu'on peut se poser pour n'importe quelle famille d'ensembles.

Problème 1.1 (Appartenance)**.** Étant donné un système $B \in \mathbb{Z}^{p\times n}$ et un vecteur $t \in \mathbb{Z}^n$, décider si $t \in \mathcal{L}(B)$.

Problème 1.2 (Inclusion)**.** Étant donnés deux systèmes $B_1, B_2 \in \mathbb{Z}^{p\times n}$, décider si $\mathcal{L}(B_1) \subseteq \mathcal{L}(B_2)$.

Problème 1.3 (Intersection)**.** Étant donnés deux bases $B_1, B_2 \in \mathbb{Z}^{p\times n}$, trouver une base pour le réseau $\mathcal{L}(B_1) \cap \mathcal{L}(B_2)$.

Problème 1.4 (Union)**.** Étant donnés deux systèmes $B_1, B_2 \in \mathbb{Z}^{p\times n}$, trouver une base pour le plus petit réseau (au sens de l'inclusion) contenant $\mathcal{L}(B_1) \cup \mathcal{L}(B_2)$.

1.2.3 Problèmes algébriques

Ils sont souvent très semblables aux précédents, mais leur énoncé met l'accent sur l'aspect algébrique.

Problème 1.5 (Calcul de base)**.** Étant donné un système $B \in \mathbb{Z}^{p\times n}$, déterminer une base pour $\mathcal{L}(B)$.

Problème 1.6 (Équivalence de bases)**.** Étant donnés deux systèmes $B_1 \in \mathbb{Z}^{p\times n}$ et $B_2 \in \mathbb{Z}^{q\times n}$, décider si ils engendrent le même réseau.

Problème 1.7 (Calcul de la forme normale d'Hermite)**.** Étant donné un système $B \in \mathbb{Z}^{p\times n}$, déterminer la forme normale d'Hermite pour $\mathcal{L}(B)$.

Problème 1.8 (Noyau entier)**.** Étant donnée une matrice $A \in \mathbb{Z}^{n\times p}$, déterminer une base pour le réseau $\{x \in \mathbb{Z}^p \mid Ax = 0\}$.

1.2.4 Problèmes euclidiens

Ces problèmes ont un énoncé qui fait intervenir la structure euclidienne. Nous rappelons que le $\mathbb{R}$–espace vectoriel engendré par un système B est désigné par $\langle B \rangle$.

Problème 1.9. [Vecteur le plus court, SVP, Décision] Étant donnés un système $B \in \mathbb{Z}^{p \times n}$, et un entier K, existe-t-il un vecteur non nul v du réseau $\mathcal{L}(B)$ qui satisfait $||v|| \leq K$?

Problème 1.10 (Vecteur le plus court, SVP, Calcul)**.** Étant donné un système $B \in \mathbb{Z}^{p \times n}$, déterminer un vecteur non nul le plus court du réseau $\mathcal{L}(B)$.

Problème 1.11 (Vecteur le plus proche, CVP, Décision)**.** Étant donnés un système $B \in \mathbb{Z}^{p \times n}$, un vecteur $t \in \langle B \rangle \cap \mathbb{Q}^n$, et un entier K, existe-t-il $x \in \mathcal{L}(B)$ vérifiant $||t - x|| \leq K$.

Problème 1.12 (Vecteur le plus proche, CVP, Calcul)**.** Étant donnés un système $B \in \mathbb{Z}^{p \times n}$, et un vecteur $t \in \langle B \rangle \cap \mathbb{Q}^n$, déterminer $x \in \mathcal{L}(B)$ tel que pour tout $y \in \mathcal{L}(B)$, on ait vérifiant $||t - x|| \leq ||t - y||$.

Les problèmes algébriques et ensemblistes sont algorithmiquement "faciles" , en ce sens qu'ils admettent des algorithmes qui les résolvent en temps polynomial déterministe, tandis que les problèmes euclidiens sont "difficiles" dans le sens de la théorie de la complexité (ils sont NP–complets ou "proches" de problèmes qui le sont). Décrivons cette dichotomie.

1.2.5 Les algorithmes de résolution pour les problèmes ensemblistes ou algébriques.

Ces problèmes se résolvent, eux, avec des outils d'algèbre linéaire, et ils sont donc de complexité polynomiale, pourvu que les entiers gardent une croissance polynomiale au cours de l'algorithme. Par exemple, l'appartenance à un réseau se résout via un système linéaire, dont on vérifie que la solution est entière. L'inclusion se résout en vérifiant l'appartenance des vecteurs de la première base au réseau engendré par la deuxième. L'intersection de deux réseaux se calcule à l'aide du réseau dual. Le calcul du plus petit réseau calculant l'union revient à calculer une base pour le réseau engendré par l'union des bases des réseaux. Le calcul de la base et le problème de l'équivalence se résolvent en calculant la forme normale d'Hermite.Ce dernier problème se résout lui-même avec des outils d'algèbre linéaire.

Mais, les solutions que nous venons de décrire ne conduisent pas toujours à des algorithmes en temps polynomial en la taille d'entrée. Il faut prouver que la taille des entiers utilisés dans ces algorithmes a une croissance polynomiale, par rapport à la taille de l'entrée, donc à la fois par rapport à $\log M$, et aux dimensions n, p. Et ce n'est pas toujours le cas... Cela dépend souvent de la qualité du système d'entrée, la qualité se mesurant par des critères euclidiens.

1.2.6 La difficulté des problèmes euclidiens.

Dans ce cas, la solution est toujours reliée à la recherche plus ou moins exhaustive dans des boules, qui contiennent un nombre de points entiers qui croit exponentiellement avec la dimension. C'est pour cela que ces problèmes sont "difficiles".

Nous rappelons des définitions de la théorie de la complexité de manière informelle. Un problème de décision est dans P s'il existe un algorithme qui le résout en temps polynomiale dans la taille de l'entrée. Un problème de décision est dans NP, lorsqu'il est possible de vérifier une réponse positive en temps polynomial. Une réduction randomisée inversée transforme toujours une instance négative en une instance négative, alors qu'elle transforme une instance positive en

une instance positive avec une probabilité p, qui vérifie $(1-p) \leq n^{-c}$, où n est la taille de l'entrée et c est une constante.

Nous rappelons maintenant les principaux résultats de complexité obtenus sur ces problèmes euclidiens. La NP-complétude de CVP a été établie en 1981 par van Emde Boas, qui a conjecturé dans le même article [81] la NP-complétude de SVP. Mais il a fallu attendre jusqu'en 1996 pour que la complexité de SVP soit aussi élucidée, et ce, sous un certain type de réduction randomisée, plus faible que les réductions déterministes usuelles.

Théorème 1.4 (Van emde Boas, [81])**.** *Le problème de décision CVP est un problème NP-complet.*

Théorème 1.5 (Ajtai, [2])**.** *Le problème de décision SVP est un problème NP-complet sous des réductions randomisées inversées.*

Existe-t-il un lien non-trivial entre la complexité de ces problèmes ? On pourrait penser au premier abord que CVP est plus dur que SVP, puisque la résolution de SVP se ramène à la résolution de CVP en choisissant $t = 0$. Mais, comme tout réseau contient 0, la réponse est donc triviale. SVP n'est donc pas une une version "homogène" de CVP. Une manière relativement triviale de réduire un problème à l'autre serait d'utiliser la NP-complétude, en essayant de simplifier les preuves déjà existantes. D'après Micciancio [55], une telle réduction ne serait pas intéressante car elle réduirait une instance de SVP à n instances de SVP. Micciancio lui-même fournit une réduction de SVP à CVP, en montrant donc que SVP n'est pas plus dur que CVP.

Par ailleurs, il faut aussi revenir sur le résultat d'Ajtai reliant le pire des cas avec le cas moyen pour SVP [1]. Selon les notes de Micciancio et Goldwasser [55] qui ont bien élucidé l'article original très technique, Ajtai a montré l'assertion suivante : S"il n'y a pas d'algorithme polynomial résolvant pour tout réseau le problème de décision SVP approximé à n'importe quel facteur polynomiale près, alors le problème calculatoire SVP est également dur à résoudre exactement lorsque le réseau est choisi aléatoirement selon une distribution facile à construire. Ce résultat a eu beaucoup d'impact en cryptographie, car la cryptographie cherche justement à fonder la sécurité des cryptosystèmes sur les instances *dures* d'un problème. Donc, si on sait qu'une instance obtenue en échantillonant une certaine distribution est très probablement dure, alors il y a un moyen de trouver des instances difficiles sur lesquelles on peut construire les cryptosystèmes de façon efficace.

La difficulté de SVP et de CVP fait des réseaux euclidiens une source potentielle de constructions de cryptosystèmes. Des tels systèmes existent, par exemple GGH, Ajtai-Dwork et NTRU. Jusque là, le seul à avoir survécu les épreuves du temps a été NTRU. Les autres ont tous été cassés, au moins dans la pratique [59]. Enfin, avec l'éventuelle entrée en scène du calcul quantique, les réseaux euclidiens semblent pour l'instant bien placés pour prendre le relais des cryptosystèmes fondés sur la difficulté de la factorisation d'entiers et du logarithme discret. À ce propos, Micciancio et Regev [56] expliquent qu'il n'y a pas encore d'algorithmes quantiques résolvant des problèmes des réseaux qui se comportent significativement mieux que les algorithmes classiques bien connus, mais qu'il faut rester prudent[50].

1.2.7 Le problème de la réduction.

Ainsi, il existe une dichotomie entre les deux classes de problèmes (Problèmes algébriques "faciles", problèmes euclidiens "difficiles"). Mais, nous avons aussi insisté sur le fait que les problèmes algébriques ne sont faciles que lorsqu'on est assuré d'une croissance polynomiale de la taille des entiers. Par ailleurs, pour la classe de problèmes euclidiens, ce n'est pas parce qu'ils

sont difficiles qu'on ne cherche pas à obtenir des algorithmes qui les résolvent de la façon la moins inefficace possible. Dans les deux cas, les algorithmes vont être bien plus efficaces si la base d'entrée possède déjà de bonnes qualités euclidiennes ; on dit alors qu'elle est réduite, avec la définition informelle suivante :

Définition 1.9 (Définition informelle de la notion de base réduite)**.** *Une base $B = (b_1, \ldots, b_p)$ formée de p vecteurs de $\mathbb{R}^n$ est une base réduite si elle est formée de vecteurs assez courts et assez orthogonaux. Ces critères se mesurent quantitativement par une majoration du défaut d'orthogonalité $\rho(B)$ et des défauts de longueur $\theta_i(B)$ définis respectivement par*

$$\rho(B) = \frac{1}{\det \mathcal{L}(B)} \prod_{i=1}^{p} ||b_i|| = \prod_{i=1}^{p} \frac{||b_i||}{||b_i^\star||}, \qquad \theta_i(B) = \frac{||b_i||}{\lambda_i(\mathcal{L}(B))}\,.$$

Le défaut d'orthogonalité $\rho(B)$ est au moins égal à 1, avec égalité seulement lorsque la base B est orthogonale. Comme un réseau ne possède pas en général de base orthogonale, on a en général $\rho(B) > 1$. La base est "assez" orthogonale lorsque son défaut d'orthogonalité $\rho(B)$ est majoré. Les défauts de longueur sont aussi au moins égaux à 1. Les égalités $\theta_i(B) = 1$ ne peuvent se produire simultanément que lorsque la base est minimale, i.e., formée par des vecteurs réalisant les minimas successifs. L'existence d'une base minimale n'est pas toujours garantie, dès que la dimension p vérifie $p \geq 5$.

Nous donnons maintenant une définition informelle de la réduction :

Définition 1.10 (Réduction)**.** *Étant donnée une base B, réduire B consiste à trouver une base équivalente et réduite.*

Une bonne notion de réduction doit établir un compromis entre la qualité de la base réduite et la complexité de l'algorithme de réduction. Un tel compromis est atteint par l'algorithme LLL, inventé par Lenstra, Lenstra et Lovász en 1982, présenté dans le chapitre suivant, et qui sera un des objets centraux de cette thèse.

Théorème 1.6 (LLL)**.** *Considérons un réel $s > 2/\sqrt{3}$. L'algorithme LLL construit à partir d'une base $B := (b_1, \ldots, b_p)$ de taille $\tau(B)$ (définie en (1.4)), une base $\hat{B}$ avec les caractéristiques suivantes*

(i) la base $\hat{B}$ est obtenue en temps polynomial en la taille $\tau(B)$ de la matrice B.

(ii) Le défaut d'orthogonalité $\rho(\hat{B})$ et les défauts de longueur $\theta_i(\hat{B})$ de la base $\hat{B}$ (définis dans la définition 1.9) satisfont

$$\rho(\hat{B}) \leq s^{p(p-1)/2} \qquad \theta_i(\hat{B}) \leq s^{p-1}$$

1.2.8 Stratégie générale de la résolution des problèmes.

Finalement, avec une notion adéquate de la réduction, une stratégie efficace pour résoudre tous les problèmes cités (ensemblistes, algébriques ou euclidiens) sera la suivante :

Problème $\Pi(B)$
$\widehat{B} :=$ **Réduction** (B) ;
$\Pi(\widehat{B})$;

Cette stratégie sera efficace si la perte de temps consacré à réduire B en $\hat{B}$ est compensée par le gain de temps mis à résoudre Π sur $\hat{B}$ plutôt que sur B.

Cette thèse est centrée sur ce problème de réduction, et sur la solution que l'algorithme LLL y apporte. Nous allons y revenir en détail dans les chapitres suivants. Mais, auparavant, nous expliquons en quoi la réduction ne permet pas seulement de résoudre les problèmes de la théorie "interne" des réseaux, mais pourquoi elle s'avère essentielle dans beaucoup de domaines de l'informatique mathématique, extérieurs a priori aux réseaux.

1.3 Problèmes algorithmiques résolus via les réseaux.

Les réseaux sont un outil de modélisation très puissant. Quand on rencontre un problème discret additif, le réflexe premier consiste à se poser la question : Y-a-t-il un réseau dessous ? Très souvent, la réponse est positive. Dans cette section nous passons en revue un certain nombre de problémes qu'on a pu modéliser par les réseaux, et où la réduction du réseau a permis de résoudre, la plupart du temps, le problème. Un certain nombre d'exemples sont issus de la cryptanalise, et d'autres de la théorie algorithmique des nombres, ou du calcul formel. Chacun de ces exemples fait intervenir une base dont la forme est bien particulière à la problématique sous-jacente. Cela doit être pris en compte à l'heure de définir ce qu'est une base *aléatoire* : on retournera à ce point dans la section 3.2 du chapitre 3.

1.3.1 Factorisation de polynômes (1).

On cherche à factoriser un polynôme $f \in \mathbb{Z}[X]$ de degré n et de norme $M = ||f||_\infty = \max |f_i|$ pour lequel on a une bonne approximation d'une racine α. On peut chercher alors à déterminer le polynôme minimal h du nombre algébrique α, qui par définition est un facteur irréductible de f. Une des premières applications de l'algorithme LLL est la solution de ce problème lorsque α est donnée par ses approximations p–adiques.

1.3.2 Factorisation de polynômes (2).

Mais on peut aussi connaître α par ses approximations complexes, obtenues par exemple avec l'algorithme de Newton, comme dans [33]. L'idée centrale est fondée sur un principe de séparation qui affirme que : il existe δ dont la taille est polynomiale en la taille de $(n, \log M)$, tel que les deux propositions suivantes sont équivalentes :

(i) Le polynôme g est multiple du polynôme minimal h de α,

(ii) Le polynôme g vérifie $|g(\bar{\alpha})| \leq \delta$.

Il s'agit alors de chercher un polynôme g vérifiant (ii). On montre que cela revient à trouver un vecteur assez court du réseau engendré par les lignes de la matrice

$$\begin{bmatrix} c & 0 & \cdots & 0 & \Re(\bar{\alpha}^0) & \Im(\bar{\alpha}^0) \\ 0 & c & \cdots & 0 & \Re(\bar{\alpha}^1) & \Im(\bar{\alpha}^1) \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & c & \Re(\bar{\alpha}^m) & \Im(\bar{\alpha}^m) \end{bmatrix}, \tag{1.5}$$

où la constante c (très petite) dépend polynomialement de M et de δ, et $m \leq n$ est le degré *supposé* du polynôme minimal de α. En effet, si $\overline{g} = (g_0, g_1, \ldots, g_n)$ désigne un vecteur court du réseau, et si g est le polynôme dont le vecteur de coefficients est $\overline{g}$, alors la norme euclidienne de $\overline{g}$ vérifie

$$||\overline{g}||^2 = |g(\bar{\alpha})|^2 + c^2||g||^2, \qquad \text{avec} \quad \left\| \sum_{i=0}^{m} a_i X^i \right\|^2 := \sum_{i=0}^{m} a_i^2$$

On voit alors que si c et $||\overline{g}||$ sont petits, alors $|g(\bar{\alpha})|^2$ l'est aussi. Le bon choix de c permet de conclure que $|g(\bar{\alpha})| \leq \delta$, et donc que nous avons trouvé un multiple de h. Il ne reste qu'à calculer le plus grand commun diviseur entre f et g pour trouver un facteur de f.

1.3.3 Approximations diophantiennes simultanées.

Il s'agit de résoudre le problème suivant :

Etant donné un n-uplet $(\alpha_1, \alpha_2, \cdots, \alpha_n)$, trouver n nombres entiers $(p_1, p_2, \cdots, p_n)$ et un nombre entier q tels que les n rationnels (p_i/q) (pour $i \in [1..n]$) forment une bonne approximation simultanée des nombres donnés α_i.

Une réponse, non constructive, à cette question a été donnée par Dirichlet [20], fondée sur le théorème de Minkowski :

Théorème 1.7 (Dirichlet)**.** *Pour tout $n \geq 1$, pour tout n-uplet $(\alpha_1, \alpha_2, \cdots, \alpha_n)$, et pour tout couple (ϵ, Q) vérifiant $\epsilon > 0$ et $Q \geq \epsilon^{-n}$, il existe des entiers $(p_1, p_2, \cdots, p_n)$ et un entier q vérifiant*

$$0 < q \leq Q \quad et \quad |q\alpha_i + p_i| < \epsilon \quad pour\ tout\ i,\ 1 \leq i \leq n.$$

Pour obtenir une version constructive de ce théorème, Lenstra, Lenstra et Lovász [46], puis Lagarias [39] considèrent le réseau engendré par les lignes $v_1, \ldots, v_{n+1}$ de la matrice

$$\begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \\ \alpha_1 & \alpha_2 & \cdots & \alpha_n & \epsilon/Q. \end{bmatrix}. \tag{1.6}$$

Un vecteur court v du réseau, qui a dans cette base $v_1, \ldots, v_{n+1}$ les composantes $(p'_1, p'_2, \cdots, p'_n, q')$, avec $q' > 0$, vérifie donc

$$||v|| = \left|\left| \sum_{i=1}^{n} p'_i v_i \right|\right| < \epsilon.$$

Alors, les entiers $(p'_1, p'_2, \cdots, p'_n, q')$ fournissent une assez bonne approximation du n-uplet considéré $(\alpha_1, \alpha_2, \cdots, \alpha_n)$, puisque

$$|q'\alpha_i + p'_i| < \epsilon \quad \text{et} \quad q' < Q.$$

La possibilité de trouver un vecteur si court, et surtout, de pouvoir le trouver efficacement est liée au choix de Q. À l'aide de l'algorithme LLL, qui fournit un vecteur dont la longueur est au plus un facteur $s^{(n-1)/2}$ du premier minimum (pour $s > 2/\sqrt{3}$), Lagarias montre le théorème suivant :

Théorème 1.8. *Pour tout n, pour tout n-uplet $(\alpha_1, \alpha_2, \cdots, \alpha_n)$ de taille M, et pour tout couple (ϵ, Q) vérifiant $\epsilon > 0$ et $Q \geq s^{n(n+1)/4}\epsilon^{-n}$, on peut construire en temps polynomial en $(\log M, n)$ des entiers $(p'_1, p'_2, \cdots, p'_n)$ et un entier q' vérifiant*

$$0 < q' \leq Q \quad et \quad |q'\alpha_i + p'_i| < \epsilon \quad pour\ tout\ i,\ 1 \leq i \leq n.$$

1.3.4 Cryptanalyse des systèmes cryptographiques fondés sur le sac-à-dos.

Ce cryptosystème est fondé sur la difficulté du problème du *sac-à-dos* :

Etant donnés n entiers positifs $(a_i)_{1\leq i\leq n}$ *–les paquets– et un entier* s *–le sac–, trouver un élément* $X = (x_i)_{1\leq i\leq n}$ *de* $\{0,1\}^n$*, solution de l'équation*

$$\sum_{i=1}^{n} a_i x_i = s,$$

si une telle solution existe, sinon indiquer qu'il n'y a pas de solution.

Ce problème est NP-complet en général, mais il est facile lorsque la suite des a_i est *super-croissante*, c'est-à-dire lorsque

$$\sum_{j=1}^{i-1} a_j < a_i$$

pour tout $2 \leq i \leq n$. Il suffit alors d'enlever à s successivement les a_i ordonnés de façon décroissante. La solution que l'on trouve ainsi est unique.

Le principe du cryptosystème de Merkle-Hellman est le suivant : la réceptrice, Alice, chosit comme clé publique la suite d'entiers $(a_i)_{1\leq i\leq n}$. Si Bob veut envoyer un message $(x_i)_{1\leq i\leq n}$ à Alice, il envoie la somme $s = \sum_{i=1}^{n} a_i x_i$. Alice doit alors trouver le message caché dans s. Bien entendu, si la suite est super-croissante, tout le monde pourra décoder, sinon, personne, même pas Alice, ne pourra le faire. La solution trouvée par Merkle et Hellman [54] consiste à appliquer la transformation $a \mapsto ca \mod m$ à une suite $(a_i)_{1\leq i\leq n}$ super-croissante, en la transformant en une suite d'apparence quelconque. Alice doit alors garder $c^{-1} \mod m$ et m en tant que clé secrète. Lorsqu'elle recevra une somme s, elle pourra alors décoder la somme $c^{-1}s \mod m$.

Ce cryptosystème a néanmoins été cassé par Shamir [66]. Puis, Lagarias et Odlyzko [40], en formulant le problème en termes de réseaux, ont montré que la plupart des problèmes de sac à dos étaient solubles pourvu que la densité du sac à dos soit suffisamment faible. On considère le réseau $\mathcal{L}$ engendré par les lignes $v_1, v_2, \ldots, v_{n+1}$ de la matrice

$$\begin{bmatrix} 1 & 0 & \cdots & 0 & -a_1 \\ 0 & 1 & \cdots & 0 & -a_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -a_n \\ 0 & 0 & \cdots & 0 & s \end{bmatrix}. \tag{1.7}$$

associée à une entrée $((a_i)_{1\leq i\leq n}, s)$du sac à dos. Si une suite $(x_i)_{1\leq i\leq n}$ est une solution du sac-à-dos, alors le vecteur

$$v = \sum_{i=1}^{n} x_i v_i + v_{n+1} = (x_1, \ldots, x_n, 0)$$

est un vecteur du réseau $\mathcal{L}$ de longueur $||v|| \leq \sqrt{n}$, qui est petite devant la taille souvent beaucoup plus grande des a_i. Avec un peu de chance (chance qui est grande pour les sac-à-dos de Merkle et Hellmann), ce vecteur solution est un vecteur le plus court du réseau, et il est tellement plus court que les autres qu'il peut se trouver même avec un algorithme d'approximation comme LLL.

1.3.5 Prédictibilité de la suite de bits produits par le générateur congruentiel linéaire

Un des générateurs pseudo-aléatoires les plus célèbres est sans doute le générateur linéaire congruentiel. On choisit un module m et un multiplicateur a, premier avec m, et une donnée x_1 de départ ; puis on considère la suite (x_i) définie par

$$x_{i+1} = a\, x_i \mod m.$$

Stern a montré [72], en améliorant les résultats de Frieze [25] que, même si aucun des paramètres a, m ni x_0 n'est connu, la suite y_i formée par une proportion assez grande des bits de poids fort des x_i est prédictible et donc que le générateur n'est pas cryptographiquement sûr. On travaille dans les réseaux X et Y engendrés respectivement par les vecteurs :

$$u_i = \begin{pmatrix} x_{i+1} - x_i \\ x_{i+2} - x_{i+1} \\ x_{i+3} - x_{i+2} \end{pmatrix} \qquad \text{et} \qquad v_i = \begin{pmatrix} y_{i+1} - y_i \\ y_{i+2} - y_{i+1} \\ y_{i+3} - y_{i+2} \end{pmatrix}.$$

Les k premiers vecteurs v_i étant donnés, on peut cherche une relation linéaire entière courte entre eux de la forme

$$\sum_{i=1}^{k} \lambda_i v_i = 0.$$

On en déduit que le vecteur de mêmes coefficients λ_i dans X est un vecteur si court .... qu'il est donc nul. Cela suppose que le réseau X soit assez "régulier", c'est-à-dire que le plus court vecteur du réseau ne soit pas trop court. Or, de manière informelle, la "plupart" des réseaux sont "assez" réguliers Si k est bien choisi en fonction de la taille présumée des données, on construit ainsi un polynôme P dont les coefficients sont les λ_i vérifiant $P(a) \equiv 0 \mod m$.

Si on réitère cette construction, on détermine ainsi une suite de l polynômes P_j qui appartiennent tous à un réseau $\mathcal{L}$ de base

$$q_0(t) = m \quad \text{et} \quad q_i(t) = t^i - a^i \quad \text{pour } i,\ 1 \le i \le k.$$

Le déterminant de ce réseau $\mathcal{L}$ est justement le nombre m cherché. Si l'on trouve le déterminant $\hat{m}$ du réseau engendré par les P_j. Le nombre $\hat{m}$ est un multiple de m qui décroit très rapidement quand l augmente ; on déduit donc la valeur de m puis ensuite une valeur très probable de a obtenue en cherchant un polynôme de premier degré dans le réseau $\mathcal{L}$.

1.3.6 Calcul de racines k-ièmes modulo n

Le problème général s'énonce ainsi :

Soient deux entiers n et $k \ge 2$. Eant donnés deux entiers x_0 et y_0, un voisinage I de x_0, un voisinage J de y_0 qui contiennent respectivement un point $x \in I$ et un point $y \in J$ vérifiant $x^k \equiv y \mod n$, on veut trouver x et y.

On veut donc deviner un couple (u, v) de *petits* entiers, solutions de l'équation $(x_0 + u)^k \equiv y_0 + v \mod n$ qui se développe en

$$x_0^k + \binom{k}{1} x_0^{k-1} u + \cdots + \binom{k}{i} x_0^{k-i} u^i + \cdots + \binom{k}{k-1} x_0 u^{k-1} + u^k - v \equiv y_0 \mod n. \qquad (1.8)$$

On pose $w_i = u^i$ pour tout i, $\quad 0 \leq i \leq k-1 \quad$ et aussi $w_k = y_0 + v - u^k$, et on travaille dans le réseau $\mathcal{L}$ des vecteurs $w = (w_0, w_1, \ldots, w_k)$ de $\mathbb{Z}^{k+1}$ vérifiant

$$\sum_{i=0}^{k-1} \binom{k}{i} x_0^{k-i} w_i - w_k \equiv 0 \mod n.$$

Le réseau $\mathcal{L}$ de déterminant n et rang $k+1$ a pour matrice

$$\begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \\ x_0^k & \binom{k}{1} x_0^{k-1} & \cdots & \binom{k}{k-1} x_0 & n \end{bmatrix}. \tag{1.9}$$

On cherche alors un point w du réseau $\mathcal{L}$ qui est proche –en un sens à préciser– du point $(0, 0, \ldots, y_0)$. Si ce réseau est suffisamment "régulier", son premier minimum $\lambda_1(\mathcal{L})$ sera proche de la moyenne géométrique des minima successifs, de l'ordre de $n^{2/(k+1)}$. Or, on peut montrer que la plupart des réseaux de ce type sont "réguliers" ; dans ce cas, l'unicité du point le plus proche permet d'affirmer que le point w trouvé par un algorithme d'approximation avec un facteur d'approximation adéquat donnera naissance au triplet (u_1, u_2, v) cherché.

1.3.7 Méthode de Coppersmith

On veut résoudre le problème suivant, très proche de celui de la section précédente :

Retrouver en temps polynomial une racine d'un polynôme modulaire quand on en connaît une fraction de ses bits.

La méthode de Coppersmith [17] permet d'obtenir le résultat suivant :
Soient $p(x)$ un polynôme de degré δ, un entier N de factorisation inconnue, et une borne $X = (1/2)N^{1/\delta - \epsilon}$. Alors, on peut trouver en temps polynomial en $(\log N, \delta, 1/\epsilon)$ toutes les racines x_0 de $p(x) = 0 \mod N$, qui vérifient de plus $|x_0| < X$.

Considérons un polynôme unitaire de la forme

$$p(x) = x^\delta + a_{\delta-1} x^{\delta-1} + \cdots + p_2 x^2 + p_1 x + p_0 = 0 \ (\text{mod } N),$$

et supposons qu'il possède une racine x_0 modulo N vérifiant en plus $|x_0| < X$. On cherche x_0. Tout d'abord on choisit un entier h tel que

$$h \geq \max\left(\frac{\delta - 1 + \epsilon}{\epsilon \delta^2}, \frac{7}{\delta}\right).$$

Et on construit une famille de polynômes qui admettent aussi x_0 comme racine : pour chaque paire (i, j) d'entiers vérifiant $0 \leq i < \delta, 1 \leq j < h$, on considère le polynôme

$$q_{ij}(x) = x^i p(x)^j, \qquad \text{qui vérifie} \quad q_{ij}(x_0) = 0 \ (\text{mod } N^j).$$

Puis on considère la matrice M suivante, carrée d'ordre $2h\delta - \delta$, construite par blocs :

(*i*) Dans la partie supérieure droite, de taille $(h\delta) \times (h\delta - \delta)$, les lignes sont indexées par un entier g tel que $0 \leq g \leq h\delta$, et les colonnes indexées par $\alpha(i, j) = h\delta + i + (j-1)\delta$ avec $0 \leq i < \delta$ et $1 \leq j < h$, de sorte que $h\delta \leq \alpha(i, j) < 2h\delta - \delta$. L'élément de la matrice M en $(g, \alpha(i, j))$ est le coefficient de x^g dans le polynôme $q_{ij}(x)$.

(*ii*) Le bloc inférieur droit est une matrice diagonale $(h\delta - \delta) \times (h\delta - \delta)$, avec la valeur N^j dans chaque colonne $\alpha(i, j)$.

(*iii*) Le bloc supérieur gauche, $(h\delta) \times (h\delta)$ est aussi une matrice diagonale, dont la valeur dans la ligne g est une approximation rationale à $X^{-g}/\sqrt{h\delta}$, où $X = N^{1/\delta - \epsilon}$ est une borne supérieure pour la solution cherchée.

(*iv*) Enfin, le bloc inférieur gauche est nul.

Suivant Coppersmith, nous illustrons cette matrice dans le cas (artificiel) $h = 3$, $\delta = 2$, avec un polynôme p de la forme $p(x) = x^2 + ax + b$ et $p(x)^2 = x^4 + cx^3 + dx^2 + ex + f$. On pose également $\tau = 1/\sqrt{h\delta}$. Nous avons

$$M = \begin{bmatrix} \tau & 0 & 0 & 0 & 0 & 0 & b & 0 & f & 0 \\ 0 & \tau X^{-1} & 0 & 0 & 0 & 0 & a & b & e & f \\ 0 & 0 & \tau X^{-2} & 0 & 0 & 0 & 1 & a & d & e \\ 0 & 0 & 0 & \tau X^{-3} & 0 & 0 & 0 & 1 & c & d \\ 0 & 0 & 0 & 0 & \tau X^{-4} & 0 & 0 & 0 & 1 & c \\ 0 & 0 & 0 & 0 & 0 & \tau X^{-5} & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & N & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & N & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & N^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & N^2 \end{bmatrix}. \tag{1.10}$$

Les lignes de cette matrice engendrent un réseau. Pour comprendre ce réseau, il convient de distinguer dans la matrice M un côté gauche et un côté droit, selon que l'on se trouve dans les premières $h\delta$ colonnes ou au delà. De la même manière, nous parlons du côté gauche et droit d'un vecteur ligne, en entendant par là que l'on se refère aux $h\delta$ premières composantes ou à celles qui restent. Il existe, dans ce réseau, un vecteur s, relié de très près à la solution cherchée x_0. En effet, à partir du vecteur ligne r suivant,

$$r_g = x_0^g, \quad (\text{pour } g \leq h\delta) \qquad r_{\alpha(i,j)} = -x_0^i y_0^j, \qquad (\text{pour } 0 \leq i < \delta \text{ et } 1 \leq j < h), \tag{1.11}$$

le produit $s := rM$ définit un vecteur ligne dont les composantes sont (à gauche, et à droite)

$$s_g = \frac{(x_0/X)^g}{\sqrt{h\delta}} \quad (\text{pour } g \leq h\delta) \quad s_{\alpha(i,j)} = q_{ij}(x_0) - x_0^i y_0^j N^j = 0, \quad (\text{pour } 0 \leq i < \delta \text{ et } 1 \leq j < h).$$

Par ailleurs, puisque $x_0/X < 1$, la norme euclidienne de s vérifie

$$||s|| = \left[\sum_g s_g^2\right]^{1/2} < \left[\sum_g (\frac{1}{\sqrt{h\delta}})^2\right]^{1/2} = 1.$$

On observe aussi que ce vecteur s est un vecteur court du sous-réseau de $\mathcal{L}(M)$ (de dimension $h\delta$) engendré par le bloc supérieur gauche (complété par une matrice nulle à droite).

Pour trouver la solution x_0, on observe que, en termes de la matrice M et des vecteurs r' et s' avec $r'M = s'$, l'espace des vecteurs r' tels que s' a un côté droit nul est un espace de dimension $h\delta$. Par ailleurs, on montre que si en plus $|s'| < 1$, l'espace des vecteurs r' a encore une dimension en moins et c'est un espace de dimension $h\delta - 1$. Dans ce cas, on peut trouver une relation de dépendance linéaire entière entre les $h\delta$ composantes non nécessairement nulles

de r', de la forme $\sum c'_g r'_g = 0$. Comme le vecteur r de (1.11) est bien dans le cas précédent, on peut trouver une relation de dépendance linéaire, qui apporte un polynôme $C(x)$ vérifiant

$$C(x_0) = \sum c_g x_0^g = 0.$$

Ainsi, nous avons une équation polynomiale en $\mathbb{Z}$ dont la solution est x_0. Il suffit alors de résoudre cette équation, par exemple avec une suite de Sturm, pour trouver x_0.

1.3.8 Cryptosystème NTRU.

Même si le cryptosystème NTRU est fondé sur l'arithmétique sur des anneaux de polynômes, on peut aussi le considérer comme un système sur les réseaux. La donnée est un petit nombre premier q et un élément $h = (h_1, h_2, \ldots, h_n)$ de $\mathbb{Z}^n$ qui vérifie $h_i \in]-q/2, +q/2[$. On considère le réseau engendré par les lignes de la matrice A carrée $(2n \times 2n)$, définie par blocs et qui fait intervenir la matrice circulante $M_n(h)$ carrée $n \times n$ sous la forme suivante

$$A_(q,h) := \begin{bmatrix} qI_n & 0_n \\ M_n(h) & I_n \end{bmatrix}, \qquad \text{avec} \quad M_n(h) := \begin{bmatrix} h_1 & h_2 & h_3 & \cdots & h_n \\ h_n & h_1 & h_2 & \cdots & h_{n-1} \\ h_{n-1} & h_n & h_1 & \cdots & h_{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ h_2 & h_3 & h_4 & \cdots & h_1 \end{bmatrix}. \tag{1.12}$$

La clé publique est la matrice **A** elle même, tandis la clé privée est un vecteur court de ce réseau. Ainsi, la sécurité du système NTRU repose sur la difficulté de trouver un petit vecteur v dans le réseau. C'est l'objet du *challenge* de reconstruction de clés de NTRU [61] de chercher de tels vecteurs courts.

Nous avons présenté les réseaux et expliqué pourquoi le problème de la réduction des réseaux est fondamental. Le chapitre suivant est donc consacré au problème de la réduction des réseaux.

Chapitre 2

La réduction des réseaux

Sommaire

Ce chapitre est consacré au problème algorithmique de la réduction des réseaux. Nous présentons le problème ainsi que des algorithmes de réduction en dimension 1, en dimension 2 et en dimension n quelconque. En dimension 1, le problème de la réduction est d'une certaine manière trivial, mais la solution à une version plus générale du problème – comment trouver une base à partir d'un système générateur – est donnée par l'algorithme d'Euclide, qui peut donc être vu comme l'algorithme de réduction de la dimension 1. L'algorithme de la dimension 2 est l'algorithme de Gauss, optimal à tous points de vue, qui peut être considéré comme une généralisation en dimension 2 de l'algorithme d'Euclide. En dimension quelcoque n, nous présentons l'algorithme LLL, le plus célèbre algorithme de réduction des réseaux, ainsi qu'une de ses variantes, la version LLL-Impair-Pair, que nous utiliserons pour proposer une analyse. Chaque section contient un résumé historique.

2.1 Algorithmes de réduction en dimension 1

Un réseau euclidien $\mathcal{L}$ entier de dimension 1 dans l'espace ambiant $\mathbb{R}$ est de la forme $\mathbb{Z}c$, où $c \in \mathbb{N}_+$. Les seules bases de $\mathcal{L}$ sont alors $\{c\}$ et $\{-c\}$ et, dans ce cas, le problème de la réduction est trivial. Mais le problème devient plus intéressant lorsque le réseau entier $\mathcal{L}$, toujours de dimension 1 dans l'espace ambiant $\mathbb{R}$, n'est plus donné par une base, mais par un système générateur formé de deux entiers u et v. Le problème s'énonce alors ainsi : étant donné une paire d'entiers $(u, v) \neq (0, 0)$, trouver une base pour le réseau $\mathcal{L}(u, v)$ engendré par u et v La proposition suivante éclaire la nature du problème.

Proposition 2.1. *Soient u, v deux entiers vérifiant $(u, v) \neq (0, 0)$. Le réseau $\mathcal{L}$ engendré par la paire (u, v) a pour base le plus grand commun diviseur d des deux entiers u, et v,*

$$\mathcal{L}(u, v) = \mathbb{Z}d \qquad \textit{où} \quad d = \operatorname{pgcd}(u, v)$$

Démonstration. Remarquons d'abord que le plus petit entier c strictement positif de $\mathcal{L}(u, v)$ engendre ce réseau $\mathcal{L}(u, v)$. En effet, la division euclidienne par défaut d'un élément w de $\mathcal{L}(u, v)$ par c, de la forme $w = mc + r$ fournit un reste $r \in [0, c[$ qui est donc nécessairement nul par le choix de c.

Montrons maintenant que c est égal au pgcd d de u et v. Puisque u et v appartiennent à $\mathcal{L}$, c est un diviseur commun à u et v, et donc c divise d. D'autre part, comme c est un élément de $\mathcal{L}(u, v)$, il s'écrit $c = xu + yv$ avec $(x, y) \neq (0, 0)$, ce qui montre que d divise c ; finalement, on a $c = d$, comme on voulait montrer. □

Le problème de calcul de base se réduit donc au calcul du plus grand commun diviseur (pgcd), qui se fait faire par un *algorithme d'Euclide*. Pour des raisons qui apparaîtront plus clairement dans la suite, on travaillera dans cette thèse avec des algorithmes d'Euclide centrés, qui utilisent des divisions euclidiennes centrées, et qu'on présente ensuite.

2.1.1 L'algorithme d'Euclide

L'algorithme d'Euclide est décrit par Euclide lui-même dans le livre 7 de ses Éléments, paru autour de l'année 300 av. J.-C.. En suivant Knuth, on peut dire que l'algorithme d'Euclide est le grand-père de tous les algorithmes, puisque c'est le seul à survivre jusqu'à nos jours (exception peut se faire sur l'algorithme d'exponentiation binaire, qui est plus ancien, mais dont les utilisateurs avaient peu fait d'effort pour lui donner une forme polie.). On pense que l'algorithme était connu par d'autres anciens, comme Eudoxe, Aristote, etc. Pour plus d'information sur les origines de l'algorithme lui-même, on peut consulter [36].

2.1.2 Divisions euclidiennes centrées.

Il existe deux divisions euclidiennes centrées. Eles travaillent toutes deux avec l'entier de plus proche du rél x, désigné par $\lfloor x \rceil$ et défini par

Considérons une paire (u, v) avec $u, v \in \mathbb{Z}$ et $v \neq 0$. La division euclidienne centrée (dite non pliée) s'écrit

$$v = mu + r \qquad \text{avec} \quad m = \left\lfloor \frac{v}{u} \right\rceil \quad \text{et} \quad r \in \left]-\frac{u}{2}, +\frac{u}{2}\right]$$

tandis que la division euclidienne centrée (dite pliée) écrit

$$v = mu + \epsilon r \qquad \text{avec} \quad m = \left\lfloor \frac{v}{u} \right\rceil, \quad r \in \left[0, +\frac{u}{2}\right], \qquad \text{et} \quad \epsilon = \text{Signe}\left(\frac{v}{u} - \left\lfloor \frac{v}{u} \right\rceil\right)$$

2.1.3 Algorithmes d'Euclide centrés.

Un algorithme d'Euclide procède par divisions et échanges. Il existe ici deux algorithmes d'Euclide centrés, l'un non plié et l'autre plié. Chacun d'eux travaille avec des paires d'entiers (v, u), et calcule une division euclidienne de v par u, en lui associant un quotient q et un reste r de sorte que

(i) $v = mu + r$, avec $r \in]-u/2, u/2]$ pour l'algorithme dit *non-plié*,

(ii) $v = mu + \epsilon r$ avec $r \in [0, u/2]$ pour l'algorithme dit *plié*.

Dans tous les cas, l'algorithme poursuit son exécution avec la paire (u, r).

L'algorithme d'Euclide (figure 2.1) reçoit en entrée une paire d'entiers (u, v) avec $v \neq 0$, pose $u_0 = u, u_1 = v$, ou $w_0 = u, w_1 = v$, et effectue une suite de divisions euclidiennes centrées

$$\text{(non pliées)} \quad u_{i-1} = m_i\, u_i + u_{i+1} \qquad \text{(pliées)} \quad w_{i-1} = \widetilde{m}_i\, w_i + \epsilon_{i+1}\, w_{i+1}.$$

Sur la même entrée (u, v), les deux suites (v_i) et (w_i) calculées par les deux versions de l'algorithme de d'Euclide satisfont $w_i = |v_i|$, et le quotient $\widetilde{m}_i$ est la valeur absolue du quotient m_i. Le nombre d'étapes est donc le même, défini par l'indice p, pour lequel $u_p = w_p = 0$. Le pgcd est donné par le module du dernier reste non nul obtenu par l'algorithme, égal donc à u_{p-1}.

EUCLIDE-CENTRÉ-NONPLIÉ(u, v)

Entrée. Une paire d'entiers (u, v)
avec $0 < |v| \leq |u|$
Sortie. Le pgcd de u et v.

1 **tant que** $v \neq 0$
2 **faire**
3 $(u, v) \leftarrow (v, u)$
4 $m \leftarrow \left\lfloor \frac{v}{u} \right\rceil$
5 $v \leftarrow v - mu$
6 **renvoyer** $|u|$

EUCLIDE-CENTRÉ-PLIÉ(u, v)

Entrée. Une paire d'entiers (u, v)
avec $0 < v \leq u$
Sortie. Le pgcd de u et v.

1 **tant que** $v \neq 0$
2 **faire**
3 $(u, v) \leftarrow (v, u)$
4 $m \leftarrow \left\lfloor \frac{v}{u} \right\rceil$
5 $\epsilon \leftarrow \text{sign}\left(\frac{v}{u} - \left\lfloor \frac{v}{u} \right\rceil\right)$
6 $v \leftarrow \epsilon(v - mu)$
7 **renvoyer** u

FIGURE 2.1 – Algorithmes d'Euclide centrés.

2.1.4 Algorithme des fractions continues centrées.

Remarquons que chacun des deux algorithmes peut se généraliser à une paire quelconque réelle (u, v). Il y a deux cas différents selon que u et v sont $\mathbb{Q}$ linéairement indépendants ou non.

Si u et v sont $\mathbb{Q}$ linéairement dépendants, alors le $\mathbb{Z}$–module engendré par u et v est un sous-groupe discret de $\mathbb{R}$; c'est donc le réseau. $\mathcal{L}(u, v)$. Alors, le quotient v/u est un rationnel et l'un ou l'autre des deux algorithmes calcule une base du réseau $\mathcal{L}(u, v)$. Si u et v sont $\mathbb{Q}$ linéairement

indépendants, alors le $\mathbb{Z}$–module engendré par u et v est un sous-groupe dense de $\mathbb{R}$, et ce n'est donc pas un réseau. Dans ce cas, d'ailleurs, les algorithmes ne terminent pas et sont utilisés pour calculer le développement en fraction continue du réel u/v.

2.1.5 Une première analyse des algorithmes d'Euclide centrés.

A chaque étape, pour $i \geq 1$, on a $|u_{i+1}| \leq |u_i|/2$, ce qui entraîne les inégalités

$$1 \leq |u_{p-1}| \leq \frac{|u_1|}{2^{p-2}} \leq \frac{|u|}{2^{p-2}}, \qquad \text{et donc} \quad p \leq 2 + \log_2 |u|$$

Le nombre de divisions de l'algorithme est donc linéaire en la taille de l'entrée. Mais bien sûr, cette borne n'est pas fine. Nous reviendrons à l'analyse plus fine de cet algorithme tout au long de cette thèse, en liaison avec la généralisation de cet algorihme à la dimension 2, que nous abordons maintenant.

2.2 Algorithmes de réduction en dimension 2

En dimension 2, l'algorithme de réduction des réseaux est l'algorithme de Gauss, qui est la généralisation naturelle de l'algorithme d'Euclide.

L'algorithme dit de Gauss est parfois attribué à Joseph Lagrange, qui a aussi étudié les formes quadratiques, mais un peu plus tôt que Gauss [26]. Dans ses premières formulations, l'algorithme est écrit dans le vocabulaire des formes quadratiques ; il cherche à réduire des formes quadratiques, de sorte à les mettre dans une forme normale, pour pouvoir ainsi les classifier.

L'algorithme de Gauss généralise l'algorithme d'Euclide, en un double sens. D'abord, parce qu'il résout aussi le problème de la réduction, tout comme l'algorithme d'Euclide, mais lorque le réseau est donné par deux vecteurs non colinéaires de $\mathbb{R}^2$. Ensuite, parce que l'algorithme de Gauss effectue les mêmes "divisions" que celles de l'algorithme d'Euclide, mais cette fois entre deux vecteurs non colinéaires. L'algorithme de Gauss est doublement optimal. La notion de base réduite en dimension 2 est très simple (et la meilleure possible), car elle coïncide avec la notion de base minimale (base formée par une paire de vecteurs réalisant le premier et le deuxième minimum). Et le nombre d'itérations de l'algorithme est linéaire en la taille des l'entrée.

Ici, nous présentons la notion de base réduite, et, tout comme nous l'avons fait pour l'algorithme d'Euclide, deux versions de l'algorithme de Gauss, l'une qui travaille sur des bases "positives" et l'autre sur des bases "aigues". Nous donnons aussi une première majoration du nombre d'itérations.

2.2.1 Bases minimales.

En dimension 2, la notion de base réduite est particulièrement simple, puisqu'il s'agit tout simplement de bases *minimales*, formées de deux vecteurs qui réalisent les deux minima successifs. En plus, il existe une caractérisation géométrique très simple de ces bases minimales.

Définition 2.1. *Une base (u, v) d'un réseau de $\mathbb{R}^2$ est* réduite *ou* minimale *si*

$$||u|| = \lambda_1(u, v) \quad \text{et } ||v|| = \lambda_2(u, v).$$

La caractérisation suivante est fondamentale pour l'algorithme de Gauss.

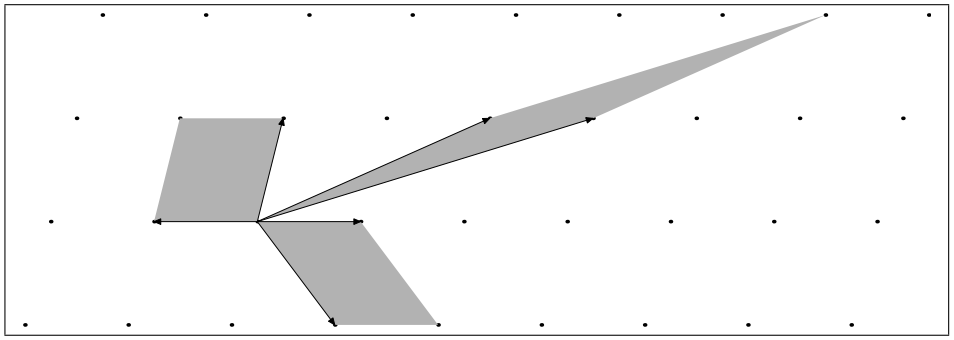

FIGURE 2.2 – Un réseau et trois de ses bases représentées par le paralléllograme qu'elles engendrent. La base de gauche est minimale (réduite), contrairement aux deux autres.

Proposition 2.2 (Caractérisation d'une base minimale)**.** *Une base $(u, v) \in \mathbb{R}^2$ est minimale si et seulement si elle vérifie les conditions suivantes :*

$$(C_1) \ : \ ||u|| \leq ||v|| \qquad\qquad (C_2) \ : \ |\tau(v,u)| \leq \frac{1}{2}$$

où $\tau(v, u)$ est le coefficient de la matrice de Gram-Schmidt défini par

$$\tau(v,u) := \frac{v \cdot u}{||u||^2}.$$

Avant de montrer cette proposition, nous allons prouver un lemme essentiel dans la suite.

Lemme 2.1. *Si une base $(u, v) \in \mathbb{R}^2$ vérifie (C_1) et (C_2) alors sa base orthogonaliée de Gram-Schmidt satisfait*

$$||v^\star|| \geq \frac{\sqrt{3}}{2}||v||$$

Démonstration. Avec la décomposition de Gram-Schmidt, le vecteur v s'écrit $v = \tau(v, u)\, u + v^\star$, et les conditions (C_1) et (C_2) entraînent

$$||v^\star||^2 = ||v||^2 - \tau(v,u)^2||u||^2 \geq \frac{3}{4}||v||^2 \qquad \text{et donc} \quad ||v^\star|| \geq \frac{\sqrt{3}}{2}||v||.$$

Ceci implique que l'angle θ entre u et v vérifie $\theta \in [\pi/3, 2\pi/3]$. □

Preuve de la proposition. Supposons d'abord que (C_1) et (C_2) sont vérifiées. Nous allons montrer que (u, v) est une base minimale. Considérons un vecteur w non nul de $\mathcal{L}(u, v)$, qui s'écrit sous la forme $w = xu + yv$ avec $(x, y) \in \mathbb{Z}^2$, avec $(x, y) \neq (0, 0)$. Considérons trois cas $y = 0, |y| = 1$ et $|y| \geq 2$. Si $y = 0$, alors $||w|| \geq ||u||$. Si $|y| \geq 2$, alors,

$$||w|| \geq |x|||v^\star|| \geq \sqrt{3}\,||v|| > ||v||.$$

Si maintenant $y = \epsilon = \pm 1$, alors

$$||w||^2 = x^2||u||^2 + 2x\epsilon(u \cdot v) \ + ||v||^2 \geq x^2||u||^2 - |x|\,||u||^2 + ||v||^2 \geq ||v||^2$$

ce qui s'achève la première partie de la preuve.

Supposons maintenant que (u, v) est minimale. La propriété (C_1) est satisfaite par définition, et il faut prouver que (C_2) est vérifiée. Si ce n'était pas le cas, on aurait $|\tau(v, u)| > 1/2$ et le vecteur $w := v - \lfloor \tau(v, u) \rceil\, u$ serait distinct de v, avec une projection orthogonale sur $\langle u \rangle$ qui satisferait

$$|\tau(w, u)u| = |\tau(v, u)u - \lfloor \tau(v, u) \rceil\, u| \leq \frac{|u|}{2} < |\tau(v, u)u|.$$

Elle serait strictement plus petite que celle de v, tandis que les projections orthogonales de w et de v sur $\langle u \rangle^{\perp}$ sont les mêmes. Il s'en suit que $||w|| < ||v||$, ce qui contredit la minimalité de (u, v). Donc, (C_2) est effectivement vérifiée, comme on voulait prouver. □

2.2.2 Bases positives et aigues.

Pour des raisons qui seront expliquées plus tard, nous désirons travailler avec deux classes de bases :

(*a*) Les bases (u, v) dites *positives*, qui vérifient la condition $\det(u, v) > 0$.

(*b*) Les bases (u, v) dites *aigües*, qui vérifient la condition $v \cdot u \geq 0$.

Ces classes sont naturelles, puisqu'il est toujours facile, à partir d'une base (u, v) quelconque, de se ramener à l'une ou l'autre de ces configurations. En effet, l'une des deux bases (u, v) ou (v, u) est positive et l'une des deux bases (u, v) ou $(u, -v)$ est aigue. La proposition 2.2 admet le corollaire suivant :

Proposition 2.3 (Caractérisations des bases minimales.)**.**
Soit (u, v) une base positive. Alors les deux conditions suivantes sont équivalentes :
- *la base (u, v) est minimale.*
- *le couple (u, v) satisfait les trois conditions suivantes :*

$$(P_1) \ : \ ||u|| \leq ||v|| \qquad (P_2) \ : \ |\tau(v, u)| \leq \frac{1}{2} \qquad (P_3) : \det(u, v) > 0.$$

Soit (u, v) une base aigüe. Alors, les deux conditions suivantes sont équivalentes :
- *la base (u, v) est minimale*
- *la paire (u, v) satisfait les deux conditions suivantes :*

$$(A_1) \ : \ ||u|| \leq ||v|| \qquad (A_2) \ : \ 0 \leq \tau(v, u) \leq \frac{1}{2}.$$

Maintenant nous présentons deux versions de l'algorithme de Gauss, travaillant sur l'un ou l'autre type de base.

2.2.3 Algorithmes de Gauss : les deux versions Gauss-positif et Gauss-aigu

Nous définissons deux versions de l'algorithme de Gauss. La version Gauss-positif travaille constamment sur des bases positives, tandis que la version Gauss-aigu travaille constamment avec des bases aigues.

Chacun des deux algorithmes cherche à satisfaire la condition $||u|| \leq ||v||$ en effectuant des échanges, et la condition sur $\tau(v, u)$ en translatant le vecteur v parallèlement à u. Ces translations diffèrent selon les versions de l'algorithme de Gauss. Pour calculer le coefficient de cette translation, l'algorithme Gauss-positif utilise la division euclidienne centrée non pliée de $(v \cdot u)$ par $||u||^2$, tandis que l'algorithme Gauss-aigu utilise la division euclidienne centrée pliée de $(v \cdot u)$ par $||u||^2$.

L'algorithme GAUSS-POSITIF**.** L'algorithme GAUSS-POSITIF reçoit en entrée une base positive et arbitraire et il travaille constamment avec une base positive ; il produit donc en sortie une base minimale positive. La condition (P_2) est satisfaite par une translation entière de la forme :

$$v := v - mu \qquad \text{avec} \quad m := \lfloor \tau(v,u) \rceil \,, \tag{2.1}$$

où $\lfloor x \rceil$ est l'entier le plus proche du réel x. Après cette translation, le nouveau coefficient $\tau(v,u)$ appartient à $]-1/2, 1/2]$. Sur l'entrée (u,v), l'algorithme GAUSS-POSITIF calcule une suite de vecteurs v_i définis par la relation

$$v_0 = u,\ v_1 = v, \qquad v_{i+1} = -v_{i-1} + m_i\, v_i \qquad \text{avec} \quad m_i := \lfloor \tau(v_{i-1}, v_i) \rceil \,. \tag{2.2}$$

Ici, chaque quotient m_i est un entier de $\mathbb{Z}$, $P(u,v) = p$ dénote le nombre d'itérations, et la paire finale (v_p, v_{p+1}) satisfait les conditions (P) de la proposition 2.3.

GAUSS-POSITIF(u,v).
Entrée. Une base positive $(u,v) \in \mathbb{R}^2$,
avec $||v|| \le ||u||, |\tau(v,u)| \le 1/2$
et $\det(u,v) > 0$.
Sortie. Une base positive minimale de $\mathcal{L}(u,v)$.

```
1  tant que ||u|| > ||v||
2      faire
3          (u, v) ← (v, −u)
4          m ← ⌊τ(v, u)⌉
5          v ← v − mu
```

GAUSS-AIGU(u,v)
Entrée. Une base aigüe $(u,v) \in \mathbb{R}^2$,
avec $||v|| \le ||u||, 0 \le |\tau(v,u) \le 1/2$.
Sortie. Une base aigüe minimale de $\mathcal{L}(u,v)$.

```
1  tant que ||u|| > ||v||
2      faire
3          (u, v) ← (v, u)
4          m ← ⌊τ(v, u)⌉
5          ε ← sign (τ(v, u) − ⌊τ(v, u)⌉)
6          v ← ε(v − mu)
```

FIGURE 2.3 – Algorithme de Gauss : Algorithmes GAUSS-POSITIF et GAUSS-AIGU

L'algorithme GAUSS-AIGU**.** L'algorithme GAUSS-AIGU reçoit en entrée une base aigüe arbitraire, travaille constamment avec des bases aigües et produit donc en sortie une base minimale aigüe. La condition (A_2) est garantie par une translation entière du type :

$$v := \epsilon(v - mu) \qquad \text{avec} \quad m := \lfloor \tau(v,u) \rceil \,, \qquad \epsilon = \operatorname{sign}\left(\tau(v,u) - m\right),$$

où $\tau(v,u)$ est défini en (2.1). Après cette transformation, le nouveau coefficient $\tau(v,u)$ satisfait $0 \le \tau(v,u)| \le (1/2)$. Sur l'entrée (u,v), l'algorithme GAUSS-AIGU calcule une séquence de vecteurs w_i définis par les relations

$$w_0 = u,\ w_1 = v, \qquad w_{i+1} = \epsilon_i(w_{i-1} - \widetilde{m}_i\, w_i)$$

$$\text{avec} \quad \widetilde{m}_i := \lfloor \tau(w_{i-1}, w_i) \rceil \,, \qquad \epsilon_i = \operatorname{sign}\left(\tau(w_{i-1}, w_i) - \lfloor \tau(w_{i-1}, w_i) \rceil\right). \tag{2.3}$$

Ici, chaque quotient $\widetilde{m}_i$ est un entier positif, $p \equiv p(u,v)$ denote le nombre d'itérations (qui est le même que pour l'algorithme GAUSS-POSITIF), et la paire finale (w_p, w_{p+1}) satisfait les conditions (A) de la proposition 2.3.

2.2.4 Comparaison entre les deux algorithmes

Ces algorithmes sont très proches, mais différents. L'algorithme Gauss-aigu peut être vu comme une version pliée de l'algorithme Gauss-positif, de la même manière que la deuxième version de l'algorithme d'Euclide est un pliage de la première. Nous reviendrons à ce point dans le chapitre 1, partie II. La proposition suivante compare ces deux algorithmes et montre qu' ils effectuent les mêmes exécutions, à un changement de signe près.

Proposition 2.4. *Considérons deux bases : une base positive (v_0, v_1), et une base aigüe (w_0, w_1) satisfaisant $w_0 = v_0$ et $w_1 = \eta_1 v_1$ avec $\eta_1 = \pm 1$. Alors, les deux suites de vecteurs (v_i) et (w_i) calculées par les deux versions de l'algorithme de Gauss, définis en (2.2) et en (2.3), satisfont $w_i = \eta_i v_i$, avec $\eta_i = \pm 1$. De plus, les deux suites de quotients (m_i) et $(\tilde{m}_i)$ sont reliées par l'égalité $\widetilde{m}_i = |m_i|$.*

Preuve. C'est une preuve par récurrence : En utilisant l'hypothèse de récurrence, on a

$$\tilde{m}_i = \lfloor \tau(w_{i-1}, w_i) \rceil = \lfloor \tau(\eta_{i-1} v_{i-1}, \eta_i v_i) \rceil = \eta_{i-1}\eta_i \lfloor \tau(v_{i-1}, v_i) \rceil = \eta_{i-1}\eta_i m_i,$$

où on a utilisé la propriété de l'entier le plus proche $\lfloor -x \rceil = -\lfloor x \rceil$. Toujours avec l'hypothèse de récurrence, on a

$$w_{i+1} = \epsilon_i(w_{i-1} - \tilde{m}_i w_i) = \epsilon_i(\eta_{i-1} v_{i-1} - (\eta_{i-1}\eta_i m_i)\eta_i v_i) = -\epsilon_i \eta_{i-1} v_{i+1},$$

ce qui achève la preuve. □

Par conséquent, lorsque l'on étudie les deux types de paramètres –paramètres d'exécution ou paramètres de sortie– les deux algorithmes sont essentiellement les mêmes, et, comme nous avons déjà dit, nous allons utiliser l'algorithme Gauss-positif pour étudier les paramètres de sortie, et l'algorithme Gauss-aigu pour les paramètres d'exécution.

2.2.5 Nombre d'itérations de l'algorithme de Gauss. Une première borne

Dans la proposition suivante, nous montrons que l'algorithme de Gauss effectue un nombre d'itérations qui est linéaire en la taille des entrées. La preuve n'est pas tout-à-fait triviale et utilise des algorithmes t-Gauss qui vont s'avérer très utiles pour l'algorithme LLL.

Nous considérons une version modifiée des algorithmes de Gauss, les algorithmes de t-Gauss, où l'on remplace la condition d'arrêt par une condition plus forte : la condition $||u|| \leq ||v||$ est remplacée par la condition $||u|| \leq t||v||$ (pour $t > 1$). Les algorithmes t–Gauss sont décrits dans la figure 2.4, et vont aussi intervenir dans la définition de l'algorithme LLL.

Proposition 2.5. *L'algorithme de Gauss effectue un nombre d'itérations linéaire en la taille des entrées. Plus précisément, sur une base d'entrée de taille M, son nombre d'itérations est au plus*

$$\log_{\sqrt{3}} M + 2$$

Démonstration. La preuve a deux étapes : d'abord, nous montrons que l'algorithme de t-Gauss termine en un nombre polynomial d'itérations dans la taille des entrées. Ensuite, nous montrons que l'algorithme de Gauss fait au plus une itération de plus que l'algorithme de t-Gauss, pour un $t < \sqrt{3}$ bien choisi.

Etudions d'abord le nombre d'itérations de t-Gauss. Supposons que l'algorithme effectue p itérations et calcule successivement la suite de vecteurs u_i, avec $u_0 = u, u_1 = v$ et, pour $i \in [\![1, p]\!]$,

$$u_{i-1} = m_i u_i + u_{i+1} \qquad \text{avec} \quad ||u_i|| < ||u_{i-1}||/t.$$

t- GAUSS-POSITIF(u, v).

Entrée. Une base positive $(u, v) \in \mathbb{R}^2$,
avec $||v|| \leq ||u||, |\tau(v, u)| \leq 1/2$
et $\det(u, v) > 0$.

Sortie. Une base positive minimale de $\mathcal{L}(u, v)$.

```
1  tant que t||u|| > |||v||
2     faire
3           (u, v) ← (v, −u)
4           m ← ⌊τ(v, u)⌉
5           v ← v − mu
```

t- GAUSS-AIGU(u, v)

Entrée. Une base aigüe $(u, v) \in \mathbb{R}^2$,
avec $||v|| \leq ||u||, 0 \leq |\tau(v, u) \leq 1/2$.

Sortie. Une base aigüe minimale de $\mathcal{L}(u, v)$.

```
1  tant que t ||u|| > ||v||
2     faire
3           (u, v) ← (v, u)
4           m ← ⌊τ(v, u)⌉
5           ϵ ← sign (τ(v, u) − ⌊τ(v, u)⌉)
6           v ← ϵ(v − mu)
```

FIGURE 2.4 – Algorithme de t-Gauss : Algorithmes t-GAUSS-POSITIF et t-GAUSS-AIGU ($t > 1$)

Nous en déduisons que

$$||u_{p-1}|| < ||u_0||/t^{p-1}\,,$$

et alors

$$p \leq 1 + \log_t \frac{||u_0||}{||u_{p-1}||} \leq 1 + \log_t M$$

car le vecteur u_{p_1} est entier et la base d'entrée est de taille M.

Maintenant, supposons que l'on applique l'algorithme de Gauss à une base de sortie (u, v) de l'algorithme de t-Gauss qui n'est pas réduite, et qui satisfait donc

$$||v|| < ||u|| \leq t||v||, \qquad |v \cdot u| \leq \frac{1}{2}||u||^2\,.$$

Alors, le premier pas de l'algorithme de Gauss échange u et v, ce qui se traduit par les relations

$$||v|| \leq t||u|| < ||v||, \qquad |u \cdot v| \leq \frac{1}{2}||v||^2\,.$$

L'étape suivante calcule le coefficient $\tau(v, u)$

$$\tau(v, u) = \frac{|v \cdot u|}{||u||^2} = \frac{|v \cdot u|}{||v||^2}\frac{||v||^2}{||u||^2} \leq \frac{t^2}{2}\,.$$

Si t vérifie $t < \sqrt{3}$, le coefficient de translation est ± 1, et le vecteur v est remplaçé par $v \pm u$, qui vérifie

$$(v \pm u) \cdot (v \pm u) = ||v||^2 \pm 2v \cdot u + ||u||^2 \geq ||v||^2 - 2|v \cdot u| + ||u||^2.$$

L'inégalité

$$||v||^2 - 2|v \cdot u| + ||u||^2 \geq ||v||^2 - ||v||^2 + ||u||^2 \geq ||u||^2,$$

entraîne que la base $(v \pm u, u)$ est réduite au bout d'une itération. Il s'en suit que l'algorithme de Gauss termine bien en un nombre d'itérations majorée par $\log_{\sqrt{3}} M + 2$, ce qu'il fallait montrer. □

2.2.6 Paramètres liés à l'exécution de l'algorithme.

La taille d'une base d'entrée $(u, v) \in \mathbb{Z}^2 \times \mathbb{Z}^2$ est définie par

$$\max\{\ell(||u||^2), \ell(||v||^2)\},$$

où $\ell(x) = 1 + \lfloor \lg x \rfloor$ est la fonction longueur binaire. Tous les calculs de l'algorithme de Gauss se font dans les matrices de Gram $G(v_i, v_{i+1})$ associée au couple (v_i, v_{i+1}). La matrice de Gram $G(u, v)$ est définie par

$$G(u, v) = \begin{pmatrix} ||u||^2 & (u \cdot v) \\ (u \cdot v) & ||v||^2 \end{pmatrix}. \tag{2.4}$$

L'initialisation de l'algorithme consiste en calculer la matrice de Gram des bases d'entrée : elle effectue le calcul de trois produits scalaires, ce qui prend un temps quadratique [1] par rapport à la taille de l'entrée (u, v). Après, tous les calculs de la *partie centrale* de l'exécution de l'algorithme sont effectués directement sur ces matrices de Gram ; plus précisément, chaque pas de l'algorithme est une division euclidienne entre les deux coefficients de la première ligne de la matrice de Gram $G(v_i, v_{i-1})$ du couple (v_i, v_{i-1}) pour obtenir le quotient m_i, suivi du calcul des nouveaux coefficients de la matrice de Gram $G(v_{i+1}, v_i)$, notamment

$$||v_{i+1}||^2 := ||v_{i-1}||^2 - 2m_i\,(v_i \cdot v_{i-1}) + m_i^2||v_i|^2, \qquad (v_{i+1} \cdot v_i) := m_i\,||v_i|^2 - (v_{i-1} \cdot v_i). \tag{2.5}$$

Le coût de la i-ème étape est donc proportionnel à $\ell(|m_i|) \cdot \ell(||v_{i-1}||^2)$, et la complexité en bits de la partie central de l'algorithme de Gauss s'exprime en fonction de

$$B(u, v) = \sum_{i=1}^{p(u,v)} \ell(|m_i|) \cdot \ell(||v_{i-1}||^2), \tag{2.6}$$

où $p(u, v)$ est le nombre d'itérations de l'algorithme de Gauss. Dans la suite, B sera appelée *complexité en bits* ou *complexité binaire.*

La complexité binaire $B(u, v)$ est l'un de nos principaux paramètres d'étude, et nous l'exprimons avec d'autres coûts plus simples. On définit trois nouveaux coûts, le coût associé aux quotients $Q(u, v)$, le coût différence $\underline{D}(u, v)$, et le coût différence approchée D :

$$Q(u, v) = \sum_{i=1}^{p(u,v)} \ell(|m_i|), \qquad \underline{D}(u, v) = \sum_{i=1}^{p(u,v)} \ell(|m_i|)\left[\ \ell(||v_{i-1}||^2) - \ell(||v_0||^2)\right], \tag{2.7}$$

$$D(u, v) := \sum_{i=1}^{p(u,v)} \ell(|m_i|) \lg \frac{||v_{i-1}||^2}{||v_0||^2}.$$

La décomposition suivante

$$B(u, v) = Q(u, v)\,\ell(||u||^2) + D(u, v) + [\underline{D}(u, v) - D(u, v)] \tag{2.8}$$

sera fondamentale dans la suite, car Q et D seront plus faciles à étudier que B, et ils seront suffisants pour notre étude, car la troisième partie de la décomposition s'avèrera d'un ordre plus petit, puisqu'elle vérifie

$$D(u, v) - \underline{D}(u, v) = \Theta(Q(u, v)).$$

Nous sommes alors conduits à étudier deux paramètres reliés au coût en bits, qui ont aussi un intérêt intrinsèque :

1. Nous considérons la multiplication naïve entre les entiers de taille M, dont la complexité en bits est $O(M^2)$.

(a) Les coûts additifs, qui fournissent une généralisation du coût Q et du coût "nombre d'itérations". Les coûts additifs sont définis comme la somme de coûts élémentaires, qui ne dépendent que des quotients m_i. Plus précisément, à partir d'un coût élémentaire positif c défini dans $\mathbb{N}$, nous considérons le coût total sur le couple (u, v) défini comme

$$C_{(c)}(u, v) = \sum_{i=1}^{p(u,v)} c(|m_i|)\,. \tag{2.9}$$

Lorsque le coût élémentaire c satisfait $c(m) = O(\log m)$, on dit que le coût additif $C_{(c)}$ est à croissance modérée.

(b) La séquence d_i (pour $i \in [1..p]$ des décroissances et la décroissance totale de la longueur $d := d_p$, définie par

$$d_i := \frac{||v_i||^2}{||v_0||^2}, \qquad d := \frac{||v_p||^2}{||v_0||^2}. \tag{2.10}$$

2.2.7 Paramètres liés à la configuration de sortie.

La base de sortie $(\hat{u}, \hat{v})$ est donc minimale. On la décrit ici via son orthogonalisée de Gram-Schmidt $(\hat{u}^\star, \hat{v}^\star)$ où $\hat{u}^\star := \hat{u}$ et $\hat{v}^\star$ est la projection orthogonale de $\hat{v}$ dans l'espace orthogonal de $< \hat{u} >$. Plus précisément, on définit les trois paramètres suivants, reliés aux minima du réseau $\mathcal{L}(u, v)$,

$$\lambda(u, v) := \lambda_1(\mathcal{L}(u, v)) = |\hat{u}|, \qquad \mu(u, v) := \frac{|\det(u, v)|}{\lambda(u, v)} = |\hat{v}^\star|, \tag{2.11}$$

$$\gamma(u, v) := \frac{\lambda^2(u, v)}{|\det(u, v)|} = \frac{\lambda(u, v)}{\mu(u, v)} = \frac{||\hat{u}||}{||\hat{v}^\star||}. \tag{2.12}$$

Le paramètre μ peut être appelé "deuxième minimum orthogonalisé", tandis que γ est exactement le défaut d'Hermite du réseau $\mathcal{L}(u, v)$. Nous expliquons à la fin du présent chapitre pourquoi ces paramètres de sortie sont importants dans l'analyse de l'algorithme LLL et les étudions en détail dans les chapitres 1 et 2.

2.3 Algorithmes de réduction en dimension n quelconque

La réduction en dimension n a d'abord été abordé comme un problème de mathématiques pures, avec un point de vue non nécessairement constructif. Les preuves du chapitre 1 sont typiques de ce point de vue : ni la preuve de Siegel, ni celle de Minkowski ne permettent de construire les objets dont on prouve l'existence. Les recherches en géométrie des nombres cherchaient à représenter un réseau par une de ses bases, celle-ci devant posséder des bonnes propriétés euclidiennes. C'est Minkowski lui-même qui a défini une des premières notions de réduction, qui porte son nom. Cette notion de réduction, qui est du point de vue mathématique la plus naturelle, présente néanmoins un inconvénient algorithmique : comme elle exige que le premier vecteur de la base soit un vecteur le plus court du réseau, elle est "difficile" à mettre en oeuvre, puisqu'on a vu dans le chapitre 1, que le calcul d'un vecteur le plus court est un problème difficile. Plus tard, on a défini d'autres notions de réduction, sans chercher le plus souvent à se poser la question d'un point de vue algorithmique : c'est le cas par exemple de la réduction de Hermite-Khorkine-Zolotareff, qui exige aussi de trouver un vecteur le plus court de réseaux définis de manière récursive, comme projetés successifs du réseau initial.

La grande percée algorithmique s'est produite en 1982, lorsque Lenstra, Lenstra et Lovász ont proposé une notion de réduction qui représentait un compromis raisonnable entre la qualité de la base réduite (moindre que celle obtenue lors des réductions précédemment décrite, mais tout de même suffisante) et le temps mis à l'obtenir, puisque l'algorithme LLL, qui construit une telle base réduite, fonctionne en temps polynomial. Ce compromis s'est avéré si fructueux qu'il a été utilisé pour résoudre des problèmes de type très divers, comme ceux que nous avons décrits dans la section 1.3 de la partie I). L'algorithme LLL s'est ainsi imposé comme un outil algorithmique majeur, en calcul formel, en théorie algorithmique des nombres, en programmation linéaire ainsi qu'en cryptanalyse. On peut sans doute dire qu'il est presque devenu une opération de base.

2.3.1 Réduction en taille : l'algorithme PROPRE.

La première idée qui vient à l'esprit pour rendre une base plus orthogonale.... est de la rapprocher de son orthogonalisée de Gram-Schmidt.

Définition 2.2 (Base propre)**.** *La base $B = (b_1, \ldots, b_p)$ est dite propre (ou réduite en taille) si les coefficients de la matrice de passage $\mathcal{P}$ de Gram-Schmidt vérifient*

$$|m_{ij}| \leq \frac{1}{2} \qquad \textit{for } 1 \leq j < i \leq p\,.$$

2.3.2 Réduction au sens de Lovász

Une base B propre est donc suffisamment proche de la base $B^\star$. Mais cela n'est pas suffisant pour assurer la bonne qualité de la base B. Il faut trouver un moyen de minorer l'angle que fait le vecteur b_i avec l'hyperplan H_{i-1} engendré par la base $B_{i-1} = (b_1, \ldots, b_{i-1})$. C'est ce qui est assuré par la condition de Lovász. L'idée de Lovász est de considérer une base B où toutes les bases dites locales sont réduites au sens de l'algorithme t-Gauss. La i-ème base locale U_i est formée par les deux vecteurs u_i, v_i, définis comme les projections de b_i, b_{i+1} sur l' orthogonal du sous-espace H_{i-1} engendré par $(b_1, b_2, \ldots, b_{i-1})$dans le sous espace H_{i+1}. Les vecteurs (u_i, v_i) sont donc définis par les relations

$$u_i = b_i^\star, \qquad v_i = b_{i+1}^\star + m_{i+1,i} b_i^\star \tag{2.13}$$

et on exige donc qu'ils satisfassent la condition de sortie de t-Gauss. Et on sait aussi la condition de sortie de l'algorithme t-Gauss induit une minoration du rapport des longueurs des vecteurs $v_i^\star = b_{i+1}^\star$ et $u_i = b_i^\star$. Tout cela conduit à la définition suivante :

Définition 2.3 (Base LLL-réduite)**.** *Une base $B = (b_1, \ldots, b_p)$ est t-réduite au sens de Lovasz si elle est propre et vérifie*

$$||b_{i+1}^\star + m_{i+1,i} b_i^\star|| \geq \frac{1}{t} ||b_i^\star||\,, \qquad \textit{pour} \quad i \in [\![1, p-1]\!] \tag{2.14}$$

ou encore, de manière équivalente

$$\ell_{i+1}^2 + + m_{i+1,i}^2 \ell_i^2 \geq \frac{1}{t^2} \ell_i^2\,, \qquad \textit{pour} \quad i \in [\![1, p-1]\!] \tag{2.15}$$

Une base $B = (b_1, \ldots, b_p)$ est réduite au sens de Siegel si elle est propre et vérifie

$$||b_{i+1}^\star|| \geq \frac{1}{s} ||b_i^\star||\,, \qquad \textit{pour} \quad i \in [\![1, p-1]\!] \tag{2.16}$$

ou encore, de manière équivalente

$$\ell_{i+1} \geq \frac{1}{s} \ell_i\,, \qquad \textit{pour} \quad i \in [\![1, p-1]\!] \tag{2.17}$$

Le lemme ci-dessous permet de montrer le lien entre les deux notions.

Lemme 2.2. *Soient s et t reliés par la relation*

$$s^2 = \frac{4t^2}{4-t^2} \qquad \left[\text{et donc} \quad s = \frac{2}{\sqrt{3}} \quad \text{pour } t = 1\right]. \tag{2.18}$$

Alors une base t-réduite au sens de Lovasz est s-réduite au sens de Siegel.

Le théorème suivant décrit la qualité d'une base réduite au sens de Lovasz. En particulier, les défauts de longueur sont bornés par une fonction exponentielle en la dimension p, et le défaut d'orthogonalité est borné par une fonction exponentielle en le carré de la dimension p,

Théorème 2.1 (LLL, 1982)**.** *Soit une base B d'un réseau $\mathcal{L}$ t-réduite au sens de Lovasz. Alors, le défaut d'Hermite $\gamma(B)$, le défaut de longueur $\theta(B)$ ou le défaut d'orthogonalité $\rho(B)$ vérifient les inégalités*

$$\gamma(B) := \frac{||b_1||^2}{(\det \mathcal{L})^{2/p}} \leq s^{p-1}, \qquad \theta(B) := \frac{||b_1||}{\lambda(\mathcal{L})} \leq s^{p-1}, \qquad \rho(B) := \frac{1}{\det \mathcal{L}} \prod_{i=1}^{p} ||b_i|| \leq s^{p(p-1)/2}, \tag{2.19}$$

qui dépendent du paramètre s, défini en fonction de t par la relation

$$s^2 = \frac{4t^2}{4-t^2} \qquad \text{et} \quad s = \frac{2}{\sqrt{3}} \quad \text{pour } t = 1.$$

Démonstration. Nous commençons par prouver l'inégalité

$$||b_i|| \leq s^{i-1}\ell_i \quad \text{pour } i \in [1,p]. \tag{2.20}$$

La relation de Siegel (2.16) prouve que

$$\ell_j \leq s^{i-j}\ell_i \text{ pour } 1 \leq j \leq i \leq d\,. \tag{2.21}$$

Le vecteur b_i satisfait alors

$$||b_i||^2 = \left|\sum_{j=1}^{i} m_{i,j} b_j^\star\right|^2 \leq \ell_i^2 \left[\frac{1}{4}\sum_{j=1}^{i-1} s^{2(i-j)} + 1\right] = \ell_i^2 \left[1 + \frac{1}{4}s^2 \frac{s^{2(i-1)}-1}{s^2-1}\right] \leq s^{2(i-1)}\ell_i^2,$$

la dernière inégalité étant vraie car $s^2 > 4/3$. Maintenant, les relations (2.20) et (2.16) prouvent les inégalités

$$||b_i|| \leq s^{i-1}\ell_i, \qquad \text{pour } 1 \leq j \leq i \leq p, \tag{2.22}$$

qui permettent de majorer le défaut d'orthogonalité

$$\rho(B) = \prod_{i=1}^{p} ||b_i|| \leq \prod_{i=1}^{p} s^{i-1}\ell_i = s^{p(p-1)/2} \prod_{i=1}^{p} \ell_i = s^{p(p-1)/2} \det(\mathcal{L}).$$

Majorons le défaut d'Hermite. En nous servant de (2.21), nous avons

$$\gamma(B) = \frac{||b_1||^2}{(\det \mathcal{L})^{2/p}} = ||b_1||^2 \cdot \left(\prod_{i=1}^{p} \frac{1}{\ell_i}\right)^{2/p} \leq ||b_1||^2 \cdot \left(\prod_{i=1}^{p} \frac{s^{i-1}}{\ell_1}\right)^{2/p} = ||b_1||^2 \cdot \left(\frac{s^{p(p-1)/2}}{\ell_1^p}\right)^{2/p} = s^{p-1},$$

ce qui établit l'inégalité.

Considérons finalement le défaut de longueur. Soit $i = k$ l'indice d'un plus petit ℓ_i. En nous servant de la proposition 1.2, puis de (2.21), nous concluons que

$$\theta(B) = \frac{\ell_1}{\lambda(\mathcal{L})} \leq \frac{\ell_1}{\ell_k} \leq s^{k-1} \leq s^{p-1},$$

ce qui majore le défaut de longueur et conclut la preuve de la proposition. □

2.3.3 Description de l'algorithme LLL(t)

L'algorithme LLL considère un réseau euclidien donné par un système B formé par p vecteurs indépendants dans l'espace ambiant $\mathbb{R}^n$. La base de sortie, notée $\hat{B}$, est t-réduite au sens de Lovasz. L'algorithme travaille sur la matrice $\mathcal{P}$ qui exprime le système B en fonction de son orthogonalisé de Gram-Schmidt $B^\star$. Rappelons que son coefficient générique $m_{i,j}$ vaut

$$m_{i,j} = \frac{b_i \cdot b_j^\star}{||b_j^\star||^2}, \qquad 1 \leq i, j \leq p.$$

L'algorithme effectue essentiellement deux types d'opérations : *réduction en taille* et *réduction de Gauss des bases locales*. À la fin de l'exécution de l'algorithme, toutes les bases locales sont réduites dans le sens de t-Gauss, et la base est propre. La base est donc t- réduite au sens de Lovasz, et donc aussi s-réduite au sens de Siegel.

Réduction en taille. On rappelle que le vecteur b_i est réduit en taille si tous les coefficients $m_{i,j}$ de la i-ème ligne de la matrice $\mathcal{P}$ satisfont $|m_{i,j}| \leq 1/2$ pour tout $j \in [\![1, i-1]\!]$. La réduction de taille du vecteur b_i s'effectue par l'algorithme `Réduction-de-taille` qui translate le vecteur b_i par rapport aux vecteurs b_j pour tous les $j \in [\![1, i-1]\!]$. Comme les coefficients sous-diagonaux jouent un rôle particulier, puisqu'ils interviennent dans les étapes de réduction de Gauss, l'opération `Réduction-de-taille`(b_i) est divisée en deux opérations :

`Réduction-de-taille-principale`(b_i) : $b_i := b_i - \lfloor m_{i,i-1} \rceil\, b_{i-1}$;

suivie de

`Réduction-de-taille-secondaire`(b_i) :
`Pour` $j := i-2$ `endescendant` 1 `faire` $b_i := b_i - \lfloor m_{i,j} \rceil\, b_j$;

Réduction de Gauss des bases locales. La i-ème base locale U_i est formée par les deux vecteurs u_i, v_i, définis en (2.13) comme les projections de b_i, b_{i+1} sur l' orthogonal du sous-espace H_{i-1} engendré par $(b_1, b_2, \ldots, b_{i-1})$ dans H_{i+1}. L'algorithme LLL(t) exécute l'algorithme t-Gauss-positif sur les bases locales U_i, mais avec les différences suivantes :

(a) Les opérations qui sont effectuées pendant le déroulement de l'algorithme Gauss-positif sur la base locale U_i sont *repercutées* sur le système (b_i, b_{i+1})

(b) L'algorithme Gauss-positif est exécuté sur la base locale U_i *mais une étape à la fois.* L'indice i des bases locales varie dans l'intervalle $[\![1, p-1]\!]$. Il commence à $i = 1$, et il est incrémenté ou décrémenté à chaque étape selon que le résultat du test $t|v_i| > |u_i|$ est positif ou négatif. Ceci définit une marche aléatoire. La longueur K de cette marche aléatoire est le nombre d'itérations,

$$\mathcal{P} := \begin{array}{c} \\ b_1 \\ b_2 \\ \vdots \\ b_i \\ b_{i+1} \\ \vdots \\ b_p \end{array} \begin{array}{c} \begin{array}{ccccccc} b_1^\star & b_2^\star & \dots & b_i^\star & b_{i+1}^\star & \dots & b_p^\star \end{array} \\ \begin{pmatrix} 1 & 0 & \dots & 0 & 0 & 0 & 0 \\ m_{2,1} & 1 & \dots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ m_{i,1} & m_{i,2} & \dots & 1 & 0 & 0 & 0 \\ m_{i+1,1} & m_{i+1,2} & \dots & m_{i+1,i} & 1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ m_{p,1} & m_{p,2} & \dots & m_{p,i} & m_{p,i+1} & \dots & 1 \end{pmatrix} \end{array} \qquad U_k := \begin{array}{c} \\ u_k \\ v_k \end{array} \begin{array}{c} \begin{array}{cc} b_k^\star & b_{k+1}^\star \end{array} \\ \begin{pmatrix} 1 & 0 \\ m_{k+1,k} & 1 \end{pmatrix} \end{array}$$

LLL (t) $\quad [t > 1]$

Input. Une base B d'un réseau $\mathcal{L}$ de dimension p.
Output. Une base réduite $\hat{B}$ de $\mathcal{L}$.

`Gram` : calculer la base orthogonale $B^\star$ et la matrice $\mathcal{P}$.
$i := 1$;
`Tant que` $i < p$ `faire`
  1– `Réduction-de-taille-principale` (b_{i+1})
  2– **Test** La base locale U_i est `réduite` ? $(|v_i| > (1/t)|u_i|$? $)$
    si oui, alors :
      `Réduction-de-taille-secondaire` (b_{i+1})
      $i := i + 1$;
    sinon :
      `Échanger` b_i et b_{i+1}
      `Recalculer` $(B^\star, \mathcal{P})$;
      `Si` $i \neq 1$ `alors` $i := i - 1$;

FIGURE 2.5 – L'algorithme LLL, la matrice $\mathcal{P}$, et les bases locales U_k.

L'algorithme LLL travaille sur la suite des longueurs ℓ_i des vecteurs du système orthogonal de Gram-Schmidt $B^\star$ et il considère les rapports r_i (appelés rapports de Siegel) des normes de deux vecteurs orthogonalisés successifs $b_i^\star$ et $b_{i+1}^\star$,

$$r_i := \frac{\ell_{i+1}}{\ell_i}, \qquad \text{avec } \ell_i := ||b_i^\star||. \tag{2.23}$$

Les étapes de réduction en taille ne modifient pas ces rapports, tandis que les échanges effectués lors des étapes de réduction de Gauss visent à minorer ces rapports de Siegel, comme on va le voir maintenant.

2.3.4 Effet des échanges de l'algorithme.

Un échange entre b_i et b_{i+1}, transforme la paire (b_i, b_{i+1}) en la paire $(\check{b}_i, \check{b}_{i+1})$. Il ne modifie pas les orthogonalisés $b_j^\star$ avec $j \notin \{i, i+1\}$; le plan orthogonal de H_{i-1} dans H_{i+1} reste inchangé, et on peut y décrire les transformations subies par la base locale (u_i, v_i), définie par les relations (2.13) (voir figure 2.6). Nous décrivons maintenant l'effet d'un tel échange sur la suite (ℓ_i).

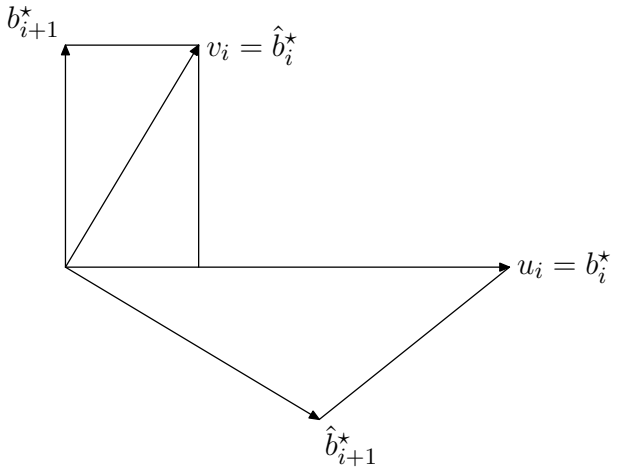

FIGURE 2.6 – L'échange des vecteurs b_i et b_{i+1}, qui deviennent $\hat{b}_i$ et $\hat{b}_{i+1}$, vu sur le sous-espace orthogonal à $(b_1, \ldots, b_{i-1})$.

Lemme 2.3. *Associons à une base B les deux quantités*

$$a := \min\{\ell_i \mid 1 \leq i \leq p\}, \qquad A := \max\{\ell_i \mid 1 \leq i \leq p\}, \tag{2.24}$$

définies comme étant le minimum et le maximum des longueurs ℓ_i des vecteurs de la base $B^\star$. Tout au long de l'algorithme LLL(t), le minimum a croît et le maximum A décroît.

Démonstration. Les longueurs ℓ_i étant invariantes lors des translations, il suffit de montrer que lors d'un échange effectué par l'algorithme, a croît et A décroît. Plus précisément, lors d'un échange entre b_i et b_{i+1}, qui transforme la paire (b_i, b_{i+1}) en la paire $(\check{b}_i, \check{b}_{i+1})$, il suffit de montrer les deux inégalités

$$\min(\check{\ell}_i, \check{\ell}_{i+1}) \geq \min(\ell_i, \ell_{i+1}) \quad \text{et} \quad \max(\check{\ell}_i, \check{\ell}_{i+1}) \leq \max(\ell_i, \ell_{i+1}). \tag{2.25}$$

Puisqu'il s'agit d'un échange, les conditions équivalentes

$$||u_i|| \geq t||v_i||, \qquad \frac{1}{t^2}\ell_i^2 \geq \ell_{i+1}^2 + m_{i+1,i}^2 \ell_i^2 \tag{2.26}$$

sont vérifiées et montrent d'abord que $\ell_i \geq \ell_{i+1}$. La relation $\check{b}^\star_i = v_i$, entraîne alors, en utilisant (2.26),

$$\ell_{i+1}^2 \leq \check{\ell}_i^2 = \ell_{i+1}^2 + m_{i+1,i}^2 \ell_i^2 \leq \frac{1}{t^2}\ell_i^2 \leq \ell_i^2. \tag{2.27}$$

Enfin, puisque l'échange ne modifie pas le réseau engendré, le déterminant est conservé, ce qui se traduit par l'égalité $\check{\ell}_i \cdot \check{\ell}_{i+1} = \ell_i \cdot \ell_{i+1}$. Ainsi, en utilisant le fait que $\check{\ell}_i \leq \ell_i$, nous concluons $\check{\ell}_{i+1} \geq \ell_{i+1}$ et la première partie de la condition (2.25) s'en suit. La conservation du produit entraîne alors l'autre partie de la condition (2.25). □

Cet intervalle $[a, A]$ joue un rôle important car il fournit une approximation pour le premier minimum $\lambda(\mathcal{L})$ du réseau (c'est-à-dire la longueur d'un plus court vecteur non nul du réseau). Plus précisément,

$$\lambda(\mathcal{L}) \leq A\sqrt{p}, \qquad \lambda(\mathcal{L}) \geq a. \tag{2.28}$$

2.3.5 Paramètres d'exécution

Ces paramètres décrivent l'exécution de l'algorithme lui-même : la longueur de la marche aléatoire (égale au nombre d'itérations K), la taille des translations entières, la taille des rationnels $m_{i,j}$ tout au long de l'exécution.

Théorème 2.2 (LLL, 1982, Daudé, Vallée 1994 [19]). *Sur une base d'entrée $B = (b_1, \ldots, b_p)$ d'un réseau $\mathcal{L}$ de $\mathbb{R}^n$, l'algorithme LLL(t) renvoie en sortie une base t-réduite au sens de Lovász de $\mathcal{L}$, après avoir effectué un nombre d'itérations K qui satisfait les diverses inégalités suivantes :*

(*i*) *Si $a := \min\{\ell_i\}, A := \max\{\ell_i\}$ désignent les longueurs du plus petit et du plus grand vecteur de l'orthogonalisée $B^\star$, alors*

$$K \leq (p-1) + p(p-1)\log_t \frac{A}{a}. \tag{2.29}$$

(*ii*) *Si les vecteurs de B sont de longueur au plus N et si le réseau $\mathcal{L}$ engendré par B a un premier minimum $\lambda(\mathcal{L})$ alors*

$$K \leq \frac{p^2}{2}\log_t \frac{N\sqrt{p}}{\lambda(\mathcal{L})}.$$

(*iii*) *Si le réseau est entier, et si les vecteurs de B sont de longueur au plus N, alors*

$$K \leq (p-1) + p(p-1)\log_t N$$

Démonstration. Considérons le sous-réseau $\mathcal{L}_j$ de $\mathcal{L}$ engendré par la base $(b_1, \ldots, b_j)$, et la quantité

$$D = \prod_{j=1}^{n-1} \det(\mathcal{L}_j) = \prod_{j=1}^{n-1}\prod_{k=1}^{j} \ell_k = \prod_{j=1}^{n-1} \ell_i^{n-i}.$$

Le nombre d'itérations de l'algorithme LLL est égal au nombre de tests de Lovász. Soit K^+ le nombre de tests dont le résultat est positif et K^- le nombre de tests dont le résultat est négatif. Nous avons $K^+ - K^- \leq p-1$, où l'inégalité provient du fait que les tests négatifs dans la première base locale ne sont pas compensés par des tests positifs. Ainsi, nous avons

$$K \leq (p-1) + 2K^-. \tag{2.30}$$

La quantité D décroît tout au long de l'algorithme. Les translations laissent D invariant. Lors d'un échange, lorsqu'un test de Lovász se revèle négatif, la valeur de D est modifiée et on désigne par $\check{D}$ la valeur de D juste après un échange. Si cet échange se produit dans la i-ème base locale, les longueurs ℓ_i et ℓ_{i+1} seront transformées respectivement en $\check{\ell}_i$ et $\check{\ell}_{i+1}$, de sorte que, comme on a vu dans (2.27), $\check{\ell}_i \leq (1/t)\ell_i$. Ainsi, parmi les déterminants des sous-réseaux $\det(\mathcal{L}_j)$, le seul qui change est $\det(\mathcal{L}_i)$, et il diminue donc d'un facteur $1/t$. Il en est donc de même pour D. Ainsi, D décroît d'un facteur t et on a $\check{D} \leq D/t$. Donc, si $\widehat{D}$ désigne la valeur finale de D, et D sa valeur initiale, on a

$$K^- \leq \log_t \frac{D}{\widehat{D}}, \qquad K \leq p - 1 + 2\log_t \frac{D}{\widehat{D}}.$$

Les majorations possibles pour K dépendent alors des majorations possibles pour le rapport $D/\widehat{D}$. Le lemme 2.3, entraîne les inégalités

$$\widehat{D} \geq a^{p(p-1)}, \qquad D \leq A^{p(p-1)}, \qquad \text{et donc} \quad \frac{D}{\widehat{D}} \leq \left(\frac{A}{a}\right)^{p(p-1)}, \tag{2.31}$$

qui impliquent la borne cherchée dans le cas (i), et les majorations

$$D \leq N^{p(p-1)/2}, \qquad D \leq 2^{Mp(p-1)/2} \tag{2.32}$$

dans le cas de (ii) et (iii). De plus, lorsque le réseau $\mathcal{L}$ est entier, on a $\widehat{D} \geq 1$, ce qui permet de conclure dans le cas (iii).

On peut aussi faire intervenir le premier minimum du réseau. La définition de la constante d'Hermite entraîne l'inégalité

$$\det(\mathcal{L}_j) \geq \frac{\lambda(\mathcal{L}_j)^j}{\gamma_j^{j/2}}.$$

Puisque $\mathcal{L}_j$ est un sous-réseau de $\mathcal{L}$, les inégalités $\lambda(\mathcal{L}_j) \geq \lambda(\mathcal{L})$ entraînent l'inégalité

$$\widehat{D} \geq \prod_{j=1}^{p-1} \frac{\lambda(\mathcal{L})^j}{\gamma_j^{j/2}} = \lambda(\mathcal{L})^{p(p-1)/2} \prod_{j=1}^{p-1} \gamma_j^{-j/2}. \tag{2.33}$$

En utilisant alors (2.32), on obtient

$$2K^- \leq \frac{p(p-1)}{2} \log_t \frac{N}{\lambda(\mathcal{L})} + \sum_{j=1}^{p-1} \log_t(\gamma_j^j),$$

La majoration $\gamma_j \leq j$ de la constante d'Hermite entraîne alors l'inégalité

$$(p-1) + \sum_{j=1}^{p-1} \log_t(\gamma_j^j) \leq \frac{p^2}{2} \log_t p,$$

ce qui, avec (2.30) conclut dans le cas (ii). □

Ainsi, l'algorithme LLL effectue un nombre d'itérations qui est polynomial en la dimension p. La notion d'entrée que nous avons utilisée jusque là est néanmoins assez large : il s'agit de bases de vecteurs de $\mathbb{R}^n$ de cardinal p. L'algorithme LLL termine bien lorsqu'il reçoit en entrée l'une de ces bases. Mais, dans la pratique algorithmique, on utilise des nombres entiers, ou des rationnels, Les opérations arithmétiques effectuées au cours de l'algorithme ont un coût, qu'il faut comparer avec la taille de l'entrée

$$\tau(B) = \Theta(pn) \cdot \log M \qquad \text{où} \quad M = \max\{b_{i,j}; \quad i \in [\![1,p]\!],\ j \in [\![1,n]\!]\}.$$

Il faut donc aussi borner aussi le nombre d'opérations arithmétiques de base lors de l'exécution de l'algorithme LLL, ce qui exige aussi de vérifier que la croissance des entiers utilisés lors de l'exécutionreste polynomiale en la taille de l'entrée. On obtient le résultat suivant.

Théorème 2.3 (LLL, 1982)**.** *Soit $\mathcal{L} \subset \mathbb{Z}^n$ un réseau donné par une base $b_1, b_2, \cdots, b_p$, et soit $B := \max\{||b_i||^2, \quad i \in [\![1,p]\!]\}$. Alors, le nombre d'opérations arithmétiques effectuées par l'algorithme LLL(t) est en $O(p^3 n \log_t B)$, et les opérandes sont des entiers dont la longueur binaire est en $O(p \log B)$.*

L'algorithme LLL fournit naturellement un algorithme d'approximation pour SVP, avec un facteur d'approximation exponentiel dans la dimension.

Théorème 2.4 (Algorithme d'approximation pour SVP)**.** *Dans un réseau de dimension $p \geq 2$, l'algorithme LLL(t) est un algorithme polynomial d'approximation pour le calcul du vecteur le plus court, avec un facteur d'approximation en s^{p-1}, où s et t sont reliés par la relation (2.18).*

2.3.6 Une variante de l'algorithme LLL : l'algorithme PAIR-IMPAIR

L'algorithme LLL original exécute des étapes de l'algorithme de Gauss sur les bases locales, mais il n'exécute pas la suite de toutes les étapes de l'algorithme sur une base locale, avant d'en traiter une autre. En effet, lorsqu'on fait un échange dans l'algorithme LLL, l'indice i est décrémenté (sauf si $i = 1$, où l'on retrouve effectivement l'algorithme de Gauss, jusqu'à ce que l'on passe en $i = 2$). L'algorithme LLL original réduit toujours la boîte non réduite de plus petit indice : il adopte la statégie de l'indice minimal. Mais, on peut imaginer beaucoup d'autres stratégies, et réduire les bases locales dans n'importe quel ordre, car l'estimation du nombre d'itérations est indépendante de l'ordre dans lequel les bases sont réduites. La fonction potentiel D reste constante sous l'effet des translations et décroit lors de chaque échange, indépendamment de l'indice où cet échange est effectué.

La variante PAIR-IMPAIR, initialement proposée par G. Villard, alterne entre deux phases. Dans la première, appelée phase IMPAIRE, l'algorithme GAUSS-AIGU est exécuté sur toutes les bases locales U_i d'indice i impair. Dans la seconde, la phase PAIRE, l'algorithme GAUSS-AIGU est exécuté sur toutes les bases locales U_i d'indice i pair. Remarquons que, lors de chaque phase, les bases locales peuvent être réduites en parallèle (puisquelles sont sans intersection). C'était d'ailleurs la motivation initiale de Villard. L'algorithme LLL PAIR-IMPAIR effectue en alternance l'une et l'autre phase jusqu'à ce que la base complète soit réduite.

LLL Pair–Impair (t) $\quad [t > 1]$

Input. Une base B d'un réseau $\mathcal{L}$ de dimension p.
Output. Une base réduite $\hat{B}$ de $\mathcal{L}$.

`Gram` : calculer la base orthogonale $B^\star$ et la matrice $\mathcal{P}$.
`Tant que` B n'est pas réduite `faire`
  `Phase Impaire` (B) :
    `Pour` $i = 1$ à $\lfloor n/2 \rfloor$ `faire`
      `Réduction-de-taille-principale` (b_{2i}) ;
      $\mathcal{M}_i :=$ t–GAUSS-AIGU (U_{2i-1}) ;
      $(b_{2i-1}, b_{2i}) := (b_{2i-1}, b_{2i})^t \mathcal{M}_i$;
    `Pour` $i = 1$ à n `faire Réduction-de-taille-secondaire` (b_i) ;
    `Recalculer` $B^\star, \mathcal{P}$;
  `Phase Paire` (B) :
    `Pour` $i = 1$ à $\lfloor (n-1)/2 \rfloor$ `faire`
      `Réduction-de-taille-principale` (b_{2i+1}) ;
      $\mathcal{M}_i :=$ t–GAUSS-AIGU (U_{2i}) ;
      $(b_{2i}, b_{2i+1}) := (b_{2i}, b_{2i+1})^t \mathcal{M}_i$;
    `Pour` $i = 1$ à n `faire Réduction-de-taille-secondaire` (b_i) ;
    `Recalculer` $B^\star, \mathcal{P}$;

FIGURE 2.7 – La variante PAIR-IMPAIR de l'algorithme LLL.

Dans une phase IMPAIRE par exemple, deux bases successives correspondent à des indices k et $k+2$, et sont de la forme $U_k := (u_k, v_k)$ et $U_{k+2} := (u_{k+2}, v_{k+2})$. Alors, la phase IMPAIRE réduit complètement ces deux bases locales, au sens de t-GAUSS-AIGU, et elle calcule deux bases locales réduites, notées $(\hat{u}_k, \hat{v}_k)$ et $(\hat{u}_{k+2}, \hat{v}_{k+2})$, qui satisfont en particulier

$$||\widehat{v}_k^\star|| = \mu(u_k, v_k), \qquad ||\widehat{u}_{k+2}|| = \lambda(u_{k+2}, v_{k+2}),$$

où les paramètres λ, μ sont définis dans (2.11). Alors, au début de la phase suivante PAIRE, on considère la base locale U_{k+1} d'indice $k+1$, alors formée (à une similitude près) des deux vecteurs

$$u_{k+1} = \widehat{v}_k^\star, \qquad v_{k+1} = \nu\, \widehat{v}_k^\star + \widehat{u}_{k+2},$$

Après la réduction en taille, le réel ν appartient à l'intervalle $[-1/2, +1/2]$. et le rapport de Siegel initial r_{k+1} de la base locale dans la phase PAIRE peut s'exprimer avec les longueurs de la sortie de la base IMPAIRE, comme

$$r_{k+1} = \frac{\lambda(u_{k+2}, v_{k+2})}{\mu(u_k, v_k)}.$$

Cela explique l'importance du rôle des paramètres λ, μ qui décrivent la sortie de l'algorithme de Gauss. Ils seront étudiés en détail dans cette thèse, tout au long de la partie III, et nous allons utiliser les résultats de cette étude dans le chapitre 1 de la partie IV pour proposer un début d'analyse de l'algorithme LLL-IMPAIR-PAIR.

Nous avons décrit les algorithmes de réduction. Maintenant, nous décrivons dans le chapitre suivant le paysage actuel de leurs analyses.

Chapitre 3

Premiers résultats sur le comportement probabiliste de l'algorithme LLL

Sommaire

L'étude du comportement probabiliste d'un algorithme comporte deux phases principales : on choisit d'abord un modèle probabiliste des entrées (le plus réaliste possible, compte-tenu des applications de l'algorithme). Les paramètres liés à l'algorithme, que ce soit des paramètres d'exécution (nombre d'itérations, complexité binaire) ou de sortie (géométrie de la sortie, qualité

de la base) deviennent alors des variables aléatoires. La deuxième phase, d'analyse proprement dite, consiste à étudier ces variables aléatoires : moyenne, variance, distribution...

Ce chapitre cherche à donner une vue d'ensemble du paysage. Il présente d'abord la problématique générale de l'analyse probabiliste d'un algorithme, et décrit assez informellement les principaux résultats déjà obtenus, concernant les analyses des algorithmes d'Euclide et de Gauss. Ensuite, il introduit divers modèles aléatoires pour les bases d'entrée (section 3.2), présente les résultats théoriques obtenus dans le modèle dit de la boule aléatoire (section 3.3). Il conclut en décrivant les expérimentations menées et les conjectures posées (section 3.4).

3.1 Analyse probabiliste d'un algorithme. L'exemple des algorithmes de réduction.

L'analyse des algorithmes est l'étude mathématique du comportement des algorithmes, tant dans le *pire des cas*, comme dans le *cas moyen* et le *meilleur des cas*. Nous nous intéressons ici aux analyses dans le cas moyen, et plus généralement aux *analyses probabilistes*, que l'on explique dans la section suivante. C'est Knuth qui a fondé le domaine, avec ses livres *The Art of Computer Programming* [35, 36, 37], parus par première fois dans les années soixante, dont l'influence dans le monde informatique est incontestable. Et les méthodes plus modernes sont décrites dans les les ouvrages de Flajolet et Sedgewick [23, 64].

3.1.1 Analyse probabiliste d'un algorithme.

Il s'agit d'élucider le comportement "générique" d'un algorithme, par opposition à son étude dans le pire des cas, qui donne des informations sur ses comportements extrêmes. Il y a deux étapes dans une telle analyse, une étape de modélisation et une étape d'analyse proprement dite. Dans la première étape, de modélisation, on cherche à caractériser l'ensemble des entrées de l'algorithme, ainsi que les paramètres que l'on veut étudier, qui peuvent décrire le comportement de l'algorithme pendant son exécution, ou à sa sortie ; pour l'exécution, les paramètres intéressants sont le nombre d'itérations, la mémoire utilisée, la complexité en bits, ... ; en ce qui concerne la sortie, ce sont des paramètres qui décrivent le résultat de l'algorithme.

D'un point de vue plus formel, on définit un espace Ω regroupant l'ensemble des entrées valides de l'algorithme, sur lequel on définit une notion de taille, et Ω_N est l'ensemble des entrées valides de taille N. Le plus souvent, c'est un ensemble fini, et on le munit d'une probabilité. Le choix de la probabilité $\mathbb{P}_N$ sur Ω_N résulte souvent d'un compromis entre la simplicité d'utilisation et le réalisme, car on souhaite aussi qu'elle modélise la distribution des données qu'on rencontre dans les applications. Sur cet espace probabilisé $(\Omega_N, \mathbb{P}_N)$, les paramètres que l'on veut étudier deviennent des variables aléatoires, et on va les étudier avec ce point de vue probabiliste : c'est l'*analyse probabiliste de l' algorithme*. Les résultats les plus simples sont obtenus lors de l'*analyse en moyenne*, où on se limite à déterminer la valeur moyenne (ou l'espérance) de ces paramètres. Si on réussit à mener cette analyse à bien, on peut alors continuer vers l'analyse en distribution, c'est à dire la détermination de la distribution de ces paramètres.

A la fois pour des considérations de simplicité et de visibilité, on ne cherche pas des résultats exacts, mais on est intéressé par le comportement probabiliste asymptotique de l'algorithme, quand la taille N des données devient grande (i.e., tend vers l'infini). On cherchera alors des équivalents asymptotiques de la moyenne et de la variance, ou bien des distributions limite.

Il faut remarquer aussi que les études dans un modèle discret sont souvent beaucoup plus délicates à mener que les études dans un modèle continu. Or, souvent, il existe une extension

naturelle de l'algorithme à des données continues, et, au moins, dans un sens informel, le modèle discret $(\Omega_N, \mathbb{P}_N)$ se "rapproche" d'un modèle continu $(\overline{\Omega}, \overline{\mathbb{P}})$ quand la taille N des données tend vers l'infini. On a parfois intérêt donc à analyser l'algorithme dans ce modèle continu, car on peut y disposer de tous les outils d'analyse, puis à opérer un retour du continu vers le discret. Ce retour peut d'ailleurs être délicat, car l'ensemble Ω peut être de mesure nulle dans $\overline{\Omega}$.

Nous décrivons maintenant l'exemple des algorithmes d'Euclide, de Gauss et celui de l'algorithme LLL.

3.1.2 L'exemple de l'algorithme d'Euclide.

L'algorithme d'Euclide (ici, celui d'Euclide centré) travaille avec des paires d'entiers (u, v), et il est naturel de choisir comme taille de la paire (u, v) le maximum de $|u|$ et de $|v|$. L'ensemble des entrées de taille N est ainsi

$$\Omega_N = \{(u, v) \in \mathbb{N}^2 \setminus (0, 0) \ : \ \max(|u|, |v|) \leq N\}.$$

Sur cet ensemble, on considère le plus souvent la probabilité uniforme (mais pas toujours). Les principaux paramètres qu'on veut étudier sont les suivants : liés à l'exécution, ce sont le nombre d'itérations, la complexité binaire, la taille des quotients m_i. L'algorithme d'Euclide calcule le pgcd, mais aussi le développement en fraction continue du rationnel u/v. Il est aussi important d'analyser la taille de la sortie de l'algorithme : taille du pgcd, ou place–mémoire occupée par le développement en fraction continue de u/v.

On a déjà expliqué dans le chapitre 2 comment l'algorithme d'Euclide se prolonge naturellement en un algorithme sur $\tilde{\mathcal{I}} = [0, 1/2]$, l'algorithme des fractions continues. Cet algorithme a été analysé de manière intensive. Et on se pose sur cet algorithme des questions de nature un peu différente. Gauss par exemple a conjecturé le premier l'existence d'une densité limite, qui décrit la distribution du réel x quand le nombre d'itérations tend vers l'infini. Cette densité, appelée densité de Gauss, a un analogue pour l'algorithme centré,

$$\psi(x) = \frac{1}{\log \phi} \left[\frac{1}{\phi + x} + \frac{1}{\phi^2 - x} \right],$$

qui va jouer un rôle important dans cette thèse, comme nous le verrons.

Il est alors tentant de chercher à utiliser ces résultats sur l'extension continue de l'algorithme pour analyser l'algorithme d'Euclide. Mais ce transfert du continu au discret est ici délicat, car les algorithmes ne se comparent pas si aisément, puisque l'algorithme des fractions continues ne termine jamais, sauf sur les rationnels, où il coincide avec l'algorithme d'Euclide.Toute une chaîne d'outils doit être utilisée : systèmes dynamiques, séries génératrices, Formule de Perron ou théorèmes taubériens... Un exemple de la différence entre les deux algorithmes est explicité par la différence entre densités limite. La densité limite qui s'installe au milieu de l'exécution de l'algorithme d'Euclide n'est pas la densité de Gauss : c'est une densité qui est reliée à celle de Gauss, mais en est distincte.

Dans le cas de l'algorithme de Gauss, le passage du continu au discret est de nature différente, comme nous allons le voir ensuite.

Les débuts de l'analyse de l'algorithme d'Euclide dans le pire des cas remontent au dix-huitième siècle, quand de Lagny [41] observe que les plus petits dénominateurs qui réalisent une certaine hauteur de fraction continue sont toujours des nombres de Fibonacci successifs. Mais les véritables analyses de l'algorithme, au sens moderne du terme, ne sont apparues qu'autour de la moitié du dix-neuvième siècle. Plusieurs analyses sont alors publiées ; on attribue un rôle

important à celle de Lamé [43] qui étudie l'algorithme d'Euclide dans sa version classique (avec division par défaut). C'est Athanase Dupré, en [22], qui a analysé le premier le pire des cas de l'algorithme d'Euclide dans sa version centrée, en suivant les idées de Binet et Lamé. De notre point de vue, Binet a eu un rôle important, car il avait déjà borné le nombre d'itérations de l'algorithme centré en 1841, avant l'analyse de Lamé. En 1844, juste après l'article de Lamé, Binet a publié [10] la solution explicite de la récurrence de Fibonacci qui, comme c'était plus ou moins évident d'après l'article de Lamé, constituait le pire des cas de l'algorithme d'Euclide. L'article de Binet a notamment inspiré Dupré pour procéder de même avec la récurrence liée à l'algorithme centré [22].

Les analyses plus fines de l'algorithme d'Euclide –analyse en moyenne et en distribution– sont bien plus récentes. L'analyse en moyenne n'a débuté que dans les années 70, avec les travaux d'Heilbronn [31] et de Dixon [21], qui ont utilisé des méthodes assez spécifiques aux algorithmes d'Euclide classiques (division par défaut ou centrée). A partir de 1995, Vallée a établi un cadre général pour analyser (presque) tous les algorithmes d'Euclide connus à ce jour, y compris vis-à-vis de leur complexité binaire [6]. L'analyse en distribution est encore plus récente ; elle débute en 1994 avec Hensley, qui montre que le nombre d'itérations de l'algorithme d'Euclide suit une distribution asymptotiquement gaussienne. Cette analyse a depuis été généralisée dans des cadres divers par Baladi et Vallée [9], puis dans le cadre de la complexité binaire par Lhote et Vallée [48], qui montrent que la complexité binaire admet aussi une loi limite gaussienne.

3.1.3 L'exemple de l'algorithme de Gauss.

Pour l'algorithme de Gauss, l'ensemble des données de taille N est un sous-ensemble de

$$\Omega_N = \{(u, v) \in \mathbb{Z}^2 \times \mathbb{Z}^2 \setminus (0, 0) \ : \ \max(||u||, ||v||) \leq N,$$

qui diffère légèrement selon la variante qu'on veut étudier. Mais, ici, il est clair qu'on peut choisir des distributions d'entrées très différentes, car le comportement de l'algorithme apparaît très sensible à la configuration géométrique de la base d'entrée. Nous allons quantifier cette dépendance, tout au long de cette thèse, par la notion de valuation.

Nous avons déjà expliqué, dans le chapitre 2, pourquoi l'algorithme de Gauss se décrivait également sur les entrées $(u, v) \in \mathbb{R}^2 \times \mathbb{R}^2$. Contrairement à l'algorithme d'Euclide, dont l'extension continue a un comportement très différent de celui de son cadre originel discret, l'algorithme de Gauss apparaît avoir un comportement similaire, dans le cadre continu et dans le cadre discret. Cette thèse va montrer que le transfert diu continu au discret, même s'il est délicat, ne se heurte pas aux mêmes difficultés conceptuelles que celles qu'on rencontre dans le cas de l'algorithme d'Euclide.

L'algorithme de Gauss a probablement été conçu par... Lagrange, et a été ensuite développé et largement utilisé par Gauss. La première analyse de l'algorithme de Gauss (dans le pire des cas) a été faite par Lagarias, qui a montré que le nombre d'itérations sur Ω_N était linéaire en la taille $\log N$ des entrées. Vallée a ensuite décrit la combinatoire du plus mauvais cas, et démontré qu'elle était reliée à celle du plus mauvais cas de l'algorithme d'Euclide centré. Les premières analyses probabilistes de l'algorithme de Gauss ont commencé autour de 1990, avec les travaux de Daudé, Flajolet, Laville et Vallée. Elles s'effectuent toutes dans un modèle uniforme. L'analyse de l'exécution se limite au nombre d'itérations, mais démontre déjà que, au moins dans ce modèle, et contrairement à ce que l'analyse dans le pire des cas pouvait suggérer, l'algorithme de Gauss a un comportement sensiblement différent de celui de l'algorithme d'Euclide. L'analyse de la géométrie de la sortie se concentre sur deux paramètres –premier minimum, défaut d'Hermite–.

Cette thèse vise à donner une analyse complète de l'algorithme de Gauss, dans un modèle réaliste, qui puisse prendre en compte la variabilité de la géométrie des bases d'entrée. Nous voulons ici analyser tous les principaux paramètree de l'algorithme, qui caractérisent aussi bien l'exécution de l'algorithme (c'est la partie II de cette thèse), que sa géométrie de sortie (c'est la partie III de la thèse). Tous ces paramètres ont déjà été décrits dans le chapitre 2, et nous les repassons en revue ici : En ce qui concerne l'exécution de l'algorithme, ce sont les coûts dits additifs (qui fournissent une extension de paramètres comme le nombre d'itérations), ou la complexité binaire. Il y a aussi trois principaux paramètres qui décrivent la configuration de sortie : le premier minimum λ, le second minimum orthogonalisé μ, et le défaut d'Hermite γ. Nous avons expliqué pourquoi ce second minimum, non encore étudié, aura sans aucun doute une grande importance dans les travaux ultérieurs sur l'algorithme LLL. Nous analyserons ici ces trois paramètres, pour une distribution d'entrée générale, non nécessairement uniforme.

Les résultats précis des analyses probabilistes existantes des algorithmes d'Euclide et de Gauss seront revisités dans la section 3.2 du chapitre 3, partie II, quand nous énoncerons nos résultats.

3.1.4 L'exemple de l'algorithme LLL.

Pour l'algorithme LLL, la difficulté apparaît dès la première phase de modélisation probabiliste : elle est liée au nombre et à la diversité des applications potentielles de la réduction des réseaux, et des formes très différentes que peuvent prendre les bases associées à ces applications. Il ne peut y avoir une modélisation réaliste unique, qui constituerait une référence absolue. Au contraire, le plus raisonnable consiste à opter pour des modélisations dédiées à chaque application potentielle. Il y a aussi, et parallèlement, des modèles probabilistes qui ne sont pas liés à des applications, mais qui sont naturels, de divers points de vue. Citons-en trois : le modèle dit de la boule, où l'on tire les vecteurs de la base d'entrée uniformément et indépendamment dans la boule unité – le modèle des réseaux aléatoires, où c'est le réseau qui est considéré comme une entrée, et non plus l'une quelconque de ses bases – enfin, les modèles dits d'Ajtai, qui capturent des instances difficiles pour l'algorithme, vont aider à capturer son comportement dans le pire des cas. Nous verrons que ce modèle, une fois paramétrisé, peut donner lieu à un modèle général qui capture à la fois les instances faciles et difficiles de l'algorithme

La deuxième phase ne s'avère pas plus facile, bien au contraire. Les quelques rares analyses existantes sont toutes très grossières, car elles considèrent l'algorithme un peu comme une boîte noire et sont impuissantes à analyser vraiment sa structure fine.

Finalement, les seules analyses existantes sont des analyses assez primitives, avec un modèle peu réaliste en entrée, et une faible prise en compte de la structure même de l'algorithme. Il faut citer à ce sujet les résultats suivants. L'analyse probabiliste de l'algorithme LLL a débuté avec le résultat de Daudé et Vallée [19] qui obtient une majoration du nombre moyen d'itérations de l'algorithme en $O(n^2 \log n)$, dans le cas du modèle de la boule unité. Dans le même esprit, Akhavi a dans sa thèse [3] étudié la probabilité qu'une base aléatoire tirée dans la boule unité soit déjà LLL-réduite. Par la suite, ce travail a été étendu à d'autres modèles probabilistes (dont aucun n'est malheureusement réaliste) par Akhavi, Marckert et Rouault [5].

Quand toute analyse apparaît aussi difficile, il est encore plus indispensable que jamais de mener une campagne extensive d'expérimentations. C'est ce que Nguyen et Stehlé [60] ont fait : ils ont obtenu des résultats très intéressants sur les principaux paramètres de l'algorithme (exécution, sortie) dans des modèles réalistes, ont énoncé des conjectures, mais ... n'ont rien prouvé !

3.2 Modèles aléatoires d'entrées pour les algorithmes de réduction.

Cette section décrit les principaux modèles aléatoires qu'on peut attacher aux entrées de l'algorithme LLL. Certains s'avèrent naturels conceptuellement, tandis que d'autres sont liés aux applications potentielles. Nous avons exhibé un bon nombre de ces applications dans la section 1.3 du chapitre 1. Comme nous avons vu, la cryptologie est une source d'inspiration particulièrement bien représentée, et il est donc important de décrire les principales classes de réseaux "cryptographiques".

3.2.1 Modèles sphériques

La façon la plus naturelle de choisir un réseau "au hasard" est de choisir indépendamment p vecteurs dans la boule unité n-dimensionnelle, avec une distribution invariante par rotation. Celui-ci est le modèle sphérique introduit par la première fois dans [19], et depuis étudié en [3, 5] (Voir section 3.3). Malheureusement, ce modèle n'apparaît pas dans les applications potentielles de la réduction (sauf peut-être dans la programmation linéaire entière), mais il constitue un modèle de référence, auquel on peut comparer les modèles inspirés des applications.

Nous considérons des distributions $\nu_{(n)}$ dans $\mathbb{R}^n$ qui sont invariantes par rotation, et qui satisfont $\nu_{(n)}(0) = 0$. Ces distributions sont appelées "distributions sphériques simples". Pour une distribution sphérique simple, la partie angulaire $\theta_{(n)} := b_{(n)}/|b_{(n)}|$ est uniformément distribuée dans la sphère unité $\mathbb{S}_{(n)} := \{x \in \mathbb{R}^n \ : \ ||x|| = 1\}$. En plus, la partie radiale $|b_{(n)}|^2$ et la partie angulaire sont indépendantes. Alors, une distribution sphérique est complètement déterminée par la distribution de sa partie radiale, notée $\rho_{(n)}$.

Ici, les distributions beta et gamma jouent un rôle important, et leurs définitions et propriétés seront rappelées dans la section 3.3.1 de ce chapitre. Nous décrivons maintenant trois exemples naturels de distributions sphériques simples.

(a) Le premier exemple d'une distribution sphérique simple est la distribution uniforme dans la boule unité $\mathcal{B}_{(n)} := \{x \in \mathbb{R}^n \mid ||x|| \leq 1\}$. Dans ce cas, la distribution radiale $\rho_{(n)}$ est égale à la distribution beta $\beta(n/2, 1)$.

(b) Un second exemple est la distribution uniforme sur la sphère unitaire $\mathbb{S}_{(n)}$, où la distribution radiale $\rho_{(n)}$ est une mésure de Dirac sur $x = 1$.

(c) Un dernier exemple est lié aux distributions gaussiennes, lorsque les n coordonnées du vecteur $b_{(n)}$ sont indépendantes et distribuées avec la loi normale standard $\mathcal{N}(0, 1)$. Dans ce cas, la distribution radiale $\rho_{(n)}$ possède une densité égale à $2\gamma_{n/2}(2t)$.

Lorsque le système $B_{p,(n)}$ est formé de p vecteurs (avec $p \leq n$) qui sont pris au hasard dans $\mathbb{R}^n$, indépendamment, et avec la même distribution sphérique simple $\nu_{(n)}$, nous disons que ce système $B_{p,(n)}$ est distribué selon un "modèle sphérique". Sous ce modèle, le système $B_{p,(n)}$ (pour $p \leq n$) est presque sûrement linéairement indépendant.

3.2.2 Notion naturelle de réseau aléatoire

Il y a une notion naturelle de réseau aléatoire, introduite par Siegel [68] en 1945. L'espace des réseaux de dimension n dans $\mathbb{R}^n$ (à homothéthie près) peut être identifié avec le quotient $X_n = SL_n(\mathbb{R})/SL_n(\mathbb{Z})$. Le groupe $G_n = SL_n(\mathbb{R})$ possède une unique mesure de Haar bi-invariante, qui se projette sur une mesure finie dans l'espace X_n. Cette mesure η_n (qu'on normalise en une mesure de probabilité) est par définition l'unique probabilité en X_n qui est invariante sous

l'action de G_n : si $A \subseteq X_n$ est mesurable et $g \in G_n$, alors $\chi_n(A) = \chi_n(gA)$. Ceci définit naturellement une notion de réseau aléatoire. Nous allons révenir sur cette notion dans l'étude du cas bidimensionnel, dans la section 2.2 de la partie III.

3.2.3 Les bases d'Ajtai

Elles ont été introduites par Ajtai pour modéliser des instances difficiles vis-à-vis de la réduction [1]. Considérons une famille d'entiers $a := (a_{i,p})$ définie pour $1 \leq i \leq p$, satisfaisant pour tout i,

$$\frac{a_{i+1,p}}{a_{i,p}} \to 0 \quad \text{lorsque } p \to \infty.$$

Historiquement, Ajtai a choisi de travailler avec la suite $a_{i,p} := 2^{(2p-i+1)^a}$, avec un réel $a > 1$, qui fournit une suite rapidement décroissante.

Une suite de bases d'Ajtai $\mathcal{B}_a := (B_p(a))$ relative à la famille $a = (a_{i,p})$ est définie comme suit : la base B_p est de dimension p et elle est formée par les vecteurs $b_{i,p} \in \mathbb{Z}^p$ de la forme

$$b_{i,p} = a_{i,p}\, e_i + \sum_{j=1}^{i-1} a_{i,j,p}\, e_j \qquad \text{avec} \qquad a_{i,j,p} = \text{rand}\left(-\frac{a_{j,p}}{2}, \frac{a_{j,p}}{2}\right) \quad \text{pour } j < i,$$

où e_j, $1 \leq j \leq p$ est la base canonique de $\mathbb{R}^p$. Comme la base est sous forme triangulaire, le coefficient $m_{i,j}$ et les longueurs ℓ_i se lisent directement sur l'entrée : le coefficient $m_{i,j}$ est égal à $a_{i,j,p}/a_{j,p}$ tandis que ℓ_i est égal à $a_{i,p}$. Ces bases sont donc déjà réduites en taille, mais en taille seulement, car, par contre, tous les rapports de Siegel $r_{i,p}$ d'entrée, égaux à $a_{i+1,p}/a_{i,p}$, tendent vers 0 lorsque p tend vers l'infini, et sont donc très loin de satisfaire la minoration de Siegel $r_{i,p} \geq 1/s$. C'est pourquoi de telles bases ont été utilisées par Ajtai en [1] pour montrer l'optimalité des bornes du pire des cas fournies en [46].

Dans la définition initiale de la distribution d'Ajtai, les longueurs ℓ_i des orthogonalisés de Siegel sont fixes. Les seules variables sont alors les coefficients $m_{i,j}$ qui sont choisis aléatoirement dans $(-1/2, +1/2)$. L'idée sous-jacente à ce modèle est que ce sont les longueurs ℓ_i qui sont importantes dans l'algorithme, et non pas les coefficients $m_{i,j}$ qui jouent un rôle moindre.

Mais, on peut aussi, si l'on veut des instances de difficulté variable, travailler avec des rapports de Siegel $r_{i,p}$ de comportement variable. Notre idée est de les choisir selon une loi puissance de la forme

$$\Pr\left[r_{i,p} \leq \frac{x}{s}\right] = x^{1+\theta_{i,p}}, \qquad \text{avec } x \in [0,1], \quad \theta_{i,p} \to -1, \quad \text{quand} \quad p \to \infty.$$

Ces distributions seront appelées (dans la thèse) distributions d'Ajtai de paramètre $\theta = (\theta_{i,p})$. Ces bases ne sont jamais réduites, car aucune des conditions de réduction de Siegel n'est satisfaite, mais elles peuvent être de difficulté variable en fonction de l'exposant de la loi puissance. La loi du rapport $r_{i,p}$ admet une densité proportionnelle à $x^{\theta_{i,p}}$, et l'exposant $\theta_{i,p}$, qui jouera un rôle très important dans la suite de cette thèse, sera appelé la *valuation* de cette distribution. Ainsi quand θ est grand, les bases sont faciles à réduire, alors que, lorsque θ tend vers -1, les bases deviennent très difficiles à réduire. La valuation est le paramètre essentiel du modèle aléatoire dans lequel nous analysons l'algorithme de Gauss, et qui généralise le modèle d'Ajtai. Il est présenté dans le chapitre 1 de la partie II.

$$\begin{bmatrix} 1 & 0 & \cdots & 0 & -a_1 \\ 0 & 1 & \cdots & 0 & -a_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -a_n \\ 0 & 0 & \cdots & 0 & s \end{bmatrix} \qquad \begin{bmatrix} c & 0 & \cdots & 0 & \Re(\bar{\alpha}^0) & \Im(\bar{\alpha}^0) \\ 0 & c & \cdots & 0 & \Re(\bar{\alpha}^1) & \Im(\bar{\alpha}^1) \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & c & \Re(\bar{\alpha}^m) & \Im(\bar{\alpha}^m) \end{bmatrix} \tag{3.1}$$

$$\begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \\ \alpha_1 & \alpha_2 & \cdots & \alpha_n & \epsilon/Q. \end{bmatrix} \qquad \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \\ x_0^k & \binom{k}{1} x_0^{k-1} & \cdots & \binom{k}{k-1} x_0 & n \end{bmatrix} \tag{3.2}$$

$$A_(q,h) := \begin{bmatrix} qI_n & 0_n \\ M_n(h) & I_n \end{bmatrix} \qquad \text{avec} \quad M_n(h) := \begin{bmatrix} h_1 & h_2 & h_3 & \cdots & h_n \\ h_n & h_1 & h_2 & \cdots & h_{n-1} \\ h_{n-1} & h_n & h_1 & \cdots & h_{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ h_2 & h_3 & h_4 & \cdots & h_1 \end{bmatrix}. \tag{3.3}$$

FIGURE 3.1 – Différents types de bases de réseaux utilisés dans les applications.

3.2.4 Réseaux des applications : Variantes des bases sac-à-dos et de ses transposées

Les bases issues d'applications sont toutes structurées, et toutes formées par des matrices identité dont on rajoute l'information particulière dans les bords. Nous les avons déjà rencontrées au chapitre 1 et elles sont rappelées dans la figure 3.1.

Ce type réunit les bases qui arrivent naturellement dans les applications suivantes :

(*a*) Les bases de type sac-à-dos sont constituées par les lignes des matrices de (3.1) de la figure 3.1. A gauche, les réseaux sac-à-dos proprement dits, les composantes $(a_1, a_2, \ldots, a_p)$ sont tirés indépendamment et uniformement dans l'intervalle $[-s, s]$. Des telles bases apparaissent souvent en cryptanalyse, dans des cryptosystèmes reposant sur la difficulté du problème du sac-à-dos. De bases semblables apparaissent en théorie des nombres pour la reconstruction du polynôme minimal (voir (3.1) à droite) et la détection de relations entières entre nombres réels.

(*b*) Les bases relatives aux transposées des matrices précédentes, décrites dans (3.2) de la figure 3.1 apparaissent dans la recherche de relations diophantiennes simultanées (dans ce cas $q \in \mathbb{Z}$, voir (3.2) à gauche) ou en géométrie discrète ($q = 1$). Des matrices semblables apparaissent dans la recherche de racines k–ıèmes (voir (3.2) à droite)

(*c*) Le cryptosystème NTRU se décrit dans le contexte des polynômes sur un corps fini, mais la clé secrète peut être vue comme la base du réseau spécifié par les colonnes de la matrice $(2p \times 2p)$ décrite dans (3.3) de la figure 3.1 où q est une petite puissance de 2 et les coefficients h_i sont des entiers de l'intervalle $]-q/2, q/2]$.

3.2.5 Modèles probabilistes continus ou discrets

Mis à part deux modèles –le modèle sphérique, ou le modèle des réseaux aléatoires– qui sont des modèles *continus*, tous les autres (le modèle d'Ajtai ou les modèles de type sac-à-dos) sont

des modèles discrets. Dans ces cas-là, il est naturel de construire des modèles probabilistes qui préservent la "forme" des matrices et qui remplacent les coefficients discrets par des coefficients continus. Ceci permet d'utiliser toutes les outils du calcul dans les analyses probabilistes, tout en obtenant des conclusions pour les modèles discrets.

(*a*) Un premier exemple est le modèle d'Ajtai relatif à la séquence $a := (a_{i,p})$, pour lequel la version continue de dimension p est la suivante

$$b_{i,p} = a_{i,p}e_i + \sum_{j=1}^{i-1} x_{i,j,p}a_{j,p}e_j \qquad \text{avec} \qquad x_{i,j,p} = \text{rand}(-1/2, 1/2) \quad \text{pour tout } j < i \leq p.$$

(*b*) On peut aussi remplacer le modèle discret associé aux bases sac-à-dos de la figure 3.1 (*a*) par le modèle continu où $A = (a_1, a_2, \ldots, a_n)$ est remplacée par un vecteur réel x uniformément distribué dans la boule $||x||_\infty \leq 1$ et I_p est remplacé par ρI_p, avec une petite constante positive $0 < \rho < 1$. Informellement, choisir des matrices aléatoires continues indépendamment et uniformément parmi les matrices de même "forme", conduit à une classe de réseaux de "forme sac-à-dos".

Remarque. Il n'est pas clair que des tels réseaux ayant une forme sac-à-dos partagent toutes les propriétés qui sont propres aux réseaux sac-à-dos qui viennent des applications, tout particulièrement l'existence d'un vecteur particulièrement court (de longueur beaucoup plus petite que la borne garantie par le théorème de Minkowski 1.3).

Réciproquement, nous pouvons associer à n'importe quel modèle continu un modèle discret : considérons un domaine $\mathcal{B} \subset \mathbb{R}^n$ avec une frontière "régulière" (continue, différentiable). Pour tout entier N, nous pouvons "remplacer" une distribution continue dans le domaine $\mathcal{B}$ relatif à une densité f de classe C^1 par la distribution dans le domaine discret

$$\mathcal{B}_N := \mathcal{B} \cap \frac{\mathbb{Z}^n}{N},$$

défini par la restriction f_N de f à $\mathcal{B}_N$. Lorsque $N \to \infty$, la distribution relative à la densité f_N tend vers la distribution relative à f, grâce au principe de Gauss, qui met en relation le volume d'un domaine $\mathcal{A} \subset \mathcal{B}$ (avec une frontière régulière $\partial\mathcal{A}$) et le nombre de points dans le domaine $\mathcal{A}_N := \mathcal{A} \cap \mathcal{B}_N$,

$$\frac{1}{N^n}\text{card}(\mathcal{A}_N) = \text{Vol}(\mathcal{A}) + O\left(\frac{1}{N}\right)\text{Area}(\partial\mathcal{A}).$$

Nous pouvons appliquer ce cadre à n'importe quel modèle sphérique simple, voire aux modèles introduits dans le cas bidimensionnel.

Dans le même esprit, on peut considérer une version discrète de la notion de réseau aléatoire : considérons l'ensemble $\mathcal{L}(n, N)$ des réseaux entiers n-dimensionnels de déterminant N. Tout réseau de $\mathcal{L}(n, N)$ peut être transformé dasn un réseau de X_n (défini en 3.2.2) par l'homothétie ψ_N de rayon $N^{-1/n}$. Goldstein et Mayer [27] montrent le résultat suivant : *Pour tout ensemble mesurable $A \subseteq X_n$ dont le bord ∂A vérifie $\chi_n[\partial A] = 0$, la fraction de réseaux de $\mathcal{L}(n, N)$ dont l'image par ψ_N appartient à A tend vers $\chi_n(A)$ lorsque N tend vers l'infini.* Autrement dit, l'image par ψ_N de la probabilité uniforme dans $\mathcal{L}(n, N)$ tend vers la mesure χ_n. Ainsi, pour engendrer des réseaux aléatoires dans un sens naturel, il suffit d'engendrer uniformément et au hasard un réseau dans $\mathcal{L}(n, N)$ pour N grand. Ceci est particulièrement facile lorsque $N = q$ est un nombre premier. En effet, lorsque q est un grand premier, la plupart des réseaux dans $\mathcal{L}(n, q)$ sont des réseaux engendrés par les lignes des matrices décrites dans la figure 3.1 (*d*), où les composantes x_i du vecteur x (avec $i \in [\![1, n-1]\!]$) sont choisies indépendamment et uniformément dans $\{0, \ldots, q-1\}$.

3.3 Analyses existantes dans les modèles sphériques.

Dans cette section, la dimension de l'espace ambiant est notée n, la dimension du réseau est notée p, et une base d'un réseau de dimension p dans $\mathbb{R}^n$ est notée $B_{p,(n)}$. La codimension g, égale par définition à $n-p$, joue ici un rôle fondamental. Nous considérons le cas où n tend vers l'infini alors que $g := g(n)$ est une fonction fixe de n (avec $g(n) \leq n$). Nous sommes intéressés dans les questions suivantes :

(*a*) Considérons un réel $s > 1$. Quelle est la probabilité $\pi_{p,(n),s}$ qu'une base aléatoire $B_{p,(n)}$ soit déjà s-réduite dans le sens de Siegel (voir 2.16) ?

(*b*) Considérons un réel $t > 1$. Quel est le nombre espéré d'itérations de l'algorithme LLL(t) pour une base aléatoire $B_{p,(n)}$?

(*c*) Quelle est la valeur moyenne du premier minimum du réseau engendré par une base aléatoire $B_{p,(n)}$?

Cette section répond à ces questions dans le cas où $B_{p,(n)}$ est choisi au hasard selon un modèle sphérique, et montre qu'il y a deux cas distincts selon la valeur de la codimension $g := n - p$.

3.3.1 Intégrales eulériennes : fonctions gamma et beta

Dans les calculs liés aux densités de probabilité de ce chapitre et de la thèse en général, nous allons rencontrer les fonctions gamma et beta. Ici nous rappelons très brièvement leurs définitions ainsi que quelques unes de leurs propriétés. Nous suivons l'exposé de Flajolet et Sedgewick dans [23]. La fonction gamma a été définie par Euler comme l'intégrale

$$\Gamma(s) = \int_0^\infty e^{-t} t^{s-1} dt,$$

où l'intégrale converge lorsque $\Re(s) > 0$. Une intégration par parties fournit l'équation fonctionnelle de base $\Gamma(s+1) = s\Gamma(s)$ Les relations $\Gamma(1) = 1$, $\Gamma(n+1) = n!$ prouvent que la fonction gamma étend la factorielle à des arguments non entiers. La valeur

$$\Gamma\left(\frac{1}{2}\right) := \int_0^\infty e^{-t} \frac{dt}{\sqrt{t}} = 2\int_0^\infty e^{-x^2} dx = \sqrt{\pi},$$

apparait souvent dans les calculs de la thèse. Par ailleurs, la fonction beta $B(a,b)$, définie pour des réels $a > 0$ et $b > 0$, est l'intégrale

$$B(a,b) := \int_0^1 x^{a-1}(1-x)^{b-1} dx,$$

qui est liée à la fonction gamma par la formule

$$B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}.$$

Chacune de ces fonctions possède une densité de probabilité qui lui est associée. Pour des nombres réels strictement positifs $a, b \in \mathbb{R}^+$, la distribution beta de paramètres (a,b), notée $\beta_{(a,b)}$ et la distribution gamma de paramètre a notée $\gamma(a)$ admetttent des densités de la forme

$$\beta_{a,b}(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}\mathbf{1}_{(0,1)}(x), \qquad \gamma_a(x) = \frac{e^{-x}x^{a-1}}{\Gamma(a)}\mathbf{1}_{[0,+\infty)}(x).$$

Les intégrales des densités beta apparaîtront dans les calculs des constantes de normalisation des densités, dans la section 1.3 de la partie II, ainsi que dans les calculs liés aux distributions des paramètres de la sortie, dans le chapitre 2 de la partie III.

3.3.2 Principaux paramètres

Soit $B_{p,(n)}$ un système linéairement indépendant de vecteurs de $\mathbb{R}^n$ dont la codimension est $g = n - p$. Soit $B^{\star}_{p,(n)}$ la base de Gram-Schmidt associée. Nous sommes intéressés par comparer les longueurs de deux vecteurs successifs du système orthogonal. Nous allons introduire plusieurs paramètres reliés à la réduction de Siegel du système $B_{p,(n)}$.

Définition 3.1. *À un système $B_{p,(n)}$ de p vecteurs de $\mathbb{R}^n$, nous associons le système de Gram-Schmidt $B^{\star}_{p,(n)}$ et la séquence $\underline{r}_{j,(n)}$ de rapports de Siegel définie par*

$$\underline{r}_{j,(n)} := \frac{\ell_{n-j+1,(n)}}{\ell_{n-j,(n)}}, \qquad \textit{for } g+1 \leq j \leq n-1,$$

ainsi que les deux autres paramètres

$$\mathcal{M}_{g,(n)} := \min\{\underline{r}^2_{j,(n)}; \quad g+1 \leq j \leq n-1\} \qquad \mathcal{I}_{g,(n)} := \min\left\{j : \underline{r}^2_{j,(n)} = \mathcal{M}_{g,(n)}\right\}.$$

Le paramètre $\mathcal{M}_{g,(n)}$ est le niveau de réduction*, et le paramètre $\mathcal{I}_{g,(n)}$ est* l'indice de pire réduction locale.

Remarque 3.1. Le rapport de Siegel $\underline{r}_{j,(n)}$ est fortement relié au rapport r_i du chapitre précédent. Il y a néanmoins deux différences : le rôle de la dimension n de l'espace ambiant apparaît nettement, et les indices i et j sont reliés via $\underline{r}_j := r_{n-j}$. Le rôle de cette "inversion du temps" sera expliquée plus tard. La variable $\mathcal{M}_{g,(n)}$ est la borne supérieure de l'ensemble des $1/s^2$ pour lesquels la base $B_{n-g,(n)}$ est s–réduite au sens de Siegel. Autrement dit, $1/\mathcal{M}_{g,(n)}$ est la borne inférieure des valeurs de s^2 pour lesquels la base $B_{n-g,(n)}$ est s–réduite au sens de Siegel. Cette variable est reliée à notre problème initial, grâce à l'égalité

$$\pi_{n-g,(n),s} := \Pr[B_{n-g,(n)}\text{est } s\text{–réduite}] = \Pr[\mathcal{M}_{g,(n)} \geq \frac{1}{s^2}],$$

et nous voulons évaluer la distribution limite (si elle existe) de $\mathcal{M}_{g,(n)}$ lorsque $n \to \infty$. La seconde variable $\mathcal{I}_{g,(n)}$ désigne le plus petit indice j pour lequel la condition de Siegel relative à l'indice $n - j$ est la plus faible. Alors, $n - \mathcal{I}_{g,(n)}$ est le plus grand indice i pour lequel la condition de Siegel relative à l'indice i est la plus faible. Cet indice est celui de la base locale la moins réduite.

Lorsque le système $B_{p,(n)}$ est choisi aléatoirement, les rapports de Siegel, le niveau de réduction et l'indice de pire réduction locale sont des variables aléatoires, bien définies lorsque $B_{p,(n)}$ est un système linéairement indépendant. Nous voulons étudier le comportement asymptotique de ces variables aléatoires (par rapport à la dimension n de l'espace ambiant), lorsque le système $B_{p,(n)}$ est distribué selon un modèle sphérique concentré, où la distribution radiale $\rho_{(n)}$ satisfait la *propriété de concentration* suivante :

Il existe une suite $(a_n)_n$ et des constantes $d_1, d_2, \alpha > 0$, $\theta_0 \in (0,1)$ telles que, pour tout n et tout $\theta \in (0, \theta_0)$, la fonction distribution $\rho_{(n)}$ satisfait

$$\rho_{(n)}\left(a_n(1+\theta)\right) - \rho_{(n)}\left(a_n(1-\theta)\right) \geq 1 - d_1 e^{-n d_2 \theta^{\alpha}}. \tag{3.4}$$

Dans ce cas, il est possible de transférer des résultats obtenus pour la distribution uniforme sur $\mathbb{S}_{(n)}$ [où la distribution radiale est une Dirac] aux distributions sphériques plus générales, pourvu que la distribution radiale soit suffisament concentrée. Cette *propriété de concentration* $\mathcal{C}$ est remplie dans les trois cas principaux de distributions sphériques simples que nous avons décrit dans la section 3.1 de ce chapitre.

3.3.3 Probabilité qu'une base d'entrée soit déjà réduite.

Nous allons tout d'abord rappeler quelques notions de probabilités, et fixer des notations avant d'énoncer le premier résultat principal.

Définition 3.2. *Une suite (X_n) de variables aléatories réelles* converge en distribution *vers la variable aléatoire réelle X si et seulement si la fonction répartition F_n de X_n converge ponctuellement à la fonction distribution F de X dans l'ensemble de points de continuité de F. Une suite (X_n) de variables aléatoires réelles converge en probabilité à une constante a si pour tout $\epsilon > 0$, la suite $\Pr[|X_n - a| > \epsilon]$ tends vers 0. Les deux situations sont respectivement notées par*

$$X_n \xrightarrow[n]{(d)} X, \qquad\qquad X_n \xrightarrow[n]{proba.} a.$$

Le premier résultat central de cette section est le suivant :

Théorème 3.1 (Akhavi, Marckert, Rouault [5] 2005)**.** *Soit $B_{p,(n)}$ une base aléatoire de codimension $g := n - p$ sous un modèle sphérique concentré. Soit $s > 1$ un paramètre réel, et supposons que la dimension n de l'espace ambiant tends vers l'infini. Alors,*

(i) Si $g := n - p$ tends vers l'infini, alors la probabilité $\pi_{p,(n),s}$ que $B_{p,(n)}$ soit s–réduite tends vers 1.

(ii) Si $g := n - p$ est constant, alors la probabilité $\pi_{p,(n),s}$ que $B_{p,(n)}$ soit s–réduite converge à une constante en $(0, 1)$ (dépendant en s et g). En plus, l'indice de pire réduction locale $\mathcal{I}_{g,(n)}$ converge en distribution.

3.3.4 Lois β pour les rapports de Siegel.

Les lois beta et gamma apparaissent très fréquemment lorsque l'on travaille avec l'orthogonalisation de Gram-Schmidt. Nous commençons par étudier les variables $Y_{j,(n)}$ définies par

$$Y_{j,(n)} := \frac{\ell^2_{j,(n)}}{|b_{j,(n)}|^2} \qquad \text{for } j \in [2..n].$$

et on montre qu'elles admettent des lois beta.

Proposition 3.1 (Akhavi, Marckert, Rouault [5] 2005)**.** *(i) Dans tout modèle sphérique, les variables $\ell^2_{j,(n)}$ sont indépendantes. En plus, la variable $Y_{j,(n)}$ suit la loi beta $\beta((n - j + 1)/2, (j-1)/2)$, pour $j \in [2..n]$, et toutes les variables de l'ensemble $\{Y_{j,(n)}, |b_{k,(n)}|^2;\ (j, k) \in [2..n] \times [1..n]\}$ sont indépendantes.*

(ii) Sous le modèle de la boule aléatoire $\mathbb{U}_n$, la variable $\ell^2_{j,(n)}$ suit la loi beta $\beta((n-j+1)/2, (j+1)/2)$.

La proposition 3.1 va permettre de montrer que, sous un modèle sphérique concentré, les lois beta et gamma vont jouer un rôle central dans l'analyse des principaux paramètres introduits dans la définition 3.2.

Notons $(\eta_i)_{i\geq 1}$ une séquence de variables aléatoires indépendantes où η_i suit une loi gamma $\gamma(i/2)$ et considérons, pour $k \geq 1$, les variables aléatoires suivantes

$$\mathcal{R}_k = \eta_k/\eta_{k+1}, \qquad \mathcal{M}_k = \min\{\mathcal{R}_j;\quad j \geq k+1\}, \qquad \mathcal{I}_k = \min\{j \geq k+1;\quad \mathcal{R}_j = \mathcal{M}_k\}.$$

Dans la suite nous allons montrer que ces variables interviennent comme les limites des variables (du même nom) de la définition 3.2 (ou 3.1). Plusieurs arguments interviennent dans cette preuve :

(*a*) Observons d'abord que, pour les indices de la forme $n-i$ avec i fixe, la variable $\underline{r}^2_{n-i,(n)}$ tends vers 1 lorsque $n \to \infty$. Il est alors convenable d'étendre la tuple $(\underline{r}_{j,(n)})$ (définie uniquement pour $j \leq n-1$) dans une suite infinie en posant $\underline{r}_{k,(n)} := 1$ pour tout $k \geq n$.

(*b*) Ensuite, la convergence

$$\mathcal{R}_j \xrightarrow[j]{a.s.} 1, \qquad \sqrt{k}(\mathcal{R}_k - 1) \xrightarrow[k]{(d)} \mathcal{N}(0,4),$$

nous amène à considérer la suite $(\mathcal{R}_k - 1)_{k\geq 1}$ comme un élément de l'espace $\mathcal{L}_q$, pour $q > 2$. Nous rappelons que

$$\mathcal{L}_q := \{x, \|x\|_q < +\infty\}, \qquad \text{with} \quad \|x\|_q := \big(\sum_{i\geq 1} |x_i|^q\big)^{1/q}, \qquad \text{for} \quad x = (x_i)_{i\geq 1}.$$

(*c*) Enfin, des résultats classiques sur des variables indépendantes distribuées selon des lois gamma et beta, ainsi que la loi des grands nombres et la proposition 3.1 prouvent que

$$\text{Pour tout } j \geq 1, \quad \underline{r}^2_{j,(n)} \xrightarrow[n]{(d)} \mathcal{R}_j. \tag{3.5}$$

Cela suggère que le minimum $\mathcal{M}_{g,(n)}$ est atteint par les variables $\underline{r}^2_{j,(n)}$ correspondantes aux plus petits indices j, motivant ainsi l'inversion temporelle faîte dans la définition 3.2 (ou autre, c'est à confirmer).

3.3.5 Le processus limite

Il est alors possible de prouver que les processus $R_{(n)} := (\underline{r}_{k,(n)} - 1)_{k\geq 1}$ convergent (en distribution) vers le processus $R := (\mathcal{R}_k - 1)_{k\geq 1}$ dans l'espace $\mathcal{L}_q$, lorsque la dimension n de l'espace ambient tend vers l'infini ∞. Puisque $\mathcal{M}_{g,(n)}$ et $\mathcal{I}_{g,(n)}$ sont des fonctionnelles continues du processus $R_{(n)}$, elles convergent aussi en distribution respectivement vers $\mathcal{M}_g$ et $\mathcal{I}_g$.

Théorème 3.2 (Akhavi, Marckert, Rouault [5] 2005)**.** *Pour toute distribution sphérique concentrée, nous avons*

(*i*) *La convergence* $(\underline{r}^2_{k,(n)} - 1)_{k\geq 1} \xrightarrow[n]{(d)} (\mathcal{R}_k - 1)_{k\geq 1}$ *est vérifiée dans tout espace* $\mathcal{L}_q$*, avec* $q > 2$.

(*ii*) *Pour tout* k *fixé, nous avons :* $\mathcal{M}_{k,(n)} \xrightarrow[n]{(d)} \mathcal{M}_k, \qquad \mathcal{I}_{k,(n)} \xrightarrow[n]{(d)} \mathcal{I}_k.$

(*iii*) *Pour toute suite* $n \mapsto g(n)$ *avec* $g(n) \leq n$ *et* $g(n) \to \infty$*, on a :* $\mathcal{M}_{g(n),(n)} \xrightarrow[n]{proba.} 1$.

Ce résultat résout le problème original et prouve le théorème 3.1. Maintenant nous donnons quelques précisions sur les processus limites $\sqrt{\mathcal{R}}_k, \sqrt{\mathcal{M}}_k$, et nous décrivons quelques propriétés des fonctions répartition F_k de $\sqrt{\mathcal{M}}_k$, qui est d'un intérêt particulier, grâce à l'égalité $\lim_{n\to\infty} \pi_{n-k,(n),s} = 1 - F_k(1/s)$.

Proposition 3.2 (Akhavi, Marckert, Rouault [5] 2005)**.** *Les processus limites* $\sqrt{\mathcal{R}}_k, \sqrt{\mathcal{M}}_k$ *admet tent des densités satisfaisant les propriétés suivantes :*

(*i*) *Pour tout* k*, la densité* φ_k *de* $\sqrt{\mathcal{R}}_k$ *est*

$$\varphi_k(x) = 2B\left(\frac{k}{2}, \frac{k+1}{2}\right) \frac{x^{k-1}}{(1+x^2)^{k+(1/2)}} \, \mathbf{1}_{[0,\infty[}(x), \qquad \textit{with} \quad B(a,b) := \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}. \tag{3.6}$$

(*ii*) *Pour tout k, les variables aléatoires $\sqrt{\mathcal{M}_k}, \mathcal{M}_k$ ont des densités, qui sont positives sur $(0,1)$ et nulles ailleurs. Les fonctions répartition F_k, G_k satisfont pour x près de 0, et pour chaque k,*

$$\Gamma\left(\frac{k+2}{2}\right) F_k(x) \sim x^{k+1}, \qquad G_k(x) = F_k(\sqrt{x}).$$

Il existe τ tel que pour chaque k, et pour $x \in [0,1]$ satisfaisant $|x^2 - 1| \leq (1/\sqrt{k})$

$$0 \leq 1 - F_k(x) \leq \exp\left[-\left(\frac{\tau}{1-x^2}\right)^2\right].$$

(*iii*) *Pour tout k, le cardinal de l'ensemble $\{j \geq k+1; \quad \mathcal{R}_j = \mathcal{M}_k\}$ est presque sûrement égal à 1.*

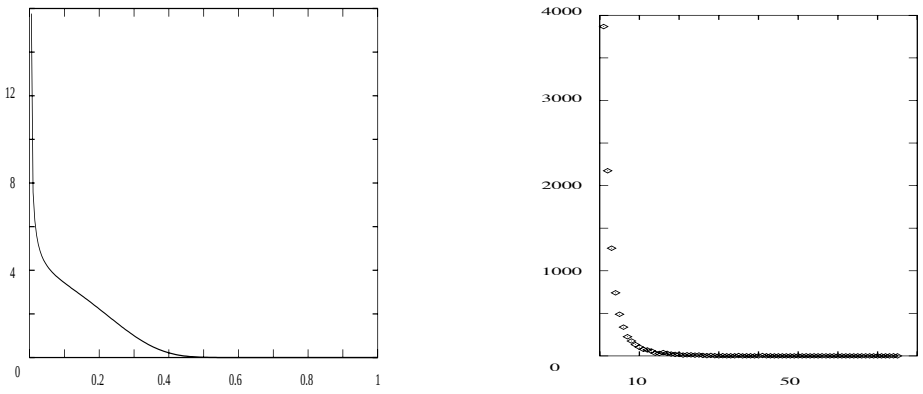

FIGURE 3.2 – À gauche : simulation de la densité de $\mathcal{M}_0$ avec 10^8 expériences. À droite : l'histogramme de $\mathcal{I}_0$, construit à partir de 10^4 simulations. Pour tout g, la suite $k \mapsto \Pr[\mathcal{I}_g = k]$ paraît decroître rapidement.

En particulier, pour un réseau plein (i.e., de dimension $p = n$),

$$\lim_{n\to\infty} \pi_{n,(n),s} \sim_{s\to\infty} 1 - \frac{1}{s}, \qquad \lim_{n\to\infty} \pi_{n,(n),s} \leq \exp\left[-\left(\frac{\tau s^2}{s^2-1}\right)^2\right] \qquad \text{lorsque } s \to 1$$

La figure 3.2 montre des expériences dans le cas d'un réseau plein ($g = 0$). Dans ce cas, la densité g_0 de $\mathcal{M}_0$ est égale à $\Theta(1/\sqrt{x})$ lorsque $x \to 0$ et tend rapidement vers 0 lorsque $x \to 1$. Par ailleurs, la même figure montre que l'indice de pire réduction pour un réseau plein est presque toujours très petit, ce qui veut dire que le premier indice i pour lequel le test du pas 2 de l'algorithme LLL (voir figure 2.5) n'est pas vérifié est un indice très proche de n.

Ces méthodes probabilistes ne fournissent aucune information sur la vitesse de convergence de $\pi_{n-g,(n)}$ vers 1 lorsque n et g tendent vers l'infini. Dans le cas du modèle de la boule aléatoire, Akhavi travaille directement avec la loi beta des variables ℓ_i et observe que

$$1 - \pi_{p,(n),s} \leq \sum_{i=1}^{p-1} \Pr[\ell_{i+1} \leq \frac{1}{s}\ell_i] \leq \sum_{i=1}^{p-1} \Pr[\ell_{i+1} \leq \frac{1}{s}] \leq \sum_{i=1}^{p-1} \exp\left[\frac{n}{2} H\left(\frac{i}{n}\right)\right] \left(\frac{1}{s}\right)^{n-i},$$

où H est la fonction entropie binaire, définie par $H(x) = -x\log x - (1-x)\log(1-x)$, pour $x \in [0,1]$, qui satisfait $0 \leq H(x) \leq \log 2$. Cela prouve la proposition suivante.

Proposition 3.3 (Akhavi [4] 2000). *Sous le modèle de la boule aléatoire, la probabilité qu'une base $B_{p,(n)}$ soit réduite satisfait, pour chaque n, pour chaque $p \leq n$ et pour chaque $s > 1$,*

$$1 - \pi_{p,(n),s} \leq \frac{1}{s-1} \ (\sqrt{2})^n \ \left(\frac{1}{s}\right)^{n-p}.$$

En particulier, pour tout $s > \sqrt{2}$, la probabilité que $B_{cn,(n)}$ soit s–réduite tend exponentiellement vers 1, pourvu que $1 - c$ soit plus grand que $1/(2 \lg s)$.

3.3.6 Une première analyse probabiliste de l'algorithme LLL

Dans le cas du modèle de la boule aléatoire, Daudé et Vallée travaillent directement avec la loi beta des variables ℓ_i et ils obtiennent des estimations, à la fois pour le nombre moyen d'itérations K et pour le premier minimum $\lambda(\mathcal{L})$. L'article [19] considère uniquement le cas des réseaux pleins (ceux pour lesquels $p = n$) mais il est facile de généraliser la preuve à une dimension $p \leq n$ quelconque. En utilisant des propriétés de la fonction beta, Daudé et Vallée obtiennent d'abord une estimation simple pour la distribution de la longueur ℓ_i,

$$\Pr[\ell_i \leq u] \leq (u\sqrt{n})^{n-i+1}.$$

Ils en déduisent des informations sur la distribution de la variable aléatoire $a := \min \ell_i$,

$$\Pr[a \leq u] \leq \sum_{i=1}^{p} \Pr[\ell_i \leq u] \leq (2\sqrt{n})u^{n-p+1}, \qquad \mathbb{E}\left[\log\left(\frac{1}{a}\right)\right] \leq \frac{1}{n-p+1}\left[\frac{1}{2}\log n + 2\right].$$

Le résultat suivant se déduit alors de la majoration (2.29), et montre que, comme précédemment, il y a deux régimes selon la position de la dimension p du réseau par rapport à la dimension n de l'espace ambiant.

Théorème 3.3 (Daudé et Vallée [19] 1994). *Sous le modèle de la boule aléatoire, le nombre moyen d'itérations K de l'algorithme LLL(t) dans la base $B_{p,(n)}$ satisfait l'inégalité*

$$\mathbb{E}_{p,(n)}[K] \leq p - 1 + \frac{p(p-1)}{n-p+1}\left(\frac{1}{\log t}\right)\left[\frac{1}{2}\log n + 2\right],$$

De plus, le premier minimum du réseau engendré par $B_{p,(n)}$ satisfait l'inégalité

$$\mathbb{E}_{p,(n)}[\lambda(\mathcal{L})] \geq \frac{n-p+1}{n-p+2}\left(\frac{1}{2\sqrt{n}}\right)^{1/(n-p+1)}$$

Dans le cas où $p = cn$, avec $c < 1$, on obtient

$$\mathbb{E}_{cn,(n)}[K] \leq \frac{cn}{1-c}\left(\frac{1}{\log t}\right)\left[\frac{1}{2}\log n + 2\right], \quad \mathbb{E}_{cn,(n)}[\lambda(\mathcal{L})] \geq \exp\left[\frac{-1}{2(1-c)n}\log(4n)\right].$$

Ce résultat montre que dans le cas général $p \leq n$, le nombre moyen d'itérations est d'ordre $O(n^2 \log n)$, mais qu'il devient d'ordre $O(n^{2-a} \log n)$ lorsque la codimension $g = n - p$ est $\Omega(n^a)$; il est en particulier d'ordre $O(n \log n)$ lorsque quand la dimension p est de la forme $p = cn$ avec $c < 1$. La valeur moyenne de la variable $[1 - \lambda(\mathcal{L})]$ devient d'ordre $O(n^{-a} \log n)$ lorsque la codimension $g = n - p$ est $\Omega(n^a)$. Elle est en particulier d'ordre $O(n^{-1} \log n)$ lorsque quand la dimension p est de la forme $p = cn$ avec $c < 1$.

3.3.7 Lois puissances pour les rapports de Siegel de la fin.

Dans le modèle sphérique, et lorsque la dimension n tend vers l'infini, toutes les bases locales (sauf peut-être les dernières) sont déjà réduites au sens de s-Siegel. Pour les bases locales de la fin, avec des indices $i := n - k$, pour $k \geq 1$ fixe, le rapport de Siegel r_{n-k} admet une densité φ_k qui est décrite dans la proposition 3.2. Dans les deux cas $x \to 0$ et $x \to \infty$, la densité φ_k a un comportement de type puissance,

$$\varphi_k(x) = \Theta(x^{k-1}), \qquad \text{pour } x \to 0, \qquad \varphi_k(x) = \Theta(x^{-k-2}) \quad \text{pour } x \to \infty.$$

Les bases locales correspondant au modèle d'Ajtai ont un rapport de Siegel qui admet une densité initiale de type puissance (voir section 3.2.3 de ce chapitre), avec un exposant θ, appelé la valuation, qui est choisi proche de -1 pour des instances vraiment difficiles. On voit ici que, dans les modèles sphériques, même les dernières bases locales, qui sont non réduites en général, rentrent dans le modèle d'Ajtai, même si elles n'en constituent pas des instances vraiment difficiles, puisque leur exposant est toujours au moins égal à 0.

3.4 Résultats expérimentaux et conjectures sur le comportement probabiliste de l'algorithme.

Nguyen et Stehlé ont utilisé leur programmation, à la fois très efficace et prouvée, de l'algorithme LLL en virgule flottante [58] pour conduire des expérimentations extensives dans deux types de bases importants : les bases d'Ajtai et les bases de type sac-à-dos. Leurs résultats sont décrits dans l'article [60]. Ces deux types de bases constituent chacun des extrêmes vis à vis des algorithmes de réduction, puisque les bases de type sac-à-dos sont assez faciles à réduire, tandis que les bases d'Ajtai (au sens historique) représentent des instances difficiles de la réduction.

On peut chercher à décrire le comportement probabiliste de l'algorithme LLL vis-à-vis de deux types de paramètres :

(*a*) les paramètres d'exécution, et en particulier le nombre d'itérations K de l'algorithme.

(*b*) les paramètres qui décrivent la configuration de sortie de l'algorithme, et en particulier, la valeur finale $\hat{r}_k$ du k-ème rapport de Siegel, le défaut d'orthogonalité ρ, le défaut d'Hermite γ, et le premier défaut de longueur θ, qui ont été définis dans le précédent chapitre 2.

On cherche en particulier à comparer les moyennes empiriques de ces variables et les bornes supérieures obtenues dans le chapitre précédent, dans les théorèmes 2.1 et 2.2, et les questions suivantes sont du plus grand intérêt :

(*a*) Ces deux grandeurs –bornes dans le pire des cas et moyennes empiriques– ont–elles le même ordre de grandeur asymptotique ?

(*b*) Le comportement de ces moyennes empiriques dépend-il du type de base choisie ?

Les figures 3.3 et 3.4 montrent quelques uns des principaux résultats expérimentaux, qui sont également décrits par Stehlé dans l'article [70]. Nous les commentons maintenant.

3.4.1 Géométrie de la sortie

La géométrie de la sortie de la base locale $\hat{U}_k$ paraît ne dépendre ni du type de bases considérées, ni de l'indice k. sauf peut-être pour des valeurs extrêmes de k. On considère le nombre complexe $\hat{z}_k$ relié à la base de sortie $\hat{U}_k := (\hat{u}_k, \hat{v}_k)$ via l'inégalité $\hat{z}_k := \hat{m}_{k,k+1} + i\hat{r}_k$. Puisque la

Principaux paramètres.	$\hat{r}_k$	γ	θ	ρ	K
Pire des cas (Bornes supérieures)	$1/s$	s^{p-1}	s^{p-1}	$s^{p(p-1)/2}$	$\Theta(Mp^2)$
Bases d'Ajtai aléatoires (Moyennes empiriques)	$1/\alpha$	α^{p-1}	$\alpha^{(p-1)/2}$	$\alpha^{p(p-1)/2}$	$\Theta(Mp^2)$
Bases sac-à-dos aléatoires (Moyennes empiriques)	$1/\alpha$	α^{p-1}	$\alpha^{(p-1)/2}$	$\alpha^{p(p-1)/2}$	$\Theta(Mp)$

FIGURE 3.3 – Comparaison entre les majorants prouvés et les moyennes empiriques des principaux paramètres. Ici, p est la dimension de la base d'entrée et M est la longueur binaire de la base d'entrée : $M := \Theta(\log N)$ où $N := \max ||b_i||^2$.

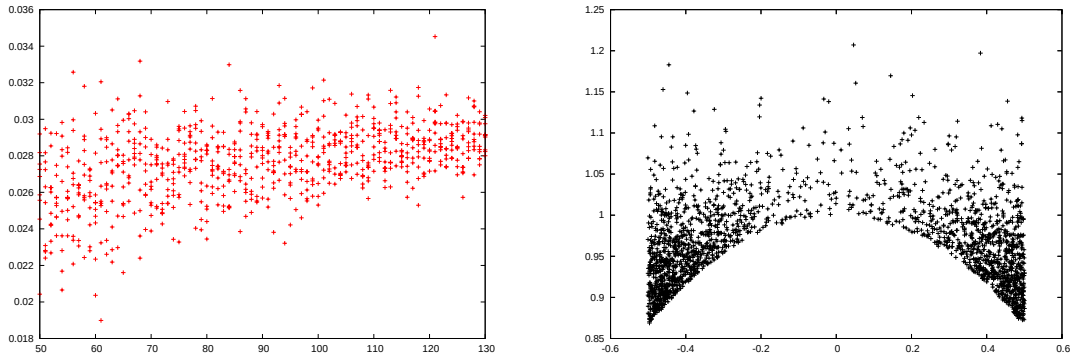

FIGURE 3.4 – À gauche : résultats expérimentaux pour $\log_2 \gamma$. La valeur expérimentale du paramètre $[1/(2p)]\,\mathbb{E}[\log_2 \gamma]$ est proche de 0.03, et donc α est proche de 1.04. À droite, la distribution de sortie des bases locales.

condition de t–Lovász, décrite dans la défintion 2.3, est satisfaite par $\hat{U}_k$, le nombre complexe $\hat{z}_k$ appartient au domaine

$$\mathcal{F}_t := \{z \in \mathbb{C}; \quad |z| \geq 1/t, \ |\Re(z)| \leq 1/2\}.$$

La géométrie d'une base locale de sortie $\hat{U}_k$ "générique" est caractérisée par une distribution qui donne un poids important aux "coins" de $\mathcal{F}_t$ définis par $\mathcal{F}_t \cap \{z; \Im z \leq 1/t\}$ (Voir figure 3.4, droite). Les moyennes empiriques des rapports de Siegel $\hat{r}_k := \Im(\hat{z}_k)$ paraissent être de la même forme que les majorants prouvés. Il apparaît un facteur α (proche de 1.04) qui remplace le facteur $s_0 = 2/\sqrt{3} \approx 1.15$ obtenu dans l'analyse du pire cas lorsque t_0 est proche de 1.

Pour le paramètre $\theta(B)$, la situation est légèrement différente. On remarque que les estimations pour le paramètre θ ne sont pas seulement une conséquence des estimations des rapports de Siegel, mais elles dépendent aussi des estimations qui relient le premier minimum et le déterminant. La plupart des réseaux sont probablement *réguliers* : ceci signifie que la valeur moyenne du rapport entre le premier minimum $\lambda(\mathcal{L})$ et $\det(\mathcal{L})^{1/p}$ est d'ordre polynomial par rapport à la dimension p. Cette propriété de régularité impliquerait alors que la moyenne empirique de paramètre θ est de la même forme que le majorant prouvé, à un facteur $\alpha^{1/2}$ près (proche de

1.02) qui remplace le rapport de Siegel s_0, proche de 1.15. Cela apparaît vérifié dans les résultats expérimentaux présentés.

Questions ouvertes. Est-ce que la constante α admet une définition mathématique, reliée par exemple au système dynamique sous-jacent (cf. chapitre 1, partie II) ? Est-il vrai que la plupart des réseaux considérés sont assez "réguliers", avec une notion adéquate de la régularité, à définir ?

3.4.2 Paramètres d'exécution

En ce qui concernant le nombre d'itérations, la situation diffère, comme on pouvait s'y attendre, selon les types de bases que l'on étudie. Dans le cas des bases d'Ajtai, le nombre d'itérations K montre expérimentalement une moyenne empirique de même ordre que le majorant prouvé, c'est -à-dire d'ordre $O(Mp^2)$ alors que, dans le cas de bases de type sac-à-dos, le nombre d'itérations K a une moyenne empirique dont l'ordre de grandeur est plus petit que celui du majorant prouvé, plus précisément d'ordre $O(Mp)$.

Question ouverte. Est-ce vrai pour toutes les bases sac-à-dos, en particulier pour celles que l'on utilise dans les applications cryptographiques ?

3.4.3 Le travail de cette thèse

A l'issue de ce chapitre, qui termine la partie I de cette thèse, nous pouvons maintenant décrire la suite de notre travail. Cette thèse présente plusieurs méthodes qui pourraient (devraient ?) nous conduire à expliquer les résultats des expérimentations de Nguyen et Stehlé. L'idée directrice est d'utiliser l'algorithme de Gauss comme un outil central pour ce propos. Et le but ultime (non atteint dans la thèse) consiste à obtenir, dans des modèles réalistes, una analyse probabiliste des principaux paramètres

(*a*) les paramètres d'exécution, et en particulier le nombre d'itérations K de l'algorithme.

(*b*) les paramètres qui décrivent la configuration de sortie de l'algorithme, et en particulier, la valeur finale $\hat{r}_k$ du k-ème rapport de Siegel, le défaut d'orthogonalité ρ, le défaut d'Hermite γ, et le premier défaut de longueur θ.

Cette thèse vise à analyser complètement ces paramètres pour préparer le travail en vue de l'analyse de l'algorithme LLL. La partie II est dédiée à l'étude de l'exécution, et la partie III à l'étude de la configuration de sortie.

Deuxième partie

Analyses de l'algorithme de Gauss : Étude probabiliste de l'exécution

Chapitre 1

Modélisations des algorithmes de Gauss : Versions complexes, Point de vue dynamique.

Sommaire

Ce chapitre a pour but de modéliser l'algorithme de Gauss en vue de l'analyse probabiliste que nous poursuivons.

Dans la première section, l'invariance par similitude de l'exécution d'un algorithme de réduction permet d'adopter une vue "projective". En dimension 2, il en résulte une version complexe de l'algorithme de Gauss, définie simplement comme l'itération d'une transformation dans le plan complexe qui généralise la transformation des fractions continues. Dans la deuxième section, nous voyons cet algorithme comme un système dynamique, et nous étudions, dans cette optique, ses principales caractéristiques : ensembles d'entrée, de sortie, dynamique du système. Il ne reste alors qu'à munir ce système dynamique d'un modèle probabiliste d'entrées. C'est l'objet de la troisième section où nous définissons le modèle probabiliste d'entrées, tant dans sa version continue que dans sa version discrète. Il s'agit du modèle de valuation r dont nous avons déjà parlé : elle capture des instances de difficulté variable et permet de décrire la transition vers l'algorithme d'Euclide.

1.1 Versions complexes des algorithmes

1.1.1 Invariance par similitude

Comme on l'a vu, d'un point de vue général, la réduction des réseaux vise à calculer des bases formées par des vecteurs assez courts et assez orthogonaux : on cherche à construre une base avec des vecteurs *relativement courts en comparaison avec les deux minima du réseau* et dont les *angles relatifs* soient proches de l'angle droit. On peut donc s'attendre à ce que tout processus de réduction de réseaux soit invariant par rotation et par homothétie de la base, c'est-à-dire par similitude, dans le sens suivant : la suite de transformations unimodulaires appliquées par l'algorithme à une base d'entrée et à la base transformée par similitude est la même, et les bases qui en résultent diffèrent par la même similitude.

Ces constatations informelles prennent forme quand on étudie l'algorithme LLL. Les transformations que l'algorithme LLL applique à chaque étape ne dépendent des coefficients $m_{i,j}$ de la matrice de Gram-Schmidt $\mathcal{P}$ et des rapports de Siegel ℓ_{i+1}/ℓ_i entre les normes des orthogonalisés. Considérons deux bases d'entrée B et C pour lesquelles chaque vecteur c_i de la base C est transformé du vecteur b_i de la base B par une similitude S ; on a donc $c_i = Sb_i$ pour $i \in [\![1, p]\!]$. Alors les matrices lignes B et C sont reliées par la relation $C = B{}^tS$ et la décomposition de Gram-Schmidt de B, qui s'écrit $B = \mathcal{P}B^\star$, conduit à la relation

$$C = (\mathcal{P}B^\star)^tS = \mathcal{P} \cdot \left(B^\star \cdot {}^tS\right).$$

Puisque $\mathcal{P}$ est triangulaire et $B^\star \cdot {}^tS$ orthogonale, comme produit de deux matrices orthogonales, l'unicité de la décomposition de Gram-Schmidt montre que C s'écrit

$$C = \mathcal{P}C^\star \qquad \text{avec} \quad C^\star = B^\star \cdot {}^tS$$

Donc, C a la même matrice de Gram-Schmidt que B, et la base orthogonale $C^\star$ est la transformée de la base $B^\star$ par la similitude S : les rapports de Siegel des deux bases B et C sont donc les mêmes. Donc la même transformation modulaire est appliquée à la base B et à la base C, qui deviennent donc, au bout d'uneétape, des matrices B' et C' vérifiant

$$B' = {}^tUB, \qquad C' = {}^tUC = {}^tU(B \cdot {}^tS) = ({}^tUB) \cdot {}^tS = B' \cdot {}^tS,$$

qui sont donc de nouveau transformées l'une de l'autre par la similitude S.

En conclusion : Sur deux bases B et C qui se déduisent l'une de l'autre par une similitude S, l'algorithme LLL effectue exactement la même suite de transformations, et les bases de sortie

$\widehat{B}$ et $\widehat{C}$ seront aussi transformées l'une de l'autre par la même similitude S. On dit en abrégé que "l'algorithme LLL est invariant par similitude". On peut donc se restreindre à la classe des bases où le premier vecteur est le premier vecteur de la base canonique. C'est une approche qui se révèle très fructueuse en dimension 2, comme nous allons le voir maintenant.

1.1.2 Versions complexes des algorithmes de Gauss.

L'invariance par similitude s'exploite bien en dimension 2. On peut considérer indifféremment que les vecteurs sont des éléments de $\mathbb{R}^2$ ou de $\mathbb{C}$, mais l'avantage du cadre complexe réside en l'existence d'une multiplication. La mutiplication par un complexe $\lambda := \rho e^{i\theta}$, de la forme $u \mapsto \lambda u$ se traduit géométriquement par une similitude de rapport ρ et d'angle θ. On définit alors une relation d'équivalence sur $\mathbb{C}^2$, en posant qu'une paire (u, v) est équivalente par similitude à (u', v') s'il existe $\lambda \in \mathbb{C}^*$ tel que $(u, v) = (\lambda u', \lambda v')$. Le quotient de $\mathbb{C}^2$ par cette relation d'équivalence est isomorphe à $\mathbb{C}$. Alors, nous identifions une base (u, v) avec l'unique base $(1, v/u)$ qui lui est équivalente. Puisque (u, v) est une base, u et v ne peuvent être colinéaires et le complexe $z = v/u$ ne peut être réel. Une telle base $(1, z)$ est dite normalisée, et l'ensemble des bases normalisées est donc en bijection avec $\mathbb{C} \setminus \mathbb{R}$.

GAUSS-POSITIF(u, v).

Entrée. Une base positive $(u, v) \in \mathbb{R}^2$,
avec $||v|| \leq ||u||, |\tau(v, u)| \leq 1/2$
et $\det(u, v) > 0$.
Sortie. Une base positive minimale de $\mathcal{L}(u, v)$.

```
1  tant que |u| > |v|
2      faire
3          (u, v) ← (v, −u)
4          m ← ⌊τ(v, u)⌉
5          v ← v − mu
```

GAUSS-AIGU(u, v)

Entrée. Une base aigüe $(u, v) \in \mathbb{R}^2$,
avec $||v|| \leq ||u||, 0 \leq |\tau(v, u) \leq 1/2$.
Sortie. Une base aigüe minimale de $\mathcal{L}(u, v)$.

```
1  tant que ||u|| > ||v||
2      faire
3          (u, v) ← (v, u)
4          m ← ⌊τ(v, u)⌉
5          ϵ ← sign (τ(v, u) − ⌊τ(v, u)⌉)
6          v ← ϵ(v − mu)
```

FIGURE 1.1 – Algorithme de Gauss : Algorithmes GAUSS-POSITIF et GAUSS-AIGU

Le complexe v/u va jouer un rôle particulier et, considérant une même base (u, v) indistinctement sur $\mathbb{C}^2$ ou $\mathbb{R}^2 \times \mathbb{R}^2$, la relation

$$\frac{v}{u} = \frac{u \cdot v}{|u|^2} + i\frac{\det(u, v)}{|u|^2}, \tag{1.1}$$

exprime la traduction entre la division complexe (membre de gauche) et les opérations sur $\mathbb{R}^2 \times \mathbb{R}^2$ (membre de droite), notamment le produit scalaire $u \cdot v$ et le déterminant $\det(u, v)$.

Il est possible de reécrire les algorithmes de Gauss sur l'entrée (u, v), uniquement en fonction de $z := v/u$. Remarquons en particulier que le coefficient de Gram-Schmidt $\tau(v, u)$ n'est autre que la partie réelle de z. Les applications suivantes de $\mathbb{C}$ dans $\mathbb{C}$ vont jouer un rôle important. J désigne l'application $z \mapsto -z$, S l'inversion-symétrie $z \mapsto -1/z$ et T la translation $z \to z + 1$. Alors l'échange de deux vecteurs $(u, v) \leftarrow (v, u)$ se traduit par l'inversion-symétrie $-S : z \mapsto 1/z$ et la translation $v \leftarrow v - mu$ se traduit par la translation $T^{-m} : z \mapsto z - m$. Partant d'un complexe z dans le disque $\mathcal{C} := \{z;\ |z| \leq 1\}$, la transformation S l'en sort et la translation T^{-m} le ramène dans la bande verticale

$$\mathcal{B} = \left\{ z \in \mathbb{H} \ :\ |\Re(z)| \leq \frac{1}{2} \right\}.$$

Donc, l'itération du corps de l'algorithme GAUSS-POSITIF sur $(1, z)$ fournit successivement les bases

$$(1, z) \to (z, -1) \to \left(z, 1 - \left\lfloor \Re\left(-\frac{1}{z}\right) \right\rceil z \right),$$

et la nouvelle base normalisée est donc

$$\left(1, \frac{1}{z} - \left\lfloor \Re\left(-\frac{1}{z}\right) \right\rceil \right).$$

De même, l'itération du corps de l'algorithme GAUSS-AIGU sur $(1, z)$ fournit successivement les bases

$$(1, z) \to (z, 1) \to \epsilon\left(\frac{1}{z}\right) \left(z, 1 - \left\lfloor \Re\left(\frac{1}{z}\right) \right\rceil z \right),$$

et la nouvelle base normalisée est donc

$$\left(1, \epsilon\left(\frac{1}{z}\right) \left(\frac{1}{z} - \left\lfloor \Re\left(-\frac{1}{z}\right) \right\rceil \right) \right).$$

Les conditions de sortie $(P_1), (P_2), (P_3)$ et les conditions $(A_1), (A_2)$ du chapitre 2, partie I, ont une jolie interprétation géométrique. Désignons par

$$\mathbb{H} := \{ z \in \mathbb{C} \ : \ \Im(z) > 0 \}$$

le demi-plan supérieur ou encore demi-plan de Poincaré, et par

$$\mathbb{H}_+ := \{ z \in \mathbb{H} \ : \ \Re(z) \geq 0 \} \qquad \mathbb{H}_- := \{ z \in \mathbb{H} \ : \ \Re(z) \leq 0 \},$$

les parties droite et gauche de ce demi-plan. De même, la bande verticale se décompose en

$$\mathcal{B} = \mathcal{B}_+ \cup \mathcal{B}_- \qquad \text{avec} \quad \mathcal{B}_+ := \mathcal{B} \cap \mathbb{H}_+ \qquad \mathcal{B}_- := \mathcal{B} \cap \mathbb{H}_-,$$

et on pose $\widetilde{\mathcal{B}} := \mathcal{B}_+ \cup J\mathcal{B}_-$. De même, le domaine des bases réduites,

$$\mathcal{F} := \left\{ z \in \mathbb{H} \ : \quad |z| \geq 1, \quad |\Re(z)| \leq \frac{1}{2} \right\}. \tag{1.2}$$

se décompose en

$$\mathcal{F} = \mathcal{F} + \cup \mathcal{F}_- \qquad \text{avec} \quad \mathcal{F}_+ := \mathcal{B} \cap \mathbb{H}_+ \qquad \mathcal{F}_- := \mathcal{B} \cap \mathbb{H}_-$$

et on pose $\widetilde{\mathcal{F}} = \mathcal{F}_+ \cup J\mathcal{F}_-$.

Finalement, les domaines d'entrée et de sortie de GAUSS-POSITIF sont respectivement $\mathcal{B} \setminus \mathcal{F}$ et $\mathcal{F}$, tandis que les domaines d'entrée et de sortie de GAUSS-AIGU. sont respectivement $\widetilde{\mathcal{B}} \setminus \widetilde{\mathcal{F}}$ et $\widetilde{\mathcal{F}}$. Le domaine $\mathcal{F}$, représenté dans la figure 1.2, intervient dans la théorie des formes modulaires ou dans la théorie de la réduction des formes quadratiques. À la frontière près, il s'agit d'un domaine fondamental pour l'action du groupe $PSL_2(\mathbb{Z})$ sur $\mathbb{H}$ par homographies. (voir [65]).

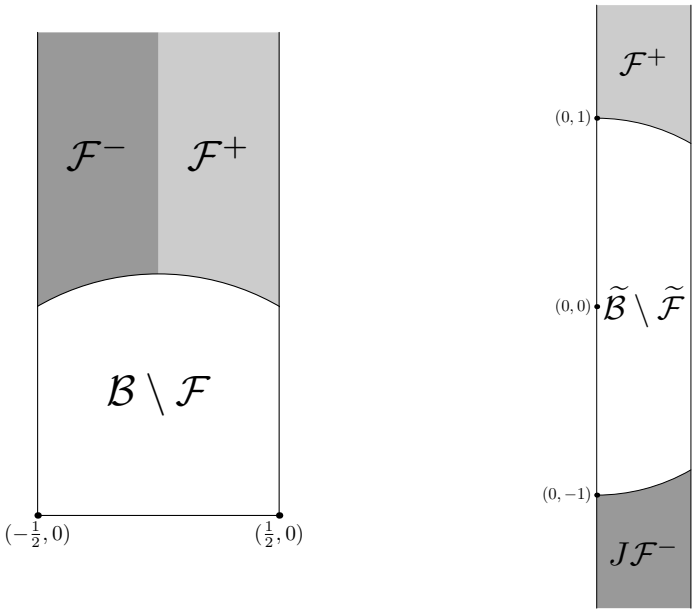

FIGURE 1.2 – Les bandes $\mathcal{B}, \tilde{\mathcal{B}}$, les domaines de sortie $\mathcal{F}, \tilde{\mathcal{F}}$ et les domaines d'entrée $\mathcal{B} \setminus \mathcal{F}, \widetilde{\mathcal{B}} \setminus \widetilde{\mathcal{F}}$.

GAUSS-POSITIF(z)
Entrée. $z \in \mathcal{B} \setminus \mathcal{F}$
Sortie. $\widehat{z} \in \mathcal{F}$
While $|z| < 1$ **do**
$z \leftarrow -\dfrac{1}{z} - \left\lfloor \Re\left(-\dfrac{1}{z}\right) \right\rceil$

GAUSS-AIGU(z)
Entrée. $z \in \widetilde{\mathcal{B}} \setminus \widetilde{\mathcal{F}}$
Sortie. $\hat{z} \in \widetilde{\mathcal{F}}$
While $|z| < 1$ **do**
$z \leftarrow \epsilon\left(\dfrac{1}{z}\right)\left(\dfrac{1}{z} - \left\lfloor \Re\left(\dfrac{1}{z}\right) \right\rceil\right)$

FIGURE 1.3 – Algorithme GAUSS-POSITIF et GAUSS-AIGU complexes.

1.1.3 Versions analogues des algorithmes d'Euclide centrés.

Cette invariance par similtitude peut aussi s'appliquer avec profit aux algorithmes d'Euclide, et en particulier à leurs versions centrées. Si au lieu de travailler avec les paires (u, v) d'entiers, l'ancien couple (u, v) et le nouveau couple (r, u), on travaille avec les rationnels v/u, l'ancien rapport $x = u/v$ et le nouveau rapport $y = r/u$, chaque division euclidienne peut être décrite par une transformation qui associe le nouveau rapport y à l'ancien rapport x, de manière que $y = V(x)$ (dans le cas de EUCLIDE-CENTRÉ-NON-PLIÉ) ou $y = \tilde{V}(x)$ (dans le cas de EICLIDE-CENTRÉ-PLIÉ).

Avec $\mathcal{I} := [-1/2, +1/2]$ et $\tilde{\mathcal{I}} := [0, 1/2]$, les transformations $V : \mathcal{I} \to \mathcal{I}$ ou $\tilde{V} : \tilde{\mathcal{I}} \to \tilde{\mathcal{I}}$ sont définies de la manière suivante

$$V(x) := \frac{1}{x} - \left\lfloor \frac{1}{x} \right\rceil, \quad \text{for } x \neq 0, \qquad V(0) = 0, \tag{1.3}$$

$$\tilde{V}(x) = \epsilon\left(\frac{1}{x}\right)\left(\frac{1}{x} - \left\lfloor \frac{1}{x} \right\rceil\right), \quad \text{for } x \neq 0, \qquad \tilde{V}(0) = 0. \tag{1.4}$$

[Ici, $\epsilon(x) := \text{sign}(x - \lfloor x \rceil)$].

La version décrite ici est définie seulement sur les rationnels. Mais nous avons expliqué dans le chapitre 2 de la partie I comment l'algorithme d'Euclide pouvait se prolonger à deux réels quelconques (u, v). Cela induit donc un prolongement des fonctions V ou $\tilde{V}$ aux intervalles réels tout entiers respectifs $\mathcal{I} := [-1/2, +1/2]$ et $\tilde{\mathcal{I}} := [0, 1/2]$.

1.2 Systèmes dynamiques

D'un point de vue général adopté en mathématiques, un système dynamique est une paire (X, T) formée d'un ensemble X et d'une transformation $T : X \to X$. L'ensemble X est appelé espace des phases, et la transformation T est appelé décalage. On s'intéresse alors à la trajectoire (appelée encore l'orbite) d'un point de $x_0 \in X$, définie par la suite des itérées de x_0 sous l'action de T,

$$\mathcal{T}(x_0) := (x_0, Tx_0, T^2x_0, \dots).$$

Quand X est un espace topologique, on étudie la convergence de la trajectoire $\mathcal{T}(x_0)$ et on cherche à comparer deux trajectoires $\mathcal{T}(x)$ et $\mathcal{T}(y)$ et à évaluer la dépendance de la trajectoire $\mathcal{T}(x)$ par rapport à la condition initiale x. L'espace des phases peut être discret, et dans ce cas, le système dynamique lui-même est qualifié de discret : c'est le cas des automates cellulaires ou des tas de sable, par exemple. L'espace X peut aussi être continu. Quand l'espace X est muni d'une mesure, on étudie les propriétés statistiques des trajectoires.

En analyse d'algorithmes, nous nous intéressons surtout à l'étude des propriétés statistiques des trajectoires qui atteignent un certain ensemble "de sortie", autrement dit, à des trajectoires finies. Dans le cas présent, les coûts que nous avons définis dans la section 3.3.2 de la partie I, modélisés par les fonctions Q et D, sont des exemples de variables aléatoires définies sur les trajectoires du système dont le comportement statistique fournit des informations qui permettent de dérouler une analyse de l'algorithme. C'est le but de cette section de changer le point de vue sur les algorithmes de Gauss et d'Euclide, et de les interpréter comme des systèmes dynamiques. Nous pourrons ainsi profiter des méthodes développées dans la théorie des systèmes dynamiques et les intégrer dans les outils propres à l'analyse de l'algorithme. C'est le principe de la méthode d'analyse dynamique, initiée par Vallée depuis une petite quinzaine d'années. L'outil fondamental que l'analyse d'algorithmes emprunte à la théorie des systèmes dynamiques est *l'opérateur de transfert.*

Dans cette section, nous présentons le formalisme des systèmes dynamiques dans notrre contexte, et nous décrivons en détail les systèmes dynamiques associés aux algorithmes Gauss-positif et Gauss-aigu. A partir du système associé à l'algorithme Gauss-aigu, nous définissons un nouveau système dynamique, le système Gauss-interne, qui modélise le noyau de l'exécution de l'algorithme de Gauss. Dans le chapitre suivant, 2.2, nous étudierons l'opérateur de transfert associé au système Gauss-interne. Ce sera l'outil fondamental pour analyser l'exécution de l'algorithme de Gauss, ce que nous ferons dans le chapitre 3 de cette partie II.

1.2.1 Premières notions sur les systèmes dynamiques.

Définition 1.1. *Une* partition topologique *d'un sous-ensemble X de $\mathbb{R}$ ou $\mathbb{C}$ est une famille dénombrable $\{X_q\}_{q\in\mathcal{Q}}$ d'ouverts disjoints telle que la réunion des adhérences de ses membres est égale à X,*

$$\cup_{q\in\mathcal{Q}}\overline{X_q} = X.$$

La définition suivante de système dynamique convient à nos objectifs.

Définition 1.2. *Un* système dynamique *est une paire (X, T) qui satisfait les conditions suivantes*

(*i*) *X, appelé* espace de phase*, est un sous-ensemble métrique compact*

(*ii*) *X possède une partition topologique $\{X_q\}_{q\in\mathcal{Q}}$,*

(*iii*) *$T : X \to X$, appelée* décalage*, est une fonction dont la restriction T_q à chaque X_q est inversible et de classe C^2.*

Étant donné un point initial $x \in X$, la suite $\mathcal{T}(x) := (x, Tx, T^2x, \dots)$ des itérées de x par T constitue la trajectoire *du point initial x.*
Un système dynamique à trou *est un triplet (X, T, Y), où (X, T) est un système dynamique, Y une partie de X, et où on tronque les trajectoires dès qu'elles arrivent dans Y.*
Un système dynamique est dit complet *lorsque toutes les branches T_q sont des surjections de $\mathcal{X}_q$ sur $\mathcal{X}$ (i.e., $T_q(X_q) = X$.*
Un système dynamique est dit markovien *lorsque, pour tout $q \in \mathcal{Q}$, le sous-ensemble $T_q(X_q)$ s'écrit comme une réunion de certains X_q.*

Définition 1.3 (Vocabulaire autour des branches inverses)**.** *Soit (X, T) un système dynamique complet de partition topologique $\{X_q\}_{q\in\mathcal{Q}}$.*

(*i*) *Les inverses $T_q^{-1} : X \to X_q$ de T sont appelées branches (inverses) primaires, et l'ensemble des branches inverses primaires est désigné par $\mathcal{H}$.*

(*ii*) *La composée de $k \geq 1$ branches inverses primaires est appelée. branche inverse de profondeur k. L'ensemble de toutes les branches inverses est donc*

$$\mathcal{H}^+ := \cup_{k\geq 1}\mathcal{H}^k, \qquad \text{avec} \quad \mathcal{H}^k = \{h_1 \circ \cdots \circ h_k \ : \ h_i \in \mathcal{H} \ \forall i \in [\![1, k]\!]\},$$

1.2.2 Les systèmes dynamiques Euclide-Centré-Non-Plié et Euclide-Centré-Plié.

Ceci nous amène naturellement aux deux systèmes dynamiques (réels) $(\mathcal{I}, V)$ et $(\tilde{\mathcal{I}}, \tilde{V})$ dont les graphes sont représentés dans la figure 1.4. On remarque que le système tildé est obtenu par un pliage du système non-tildé (ou non-plié), d'abord par rapport à l'axe des abscisses, puis par rapport à l'axe des ordonnées), comme il est expliqué en détail dans [11]. Le premier système (resp. algorithme) s'appelle système (resp. algorithme) Euclide-Plié, alors que le deuxième s'appelle, lui, système (resp. algorithme) Euclide-NonPlié.

Le système Euclide-Centré-Plié est complet et le système Euclide-Centré-Non-Plié est markovien. C'est pourquoi nous préferons le premier, car il a une structure plus claire. En particulier,

Proposition 1.1. [Hurwitz] *L'algorithme* Euclide-Centré-Plié *est un système complet dont l'ensemble des branches inverses primaires est*

$$\mathcal{H} := \{h_{\langle m,\epsilon\rangle}; \quad (m, \epsilon) \geq (2, +1)\}.$$

il existe une caractérisation de $\mathcal{H}^+$ due à Hurwitz qui fait intervernir le nombre d'or $\phi = (1 + \sqrt{5})/2$:

$$\mathcal{H}^+ = \{h(z) = \frac{az+b}{cz+d} \ : \ (a, b, c, d) \in \mathbb{Z}^4, b \geq 1, d \geq 2, ac \geq 0, \\ |ad - bc| = 1, |a| \leq \frac{|c|}{2}, b \leq \frac{d}{2}, -\frac{1}{\phi^2} < \frac{c}{d} < \frac{1}{\phi}\}. \quad (1.5)$$

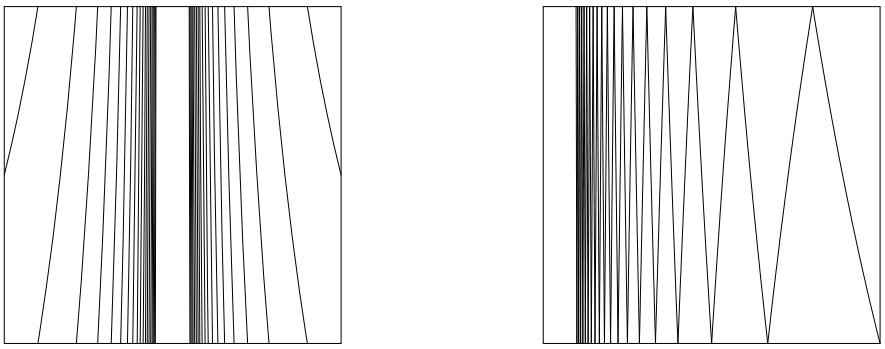

FIGURE 1.4 – Les deux systèmes dynamiques sous-jacents à l'algorithme d'Euclide centré.

Ce résultat est prouvé dans la section 1.2.8.

Nous présentons les systèmes dynamiques associés aux algorithmes GAUSS-POSITIF et GAUSS-AIGU, et nous expliquons pourquoi nous préférons l'un ou l'autre de ces systèmes. Nous présentons l'espace des phases, le décalage, le trou, les branches inverses et les domaines remarquables. Nous construisons, à partir du système dynamique GAUSS-AIGU, un autre système, le système GAUSS-INTERNE, qui capture le noyau de l'exécution de l'algorithme de Gauss, et qui sera fondamental dans les études de la complexité en bits.

Les systèmes GAUSS-POSITIF et GAUSS-AIGU sont des systèmes dynamiques à trous. Le trou de ces systèmes correspond à l'ensemble de sortie formé de bases réduites. L'espace des phases est donc naturellement divisé en deux parties : l'ensemble d'entrée, formé de bases non réduites, et l'ensemble de sortie, formé des bases réduites.

1.2.3 Le système GAUSS-POSITIF.

Nous ne le détaillerons pas vraiment, car nous l'utiliserons seulement pour étudier la configuration de sortie. L'ensemble des complexes de sortie est

$$\mathcal{F} := \left\{ z \in \mathbb{H} \ : \quad |z| \geq 1, \quad |\Re(z)| \leq \frac{1}{2} \right\}. \tag{1.6}$$

Ce domaine, représenté est dans la figure 1.2, intervient dans la théorie des formes modulaires ou dans la théorie de la réduction des formes quadratiques. À la frontière près, il s'agit d'un domaine fondamental pour l'action du groupe $PSL_2(\mathbb{Z})$ sur $\mathbb{H}$ par homographies. (voir [65]). Nous aurons alors besoin de la caractérisation suivante :

Proposition 1.2. *Soit $\mathcal{G}$ l'ensemble des branches inverses de l'algorithme* GAUSS-POSITIF*,*

$$\mathcal{G} = \{h : z \mapsto (az+b)/(cz+d)\} \mid h = h_{[m_k]} \circ h_{[m_{k-1}]} \circ \cdots \circ h_{[m_2]} \circ h_{[m_1]}\}$$

envoyant l'ensemble des sorties $\mathcal{F}$ dans l'ensemble des entrées $\mathcal{B} \setminus \mathcal{F}$ est en bijection avec l'ensemble $\mathcal{Q}$ des quadruplets $(a,b,c,d) \in \mathbb{Z}^4$ avec $ad-bc=1$ et telles que $c \geq 1$ et $|a| \leq |c|/2$. Par ailleurs, il existe une bijection entre cet ensemble $\mathcal{Q}$ et l'ensemble $\mathcal{P} := \{(c,d) \mid c \geq 1,\ pgcd(c,d) = 1\}$. En plus, pour chaque paire (a,c), $c \geq 2$, dans l'ensemble

$$\mathcal{C} := \{(a,c) \mid \frac{a}{c} \in [-1/2, 1/2],\ c \geq 1,\ pgcd(a,c) = 1\},$$

l'homographie ayant par coefficients (a, c) *peut s'écrire* $h = h_{(a,c)} \circ T^q$ *avec* $q \in \mathbb{Z}$ *et* $h_{(a,c)}(z) = (az + b_0)/(cz + d_0)$*, avec* $|b_0| \leq |a/2|$*,* $|d_0| \leq |c/2|$*.*

Démonstration. La bijection entre $\mathcal{Q}$ et $\mathcal{P}$ est une conséquence de la proposition 1.4 de la partie III. Par ailleurs, si $h \in \mathcal{G}$ est une homographie de la forme $h(z) = (az + b)/(cz + d)$ avec un couple $(a, c) \in \mathcal{C}$ avec $c \geq 2$, le couple (b, d) est une solution particulière de l'équation de Bézout $ad - bc = 1$. Puisque $c \geq 2$, la proposition 1.4 assure qu'il existe une unique solution (b_0, d_0) avec $|b_0| \leq |a/2|$, $|d_0| \leq |c/2|$. Cette solution correspond à l'homographie $h_{(a,c)}(z) = (az+b_0)/(cz+d_0)$. Il suffit alors de poser $q = (d - d_0)/c$ pour avoir $h = h_{(a,c)} \circ T^q$. Maintenant, si $(a, c) = (0, 1)$, on peut poser $h_{(a,c)}(z) = -1/z$ et dans ce cas $h = h_{(a,c)} \circ T^d$. □

La notion suivante a un rôle important dans l'étude géométrique de la sortie, dans le chapitre 1 de la partie III.

Définition 1.4. *Le* feston *de* (a, c) *est l'ensemble des images du domaine fondamental par toutes les homographies de même couple* (a, c)*,*

$$\mathcal{F}_{(a,c)} = h_{(a,c)} \circ \left(\bigcup_{q \in \mathbb{Z}} T^q(\mathcal{F}) \right). \tag{1.7}$$

Des illustrations des festons se trouvent dans la figure 1.7 (droite). Dans le cas des couples $(1, 2)$ et $(-1, 2)$, il n'y a que des demi-festons, en accord avec la caractérisation des branches donnée dans la proposition 1.2.

1.2.4 Le système Gauss-aigu

Ainsi, l'espace des phases du système Gauss-aigu est la bande

$$\tilde{\mathcal{B}} = \left\{ z \in \mathbb{C} \ : \quad \Im(z) \neq 0, \quad 0 \leq \Re(z) \leq \frac{1}{2} \right\} = \mathcal{B}_+ \cup J\mathcal{B}_-$$

et le décalage du système est la transformation

$$\tilde{U}(z) = \epsilon\left(\frac{1}{z}\right) \left(\frac{1}{z} - \left\lfloor \Re\left(\frac{1}{z}\right) \right\rceil \right) \qquad \text{avec} \quad \epsilon(z) := \operatorname{sign}(\Re(z) - \lfloor \Re(z) \rceil). \tag{1.8}$$

Le domaine de sortie est l'ensemble

$$\tilde{\mathcal{F}} = \left\{ z \in \mathbb{C}; \quad |z| \geq \leq \Re(z) \leq \frac{1}{2} \right\} = \mathcal{F}_+ \cup J\mathcal{F}_-. \tag{1.9}$$

L'expression du décalage $\tilde{U}$ dépend de la valeur prise par la fonction partie entière et par le signe. Cela conduit à la description de la partition topologique :

Lemme 1.1. *Les ensembles*

$$\mathcal{B}_{m,\epsilon} = \left\{ z \in \mathbb{C} \ : \ \left\lfloor \Re\left(-\frac{1}{z}\right) \right\rfloor = m, \ \ \varepsilon\left(\frac{1}{z}\right) = \epsilon \right\} \cap \mathcal{B}, \qquad \textit{pour } (m, \epsilon) \geq (0, 1)$$

forment une partition topologique pour le système $(\tilde{\mathcal{B}}, \tilde{U})$ *associé au système* Gauss-aigu*.*

Démonstration. On observe que

$$\left\lfloor \Re\left(\frac{1}{z}\right)\right\rfloor = m \quad \text{et} \quad \varepsilon\left(\frac{1}{z}\right) = 1 \iff m \leq \Re\left(\frac{1}{z}\right) < m + \frac{1}{2}$$
$$\left\lfloor \Re\left(\frac{1}{z}\right)\right\rfloor = m \quad \text{et} \quad \varepsilon\left(\frac{1}{z}\right) = -1 \iff m - \frac{1}{2} \leq \Re\left(\frac{1}{z}\right) < m.$$

Cela montre que l'ensemble $\mathcal{B}_{(m,\epsilon)}$ est le transformé d'une bande verticale par l'inversion-symétrie $z \mapsto 1/z$. L'effet d'une telle transformation est illustré sur la figure 1.5, et l'image de l'ensemble des bandes verticales est décrite dans la figure 1.6.

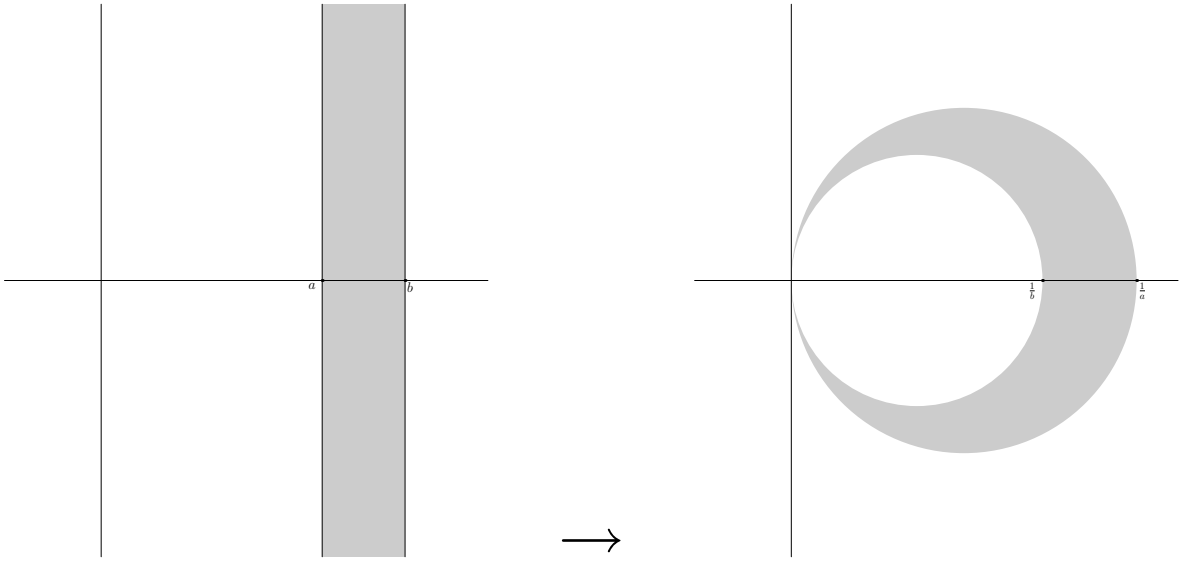

FIGURE 1.5 – Transformation d'une bande verticale par l'inversion $z \mapsto 1/z$.

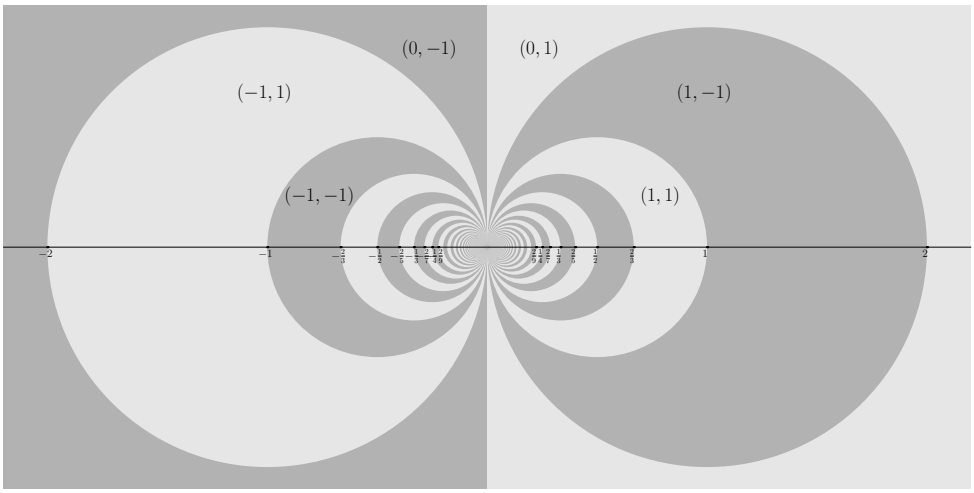

FIGURE 1.6 – Image des bandes verticales $m - \frac{1}{2} \leq \Re(z) < m$ et $m \leq \Re(z) < m + \frac{1}{2}$ (en gris et gris foncé respectivement) par une inversion $z \mapsto 1/z$

Pour déterminer les indices des bandes verticales qui conviennent, on écarte d'abord tous les couples (m, ϵ) pour lesquels les bandes verticales sont dans le demi-plan droit. On ne conserve

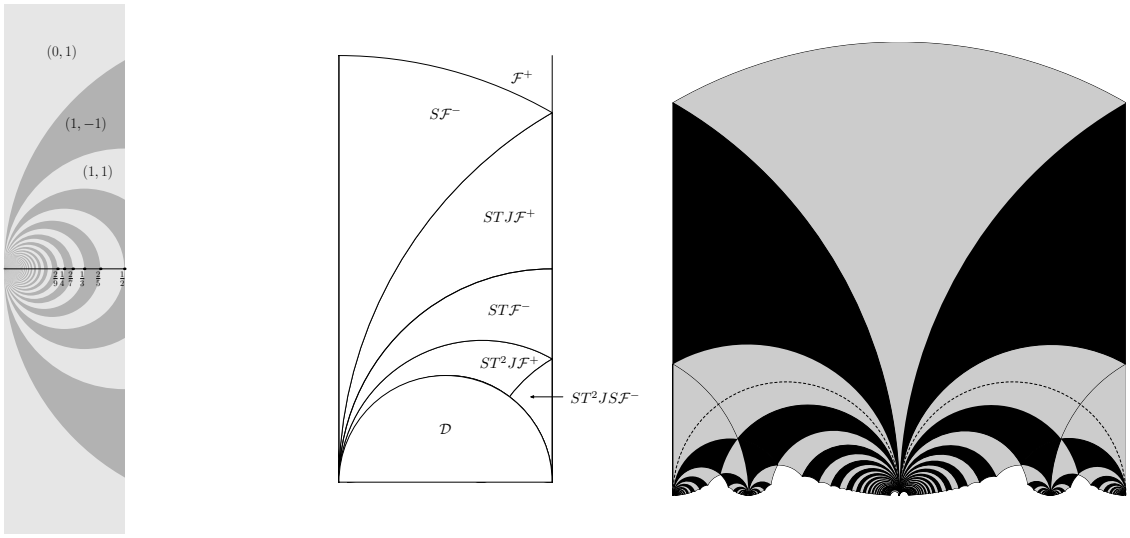

FIGURE 1.7 – À gauche, partition topologique de $\tilde{\mathcal{B}}$. Au milieu, la décomposition de $\tilde{\mathcal{B}} \setminus \mathcal{D}$. À droite, on montre que le disque $\mathcal{D}$ n'est pas compatible avec la géométrie des transformées des domaines fondamentaux $\mathcal{F}$.

donc que les couples $(m, \epsilon) \geq (0, 1)$. Comme les images des bandes associées aux couples $(m, \epsilon) \geq (0, 1)$ contiennent toutes des points de $\mathbb{H}^+$ et $\mathbb{H}^-$ arbitrairement proches de l'origine, il faut conserver tous les couples $(m, \epsilon) \geq (0, 1)$. □

1.2.5 La définition du système GAUSS-INTERNE.

Dans la section précédente, nous avons déterminé la partition topologique du système GAUSS-AIGU. On remarque que les restrictions $\tilde{U}_{(m,\epsilon)}$ de $\tilde{U}$ à $\tilde{\mathcal{B}}_{(m,\epsilon)}$ sont surjectives si et seulement si $(m, \epsilon) \geq (2, 1)$ (dans l'ordre lexicographique des couples), ce qui est confirmé par le calcul. Par ailleurs, la relation

$$\mathcal{D} := \bigcup_{(m,\epsilon)\geq(2,1)} \tilde{\mathcal{B}}_{(m,\epsilon)} = \left\{ z \notin \mathbb{R} \ : \ \Re\left(\frac{1}{z}\right) \geq 2 \right\}. \tag{1.10}$$

montre que la réunion des éléments de la partition topologique où la restriction du décalage est surjective est égale au disque $\mathcal{D}$ du plan complexe dont le diamètre est l'intervalle $[0, 1/2]$, privé de cet intervalle.

Par ailleurs, la géométrie de $\widetilde{\mathcal{B}} \setminus \mathcal{D}$ est compatible avec la géométrie de $\tilde{\mathcal{F}}$, puisque le domaine $\widetilde{\mathcal{B}} \setminus \mathcal{D}$ s'écrit comme la réunion de six transformées du domaine fondamental $\mathcal{F}$,

$$\tilde{\mathcal{B}} \setminus \mathcal{D} = \bigcup_{h\in\mathcal{K}} h(\tilde{\mathcal{F}}) \qquad \text{avec} \quad \mathcal{K} := \{I, S, STJ, ST, ST^2J, ST^2JS\}, \tag{1.11}$$

comme le montre la figure 1.7, au milieu. Puisque $\tilde{\mathcal{B}}$ est une réunion de transformées du domaine fondamental $\tilde{\mathcal{F}}$, cela montre que le disque $\mathcal{D}$ est aussi une réunion de transformées du domaine fondamental $\tilde{\mathcal{F}}$. On note que la situation est différente pour l'algorithme GAUSS-POSITIF, puisque la frontière de $\mathcal{D}$ se trouve "au milieu" des transformées du domaine fondamental $\mathcal{F}$ (voir figure 1.7, droite).

Comme la figure 1.8 le montre, il y a deux parties dans l'exécution de l'algorithme GAUSS-AIGU, selon la position actuelle de la base courante z_i par rapport au disque $\mathcal{D}$ de diamètre $[0, 1/2]$ dont l'équation alternative est

$$\mathcal{D} := \{z; \qquad \Re\left(\frac{1}{z}\right) \geq 2\}.$$

Tant que z_i appartient à $\mathcal{D}$, le quotient (m_i, ϵ_i) satisfait $(m_i, \epsilon_i) \geq (2, +1)$ (par rapport à l'ordre lexicographique) et chaque étape de l'algorithme utilise la transformation associée à l'une des branches de l'ensemble

$$\mathcal{H} := \{h_{\langle m,\epsilon\rangle}; \quad (m, \epsilon) \geq (2, +1)\}$$

de sorte que $\mathcal{D}$ peut s'écrire

$$\mathcal{D} = \bigcup_{h\in\mathbb{H}+} h(\tilde{\mathcal{B}} \setminus \mathcal{D}).$$

GAUSS-INTERNE(z)
Entrée. Un nombre complexe dans $\mathcal{D}$.
Sortie. Un nombre complexe dans $\tilde{\mathcal{B}} \setminus \mathcal{D}$.

`Tant que` $z \in \mathcal{D}$ `faire` $z := \tilde{U}(z)$;

GAUSS-FINAL(z)
Entrée. Un nombre complexe dans $\tilde{\mathcal{B}} \setminus \mathcal{D}$.
Sortie. Un nombre complexe dans $\tilde{\mathcal{F}}$.

`Tant que` $z \notin \tilde{\mathcal{F}}$ `faire` $z := \tilde{U}(z)$;

GAUSS-AIGU(z)
Entrée. Un nombre complexe dans $\tilde{\mathcal{B}} \setminus \tilde{\mathcal{F}}$.
Sortie. Un nombre complexe dans $\tilde{\mathcal{F}}$.

GAUSS-INTERNE (z) ;
GAUSS-FINAL (z) ;

FIGURE 1.8 – La décomposition de l'algorithme GAUSS-AIGU.

1.2.6 Les propriétés du système GAUSS-INTERNE.

A partir du système de l'algorithme GAUSS-AIGU, on définit donc un domaine $\mathcal{D}$ qui est la réunion de tous les éléments de la partition topologique où la restriction du décalage est surjective. Et on définit le trou comme le complémentaire de ce domaine par rapport à l'espace de phase. Le trou du système est ici l'ensemble $\tilde{\mathcal{B}} \setminus \mathcal{D}$, et on garde le même décalage $\tilde{U}$. Le système GAUSS-INTERNE possède la partition topologique

$$\{\tilde{\mathcal{B}} \setminus \mathcal{D}\} \cup \{\tilde{\mathcal{B}}_{(m,\epsilon)}\}_{(m,\epsilon)\geq(2,1)}$$

et son ensemble de branches primaires est

$$\mathcal{H} := \{h_{(m,\epsilon)} : \tilde{\mathcal{B}}_{(m,\epsilon)} \to \tilde{\mathcal{B}} \ : \ (m, \epsilon) \geq (2, 1)\}. \tag{1.12}$$

Le système GAUSS-INTERNE a une structure remarquable puisque ses branches primaires sont toujours composables. Cet ensemble $\mathcal{H}$ est exactement le même que celui du système $(\tilde{\mathcal{I}}, \tilde{V})$

associé à l'algorithme EUCLIDE-CENTRÉ-PLIÉ, Ainsi, l'algorithme GAUSS-INTERNE peut-être vu comme un relèvement de l'algorithme EUCLIDE-CENTRÉ-PLIÉ.

Proposition 1.3. *L'ensemble des homographies utilisées dans l'algorithme* GAUSS-INTERNE *est le même que l'ensemble de celles qui sont utilisées dans l'algorithme*EUCLIDE-CENTRÉ-PLIÉ. *C'est l'ensemble $\mathcal{H}^+$ où*

$$\mathcal{H} = \{h_{(m,\epsilon)} : \tilde{\mathcal{B}}_{(m,\epsilon)} \to \tilde{\mathcal{B}} \ : \ (m,\epsilon) \geq (2,1)\}.$$

Il existe une caractérisation de $\mathcal{H}^+$ due à Hurwitz qui fait intervernir le nombre d'or $\phi = (1+\sqrt{5})/2$:

$$\mathcal{H}^+ = \{h(z) = \frac{az+b}{cz+d} \ : \ (a,b,c,d) \in \mathbb{Z}^4, b \geq 1, d \geq 2, ac \geq 0, \\ |ad-bc| = 1, |a| \leq \frac{|c|}{2}, b \leq \frac{d}{2}, -\frac{1}{\phi^2} < \frac{c}{d} < \frac{1}{\phi}\}. \quad (1.13)$$

Ils font donc remarquer que seules les étapes finales de l'algorithme GAUSS-AIGU utilisent des homographies qui n'existent pas dans l'algorithme EUCLIDE-CENTRÉ-PLIÉ.

1.2.7 Liens entre les algorithmes de Gauss et les algorithmes d'Euclide centré

Bien sûr, il y a des connexions étroites entre U et $-V$ d'un côté, et entre $\tilde{U}$ et $\tilde{V}$ d'un autre : même si les systèmes dynamiques complexes $(\mathcal{B}, U)$ et $(\tilde{\mathcal{B}}, \tilde{U})$ sont définis dans des bandes formées par des nombres complexes z non-réels (c'est-à-dire, $\Im z \neq 0$), rien n'empêche de les étendre aux entrées réelles puisque les décalages s'évaluent sans problème : cela définit deux nouveaux systèmes dynamiques $(\underline{\mathcal{B}}, \underline{U})$ et $\underline{\tilde{\mathcal{B}}}, \underline{\tilde{U}})$, avec $\underline{\mathcal{B}} = \mathcal{B} \cup \mathcal{I}$ et $\underline{\tilde{\mathcal{B}}} = \tilde{\mathcal{B}} \cup \tilde{\mathcal{I}}$ et les systèmes réels $(\mathcal{I}, -V)$ et $(\tilde{\mathcal{I}}, \tilde{V})$ ne sont que la restriction des systèmes complexes étendus aux entrées réelles. Dans ces systèmes complexes étendus, les trous $\mathcal{F}, \tilde{\mathcal{F}}$ ne sont plus nécessairement atteints par les entrées réelles irrationnelles, puisque les orbites restent réelles et irrationnelles. En particulier elles n'atteignent pas 0. Par contre, les trajectoires des nombres rationnels atteignent toujours 0, d'où elles sont envoyées vers $i\infty$, point qui peut être naturellement incorporé à $\mathcal{F}$ et $\tilde{\mathcal{F}}$. Ainsi, la manière qui semble la plus appropriée d'étendre les systèmes gaussiens aux entrées réelles est de les étendre aux réels rationnels tout en incorporant le point $i\infty$ à $\mathcal{F}$ et à $\tilde{\mathcal{F}}$.

1.2.8 Propriétes des DFC des Algorithmes EUCLIDE-PLIÉ et GAUSS-INTERNE. Le résultat d'Hurwitz

Pour établir (1.13), on désigne par HUR l'ensemble défini par le membre de droite de (1.13) et on montre les deux inclusions mutuelles, en commençant par $\subseteq$ et en concluant $\supseteq$.

($\subseteq$) On prouve, par récurrence sur k, l'inclusion $\mathcal{H}^k \subseteq$ HUR. Pour$k = 1$, les homographies

$$h_{m,\epsilon}(z) = \frac{1}{\epsilon z + m}, \qquad \text{pour} \quad (m,\epsilon) \geq (2,1)$$

sont éléments de HUR. En effet, on a $a = 0$, $b = 1$, $c = \epsilon$ et $d = m$, et onvérifie immédiatement que

$$b \geq 1, \qquad d = m \geq 2, \qquad |ad-bc| = |-\epsilon| = 1, \qquad |a| = 0 \leq \frac{1}{2} = \frac{|c|}{2}, \qquad b = 1 \leq \frac{d}{2}, \quad \text{et } c \neq 0.$$

On vérifie aussi que $c/d = \epsilon/m$ appartient à $]-1/\phi^2, 1/\phi[$. Lorsque $m \geq 3$, $\epsilon/m \in [-1/3, 1/3] \subset]-1/\phi^2, 1/\phi[$. Si $m = 2$ alors $\epsilon = 1$ et $0 < 1/2 < 1/\phi$. Le cas de base est donc vérifié.
Supposons que $k \geq 2$ et que pour tout $j < k$, $\mathcal{H}^j \subset \text{HUR}$. Alors, toute homographie h de $\mathcal{H}^k$ s'écrit $g \circ h_{m,\epsilon}$ où $g \in \mathcal{H}^{k-1} \subset \text{HUR}$ et $h_{m,\epsilon} \in \mathcal{H}$. Ainsi,

$$h(z) = g \circ h_{m,\epsilon}(z) = \frac{a\left(\frac{1}{\epsilon z+m}\right) + b}{c\left(\frac{1}{\epsilon z+m}\right) + d} = \frac{a'z + b'}{c'z + d'}$$

avec $a' = \epsilon b$, $b' = a + bm$, $c' = \epsilon d$ et $d' = c + dm$. On vérifie que $|a'd' - b'c'| = |(ad - bc) \cdot (-\epsilon)| = 1$. De même, on observe que

$$a'c' = bd \geq 0, \qquad |a'| = b \leq \frac{d}{2} = \frac{|c'|}{2}, \qquad c' \neq 0, \qquad d' = d(\frac{c}{d} + m) \geq d \geq 2$$

où dans la dernière inégalité on s'est servi du fait que $c/d > -1/\phi^2$. la relation

$$\frac{d'}{c'} = \frac{1}{\epsilon}\left(\frac{c}{d} + m\right),$$

et les deux inégalités

$$(\text{si } \epsilon = 1) \quad \frac{d'}{c'} = m + \frac{c}{d} > 2 - \frac{1}{\phi^2} = \phi, \qquad (\text{ si } \epsilon = -1), \quad \frac{d'}{|c'|} = m + \frac{c}{d} > 3 - \frac{1}{\phi^2} = \phi^2,$$

prouvent que $c'/d' \in]-1/\phi^2, 1/\phi[$.
Montrer que $b' \leq d'/2$ ne pose pas de difficulté car toutes les homographies de $\mathcal{H}$, et donc ses composées, laissent stable l'intervalle $[0, 1/2]$, d'où, en particulier, $h(0) = b'/d' \in [0, 1/2]$. Par ailleurs, $b' \neq 0$ car autrement on aurait $a = -mb$, ce qui contredirait la primalité relative de a et b. On en déduit que

$$b' \leq \frac{d'}{2}, \qquad \text{et} \qquad b' \geq 1.$$

Ainsi, $\mathcal{H}^k \subset \text{HUR}$, et finalement,

$$\mathcal{H}^+ = \bigcup_{k \geq 1} \mathcal{H}^k \subseteq \text{HUR}.$$

($\supseteq$) Soit HUR_n le sous-ensemble de HUR contenant les homographies $h(z) = (az+b)/(cz+d)$ dont le coefficient c vérifie $|c| \leq n$. Nous allons montrer que pour tout $n \geq 1$, $\text{HUR}_n \subseteq \mathbb{H}^+$. On procède par induction comme précédamment. Considérons le cas $n = 1$. Dans ce cas, nous avons $|c| = 1$, et donc $a = 0$ car $|a| \leq |c|/2 = 1/2$. Si $c = 1$, nous avons $d > \phi$ et donc $d \geq 2$, et si $c = -1$, alors $d > \phi^2$ et donc $d \geq 3$. Finalement, puisque $b \leq d/2$, nous avons $b = 1$ dans les deux cas précédents. Ainsi, les homographies de HUR_1 sont de la forme

$$h(z) = \frac{1}{cz + d} \qquad \text{avec} \quad (d, c) \geq (2, 1), c = \pm 1,$$

et ces homographies sont exactement les homographies de $\mathcal{H}$. Ainsi, $\text{HUR}_1 \subset \mathcal{H}^+$.
Soit $n \geq 2$ et supposons que $\text{HUR}_{k-1} \subset \mathcal{H}^+$ pour tout $k \leq n$. Soit $h(z) = (az + b)/(cz + d)$ une homographie de HUR_n. La fraction c/d appartient à $]-1/\phi^2, 1/\phi[\backslash\{0\}$.
Etape 1. Nous montrons que l'on peut l'écrire sous la forme

$$\frac{c}{d} = \frac{\epsilon}{m + \dfrac{c'}{d'}} \qquad \text{avec } \frac{c'}{d'} \in \left]-\frac{1}{\phi^2}, \frac{1}{\phi}\right[\setminus \{0\}.$$

Comme l'intervalle $[-1/\phi^2, 1/\phi[$ est de longueur 1, les intervalles $]g-1/\phi^2, g+1/\phi]$, pour $g \geq 2$ forment une partition de $]2-1/\phi^2, +\infty[$. Il y a deux cas selon le signe de c, et on pose $\epsilon = \text{signe}(c)$. Si $\epsilon = 1$, alors $d/c > \phi$ et, comme $\phi = 2 - (1/\phi^2)$, il existe un unique $m \geq 2$ tel que

$$\frac{c'}{d'} = \frac{d}{c} - m \ \in \left]-\frac{1}{\phi^2}, \frac{1}{\phi}\right[\setminus \{0\}.$$

Si $\epsilon = -1$, alors $d/|c| > \phi^2$ et, comme $\phi^2 = 3 - (1/\phi^2)$, il existe $m \geq 3$ tel que

$$\frac{c'}{d'} = \frac{d}{|c|} - m \ \in \left]-\frac{1}{\phi^2}, \frac{1}{\phi}\right[\setminus \{0\}.$$

Finalement, dans tous les cas, la fraction c/d s'écrit sous la forme cherchée, le couple $(m, \epsilon) \geq (2, 1)$ étant déterminé par les conditions

$$\epsilon = \text{signe}(c) \qquad \text{et} \qquad m = \left\lfloor \left|\frac{d}{c}\right| + \frac{1}{\phi^2} \right\rfloor.$$

Etape 2. Par ailleurs, en utilisant le couple (m, ϵ) obtenu précedemment, on décompose l'homographie $h(z)$ sous la forme

$$h(z) = (h \circ h_{m,\epsilon}^{-1}) \circ h_{m,\epsilon}(z)$$

et il reste à montrer que $h \circ h_{m,\epsilon}^{-1}$ appartient à HUR_{n-1}. En développant $h \circ h_{m,\epsilon}^{-1}(z)$, on obtient

$$h \circ h_{m,\epsilon}^{-1}(z) := \frac{a'z + b'}{c'd + d'} = \frac{a(\epsilon(\frac{1}{z} - m)) + b}{c(\epsilon(\frac{1}{z} - m)) + d}$$

où le quadruplet (a', b', c', d') vérifie

$$a' = b - m\epsilon a, \quad b' = \epsilon a, \quad c' = d - m\epsilon c, \quad d' = \epsilon c.$$

Il faut vérifier que les entiers a', b', c', d' satisfont les conditions d'appartennance à HUR_{n-1}. Tout d'abord, puisque les déterminants de h et $h_{m,\epsilon}^{-1}$ valent ± 1, leur produit, égal au déterminant $a'd' - b'c'$ aussi. Cela montre aussi en particulier les couples de $\{a', d'\} \times \{b', c'\}$ sont formées par des entiers premiers entre eux.

Par ailleurs, par construction $c'/d' \in]-1/\phi^2, 1/\phi]$. Comme $1/\phi^2 \notin \mathbb{Q}$, l'inégalité de gauche est stricte. En plus, le cas $c' = 0$ ne peut pas arriver car dans ce cas on aurait $d = m\epsilon c$, ce qui contredit la primalité relative de c' et d'.
Puisque, par définition, ϵ est le signe de c, alors on a $d' = |c| \geq 1$. Or, si $d' = 1$, alors $-1/\phi^2 < c' < 1/\phi$, donc $c' = 0$, ce qui n'est pas possible. Ainsi, $d' \geq 2$. En plus, nous avons,

$$|c'| < \frac{d'}{\phi} = \frac{|c|}{\phi} < |c| \leq n$$

et donc $|c'| \leq n - 1$ comme on souhaitait.
Maintenant, l'égalité $b' = 0$ implique $|a'd'| = 1$, et alors $d' = 1$, ce qui est impossible, comme on l'a déjà vu. Donc $b' = |a| \geq 1$. En plus,

$$\left|\frac{b'}{d'}\right| = \left|\frac{a}{c}\right| \leq \frac{1}{2}.$$

Il ne reste plus qu'à montrer que a'/c' appartient à $[0, 1/2]$. Supposons d'abord que la fraction b'/d' appartient à $]0, 1/2[$. Dans ce cas, si a'/c' n'appartient pas à $[0, 1/2]$, alors soit $0/1$ soit $1/2$

est compris entre a'/c' et b'/d'. Mais, l'égalité $|a'd' - b'c'| = 1$ montre que les fractions a'/c' et b'/d' sont adjacentes, et donc que les fractions strictement comprises entre a'/c' et b'/d' ont un dénominateur $\geq |c'| + d' \geq 3$. Il y a contradiction.
Supposons maintenant que $b'/d' = 1/2$, ce qui implique $b' = 1, d' = 2$. Dans ce cas, la relation $c'/2 \in]-1/\phi^2, 1/\phi]$ entraîne que c' ne peut être égal qu'à 0 ou 1. On sait que $c' = 0$ est exclu. il ne reste que le cas $c' = 1$. L'égalité $|a'd' - b'c'| = 1$ entraîne alors $|2a' - 1| = 1$, et donc soit $a' = 0$ ou $a' = 1$. Mais si $a' = 1$, alors on a $b = m + 1$, $1 + 2m = d$ et donc $b/d > 1/2$, ce qui est une contradiction. Donc, nécessairement $a'/c' = 0/1$.
On a donc montré que toutes les conditions sur les coefficients (a', b', c', d') de l'homographie $h \circ h_{m,\epsilon}^{-1}$ étaient satisfaites. Cela montre que cette homographie est élément de HUR_{n-1}.

Etape 3. L'hypothèse de récurrence prouve alors que $h \circ h_{m,\epsilon}^{-1}$ est élément de $\mathcal{H}^+$, et donc qu'il en est de même de $h = (h \circ h_{m,\epsilon}^{-1}) \circ h_{m,\epsilon}$. Finalement, pour tout $n \geq 1$, l'inclusion $\text{HUR}_n \subseteq \mathcal{H}^+$ est vraie et montre l'inclusion

$$\text{HUR} = \bigcup_{n \geq 1} \text{HUR}_n \subseteq \mathcal{H}^+,$$

ce qui achève la preuve.

1.2.9 Propriétes des DFC des Algorithmes EUCLIDE-PLIÉ et GAUSS-INTERNE- Propriétés des continuants.

Rappelons que l'algorithme de Gauss produit une suite de bases $(u_0, v_0), (u_1, v_1), \ldots, (u_p, v_p)$ où (u_0, v_0) est la base d'entrée, (u_i, v_i) est la base courante après i réductions, et (u_p, v_p) est la base de sortie, p étant le nombre total de réductions. Les bases satisfont toujours $u_i = v_{i-1}$. Lorsqu'on voit ces bases à similitude près, l'algorithme produit une suite de nombres complexes$z_0, z_1, \ldots, z_p$ vérifiant $z_i := v_i/u_i$

Pour passer d'une base (u_i, v_i) à la base suivante (u_{i+1}, v_{i+1}), l'algorithme applique une transformation unimodulaire, et écrit la base d'entrée en fonction de la i-ème base, obtenue après i étapes sous la forme

$$\begin{pmatrix} v_0 \\ u_0 \end{pmatrix} = \mathcal{M}_i \begin{pmatrix} v_i \\ u_i \end{pmatrix}, \qquad \text{avec} \quad \mathcal{M}_i := \begin{pmatrix} 0 & 1 \\ \epsilon_1 & m_1 \end{pmatrix} \cdots \begin{pmatrix} 0 & 1 \\ \epsilon_i & m_i \end{pmatrix} =: \begin{pmatrix} a_i & b_i \\ c_i & d_i \end{pmatrix}. \quad (1.14)$$

Dans le cadre complexe, on fait correspondre à la la matrice de lignes $(0, 1), (\epsilon_i, m_i)$ l'homographie $h_{m_i,\epsilon_i} \in \mathcal{H}$. Ainsi, l'homographie qui correspond à la matrice $\mathcal{M}_i$ est

$$h_i = h_{m_1,\epsilon_1} \circ \cdots \circ h_{m_i,\epsilon_i}$$

On considère aussi l'application inverse $g_i := h_i^{-1}$, et les expressions

$$h_i(z) = \frac{a_i z + b_i}{c_i z + d_i}, \qquad g_i(z) = h_i^{-1}(z) = \frac{d_i z - b_i}{-c_i z + a_i}. \quad (1.15)$$

Alors, le complexe d'entrée $z = z_0$ s'exprime simplement en fonction du complexe z_i obtenu à la i-ème étape,

$$h_i(z_i) = z, \qquad \text{de sorte que} \qquad z_i = g_i(z).$$

On peut donc dire que, au cours de son exécution, l'algorithme de Gauss calcule le développement en fraction continue du complexe d'entrée z.

1.2.10 Expression complexe des principaux paramètres liés à l'exécution

Dans l'étude de la complexité binaire B définie dans la section 2.2.6, page 38, nous sommes principalement intéressés par les coûts Q et D définis par

$$Q(u,v) = \sum_{i=1}^{P(u,v)} \ell(m_i), \qquad D(u,v) := \sum_{i=1}^{P(u,v)} \ell(m_i) \lg \frac{||u_{i-1}||^2}{||u_0||^2}.$$

Cela nous mène à une étude plus générale, celle des coûts dits additifs, classe à laquelle appartiennent le coût Q et le nombre d'itérations P. Comme nous avons dit dans la section 2.2.6, le coût additif C_c est défini à partir d'un coût élémentaire c par

$$C_{(c)}(u,v) = \sum_{i=1}^{P(u,v)} c(m_i).$$

À ce stade, il est pratique de définir les coûts additifs directement sur les branches inverses. Pour cela, on étend le coût élémentaire c, de sorte que

$$c(h_{m_i,\epsilon_i}) = c(m_i) \qquad \text{et} \qquad c(h \circ g) = c(h) + c(g),$$

pour toutes branches primaires $h_{m_i,\epsilon_i} \in \mathcal{H}$ et branches inverses $h, g \in \mathcal{H}^+$.

Ces coûts sont invariants par similitude, c'est-à-dire,

$$X(\lambda u, \lambda v) = X(u,v) \qquad \text{pour} \quad X \in \{Q, D, P, C_{(c)}\}.$$

Si, avec un léger abus de notation, nous posons $X(z) := X(1,z)$, les coûts qui nous intéressent deviennent des coûts définis sur $\mathbb{C}$, et il est utile de déterminer leurs expressions complexes

Proposition 1.4. *Soit z un complexe d'entrée de l'algorithme* Gauss-interne*, et soit $(z_i)_{i=0}^p$ la suite de bases calculées par l'algorithme, avec $z_0 = z$. On désigne par $h_i(z) = (a_i z + b_i)/(c_i z + d_i)$ l'homographie des i premières itérations de l'algorithme, définie par $z = h_i(z_i)$, et on désigne par h l'homographie totale ($h := h_p$) avec $h(z) = (az+b)/(cz+d)$. Alors :*

(i) Alors, les coûts $C_{(c)}$ et D s'expriment uniquement en fonction de z et des coefficients des homographies h_i, sous la forme suivante

$$C_{(c)}(z) = \sum_{i=1}^{p} c(m_i) \qquad D(z) = \sum_{i=1}^{p} \ell(m_i) \cdot \lg |c_{i-1} z - a_{i-1}|^2, \qquad m_i = \left\lfloor \frac{d_i}{d_{i-1}} + \frac{1}{\phi^2} \right\rfloor.$$

(ii) La suite $\{(m_i, \epsilon_i)\}_{i=1}^p$, avec $\epsilon_i = signe(c_i)$, forme le développement en fraction continue centrée du nombre $h(0) = b/d$. La suite des rationnels a_i/c_i,complétée par le rationnel $h(\infty) = a/c$ forme la suite des convergents (par rapport au développement en fraction continue centrée) du rationnel b/d.

Démonstration. Tout d'abord, nous avons vu dans la preuve d'Hurwitz (caractérisation des branches inverses de Gauss-interne) que le quotient m_i et le signe ϵ_i calculés à la i-ième itération de l'algorithme s'expriment en fonction des coefficients d_i et c_i par

$$\epsilon_i = \text{signe}(c_i), \qquad m_i = \left\lfloor \frac{d_i}{|c_i|} + \frac{1}{\phi^2} \right\rfloor, \qquad |c_i| = d_{i-1},$$

et qui correspondent au développement en fraction continue de $h_p(0)$. Cela établit le résultat pour le coût $C_{(c)}$.

Par ailleurs, la variable D fait intervenir aussi les quotients $||u_{i-1}||/||u_0||^2$. D'après la forme matricielle (1.14), la définition de $z_i = v_i/u_i$, de h_i, de son inverse g_i, et avec (1.15), nous obtenons

$$\left|\frac{u_i}{u_0}\right|^2 = \frac{|u_i|^2}{|c_i v_i + d_i u_i|^2} = \frac{1}{|c_i z_i + d_i|^2} = |h_i'(z_i)| = \frac{1}{|g_i'(z)|} = |c_i z - a_i|^2,$$

d'où l'expression de $D(z)$. □

1.2.11 Géométrie des ensembles $h(\mathcal{B} \setminus \mathcal{D})$ et des ensembles $h(\mathcal{D})$.

L'algorithme Gauss-interne a une belle géométrie et utilise les ensembles $h(\mathcal{D})$ et $h(\widetilde{\mathcal{B}} \setminus \mathcal{D})$.

Ensembles $h(\mathcal{D})$. Si R est le nombre d'itérations de l'algorithme Gauss-interne, le domaine $[R \geq k+1]$ contient les complexes z pour lesquels $\tilde{U}^k(z)$ sont encore dans $\mathcal{D}$. Un tel domaine s'écrit donc

$$[R \geq k+1] = \bigcup_{h \in \mathcal{H}^k} h(\mathcal{D}), \tag{1.16}$$

qui est présenté dans la figure 1.9. Le disque $h(\mathcal{D})$ pour $h \in \mathcal{H}^+$ est le disque dont le diamètre est l'intervalle $[h(0), h(1/2)] = h(\tilde{\mathcal{I}})$. Dans le système Euclide-Centré-Plié, l'intervalle $h(\tilde{\mathcal{I}})$ (relatif à une LFT $h \in \mathcal{H}^k$) est appelé intervalle fondamental de profondeur k : il contient toutes les transformées des nombres réels de l'intervalle $\tilde{\mathcal{I}}$ qui ont le même développement en fraction continue de hauteur k. C'est pourquoi le disque $h(\mathcal{D})$ est ici appelé disque fondamental.

La figure 1.9 montre de manière frappante l'efficacité de l'algorithme, et pose des questions

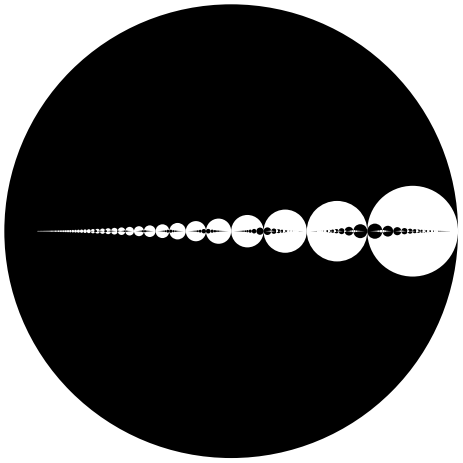

Figure 1.9 – Les domaines $[R = k]$ pour $k \geq 1$, alternativement en noir et blanc.

naturelles : est-il possible d'estimer la probabilité de l'événement $[R \geq k+1]$? Est-il vrai qu'elle décroît géométriquement ? Avec quel rapport ? Nous retournerons à ces questions dans les chapitres suivants.

Ensembles $h(\widetilde{\mathcal{B}} \setminus \mathcal{D})$. Ces ensembles regroupent tous les nombres complexes z de $\mathcal{D}$ pour lesquels l'algorithme Gauss-Interne utilise la même homographie h. Ce sont des triangles curvilignes pour la géométrie du demi-plan de Poincaré, plus précisément décrits comme suit.

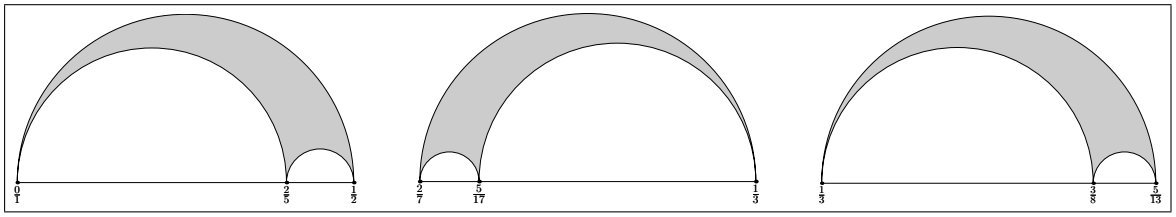

FIGURE 1.10 – Quelques exemples des configurations des domaines $h(\widetilde{\mathcal{B}} \setminus \mathcal{D})$.

Définition 1.5. *On désigne par $\mathcal{D}_{\alpha,\beta}$ le disque du plan complexe dont le diamètre est l'intervalle réel $[\alpha, \beta]$.*

Proposition 1.5. *Soit $h \in \mathcal{H}^+$ telle que $h(z) = (az+b)/(cz+d)$. Alors, l'ensemble $h(\widetilde{\mathcal{B}} \setminus \mathcal{D})$ est, à la frontière près, égale à la différence ensembliste d'un grand disque et de deux petits disques qu'il contient. Plus précisément, $h(\widetilde{\mathcal{B}} \setminus \mathcal{D})$ s'écrit sous la forme*

$$h(\widetilde{\mathcal{B}} \setminus \mathcal{D}) = \mathcal{D}_{\alpha,\beta} \setminus (\mathcal{D}_{\alpha,\delta} \cup \mathcal{D}_{\delta,\beta}) \qquad \text{avec} \quad \{\alpha, \beta, \delta\} = \{h(i\infty), h(0), h(1/2)\}.$$

Les diamètres de ces trois disques sont égaux respectivement à

$$\frac{1}{|c|d}, \qquad \frac{2}{|c|(c+2d)}, \qquad \frac{1}{d(c+2d)}$$

et le diamètre du grand disque est

$$\frac{1}{cd} \quad (si\ c > 0) \qquad \frac{2}{|c|(2d+c)} \quad (si\ c < 0)$$

Preuve. L'homographie $h(z) = (az+b)/(cz+d)$ transforme les droites $\Re z = 0$ et $\Re z = 1/2$ ainsi que le bord du disque $\mathcal{D}$ en circonférences, dont les diamètres sont des intervalles réels déterminés par $h(i\infty), h(0), h(1/2)$. On conclut par une étude de cas selon la position relative de ces trois points. □

Quelques exemples de domaines $h(\widetilde{\mathcal{B}} \setminus \mathcal{D})$ sont donnés dans la figure 1.10.

1.3 Modèles probabilistes d'étude.

L'invariance par similitude nous mène naturellement à considérer des modèles probabilistes où $|u|$ et $z := v/u$ sont des variables aléatoires indépendantes. En pratique, on se donnera la taille de $|u|$ en tant que paramètre fixe du modèle, et il suffira alors de spécifier un modèle probabiliste pour le complexe z, ayant par ensemble fondamental le domaine d'entrée de l'algorithme en étude. Ce domaine d'entrée correspond à $\mathcal{B} \setminus \mathcal{F}$ pour GAUSS-POSITIF, à $\widetilde{\mathcal{B}} \setminus \widetilde{\mathcal{F}}$ pour GAUSS-AIGU, et enfin à $\mathcal{D}$ pour GAUSS-INTERNE.

1.3.1 Modèles continus.

Le modèle probabiliste continu est spécifié par une densité de probabilité. Ce sont des densités $\underline{f}(x, y) = f(z)$ ayant par domaine un ensemble d'entrée $\mathcal{X}$ pouvant être $\mathcal{B} \setminus \mathcal{F}$, $\widetilde{\mathcal{B}} \setminus \widetilde{\mathcal{F}}$ ou $\mathcal{D}$ selon le cas. Bien sûr, ces densités peuvent être vues indistinctement comme des fonctions de $(x, y) \in \mathbb{R}^2$ ou de $z = x + iy \in \mathbb{C}$, mais en précisant tout de même que l'intégration se fera toujours au sens de $\mathbb{R}^2$.

1.3.2 Modèles discrets.

Nous n'allons considérer un modèle discret que pour l'algorithme GAUSS-INTERNE. On se restreint à des entrées (u, v) entières, pour lesquelles $z := v/u$ est élément de $\mathcal{D}$. Le modèle est paramétré par la longueur du vecteur u supposé toujours plus long que v, et nous allons considérer uniquement les vecteurs u de la forme $(N, 0)$ où N est un entier [2]. L'ensemble fondamental du modèle discret des entrées de longueur N est

$$\Omega_N := \left\{ \omega = \frac{v}{u} \in \mathcal{D} \ : \ u = (N, 0),\ v = (a, b),\ a, b, N \ \in \mathbb{Z},\ b \neq 0 \right\}. \tag{1.17}$$

La probabilité sur le modèle discret est définie d'une manière générique à partir du modèle continu : étant données une densité de probabilité f sur $\mathcal{D}$, nous définissons une version discrète f_N sur Ω_N de la manière suivante : pour tout $\omega \in \Omega_N$, on désigne par c_ω, le carré ouvert [3] de centre ω et de côtés $1/N$, parallèles aux axes, et on définit (presque partout dans $\mathcal{D}$) la fonction f_N par

$$f_N(z) = f(\omega) \qquad \text{pour tout } z \in c_\omega.$$

De cette manière, on obtient une famille de fonctions f_N, définies presque partout sur $\mathcal{D}$, et qui s'approchent de f quand $N \to \infty$, pourvu que f soit suffisamment régulière.

1.3.3 Calculs d'espérance dans le discret et le continu.

Pour une variable aléatoire X définie sur $\mathcal{D}$, nous allons définir une version discrète X_N de la même manière que nous l'avons fait pour la densité f, c'est à dire

$$X_N(z) = X(\omega) \qquad \text{pour tout } z \in c_\omega.$$

Par ailleurs, nous adoptons les notations suivantes pour les intégrales dans le modèle continu et discret : A une partie $A \subset \mathcal{X}$, nous associons la partie A_N définie par

$$A_N = \bigcup_{\omega \in A \cap \Omega_N} c_\omega,$$

et nous posons

$$I[X, A] = \iint_A X(z)\, f(z) dxdy, \qquad I_N[X, A] = \iint_{A_N} X_N(z)\, f_N(z) dxdy.$$

Les espérances dans le modèle continu et discret sont données alors par

$$\mathbb{E}(X) = \frac{I[fX, \mathcal{X}]}{I[f, \mathcal{X}]} \qquad \text{et} \qquad \mathbb{E}_N X = \frac{I[f_N X, \mathcal{X}]}{I[f_N, \mathcal{X}]}.$$

2. Cette restriction élimine certes des entrées valides, par exemple les entrées où le vecteur u est de longueur irrationnelle, ce qui est bien possible pour un vecteur à coordonnées entières.

3. Le fait qu'il soit ouvert ou fermé ne jouera à posteriori aucun rôle, on a juste voulu éviter que la définition de f_N puisse être contradictoire.

1.3.4 Modèles liés à une valuation.

Nous cherchons à paramétriser le modèle probabiliste d'entrée, en quantifiant la difficulté qu'ont les bases à se laisser réduire. Notre outil pour quantifier cette difficulté sera la valuation. Une densités f sur $\mathcal{X}$ est dite de valuation r, si elle s'écrit

$$f(z) = |\Im(z)|^r g(z) \qquad r > -1, \quad \text{où } g(z) \text{ est intégrable sur } \mathcal{X} \text{ et } g(z) > 0 \text{ si } \Im(z) = 0$$

Un cas particulier de densité de valuation r est la densité f_r standard de valuation r, où la fonction g definie ci-dessus est constante, désignée par f_r, $f_r(z) = |\Im(z)|^r$

Ce modèle à valuation possède plusieurs propriétés intéressantes, comme on l'a déjà expliqué dans le chapitre 3 de la partie I. On les répète maintenant rapidement :

(*i*) Il s'agit tout d'abord d'un modèle qui apparaît naturellement dans les bases locales de LLL, comme le montre Akhavi [3]. Cela a été précisé par la suite dans [4, 5] , et décrit précisément dans le chapitre 3 de la partie II. Plus précisément, lorsqu'on tire aléatoirement des vecteurs dans la boule unité de $\mathbb{R}^n$, les distributions de probabilité dans les bases locales d'indice $n - j$ présentent des distributions qui sont "presque" de valuation j. Dans ce cas, la valuation j est positive.

(*ii*) Les valuations négatives (en particulier proches de -1) présentent aussi un grand intérêt. Lorsque r se rapproche de -1, la densité donne de plus en plus de poids aux bases dont les vecteurs sont colinéaires. Elles permettent ainsi de simuler une transition vers l'algorithme d'Euclide, que comme nous avons vu correspond à l'algorithme de Gauss dans des entrées collinéaires. Cette transition correspond à l'étude de la complexité limite lorsque le paramètre r tend vers -1.

(*iii*) Le paramètre r paramétrise aussi la difficulté des instances d'entrée de l'algorithme de Gauss, comme on le voit de façon frappante dans la figure 1.9. Lorsque r s'approche de -1, il est naturel de penser que la complexité moyenne de l'algorithme augmente, et c'est une question raisonnable de vouloir quantifier la dépendance de la complexité avec r. Nous allons aborder ce point dans le chapitre 3 de cette partie II.

(*iv*) Enfin, la densité standard de valuation r possède des vertus mathématiques intéressantes, puisque sa dépendance uniquement en $y = \Im(z)$ la rend facile à utiliser, et permet des calculs explicites. Cela conduit à trouver des relations entre la densité de sortie de l'algorithme et les séries d'Eisenstein, sujet qu'on abordera dans le chapitre 2 de la partie III.

(*v*) Enfin, les résultats précis que nous obtenons en dimension 2 pourront sans doute être transféré dans un modèle en dimension quelconque, qui généralise les bases d'Ajtai. Nous avopns décrit ce modèle dans le chapitre 3 de la partie I.

1.3.5 Quelques calculs avec la densité de valuation r.

Proposition 1.6. *Soit f_r la densité standard de valuation r, définie par $f_r(x, y) := |y|^r$ et ν_r la mesure de densité r définie dans $S \subset \mathbb{R}^2$ par*

$$\nu_r[S] = I[f_r, S] = \iint_S |y|^r dydx.$$

Alors,

(i) La mesure d'un disque C_ρ de rayon ρ centré sur l'axe des abscisses vérifie

$$\nu_r[C_\rho] = \frac{1}{2(r+1)} \rho^{r+2} B((r+3)/2, 1/2).$$

(ii) On peut définir une mesure de probabilité sur les ensembles mesurables de $\mathcal{D}$ par

$$\mathbb{P}_{(r)}[E] = \frac{\nu_r[E]}{A_0(r)}, \qquad E \subset \mathcal{D},$$

où $A_0(r)$ est la constante de normalisation définie par

$$A_0(r) = \nu_r[\mathcal{D}] = \nu_r[C_{1/4}] = \frac{1}{2(r+1)} \left(\frac{1}{4}\right)^{r+2} B((r+3)/2, 1/2)$$

qui est un $\Theta(r+1)^{-1}$ pour $r \to -1$.

(iii) La mesure normalisée d'un disque de rayon ρ ou de diamètre δ est donc

$$\mathbb{P}_{(r)}[C_\rho] = (4\rho)^{r+2} = (2\delta)^{r+2}.$$

La proposition suivante présente une estimation de la mesure des domaines $h(\widetilde{\mathcal{B}} \setminus \mathcal{D})$, décrits dans la section 1.2.11, et qui seront essentiels dans notre étude du chapitre 3 de cette partie II.

Proposition 1.7. *Soit $h \in \mathcal{H}^+$ avec $h(z) = (az+b)/(cz+d)$. La mesure de probabilité de l'ensemble $h(\widetilde{\mathcal{B}} \setminus \mathcal{D})$ selon la densité standard de valuation $r > -1$ s'exprime en fonction de d, de $\theta = c/d$ et de la fonction $\mathcal{E}_r$ définie pour $x \in [0,1]$ par*

$$\mathcal{E}_r(x) = 1 - x^{r+2} - (1-x)^{r+2}, \tag{1.18}$$

différemment selon le signe de c et donc de θ,

$$\mathbb{P}_{(r)}[h(\widetilde{\mathcal{B}} \setminus \mathcal{D})] = \left(\frac{2}{\theta d^2}\right)^{r+2} \mathcal{E}_r\left(\frac{\theta}{2+\theta}\right), \qquad \textit{si } \theta > 0 \tag{1.19}$$

$$\mathbb{P}_{(r)}[h(\widetilde{\mathcal{B}} \setminus \mathcal{D})] = \left(\frac{4}{|\theta|(2+\theta)d^2}\right)^{r+2} \mathcal{E}_r\left(\frac{|\theta|}{2}\right), \qquad \textit{si } \theta < 0 \tag{1.20}$$

Dans tous les cas, on a les majorations suivantes

$$\mathbb{P}_{(r)}[h(\widetilde{\mathcal{B}} \setminus \mathcal{D})] \le 2(r+1) \left(\frac{K}{|\theta| d^2}\right)^{r+2} |\theta| \log|\theta| \qquad \textit{pour une constante } K = \frac{4}{\phi}, \tag{1.21}$$

$$I[f_r, h(\widetilde{\mathcal{B}} \setminus \mathcal{D})] \le L \left(\frac{1}{|\theta| d^2}\right)^{r+2} |\theta| \log|\theta| \qquad \textit{pour une constante } L \textit{ bornée pour } r \to -1. \tag{1.22}$$

Preuve. Considérons trois disques centrés sur l'axe des réels, tels que le diamètre D du plus grand est égal à la somme des diamètres des deux autres. Désignons donc par $D, \delta, D-\delta$ (avec $\delta \in [0, D]$) ces diamètres. Alors, la mesure selon $\mathbb{P}_{(r)}$ du disque le plus grand privé des deux autres vaut, d'après le point (iii) de la proposition 1.6,

$$(2D)^{r+2} - (2\delta)^{r+2} - (2D - 2\delta)^{r+2} = (2D)^{r+2}\, \mathcal{E}_r\left(\frac{\delta}{D}\right),$$

où $\mathcal{E}_r(x) = 1 - x^{r+2} - (1-x)^{r+2}$. Les ensembles $h(\widetilde{\mathcal{B}} \setminus \mathcal{D})$ sont des telles différences de disques, avec des diamètres donnés par la proposition 1.5,

$$\frac{1}{cd} > \frac{2}{c(2d+c)} > \frac{1}{d(2d+c)}, \qquad \text{si } c > 0$$

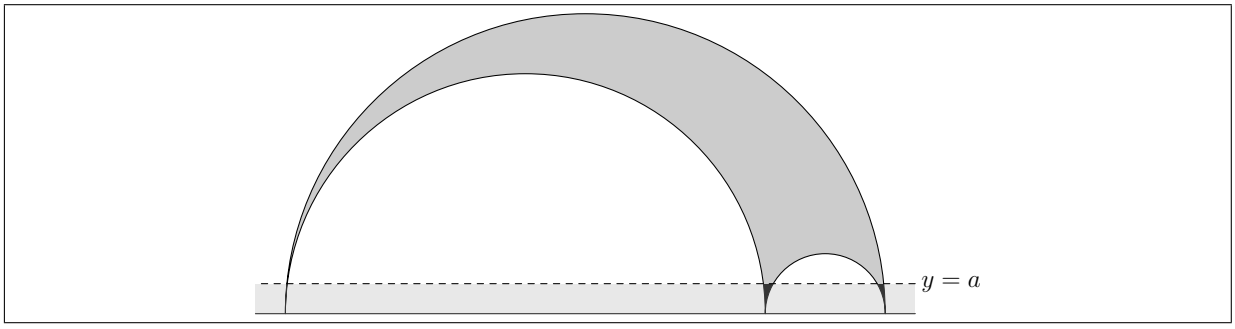

FIGURE 1.11 – Section du domaine $h(\widetilde{\mathcal{B}} \setminus \mathcal{D})$ en dessous de la droite $y = a$ où a est la moitié du rayon du plus petit disque (cf. preuve proposition 1.8).

$$\frac{2}{|c|(2d+c)} > \frac{1}{|c|d} > \frac{1}{d(2d+c)}, \qquad \text{si } c < 0$$

On en déduit immédiatement (1.19) et (1.20) en posant $\theta = c/d$.

Pour majorer $\mathbb{P}_{(r)}[h(\widetilde{\mathcal{B}} \setminus \mathcal{D})]$, on remarque d'abord que, pour $x \in [0,1]$ fixe, la dérivée de $r \mapsto \mathcal{E}_r(x)$ est égale à

$$-x^{r+2}\log x - (1-x)^{r+2}\log(1-x)$$

et est donc toujours inférieure à la fonction entropie $\mathcal{E}(x)$. Donc, puisque $\mathcal{E}_r(x) = 0$ pour $r = -1$, on a toujours $\mathcal{E}_r(x) \leq (r+1)\mathcal{E}(x)$. De plus, on observe aussi les inégaités suivantes dues aux conditions d'Hurwitz :

$$\frac{1}{2+\theta} \leq \frac{1}{\phi} \quad (\theta < 0), \qquad \frac{\theta}{2+\theta} \leq 1 - \frac{2}{\phi^2} < \frac{1}{2} \qquad (\theta > 0), \qquad \frac{|\theta|}{2} \leq \frac{1}{2\phi^2} < \frac{1}{2} \quad (\theta < 0)$$

Comme par ailleurs, la fonction entropie binaire vérifie $\mathcal{E}(x) \leq 2x|\log x|$ pour tout $x \in [0, 1/2]$., on a le résultat cherché. □

Proposition 1.8. *Pour tout $\ell \geq 0$, Il existe une constante $K > 0$ telle que, pour toute branche primaire $h \in \mathcal{H}$, associée au couple (m, ϵ), la mesure du domaine $h(\widetilde{\mathcal{B}} \setminus \mathcal{D})$ par rapport à la densité $g_\ell := \big||y|^{-1}\log^\ell |y|\big|$ vérifie*

$$I\left[\big||y|^{-1}\log|y|\big|, h(\widetilde{\mathcal{B}} \setminus \mathcal{D})\right] \leq K\,\frac{1}{m^2}\,\log^{\ell+1} m$$

Démonstration. On coupe le domaine $h(\widetilde{\mathcal{B}} \setminus \mathcal{D})$ avec la droite déquation $y = a$, où a est égal à la moitié du rayon du plus petit disque, comme le montre la figure 1.11. La proposition précédente montre que l'aire de la partie supérieure (au-dessus de la droite) est en $O(m^{-2}\log^{\ell+1} m)$, tandis que l'aire de la partie inférieure est formée de 3 pointes. On estime l'aire de chaque pointe, ou plutôt d'ailleurs des demi-pointes, déterminées par la verticale, tangente à un disque, le bord de ce disque, et la droite horizontale $y = a$. La demi-pointe peut donc être bordée par le petit disque de rayon $r = 2a$, ou par le disque moyen, de rayon R. Supposons que la pointe est centrée à l'origine et relative à un cercle de rayon ρ. Remarquons alors que l'abscisse x_a du point d'intersection du disque et de la droite horizontale, s'écrit $x_a = \rho - (\rho^2 - a^2)^{1/2}$. La mesure d'une demi-pointe positive (par rapport à la densité g_ℓ) limitée par un disque de rayon $\rho > a$ (par rapport à la densité g_ℓ) se calcule en fonction de x_a, sous la forme suivante

$$\int_0^{x_a} dx \int_{(2\rho x - x^2)^{1/2}}^{a} \left(\frac{-\log^\ell y}{y}\right) dy = \frac{1}{\ell+1}\int_0^{x_a} \left[\frac{1}{2^\ell}\log^{\ell+1}(2\rho x - x^2) - \log^{\ell+1} a\right] dx$$

et est donc d'ordre $x_a\, O(\log^{\ell+1} x_a, \log \rho)$.

Selon que l'on considère une demi-pointe relative au petit disque, ou au disque moyen, on a $x_a = O(a)$ ou $x_a = O(a^2/R)$. De plus, pour une branche primaire relative à (m, ϵ), le disque moyen est de rayon $\Theta(1/m)$ tandis que le petit rayon est d'ordre $\Theta(1/m)^2$. Dans tous les cas, on a donc $x_a = O(1/m)^2$, et l'aire inférieure est donc aussi en $O(m^{-2} \log^{\ell+1} m)$. □

Nous avons décrit dans ce chapitre nos premiers outils (Systèmes dynamiques, mesures utilisées.) Dans le prochain chapitre, nous introduisons notre outil fondamental, l'opérateur de transfert, et expliquons comment il peut être utilisé dans l'analyse des paramètres d'exécution de l'algoritme Gauss-interne.

Chapitre 2

Opérateurs de transfert et séries génératrices.

Sommaire

David Ruelle [63] a introduit l'opérateur de transfert pour étudier les trajectoires périodiques. Le contexte de l'analyse des algorithmes est différent, puisqu'il s'intéresse plutôt aux trajectoires finies. C'est Brigitte Vallée qui a adapté l'outil des opérateurs de transfert au contexte de l'analyse d'algorithmes.

Ce chapitre présente toutes les bases qui seront utiles pour l'analyse de l'exécution de l'algorithme de Gauss, qui sera faite au chapitre suivant. Nous commençons par faire quelques rappels d'analyse fonctionnelle élémentaire (section 2.1). Puis, nous définissons les opérateurs de transfert qui nous seront utiles (section 2.2) et expliquons comment ils interviennent dans nos analyses (section 2.3). C'est le quasi-inverse de l'opérateur de transfert qui est omni-présent et ses propriétés sont reliées aux propriétés spectrales de l'opérateur. C'est pourquoi, nous rappelons dans la section 2.4 quelques éléments de la théorie spectrale des opérateurs, que nous appliquons, dans la section 2.5 aux opérateurs de transfert. La section 2.6 rassemble toutes les propriétés importantes du quasi-inverse qui seront essentielles dans le chapitre 3 suivant.

2.1 Notions de base d'analyse fonctionnelle.

Cette section a pour but de rappeler quelques éléments de la théorie d'opérateurs linéaires bornés sur des espaces de Banach. Nous suivons Kato [34]. Nous supposons connue du lecteur les éléments des espaces vectoriels normés.

2.1.1 Définitions de base.

Soit $(E, ||\cdot||_E)$ un espace de Banach sur un corps $\mathbb{K}$ ($\mathbb{R}$ ou $\mathbb{C}$), c'est-à-dire un espace vectoriel normé complet. Un *opérateur (linéaire)* (sur E) est une application $\mathbf{T} : E \to E$ telle que

$$\mathbf{T}(\alpha_1 x_1 + \alpha_2 x_2) = \alpha_1 \mathbf{T}(x_1) + \alpha_2 \mathbf{T}(x_2)$$

pour tous x_1 et x_2 dans E et tous scalaires $\alpha_1, \alpha_2 \in \mathbb{K}$. L'ensemble des opérateurs linéaires sur E, muni de la multiplication par scalaire et de la somme

$$(\alpha \mathbf{T})x = \alpha \mathbf{T}x \qquad (\mathbf{S} + \mathbf{T})x = \mathbf{S}x + \mathbf{T}x,$$

est un espace vectoriel. Si on y rajoute en plus la composition des 'opérateurs définie par

$$\mathbf{ST} = \mathbf{S} \circ \mathbf{T}, \tag{2.1}$$

où $\circ$ est l'opération de composition, on en fait une *algèbre*. Nous munissons cette algèbre d'une norme subordonnée à la norme de l'espace de base $(E, ||\cdot||)$, définie pour $\mathbf{T} : E \to E$ par

$$||\mathbf{T}|| = \sup_{x \in E \setminus \{0\}} \frac{||\mathbf{T}x||}{||x||}. \tag{2.2}$$

Cette norme subordonnée est une norme au sens usuel, et elle vérifie en outre la propriété

$$||\mathbf{ST}|| \leq ||\mathbf{S}|| \cdot ||\mathbf{T}||,$$

pour tous opérateurs linéaires $\mathbf{S} : E \to E$ et $\mathbf{T} : E \to E$. Lorsque $||\mathbf{T}|| < +\infty$, on dit que $\mathbf{T}$ est un opérateur borné. Grâce aux propriétés de la norme (2.2), les opérateurs linéaires bornés sur E forment un espace vectoriel normé, qui est un espace de Banach, voire une algèbre de Banach lorsqu'on considère le produit 2.1. On note $\mathfrak{B}(E)$ cette algèbre de Banach.

Dans la suite, le mot "opérateur" voudra dire "opérateur linéaire borné" sauf indication contraire. Dans une algèbre de Banach, on peut définir naturellement définir des séries entières d'opérateurs. La série

$$\mathbf{S} = \sum_{k \geq 0} \mathbf{T}^k$$

où $\mathbf{T}^k$ denote la composée de $\mathbf{T}$ k fois avec lui-même, est normalement convergente lorsque $||\mathbf{T}|| < 1$: en effet,

$$||\mathbf{S}|| \leq \sum_{k \geq 0} ||\mathbf{T}||^k = \frac{1}{1 - ||\mathbf{T}||}.$$

Dans ce cas, sa somme $\mathbf{S}$ appartient à $\mathfrak{B}(E)$ et vérifie $\mathbf{ST} = \mathbf{TS} = \mathbf{S} - \mathbf{I}$ où $\mathbf{I}$ est l'opérateur identité. Un opérateur $\mathbf{T} \in \mathfrak{B}(E)$ est *inversible* lorsqu'il possède un inverse $\mathbf{T}^{-1} \in \mathfrak{B}(E)$ borné. Nous avons donc

$$\mathbf{S} = (\mathbf{I} - \mathbf{T})^{-1}$$

et l'opérateur $(\mathbf{I} - \mathbf{T})^{-1} \in \mathfrak{B}(E)$ est appelé *opérateur quasi-inverse* de $\mathbf{T}$.

2.1.2 Opérateur adjoint

L'espace *dual* de l'espace vectoriel normé E est l'ensemble E^* des applications linéaires de E dans $\mathbb{K}$. L'espace dual E^* est un espace de Banach, lorsqu'on le munit de la norme des applications linéaires, c'est-à-dire de la norme

$$||g||_{E^*} = \sup_{u \in E \setminus \{0\}} \frac{|g(u)|}{||u||_E} \qquad g \in E^*.$$

Nous définissons le *crochet de dualité* entre un élément de E et un élément du dual E^* par

$$\langle u, g \rangle = g(u) \qquad u \in E,\ g \in E^*.$$

L'application $(u, g) \mapsto \langle u, g \rangle$ est bilinéaire. Pour un opérateur $\mathbf{T} \in \mathfrak{B}(E)$ nous définissons son opérateur *adjoint* $\mathbf{T}^* :\ E^* \to E^*$ comme l'unique opérateur tel que

$$\langle \mathbf{T}u, g \rangle = \langle u, \mathbf{T}^* g \rangle \qquad \text{pour tout}\ \ u \in E,\ g \in E^*.$$

Cet opérateur appartient à $\mathfrak{B}(E^*)$, et il vérifie, pour la norme subordonnée correspondante (cf. (2.2)),

$$||\mathbf{T}|| = ||\mathbf{T}^*||.$$

Nous allons voir plus tard que les propriétés spectrales d'un opérateur et de son adjoint sont étroitement liées.

2.1.3 Opérateurs dépendant d'un paramètre.

Dans la suite, nous allons considérer des fonctions $\mathbf{T}_t := \mathbf{T}(t)$ d'une variable réelle ou complexe t à valeurs dans $\mathfrak{B}(E)$. Nous allons les appeler également *opérateurs dépendant d'un paramètre*. La dépendance continue voire holomorphe d'un opérateur par rapport à son paramètre se définit de façon naturelle, la convergence dans $\mathfrak{B}(E)$ étant régie par la norme (2.2). Nous allons condenser cela dans la définition suivante.

Définition 2.1. *Soit $\mathbf{T}_t$ une fonction d'une variable complexe t à valeurs dans $\mathfrak{B}(E)$. On dit que $\mathbf{T}_t$ est continue en t_0 si*

$$\lim_{h\to 0} ||\mathbf{T}_{t_0+h} - \mathbf{T}_{t_0}|| = 0,$$

et qu'elle est continue si elle est continue dans tout point de son domaine. Par ailleurs, $\mathbf{T}_t$ est dérivable en t_0 si la limite

$$\lim_{h\to 0} \frac{\mathbf{T}_{t_0+h} - \mathbf{T}_{t_0}}{h}$$

existe dans $\mathfrak{B}(E)$, et elle est holomorphe dans un point t_0 si elle est dérivable dans un voisinage ouvert de t_0. De même, on dit encore que $\mathbf{T}_t$ est dérivable ou holomorphe si elle l'est dans chaque point à l'intérieur de son domaine.

On admettra que toute fonction holomorphe est égale à sa série de Taylor au voisinage de tout point de son domaine, autrement dit, que toutes les fonctions holomorphes sont des fonctions analytiques. Les propriétés algébriques de l'espace $\mathfrak{B}(E)$ invitent naturellement à généraliser les identités connues pour les dérivées des fonctions réelles usuelles.

Proposition 2.1. *Soient $\mathbf{S}_t$ et $\mathbf{T}_t$ des opérateurs dépendant analytiquement du paramètre t. Alors,*

(i) La dérivée du produit (défini en (2.1)) vérifie une identité analogue à celle de la dérivée classique des fonctions réelles ou complexes, à savoir

$$\frac{d}{dt}\mathbf{S}_t\mathbf{T}_t = \frac{d\mathbf{S}_t}{dt}\mathbf{T}_t + \mathbf{S}_t\frac{d\mathbf{T}_t}{dt}.$$

(ii) L'identité précédente se généralise pour tout $k \geq 1$ en

$$\frac{d}{dt}\mathbf{T}_t^k = \sum_{j=0}^{k-1} \mathbf{T}_t^j \circ \frac{d\mathbf{T}_t}{dt} \circ \mathbf{T}_t^{k-j-1}.$$

(iii) Supposons de plus que la norme de l'opérateur $\mathbf{T}_t$ vérifie $||\mathbf{T}_t|| < 1$. Alors, le quasi-inverse $(\mathbf{I} - \mathbf{T}_t)^{-1}$ dépend analytiquement de t, et

$$\frac{d}{dt}(\mathbf{I} - \mathbf{T}_t)^{-1} = (\mathbf{I} - \mathbf{T}_t)^{-1} \circ \frac{d\mathbf{T}_t}{dt} \circ (\mathbf{I} - \mathbf{T}_t)^{-1}.$$

Démonstration. La propriété (i) découle du fait que l'application $(\mathbf{S}, \mathbf{T}) \mapsto \mathbf{S}\circ\mathbf{T}$ de $\mathfrak{B}(E)\times\mathfrak{B}(E)$ dans $\mathfrak{B}(E)$ est bilinéaire, en analogie avec le cas des fonctions réelles avec la multiplication usuelle. La propriété (ii) découle immédiatement de la propriété (i) par récurrence. Dans le cas de (iii), nous avons par définition,

$$\frac{d}{dt}(\mathbf{I} - \mathbf{T}_t)^{-1} = \frac{d}{dt}\sum_{k\geq 0}\mathbf{T}_t^k = \sum_{k\geq 0}\frac{d}{dt}\mathbf{T}_t^k \ ,$$

et il suffit d'appliquer (ii) et d'arranger les indices des sommes pour invoquer encore une fois la définition des opérateurs quasi-inverse et conclure. □

2.2 Opérateurs de transfert.

Dans cette section nous allons introduire les opérateurs de transfert d'un point de vue formel. Nous commençons par l'opérateur de Perron-Frobenius, qui est à l'origine de l'opérateur de transfert. Ensuite, nous introduisons les opérateurs à une branche, et nous en étudions les propriétés de composition. Enfin, nous définissons l'opérateur de transfert proprement dit. Dans toute cette section, nous supposons donné un système dynamique (X, T) *complet* (cf. définition 1.2).

2.2.1 Transformateur de densité.

L'opérateur de transfert généralise l'opérateur transformateur de densité, ou opérateur de Perron-Frobenius, qui est lui-même central dans l'étude des systèmes dynamiques pour l'étude de l'évolution des densités, comme nous l'expliquons maintenant.

L'étude des systèmes dynamiques se concentre sur l'étude des trajectoires d'un point x de l'espace des phases sous l'action du décalage T. Or, l'étude directe de ces trajectoires est souvent compliquée, par la sensibilité aux conditions initiales, ou par les discontinuités du décalage. Un exemple est fourni par le système dynamique associé à l'algorithme d'Euclide, dont le décalage est représenté dans la figure 1.4. Pour contourner ces difficultés on remplace l'étude directe des trajectoires, par l'étude globale de ces trajectoires sous un modèle probabiliste, qui est conditionnée par létude de l'évolution d'une densité sur l'espace des phases. On se donne une densité f_0 et on s'intéresse à la densité f_1 qui résulte de l'"application du décalage sur l'espace de phase, : on définit ainsi une suite (f_k) de densités, qui ont très souvent un comportement plus régulier et compréhensible que les trajectoires. On peut dire qu'on remplace l'étude directe *d'une* trajectoire par l'étude *probabiliste* de *l'ensemble* des trajectoires.

L'opérateur de Perron-Frobenius ou transformateur de densité, désigné par $\mathbf{X}$, est l'opérateur qui à une densité f associe la densité $\mathbf{X}[f]$ qui s'installe sur l'espace des phases X après avoir appliqué le décalage Ainsi, avec la notation précédante, nous étudions la suite des densités $\{f_i\}_{i\geq 0}$ avec

$$f_0 = f, \qquad f_{i+1} = \mathbf{X}[f_i].$$

L'opérateur de Perron-Frobenius a une forme explicite agréable dans le cas d'un système complet, ce qui est notre cas. On considère un ensemble mesurable $B \subset X$. Étant donné une densité f, la mesure de cet ensemble B par rapport à la densité $\mathbf{X}[f]$ est égale à la somme des mesures selon f des préimages de B par le décalage [4]. On dit informellement qu'il y a un flux de densité. Plus précisément, ces mesures font intervenir le jacobien $\mathrm{Jac}(h)$ de la branche inverse h

$$\int_B \mathbf{X}[f](x)dx = \sum_{h\in\mathcal{H}} \int_{h(B)} f(x)dx = \sum_{h\in\mathcal{H}} \int_B \mathrm{Jac}(h)(x)\cdot f\circ h(x)dx,$$

où $\mathcal{H}$ est l'ensemble des branches primaires. En intervertissant somme et intégrale, on obtient

$$\int_B \mathbf{X}[f](x)dx = \int_B \sum_{h\in\mathcal{H}} \mathrm{Jac}(h)(x)\cdot f\circ h(x)dx,$$

et puisque B est arbitraire, l'opérateur transformateur de densité vérifie

$$\mathbf{X}[f] = \sum_{h\in\mathcal{H}} \mathbf{X}_{[h]}[f], \qquad \text{où} \quad \mathbf{X}_{[h]}[f] := \mathrm{Jac}(h)\cdot f\circ h$$

4. La complétude du système intervient essentiellement dans l'expression des préimages, plus simples que dans le cas d'un système non complet.

est l'opérateur associé à la branche h.

2.2.2 Opérateur de transfert.

On généralise les opérateurs précédents en élevant le jacobien à la puissance s (pour un complexe s),

$$\mathbf{X}_{s,[h]}[f] = \mathrm{Jac}(h)^s \cdot f \circ h, \qquad \mathbf{X}_s = \sum_{h\in\mathcal{H}} X_{s,[h]}$$

si bien qu'on retrouve les opérateurs précédents lorsque $s = 1$. L'opérateur $\mathbf{X}_s$ s'appelle l'opérateur de transfert, et pour $s = 1$, on retrouve l'opérateur transformateur de densité $\mathbf{X}$. L'addition du paramètre s, dont l'idée remonte à Ruelle, va s'avérer très puissante dans la suite. En analyse d'algorithmes, cela permettra d'engendrer des séries génératrices.

Ces opérateurs $\mathbf{X}_{s,[h]}$ peuvent aussi être définis pour une branche inverse de profondeur quelconque, et ils présentent une propriété de composition remarquable. Pour deux branches inverses h et g, l'égalité

$$\mathbf{X}_{s,[h]} \circ \mathbf{X}_{s,[g]}[f] = \mathbf{X}_{s,[h]}\left[\mathrm{Jac}(g)^s \cdot f \circ g\right] = \mathrm{Jac}(h)^s[\mathrm{Jac}(g)\circ h]^s \cdot f \circ g \circ h,$$

et la règle de dérivation de Leibnitz qui conduit à l'égalité $\mathrm{Jac}(g)\circ h \cdot \mathrm{Jac}(h) = \mathrm{Jac}(g\circ h)$ démontrent la relation

$$\mathbf{X}_{s,[h]} \circ \mathbf{X}_{s,[g]} = \mathbf{X}_{s,[g\circ h]}.$$

Ceci montre que le k-ème itéré de l'opérateur de transfert est égal à la somme des opérateurs élémentaires sur les branches de hauteur k,

$$\mathbf{X}_s^k = \sum_{h\in\mathcal{H}^k} \mathbf{X}_{s,[h]},$$

et que l'opérateur quasi inverse est la somme des opérateurs élémentaires sur les branches inverses d'hauteur quelconque, c'est-à-dire sur les branches de $\mathbb{H}^*$,

$$(I - \mathbf{X}_s)^{-1} = \sum_{k\geq 0} \mathbf{X}_s^k = \sum_{h\in\mathcal{H}^*} \mathbf{X}_{s,[h]}.$$

Souvent nous allons travailler avec l l'ensemble $\mathcal{H}^+ = \mathcal{H}^* \setminus \{\mathrm{Id}\}$, et dans ce cas l'opérateur en jeu sera

$$(I - \mathbf{X}_s)^{-1} \circ \mathbf{X}_s = \mathbf{X}_s \circ (I - \mathbf{X}_s)^{-1} = \sum_{h\in\mathcal{H}^+} \mathbf{X}_{s,[h]}.$$

2.2.3 Opérateur de transfert avec coût.

L'opérateur $\mathbf{X}_s$ est l'opérateur de transfert utilisé dans la théorie classique des systèmes dynamiques. Le paramètre s y est associé à la taille de départ des orbites. En analyse d'algorithmes, on cherche à relier cette taille avec le coût d'exécution de l'algorithme. On y parvient typiquement par l'usage des séries génératrices bivariées.

Dans le cadre de l'analyse dynamique, ces séries sont engendrées en utilisant les opérateurs de transfert *avec coût* ou *pondérés*,

$$\mathbf{X}_{s,w}^{(c)} = \sum_{h\in\mathcal{H}} \mathbf{X}_{s,w,[h]}^{(c)}, \qquad \text{où} \quad \mathbf{X}_{s,w,[h]}^{(c)} = e^{wc(h)}\mathbf{X}_{s,[h]},$$

associés à des coûts additifs, (cf. section 1.2.10), que comme on a vu sont définis à partir d'une fonction de coût élémentaire positive $c : \mathcal{H} \to \mathbb{N}$ qui quantifie le coût d'une itération. Ces opérateurs avec coût continuent à jouir des mêmes propriétés de composition que les opérateurs classiques, grâce à la propriété d'additivité de c, qui vérifie $c(g \circ h) = c(g) + c(h)$ pour deux branches inverses $g, h \in \mathcal{H}^+$. Nous avons

$$\mathbf{X}^{(c)}_{s,w,[h]} \circ \mathbf{X}^{(c)}_{s,w,[g]} = (e^{wc(h)}\mathbf{X}_{s,[h]}) \circ (e^{wc(g)}\mathbf{X}_{s,[g]}) = e^{w(c(h)+c(g))}\mathbf{X}_{s,[h]} \circ \mathbf{X}_{s,[g]}$$

et

$$e^{w(c(h)+c(g))}\mathbf{X}_{s,[h]} \circ \mathbf{X}_{s,[g]} = e^{wc(g\circ h)}\mathbf{X}_{s,[g\circ h]} = \mathbf{X}^{(c)}_{s,w,[g\circ h]}.$$

En conséquence, nous avons les identités suivantes pour les opérateurs quasi-inverse,

$$(I - \mathbf{X}^{(c)}_{s,w})^{-1} = \sum_{k\geq 0} (\mathbf{X}^{(c)}_{s,w})^k = \sum_{h\in\mathcal{H}^*} \mathbf{X}^{(c)}_{s,w,[h]},$$

et

$$(I - \mathbf{X}^{(c)}_{s,w})^{-1} \circ \mathbf{X}^{(c)}_{s,w} = \mathbf{X}^{(c)}_{s,w} \circ (I - \mathbf{X}^{(c)}_{s,w})^{-1} = \sum_{h\in\mathcal{H}^+} \mathbf{X}^{(c)}_{s,w,[h]}.$$

Ce dernier opérateur, associé à l'ensemble $\mathcal{H}^+$, apparaîtra systématiquement dans les séries génératrices bivariées de nos algorithmes.

2.2.4 Opérateur de transfert de l'algorithme Euclide-plié.

L'ensemble des branches primaires de l'algorithme Euclide-plié est

$$\mathcal{H}_{[0,1]} = \left\{ h_{m,\epsilon} \ : [0,1] \to [0,1],\ h_{m,\epsilon}(x) = \frac{1}{\epsilon x + m} \ :\ m \in \mathbb{Z},\ \epsilon \in \{1,-1\},\ (m,\epsilon) \geq (2,1) \right\},$$

et le jacobien de la branche $h_{m,\epsilon}$ vaut

$$\mathrm{Jac}\left(h_{m,\epsilon}\right)(x) = |h'_{m,\epsilon}(x)| = \frac{1}{(m+\epsilon x)^2}.$$

Ainsi, l'opérateur pondéré associé au coût c est donné par

$$\mathbf{H}_{s,w}[F](x) = \sum_{(m,\epsilon)\geq(2,1)} \left(\frac{1}{m+\epsilon x}\right)^{2s} \cdot e^{wc(m,\epsilon)} \cdot F\left(\frac{1}{m+\epsilon x}\right). \tag{2.3}$$

2.2.5 Opérateur de transfert de l'algorithme Gauss-interne.

Considérons maintenant l'algorithme Gauss-interne. L'ensemble des homographies qui envoient $\tilde{\mathcal{B}} \setminus \mathcal{D}$ dans $\mathcal{D}$ est $\mathcal{H}^+$ avec

$$\mathcal{H} = \left\{ h_{m,\epsilon} \ : \widetilde{\mathcal{B}} \to \widetilde{\mathcal{B}},\ h_{m,\epsilon}(z) = \frac{1}{\epsilon z + m} \ :\ m \in \mathbb{Z},\ \epsilon \in \{1,-1\},\ (m,\epsilon) \geq (2,1) \right\},$$

Le calcul du jacobien d'une fonction $h \in \mathcal{H}$ demande un certain soin : il faut bien observer que ces fonctions sont de variable et à valeurs complexes, mais que *l'intégration est faîte au sens* $\mathbb{R}^2$. Le calcul du jacobien doit donc considérer h comme une fonction vectorielle, définie sur un sous-ensemble de $\mathbb{R}^2$, et à valeurs dans $\mathbb{R}^2$. Le calcul est résumé dans le lemme suivant.

Lemme 2.1 (Jacobien d'une homographie vue comme fonction vectorielle)**.** *Soit h une homographie à coefficients réels et soit*

$$\underline{h}(x,y) = (\Re(h(x+iy)), \Im(h(x+iy))),$$

son interprétation comme fonction $\mathbb{R}^2 \to \mathbb{R}^2$. Alors, le jacobien de $\underline{h}$ est donné par

$$Jac(\underline{h})(x,y) = \left|h'(z)\right|^2,$$

où $z = x + iy$.

Démonstration. En effet, la fonction $\underline{h}$ s'écrit

$$\underline{h}(x,y) = \Phi^{-1} \circ \mathrm{Diag}_h \circ \Phi(x,y),$$

où

$$\Phi(x,y) = (x+iy, x-iy), \qquad \mathrm{Diag}_h(u,v) = (h(u), h(v)),$$

et donc

$$\Phi^{-1}(u,v) = \left(\frac{u+v}{2}, \frac{u-v}{2}\right).$$

Par la règle de Leibnitz,

$$\mathrm{Jac}(\underline{h}) = \left(\mathrm{Jac}(\Phi^{-1}) \circ \mathrm{Diag}_h \circ \Phi\right) \cdot (\mathrm{Jac}(\mathrm{diag}_h) \circ \Phi) \cdot \mathrm{Jac}(\Phi).$$

Or, Φ est une fonction linéaire, de même que son inverse Φ^{-1}. Leurs jacobiens sont donc constants et l'un est l'inverse de l'autre. Par ailleurs,

$$\mathrm{Jac}(\mathrm{Diag}_h)(u,v) = \begin{vmatrix} h'(u) & 0 \\ 0 & h'(v) \end{vmatrix} = |h'(u) \cdot h'(v)|, \tag{2.4}$$

et donc,

$$\mathrm{Jac}(\underline{h})(x,y) = \mathrm{Jac}(\mathrm{Diag}_h)(\Phi(x,y)) = \mathrm{Jac}(\mathrm{Diag}_h)(x+iy, x-iy). \tag{2.5}$$

En posant $z = x + iy$, et avec (2.5) et (2.4) on obtient

$$\mathrm{Jac}(\underline{h})(x,y) = \mathrm{Jac}(\mathrm{diag}_h)(z,\bar{z}) = |h'(z) \cdot h'(\bar{z})| = |h'(z)| \cdot |h'(\bar{z})|$$

mais puisque l'homographie h est à coefficients réels, on a l'égalité $h'(\bar{z}) = \overline{h'(z)}$, qui montre la relation

$$\mathrm{Jac}(\underline{h})(x,y) = |h'(z) \cdot \overline{h'(z)}| = |h'(z)|^2$$

et qui achève la preuve. □

Dans le cas de l'algorithme Gauss-interne, le jacobien s'écrit

$$|h'_{m,\epsilon}(z)|^2 = \left|\frac{1}{m+\epsilon z}\right|^4 = \left(\frac{1}{m+\epsilon z}\right)^2 \cdot \left(\frac{1}{m+\epsilon \bar{z}}\right)^2, \tag{2.6}$$

et l'opérateur de transfert naturel serait donc défini comme suit

$$\check{\mathbf{H}}_{s,w}[f](z) := \sum_{(m,\epsilon)\geq(2,1)} \left|\frac{1}{m+\epsilon z}\right|^{4s} \cdot e^{wc(m,\epsilon)} \cdot f \circ h(z) \tag{2.7}$$

Or, cet opérateur possède un défaut majeur pour nos propos : il ne préserve pas l'analyticité, à cause du module présent dans le jacobien (2.6)Or, nous voudrions que l'opérateur agisse sur les fonctions analytiques. On contourne cette difficulté en considérant le côté droit de (2.6), et en y remplaçant z et $\bar{z}$ par u et v. On définit ainsi un opérateur associé à la branche h agissant sur des fonctions à deux variables $(u, v) \to F(u, v)$,

$$\underline{\mathbf{H}}_{s,w,[h]}[F](u,v) = h'(u)^{s/2}\, h'(v)^{s/2} \cdot e^{wc(h)}\, F(h(u), h(v))$$

et finalement l'opérateur de transfert

$$\underline{\mathbf{H}}_{s,w}[F](u,v) = \sum_{(m,\epsilon)\geq(2,1)} \left(\frac{1}{m+\epsilon u}\right)^s \cdot \left(\frac{1}{m+\epsilon v}\right)^s \cdot e^{wc(m,\epsilon)} \cdot F\left(\frac{1}{m+\epsilon u}, \frac{1}{m+\epsilon v}\right), \quad (2.8)$$

est maintenant bien adapté à nos besoins, comme nous le verrons plus tard.

2.2.6 Premières propriétes de l'opérateur de transfert de Gauss-interne.

Nous présentons ici les premières propriétés de l'opérateur $\underline{\mathbf{H}}_{s,w}$ qui justifient son utilité. Il étend en un sens précis l'opérateur $\mathbf{H}_{s,w}$, et il permet de travailler efficacement avec les densités à valuation, qui est, nous le rappelons, un de nos outils essentiels.

Proposition 2.2. *L'opérateur $\underline{\mathbf{H}}_{s,w}$ satisfait trois principales propriétés.*

(i) Il étend l'opérateur de transfert $\mathbf{H}_{s,w}$ associé à l'algorithme Euclide-plié. *En effet, lorsque le couple (u, v) est un couple diagonal (x, x) avec x réel, on a,*

$$\underline{\mathbf{H}}_{s,w}[F](x,x) = \mathbf{H}_{s,w}[f](x)$$

pour toute fonction F dans le domaine de $\underline{\mathbf{H}}_{s,w}$ et pour son application diagonale f définie par $f(x) := F(x, x)$.

(ii) Il généralise l'opérateur de transfert de l'algorithme Gauss-interne, *défini en (2.7), dans le sens que pour un couple (u, v) "conjugué" de la forme $(z, \bar{z})$, on a l'égalité*

$$\underline{\mathbf{H}}_{s,w}[F](z,\bar{z}) = \check{\mathbf{H}}_{s,w}[f](z),$$

pour toute fonction F dans le domaine de $\underline{\mathbf{H}}_{s,w}$ et f définie à partir de F par la relation $f(z) := F(z, \bar{z})$.

(iii) On a l'identité suivante

$$\underline{\mathbf{H}}_{s,w}[F](z,\bar{z}) = |z-\bar{z}|^r\, \underline{\mathbf{H}}_{s+r,w}[G](z,\bar{z}),$$

pour toute fonction F de la forme

$$F(u,v) = |u-v|^r G(u,v). \quad (2.9)$$

Preuve. Les propriétés (i) et (ii) se vérifient immédiatement. Pour la propriété 3, on remarque d'abord, l'identité due au fait que h est à coefficients réels,

$$|h(z) - h(\bar{z})| = |z-\bar{z}| \cdot \left|h'(z)h'(\bar{z})\right|^{1/2} = |z-\bar{z}| \cdot \left(h'(z)h'(\bar{z})\right)^{1/2}.$$

La forme particulière de F (2.9) entraîne alors la relation

$$F(h(z),h(\bar{z})) = |h(z)-h(\bar{z})|^r\, G(h(z),h(\bar{z})) = |z-\bar{z}|^r\, h'(z)^{r/2}\, h'(\bar{z})^{r/2}\, G(h(z),h(\bar{z})),$$

et donc

$$\underline{\mathbf{H}}_{s,w,[h]}[F](z,\bar{z}) = h'(z)^{s/2}h'(\bar{z})^{s/2}\cdot e^{wc(h)}F(h(z),h(\bar{z})) = |z-\bar{z}|^r\underline{\mathbf{H}}_{s+r,w}[G](z,\bar{z}).$$

□

Dans les prochaines sections nous étudions l'opérateur $\underline{\mathbf{H}}_{s,w}$, très souvent en référence à . l'opérateur $\mathbf{H}_{s,w}$.

2.2.7 Fonctionnelles W et Δ.

Dans les analyses qui vont suivre, les opérateurs quasi-inverse ainsi que deux fonctionnelles de "pondération" vont jouer des rôles importants.

Le premier est la fonctionnelle W, avec son cas particulier W_0, définies par

$$W\underline{\mathbf{H}}_{s,w} = \frac{\partial}{\partial w}\underline{\mathbf{H}}_{s,w}, \qquad W_0\underline{\mathbf{H}}_{s,w} = \frac{\partial}{\partial w}\underline{\mathbf{H}}_{s,w}\Big|_{w=0}.$$

Le but de W est de pondérer les termes de l'opérateur par le coût associé à la branche correspondante. Cet opérateur sera utile pour calculer les moyennes et les autres moments des coûts additifs.

La deuxième fonctionnelle a aussi par but de pondérer les termes de la suite, mais cette fois par la taille de l'entrée associée à la branche . La fonctionnelle Δ est définie par

$$\Delta\underline{\mathbf{H}}_{s,w} = \frac{\partial}{\partial s}\underline{\mathbf{H}}_{s,w}.$$

Ainsi, les paramètres s et w de l'opérateur avec coût $\mathbf{X}_{s,w}$ joueront le rôle des "marqueurs formels" de la taille et le coût de l'entrée, en directe analogie avec les séries génératrices de l'analyse d'algorithmes classique.

2.3 Séries génératrices et opérateurs de transfert.

Dans cette section nous mettons en rapport séries génératrices et opérateurs de transfert. C'est le premier pas vers l'analyse de l'algorithme.

2.3.1 Omni-présence du quasi-inverse.

L'opérateur quasi-inverse apparaîtra systématiquement dans les expressions des séries génératrices. Cela est naturel puisque le quasi-inverse engendre toutes les exécutions possibles de l'algorithme, ou encore, en langage dynamique, toutes les trajectoires possibles du système dynamique. C'est donc légitime d'espérer qu'il puisse servir à exprimer les séries génératrices, et qu'il puisse ainsi jouer le rôle d'opérateur générateur. Il sera étudié d'un point de vue analytique dans la section 2.6, et les théorèmes techniques qu'on y prouve seront la base de l'étude de complexité du chapitre 3.

2.3.2 Densité de sortie.

La densité dite de sortie est celle qui s'installe sur l'espace de sortie de l'algorithme, quand on s'est donné une densité dite d'entrée sur l'espace des entrées. La densité de sortie s'exprime en fonction du quasi-inverse de l'opérateur de transfert, comme le montre la proposition suivante.

Proposition 2.3. *Soit $F \in B_\infty(\mathcal{V})$ et soit f son application diagonale complexe, $f(z) = F(z, \bar{z})$. Si f est la densité d'entrée de l'algorithme de Gauss complexe, alors la densité de sortie est donnée par l'application diagonale complexe de*

$$\widehat{F} = \mathbf{G}_{2,\mathrm{Id}}[F] \qquad \textit{avec} \quad \mathbf{G}_{2,\mathrm{Id}} := \underline{\mathbf{H}}_2 \circ (\underline{I} - \underline{\mathbf{H}}_2)^{-1} \tag{2.10}$$

Si la densité d'entrée f est de valuation r, liée à une application F de la forme $F(u, v) = |u - v|^r L(u, v)$ avec $L \in B_\infty(\mathcal{V})$, alors la densité de sortie est donnée par l'application diagonale complexe de

$$\widehat{F} = |y|^r \mathbf{G}_{2+r,\mathrm{Id}}[L] \tag{2.11}$$

Preuve. En effet, si $\hat{f}$ est la densité de sortie, la mesure d'un ensemble mesurable $A \subset \widetilde{\mathcal{B}} \setminus \mathcal{D}$ est

$$\iint_A \hat{f}(\hat{z}) d\hat{x} d\hat{y} = \sum_{h \in \mathcal{H}^+} \iint_{h(A)} f(z) dx dy = \sum_{h \in \mathcal{H}^+} \iint_A |h'(\hat{z})|^2 f \circ h(\hat{z}) d\hat{x} d\hat{y}$$

où nous nous sommes servis du lemme 2.1 pour le calcul du jacobien. Par ailleurs, par hypothèse nous avons

$$f \circ h(\hat{z}) = F(h(\hat{z}), h(\bar{\hat{z}})),$$

et donc,

$$\iint_A \hat{f}(\hat{x}, \hat{y}) d\hat{x} d\hat{y} = \iint_A \sum_{h \in \mathcal{H}^+} |h'(\hat{z})|^2 F(h(\hat{z}), h(\bar{\hat{z}})) d\hat{x} d\hat{y} = \iint_A \underline{\mathbf{H}}_2 \circ (\underline{I} - \underline{\mathbf{H}}_2)^{-1}[F](\hat{z}, \bar{\hat{z}}) d\hat{x} d\hat{y}$$

ce qui montre que

$$\hat{f}(\hat{x}, \hat{y}) = \underline{\mathbf{H}}_2 \circ (\underline{I} - \underline{\mathbf{H}}_2)^{-1}[F](\hat{z}, \bar{\hat{z}}),$$

et donc (2.10). □

Nous n'avons pas donné le résultat analogue pour les algorithmes Gauss-positif et Gauss-aigu, puisque pour ces algorithmes nous n'avons pas défini d'opérateur. Nous donnons néanmoins une preuve sans opérateurs dans le théorème 2.1 du chapitre 1, partie III, où on étend les résultats sur la densité de sortie, en établissant un lien avec les séries d'Eisenstein.

2.3.3 Série génératrice des moments d'un coût additif C

Proposition 2.4 (Série génératrice des coûts additifs)**.** *La série génératrice des moments d'un coût additif C de coût élémentaire modéré c se décrit en fonction de l'opérateur $\underline{\mathbf{H}}_{s,w}$ de la manière suivante. Désignant par $\mathbf{G}_{s,w}$ l'opérateur défini par*

$$\mathbf{G}_{s,w} = \underline{\mathbf{H}}_{s,w} \circ (I - \underline{\mathbf{H}}_{s,w})^{-1}, \tag{2.12}$$

on a

$$\mathbb{E}_{\langle f \rangle}(e^{wC}) = \iint_{\widetilde{\mathcal{B}} \setminus \mathcal{D}} \mathbf{G}_{2,w}[F](z, \bar{z})\, dxdy \tag{2.13}$$

où F est donnée par $F(x + iy, x - iy) = f(x, y)$. En particulier, dans le cas d'une densité F de valuation r, s'écrivant $F(u, v) = |u - v|^r L(u, v)$ avec $L(u, u) \neq 0$, on a

$$\mathbb{E}_{\langle f \rangle}(e^{wC}) = \iint_{h(\widetilde{\mathcal{B}} \setminus \mathcal{D})} |y|^r \mathbf{G}_{2+r,w}[L](z, \bar{z})\, dxdy.$$

Démonstration. La quantité $C(z)$ ne dépend que de la branche inverse $h \in \mathcal{H}^+$ associée à z. L'expression $C(h)$ ne contient donc pas d'ambigüité. La série génératrice des moments du coût C peut s'écrire en termes des branches inverses de l'algorithme de Gauss, de la manière suivante :

$$\mathbb{E}_{\langle f\rangle}(e^{wC}) = \sum_{h\in\mathcal{H}^+} e^{wC(h)} \cdot \mathbb{P}_{\langle f\rangle}\left[h(\widetilde{\mathcal{B}} \setminus \mathcal{D})\right], \tag{2.14}$$

où $\mathbb{P}_{\langle f\rangle}[h(\widetilde{\mathcal{B}} \setminus \mathcal{D})]$ est la probabilité d'avoir une entrée z dont la branche associée est h. Avec le changement de variable $(x, y) = \underline{h}(\hat{x}, \hat{y})$, cette probabilité s'exprime comme

$$\mathbb{P}_{\langle f\rangle}\left[h(\widetilde{\mathcal{B}} \setminus \mathcal{D})\right] = \iint_{h(\widetilde{\mathcal{B}}\setminus\mathcal{D})} f(x,y)dxdy = \iint_{\widetilde{\mathcal{B}}\setminus\mathcal{D}} |h'(\hat{z})|^2 f \circ \underline{h}(\hat{x},\hat{y})d\hat{x}d\hat{y}.$$

En utilisant l'opérateur $\underline{\mathbf{H}}_{2,w,[h]}$, on remarque l'égalité

$$e^{wC(h)} \cdot \mathbb{P}_{\langle f\rangle}\left[h(\widetilde{\mathcal{B}} \setminus \mathcal{D})\right] = \iint_{\widetilde{\mathcal{B}}\setminus\mathcal{D}} \underline{\mathbf{H}}_{2,w,[h]}[F](\hat{z},\bar{\hat{z}})\, d\hat{x}d\hat{y},$$

et en utilisant la relation ()

$$\mathbb{E}_{\langle f\rangle}(e^{wC}) = \iint_{\widetilde{\mathcal{B}}\setminus\mathcal{D}} \underline{\mathbf{H}}_{2,w} \circ (\mathbf{I} - \underline{\mathbf{H}}_{2,w})^{-1}[F](\hat{z},\bar{\hat{z}}).$$

La preuve est ainsi achèvée. □

2.3.4 Espérance d'un coût additif.

Si l'on veut trouver les différents moments de la c-variable aléatoire C, on dérive par rapport à w et on égale w à 0. Il suffit donc d'utiliser la fonctionnelle W_0, qui effectue exactement cette tâche, :

$$W_0\mathbf{X}_{s,w,[h]} := \frac{\partial}{\partial w}\, \mathbf{X}_{s,w,(c),[h]}\big|_{w=0} = c(h)\mathbf{X}_{s,[h]} =: W_{(c)}\mathbf{X}_{s,[h]}.$$

Cette fonctionnelle linéaire permet de définir l'opérateur générateur du coût additif

$$\mathbf{G}_{s,C} := W_{(c)}[\underline{\mathbf{H}}_s \circ (I - \underline{\mathbf{H}}_s)^{-1}] = (\underline{I} - \underline{\mathbf{H}}_s)^{-1} \circ W_{(c)}[\underline{\mathbf{H}}_s] \circ (\underline{I} - \underline{\mathbf{H}}_s)^{-1},$$

qui permet d'écrire l'espérance du coût additif C sous la forme alternative

$$\mathbb{E}_{\langle f\rangle}(C) = \iint_{\widetilde{\mathcal{B}}\setminus\mathcal{D}} \mathbf{G}_{2,C}[F](z,\bar{z})dxdy \tag{2.15}$$

et aussi, dans le cas où F s'écrit sous la forme $F(u, v) = |u - v|^r L(u, v)$,

$$\mathbb{E}_{\langle f\rangle}(C) = \iint_{\widetilde{\mathcal{B}}\setminus\mathcal{D}} |y|^r \mathbf{G}_{2+r,C}[L](z,\bar{z})dxdy. \tag{2.16}$$

2.3.5 Espérance du coût D

Dans le cas du coût D, nous considérons l'espérance directement.

Proposition 2.5 (Espérance du coût D)**.** *L'espérance du coût D se décrit en fonction de l'opérateur $\mathbf{H}_{s,w}$ de la manière suivante. Si*

$$\mathbf{G}_{s,D} = (\underline{I} - \underline{\mathbf{H}}_s)^{-1} \circ W_{(\ell)}[\underline{\mathbf{H}}_s] \circ (\underline{I} - \underline{\mathbf{H}}_s)^{-1} \circ \Delta[\underline{\mathbf{H}}_s] \circ (\underline{I} - \underline{\mathbf{H}}_s)^{-1},$$

où ℓ est la longueur binaire, alors

$$\mathbb{E}_{\langle f\rangle}(D) = \iint_{\widetilde{\mathcal{B}}\setminus\mathcal{D}} \mathbf{G}_{2,D}[F](z,\bar{z})dxdy = \iint_{\widetilde{\mathcal{B}}\setminus\mathcal{D}} y^r \mathbf{G}_{2+r,D}[L](z,\bar{z})dxdy. \tag{2.17}$$

où F est donnée par $F(x+iy, x-iy) = f(x,y)$.

Démonstration. En effet,

$$\mathbb{E}_{\langle f\rangle}(D) = \iint_{\mathcal{D}} D(z)F(z,\bar{z})dxdy = \sum_{h\in\mathcal{H}^+} \iint_{h(\widetilde{\mathcal{B}}\setminus\mathcal{D})} D(z)F(z,\bar{z})dxdy, \tag{2.18}$$

et, en utilisant la proposition 2.1 sur le jacobien de h,

$$\iint_{h(\widetilde{\mathcal{B}}\setminus\mathcal{D})} D(z)F(z,\bar{z})dxdy = \iint_{\widetilde{\mathcal{B}}\setminus\mathcal{D}} |h'(\hat{z})|^2 D(h(\hat{z}))F(h(\hat{z}),h(\bar{\hat{z}}))d\hat{x}d\hat{y}.$$

D'après XX, lorsque la hauteur de h est égale à p, le coût $D(h(\hat{z})$ s'exprime comme une somme de p termes, le i-ème terme faisant intervenir le coût $\ell(m_i)$ et l'homographie h_{i-1} regroupant les $i-1$ premières étapes de l'algorithme, ainsi que le complexe z_{i-1} associé à la $(i-1)$-ème base construite dans l'algorithme. On raffine la décomposition de l'homographie h en écrivant

$$z_{i-1} = h_{m_i,\epsilon_i} \circ \hat{h}_{p-i}(\hat{z}), \qquad \text{de sorte que} \quad h = h_{i-1} \circ h_{m_i,\epsilon_i} \circ \hat{h}_{p-i}(\hat{z}),$$

fait intervenir l'homographie $\hat{h}_{p-i}$ qui regroupe les $p-i$ étapes de la fin. Alors

$$|h'(\hat{z})|^2 D(h(\hat{z})) = \sum_{i=1}^{p} \ell(m_i)|h'(\hat{z})|^2 \, |h'_{i-1}(h_{m_i,\epsilon_i} \circ \hat{h}_{p-1}(\hat{z}))|. \tag{2.19}$$

En utilisant la dérivée d'une composition, le terme général de la somme (2.19) devient

$$\left[|\hat{h}'_{p-i}(\hat{z})|^2\right] \cdot \left[(\ell(m_i)|h'_{m_i,\epsilon_i}(z_i)|^2\right] \cdot \left[\lg|h'_{i-1}(z_i)| \cdot |h'_{i-1}(z_{i-1})|^2\right].$$

En utilisant les opérateurs élémentaires, ce terme (2.18) s'écrit,

$$\mathbf{H}_{2,[\hat{h}_{p-i}]} \circ W_{(c)} \mathbf{H}_{2,[h_{m_i,\epsilon_i}]} \circ \Delta\mathbf{H}_{2,[h_{i-1}]}[F](\hat{z},\bar{\hat{z}}).$$

En sommant sur touts les i et toutes les hauteurs possibles p, et en utilisant la linéarite des fonctionnelles W et Δ, on obtient

$$\sum_{h\in\mathcal{H}^+} |h'(\hat{z})|^2 \, D(h(\hat{z})) \, F(h(\hat{z}),h(\bar{\hat{z}})) = \sum_{p\geq 1} \sum_{i=1}^{p} \mathbf{H}_2^{p-i} \circ \mathbf{H}_2^{(c)} \circ \Delta\mathbf{H}_2^{i-1}[F](\hat{z},\bar{\hat{z}}).$$

En intervertissant les sommes, on obtient

$$\sum_{p\geq 1} \sum_{i=1}^{p} \mathbf{H}_2^{p-i} \circ \mathbf{H}_2^{(c)} \circ \Delta\mathbf{H}_2^{i-1} = \left(\sum_{p\geq i} \mathbf{H}_2^{p-i}\right) \circ \mathbf{H}_2^{(c)} \circ \left(\sum_{i\geq 1} \Delta\mathbf{H}_2^{i-1}\right)$$

puis, en utilisant encore la linéarité de Δ, en répérant les quasi-inverses, et en utilisant l'identité (conséquence de la proposition 2.1)

$$\Delta(I - \mathbf{H}_2)^{-1} = (I - \mathbf{H}_2)^{-1} \circ \Delta\mathbf{H}_2^{-1} \circ (I - \mathbf{H}_2)^{-1},$$

on obtient

$$\sum_{p\geq 1}\sum_{i=1}^{p} \mathbf{H}_2^{p-i} \circ \mathbf{H}_2^{(c)} \circ \Delta\mathbf{H}_2^{i-1} = (I - \mathbf{H}_2)^{-1} \circ \Delta\mathbf{H}_2^{-1} \circ (I - \mathbf{H}_2)^{-1}$$

qui permet de conclure la démonstration. □

2.4 Propriétés spectrales des opérateurs.

Jusque là, nous avons défini formellement les opérateurs de transfert et nous avons exprimé les quantités qui nous intéressent en fonction de ces opérateurs. Les expressions sont pour l'instant formelles, algébriques, et ne comportent pas d'information *quantitative*. C'est le but de la présente section d'introduire de l'analyse, de sorte que l'on puisse obtenir des informations quantitatives. Il y a deux ingrédients principaux, qui sont liés : les espaces fonctionnels et la théorie spectrale. Les espaces fonctionnels fournissent à la fois l'espace vectoriel sur lequel l'opérateur agit, mais aussi une norme, qui donne à la fois une notion de taille et de distance. La norme associée à l'espace fonctionnel donne les moyens de *mesurer la taille* des objets, et donc d'approximer les quantités en jeu. La théorie spectrale permet de *décomposer* les opérateurs. Les deux aspects vont se croiser lorsqu'il s'agira de construire une décomposition en une partie *dominante* et une partie *dominée*.

Dans cette section nous allons rappeler quelques éléments de théorie spectrale, concernant les opérateurs compacts sur un espace de Banach. On commence par discuter le choix d'un espace fonctionnel adapté.

2.4.1 Espaces fonctionnels.

Les espaces fonctionnels usuels sont toujours des espaces de Banach. Le choix d'un espace fonctionnel est dirigé par deux principes : tout d'abord, il doit contenir les fonctions étudiées. Ensuite, il doit avoir suffisamment de structure pour que les opérateurs de transfert y possèdent de bonnes propriétés, notamment spectrales. Donc, l'espace ne doit être, ni trop petit, car il doit contenir les fonctions étudiées, ni trop grand, car le spectre est alors trop grand lui aussi, sans propriétés spectrales intéressantes. Dans notre contexte, les fonctions étudiées sont les densités d'entrée qui sétendent en des fonctions de deux variables. Afin d'obtenir naturellement des opérateurs de transfert compacts, on va travailler avec des fonctions analytiques de deux variables $G(u, v)$, sans que cela limite la possibilité d'utiliser des fonctions de la forme

$$F(u, v) = |u - v|^r \cdot G(u, v),$$

grâce à la propriété (*iii*) de la proposition 2.2. Il nous reste maintenant à préciser cet espace fonctionnel, et c'est l'objet de la définition suivante.

Définition 2.2 (Espaces fonctionnels $A_\infty(\mathcal{V}), B_\infty(\mathcal{V})$)**.** *Soit $\mathcal{V}$ un disque ouvert sur le plan complexe. L'espace $A_\infty(\mathcal{V})$ est l'espace de Banach des fonctions complexes d'une variable, analytiques sur $\mathcal{V}$ et continues sur $\bar{\mathcal{V}}$, avec la norme*

$$||f|| = \sup_{u\in\bar{\mathcal{V}}} |f(u)|.$$

L'espace $B_\infty(\mathcal{V})$ est l'espace de Banach des fonctions complexes à deux variables, analytiques sur $\mathcal{V}\times\mathcal{V}$ et continues sur $\overline{\mathcal{V}}\times\overline{\mathcal{V}}$, avec la norme

$$||F|| = \sup_{(u,v)\in\overline{\mathcal{V}}\times\overline{\mathcal{V}}} |F(u,v)|.$$

Les espaces fonctionnels $A_\infty(\mathcal{V}), B_\infty(\mathcal{V})$ appartiennent à la classe des espaces *nucléaires d'ordre* 0. Ils ont été étudiés par Grothendieck dans les années 50 ([29, 28]). Ils sont remarquables dans le sens que tout opérateur borné agissant sur l'un de ces espaces est lui-même nucléaire, et possède automatiquement des propriétés très fortes : il est notamment compact. Nous allons rentrer dans les détails plus tard dans le chapitre, en spécifiant le disque $\mathcal{V}$ (qui dépend essentiellement de la géométrie des branches primaires) et en prouvant que l'opérateur $\mathbf{H}_{s,w}$ agit sur $A_\infty(\mathcal{V})$, l'opérateur $\underline{\mathbf{H}}_{s,w}$ agit sur $B_\infty(\mathcal{V})$ et que ce sont des opérateurs bornés.

On dit que l'opérateur $\mathbf{T}$ est *compact*, lorsque l'adhérence de l'image par $\mathbf{T}$ de la boule unité ouverte est un sous-espace compact de E. Nous allons insister sur les propriétés spécifiques de ce genre d'opérateur, puisque l'opérateur de transfert en est un, comme on verra.

2.4.2 Valeurs propres, vecteurs propres.

A un opérateur $\mathbf{T}\in\mathfrak{B}(E)$, on associe l'opérateur $(\mathbf{T}-\zeta\mathbf{I})$. L'ensemble des $\zeta\in\mathbb{C}$ pour lesquels cet opérateur est non inversible s'appelle *spectre* de $\mathbf{T}$, et il noté $\sigma(\mathbf{T})$. La fonction $\zeta\to(\mathbf{T}-\zeta\mathbf{I})^{-1}$ est la *résolvante* de $\mathbf{T}$, et elle est définie sur *l'ensemble résolvant* $\rho(\mathbf{T})=\mathbb{C}\setminus\sigma(\mathbf{T})$. En dimension infinie, un opérateur peut être non inversible car il est soit non injectif, soit non surjectif. On dit que $\lambda\in\mathbb{C}$ est une *valeur propre* de $\mathbf{T}$ lorsque l'opérateur $(\mathbf{T}-\lambda\mathbf{I})$ est non injectif : ll existe alors $f\in E\setminus\{0\}$, vérifiant $\mathbf{T}f=\lambda f$. Une telle fonction f est appelée *fonction propre*. De façon équivalente, λ est une valeur propre si le noyau de $\mathbf{T}-\lambda\mathbf{I}$, appelé l'*espace propre* associé à λ, est de dimension $d\geq 1$. Cette dimension d est la *multiplicité géométrique* de λ.

Le rayon spectral de l'opérateur $\mathbf{T}$, désigné par $r(\mathbf{T})$ est défini par

$$r(\mathbf{T}) := \sup\{|\lambda|;\quad \zeta\in\sigma(\mathbf{T})\},$$

et le théorème du rayon spectral affirme que

$$r(\mathbf{T}) = \lim_{k\to\infty} ||\mathbf{T}^k||^{1/k},$$

et ceci, pour toute norme $||.||$.

Lorsque $\mathbf{T}$ est un opérateur compact, le spectre privé de la valeur 0 est un ensemble discret, et n'est formé que de valeurs propres propres isolées.

2.4.3 Décomposition spectrale.

Comme on l'a déjà dit, pour étudier plus facilement les opérateurs, on cherche à les décomposer en opérateurs plus simples, en utilisant notamment des projecteurs. Soit $\mathbf{P}$ un opérateur de $\mathfrak{B}(E)$ idempotent (ou projecteur), vérifiant $\mathbf{P}^2=\mathbf{P}$. Alors, l'espace E se décompose sous la forme

$$E = M\oplus N$$

où $M=\mathbf{P}E$ et $N=(\mathbf{I}-\mathbf{P})E$. Les espaces M et N sont des sous-espaces fermés de E. De façon réciproque, lorsque E se décompose sous la forme

$$E = M\oplus N,\qquad \text{avec } M,\ N \text{ des sous-espaces vectoriels fermés de } E,$$

il existe un opérateur $\mathbf{P} \in \mathfrak{B}(E)$ idempotent pour lequel $\mathbf{P}E = M,\ (\mathbf{I}-\mathbf{P})E = N$.

On cherche maiontenant à définir des projecteurs à partir du spectre $\sigma(\mathbf{T})$ d'un opérateur compact $\mathbf{T}$. La compacité de $\mathbf{T}$ garantit que tout élément non-nul du spectre $\lambda \in \sigma(\mathbf{T})$ y est isolé, et il existe donc une courbe $\mathcal{C}$ simple et régulière sur le plan complexe, entourant λ et isolant λ du reste du spectre $\sigma(\mathbf{T}) \setminus \{\lambda\}$. L'opérateur $\mathbf{P}$ défini par

$$\mathbf{P} := \frac{1}{2\pi i}\int_{\mathcal{C}} (\mathbf{T}-\zeta\mathbf{I})^{-1} d\zeta,$$

est un projecteur, qui ne dépend pas de la courbe $\mathcal{C}$ et seulement de la valeur propre λ : c'est le projecteur associé à λ. La dimension de l'espace $\mathbf{P}E$ est la *multiplicité algébrique* de λ, qui est toujours supérieure ou égale à la multiplicité géométrique. Les valeurs propres non-nulles des opérateurs compacts sont toutes de multiplicité algébrique finie. Lorsqu'une valeur propre λ possède une multiplicité algébrique égale à la multiplicité géométrique, elle est dite semi-simple. Une valeur propre est simple si sa multiplicité algébrique (et donc sa multiplicité géométrique) vaut 1. Une valeur propre simple est toujours semisimple.

Nous considérons maintenant le cas, essentiel pour la suite, où l'opérateur $\mathbf{T}$ a une unique valeur propre dominante : Cela signifie qu'il existe une valeur propre λ de multiplicité algébrique égale à 1, qui vérifie : Pour tout $\zeta \in \sigma(\mathbf{T}) \setminus \{\lambda\}$, le module $|\zeta|$ vérifie $|\zeta| < |\lambda|$.

2.4.4 Opérateurs compacts avec une unique valeur propre dominante.

Dans ce cas, la décomposition $E = \mathbf{P}E \oplus (\mathbf{I}-\mathbf{P})E$ définit deux sous espaces stables par $\mathbf{T}$. De plus, puisque $\mathbf{P}E$ est de dimension 1, et est stable par $\mathbf{T}$, on a $\mathbf{T}\circ\mathbf{P} = \lambda\mathbf{P} = \mathbf{P}\circ\mathbf{T}$. Posant $\mathbf{N} = \mathbf{T}\circ(\mathbf{I}-\mathbf{P}) = (\mathbf{I}-\mathbf{P})\circ\mathbf{T}$, l'opérateur $\mathbf{T}$ se décompose en

$$\mathbf{T} = \lambda\mathbf{P} + \mathbf{N}. \tag{2.20}$$

Remarquons que, par défintion, le spectre de $\mathbf{N}$ vérifie $\sigma(\mathbf{N}) = \sigma(\mathbf{T}) \setminus \{\lambda\}$ et qu'on a l'égalité $\mathbf{N}\circ\mathbf{P} = \mathbf{P}\circ\mathbf{N} = \mathbf{0}$. La décomposition (2.20) permet alors d'obtenir une décomposition pour toutes les puissances de $\mathbf{T}$, pour tout $k \geq 1$

$$\mathbf{T}^k = \lambda^k\mathbf{P} + \mathbf{N}^k. \tag{2.21}$$

Le rayon spectral de $\mathbf{N}$ l'inégalité $r(\mathbf{N}) \leq \rho|\lambda|$, avec $\rho < 1$, et le thórème du rayon spectral montre que, pour k suffisamment grand, on a $||\mathbf{N}^k|| \leq (\rho')^k|\lambda|^k$, pour un certain ρ' vérifiant $\rho < \rho' < 1$. Cela montre donc que le premier terme de la décomposition (2.21) domine strictement le second.

Si maintenant $||\mathbf{T}|| = \rho < 1$, le théorème du rayon spectral montre l'inégalité $r(\mathbf{T}) \leq \rho < 1$, et donc, à la fois λ et le spectre $\sigma(\mathbf{N})$ sont inclus dans le disque $\{|z| \leq \rho\}$. Alors, le théorème du rayon spectral montre l'inégalité $||\mathbf{N}^k|| < 1$ pour k suffisamment grand, et donc l'existence des deux quasi-inverses $(I-\mathbf{T})^{-1}, (I-\mathbf{N})^{-1}$, et finalement la décomposition,

$$(\mathbf{I}-\mathbf{T})^{-1} = \frac{\lambda}{1-\lambda}\mathbf{P} + \mathbf{N}.$$

Puisque $\mathbf{P}E$ est de dimension 1, il est engendré par une fonction propre relative à λ, que nous désignons par ψ. Alors, pour toute fonction f de E, il existe un complexe $\nu[f]$ pour lequel on a $\mathbf{P}[f] = \psi\cdot\nu[f]$. Cela définit un élément ν du dual E^* de E. On a, pour tout $f \in E$,

$$\mathbf{P}[\mathbf{T}f] = \psi\cdot\nu[\mathbf{T}f], \qquad \text{mais aussi} \quad \mathbf{P}[\mathbf{T}f] = \lambda\mathbf{P}[f] = \lambda\,\psi\cdot\nu[f]. \tag{2.22}$$

Cela montre l'égalité $\nu[\mathbf{T}f] = \lambda\nu[f]$, et la définition de l'opérateur adjoint T^* prouve que ν est une mesure propre pour T^*. Si on normalise ν en exigeant $\nu[1] = 1$, alors le vecteur propre ψ est unique et on a aussi la relation $\nu[\psi] = 1$.

2.4.5 Existence d'une valeur propre dominante

Dans cette section, nous énonçons le théorème de Krasnoselskii [38], qui donne des conditions suffisantes pour l'existence d'une valeur propre dominante. Ces conditions font intervenir les opérateurs u_0-positifs, qui généralisent les opérateurs positifs de la dimension finie et possèdent donc des propriétés spectrales dominantes.

Un sous-ensemble K d'un espace de Banach réel B est appelé un *cône propre* si

(*i*) pour tout réel $\varrho > 0$ et tout vecteur f de K, $\varrho f \in K$,

(*ii*) $K \cap -K = \{0\}$.

Un cône propre est appelé *reproductif* si $B = K - K$, c'est-à-dire si tout élément f de B s'écrit comme la différence de deux éléments de K.

Soit K un cône propre, reproductif et d'intérieur $\mathring{K}$ non vide. On dit que $\mathbf{T} : B \to B$ est *positif* (par rapport au cône K) si $\mathbf{T}(K)$ est inclus dans K. Soit u_0 un élément de $\mathring{K}$; on dit que l'opérateur positif $\mathbf{T}$ est u_0-*positif* par rapport au cône K si, pour tout élément f non nul de K, il existe un entier p et deux réels α et β strictement positifs pour lesquels

$$\beta u_0 \leq \mathbf{T}^p f \leq \alpha u_0, \tag{2.23}$$

et l'ordre est défini en relation avec K : $f \leq g$ si et seulement si $g - f \in K$.

Alors, le théorème de Krasnoselskii [38] s'énonce ainsi : *Tout opérateur* $\mathbf{T}$ *compact et* u_0*-positif satisfait à une propriété de type Perron-Frobenius : il a une unique fonction propre* g *dans* $\mathring{K}$ *et la valeur propre associée* λ *est simple, et strictement plus grande en valeur absolue que les autres valeurs propres.*

2.4.6 Théorie de la perturbation

Les objets spectraux d'un opérateur $\mathbf{T}_t$ qui dépend analytiquement d'un paramètre jouissent elles aussi de cette dépendance analytique, pourvu qu'elles soient elles-mêmes bien définies. C'est ce que dit la théorie de la perturbation, dont nous énonçons ici les principaux résultats directement appliqués à notre contexte.

Proposition 2.6. *Soit* $\mathbf{T}_t$ *un opérateur qui vérifie les deux propriétés*

(*i*) *L'application* $t \mapsto \mathbf{T}_t$ *est analytique dans un voisinage de* $t = t_0$

(*ii*) *Pour* $t = t_0$*, l'opérateur* $\mathbf{T}_t$ *admet une valeur propre simple* λ *qui est isolée dans* $\sigma(\mathbf{T}_t)$*. On désigne par* $\mathbf{P}, \mathbf{N}$ *les opérateurs associés à la décomposition spectrale de* $\mathbf{T}_{t_0}$*.*

Alors, il existe un voisinage $\mathcal{V}$ *de* t_0 *sur lequel les propriétés suivantes sont vérifiées :*

(*i*) *L'opérateur* $\mathbf{T}_t$ *admet une valeur propre simple* $\lambda(t)$*, isolée dans* $\sigma(\mathbf{T}_t)$*, qui vérifie* $\lambda(t_0) = \lambda$*, qui définit des opérateurs* $\mathbf{P}_t, \mathbf{N}_t$ *associés à la décomposition spectrale de* $\mathbf{T}_t$*.*

(*ii*) *Les applications* $t \mapsto \lambda(t),\ t \mapsto \mathbf{P}_t,\ t \mapsto \mathbf{N}_t$ *sont analytiques sur* $\mathcal{V}$*,*

(*iii*) *Si* $\lambda'(t_0) \neq 0$*, l'égalité* $\lambda(t) = \lambda$ *définit une courbe analytique* $z \mapsto t(z)$ *sur laquelle* $\lambda(t(z)) = \lambda$*.*

Preuve. Ces propriétés sont abordées dans la théorie de la perturbation analytique, dans [34], où on renvoie le lecteur pour les preuves. □

2.5 Propriétés spectrales des opérateurs de transfert des systèmes de Gauss et d'Euclide.

On rappelle que les deux opérateurs de transfert que nous voulons étudier sont définis par

$$\mathbf{H}_{s,w}[f](x) = \sum_{(m,\epsilon)\geq(2,1)} \left(\frac{1}{m+\epsilon x}\right)^{2s} \cdot e^{wc(m,\epsilon)} \cdot f\left(\frac{1}{m+\epsilon x}\right). \tag{2.24}$$

(pour l'algorithme d'Euclide), et par

$$\underline{\mathbf{H}}_{s,w}[F](u,v) = \sum_{(m,\epsilon)\geq(2,1)} \left(\frac{1}{m+\epsilon u}\right)^{s} \cdot \left(\frac{1}{m+\epsilon v}\right)^{s} \cdot e^{wc(m,\epsilon)} \cdot F\left(\frac{1}{m+\epsilon u}, \frac{1}{m+\epsilon v}\right), \tag{2.25}$$

(pour l'algorithme de Gauss).

On omettra l'indice w quand il est égal à 0, et omettra l'indice s quand il est égal à 1.

2.5.1 Espaces fonctionnels adéquats.

Il faut d'abord trouver un espace fonctionnel adéquat, pour chacun des opérateurs. Comme nous l'avons annoncé, ce sera un espace fonctionnel du type $A_\infty(\mathcal{V})$ pour $\mathbf{H}_{s,w}$ et du type $B_\infty(\mathcal{V})$ pour $\underline{\mathbf{H}}_{s,w}$, qui sont des espaces nucléaires, où tout opérateur borné est compact. Il reste à préciser le disque $\mathcal{V}$, qui est choisi en fonction des propriétés de l'ensemble $\mathcal{H}$ des branches. Avec un choix adéquat de $\mathcal{V}$, les opérateurs seront bornés et donc compacts.

Définition 2.3 (Espaces fonctionnels $A_\infty(\mathcal{V}), B_\infty(\mathcal{V})$)**.** *On désigne par $\mathcal{V}$ le disque ouvert de centre $1/4$ et rayon $1/2$. L'espace $A_\infty(\mathcal{V})$ est l'espace de Banach des fonctions complexes d'une variable, analytiques sur $\mathcal{V}$ et continues sur $\bar{\mathcal{V}}$, avec la norme*

$$||f|| = \sup_{u\in\bar{\mathcal{V}}} |f(u)|.$$

L'espace $B_\infty(\mathcal{V})$ est l'espace de Banach des fonctions complexes à deux variables, analytiques sur $\mathcal{V}\times\mathcal{V}$ et continues sur $\overline{\mathcal{V}}\times\overline{\mathcal{V}}$, avec la norme

$$||F|| = \sup_{(u,v)\in\overline{\mathcal{V}}\times\overline{\mathcal{V}}} |F(u,v)|.$$

2.5.2 Propriétés analytiques des branches du système Gauss-interne.

Le choix du disque $\mathcal{V}$ n'est arbitraire ; il est là pour assurer aux branches inverses du système Gauss-interne de bonnes propriétés. Dans le lemme suivant, nous établissons des propriétés importantes des branches du système Gauss-interne. Elles permettront d'assurer que l'espace $B_\infty(\mathcal{V})$ est un espace nucléaire au sens de Grothendieck, où tout opérateur qui y agit est lui-même nucléaire, et donc compact.

Proposition 2.7 (Propriétés des branches primaires)**.** *L'ensemble $\mathcal{H}$ du système* Gauss-interne *(qui coincide avec celui de l'algorithme* Euclide-plié*) satisfait les propriétés suivantes :*

(i) Toute homographie $h\in\mathcal{H}$ envoie le disque $\overline{\mathcal{V}}$ dans son intérieur $\mathcal{V}$,

$$h(\overline{\mathcal{V}}) \subset \mathcal{V} \qquad \text{pour tout } h\in\mathcal{H}.$$

(ii) Il existe $\delta < 1$ tel que, pour toute branche $h \in \mathcal{H}$ et pour tout $(u, v) \in \overline{\mathcal{V}} \times \overline{\mathcal{V}}$ on a

$$|h(u) - h(v)| \leq \delta_h \, |u - v| \qquad \text{avec } \delta_h \leq \delta < 1.$$

(iii) Il existe $K > 0$ telle que pour toute branche $h \in \mathcal{H}$, et pour tout u, v réels de $\overline{\mathcal{V}}$,

$$\left|\frac{h''(u)}{h'(u)}\right| \leq K, \qquad \left|\frac{h'(v)}{h'(u)}\right| \leq e^{K|u-v|}$$

Remarque. C'est la propriété (i) qui est la plus importante et qui permet de montrer que les opérateurs $\mathbf{H}_{s,w}, \underline{\mathbf{H}}_{s,w}$ sont bornés dans $A_\infty(\mathcal{V})$ ou $B_\infty(\mathcal{V})$ (selon le cas), et qu'ils sont donc compacts.

Démonstration. Preuve de (i). Comme les homographies de $\mathcal{H}$ envoient disques en disques, réels en réels et vérifient $\overline{h(z)} = h(\bar{z})$ pour tout z, il suffit de vérifier que l'homographie $h(z) = 1/(m + \epsilon z)$ envoie l'intervalle fermé $\overline{\mathcal{V}} \cap \mathbb{R}$, dans l'intervalle ouvert correspondant. L'intervalle $\overline{\mathcal{V}} \cap \mathbb{R}$ est égal à $]-1/4, 3/4[$. Lorsque $(m, \epsilon) \geq (3, -1)$, les deux points $h(-1/4)$ et $h(3/4)$ sont tous deux éléments de $]-1/4, 3/4[$, comme on le vérifie facilement. Lorsque $(m, \epsilon) = (2, 1)$, on a $h(-1/4) = 4/7 < 3/4$ et $h(3/4) = 4/11 < 3/4$. Donc, l'intervalle $[-1/4, 3/4] = \overline{\mathcal{V}} \cap \mathbb{R}$ est envoyé dans l'intervalle $]-1/4, 3/4[$ et donc $\overline{\mathcal{V}}$ est envoyé en $\mathcal{V}$, comme on voulait prouver.

Preuve de (ii). Un calcul direct, déjà utilisé, montre que toute homographie de déterminant ± 1 vérifie

$$\left|\frac{h(u) - h(v)}{u - v}\right| = \sqrt{|h'(u)| \cdot |h'(v)|},$$

pour tous les $u, v \in \mathbb{C}$ distincts, et pour $u = v$ par continuité. Dans le cas d'une branche de $\mathcal{H}$, primaires, désignée par $h_{m,\epsilon}$, il prend la forme

$$\left|\frac{h_{m,\epsilon}(u) - h_{m,\epsilon}(v)}{u - v}\right| = \frac{1}{|\epsilon u + m| \cdot |\epsilon v + m|}. \tag{2.26}$$

Lorsque $\epsilon = 1$, le côté droit est maximum pour $u = v = -1/4$. Nous avons,

$$\left|\frac{h_{m,1}(u) - h_{m,1}(v)}{u - v}\right| \leq \frac{1}{(m - 1/4)^2} \leq \frac{16}{49}$$

la borne donnée correspondant au cas $m = 2$. Par ailleurs, si $\epsilon = -1$, le côté droit de (2.26) est maximum dans $\overline{\mathcal{V}} \times \overline{\mathcal{V}}$ pour $u = v = 3/4$. En conséquence,

$$\left|\frac{h_{m,-1}(u) - h_{m,-1}(v)}{u - v}\right| \leq \frac{1}{(m - 3/4)^2} \leq \frac{16}{81}$$

la borne donnée correspondant maintenant au cas $m = 3$. En total, nous avons montré que nous pouvons prendre pour établir (ii)

$$\delta_{h_{m,+1}} = \frac{1}{(m - 1/4)^2}, \qquad \delta_{h_{m,-1}} = \frac{1}{(m - 3/4)^2}, \qquad \text{et finalement} \quad \delta = \frac{16}{49},$$

Preuve de (iii). Nous avons, pour une branche $h(z) = 1/(m + \epsilon z)$,

$$\left|\frac{h''(u)}{h'(u)}\right| = \frac{2}{|(m + \epsilon u)|} \leq \frac{2}{||m| - |u||} \leq \frac{8}{5}, \qquad \text{car } m \geq 2 \text{ et } |u| \leq 3/4.$$

La propriété (iv) est à prouver pour $u, v \in \overline{\mathcal{V}}$ réels : cela garantit que la fonction $u \mapsto |h'(u)|$ est dérivable, puisque le signe de $h'(u)$ ne change pas dans $\overline{\mathcal{V}} \cap \mathbb{R}$. Par le théorème des accroissements finis, appliqué à la fonction $x \mapsto \log(h'(x))$, nous obtenons

$$|\log|h'(u)| - \log|h'(v)|| = \left|\frac{h''(w)}{h'(w)}\right| \cdot |u - v| \qquad \text{pour } w \in [u, v],$$

et donc, avec (iii), on obtient $\log|h'(u)| - \log|h'(v)| \leq K|u - v|$. Il suffit alors de prendre l'exponentielle des deux côtés pour conclure. □

Comme nous l'avons dit, cette proposition permet de montrer que les opérateurs $\mathbf{H}_{s,w}$ et $\underline{\mathbf{H}}_{s,w}$ sont compacts, pourvu qu'ils soient bornés sur $B_\infty(\mathcal{V})$, ce que nous discutons maintenant.

2.5.3 Domaine de définition des opérateurs.

Les opérateurs $\mathbf{H}_{s,w}$ et $\underline{\mathbf{H}}_{s,w}$ dépendent d'un coût 'élémentaire c, et les propriétés de ces opérateurs dépendent de la croissance de ce coût. La classe de coûts définie ci-dessous contient tous les coûts intéressants, et permettra aux opérateurs associé d'être bornés sur $A_\infty(\mathcal{V})$, $B_\infty(\mathcal{V})$ (et donc compacts).

Définition 2.4. [Coût à croissance modérée]. *On dit que le coût élémentaire $c : \mathbb{N} \to \mathbb{N}$ est à croissance modérée si*

$$p_c^{(-)} := \liminf_{m\to\infty} \frac{c(m)}{\log(m)} \qquad et \qquad p_c^{(+)} = \limsup_{m\to\infty} \frac{c(m)}{\log(m)}$$

sont finis. L'ensemble $\mathfrak{H}_{(c)}$ et sa frontière $\mathcal{F}_{(c)}$ sont alors définis par

$$\mathfrak{H}_{(c)} = \left\{(s, w) \in \mathbb{C}^2 \ : \ \Re s > \frac{1}{2}\left[1 + p_c^{(\text{signe}(\Re w))} \cdot \Re w\right]\right\}, \qquad \mathcal{F}_{(c)} := \text{ Frontière } \mathfrak{H}_{(c)} \tag{2.27}$$

Proposition 2.8. *On considère un coût élémentaire c modéré, et le domaine $\mathfrak{H}_{(c)}$ correspondant. Les opérateurs de transfert $\mathbf{H}_{s,w}$ et $\underline{\mathbf{H}}_{s,w}$ définis en (2.24) et (2.25) possédent les propriétés suivantes :*

- (i) *Pour $(s, w) \in \mathfrak{H}_{(c)}$, l'opérateur $\mathbf{H}_{s,w}$ agit sur $A_\infty(\mathcal{V})$, l'opérateur $\underline{\mathbf{H}}_{s,w}$ agit sur $B_\infty(\mathcal{V})$, et ce sont des opérateurs bornés et compacts.*
- (ii) *Lorsque le couple (s, w) réel tend vers un point de la frontière $\mathcal{F}_{(c)}$, alors la norme de l'opérateur $\underline{\mathbf{H}}_{s,w}$ tend vers l'infini.*

Démonstration. On prouve d'abord (i) dans le cas de l'opérateur généralisé. Soit $F \in B_\infty(\mathcal{V})$ et un point (s, w) satisfaisant (2.27). Pour prouver que $\underline{\mathbf{H}}_{s,w}[F]$ est une fonction de $B_\infty(\mathcal{V})$, nous montrons que chaque opérateur composante défini par

$$\underline{\mathbf{H}}_{s,w,[h]}[F](u, v) := \frac{1}{(m + \epsilon u)^s} \frac{1}{(m + \epsilon v)^s} e^{wc(m)} F(h(u), h(v)) \tag{2.28}$$

agit sur $B_\infty(\mathcal{V})$, et que la série est normalement convergente. La norme de l'opérateur $\underline{\mathbf{H}}_{s,w,[h]}$ s'étudie comme suit. En posant $\Re(w) = \tau$, nous avons

$$|\exp(wc(m))| = \exp(\tau c(m)) = m^{\tau \frac{c(m)}{\log m}} \quad \text{ et } \quad |F(h(u), h(v))| \leq ||F|| O(1),$$

et par ailleurs, si $s = \sigma + it$, pour tout $u \in \mathcal{V}$,

$$\frac{1}{|(m+\epsilon u)^s|} = \frac{1}{|m^s| \cdot |(1+\epsilon u/m)^s|} = \frac{1}{m^\sigma} \frac{1}{|(1+\epsilon u/m)^s|}$$

et, lorsque $m \to \infty$,

$$\begin{aligned} |(1+\epsilon u/m)^{-s}| &= |\exp(-s \log(1+\epsilon u/m)|) \\ &= \exp(-\sigma \cdot \log|1+\epsilon u/m| + t \cdot \arg(1+\epsilon u/m)) \\ &= |1+\epsilon u/m|^{-\sigma} \cdot \exp(t \cdot O(1)) \\ &= O_s(1), \end{aligned}$$

où la constante dans $O_s(1)$ dépend seulement de s. Ainsi,

$$\frac{1}{|(m+\epsilon u)^s|} = O_s\left(\frac{1}{m^\sigma}\right).$$

Finalement, la norme de l'opérateur composante $\underline{\mathbf{H}}_{s,w,[h]}$ associé à un couple (m, ϵ) satisfait

$$||\underline{\mathbf{H}}_{s,w,[h]}|| = O_s\left(m^{-e(s,w)}\right) \qquad \text{avec} \quad e(s,w) = 2\Re s - \Re w \frac{c(m)}{\log m}.$$

Cela montre que la série (2.28) est normalement convergente lorsque $e(s,w)$ est supérieur à 1, et donc ausitôt que s et w vérifient (2.27). Comme la convergence normale assure la conservation de l'analyticité et de la continuité, l'opérateur $\underline{\mathbf{H}}_{s,w}$ agit sur $B_\infty(\mathcal{V})$, et y est borné. D'après la propriété de nucléarité, c'est aussi un opérateur compact.

Preuve de (ii). On remarque, que la preuve précédente peut se préciser quand s et w sont réels. Alors, les O des estimations précédentes se transforment en Θ, et on peut prouver que

$$||\underline{\mathbf{H}}_{s,w,[h]}|| = \Theta_s\left(m^{-e(s,w)}\right) \qquad \text{avec} \quad e(s,w) = 2\Re s - \Re w \frac{c(m)}{\log m}.$$

En faisant agir l'opérateur sur la fonction constante égale à 1, on déduit alors que la norme de l'opérateur $\underline{\mathbf{H}}_{s,w}$ lui-même est $\Theta(m^{-e(s,w)})$. Comme $e(s,w)$ tend vers 1 quand (s,w) tend vers un point de $\mathcal{F}_{(c)}$, cela conclut la preuve de (ii). □

2.5.4 Existence d'une valeur propre dominante pour s, w réels.

La propriété spectrale essentielle des deux opérateurs $\mathbf{H}_{s,w}$ et $\underline{\mathbf{H}}_{s,w}$ est l'existence d'une valeur propre dominante unique, simple et isolée du reste du spectre. C'est le théorème de Krasnoselskii qui va prouver ce résultat. Cela permettra de décomposer comme nous l'avons annoncé en (2.21) l'opérateur en une partie "dominante" et une autre partie "dominée".

Proposition 2.9. *On considère un coût élémentaire $c : \mathbb{N} \to \mathbb{N}$ de croissance modérée et un couple $(s,w) \in \mathfrak{H}_{(c)}$. Alors, pour un couple (s,w) réel,*

(i) l'opérateur $\mathbf{H}_{s,w}$ possède une unique valeur propre dominante $\lambda(s,w)$ réelle, associée à une fonction propre $\psi_{s,w}$ strictement positive. Le vecteur propre $\nu_{s,w}$ de l'opérateur adjoint $\mathbf{H}^*_{s,w}$ est une mesure de Radon positive, et, lorsqu'on on les normalise, on a $\nu_{s,w}[1] = 1$ et $\nu_{s,w}[\psi_{s,w}] = 1$. La paire $(\psi_{s,w}, \nu_{s,w})$ est unique.

(*ii*) *l'opérateur* $\underline{\mathbf{H}}_{s,w}$ *possède une unique valeur propre dominante* $\underline{\lambda}(s,w)$*, réelle pour* s, w *réels, associée à une fonction propre* $\underline{\psi}_{s,w}$ *strictement positive. Le vecteur propre* $\underline{\nu}_{s,w}$ *de l'opérateur adjoint* $\underline{\mathbf{H}}^*_{s,w}$ *est une mesure de Radon positive, et, lorsqu'on on les normalise, on a* $\underline{\nu}_{s,w}[1] = 1$ *et* $\underline{\nu}_{s,w}[\underline{\psi}_{s,w}] = 1$*. La paire* $(\underline{\psi}_{s,w}, \underline{\nu}_{s,w})$ *est unique.*

Démonstration. Nous reprenons essentiellement la preuve donnée dans l'article [15], er nous l'adaptons à un opérateur mondéré. Il s'agit de montrer que l'opérateur $\underline{\mathbf{H}}_{s,w}$ est u_0-positif lorsque $s, w \in \mathbb{R}$ et ensuite d'appliquer le théorème de Krasnoselskii. Soit $\mathcal{J} = \mathcal{V} \cap \mathbb{R}$ et considérons le sous-espace vectoriel réel $B^{\mathbb{R}}_\infty(\mathcal{V})$ de $B_\infty(\mathcal{V})$, qui contient les fonctions définies sur $\mathcal{V}$ et à valeurs réelles sur $\mathcal{J} \times \mathcal{J}$. Pour $a > 0$ on considère l'ensemble

$$K_a = \{F \in B^{\mathbb{R}}_\infty(\mathcal{V}) \ : \ \forall u, v \in \mathcal{J},\ 0 \le F(u,v) \le e^{a|u-v|} F(u,u)\}.$$

Nous allons montrer que K_a est un cône propre, reproductif et d'intérieur non vide. C'est un cône propre puisque si $\rho > 0$ et $F \in K_a$, alors évidemment $\rho F \in K_a$, et puisque $F, -F \in K_a$ implique $F = 0$. C'est un cône reproductif puisqu'à toute fonction $F \in B^{\mathbb{R}}_\infty(\mathcal{V})$ nous pouvons associer $F + R \cdot 1$, qui est une fonction de K_a dès que

$$R \ge M + \frac{c}{a} \ge \sup_{(u,v) \in \mathcal{J} \times \mathcal{J}} \frac{F(u,v) - e^{a|u-v|} F(u,u)}{e^{a|u-v|} - 1} \tag{2.29}$$

puisque dans ce cas

$$(e^{a|u-v|} - 1)R \ge F(u,v) - e^{a|u-v|} F(u,u) \text{ pour tout } (u,v) \in \mathcal{J} \times \mathcal{J}$$

et donc

$$e^{a|u-v|}(F(u,u) + R) \ge F(u,v) + R \text{ pour tout } (u,v) \in \mathcal{J} \times \mathcal{J},$$

ce qui montre que $F + R \cdot 1$ appartient à K_a. Le cône K_a est d'intérieur non vide car la fonction $F(u,v) := \exp[(a/2)(u-v)]$ est dans l'intérieur de K_a.

Le membre droit de (2.29) est fini puisque pour toute fonction $G \in B^{\mathbb{R}}_\infty(\mathcal{V})$:

$$|e^{a|u-v|} G(u,u) - G(u,v)| \le (e^{a|u-v|} - 1)(M + c/a). \tag{2.30}$$

En effet,

$$|e^{a|u-v|} G(u,u) - G(u,v)| \le (e^{a|u-v|} - 1)|G(u,u)| + |G(u,u) - G(u,v)|.$$

Comme G est analytique dans $\mathcal{V} \times \mathcal{V}$ et continue en $\overline{\mathcal{V}} \times \overline{\mathcal{V}}$ (ainsi que ses dérivées partielles), les grandeurs

$$M = \sup_{u \in \mathcal{J}} |G(u,u)|, \qquad c = \sup_{u,v \in \mathcal{J}} \left| \frac{\partial G}{\partial v}(u,v) \right|,$$

sont finies. On peut alors utiliser l'inégalité classique $e^x - 1 \ge x$ pour obtenir

$$(e^{a|u-v|} - 1)|G(u,u)| + |G(u,u) - G(u,v)| \le (e^{a|u-v|} - 1)M + c|u-v| \le (e^{a|u-v|} - 1)(M + c/a).$$

Pour établir la positivité de $\underline{\mathbf{H}}_{s,w}$ nous allons montrer que $\underline{\mathbf{H}}_{s,w}$ envoie K_a en $K_{\delta' a}$ avec $\delta' < 1$. En nous servant de la proposition 2.7, propriété (*ii*), nous pouvons fixer un réel positif $\delta < 1$ pour lequel toute branche $h \in \mathcal{H}$ vérifie

$$|h(u) - h(v)| \le \delta |u - v| \quad \text{pour tout } u, v \in \overline{\mathcal{V}}. \tag{2.31}$$

On fixe aussi un réel $\delta' \in]\delta, 1[$ quelconque. Une fonction $F \in K_a$. vérifie, grâce à (2.31), et pour tous $u, v \in \mathcal{V}$.

$$F(h(u), h(v)) \leq e^{a|h(u)-h(v)|} F(h(u), h(u)) \leq e^{a|h(u)-h(v)|} F(h(u), h(u)) \leq e^{a\delta|u-v|} F(h(u), h(u)).$$

Par ailleurs, si $u, v \in \mathcal{J}$ et $F \in K_a$, nous avons

$$\underline{\mathbf{H}}_{s,w}[F](u,v) = \sum_{h\in\mathcal{H}} h'(u)^{s/2}\, h'(v)^{s/2}\, e^{wc(h)} F(h(u), h(v)) \tag{2.32}$$

$$\leq e^{a\delta|u-v|} \sum_{h\in\mathcal{H}} h'(u)^{s/2} h'(v)^{s/2} e^{wc(h)} F(h(u), h(u)), \tag{2.33}$$

or, grâce à la propriété (iv) de la proposition 2.7, on montre que

$$0 \leq \underline{\mathbf{H}}_{s,w}[F](u,v) \leq e^{(a\delta+Ks/2)|u-v|} \underline{\mathbf{H}}_{s,w}[F](u,u) \leq e^{a\delta'|u-v|}, \tag{2.34}$$

et alors $\underline{\mathbf{H}}_{s,w}$ est un opérateur positif de $K_{\delta' a}$. Il ne reste qu'à montrer que $\underline{\mathbf{H}}_{s,w}$ est 1-positif : pour toute fonction $F \in K_a$ non identiquement nulle, il existe $p \in \mathbb{N}$, et $\alpha, \beta > 0$ tels que

$$\underline{\mathbf{H}}^p_{s,w}[F] - \alpha \cdot 1 \in K_a \quad \text{et} \quad \beta \cdot 1 - \underline{\mathbf{H}}_{s,w}[F] \in K_a,$$

ou de façon équivalente, pour tous $u, v \in \mathcal{J}$,

$$\alpha \leq \frac{e^{a|u-v|} \underline{\mathbf{H}}^p_{s,w}[F](u,u) - \underline{\mathbf{H}}^p_{s,w}[F](u,v)}{e^{a|u-v|} - 1} \leq \beta, \tag{2.35}$$

Or, pour tout $p \geq 0$, l'inégalité (2.30- montre l'existence de β, puisque $\underline{\mathbf{H}}^p_{s,w}[F] \in B^{\mathbb{R}}_\infty(\mathcal{V})$ quand $F \in K_a$. Pour prouver l'existence de α, on commence par montrer que si $F \neq 0$, alors il existe $p \in \mathbb{N}$ tel que

$$m_p := \inf_{u\in\mathcal{J}} \underline{\mathbf{H}}^p_{s,w}[F](u,u) > 0. \tag{2.36}$$

En effet, si ce n'était pas le cas, il existerait, pour chaque $p \in \mathbb{N}$, un réel $u_p \in \mathcal{J}$ qui vérifie, pour toute homographie $h \in \mathcal{H}^+$ de hauteur $|h| = p$, l'égalité $F(h(u_p), h(u_p)) = 0$. L'ensemble $\{h(u_p) \ : \ |h| = p, \ p \geq 0\}$ est dense en $\mathcal{J}$ car les intervalles fondamentaux ont des longueurs $\leq \delta^p$ pour $|h| = p$, et ils forment une partition de $\mathcal{J}$. Comme de plus la fonction diagonale $u \mapsto F(u,u)$ est analytique, cela entraîne la nullité de la fonction $u \mapsto F(u,u)$, dans $\mathcal{V}$, et donc, grâce à la définition du cône K_a, la nullité de $F(u,v)$ dans $\mathcal{J} \times \mathcal{J}$. Or, pour chaque $u_0 \in \mathcal{J}$, $F(u_0, v)$ est une fonction analytique dans $\mathcal{V}$, nulle dans $\mathcal{J} \times \mathcal{V}$. En fixant $v_0 \in \mathcal{V}$, $F(u, v_0)$ est également nulle pour $u \in \mathcal{J}$, et donc dans tout $\mathcal{V}$. On conclut que F est la fonction nulle, ce qui contredit notre hypothèse.

Soit p un entier vérifiant (2.36). On choisit $a = sA/(4(\delta - \delta'))$, de manière à ce que (2.34) soit vérifiée. Alors, l'opérateur $\underline{\mathbf{H}}^p_{s,w}$ envoie K_a sur $K_{\delta' a}$ pour chaque $p \geq 1$. Ainsi,

$$\frac{e^{a|u-v|} \underline{\mathbf{H}}^p_{s,w}[F](u,u) - \underline{\mathbf{H}}^p_{s,w}[F](u,v)}{e^{a|u-v|} - 1} \geq \frac{\underline{\mathbf{H}}^p_{s,w}[F](u,u)(e^{a|u-v|} - e^{a\delta'|u-v|})}{e^{a|u-v|} - 1}$$

et en choisissant

$$\alpha := m_p \inf_{(u,v)\in\mathcal{J}\times\mathcal{J}} \frac{e^{a|u-v|} - e^{a\delta'|u-v|}}{e^{a|u-v|} - 1}$$

l'inégalité (2.35) est vérifiée.

Nous avons donc vérifié toutes les hypothèses du théorème de Krasnoselskii. Il existe donc un unique vecteur propre dans l'intérieur de K_a, et la valeur propre associée est simple (de multiplicité algébrique 1) et elle est, en module, strictement supérieure à toutes les autres valeurs propres. La dernière assertion est un cas particulier de ce que nous avons montré dans la section 2.4.4. □

2.5.5 Propriétés spectrales dominantes.

La proposition suivante établit maintenant un lien entre les objets spectraux dominants des opérateurs $\underline{\mathbf{H}}_{s,w}$ et $\underline{\mathbf{H}}_{s,w}$.

Proposition 2.10. *Lorsque s et w sont réels, les propriétés spectrales dominantes de l'opérateur à deux variables $\underline{\mathbf{H}}_{s,w}$ sont reliées à celles de l'opérateur à une variable $\mathbf{H}_{s,w}$. Plus précisément,*

(i) Les deux opérateurs ont la même valeur propre dominante $\lambda(s,w)$,

(ii) La restriction de la fonction propre dominante $\underline{\psi}_{s,w}$ de $\underline{\mathbf{H}}_{s,w}$ à la diagonale de $\mathcal{V}\times\mathcal{V}$ coincide avec la fonction propre dominante $\psi_{s,w}$ de $\mathbf{H}_{s,w}$.

(iii) La fonction propre $\underline{\psi}_{s,w}$ s'exprime en fonction de $\psi_{s,w}$:

$$\underline{\psi}_{s,w}(u,v) = \int_0^1 \beta_{s/2,s/2}(t)\psi_{s,w}(u+(v-u)t)dt, \tag{2.37}$$

où $\beta_{a,b}$ est la densité β classique,

$$\beta_{a,b}(y) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} t^{a-1}(1-t)^{b-1} \qquad a>0,\, b>0.$$

De plus, $\underline{\psi}_{s,w}(z,\bar{z})$ est un réel strictement positif.

(iv) Le vecteur propre dominant $\underline{\nu}_{s,w}$ de l'opérateur adjoint $\underline{\mathbf{H}}^*_{s,w}$ est égal au vecteur propre dominant $\nu_{s,w}$ de l'opérateur $\mathbf{H}^*_{s,w}$ dans le sens suivant : pour une fonction F de $B_\infty(\mathcal{V})$ dont l'application diagonale est f, on a

$$\underline{\nu}_{s,w}[F] = \nu_{s,w}[f].$$

Démonstration. Nous rappelons le lien essentiel entre les deux opérateurs $\mathbf{H}_{s,w}$ et $\underline{\mathbf{H}}_{s,w}$ prouvé dans la proposition 2.2. Pour toute fonction F, d'application diagonale f définie par $f(u) := F(u,u)$, on a, pour tout $k \geq 1$

$$\underline{\mathbf{H}}^k_{s,w}[F](u,u) = \mathbf{H}^k_{s,w}[f](u). \tag{2.38}$$

Par ailleurs, la section précédente permet d'appliquer la décomposition (2.21) aux deux opérateurs $\mathbf{H}_{s,w}$ et $\underline{\mathbf{H}}_{s,w}$, et comme les fonctions propres $\psi_{s,w}$ et $\underline{\psi}_{s,w}$ sont strictement positives respectivement sur $\mathcal{J}$ et sur $\mathcal{J}\times\mathcal{J}$, on a

$$\underline{\mathbf{H}}^k_{s,w}[F](x,x) = \underline{\lambda}(s,w)^k \underline{\psi}_{s,w}(x,x)\underline{\nu}_{s,w}[F]\left[1+O(\underline{\rho}^k)\right], \tag{2.39}$$

$$\mathbf{H}^k_{s,w}[f](x) = \lambda(s,w)^k \psi_{s,w}(x)\nu_{s,w}[f]\left[1+O(\rho^k)\right],$$

où $\underline{\rho}, \rho$ sont liés au saut spectral des opérateurs. On en déduit d'abord, avec (2.38), l'égalité

$$\underline{\lambda}(s,w) = \lim_{k\to\infty}\left(\underline{\mathbf{H}}^k_{s,w}[1](x,x)\right)^{1/k} = \lim_{k\to\infty}\left(\mathbf{H}^k_{s,w}[1](x)\right)^{1/k} = \lambda(s,w),$$

puis, avec l'égalité $\underline{\nu}_{s,w}[1] = \nu_{s,w}[1] = 1$

$$\underline{\psi}_{s,w}(x,x) = \lim_{k\to\infty}\frac{\underline{\mathbf{H}}^k_{s,w}[1](x,x)}{\lambda(s,w)^k} = \lim_{k\to\infty}\frac{\mathbf{H}^k_{s,w}[1](x)}{\lambda(s,w)^k} = \psi_{s,w}(x).$$

Enfin,

$$\underline{\nu}_{s,w}[F] = \lim_{k\to\infty} \frac{\underline{\mathbf{H}}^k_{s,w}[1](x,x)}{\lambda(s,w)^k \underline{\psi}_{s,w}(x,x)} = \lim_{k\to\infty} \frac{\mathbf{H}^k_{s,w}[1](x)}{\lambda(s,w)^k \psi_{s,w}(x)} = \nu_{s,w}[f].$$

Il reste donc à prouver (*iii*). Nous suivons la preuve de Vallée ([77, Théorème 5]). Dans l'intégrale (2.37), nous effectuons le changement de variable $z = u + (v-u)t$ en obtenant

$$\underline{\psi}_{s,w}(u,v) = \frac{\Gamma(s)}{\Gamma(s/2)\Gamma(s/2)} \int_\gamma \psi_{s,w}(z) \frac{(z-u)^{s/2-1}(v-z)^{s/2-1}}{(v-u)^{s-1}} dz, \tag{2.40}$$

où γ est le segment de droite joignant u et v. Puisque la fonction à intégrer est holomorphe, nous pouvons remplacer γ par un chemin simple quelconque reliant u à v. Vérifions maintenant que $\underline{\psi} \equiv \underline{\psi}_{s,w}$ définit à partir de $\psi \equiv \psi_{s,w}$ un vecteur propre de $\underline{\mathbf{H}}_{s,w}$. Commençons par évaluer $\underline{\mathbf{H}}_{s,w,[h]}[\underline{\psi}]$, où h est une branche primaire quelconque. Cette expression fait intervenir le prolongement analytique $\tilde{h}$ de $|h'|$, sous la forme

$$e^{wc(h)} \frac{\tilde{h}(u)|^{s/2}\, \tilde{h}(v)^{s/2}}{[h(v)-h(u)]^{s-1}} \int_\delta \psi(z)[z-h(u)]^{s/2-1}[h(v)-z]^{s/2-1} dz, \tag{2.41}$$

où δ est un chemin simple, reliant $h(u)$ à $h(v)$. Comme l'intégrale en jeu ne dépend que des extrémités de δ, nous pouvons choisir pour δ l'image par h d'un chemin γ reliant simplement u à v. Le changement de variable $z = h(t)$ dans (2.41), ainsi que les relations

$$dz = h'(t)dt = \epsilon(h)\tilde{h}(t)dt; \qquad h(a) - h(b) = \epsilon(h)[\tilde{h}(a)\cdot\tilde{h}(b)]^{1/2}(a-b),$$

(où $\epsilon(h)$ est le signe de h'), valables pour tout a et b, permettent de réécrire (2.41) sous la forme

$$\frac{1}{(v-u)^{s-1}} \int_\gamma e^{wc(h)}\, \tilde{h}(t)^s\, \psi\circ h(t)\, (t-u)^{s/2-1}\, (v-t)^{s/2-1} dt$$
$$= \frac{1}{(v-u)^{s-1}} \int_\gamma \mathbf{H}_{s,w,[h]}[\psi](t)(t-u)^{s/2-1}(v-t)^{s/2-1} dt. \tag{2.42}$$

Pour $\underline{\psi}$ définie à partir de ψ par (2.40), nous avons donc établi la relation

$$\underline{\mathbf{H}}_{s,w,[h]}[\underline{\psi}](u,v) = \frac{\Gamma(s)}{\Gamma(s/2)\Gamma(s/2)} \int_\gamma \mathbf{H}_{s,w,[h]}[\psi(t) \frac{(t-u)^{s/2-1}(v-t)^{s/2-1}}{(v-u)^{s-1}} dt$$

valable pour toute branche primaire h. En sommant sur toutes les branches primaires $h \in \mathcal{H}$, nous obtenons

$$\underline{\mathbf{H}}_{s,w}[\underline{\psi}](u,v) = \frac{\Gamma(s)}{\Gamma(s/2)\Gamma(s/2)} \int_\gamma \mathbf{H}_{s,w}[\psi](t) \frac{(t-u)^{s/2-1}(v-t)^{s/2-1}}{(v-u)^{s-1}} dt.$$

Ainsi, si ψ est un vecteur propre de $\mathbf{H}_{s,w}$, avec la valeur propre λ, alors $\underline{\psi}$ est un vecteur propre de $\underline{\mathbf{H}}_{s,w}$ relatif à λ, comme nous voulions montrer. □

Proposition 2.11. *Soit c un coût élémentaire à croissance modérée, non identiquement nul et à valeurs dans $\mathbb{N}$. Alors*

(i) la valeur propre dominante $\lambda(s,w)$ ainsi que ses dérivées partielles possèdent des expressions explicites pour $(s,w) \in \mathfrak{H}_{(c)}$

$$\begin{aligned} \lambda(s,w) &= \sum_{h\in\mathcal{H}} \exp[wc(h)] \cdot I_h(s,w), && (2.43)\\ \lambda'_w(s,w) &= \sum_{h\in\mathcal{H}} c(h)\exp[wc(h)] \cdot I_h(s,w), && (2.44)\\ \lambda'_s(s,w) &= \sum_{h\in\mathcal{H})} \exp[wc(h)] \cdot J_h(s,w). && (2.45) \end{aligned}$$

où les intégrales $I_{m,\epsilon}(s,w)$ et $J_{m,\epsilon}(s,w)$, données par

$$\begin{aligned} I_h(s,w) &:= \int_{\tilde{\mathcal{I}}} |h'(t)|^s \cdot f_{s,w}\circ h(t) d\nu_{s,w},\\ J_h(s,w) &:= \int_{\tilde{\mathcal{I}}} \log|h'(t)| \cdot |h'(t)|^s \cdot f_{s,w}\circ h(t) d\nu_{s,w}, \end{aligned} \qquad (2.46)$$

font intervenir la fonction propre dominante $\psi_{s,w}$ et le vecteur propre dominant $\nu_{s,w}$ de l'opérateur adjoint $\mathbf{H}^*_{s,w}$.

(ii) La fonction $(s,w) \mapsto \lambda(s,w)$ est analytique pour tout (s,w) à l'intérieur de $\mathfrak{H}_{(c)}$. Lorsque s et w sont réels, elle est strictement décroissante en s et strictement croissante en w. Quand (s,w) tend vers un point de $\mathcal{F}_{(c)}$, la valeur propre $\lambda(s,w)$ tend vers $+\infty$. Quand w tend vers $-\infty$ à s fixé, alors $\lambda(s,w)$ tend vers 0. Quand s tend vers $+\infty$à w fixé, alors $\lambda(s,w)$ tend vers 0.

Preuve. On considère l'identité

$$\mathbf{H}_{s,w}[\psi_{s,w}] = \lambda(s,w)\psi_{s,w} \qquad (2.47)$$

Toutes les expressions qui apparaissent dans (2.47) sont analytiques par rapport à s et à w. On peut alors dériver (2.47) d'abord par rapport à s, puis par rapport à w pour obtenir

$$\begin{aligned} \frac{d\mathbf{H}_{s,w}}{ds}[\psi_{s,w}] + \mathbf{H}_{s,w}[\frac{d\psi_{s,w}}{ds}] &= \lambda'_s(s,w)\psi_{s,w} + \lambda(s,w)\frac{d\psi_{s,w}}{ds} && (2.48)\\ \frac{d\mathbf{H}_{s,w}}{dw}[\psi_{s,w}] + \mathbf{H}_{s,w}[\frac{d\psi_{s,w}}{dw}] &= \lambda'_w(s,w)\psi_{s,w} + \lambda(s,w)\frac{d\psi_{s,w}}{dw}. && (2.49) \end{aligned}$$

En intégrant (2.47), (2.48) et (2.49) par rapport à la mesure propre $\nu_{s,w}$ de l'opérateur adjoint $\mathbf{H}^*_{s,w}$, on obtient respectivement (2.43), (2.44) et (2.45). Par définition, $\nu_{s,w}$ satisfait l'égalité

$$\int_{\tilde{\mathcal{I}}} \mathbf{H}_{s,w}[g](t) d\nu_{s,w}(t) = \lambda(s,w)\int_{\tilde{\mathcal{I}}} g(t) d\nu_{s,w}(t) \quad \text{pour tout } g \in B_\infty(\mathcal{V}). \qquad (2.50)$$

En posant $g = \psi_{s,w}$ dans (2.50), nous obtenons l'expression correspondante à intégrer (2.47). La fonction $\psi_{s,w}$ ayant une intégrale égale à 1 par rapport à $\nu_{s,w}$, il s'en suit que

$$\lambda(s,w) = \int_{\tilde{\mathcal{I}}} \mathbf{H}_{s,w}[\psi_{s,w}](t) d\nu_{s,w}(t) = \sum_{h\in\mathcal{H}} \exp[wc(h)] \cdot I_h(s,w),$$

où dans la dernière égalité nous avons interverti série et intégrale.

Par ailleurs, en prenant $g = \frac{d}{dw}\psi_{s,w}$ en (2.50), et en intégrant (2.48), on obtient

$$\lambda'_s(s,w) = \lambda'_s(s,w)\int_{\tilde{\mathcal{I}}} \psi_{s,w}(t)d\nu_{s,w}(t) = \int_{\tilde{\mathcal{I}}} \mathbf{H}_{s,w}[\psi_{s,w}](t)d\nu_{s,w}(t),$$

où dans la dernière égalité nous avons encore interverti somme et intégrale. Le cas de (2.45) est analogue. Ainsi, nous avons établi (i).

Pour prouver (ii), on fait usage de (i). Lorsque (s,w) sont réels, les fonctions $f_{s,w}$ et $\nu_{s,w}$ sont strictement positives. Il suffit alors d'observer que l'intégrale $I_h(s,w)$ est l'intégrale d'un produit de fonctions strictement positives, et que $J_h(s,w)$ est l'intégrale du produit de la fonction strictement négative $\log|h'(t)|$ et de fonctions strictement positives. Ainsi, pour tout $h \in \mathcal{H}$, on a les deux inégalités strictes : $I_h(s,w) > 0$ et $J_h(s,w) < 0$. Cela entraîne l'inégalité stricte $\lambda'_s(s,w) < 0$ pour tous (s,w) réels, et, pourvu que le coût élémentaire c prenne des valeurs dans $\mathbb{N}$ sans être identiquement nul, n l'inégalité stricte $\lambda'_w(s,w) > 0$, comme voulu. □

2.5.6 Objets spectraux dominants pour $(s,w) = (1,0)$.

Proposition 2.12. *La densité invariante de l'opérateur* $\mathbf{H}$ *associé au système* Euclide-plié*, est donnée par*

$$\psi(x) = \frac{1}{\log\phi}\left(\frac{1}{\phi+x} + \frac{1}{\phi^2-x}\right).$$

La densité invariante de l'opérateur $\underline{\mathbf{H}}$ *du système* Gauss-interne*, est reliée à la précédente et s'écrit,*

$$\underline{\psi}(u,v) = \frac{1}{\log\phi}\frac{1}{v-u}\left(\log\frac{\phi+v}{\phi+u} - \log\frac{\phi^2-v}{\phi^2-u}\right) \tag{2.51}$$

pour $u \neq v$ *et* $\underline{\psi}(u,u) = \psi(u)$*. L'application linéaire définissant le projecteur* $\underline{\mathbf{P}}$ *vérifie*

$$J[F] = \int_{\tilde{\mathcal{I}}} F(x,x)dx. \tag{2.52}$$

Preuve. Nous commençons par prouver que $\psi(x)$ est la densité invariante de $\mathbf{H}$. Comme

$$\mathbf{H}[\psi](x) = \sum_{(m,\epsilon)\geq(2,1)} \frac{1}{(m+\epsilon x)^2}\cdot\psi\left(\frac{1}{m+\epsilon x}\right), \tag{2.53}$$

nous cherchons à évaluer les expressions de la forme $x^{-2}\psi(x^{-1})$. Comme le nombre d'or ϕ vérifie $\phi^2-\phi-1=0$, on a aussi

$$\phi^2+\phi = \phi^3, \qquad -\frac{1}{\phi^2} = \frac{1}{\phi} - 1 \tag{2.54}$$

Alors,

$$\log\phi\frac{1}{x^2}\psi\left(\frac{1}{x}\right) = \frac{1}{x^2}\left[\frac{1}{\phi+\frac{1}{x}} + \frac{1}{\phi^2-\frac{1}{x}}\right] = \frac{1}{x}\left[\frac{1}{\phi x+1} + \frac{1}{\phi^2 x-1}\right] = \frac{\phi^3}{(\phi x+1)(\phi^2 x-1)}$$

soit encore

$$\log\phi\frac{1}{x^2}\psi\left(\frac{1}{x}\right) = \frac{1}{(x+\phi^{-1})(x-\phi^{-2})} = \frac{1}{x-\phi^{-2}} - \frac{1}{x+\phi^{-1}} = \frac{1}{x-1+\phi^{-1}} - \frac{1}{x+\phi^{-1}}.$$

ce qui montre que la série (2.53) est téléscopique. En plus, puisque c'est une série absolument convergente, on peut évaluer d'abord les termes avec $\epsilon = 1$, puis ceux avec $\epsilon = -1$ pour obtenir $\psi(x)$, après s'être servi à nouveau des relations (2.54).

La densité invariante $\underline{\psi}$ de l'opérateur $\underline{\mathbf{H}}$ de l'algorithme GAUSS-INTERNE se détermine à l'aide de la proposition 2.10. La densité $\underline{\psi}$ vérifie

$$\underline{\psi}(u,v) = \int_0^1 \psi(u + (v-u)x)dx.$$

On a donc $\underline{\psi}(u,u) = \psi(u)$ et (2.51) lorsque $u \neq v$.

La proposition 2.10 relie J à son analogue en une variable ν, par la relation $J[F] = \nu[f]$ où f est la diagonale de F, définie par $f(x) := F(x,x)$ pour $x \in \tilde{\mathcal{I}}$. La forme linéaire ν est explicite, de la forme $\nu[f] = \int_{\mathcal{I}} f(x)dx$. En effet, puisque $\mathbf{H}$ est un transformateur de densité, en utilisant la décomposition donnée dans la section 2.4.4, on déduit

$$\int_{\tilde{\mathcal{I}}} \mathbf{H}^k[f](x)dx = \int_{\tilde{\mathcal{I}}} f(x)dx = \nu[f] \cdot \int_{\tilde{\mathcal{I}}} \psi(x)dx + \int_{\tilde{\mathcal{I}}} O(\varrho^k).$$

Par ailleurs ψ est normalisée de sorte que son intégrale vaut 1. Il ne reste qu'à faire tendre k à l'infini pour obtenir (2.52). □

2.5.7 Entropie

L'entropie modélise le concept de *surprise espérée*, et elle se comprend dans le contexte des *sources dynamiques* (voir [79, 15]). Dans le contexte de cette thèse, elle intervient juste en tant que constante structurelle dans la décomposition de l'opérateur quasi-inverse.

Définition 2.5. *L'entropie associée au système* EUCLIDE-PLIÉ [5] *s'exprime en fonction du décalage $\tilde{V}$, et de la densité invariante du système.*

$$h(\mathcal{E}) = \int_{\tilde{\mathcal{I}}} \log|\tilde{V}'(x)| \cdot \psi(x)dx. \tag{2.55}$$

La proposition suivante lie l'entropie avec la dérivée par rapport à s de la valeur propre dominante.

Proposition 2.13. *L'entropie du système dynamique* EUCLIDE-PLIÉ *s'exprime de deux manières différentes en fonction de la dérivée par rapport à s de la valeur propre dominante, ou en fonction de l'opérateur transformateur de densité $\mathbf{H}$ et de la densité invariante $\underline{\psi}_{s,w}$,*

$$h(\mathcal{E}) = -\lambda_s'(1,0), \qquad \textit{ou aussi} \quad h(\mathcal{E}) = -J\left[(\Delta\underline{\mathbf{H}})[\underline{\psi}]\right].$$

Preuve. En effet, d'après (2.45), nous avons, avec un changement de variable $u = h(x)$, pour lequel $\tilde{V}'(u) = 1/|h'(x)|$,

$$\lambda_s'(1,0) = \sum_{h\in\mathcal{H}} \int_{\tilde{\mathcal{I}}} |h'(x)| \cdot \log|h'(x)| \cdot \psi \circ h(x)dx,$$

$$\lambda_s'(1,0) = \sum_{h\in\mathcal{H}} \int_{h(\tilde{\mathcal{I}})} \log|\tilde{V}'(u)| \cdot \psi(u)du = \int_{\tilde{\mathcal{I}}} \log|\tilde{V}'(u)| \cdot \psi(u)du,$$

d'où le résultat. □

5. Cette formule, due à Rohlin, n'est pas générale : elle dépend du fait que le déterminant des branches inverses de EUCLIDE-PLIÉ est constant

2.5.8 Espérance limite d'un coût élémentaire.

Dans cette section nous définissons une autre constante structurelle qui interviendra, en plus de l'entropie, dans l'expression de l'opérateur quasi-inverse. L'espérance limite du coût c est l'espérance du coût lors d'une itération du système, lorsque le nombre d'itérations déjà effectuées tend vers l'infini.

Définition 2.6. *Considérons le système* Euclide-plié*, ainsi qu'un coût élémentaire de croissance modérée c. On définit l'*espérance limite *de c par*

$$\mathbb{E}(c) = \sum_{h\in\mathcal{H}} c(h) \cdot \int_{h(\tilde{\mathcal{I}})} \psi(x)dx. \tag{2.56}$$

Tout comme l'entropie, l'espérance canonique s'exprime en fonction de la valeur propre dominante.

Proposition 2.14. *L'espérance limite d'un coût c élémentaire modéré s'exprime en fonction de la valeur propre dominante, ou en fonction de l'opérateur transformateur de densité,*

$$\mathbb{E}(c) = \lambda'_w(1,0), \qquad \textit{ou aussi} \quad \mathbb{E}(c) = J\big[(W_{(c)}\underline{\mathbf{H}})[\underline{\psi}]\big].$$

Preuve. La preuve est analogue à celle de la proposition 2.13. D'après (2.44), nous avons

$$\lambda'_w(1,0) = \sum_{h\in\mathcal{H}} c(h) \cdot \int_{\tilde{\mathcal{I}}} |h'(x)| \cdot \psi\circ h(x)dx = \sum_{h\in\mathcal{H}} c(h) \cdot \int_{h(\tilde{\mathcal{I}})} \psi(u)du = \mathbb{E}(c),$$

comme on voulait montrer. □

2.5.9 Propriétés de maximum de la valeur propre dominante

Nous allons prouver ici une propriété de maximalité de la valeur propre, qui sera utile lors de l'étude des coûts additifs.

Proposition 2.15. *Soit c un coût élémentaire de croissance modérée. Soit $(s,w) \in \mathfrak{H}_{(c)}$. Alors*

(i) Toute valeur propre λ de l'opérateur $\mathbf{H}_{s,w}$ a un module $|\lambda|$ qui vérifie

$$|\lambda(s,w)| \le \lambda(\sigma,\tau), \qquad \textit{pour } (\Re(s),\Re(w)) = (\sigma,\tau).$$

(i) Lorsque $s=\sigma$ est réel, l'égalité $|\lambda| = \lambda(\sigma,\tau)$ n'est vérifiée que si $\Im(w) = k\cdot 2\pi/d$, où d est le pgcd des valeurs du coût élémentaire c.

(iii) Si le coût c est primitif, (i.e., le pgcd des valeurs de c est égal à 1), alors le rayon spectral de $\underline{\mathbf{H}}_{s,w}$ est strictement inférieur à $\lambda(\sigma,\tau$.

Preuve. Nous suivons la preuve donnée en [79]. Soit $(s,w) \in \mathfrak{H}_{(c)}$, et posons $\sigma = \Re s, \tau = \Re(w)$. Soit λ une valeur propre de $\mathbf{H}_{s,w}$ et f un vecteur propre associé. De même, soit $\lambda(\sigma,\tau)$ la valeur propre dominante de l'opérateur $\mathbf{H}_{\sigma,\tau}$ et $f_{\sigma,\tau}$ un vecteur propre dominant, strictement positif. On peut supposer (quitte à multiplier les fonctions propres par des scalaires adéquats) que la fonction

$$\eta(x) := \frac{f(x)}{f_{\sigma,\tau}(x)},$$

est de module au plus 1 dans $\tilde{\mathcal{I}}$ et qu'elle atteint le module 1 au point x_0. On a toujours

$$|\lambda f(x_0)| = |\mathbf{H}_{s,w}[f](x_0)| = \left| \sum_{h\in\mathcal{H}} |h'(x_0)|^s e^{wc(h)} f\circ h(x_0) \right|,$$

et

$$\left| \sum_{h\in\mathcal{H}} |h'(x_0)|^s e^{wc(h)} f\circ h(x_0) \right| \leq \sum_{h\in\mathcal{H}} |h'(x_0)|^\sigma e^{\tau c(h)} |f\circ h(x_0)|,$$

et puis finalement,

$$\sum_{h\in\mathcal{H}} |h'(x_0)|^\sigma e^{\tau c(h)} |f\circ h(x_0)| \leq \sum_{h\in\mathcal{H}} |h'(x_0)|^\sigma e^{\tau c(h)} f_{\sigma,\tau}\circ h(x_0) = \lambda(\sigma,\tau) f_{\sigma,\tau}(x_0).$$

Grâce au choix de x_0, nous concluons que $|\lambda(s,w)| \leq \lambda(\sigma,\tau)$, comme voulu.

Maintenant, si l'égalité $|\lambda| = \lambda(\sigma,\tau)$ a lieu, alors toutes les ingalités deviennent des égalités

$$|\lambda|\,|f(x_0)| = \left| \sum_{|h|=1} |h'(x_0)|^\sigma e^{wc(h)} f\circ h(x_0) \right| =$$
$$\sum_{|h|=1} |h'(x_0)|^\sigma e^{\tau c(h)}\,|f\circ h(x_0)| = \sum_{|h|=1} |h'(x_0)|^\sigma e^{\tau c(h)} f_{\sigma,\tau}\circ h(x_0) = \lambda(\sigma,\tau) f_{\sigma,\tau}(x_0). \quad (2.57)$$

Grâce au choix de x_0, l'égalité

$$|f\circ h(x_0)| = f_{\sigma,\tau}\circ h(x_0) \quad (2.58)$$

est vérifiée pour tout branche primaire h. C'est aussi vrai pour des branches inverses arbitraires, car on peut écrire une suite d'égalités analogues à (2.57) en itérant l'opérateur $\mathbf{H}_{s,w}$. Puisque l'ensemble des $h(x_0)$ est dense en $\tilde{\mathcal{I}}$, l'égalité (2.58) est vérifiée partout dans $\tilde{\mathcal{I}}$, et donc que ηest de module 1 partout dans $\tilde{\mathcal{I}}$. Cela entraîne, avec (2.57), que

$$\left| \sum_{|h|=1} |h'(x)|^\sigma e^{wc(h)} f\circ h(x) \right| = \sum_{|h|=1} |h'(x)|^\sigma e^{\tau c(h)}\,|f\circ h(x)|\,, \qquad \text{pour tout } \; x\in\tilde{\mathcal{I}}.$$

Ainsi, posant $a_h(x) := |h'(x)|^s e^{wc(h)} f\circ h(x)$, l'égalité précédente s'écrit$|\sum_h a_h(x)| = \sum |a_h(x)|$. Il existe donc une fonction θ (de module 1), indépendante de la branche h, pour laquelle $a_h(x) = \theta(x)|a_h(x)|$ pour toute branche h. Dans notre cadre, cette fonction $\theta(x)$ coincide avec $\eta(x)$ et l'égalité

$$e^{i\Im(w)c(h)}\eta\circ h(x) = \eta(x)$$

est vraie pour toute branche primaire h. En évaluant cette égalité en l'unique point fixe x_h de chaque branche h, on a

$$e^{i\Im(w)c(h)}\eta(x_h) = \eta(x_h) \qquad \text{et donc} \qquad e^{i\Im(w)c(h)} = 1,$$

puisque $\eta(x_h)\neq 0$. Comme cette dernière égalité est valide pour toute branche primaire h, on a donc, si d est le pgcd des $c(h)$ pour $h\in\mathcal{H}$,

$$e^{i\Im(w)d} = 1$$

Il faut donc que $\Im(w) = 2k\pi/d \mod 2\pi$ pour un certain $k\in\mathbb{Z}$. Dans ce cas-là, on remarque que l'opérateur $\mathbf{H}_{s,w}$ coincide avec $\mathbf{H}_{s,\tau}$, et donc que λ est égale à $\lambda(s,\tau)$. Cela achève la preuve. □

2.6 Le quasi-inverse.

Nous avons vu que tous les objets essentiels à nos analyses font intervenir les quasi-inverses, et des quasi-inverses généralisés. C'est pourquoi nous leur consacrons cette section, qui contient les théorèmes techniques les plus importants de la partie II, qui permettront en particulier de prouver les théorèmes **A** et **B**.

2.6.1 Région d'analyticité.

Tout d'abord, nous décrivons la région où le quasi-inverse $(\underline{I} - \underline{\mathbf{H}}_{s,w})^{-1}$ de l'opérateur $\underline{\mathbf{H}}_{s,w}$ est analytique. L 'ensemble des couples réels $(s,w) \in \mathfrak{H}_{(c)}$ qui vérifient $\lambda(s,w) = 1$ est une courbe bien définie Λ située à l'intérieur de $\mathfrak{H}_{(c)}$. Cela est dû aux bonnes propriétés de la fonction valeur propre dominante. En effet, pour un s fixé, la valeur propre $\lambda(s,w)$ est bien définie pour $w \in]-\infty, (1/p_c)(s-1/2)]$. La fonction $w \to \lambda(s,w)$ est strictement croissante. De plus, $\lambda(s,w)$ tend vers 0 quand w tend vers $-\infty$ et $\lambda(s,w)$ tend vers l'infini quand w tend vers le point frontière $(1/p_c)(s-1/2)$. Il existe donc une unique valeur $w(s)$ de w pour laquelle on a l'égalité $\lambda(s,w(s)) = 1$. Puisque $\lambda(1,0) = 1$, cette valeur $w(s)$ est strictement positive quand $s > 1$. Donc, pour tout couple (s,w) strictement à gauche de la courbe Λ, la valeur propre $\lambda(s,w)$ vérifie $\lambda(s,w) < 1$, et le quasi-inverse $(\underline{I} - \underline{\mathbf{H}}_{s,w})^{-1}$ est bien défini dans ce domaine et y est analytique. C'est également vrai pour tout couple (s,w) non nécessairement réel pour lequel le couple $(\Re s, \Re w)$ associé est a gauche de la courbe Λ.

2.6.2 Pôles du quasi-inverse.

Le résultat suivant décrit ce qui se passe sur la courbe Λ.

Proposition 2.16 (Pôles du quasi-inverse)**.** *Le quasi-inverse $(\underline{I} - \underline{\mathbf{H}}_{s,w})^{-1}$ de l'operateur $\underline{\mathbf{H}}_{s,w}$ vérifie les propriétés suivantes :*

(*i*) *Supposons que c est un coût à croissance modéree, et fixons un s vérifiant $s > 1$. Alors, il existe un unique $w > 0$ pour lequel $\lambda(s,w) = 1$. La fonction qui à w associe le quasi-inverse $(\underline{I} - \underline{\mathbf{H}}_{s,w})^{-1}$ de l'opérateur $\underline{\mathbf{H}}_{s,w}$ possède un pôle au point $w = w(s)$ pour lequel $(s,w) \in \mathfrak{H}_{(c)}$ et $\lambda(s,w) = 1$, et, pour w proche de $w(s)$, on a :*

$$(I - \underline{\mathbf{H}}_{s,w})^{-1}[F](u,v) \sim \frac{1}{w - w(s)} \frac{1}{\lambda'_w(s,w(s))} \cdot \underline{\mathbf{P}}_{s,w(s)}[F](u,v) \tag{2.59}$$

pour toute fonction $F \in B_\infty(\mathcal{V})$ et $(u,v) \in \overline{\mathcal{V}} \times \overline{\mathcal{V}}$. Plus précisément, pour toute fonction F dont la diagonale f est strictement positive, et pour tout $z \in \overline{\mathcal{V}}$. et pour w proche de $w(s)$, on a :

$$(I - \underline{\mathbf{H}}_{s,w})^{-1}[F](z,\bar{z}) = \frac{1}{w - w(s)} \frac{1}{\lambda'_w(s,w(s))} \cdot \underline{\mathbf{P}}_{s,w(s)}[F](z,\bar{z}) \Big[1 + (w - w(s)) R_s(w,\bar{z}) \Big],$$

où R_s est une fonction bornée quand $z \in \overline{\mathcal{V}}$ et w proche de $w(s)$.

(*ii*) *Fixons $w = 0$. Alors $\lambda(1) = 1$ et, pour r proche de -1, on a :*

$$(I - \underline{\mathbf{H}}_{2+r})^{-1}[F](u,v) \sim \frac{1}{r+1} \frac{1}{h(\mathcal{E})} \cdot J[F]\, \underline{\psi}(u,v) \tag{2.60}$$

pour toute fonction $F \in B_\infty(\mathcal{V})$ et $(u,v) \in \overline{\mathcal{V}} \times \overline{\mathcal{V}}$, et où $h(\mathcal{E}) \approx 3,41831$ est l'entropie du système Euclide-plié*, définie en (2.55), $\mathbb{E}(c)$ est l'espérance limite du coût c, définie en*

(2.56), $J[F]$ est le vecteur propre dominant de l'opérateur adjoint de $\mathbf{H}_s$, évalué en F, et $\underline{\psi}$ est la densité invariante de l'opérateur transformateur de densité $\underline{\mathbf{H}}_1$. Plus précisément, pour toute fonction F dont la diagonale f est strictement positive, et pour tout $z \in \overline{\mathcal{V}}$. et pour r proche de -1, on a :

$$(I - \underline{\mathbf{H}}_{2+r})^{-1}[F](z,\bar{z}) = \frac{1}{r+1}\frac{1}{h(\mathcal{E})} \cdot J[F]\,\underline{\psi}(z,\bar{z})\left[1 + (r+1)R(r,z,\bar{z})\right]$$

où R est une fonction bornée quand $z \in \overline{\mathcal{V}}$ et r proche de -1.

Démonstration. Preuve de (i). Fixons $s \in \mathbb{R}$ et considérons le point $w(s)$ de la courbe Λ d'abscisse s. Puisque $\lambda(1,0) = 1$, cette valeur $w(s)$ est strictement positive quand $s > 1$. Alors, la décomposition de l'opérateur quasi-inverse induite par la décomposition spectrale de $\mathbf{H}_{s,w}$,

$$(\underline{I} - \underline{\mathbf{H}}_{s,w})^{-1} = \frac{\lambda(s,w)}{1-\lambda(s,w)}\underline{\mathbf{P}}_{s,w} + (\underline{I} - \underline{\mathbf{N}}_{s,w})^{-1},$$

montre que $w \mapsto (\underline{I} - \underline{\mathbf{H}}_{s,w})^{-1}$ possède un pôle en $w = w(s)$. En effet, le théorème de perturbation (voir proposition 2.6) montre qu'il existe un voisinage complexe de $w = w(s)$, pour lequel tous les objets de la décomposition spectrale $\lambda(s,w)$, $\underline{\mathbf{P}}_{s,w}$, $\underline{\mathbf{N}}_{s,w}$ définissent des fonctions analytiques de w (à s fixé). On observe de plus que le quasi-inverse $(\underline{I} - \underline{\mathbf{N}}_{s,w})^{-1}$ est lui-même analytique, puisque son rayon spectral est strictement inférieur à 1. Par ailleurs, au voisinage de $w = w(s)$, on a

$$1 - \lambda(s,w) \sim -\lambda'_w(s, w(s)\,(w - w(s)).$$

Cela montre le résultat annoncé. La preuve est du même style pour (ii). □

2.6.3 Extension de la méromorphie du quasi-inverse

Dans le cas d'un coût c entier, on a l'égalité des deux opérateurs $\underline{\mathbf{H}}_{s,w}$ et $\underline{\mathbf{H}}_{s,w'}$ aussitôt que $w = w'$ modulo $2i\pi$. Il suffit donc de travailler avec les complexes w pour lesquels $\Im w \in [-\pi, +\pi]$. On pourra aussi travailler par la suite avec le paramètre $u = e^w$.

Proposition 2.17. *Pour $s > 1$ fixé, on considère l'unique valeur $w(s)$ de w pour laquelle $\lambda(s,w) = 1$. Le quasi-inverse de l'opérateur $\underline{\mathbf{H}}_{s,w}$ vérifie les propriétés suivantes*

- (i) *L'application $w \mapsto (\underline{I} - \underline{\mathbf{H}}_{s,w})^{-1}$ définit une fonction méromorphe de la variable w, dans une bande $\Re w \leq w(s) + \rho(s)$, avec $\rho(s) > 0$.*
- (ii) *Si de plus, le coût c est primitif (le pgcd des valeurs de c est 1), alors $w(s)$ est le seul pôle de $w \mapsto\mapsto (\underline{I} - \underline{\mathbf{H}}_{s,w})^{-1}$ dans la bande $\Re w \leq w(s) + \rho(s)$.*
- (iii) *Si le coût n'est pas primitif, de pgcd d, alors les seuls pôles de $w \mapsto\mapsto (\underline{I} - \underline{\mathbf{H}}_{s,w})^{-1}$ dans la bande $\Re w \leq w(s) + \rho(s)$ sont les points $w_k := w(s) + 2ik\pi/d$*

Preuve. Grâce à la décomposition spectrale, l'opérateur $(\underline{I} - \underline{\mathbf{H}}_{s,w})^{-1}$ s'écrit sous la forme

$$(\underline{I} - \underline{\mathbf{H}}_{s,w})^{-1} = \frac{\lambda(s,w)}{1-\lambda(s,w)}\underline{\mathbf{P}}_{s,w} + (\underline{I} - \underline{\mathbf{N}}_{s,w})^{-1}, \tag{2.61}$$

Dans un voisinage de $w(s)$, les principes de la perturbation (voir la proposition 2.6) s'appliquent et tous les objets qui interviennent dans la décomposition spectrale sont analytiques par rapport à w. Il existe donc un voisinage complexe $\mathcal{W}$ de $w(s)$ pour lequel les deux conditions sont

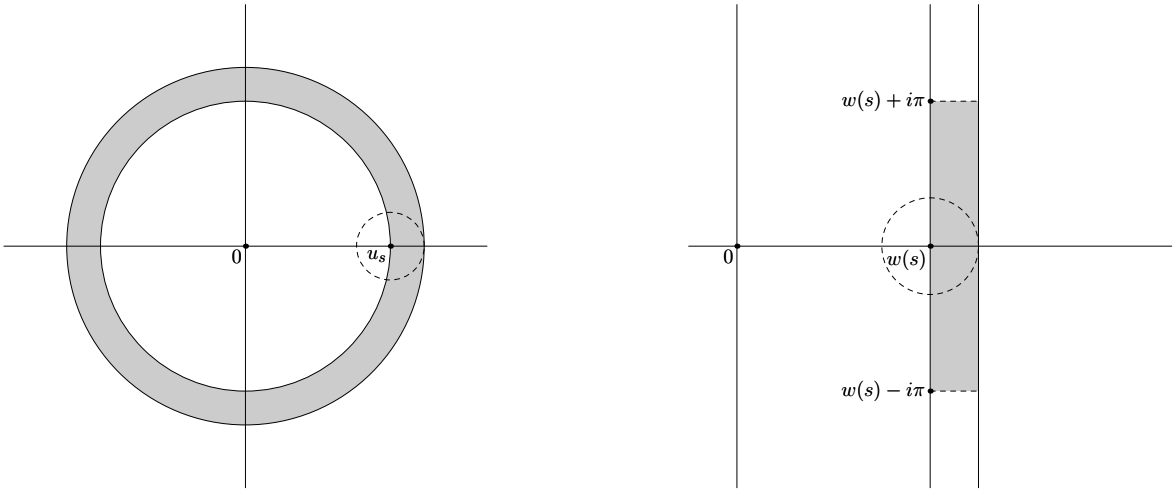

FIGURE 2.1 – Configuration du théorème 2.17. On veut montrer que l'opérateur quasi inverse est méromorphe dans la zone grise. Les deux diagrammes montrent la même situation, celui de gauche en fonction de u, celui de droite en fonction de $w = \log u$.

remplies :(i) $w \mapsto 1 - \lambda(s, w)$ s'annulle seulement en $w(s)$, – (ii) le rayon spectral $r(\underline{\mathbf{N}}_{s,w})$ est strictement inférieur à 1. Alors, chacun des deux termes de la décomposition spectrale définit une application méromorphe sur ce voisinage $\mathcal{W}$. Et donc, l'application $w \mapsto (\underline{I} - \underline{\mathbf{H}}_{s,w})^{-1}$ est méromorphe sur ce voisinage $\mathcal{W}$, avec un seul pôle en $w = w(s)$. On peut toujours supposer que ce voisinage $\mathcal{W}$ contient un rectangle de la forme $[w(s) + \rho_1(s)] \times [-\tau(s), +\tau(s)]$.

Supposons que c est primitif. Alors, d'après la proposition 2.15, sur le segment de droite $\Re w = w(s), \Im w \in [-\pi - \tau(s)] \cup [\tau(s), \pi]$, le rayon spectral $r(\underline{\mathbf{H}}_{s,w})$ de $\underline{\mathbf{H}}_{s,w}$ est strictement inférieur à $\lambda(s, w(s)) = 1$. La fonction $w \mapsto r(\underline{\mathbf{H}}_{s,w})$ est continue sur un compact, et donc il existe $\rho_2(s) > 0$ pour lequel la fonction $r(\underline{\mathbf{H}}_{s,w})$ est strictement inférieure à 1 sur le rectangle $[w(s), w(s)+\rho_2(s)] \times [-\pi-\tau(s)] \cup [\tau(s), \pi]$. Dans ce rectangle donc, l'application $w \mapsto (\underline{I}-\underline{\mathbf{H}}_{s,w})^{-1}$ est analytique. En choisissant $\rho(s) := \min(\rho_1(s), \rho_2(s))$, on a bien le résultat cherché. □

2.6.4 Méromorphie des quasi-inverses et intégration sur le domaine $\widetilde{\mathcal{B}} \setminus \mathcal{D}$.

Nous avons étudié les quasi-inverses, c'est un premier point. Mais, les principales séries génératrices ne s'expriment pas comme des quasi-inverses, ou des quasi-inverses généralisés. Ils s'expriment comme l'intégrale sur $\widetilde{\mathcal{B}} \setminus \mathcal{D}$ de quasi-inverses (possiblement généralisés). Il y a donc a priori deux questions non encore résolues :

(a) Ces quasi-inverses sont-ils bien définis, puisque les opérateurs n'agissent que sur des fonctions définies sur $\mathcal{V}$, et que bien sûr $\widetilde{\mathcal{B}} \setminus \mathcal{D}$ n'est pas inclus dans $\mathcal{V}$?

(b) Les quasi–inverses définissent des fonctions méromorphes, mais la méromorphie est-elle conservée par passage à l'intégrale sur $\widetilde{\mathcal{B}} \setminus \mathcal{D}$?

Nous allons répondre maintenant à ces questions, dans la proposition suivante.

Proposition 2.18. [Conservation de la méromorphie par intégration sur $\widetilde{\mathcal{B}} \setminus \mathcal{D}$.] *Nous considérons deux cas, un pour chaque variable, la variable s (ou r défini par $r = s - 2$) ou la variable w.*

(*i*) [Cas de la variable $r = s - 2$ pour s proche de 1] *Supposons que A_r soit une fonction de $B_\infty(\mathcal{V})$ définie sur $\overline{\mathcal{V}} \times \overline{\mathcal{V}}$, et méromorphe par rapport à la variable r en $r = -1$, qui admet, autour de $r = -1$, le développement*

$$A_r(z,\bar{z}) = \frac{a}{(r+1)^e}\underline{\psi}(z,\bar{z})\Big[1 + (r+1)R_r(z,\bar{z})\Big],$$

où R_r est une fonction de $B_\infty(\mathcal{V})$. Soit X une fonctionnelle qui peut être l'identité, ou la dérivation Δ par rapport à r. Alors, la fonction $B_r := |y|^r X\underline{\mathbf{H}}_{2+r}[A_r]$ est définie sur $\widetilde{\mathcal{B}} \setminus \mathcal{D} \times \widetilde{\mathcal{B}} \setminus \mathcal{D}$ et vérifie, autour de $r = -1$

$$\iint_{\widetilde{\mathcal{B}}\setminus\mathcal{D}} B_r(z,\bar{z})dxdy = \left[\frac{a}{(r+1)^e}\iint_{\widetilde{\mathcal{B}}\setminus\mathcal{D}} \frac{1}{|y|}\underline{\psi}(z,\bar{z})dxdy\right]\Big[1 + (r+1)T_r\Big],$$

pour une fonction T_r analytique par rapport à r au voisinage de $r = -1$.

(*i*) [Cas de la variable w]. *Supposons que A_w soit une fonction de $B_\infty(\mathcal{V})$, et méromorphe par rapport à la variable w en w_0 et qui admet, autour de $w = w_0$, le développement*

$$A_w(z,\bar{z}) = \frac{a}{(w-w_0)^e}\phi(z,\bar{z})\Big[1 + (w-w_0)R_w(z,\bar{z})\Big],$$

où R_w est une fonction de $B_\infty(\mathcal{V})$. Soit X une fonctionnelle qui peut être l'identité, ou la dérivation W par rapport à w. Alors, pour tout nombre réel r fixé vérifiant $r > -1$, la fonction $B_w := |y|^r X\underline{\mathbf{H}}_{2+r,w}[A_w]$ est définie sur $\widetilde{\mathcal{B}} \setminus \mathcal{D} \times \widetilde{\mathcal{B}} \setminus \mathcal{D}$ et vérifie, autour de $w = w_0$

$$\iint_{\widetilde{\mathcal{B}}\setminus\mathcal{D}} B_w(z,\bar{z})dxdy = \left[\frac{a}{(w-w_0)^e}\iint_{\widetilde{\mathcal{B}}\setminus\mathcal{D}} \underline{\mathbf{H}}_{2,w}\left[|y|^r\phi\right](z,\bar{z})dxdy\right]\Big[1 + (w-w_0)T_w\Big],$$

pour une fonction T_w analytique par rapport à w au voisinage de $w = w_0$.

La preuve de cette proposition est fondée sur la proposition suivante.

Proposition 2.19. *Soit F une fonction de $B_\infty(\mathcal{V})$. Alors,*

(*i*) *La fonction $(z,\bar{z}) \mapsto \underline{\mathbf{H}}_2[|y|^{-1}F](z,\bar{z})$ est intégrable sur $\widetilde{\mathcal{B}} \setminus \mathcal{D}$.*

(*ii*) *La fonction $(z,\bar{z}) \mapsto \Delta\underline{\mathbf{H}}_2[|y|^{-1}F](z,\bar{z})$ est intégrable sur $\widetilde{\mathcal{B}} \setminus \mathcal{D}$.*

(*iii*) *Pour tout $r > -1$, et tout w, pour lesquels $(2+r, w)$ appartient à $\mathfrak{H}_{(c)}$, la fonction $(z,\bar{z}) \mapsto W\underline{\mathbf{H}}_{2,w}[|y|^r F](z,\bar{z})$ est intégrable sur $\widetilde{\mathcal{B}} \setminus \mathcal{D}$.*

Démonstration. Nous commençons par rappeler les trois égalités essentielles, valables pour toute fonction L de $B_\infty(\mathcal{V})$.

$$\iint_{\widetilde{\mathcal{B}}\setminus\mathcal{D}} \underline{\mathbf{H}}_2[L](z,\bar{z})dxdy = \sum_{h\in\mathcal{H}} \iint_{h(\widetilde{\mathcal{B}}\setminus\mathcal{D})} L(z,\bar{z})dxdy.$$

$$\iint_{\widetilde{\mathcal{B}}\setminus\mathcal{D}} \Delta\underline{\mathbf{H}}_2[L](z,\bar{z})dxdy = \sum_{h\in\mathcal{H}} \iint_{h(\widetilde{\mathcal{B}}\setminus\mathcal{D})} |\log|y||\, L(z,\bar{z})dxdy.$$

$$\iint_{\widetilde{\mathcal{B}}\setminus\mathcal{D}} W\underline{\mathbf{H}}_{2,w}[L](z,\bar{z})dxdy = \sum_{h\in\mathcal{H}} c(h)e^{wc(h)} \iint_{h(\widetilde{\mathcal{B}}\setminus\mathcal{D})} L(z,\bar{z})dxdy.$$

Pour montrer que les intégrales en jeu sont convergentes, il suffit donc de montrer que les séries des membres de droite sont convergentes, quand la fonction à intégrer sur $h(\widetilde{\mathcal{B}} \setminus \mathcal{D})$ est de la forme

$$|y|^{-1}F, \quad |y|^r F, \quad \left|\log|y|^\ell\right| |y|^{-1}F, (\text{pour } \ell = 1, 2)$$

avec une fonction F de $B_\infty(\mathcal{V})$. Une telle fonction F étant bornée sur $\mathcal{D}$, elle n'intervient pas de fait dans la preuve, et il suffit de montrer que les quatre séries

$$\sum_{h\in\mathcal{H}} I[|y|^{-1}, h(\widetilde{\mathcal{B}} \setminus \mathcal{D})], \quad \sum_{h\in\mathcal{H}} I[|y|^{-1}\left|\log|y|\right|, h(\widetilde{\mathcal{B}} \setminus \mathcal{D})], \quad \sum_{h\in\mathcal{H}} I[|y|^{-1}\left|\log^2|y|\right|, h(\widetilde{\mathcal{B}} \setminus \mathcal{D})],$$

$$\sum_{h\in\mathcal{H}} c(h)e^{wc(h)} I[|y|^r, h(\widetilde{\mathcal{B}} \setminus \mathcal{D})]$$

sont convergentes. Cela découle immédiatement des deux propositions 1.7 et 1.8. □

Nous prouvons maintenant la proposition 2.18

Démonstration. Il suffit de prouver l'item (i), le second se démontrant de même. La fonctionnelle X peut être ici l'identité ou la dérivation Δ. Lorsque r est proche de -1, la fonction A_r est définie sur $\mathcal{D}\times\mathcal{D}$, la fonction $B_r := |y|^r X\underline{\mathbf{H}}_{2+r}[A_r]$ est bien définie pour $(z,\bar{z}) \in \widetilde{\mathcal{B}} \setminus \mathcal{D} \times \widetilde{\mathcal{B}} \setminus \mathcal{D}$, et elle vérifie

$$B_r(z,\bar{z}) = \frac{a}{(r+1)^e}|y|^r\left[X\underline{\mathbf{H}}_{2+r}[\underline{\psi}] + (r+1)X\underline{\mathbf{H}}_{2+r}[R_r\underline{\psi}]\right](z,\bar{z})$$

On a, pour une fonction $F \in B_\infty(\mathcal{V})$,

$$|y|^r X\underline{\mathbf{H}}_{2+r}[F] = X\underline{\mathbf{H}}_2[|y|^r F] \quad \text{avec} \quad |y|^r = |y|^{-1} + (r+1)O(g(y)), \qquad g(y) = |y|^{-1}|\log|y||$$

quand $y \in]0,1]$, le O faisant intervenir une constante indépendante de r. On a donc

$$B_r = \frac{a}{(r+1)^e}\left[X\underline{\mathbf{H}}_2[|y|^{-1}\underline{\psi}] + O(r+1)X\underline{\mathbf{H}}_2[g\underline{\psi}] + (r+1)X\underline{\mathbf{H}}_2[|y|^{-1}R_1\underline{\psi}] + O(r+1)^2X\underline{\mathbf{H}}_2[R_1\underline{\psi}g]\right].$$

On remarque que R_r et $\underline{\psi}$ sont toutes les deux bornées sur $\mathcal{D}\times\mathcal{D}$, et restent bornées aussi quand r tend vers -1, et donc, les inégalités

$$|X\underline{\mathbf{H}}_2[R_r\psi|y|^{-1}]| \le KX\underline{\mathbf{H}}_2[|y|^{-1}], \qquad |X\underline{\mathbf{H}}_2[R_r\psi g]| \le KX\underline{\mathbf{H}}_2[g]|$$

permettent de les faire disparaître du jeu. Il suffit de montrer donc que les fonctions $X\underline{\mathbf{H}}_2[|y|^{-1}]$ et $X\underline{\mathbf{H}}_2[g]$ sont intégrables sur $\widetilde{\mathcal{B}} \setminus \mathcal{D}$. C'est justement ce que nous venons de prouver □

Tous les résultats sont maintenant en place pour pouvoir utiliser les quasi-inverses dans l'analyse de l'exécution de l'algorithme de Gauss, que nous effectuons dans le chapitre suivant.

Chapitre 3

Analyse des paramètres d'exécution de l'algorithme de Gauss

Sommaire

Le présent chapitre propose une analyse fine de l'algorithme de Gauss. On travaille sur l'ensemble des entrées

$$\Omega_N := \left\{ \omega = \frac{v}{u} \in \mathcal{D} \ : \ u = (N, 0),\ v = (a, b),\ a, b, N \ \in \mathbb{Z},\ b \neq 0 \right\}, \tag{3.1}$$

et on cherche à décrire le comportement de l'algorithme GAUSS-INTERNE lorsque le paramètre N tend vers l'infini. Nous allons centrer notre étude sur le comportement probabiliste de l'algorithme, mais nous commençons par rappeler le comportement de l'algorithme dans le pire des cas (section 3.1). Puis nous énonçons les principaux résultats de cette thèse qui décrivent l'exécution de l'algorithme de Gauss (théorèmes $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$) et nous les mettons en regard avec les résultats analogues déjà obtenus pour l'algoritme d'Euclide (section 3.2). Puis nous prouvons ces théorèmes, d'abord les deux théorèmes qui s'établissent dans le modèle continu (théorème **A** dans la section 3.3, puis théorème **B** dans la section 3.4). La section 3.5 finale est dédiée aux preuves des deux derniers thèorèmes $\mathbf{C}, \mathbf{D}$, dans le modèle discret.

3.1 Étude de l'algorithme de Gauss dans le pire des cas

Nous étudions trois variables principales, le nombre d'itérations P, les coûts C et D.

3.1.1 Nombre d'itérations.

Nous commençons par le nombre d'itérations. L'étude du nombre maximum d'itérations est la version complexe du résultat de Vallée, qu'elle a initialement prouvé dans le cadre vectoriel classique [76].

Proposition 3.1 (Vallée [76] 1991)**.** *Dans l'ensemble Ω_N, le nombre maximum P_N d'itérations P de l'algorithme* GAUSS-INTERNE*, satisfait*

$$P_N \sim \frac{1}{2} \log_{1+\sqrt{2}} N.$$

Démonstration. Tout d'abord, on observe que l'inclusion

$$[P \geq k+1] \subset \left\{ z; \quad |\Im(z)| \leq \frac{1}{2} \left(\frac{1}{1+\sqrt{2}} \right)^{2k-1} \right\} \tag{3.2}$$

nous fournit le résultat. En effet, tout complexe $z = (1/N)(a+ib)$ de Ω_N satisfait $b \neq 0$ et a a une partie imaginaire de module au moins $1/N$. Donc, le complexe z appartient au domaine $[P \leq k]$ dès que $N < 2(1+\sqrt{2})^{2k-1}$, ou dès que

$$k > \frac{1}{2} \left(1 + \log_{(1+\sqrt{2})} \frac{N}{2} \right).$$

Le plus petit k qui satisfait (3.1.1) est la borne supérieure pour le nombre d'itérations ; elle vérifie,

$$P_N = \left\lceil \frac{1}{2} \left(1 + \log_{(1+\sqrt{2})} \frac{N}{2} \right) \right\rceil.$$

Maintenant nous prouvons la relation (3.2). D'après (1.16) le domaine $[R \geq k+1]$ est l'union des transformées $h(\mathcal{D})$ pour $h \in \mathcal{H}^k$, où $\mathcal{D}$ et $\mathcal{H}$ sont définis respectivement dans (1.10) et (1.12). Tous les disques $h(\mathcal{D})$ ont leur centre dans l'axe réel, Le plus grand d'entre eux correspond à l'homographie h qui est obtenue en composant k fois la branche primaire $h_{m,\epsilon}$ de coefficient minimal $(m, \epsilon) = (2, +1)$. Dans ce cas, les coefficients (c, d) de $h(z) = (az+b)/(cz+d)$ sont les termes A_k, A_{k+1} de la séquence définie par la condition initiale $A_0 = 0, A_1 = 1$ et la récurrence $A_{k+1} = 2A_k + A_{k-1}$, qui satisfait $A_k \geq (1+\sqrt{2})^{k-2}$. Alors, le plus grand disque a par rayon

$$\frac{1}{2} \left| h(0) - h(\frac{1}{2}) \right| = \frac{1}{2d(2d+c)} = \frac{1}{2A_{k+1}(2A_{k+1}+A_k)} = \frac{1}{2A_{k+1}A_{k+2}}.$$

La matrice associée à l'homographie $h_{2,1}$ a pour valeur propre dominante $(1+\sqrt{2})$, ce qui implique que, asymptotiquement, on a $A_k \sim (1+\sqrt{2})^k$, de sorte que le plus grand disque a un rayon au plus égal à $(1/2)(1+\sqrt{2})^{1-2k}$, comme on voulait montrer. □

3.1.2 Comportement des fonctions Q et D.

Nous décrivons le comportement de Q, D à l'intérieur d'un domaine $h(\widetilde{\mathcal{B}} \setminus \mathcal{D})$ qui regroupe tous les complexes z pour lequel l'algorithme utilise la même branche $h \in \mathcal{H}^+$.

Proposition 3.2. *Soit $h \in \mathcal{H}^+$, avec $h(z) = (az+b)/(cz+d)$. Alors, l'ordre de grandeur des fonctions Q et D à l'intérieur du domaine $h(\widetilde{\mathcal{B}} \setminus \mathcal{D})$ satisfait*

$$|Q(z)| = O(\log d), \qquad |D(z)| = O((\log d)^2).$$

De plus, Q est constante et la différentielle ΔD de la fonction D vérifie sur $h(\widetilde{\mathcal{B}} \setminus \mathcal{D})$

$$||\Delta D|| = O(\log d)\frac{1}{\rho_h} \qquad \text{où } \rho_h \text{ est le diamètre du grand disque de } h(\widetilde{\mathcal{B}} \setminus \mathcal{D}).$$

Démonstration. Tout d'abord, on rappelle les expressions de Q, D obtenues précédemment dans la section 1.2.10 du chapitre 1,

$$Q(z) = \sum_{i=1}^{P(z)} \ell(m_i) \qquad D(z) = \sum_{i=1}^{P(z)} \ell(m_i) \cdot \lg |c_{i-1}z - a_{i-1}|^2, \qquad m_i = \left\lfloor \frac{d_i}{|c_i|} + \frac{1}{\phi^2} \right\rfloor$$

où (a_i, c_i) sont les coefficients de l'homographie h_i qui regroupe les i premières étapes de l'algorithme. On sait aussi que la suite des rationnels a_i/c_i, complétée par a/c forme la suite des convergents de b/d. Le quotient m_i vérifie

$$m_i \leq \frac{d_i}{|c_i|} + \frac{1}{\phi^2} \leq 2\frac{d_i}{|c_i|} = 2\frac{d_i}{d_{i-1}},$$

où la dernière égalité utilise le fait que $|c_i| = d_{i-1}$, avec $d_0 = 1$. Donc,

$$\ell(m_i) \leq 1 + \lg m_i \leq 1 + \lg\left(2\frac{d_i}{d_{i-1}}\right) = (1+\lg 2) + \lg\left(\frac{d_i}{d_{i-1}}\right) = 2 + \lg\left(\frac{d_i}{d_{i-1}}\right).$$

Ainsi, on obtient, en utilisant le résultat précédent

$$Q(z) = \sum_{i=1}^{P(z)} \ell(m_i) \leq 2 \cdot P(z) + \sum_{i=1}^{P(z)} \lg\left(\frac{d_i}{d_{i-1}}\right) \leq 2 \cdot (\lg d + 2) + \lg\left(\frac{d}{d_0}\right) = 3 \cdot \lg d + 4,$$

ce qui établit en particulier que $Q(z) = O(\log d)$.

Dans le cas de $D(z)$, nous commençons par étudier D en un point particulier de $h(\widetilde{\mathcal{B}} \setminus \mathcal{D})$, qui est le point $h(0) = b/d$, et on remarque que

$$D(\frac{b}{d}) = \sum_{i=1}^{P} \ell(m_i) \left(\lg c_i^2 + \left| \lg \left| \frac{b}{d} - \frac{a_{i-1}}{c_{i-1}} \right|^2 \right| \right).$$

Comme b/d et a_{i-1}/c_{i-1} sont tous les deux éléments de $[0, 1/2]$, on a

$$\frac{1}{c_i d} \leq \left| \frac{b}{d} - \frac{a_{i-1}}{c_{i-1}} \right| \leq \frac{1}{2},$$

donc

$$\left|\lg\left|\frac{b}{d}-\frac{a_{i-1}}{c_{i-1}}\right|\right| = \lg\left|\frac{b}{d}-\frac{a_{i-1}}{c_{i-1}}\right|^{-1} \leq \lg(|c_{i-1}|d) \leq \lg(|c_{i-1}|d) \leq 2\lg d.$$

De même, $\lg c_{i-1}^2 \leq 2\lg d$, et donc, en utilisant le résultat sur Q,

$$D(b/d) \leq 2Q(z)\lg d = O((\log d)^2).$$

Étudions maintenant la variation de la fonction D sur $h(\widetilde{\mathcal{B}} \setminus \mathcal{D})$. Ecrivons D sous la forme

$$D(x,y) = \sum_{i=1}^{p} \ell(m_i)\lg c_{i-1}^2 - \sum_{i=1}^{p)} \ell(m_i)D_i(x,y) \qquad \text{avec} \quad D_i(x,y) := \lg\left((x-x_i)^2+y^2\right),$$

où l'on a posé $x_i = a_{i-1}/c_{i-1}$. Ainsi, la différentielle de D vérifie, grâce à la linéarité,

$$||\Delta D(z)|| \leq \sum_{i=1}^{P} \ell(m_i)||\Delta D_i||. \tag{3.3}$$

Nous allons borner $||\Delta D_i(z)||$ uniformément en i. Nous avons

$$\frac{\partial D_i}{\partial x}(x,y) = \frac{2(x-x_i)}{(x-x_i)^2+y^2} \quad \text{et} \quad \frac{\partial D_i}{\partial y}(x,y) = \frac{2y}{(x-x_i)^2+y^2},$$

Ainsi,

$$||\Delta D_i(z)|| = \frac{2}{|z-x_i|}.$$

Ppuisque $x_i = a_{i-1}/c_{i-1}$ est un convergent de $h(0) = b/d$ distinct de a/c, il est sur l'axe réel, à l'extérieur du diamètre majeur de $h(\widetilde{\mathcal{B}} \setminus \mathcal{D})$, et le dénominateur c_{i-1} est strictement inférieur à c. Ainsi, nous avons

$$|z-x_i| \geq \min\left\{\left|\frac{a}{c}-x_i\right|, \left|\frac{b}{d}-x_i\right|, \left|\frac{a+2b}{c+2d}-x_i\right|\right\} \geq$$
$$\min\left\{\frac{1}{|c_{i-1}c|}, \frac{1}{|c_{i-1}|d}, \frac{1}{|c_{i-1}|(c+2d)}\right\} \geq \frac{1}{|c|(c+2d)}.$$

Nous comparons cette borne au rayon ρ_h du disque majeur. et, grâce aux relations d'Hurwitz,

$$|z-x_i| \geq \frac{1}{2\rho_h} \quad (c<0) \qquad |z-x_i| \geq \frac{1}{\phi^2}\frac{1}{\rho_h} \quad (c>0)$$

Ainsi, pour $z \in h(\widetilde{\mathcal{B}} \setminus \mathcal{D})$

$$||\Delta D_i(z)|| = O\left(\frac{1}{\rho_h}\right),$$

et donc, utilisant (3.3), et rappelant que $Q(z)$ est un $O(\log d)$ sur $h(\widetilde{\mathcal{B}} \setminus \mathcal{D})$, nous obtenons

$$||\Delta D(z)|| \leq Q(z) \cdot O\left(\frac{1}{\rho_h}\right) = O(\log d)\frac{1}{\rho_h}$$

pour tout $z \in h(\widetilde{\mathcal{B}} \setminus \mathcal{D})$, comme voulu. Comme $h(\widetilde{\mathcal{B}} \setminus \mathcal{D})$ est inclus dans le grand disque, donr le diamètre est $2\rho_h$, on en déduit que $|D(z)-D(z')|$ est en $O(\log d)$. On conclut alors facilement que l'ordre de $D(z)$ dans $h(\widetilde{\mathcal{B}} \setminus \mathcal{D})$ est $O((\log d)^2)$, puisque $D(b/d) = O((\log d)^2)$ et que la variation à l'intérieur du domaine est en $O(\log d)$. Cela conclut la preuve. □

3.1.3 Comportement d'un coût additif C et de la complexité binaire dans le pire des cas.

On étudie maintenant des coûts additifs C à croissance modérée, et la complexité binaire B, définis dans la section 3.3.2) de la partie I.

Proposition 3.3. *Dans l'ensemble Ω_N, la valeur maximale C_N d'un coût additif C à croissance modérée est en $\Theta(\log N)$. La valeur maximum B_N de la complexité binaire B dans $\tilde{\Omega}_N$ est $\Theta((\log N)^2)$.*

Démonstration. On commence par les coûts additifs, puis on étudie la complexité binaire.

Coûts additifs. Puisque nous étudions les coûts c à croissance modérée, il suffit d'étudier le coût C relatif au coût élémentaire $c(q) := \log q$. Tout d'abord, on observe que

$$|q_i| = m_i = \left\lfloor \frac{d_i}{|c_i|} + \frac{1}{\phi^2} \right\rfloor \leq \frac{d_i}{|c_i|} + \frac{1}{\phi^2} \leq 2\frac{d_i}{|c_i|} = 2\frac{d_i}{d_{i-1}},$$

où la dernière égalité utilise le fait que $|c_i| = d_{i-1}$, avec $d_0 = 1$. Par ailleurs,

$$\ell(|q_i|) \leq 1 + \lg|q_i| \leq 1 + \lg\left(2\frac{d_i}{d_{i-1}}\right) = (1 + \lg 2) + \lg\left(\frac{d_i}{d_{i-1}}\right) = 2 + \lg\left(\frac{d_i}{d_{i-1}}\right).$$

Ainsi, pour $Q(z)$ on obtient,

$$Q(z) = \sum_{i=1}^{P(z)} \ell(|q_i|) \leq 2 \cdot P(z) + \sum_{i=1}^{P(z)} \lg\left(\frac{d_i}{d_{i-1}}\right) \leq 2 \cdot (\lg d + 2) + \lg\left(\frac{d}{d_0}\right) = 3 \cdot \lg d + 4,$$

ce qui établit en particulier que $Q(z) = O(\log d)$.

Complexité binaire. Dans ce cas, le résultat est obtenu grâce à l'équation (2.8). □

3.2 Analyse probabiliste de l'algorithme GAUSS-INTERNE. Comparaison avec celui de EUCLIDE-PLIÉ.

Nous voulons comparer le comportement des algorithmes GAUSS-INTERNE et EUCLIDE-PLIÉ. Nous rappelons d'abord les résultats principaux sur le comportement probabiliste de l'algorithme EUCLIDE-PLIÉ.

3.2.1 Principaux résultats de l'analyse de l'Algorithme EUCLIDE-PLIÉ. Analyse en moyenne.

Nous reprenons la discussion du début du chapitre 3 de la partie I, et ous présentons les résultats connus sur l'analyse en moyenne de l'algorithme EUCLIDE-PLIÉ.

Théorème 3.1. (Vallée, Akhavi and Vallée) (1995-2000) *Dans l'ensemble ω_N formé par les paires d'entrées (u, v) pour lesquelles $u/v \in \tilde{\mathcal{I}}$ et $|v| \leq N$, le nombre moyen d'itérations P, la valeur moyenne d'un coût C à croissance modérée, la valeur moyenne de la complexité en bits B de l'algorithme* EUCLIDE-PLIÉ *satisfont, lorsque $N \to \infty$,*

$$\mathbb{E}_N[P] \sim \frac{2\log 2}{h(\mathcal{E})} \lg N, \quad \mathbb{E}_N[C_{(c)}] \sim \frac{2\log 2}{h(\mathcal{E})} \mathbb{E}[c] \lg N, \quad \mathbb{E}_N[B] \sim \frac{\log 2}{h(\mathcal{E})} \mathbb{E}[\ell] \lg^2 N.$$

Ici, $h(\mathcal{E})$ est l'entropie du système Euclide-plié *décrite en (2.55) et $\mathbb{E}[c]$ désigne la valeur moyenne du coût élémentaire c par rapport à la densité invariante ψ définie en (2.56). En particulier, lorsque c est la longueur binaire, il y a une formule close pour $\mathbb{E}[\ell]$, qui est une constante de type Khinchin, de la forme*

$$\mathbb{E}[\ell] := \sum_{h\in\mathcal{H}} \int_{h(\tilde{\mathcal{I}})} \ell(h)\psi(x)dx = \frac{1}{\log\phi} \log \prod_{k\geq 1} \frac{2^k\phi^2+\phi}{2^k\phi^2-1} \approx 2.02197. \tag{3.4}$$

Les constantes qui apparaissent dans les valeurs moyennes –l'entropie $h(\mathcal{E})$ ou l'espérance limite $\mathbb{E}[c]$– sont des constantes du système dynamique, qui représente l'extension continue de l'algorithme. On a ainsi une manifestation du transfert du continu au discret dont nous avons déjà parlé. Cela se confirmera dans les analyses de l'algorithme de Gauss.

3.2.2 Principaux résultats de l'analyse de l'Algorithme Euclide-plié. Analyse en distribution.

Il existe aussi des résultats plus précis, en distribution, qui montrent que tous ces coûts $P, C_{(c)}$, ainsi qu'une version régularisée de B, suivent des lois asymptotiquement gaussiennes pour $N \to \infty$.

Théorème 3.2. (Hensley, Baladi, Lhote, Vallée) (1994-2007) *Considérons l'ensemble ω_N formé par les paires d'entrées (u, v) de l'algorithme* Euclide-plié*, vérifiant $u/v \in \tilde{\mathcal{I}}$ et $|v| \leq N$. Alors :*

(*i*) *Un coût additif $C_{(c)}$ associé à un coût élémentaire modéré suit une loi asymptotiquement gaussienne, avec une vitesse de convergence en $O(1/\sqrt{\lg N})$:*

$$\mathbb{P}_N\left[(u,v);\ \frac{C_{(c)}(u,v)-\mathbb{E}[c]\cdot\lg N}{\rho(c)\cdot\sqrt{\lg N}} \leq y\right] = \frac{1}{2\pi}\int_{-\infty}^{y} e^{-x^2/2}dx + O\left(\frac{1}{\sqrt{\lg N}}\right),$$

où $\mathbb{E}[c]$ est l'espérance limite définie en (2.56), et où la constante $\rho(c)$ est liée à la valeur propre dominante de l'opérateur $\mathbf{H}_{s,w}$. L'espérance et la variance de $C_{(c)}$ vérifient

$$\mathbb{E}(C_{(c)}) = \mathbb{E}[c]\cdot\lg N + \mu_1[c] + O(N^{-\alpha}) \qquad \mathbb{V}(C) = \rho(c)\cdot\lg N + \rho_1(c) + O(N^{-\alpha})$$

où α est une constante positive indépendante de c.

(*ii*) *La complexité en bits B de l'algorithme* Euclide-plié *vérifie*

$$\mathbb{E}(B) = \mathbb{E}(\ell)\cdot\lg^2 N\left(1+O\left(\frac{1}{\lg N}\right)\right) \qquad \mathbb{V}(B) = \rho_0(\ell)\cdot\lg^3 N\left(1+O\left(\frac{1}{\lg N}\right)\right),$$

où $\mathbb{E}(\ell)$ possède la forme close (3.4), et $\rho(\ell)$ est la constante correspondante au coût additif dont $c \equiv \ell$, qui correspond au coût additif $Q(u, v)$.

(*ii*) *La complexité en bits $\check{B}$ associée à l'algorithme d'Euclide* étendu *possède un distribution asymptotiquement gaussienne, avec une vitesse de convergence en $O(1/(\lg N)^{1/3}$.*

Comme on le voit, la plupart des coûts naturels intervenant dans l'algorithme d'Euclide ont une distribution qui est asymptotiquement gaussienne. C'est un problème ouvert de montrer que cela est vrai pour la complexité binaire de ll'algorithme non étendu.

3.2.3 Euclide et Gauss : Ressemblances, différences.

Le reste de ce chapitre est consacré à la comparaison entre les deux algorithmes de Gauss et d'Euclide. Que peut-on espérer du comportement probabiliste de l'algorithme de Gauss ? D'un côté, il y a une grande ressemblance entre les deux algorithmes, puisque l'algorithme de Gauss peut se voir comme une généralisation formelle de l'algorithme d'Euclide. D'un autre côté, des différences importantes apparaissent lorsque l'on considère le comportement des deux algorithmes. En effet, les algorithmes d'Euclide, et en particulier EUCLIDE-PLIÉ terminent uniquement sur des entrées rationnelles (qui vont dans le trou $\{0\}$), alors que l'algorithme de Gauss termine toujours, sauf pour des entrées irrationnelles réelles. Néanmoins, une partie de ces différences disparaissent, lorsque l'on se restreint à des entrées rationnelles, réelles ou complexes. Dans ce cas, les deux algorithmes terminent, et il est important de comparer le comportement de ces algorithmes et de voir s'il existe une transition de l'un vers l'autre.
Nous présentons maintenant et expliquons les principaux résultats que nous obtenons dans cette thèse sur l'algorithme de Gauss.

3.2.4 Les résultats de ce chapitre.

Nous présentons ici quatre principaux résultats sur l'analyse de l'algorithme de Gauss. Le premier résultat montre une première différence importante entre les deux algorithmes. Il exhibe une loi géométrique pour une classe de coûts additifs, et constitue une généralisation importante d'un résultat de Daudé, Flajolet, Vallée, qui avaient seulement étudié le cas du nombre d'itérations.

Théorème A. (Vallée et Vera, [74] 1994-2007) *Considérons un coût additif $C_{(c)}$ lié à un coût c de croissance modérée, supposé de plus primitif. Alors, pour toute densité d'entrée f de valuation $r > -1$, le coût $C_{(c)}$ suit une loi asymptotiquement géométrique. La raison de cette loi géométrique dépend du coût c et de la valuation r, et non de la fonction f elle-même. On a :*

$$\mathbb{P}_{\langle f\rangle}[C_{(c)} = k] \sim_{k\to\infty} a(f,r)\,\mu(c,r)^k. \tag{3.5}$$

La raison $\mu(c,r)$ se définit à partir de la valeur propre dominante $\lambda(2+r,w)$ de l'opérateur $\mathbf{H}_{s,w,(c)}$ défini en () et vérifie $\lambda(2+r,-\log\mu(c,r)) = 1$. La raison $\mu(c,r)$ tend vers 1 quand r tend vers -1 et $\log\mu(c,r)$ est $\Theta(r+1)$ quand $r \to -1$.

Pour l'algorithme d'Euclide, ces mêmes coûts admettent une distribution limite gaussienne, alors qu'on est en présence ici d'une loi géométrique. Il faut néanmoins prendre en compte que le modèle probabiliste n'est pas du même genre (continu dans le cas de l'algorithme de Gauss, discret dans le cas de l'algorithme d'Euclide). Le retour au modèle discret n'est pas fait dans cette thèse, en toute généralité. Mais c'est un retour que Daudé, Flajolet et Vallée avait effectué dans un cas particulier (nombre d'itérations dans un modèle uniforme, de valuation nulle), et ils avaient démontré que la distribution de la variable P_N (nombre d'itérations dans le modèle discret de taille N) convergeait vers la distribution de la variable P (nombre d'itérations dans le modèle continu).

Le résultat suivant étudie les coûts C et D de l'algorithme de Gauss, dans le modèle continu.

Théorème B (Étude en moyenne de la complexité, modèle d'entrées continu, $r \to -1$)**.** *Considérons l'algorithme* GAUSS-INTERNE *; où les entrées sont distribuées dans le disque $\mathcal{D}$ selon la densité standard de valuation $r > -1$. Alors*

(*i*) *L'espérance* $\mathbb{E}_r(C)$ *d'un coût additif à croissance modérée, et l'espérance* $\mathbb{E}_r(D)$ *du coût* D *sont bien définies et satisfont, quand* $r \to -1$,

$$\mathbb{E}_r(C) \sim \frac{1}{r+1}\frac{\mathbb{E}(c)}{h(\mathcal{E})}, \qquad \mathbb{E}_r(D) \sim -\frac{1}{(r+1)^2}\frac{1}{\log 2}\frac{\mathbb{E}(\ell)}{h(\mathcal{E})}.$$

Ici, $h(\mathcal{E})$ *désigne l'entropie du système d'Euclide centré,* ℓ *désigne la longueur binaire, et* $\mathbb{E}(c)$ *désigne la valeur moyenne limite du coût* c.

(*ii*) *Quand* r *tend vers* -1, *la densité de sortie associée à une densité initiale de valuation* r *tend vers*

$$\frac{1}{h(\epsilon)}\frac{1}{y}\underline{\psi}(z,\bar{z}),$$

où $\underline{\psi}$ *est la densité invariante pour l'opérateur généralisé* $\underline{\mathbf{H}}$, *dont une expression est*

$$\underline{\psi}(z,\bar{z}) = \frac{1}{\log\phi}\frac{1}{z-\bar{z}}\left(\log\frac{\phi+z}{\phi+\bar{z}} - \log\frac{\phi^2-z}{\phi^2-\bar{z}}\right)$$

Ce résultat fournit donc une analyse en moyenne des coûts C, D dans le modèle continu, et décrit le comportement des valeurs moyennes lorsque la valuation r tend vers -1. Nous considérons maintenant le modèle discret associé à une taille N, quand la taille N tend vers l'infini, d'abord avec une valuation fixe $r > -1$, puis, ensuite, avec une valuation r qui tend vers -1 en même temps que N tend vers l'infini. Plus précisément, nous obtenons les résultats suivants :

Théorème C (Complexité en bits dans le modèle d'entrées discret, valuation fixe)**.** *Considérons l'algorithme* Gauss-interne *travaillant sur l'ensemble* Ω_N *des bases d'entrées* (u,v) *entières vérifiant les trois conditions :* $u = (N,0)$, $\ell(||v||^2) \leq N^2$ *et* $v/u \in \mathcal{D}$, *et se distribuant selon la densité standard de valuation* $r > -1$.

(*i*) *Considérons un coût* X *défini sur* Ω_N, X *pouvant être un coût additif* C *à croissance modérée ou le coût* D *associé à la complexité binaire. Alors, quand* $N \to \infty$, *la valeur moyenne* $\mathbb{E}_{r,N}(X)$ *du coût* X *tend vers la valeur moyenne* $\mathbb{E}_r(X)$ *du coût* X. *Plus précisément,*

$$\mathbb{E}_{r,N}(X) = \mathbb{E}_r(X) + O\left(\frac{(\log N)^{e(X)+1}}{N^{r+1}}\right),$$

où l'exposant $e(X)$ *dépend du coût* X *et satisfait* $e(C) = 1$, $e(D) = 2$.

(*ii*) *Pour tout valuation fixée* $r > -1$, *lorsque* $N \to \infty$, *la valeur moyenne* $\mathbb{E}_{r,N}(B)$ *de la complexité binaire est donc asymptotique à une fonction linéaire de* $\log N$ *et satisfait,*

$$\mathbb{E}_{r,N}(B) \sim 2\mathbb{E}_r(Q)\cdot\ell(N),$$

où Q *est le coût additif associé au coût élémentaire longueur binaire, et* $\mathbb{E}_r(Q)$ *est l'espérance de* Q *dans le modèle continu de valuation* r.

Le dernier résultat est particulièrement important puisqu'il fournit une analyse de la transition avec l'algorithme d'Euclide. On montre que, si la convergence de la taille N vers l'infini et celle de la valuation r vers -1 se fait de manière "raisonnable" (de sorte que l'espérance $\mathbb{E}_{r,N}(()X)$ dans le modèle discret soit asymptotiquement équivalente à l'espérance $\mathbb{E}_r(()X)$ dans le modèle continu), alors, il y a une bonne transition vers l'algorithme d'Euclide : la complexité limite en bits de l'algorithme de Gauss est de même ordre que la complexité en bits de l'algorithme d'Euclide.

Théorème D (Complexité en bits, cas limite $r \to -1$, transition vers algorithme d'Euclide)**.** *Considérons l'algorithme* Gauss-interne*, travaillant sur l'ensemble Ω_N des bases entières (u, v) vérifiant les trois conditions : $u = (N, 0)$, $\ell(||v||^2) \leq N^2$ et $v/u \in \mathcal{D}$, et se distribuant selon la densité standard de valuation $r > -1$. Considérons un coût X défini sur Ω_N, X pouvant être un coût additif C à croissance modérée ou le coût D associé à la complexité binaire.*

(*i*) *Lorsque la taille* $\log N$ *tend vers l'infini et la valuation* r *tend vers* -1*, avec* $(r+1)\log N = \Omega(1)$*, alors la valeur moyenne* $\mathbb{E}_{r,N}(X)$ *satisfait*

$$\mathbb{E}_{r,N}(X) = \mathbb{E}_r(X)\left[1 + O\left(\frac{(\log N(r+1))^{e(X)+1}}{N^{r+1}}\right) + O\left(\frac{1}{N^{r+1}-1}\right)\right], \tag{3.6}$$

où l'exposant $e(X)$ *dépend dans le coût* X *et satisfait* $e(C) = 1$*,* $e(D) = 2$*.*

(*ii*) *Lorsque* $r + 1 =: (\log N)^{-\alpha}$ *avec* $(\log N)^\alpha \to \infty$ *(avec* $0 < \alpha < 1$*), alors les valeurs moyennes* $\mathbb{E}_{r,N}(C)$*,* $\mathbb{E}_{r,N}(D)$ *et* $\mathbb{E}_{r,N}(B)$ *satisfont*

$$\mathbb{E}_{r,N}(C) \sim \frac{\mathbb{E}[c]}{h(\mathcal{E})}(\log N)^\alpha, \qquad \mathbb{E}_{r,N}(D) \sim \frac{\mathbb{E}[\ell]}{h(\mathcal{E})}(\log N)^{2\alpha}, \tag{3.7}$$

$$\mathbb{E}_{r,N}(B) \sim 2\frac{\mathbb{E}[\ell]}{h(\mathcal{E})}(\log N)^{1+\alpha}. \tag{3.8}$$

(*iii*) *Lorsque* $(r+1)\log N = \Theta(1)$*, alors les valeurs moyennes* $\mathbb{E}_{r,N}(C)$*,* $\mathbb{E}_{r,N}(D)$ *et* $\mathbb{E}_{r,N}(B)$ *satisfont*

$$\mathbb{E}_{r,N}(C) = \Theta(\log N), \qquad \mathbb{E}_{r,N}(D) = \Theta((\log N)^2), \qquad \mathbb{E}_{r,N}(B) = \Theta((\log N)^2). \tag{3.9}$$

Nous commençons notre étude par la distribution des coûts additifs, ensuite nous abordons l'étude des variables Q et D quand $r \to -1$ dans le modèle continu, pour finir avec l'étude de la complexité en bits dans le modèle discret.

3.3 Étude de la distribution des coûts additifs

Ici, nous voulons analyser les coûts additifs décrits dans la section 2.2.6, et exprimés dans le cadre complexe dans la proposition 1.4. La preuve fait usage des propriétés analytiques de l'opérateur quasi-inverse. Elle donne une généralisation intéressante d'un résultat que Daudé, Flajolet, Vallée avaient déjà obtenu par en 1994, dans le cas du nombre d'itérations.

3.3.1 Etude générale d'un coût additif.

Nous faisons la preuve du théorème **A**. Soit f une densité de valuation r sur $\mathcal{D}$, et F son extension diagonale à deux variables, qui s'exprime donc sous la forme

$$F(z, \bar{z}) = |y|^r L(z, \bar{z}), \qquad \text{avec} \quad L(u, u) \neq 0$$

Alors, la série génératrice de probabilités du coût $C_{(c)}$ pour cette densité initiale f est

$$B(u) := \mathbb{E}_{\langle f\rangle}[u^{C_c}] = \sum_{k\geq 0} \mathbb{P}_{\langle f\rangle}[C_{(c)} = k]\, u^k. \tag{3.10}$$

C'est une série entière dont le rayon de convergence autour de l'origine est au moins égal à 1. En posant $u = e^w$, on retrouve la série génératrice des moments pour $C_{(c)}$, qu'on a exprimé en fonction de l'opérateur $\underline{\mathbf{H}}_{s,w}$ comme

$$\mathbb{E}[e^{wC_c}] = \iint_{\widetilde{\mathcal{B}}\setminus\mathcal{D}} |y|^r \underline{\mathbf{H}}_{2+r,w} \circ (\mathbf{I} - \underline{\mathbf{H}}_{2+r,w})^{-1}[G](\hat{z}, \bar{\hat{z}}) d\hat{x} d\hat{y}. \tag{3.11}$$

Le coût c est un coût général à croissance modérée, mais on a vu, que selon que le pgcd d de ses valeurs est égal à 1 ou non, les propriétés sont différentes. En posant $\hat{c} := c/d$, le coût $\hat{c}$ devient primitif, et nous pouvons donc nous limiter à ce cas. On considère donc dans le corps de la preuve un coût primitif, et on reviendra à un coût quelconque à la fin.

Alors, les propositions 2.16, 2.17 et 2.18 du chapitre précédent prouvent ce qui suit : Il existe un unique $w := w_r$ pour lequel la valeur propre $\lambda(2 + r, w) = 1$. De plus, $B(u)$ a un pôle en $u = e^{w_r}$, et qu'il existe un $R > e^{w_r}$ pour lequel $B(u)$ est méromorphe dans le disque $|u| \leq R$, avec comme unique pôle dans ce disque le pôle simple en $u = e^{w_r}$. On a de plus

$$B(e^w) = \left[\frac{1}{(w - w_r)} \frac{\nu_r[L]}{\lambda'_w(2 + r, w_r)} \iint_{\widetilde{\mathcal{B}}\setminus\mathcal{D}} \underline{\mathbf{H}}_{2,w} \left[|y|^r \phi_r\right](z, \bar{z}) dx dy\right] \left[1 + (w - w_r)T_w\right].$$

Ici, T_w est une fonction analytique par rapport à w au voisinage de $w = w_0$, ϕ_r est la fonction propre dominante normalisée de l'opérateur $\underline{\mathbf{H}}_{2+r,w_r}$. De plus, ν_r est la mesure dominante de l'opérateur $\underline{\mathbf{H}}^*_{2+r,w_r}$, et L est la fonction associée à la densité initiale f de densité r.

Pour pouvoir profiter de cette expression et obtenir des informations sur les coefficients de la série $B(u)$ qui sont justement les probabilités $\mathbb{P}_{\langle f\rangle}[C_{(c)} = k]$ de (3.10), nous appliquons le théorème suivant.

Théorème 3.3 (Flajolet et Sedgewick, [23, Théorème IV.10, p. 258])**.** *Soit $B(u)$ une fonction méromorphe dans un disque fermé $|u| \leq R$, analytique en $|u| = R$, ayant un unique pôle $u_0 \neq 0$ à l'intérieur du disque, réel et positif. Alors, le coefficient b_k de son développement en série de Taylor vérifie*

$$b_k \sim \frac{1}{u_0^{k+1}} \mathrm{Res}(B(u) \ ; \ u = u_0)$$

Ici, le pôle est en $u_0 := e^{w_r}$ et le résidu de la fonction $u \mapsto B(u)$ en $u_0 = e^{w_r}$ est égal à $e^{w_r}\mathrm{Res}(B(e^w) \ ; \ w = w_r)$ et donc le résidu vérifie

$$A((r, f) := \mathrm{Res}(B(u) \ ; \ u = e^{w_r}) = \frac{e^{w_r}\, \nu_r[L]}{\lambda'_w(2 + r, w_r)} \iint_{\widetilde{\mathcal{B}}\setminus\mathcal{D}} \underline{\mathbf{H}}_{2,w} \left[|y|^r \phi_r\right](z, \bar{z}) dx dy$$

Remarquons qu'il dépend de r et de L, c'est-à-dire de l'expression complète de la densité f.

Donc, nous avons trouvé une estimation asymptotique pour le coefficient de la série génératrice $B(u)$, et donc

$$\mathbb{P}_{\langle f\rangle}[C_{(c)} = k] \sim_{k\to\infty} A(r, f)\, e^{-kw_r},$$

comme on voulait montrer. Par ailleurs, le théorème des fonctions implicites s'applique et prouve ce qui suit : puisque $\lambda(s, w)$ est une fonction dérivable en s et w dont la dérivée $w \mapsto \lambda'_w(2+r, w)$ ne s'annule pas, l'équation $\lambda(r + 2, w_r) = 1$ définit implicitement, et de façon unique, la fonction dérivable w_r, pour r proche de -1. La relation $\lambda(1, 0) = 1$, prouve que $w_{-1} = 0$. Donc, le développement de Taylor de w_r autour de $r = -1$ montre que $w_r = \Theta(r + 1)$, comme on voulait montrer.

Enfin, si l'on veut revenir à un coût général non nécessairement primitif, pour lequel d est le pgcd des valeurs, il suffit de remplacer c par cd, et donc le coût total C par Cd. On a donc dans ce cas

$$\mathbb{P}_{\langle f\rangle}[C_{(c)} = dk] \sim_{k\to\infty} A(r,f)\, e^{-kw_r}, \qquad \mathbb{P}_{\langle f\rangle}[C_{(c)} = k'] \ = 0 \qquad \text{pour} \ \ k' \notin \{dN\}$$

Cela conclut la preuve du théorème **A**.

3.3.2 Cas particulier du nombre d'itérations.

Dans le cas particulier du coût constant $c = 1$, l'opérateur $\underline{\mathbf{H}}_{s,w,(c)}$ s'exprime plus simplement et est égal à $e^w \cdot \underline{\mathbf{H}}_s$. La valeur $w_{(1)}(s)$ est définie par la relation $e^w\lambda(s) = 1$. Cela entraîne que la raison $e^{-w_{r,(1)}}$ dans (3.5) est tout simplement égale à $\lambda(2+r)$.

Dans ce cas, il existe une expression alternative pour le nombre moyen d'itérations de l'algorithme Gauss-interne, expression qui résulte de la caractérisation de Hurwitz, rappelée dans la proposition 1.1 de cette partie.

Théorème 3.4 (Daudé, Flajolet, Vallée,[18, 77] 1994-1996)**.** *Considérons le modèle continu avec la densité standard de valuation $r > -1$. Alors, l'espérance du nombre d'itérations P de l'algorithme* Gauss-interne *admet l'expression suivante*

$$\begin{aligned}\mathbb{E}_{(r)}[P] &= \frac{1}{A_0(r)} \iint_{\mathcal{D}} |y|^r (I - \underline{\mathbf{H}}_{2+r})^{-1}[1](z,\bar z)dxdy \\ &= \frac{2^{r+2}}{\zeta(2r+4)} \sum_{\substack{(c,d)=1 \\ d\phi<c<d\phi^2}} \left(\frac{1}{cd}\right)^{r+2}\end{aligned}$$

En plus, pour toute valuation $r > -1$, le nombre d'itérations suit une loi géométrique de raison $\mu(1,r) = \lambda(2+r)$

$$\mathbb{P}_{(r)}[K \geq k+1] \sim_{k\to\infty} a(r)\, \lambda(2+r)^k$$

où $\lambda(s)$ est la valeur propre dominante de l'opérateur $\underline{\mathbf{H}}_s$ et $\tilde a(r)$ s'exprime en fonction de l'opérateur $\underline{\mathbf{P}}_s$ de projection sur le sous-espace propre relatif à $\lambda(s)$, sous la forme

$$a(r) = \frac{1}{A_0(r)} \iint_{\mathcal{D}} |y|^r \underline{\mathbf{P}}_{2+r}[1](z,\bar z)dxdy.$$

Preuve. Par définition, l'évenement $[P \geq k+1]$ regroupe tous les complexes z qui sont encore dans le disque $\mathcal{D}$ au bout de k itérations. On a donc

$$[P \geq k+1] = \bigcup_{h\in\mathcal{H}^k} h(\mathcal{D}).$$

La probabilité de l'événement $[P \geq k+1]$ peut s'écrire

$$\mathbb{P}_{(r)}[K \geq k+1] = \frac{1}{A_0(r)} \sum_{h\in\mathcal{H}^k} \iint_{h(\mathcal{D})} |y|^r dxdy = \frac{1}{A_0(r)} \iint_{\mathcal{D}} \left(\sum_{h\in\mathcal{H}^k} |h'(z)|^{2+r} \right) y^r dxdy \quad (3.12)$$

On retrouve l'opérateur $\underline{\mathbf{H}}_{2+r}$, et finalement

$$\mathbb{P}_{(r)}[P \geq k+1] = \frac{1}{A_0(r)} \iint_{\mathcal{D}} |y|^r \underline{\mathbf{H}}^k_{2+r}[1](z)dxdy, \quad (3.13)$$

où $A_0(r)$ est la mesure de $\mathcal{D}$ par rapport à la densité standard de valuation r, définie dans..... Maintenant, grâce à la décomposition spectrale de $\underline{\mathbf{H}}_{s,w}$, donnée dans la proposition 2.10 nous avons, pour un $\alpha \in]0,1[$,

$$\underline{\mathbf{H}}^k_{2+r}[1](z,\bar{z}) = \lambda^k(2+r)\underline{\mathbf{P}}_{2+r}[1](z,\bar{z})(1+O(\alpha^k)).$$

En intégrant cette dernière relation sur $\mathcal{D}$ par rapport à la densité standard de valuation r, et en utilisant (3.13), on achève la preuve.

L'espérance d'une variable aléatoire entière positive vérifie

$$\mathbb{E}_{(r)}[P] = \sum_{k\geq 0} \mathbb{P}_{(r)}[P \geq k+1].$$

En remplaçant la probabilité $\mathbb{P}_{(r)}[P \geq k+1]$ par son expression intégrale (3.13), nous obtenons

$$\mathbb{E}_{(r)}[P] = \frac{1}{A_0(r)} \iint_{\mathcal{D}} |y|^r \sum_{k\geq 0} \underline{\mathbf{H}}^k_{2+r}[1](z,\bar{z})dxdy = \frac{1}{A_0(r)} \iint_{\mathcal{D}} |y|^r (\underline{I} - \underline{\mathbf{H}}_{2+r})^{-1}[1](z,\bar{z})dxdy,$$

comme souhaité.

Mais ce n'est pas la preuve originale, car les auteurs ne travaillaient pas avec les opérateurs généralisés. Nous décrivons maintenant la preuve originale. Les auteurs remarquent que le domaine $h(\mathcal{D})$ est un disque de diamètre $[h(0), h(1/2)]$, et donc

$$\mathbb{P}_{(r)}[P \geq k+1] = 2^{r+2} \sum_{h\in\mathcal{H}^k} \left| h(0) - h(\frac{1}{2}) \right|^{r+2}, \qquad \mathbb{E}_{(r)}[P] = 2^{r+2} \sum_{h\in\mathcal{H}^+} \left| h(0) - h(\frac{1}{2}) \right|^{r+2}.$$

Grâce à la caractérisation de Hurwitz (1.13), qui dit que l'ensemble d'homographies $\mathcal{H}^+$ est en correspondance avec les couples (c,d) d'entiers premiers entre eux et tels que $-d/\phi^2 < c < d/\phi$ avec $c \neq 0$, on obtient finalement

$$\mathbb{E}_{(r)}[P] = \frac{2^{r+2}}{\zeta(2r+4)} \sum_{\substack{(c,d)=1 \\ \phi d < c < (\phi^2 d)}} \left(\frac{1}{cd}\right)^{r+2}$$

comme voulu. □

À l'heure actuelle, on ne connaît pas de formule explicite pour la valeur propre dominante $\lambda(s)$ sauf en $s = 1$. Néanmoins, Loick Lhote [47] a démontré qu'on pouvait calculer cette valeur propre dominante en temps polynomial. Des valeurs numériques sont fournies dans le cas de la densité uniforme (lorsque $r = 0$) dans [47] :

$$\mathbb{E}_{(0)}[K] \approx 1,08922, \quad \lambda(2) \approx 0,07738.$$

3.4 Analyse en moyenne des paramètres Q et D dans un modèle continu pour une valuation $r \to -1$

Cette section est consacrée à la preuve du théorème **B**. Il faudrait tout d'abord, vérifier que les intégrales $I[Xf_r, \mathcal{D}]$ pour $X \in \{\mathrm{Id}, C, D\}$ sont convergentes pour une valuation r fixe, ce qui n'est pas clair du tout, pour les coûts C et D. Les fonctions Q et D à intégrer ne restent pas bornées au voisinage de l'axe (comme nous l'ont montré les résultats du lemme 3.2), et la fonction $f_r = |y|^r$ tend aussi vers l'infini (pour $r < 0$) quand y tend vers 0. Mais les résultats de la proposition 2.18 nous le montrent justement. Il reste donc à estimer le comportement des espérances et de la densité de sortie quand la valuation r tend vers -1.

3.4.1 Densité de sortie.

La proposition 2.3 montre que la densité de sortie $\hat{f}_r$ sur $\widetilde{\mathcal{B}} \setminus \mathcal{D}$ associée à la densité standard de valuation r s'exprime sous la forme

$$\hat{f}_r(z) = |y|^r \mathbf{G}_{2+r,\mathrm{Id}}[1](z,\bar{z}) \qquad \text{avec} \quad \mathbf{G}_{2+r,\mathrm{Id}} = \underline{\mathbf{H}}_{2+r} \circ (\underline{I} - \underline{\mathbf{H}}_{2+r})^{-1}.$$

Lorsque r est proche de -1, la proposition 2.16 prouve que

$$A_r(z,\bar{z}) := (\underline{I} - \underline{\mathbf{H}}_{2+r})^{-1}[1](z,\bar{z}) = \frac{1}{r+1}\,\frac{1}{h(\mathcal{E})}\underline{\psi}(z,\bar{z})\Big[1 + (r+1)R_r(z,\bar{z})\Big]$$

Cette fonction A_r est définie sur $\mathcal{D} \times \mathcal{D}$ et donc, la proposition 2.18 montre que la fonction $\hat{f}_r := |y|^r \underline{\mathbf{H}}_{2+r}[A_r]$ est bien définie pour $z \in \widetilde{\mathcal{B}} \setminus \mathcal{D}$, et que pour r proche de -1, elle vérifie

$$\hat{f}_r(z,\bar{z}) = \frac{1}{r+1}\,\frac{1}{h(\mathcal{E})}\,\frac{1}{|y|}\underline{\psi}(z,\bar{z})\big[1 + (r+1)T_r(z,\bar{z})\big],$$

où T_r est analytique sur $\widetilde{\mathcal{B}} \setminus \mathcal{D}$ et reste bornée au voisinage de $r = -1$. De plus, $\hat{f}_r$ est intégrable, et au voisinage de $r = -1$, son intégrale vérifie

$$\iint_{\widetilde{\mathcal{B}}\setminus\mathcal{D}} \hat{f}_r(z,\bar{z})dxdy = \left[\frac{1}{h(\mathcal{E})}\frac{1}{(r+1)}\iint_{\widetilde{\mathcal{B}}\setminus\mathcal{D}}\frac{1}{|y|}\underline{\psi}(z,\bar{z})dxdy\right]\Big[1 + (r+1)Q_r\Big],$$

pour une fonction Q_r analytique par rapport à r au voisinage de $r = -1$

3.4.2 Espérances des coûts C et D

Dans le chapitre 2 de cette partie II, nous avons exprimé les espérances des coûts C et D en fonction de quasi-inverses généralisés, sous la forme

$$\mathbb{E}_r(C) = \frac{1}{A_0(r)}\iint_{\widetilde{\mathcal{B}}\setminus\mathcal{D}} |y|^r\, \mathbf{G}_{2+r,C}[1](z,\bar{z})dxdy, \tag{3.14}$$

$$\mathbb{E}_r(D) = \frac{1}{A_0(r)}\iint_{\widetilde{\mathcal{B}}\setminus\mathcal{D}} |y|^r\, \mathbf{G}_{2+r,D}[1](z,\bar{z})dxdy, \tag{3.15}$$

où les opérateurs $\mathbf{G}_{s,C}$ et $\mathbf{G}_{s,D}$ sont définis comme

$$\mathbf{G}_{s,C} := (\underline{I} - \underline{\mathbf{H}}_s)^{-1} \circ W_{(c)}[\underline{\mathbf{H}}_s] \circ (\underline{I} - \underline{\mathbf{H}}_s)^{-1},$$

$$\mathbf{G}_{s,D} = (\underline{I} - \underline{\mathbf{H}}_s)^{-1} \circ W_{(\ell)}[\underline{\mathbf{H}}_s] \circ (\underline{I} - \underline{\mathbf{H}}_s)^{-1} \circ \Delta[\underline{\mathbf{H}}_s] \circ (\underline{I} - \underline{\mathbf{H}}_s)^{-1},$$

Coût C. Dans le cas du coût C, nous considérons la décomposition

$$|y|^r \mathbf{G}_{2+r,C}[F](z,\bar{z}) = |y|^r \underline{\mathbf{H}}_{2+r}[B_r] + |y|^r W \underline{\mathbf{H}}_{2+r}[A_r](z,\bar{z})$$

avec $A_r := (\underline{I} - \underline{\mathbf{H}}_{2+r})^{-1}[F](z,\bar{z})$ et $B_r = \mathbf{G}_{2+r,C}[F](z,\bar{z})$. Lorsque $z \in \mathcal{D}$, les fonctions A_r et B_r définissent des fonctions méromorphes en $r = -1$, en application de la proposition 2.16 Compte-tenu de la proposition 2.18 les transformées

$$|y|^r \underline{\mathbf{H}}_{2+r}[B_r], \qquad |y|^r W \underline{\mathbf{H}}_{2+r}[A_r](z,\bar{z})$$

sont bien définies sur $\widetilde{\mathcal{B}} \setminus \mathcal{D}$, et leur intégrale sur $\widetilde{\mathcal{B}} \setminus \mathcal{D}$ définit une fonction méromorphe en $r = -1$. Il en est de même de la somme de ces deux termes, et donc $|y|^r \mathbf{G}_{2+r,C}[1](z, \bar{z})$ est bien définie sur $\widetilde{\mathcal{B}} \setminus \mathcal{D}$, et son intégrale sur $\widetilde{\mathcal{B}} \setminus \mathcal{D}$ définit une fonction méromorphe en $r = -1$.

Le terme dominant dans le développement asymptotique est donné par l'intégrale du terme dominant dans le développement asymptotique de $|y|^r \mathbf{G}_{2+r,C}[1](z, \bar{z})$ en $r = 1$, qu'on calcule maintenant. Commençons par étudier

$$\mathbf{G}_{2+r,C}[1] = (\underline{I} - \underline{\mathbf{H}}_{2+r})^{-1} \circ W \underline{\mathbf{H}}_{2+r}[A_r] \sim_{r \to -1} \frac{1}{(r+1)} \frac{1}{h(\mathcal{E})} J\left[W_{(c)} \underline{\mathbf{H}}[A_r]\right] \underline{\psi},$$

où A_r est elle-même de la forme $(\underline{I} - \underline{\mathbf{H}}_{2+r})^{-1}[F]$, on a

$$A_r \sim_{r \to -1} \frac{1}{(r+1)} \frac{1}{h(\mathcal{E})} J[F] \underline{\psi} \qquad J\left[(W_{(c)} \underline{\mathbf{H}})[F_0]\right] \sim \frac{1}{(r+1)} \frac{1}{h(\mathcal{E})} J[F] J\left[(W_{(c)} \underline{\mathbf{H}})[\underline{\psi}]\right]$$

La proposition 2.14 prouve l'égalité $J\left[(W_{(c)} \underline{\mathbf{H}})[\underline{\psi}]\right] = \mathbb{E}[c]$. En tout, nous avons prouvé que

$$\mathbf{G}_{2+r,C}[F] \sim_{r \to -1} \frac{1}{(r+1)^2} \frac{\mathbb{E}(c)}{h(\mathcal{E})^2} J[F] \underline{\psi}. \tag{3.16}$$

Et donc, en intégrant,

$$\mathbb{E}_r(C) \sim \frac{1}{(r+1)^2} \frac{\mathbb{E}(c)}{h(\mathcal{E})} J[1] \iint_{\widetilde{\mathcal{B}} \setminus \mathcal{D}} \frac{1}{y} \underline{\psi}(z, \bar{z}) dx dy = \frac{1}{(r+1)} \frac{\mathbb{E}(c)}{h(\mathcal{E})}.$$

Coût D. Nous procédons de manière analogue, en décomposant l'opérateur $|y|^r \mathbf{G}_{2+r,D}$ en deux termes, de la forme

$$|y|^r \mathbf{G}_{2+r,D}[F](z, \bar{z}) = |y|^r \underline{\mathbf{H}}_{2+r}[C_r] + |y|^r W \underline{\mathbf{H}}_{2+r}[D_r](z, \bar{z})$$

où

$$C_r(z, \bar{z}) = \mathbf{G}_{2+r,D}[1](z, \bar{z}), \quad D_r = (\underline{I} - \underline{\mathbf{H}}_{2+r})^{-1} \circ \Delta \underline{\mathbf{H}}_{2+r} \circ (\underline{I} - \underline{\mathbf{H}}_{2+r})^{-1}[F](z, \bar{z}).$$

La preuve est la même que précedemment et montre que $|y|^r \mathbf{G}_{2+r,D}[1](z, \bar{z})$ est bien définie sur $\widetilde{\mathcal{B}} \setminus \mathcal{D}$, et que son intégrale sur $\widetilde{\mathcal{B}} \setminus \mathcal{D}$ définit une fonction méromorphe en $r = -1$. On calcule maintenant le terme dominant dans le développement asymptotique. Commençons par étudier

$$\Delta \underline{\mathbf{H}}_{2+r} \circ (\underline{I} - \underline{\mathbf{H}}_{2+r})^{-1}[F] \sim \Delta \underline{\mathbf{H}} \left[\frac{1}{(r+1)} \frac{1}{h(\mathcal{E})} J[F] \underline{\psi}\right] = \frac{1}{(r+1)} \frac{1}{h(\mathcal{E})} J[F] \left(\Delta \underline{\mathbf{H}}\right)[\underline{\psi}].$$

En réutilisant le résultat obtenu pour C, nous trouvons

$$(\underline{I} - \underline{\mathbf{H}}_{2+r})^{-1} \circ W_{(\ell)}[\underline{\mathbf{H}}_{2+r}] \circ (\underline{I} - \underline{\mathbf{H}}_{2+r})^{-1}[\Delta_s \underline{\mathbf{H}}[\underline{\psi}]] \sim \frac{1}{(r+1)^2} \frac{\mathbb{E}(\ell)}{h(\mathcal{E})^2} J[\Delta \underline{\mathbf{H}}[\underline{\psi}]] \underline{\psi}$$

et d'après 2.13, nous avons $J[\Delta \underline{\mathbf{H}}[\underline{\psi}]] = h(\mathcal{E})$. En conclusion, nous avons prouvé que

$$\mathbf{G}_{2+r,D}[F] \sim_{r \to -1} \frac{1}{(r+1)^3} \frac{\mathbb{E}(\ell)}{h(\mathcal{E})^2} J[F] \underline{\psi}. \tag{3.17}$$

Et, en intégrant,

$$\mathbb{E}_r(D) \sim (r+1) \cdot \frac{1}{(r+1)^3} \frac{\mathbb{E}(\ell)}{h(\mathcal{E})^2} J[1] \iint_{\widetilde{\mathcal{B}} \setminus \mathcal{D}} \frac{1}{y} \underline{\psi}(z, \bar{z}) dx dy = \frac{1}{(r+1)^2} \frac{\mathbb{E}(\ell)}{h(\mathcal{E})}.$$

3.5 Etude dans le modèle discret.

La preuve des deux théorèmes **C** et **D** possède une grande partie commune, correspondant aux principes, et aux lemmes qui donnent les estimations cherchées. Après avoir effectué ces calculs communs, nous présenterons la preuve spécifique de chaque théorème.

3.5.1 Cadre général.

Nous rappelons d'abord que, pour tout $z \in \Omega_N$, le coût B se décompose en trois termes,

$$B(z) = Q(z)\ell(N) + D(z) + (\underline{D}(z) - D(z)),$$

Par ailleurs, la variable $\underline{D} - D$ satisfait $|\underline{D}(z) - D(z)| \leq 2Q(z)$ et donc

$$|B(z) - \ell(N)Q(z) - D(z)| \leq 2\mathbb{E}_{r,N}(Q).$$

Il est donc suffisant détudier les variables Q et D.

Le disque $\mathcal{D}$ est réunion des domaines $h(\tilde{\mathcal{B}} \setminus \mathcal{D})$ pour $h \in \mathcal{H}^+$, et les variables C, Q, D ont une définition locale qui dépend du domaine. Il y a deux remarques. La variable générique à étudier, désignée par X et pouvant varier dans l'ensemble $\{\mathrm{Id}, C, D\}$ est une fonction définie par morceaux sur chaque domaine $h(\widetilde{\mathcal{B}} \setminus \mathcal{D})$. Donc, même dans le modèle continu, le coût X n'est pas continu, et il faut éventuellement étudier ses discontinuités sur les frontières des domaines $h(\widetilde{\mathcal{B}} \setminus \mathcal{D})$. Rappelons que toutes les frontières des domaines $h(\widetilde{\mathcal{B}} \setminus \mathcal{D})$ sont de la forme $h(\mathcal{C})$ où $\mathcal{C}$ est la circonférence qui borde le disque $\mathcal{D}$.

L'étude dans le modèle discret apporte des complications supplémentaires, car la densité elle-même, mais aussi les coûts dans le modèle discret deviennent discontinus, même à l'intérieur des domaines $h(\widetilde{\mathcal{B}} \setminus \mathcal{D})$ où la version des coûts dans le modèle continu et la densité étaient continues.

Rappelons que Ω_N est l'ensemble des points

$$\omega = \frac{a}{N} + i\frac{b}{N}, \qquad \omega \in \mathcal{D},$$

et qu'on associe à chaque $\omega \in \Omega_N$ un carré c_ω de centre ω et de côté $1/N$. Alors, les coûts à étudier X ou la densité f_r ont deux versions : une version dans le modèle continu, et une version dans le modèle discret, définie par :

$$f_{r,N}(z) = f_r(\omega), \qquad X_N(z) = X(\omega) \qquad \text{pour tout } z \in c_\omega.$$

Compte-tenu de l'importance de la partition topologique $h(\widetilde{\mathcal{B}} \setminus \mathcal{D})$ pour la définition des coûts, on introduit un découpage de Ω_N en trois domaines.

$$\Omega_N^{(\mathrm{b})} = \{w \in \Omega_N; \quad c_\omega \cap \mathcal{C} \neq \emptyset\}, \qquad \Omega_N^{(\mathrm{f})} = \{w \in \Omega_N; \quad \exists\, h \in \mathcal{H}^+, c_\omega \cap h(\mathcal{C}) \neq \emptyset\}, \tag{3.18}$$

$$\Omega_N^{(\mathrm{i})} = \Omega_N \setminus \left(\Omega_N^{(\mathrm{b})} \cup \Omega_N^{(\mathrm{f})}\right) = \{\omega \in \Omega_N; \quad \exists\, h \in \mathcal{H}^+, c_\omega \subset h(\widetilde{\mathcal{B}} \setminus \mathcal{D})\}. \tag{3.19}$$

On considère les domaines du modèle continu qui leurs sont associés : Ce sont le domaine total $\mathcal{D}_N$, le domaine des frontières $\mathcal{D}^{(\mathrm{f})}$, le domaine des points intérieurs $\mathcal{D}_N^{(\mathrm{i})}$ et enfin le domaine du bord $\mathcal{D}_N^{(\mathrm{b})}$, définis comme suit :

$$\mathcal{D}_N = \bigcup_{\omega \in \Omega_N} c_\omega; \qquad \mathcal{D}_N^{(\mathrm{f})} = \bigcup_{\omega \in \Omega_N^{(\mathrm{f})}} c_\omega\,, \qquad \mathcal{D}_N^{(\mathrm{i})} = \bigcup_{\omega \in \Omega_N^{(\mathrm{i})}} c_\omega\,, \qquad \mathcal{D}_N^{(\mathrm{b})} = \bigcup_{\omega \in \Omega_N^{(\mathrm{b})}} c_\omega\,. \tag{3.20}$$

3.5.2 Début de la preuve.

Nous introduisons les intégrales suivantes, dans le modèle continu et dans le modèle discret,

$$I[Y, \mathcal{X}] = \iint_{\mathcal{X}} Y(z) dxdy,$$

et pour un domaine $\mathcal{X}_N$ qui est une réunion de carrés c_ω pour lesquels $\omega \in \mathcal{X}$,

$$I_N[Y, \mathcal{X}_N] = \iint_{\mathcal{X}_N} Y_N(z) dxdy = \sum_{\omega \in \mathcal{X}} Y(\omega) f_r(\omega).$$

Les espérances à étudier s'écrivent

$$\mathbb{E}_r[X] := \frac{I[Xf_r, \mathcal{D}]}{I[f_r, \mathcal{D}]}, \qquad \mathbb{E}_{r,N}[X] := \frac{I_N[Xf_r, \mathcal{D}_N]}{I_N[f_r, \mathcal{D}_N]},$$

et la différence entre les espérances s'écrit

$$\left| \frac{I[Xf_r, \mathcal{D}]}{I[f_r, \mathcal{D}]} - \frac{I_N[Xf_r, \mathcal{D}_N]}{I_N[f_r, \mathcal{D}_N]} \right|$$

$$= \frac{1}{I[f_r, \mathcal{D}]\, I_N[f_r, \mathcal{D}_N]} \left| I[Xf_r, \mathcal{D}]\, I_N[f_r, \mathcal{D}_N] - I[f_r, \mathcal{D}]\, I_N[Xf_r, \mathcal{D}_N] \right|$$

$$\leq \frac{1}{I[f_r, \mathcal{D}]\, I_N[f_r, \mathcal{D}_N]} I[Xf_r, \mathcal{D}]\ |I_N[f_r, \mathcal{D}_N] - I[f_r, \mathcal{D}]| + I[f_r, \mathcal{D}]\ |I_N[Xf_r, \mathcal{D}_N] - I[Xf_r, \mathcal{D}]| \tag{3.21}$$

Il est donc suffisant, pour chaque coût X, d'évaluer la différence

$$I_N[Xf_r, \mathcal{D}_N] - I[Xf_r, \mathcal{D}].$$

Commençons donc par le cas du coût identité.

Lemme 3.1. *Les intégrales relatives à la variable identité vérifient*

$$I[f_r; \mathcal{D}_N] = \Theta\left(\frac{1}{r+1}\right)\left[1 - \frac{1}{N^{r+1}}\right] \qquad I[f_r; \mathcal{D}] = \Theta\left(\frac{1}{r+1}\right) \tag{3.22}$$

et leur différence vérifie

$$|I[f_r; \mathcal{D}_N] - I[f_r; \mathcal{D}]| = \frac{1}{N^{r+1}} O\left(\frac{1}{r+1}\right). \tag{3.23}$$

La différence symétrique entre les deux ensembles $\mathcal{D}$ et $\mathcal{D}_N$ sur lesquels nous intégrons,

$$\mathcal{D} \Delta \mathcal{D}_N = \mathcal{A}_N \cup \mathcal{D}_N^{(\mathrm{b})}$$

fait intervenir la bande horizontale $\mathcal{A}_N$ autour de l'axe $y = 0$ et le bord extérieur $\mathcal{D}_N^{(\mathrm{b})}$ de $\mathcal{D}$. Il faut donc estimer donc chacune des trois expressions

$$I[Xf_r, \mathcal{A}_N], \qquad I_N[Xf_r, \mathcal{D}_N^{(\mathrm{b})}] \qquad I_N[Xf_r, \mathcal{D}_N \setminus \mathcal{D}_N^{(\mathrm{b})}] - I[Xf_r, \mathcal{D}_N \setminus \mathcal{D}_N^{(\mathrm{b})}]. \tag{3.24}$$

Nous commencons par la première expression, puis nous nous concentrons sur la troisième expression et nous verrons que la deuxième expression va s'évaluer lors de la preuve.

3.5.3 Contribution des coûts au voisinage de l'axe – Un lemme utile.

Il faut donc travailler au voisinage de l'axe, et introduisant la bande horizontale $\mathcal{A}_N$ définie par

$$\mathcal{A}_N = \{z \in \mathcal{D} \ : \ |\Im(z)| < 1/(2N)\}, \tag{3.25}$$

Les fonctions Q et D sont définies par morceaux, sur chaque domaine $h(\widetilde{\mathcal{B}} \setminus \mathcal{D})$, et leur définition dépend de l'homographie h. Le premier travail consiste donc à préciser la position des domaines $h(\widetilde{\mathcal{B}} \setminus \mathcal{D})$ par rapport à la bande horizontale $\mathcal{A}_N$. Nous établissons ici un résultat qui décrit deux types d'homographies différents, selon que le domaine $h(\widetilde{\mathcal{B}} \setminus \mathcal{D})$ est inclus dans la bande $\mathcal{A}_N$ ou non.

Lemme 3.2. *Pour tout $N \geq 1$, on considère le sous-ensemble $\mathcal{H}_N^+$ de $\mathcal{H}^+$ défini comme suit :*

$$\mathcal{H}_N^+ := \{h \in \mathcal{H}^+ \ : \ h(\widetilde{\mathcal{B}} \setminus \mathcal{D}) \subset \mathcal{A}_N\}.$$

Si l'homographie h définie par $h(z) = (az+b)/(cz+d)$ est élément de $\mathcal{H}_N^+$, alors le couple (c,d) appartient au domaine $\mathcal{T}_N$ défini par

$$\mathcal{T}_N := \left\{(c,d); \quad |\theta(c,d)| \leq \frac{1}{\phi} \quad et \quad d^2 > N\phi\right\} \qquad avec \quad \theta(c,d) := \frac{c}{d}.$$

Si h n'appartient pas à $\mathcal{H}_N^+$, alors on a l'inégalité $d \leq N$, et aussi l'inclusion

$$\mathcal{H}^+ \setminus \mathcal{H}_N \subset \sum_{k=1}^{k_N} \mathcal{H}^k \qquad avec \quad k_N = O(\log N).$$

Démonstration. Soit $h \in \mathcal{H}^+$ telle que $h(z) = (az+b)/(cz+d)$. La contrainte $c/d| \in]0, 1/\phi$ est liée à la caractérisation de Hurwitz (1.13). Par ailleurs, l'ensemble $h(\widetilde{\mathcal{B}} \setminus \mathcal{D})$ est inclus dans $\mathcal{A}_N$ si et seulement si le diamètre du grand disque est strictement plus petit que $1/N$. L'expression du grand diamètre est donnée dans la proposition 1.5 en fonction de (c,d). La condition d'inclusion dans $\mathcal{A}_N$ devient alors :

$$\frac{1}{cd} < \frac{1}{N} \quad \text{si} \quad c > 0, \quad \text{et} \quad \frac{2}{|c|(2d+c)} < \frac{1}{N} \quad \text{si} \quad c < 0. \tag{3.26}$$

Cette condition (3.26) entraîne toujours la ondition $|c|d > N$. Cela est évident lorsque $c > 0$. Lorsque $c < 0$, l'inégalité $2d + c < 2d$ montre que la condition (3.26) implique aussi $|c|d > N$. On conclut en remarquant que la condition $|c|d > N$ implique la condition $d^2 > N\phi$. □

3.5.4 Contribution au voisinage de l'axe – Le résultat.

Nous donnons maintenant une estimation des intégrales au voisinage de l'axe.

Proposition 3.4. *Pour toute valuation $r > -1$ fixée, et pour $X \in \{\mathrm{Id}, C, D\}$, on a*

$$I[Xf_r, \mathcal{A}_N] = \iint_{\mathcal{A}_N} X(z)\, f_r(z) dxdy = O\left(\frac{1}{r+1}\right) \frac{(\log N)^{e(X)}}{N^{r+1}}$$

où l'exposant $e(X)$ dépend du coût : $e(\mathrm{Id}) = 0$, $e(C) = 1$ et $e(D) = 2$. Les variables C et D sont donc intégrables sur le disque $\mathcal{D}$ par rapport à la densité de valuation $r > -1$.

Démonstration. L'intégrale se décompose en deux sommes,

$$I[Xf_r, \mathcal{A}_N] = \sum_{h\in\mathcal{H}_N^+} I[Xf_r, h(\widetilde{\mathcal{B}}\setminus\mathcal{D})] + \sum_{h\in\mathcal{H}^+\setminus\mathcal{H}_N^+} I[Xf_r, h(\widetilde{\mathcal{B}}\setminus\mathcal{D})\cap\mathcal{A}_N], \tag{3.27}$$

et la deuxième se majore aisément, car pour une homographie qui n'est pas dans $\mathcal{H}_N^+$, le coût X satisfait $X(z) \le (\log N)^{e(X)}$ et donc

$$I[Xf_r, h(\widetilde{\mathcal{B}}\setminus\mathcal{D})\cap\mathcal{A}_N] \le (\log N)^{e(X)}\, I[f_r, h(\widetilde{\mathcal{B}}\setminus\mathcal{D})\cap\mathcal{A}_N],$$

ainsi,

$$\sum_{h\in\mathcal{H}^+\setminus\mathcal{H}_N^+} I[Xf_r, h(\widetilde{\mathcal{B}}\setminus\mathcal{D})\cap\mathcal{A}_N] \le (\log N)^{e(X)}\, I[f_r, \mathcal{A}_N] \le \frac{1}{r+1}\frac{(\log N)^{e(X)}}{(2N)^{r+1}}, \tag{3.28}$$

où on a majoré la mesure de $\mathcal{A}_N$ par celle du rectangle qui l'enveloppe.

Étudions maintenant la première somme de (3.27). Le terme général se majore à l'aide du lemme 3.2, qui nous fournit

$$I[Xf_r, h(\widetilde{\mathcal{B}}\setminus\mathcal{D})] \le (\log d)^{e(X)}\, I[f_r, h(\widetilde{\mathcal{B}}\setminus\mathcal{D})].$$

Par ailleurs, la proposition (1.7) estime les mesures des domaines $h(\widetilde{\mathcal{B}}\setminus\mathcal{D})$,

$$I[f_r, h(\widetilde{\mathcal{B}}\setminus\mathcal{D})] \le L\cdot\left(\frac{1}{|\theta|d^2}\right)^{r+2} |\theta||\log|\theta||,$$

tandis que le lemme 3.2 montre qu'il suffit de sur les couples $(c,d)\in\mathcal{T}_N$, pour obtenir la majoration

$$\sum_{h\in\mathcal{H}_N^+} I[Xf_r, h(\widetilde{\mathcal{B}}\setminus\mathcal{D})] \le LS_N \qquad \text{avec} \quad S_N := \sum_{(c,d)\in\mathcal{T}_N} \frac{1}{(|\theta|d^2)^{r+2}}(\log d)^{e(X)}|\theta||\log|\theta||. \tag{3.29}$$

Il s'agit maintenant d'évaluer cette somme S_N en la comparant à une intégrale. Travaillant d'abord pour d fixe, on remarque que la somme de Riemann

$$\frac{1}{d}\sum_{|\theta|\le 1/\phi} f(\theta) \qquad \text{avec} \quad f(\theta) := \frac{1}{|\theta|^{r+1}}|\log|\theta||$$

est majorée par l'intégrale impropre convergente, et donc (pour $r < -1/2$)

$$J_r := 2\int_0^{1/\phi} \frac{|\log|\theta||}{\theta^{r+1}}d\theta \le J := 2\int_0^{1/\phi} \frac{|\log|\theta||}{\sqrt{\theta}}d\theta$$

car la fonction f est décroissante. Donc S_N est majorée par la somme

$$S_N \le J\sum_{d>\lfloor\sqrt{N\phi}\rfloor} \frac{1}{d^{2r+3}}(\log d)^{e(X)}$$

qu'on peut aussi majorer par l'intégrale correspondante, car la fonction considérée est elle aussi décroissante. Finalement, on a montré que

$$\sum_{h\in\mathcal{H}_N^+} I[Xf_r, h(\widetilde{\mathcal{B}}\setminus\mathcal{D})] \le LJ\int_{\lfloor\sqrt{N\phi}\rfloor} \frac{1}{y^{2r+3}}(\log y)^{e(X)}dy$$

Cette dernière intégrale se calcule aisément par parties et on trouve

$$\sum_{h\in\mathcal{H}_N^+} I[Xf_r, h(\widetilde{\mathcal{B}}\setminus\mathcal{D})] = O\left(\frac{1}{r+1}\right)\frac{(\log N)^{e(X)}}{N^{r+1}}. \tag{3.30}$$

ce qui avec la majoration de (3.28) termine la preuve de la proposition. □

Nous considérons maintenant la troisième expression définie en (3.24), et relative aux points intérieurs.

3.5.5 Contribution des points intérieurs.

Nous travaillons maintenant avec les points intérieurs du domaine $\mathcal{D}_N^{(\mathrm{b})}$ défini en (3.18) et (3.20) et montrons l'estimation suivante.

Lemme 3.3. *Considérons un coût X qui peut être un coût additif C, ou la variable D. La différence entre les intégrales du modèle discret et du modèle continu vérifie*

$$I_N[Xf_r, \mathcal{D}_N\setminus\mathcal{D}_N^{(b)}] - I[Xf_r, \mathcal{D}_N\setminus\mathcal{D}_N^{(b)}] =$$

$$O(1)\frac{(\log N)^{e(X)+1}}{N} \quad (si\ r\geq 0) \qquad O(1)\frac{(\log N)^{e(X)+1}}{N^{r+1}} \quad (si\ -1<r\leq 0)$$

Démonstration. La fonction à évaluer est $|f_r(z)X(z) - f(\omega)X(\omega)|$ se majore sous la forme

$$|f_r(z)X(z) - f(\omega)X(\omega)| \leq G(z) + L(z)$$

avec

$$G(z) = X(\omega)|f_r(z) - f_r(\omega)|, \qquad L(z) = f_r(z)|X(z) - X(\omega)|,$$

et il faut en chercher son maximum sur un carré c_ω de centre $\omega = (x_\omega, y_\omega)$ et de côté $1/(2N)$. Nous estimons les deux termes $G(z)$ et $L(z)$ séparément :

Terme $G(z)$. Le premier terme $G(z)$ est majoré par $(\log N)^{e(X)}|f_r(z)-f_r(\omega)|$. il est donc suffisant d'étudier la différence $|f_r(z) - f_r(\omega)|$. Grâce à la symétrie des fonctions intégrées et du disque $\mathcal{D}_N$ par rapport à l'axe des abscisses, nous pouvons nous restreindre à travailler sur $\mathcal{D}_N^+$. Le cas $r = 0$ est trivial puisque dans ce cas la fonction f_r est constante. On suppose donc $r \neq 0$ dans la suite et alors, pour tout $z \in c_\omega$, on a

$$|f_r(z) - f_r(\omega)| \leq |z-\omega|\cdot\sup\{||\Delta f_r(u,v)||;\ (u,v)\in c_\omega\},$$

où la différentielle $\Delta f_r(u,v)$ satisfait $\Delta f_r(u,v) = rv^{r-1}$. On a toujours, pour tout couple (y,v) d'ordonnées de points de c_ω, la relation

$$v \leq y_\omega + \frac{1}{2N} \leq 3y,$$

ce qui entraîne,

$$(\text{pour } r>1)\quad v^{r-1}\leq 3^{r-1}y^{r-1}, \qquad (\text{pour } r\leq 1)\quad v^{r-1}\leq\left(\frac{1}{3}\right)^{r-1}y^{r-1}$$

Dans tous les cas, on a donc

$$|f_r(z) - f_r(\omega)| \ = O(1)\frac{1}{N}|r|y^{r-1}.$$

En intégrant sur $\mathcal{D}_N^+$, il y a de nouveau deux cas

$$\iint_{\mathcal{D}_N^+} |r|y^{r-1}dxdy = O(1) \quad (\text{si } r > 0) \qquad \iint_{\mathcal{D}_N^+} |r|y^{r-1}dxdy = O\left(\frac{1}{N^r}\right) \quad (\text{si } r < 0)/$$

L'intégrale $I[G, \mathcal{D}_N \setminus \mathcal{D}_N^{(\mathrm{b})}]$ admet une estimation de la forme

$$O(1)\frac{(\log N)^{e(X)}}{N} \quad (\text{si } r > 0) \qquad O(1)\frac{(\log N)^{e(X)}}{N^{r+1}} \quad (\text{si } -1 < r < 0) \tag{3.31}$$

Terme $L(z)$. Pour le second terme $L(z)$, il y a deux cas selon que z est élément de $\mathcal{D}_N^{(\mathrm{i})}$ (domaine des points intérieurs) ou élément de $\mathcal{D}_N^{(\mathrm{f})}$ (domaine des points frontières). Ces domaines sont définis en (3.18) et (3.20) .

Premier cas $z \in \mathcal{D}_N^{(i)}$. Puisque les coûts Id et C sont constants dans $h(\widetilde{\mathcal{B}} \setminus \mathcal{D})$, la question se pose seulement pour le coût $X = D$ et la variation $|D(z) - D(\omega)|$ vérifie

$$|D(z) - D(\omega)| \leq |z - \omega| \cdot \sup\{||\Delta D(u, v)||;\ (u, v) \in c_\omega\},$$

où la différentielle ΔD satisfait pour $z \in h(\widetilde{\mathcal{B}} \setminus \mathcal{D})$, l'inégalité

$$\Delta D(u, v) = O(\log d)\frac{1}{\rho_h} \qquad \text{où } \rho_h \text{ est le diamètre du grand disque.}$$

La mesure $I[L, h(\widetilde{\mathcal{B}} \setminus \mathcal{D}) \cap \mathcal{A}_N^{\mathrm{c}}]$ se majore ainsi :

$$I[L, h(\widetilde{\mathcal{B}} \setminus \mathcal{D}) \cap \mathcal{A}_N^{\mathrm{c}}] \leq \sup\{f_r(z;\ \ z \in h(\widetilde{\mathcal{B}} \setminus \mathcal{D}) \cap \mathcal{A}_N^{\mathrm{c}}\} \cdot \frac{\log d}{\rho_h} I[\mathrm{Id}, h(\widetilde{\mathcal{B}} \setminus \mathcal{D})] \tag{3.32}$$

Lorsque $r \geq 0$, la densité f_r est $O(1)$. Lorsque $r < 0$, la fonction $(x, y) \mapsto y^r$ atteint son maximum lorsque $|y| = 1/2N$ et

$$|f_r(z)| = O\left(\frac{1}{N^r}\right).$$

Par ailleurs, on a $I[\mathrm{Id}, h(\widetilde{\mathcal{B}} \setminus \mathcal{D})] = O(\rho_h^2)$ et finalement,

$$I[L, h(\widetilde{\mathcal{B}} \setminus \mathcal{D}) \cap \mathcal{A}_N^{\mathrm{c}}] = O\left(\frac{1}{N}\right)\rho_h \log d \quad (\text{si } r \geq 0)$$

$$I[L, h(\widetilde{\mathcal{B}} \setminus \mathcal{D}) \cap \mathcal{A}_N^{\mathrm{c}}] = O\left(\frac{1}{N^r}\right)\rho_h \log d \quad (\text{si } -1 < r \leq 0)$$

Le domaine $\mathcal{D}_N^{(\mathrm{i})}$ vérifie

$$\mathcal{D}_N^{(\mathrm{i})} \subset \bigcup_{h \in \mathcal{H}^+ \setminus \mathcal{H}_N^+} h(\widetilde{\mathcal{B}} \setminus \mathcal{D}) \cap \mathcal{A}_N^{\mathrm{c}}] \qquad \text{et donc} \quad I[D, \mathcal{D}_N^{(\mathrm{i})}] = \sum_{h \in \mathcal{H}^+ \setminus \mathcal{H}_N^+} O(\rho_h \log d) = O(\log N)^2.$$

on obtient finalement

$$I[L,\mathcal{D}_N^{(\mathrm{i})}] = O(1)\frac{(\log N)^2}{N} \quad (\text{si } r \geq 0) \qquad I[L,\mathcal{D}_N^{(\mathrm{i})}] = O(1)\frac{(\log N)^2}{N^{r+1}} \quad (\text{si } -1 < r \leq 0) \quad (3.33)$$

Deuxième cas $z \in \mathcal{D}_N^{(\mathrm{f})}$. La différence $f_r(z)|X(z) - X(\omega)|$ est majorée par $(\log N)^{e(X)}|f_r(z)|$. Par ailleurs, le domaine $\mathcal{D}_N^{(\mathrm{f})}$ est inclus dans une reunion de couronnes bordant les circonférences $h(\mathcal{C})$: Désignons par ρ_h le rayon du disque $h(\mathcal{D})$ et par τ_h son centre, et considérons la couronne circulaire $\mathcal{C}_{(N,h)}$ de centre τ_h et de rayons $\rho_h - \sqrt{2}/N$ et $\rho_h + \sqrt{2}/N$,

$$\mathcal{C}_{(N,h)} = \left\{ z \in \mathbb{C} \ : \ \rho_h - \frac{\sqrt{2}}{N} \leq |z - \tau_h| < \rho_h + \frac{\sqrt{2}}{N}, \quad |\Im(z)| > \frac{1}{N} \right\}.$$

Comme le diamètre de c_ω est au plus égal à $\sqrt{2}/N$, on a alors l'inclusion

$$\mathcal{D}_N^{(\mathrm{f})} \subset \bigcup_{h \in \mathcal{H}^+ \setminus \mathcal{H}_N^+} \mathcal{C}_{(N,h)}.$$

La mesure $I[f_r, \mathcal{C}_{(N,h)}]$ se majore ainsi :

$$I[f_r, \mathcal{C}_{(N,h)}] \leq \sup\{f_r(z; \quad z \in \mathcal{C}_{(N,h)}\} \cdot I[\mathrm{Id}, \mathcal{C}_{(N,h)}]. \quad (3.34)$$

Lorsque $r \geq 0$, la densité f_r est $O(1)$. Lorsque $r < 0$, la fonction $(x, y) \mapsto y^r$ atteint son maximum lorsque $|y| = 1/2N$ et

$$|f_r(z)| = O\left(\frac{1}{N^r}\right).$$

Par ailleurs, il suffit de borner trivialement l'aire de la couronne tronquée par celle de la couronne non-tronquée. Nous avons

$$I[\mathrm{Id}, \mathcal{C}_{(N,h)}] \leq \pi \left[\left(\rho_h + \frac{\sqrt{2}}{N} \right)^2 - \left(\rho_h - \frac{\sqrt{2}}{N} \right)^2 \right], \qquad \text{et donc} \quad I[\mathrm{Id}, \mathcal{C}_{(N,h)}] = O\left(\frac{\rho_h}{N}\right)$$

Avec ces informations, (3.34) se traduit en

$$I[f_r, \mathcal{C}_{(N,h)}] = O\left(\frac{\rho_h}{N}\right) \quad (\text{si } r \geq 0) \qquad I[f_r, \mathcal{C}_{(N,h)}] = O\left(\frac{\rho_h}{N^{r+1}}\right) \quad (\text{si } -1 < r \leq 0)$$

En utilisant alors la deuxième partie du lemme 3.2, on déduit une estimation pour l'intégrale $I[L, \mathcal{D}_N^{(\mathrm{f})}]$ de la forme

$$I[L,\mathcal{D}_N^{(\mathrm{f})}] = O(1)\frac{(\log N)^{e(X)+1}}{N} \quad (\text{si } r \geq 0), \quad I[L,\mathcal{D}_N^{(\mathrm{f})}] = O(1)\frac{(\log N)^{e(X)+1}}{N^{r+1}} \quad (\text{si } -1 < r \leq 0) \quad (3.35)$$

En regroupant les résultats correspondant aux trois étapes de la preuve, donnant lieu aux trois estimations (3.31, 3.33, 3.35), on obtient bien le résultat cherché. □

3.5.6 Preuve du théorème C.

Nous allons établir donc le théorème **C**. La distance entre les deux espérances (continue et discrète) satisfait, d'après (3.21),

$$|\mathbb{E}_{r,N}(X) - \mathbb{E}_r(X)| \leq \frac{\mathbb{E}_r(X)}{I_N[f_r, \mathcal{D}_N]} \, |I_N[f_r, \mathcal{D}_N] - I[f_r, \mathcal{D}]| + I[f_r, \mathcal{D}] \, |I_N[Xf_r, \mathcal{D}_N] - I[Xf_r, \mathcal{D}]| \, .$$

Supposons à présent que r est fixe. Dans ce cas, et grâce au théorème **B** et au lemme 3.1, nous avons

$$\mathbb{E}_r(X) = \Theta(1), \quad I_N[f_r, \mathcal{D}_N] = \Theta(1), \quad I[f_r, \mathcal{D}] = \Theta(1), \quad |I_N[f_r, \mathcal{D}_N] - I[f_r, \mathcal{D}]| = O\left(\frac{1}{N^{r+1}}\right).$$

En plus, en regroupant les deux lemmes 3.4 et 3.3, on montre que

$$I_N[Xf_r, \mathcal{D}_N] - I[Xf_r, \mathcal{D}] = \frac{1}{r+1} \frac{(\log N)^{e(X)}}{N^{r+1}} O\left(\max\{1, (r+1)\log N\}\right) = O\left(\frac{(\log N)^{e(X)+1}}{N^{r+1}}\right),$$

ce qui implique que

$$|\mathbb{E}_{r,N}(X) - \mathbb{E}_r(X)| = O\left(\frac{(\log N)^{e(X)+1}}{N^{r+1}}\right),$$

d'où le résultat.

3.5.7 Preuve du théorème D.

La preuve du théorème **D** utilise les deux thèorèmes précédents **C** et **D**. D'après (3.21), nous avons une majoration de la différence entre les deux espérances, continue et discrète, de la forme

$$|\mathbb{E}_{r,N}(X) - \mathbb{E}_r(X)| \leq \mathbb{E}_r(X) \left(\frac{|I_N[f_r, \mathcal{D}_N] - I[f_r, \mathcal{D}]|}{I_N[f_r, \mathcal{D}_N]} + \frac{I[f_r, \mathcal{D}]}{\mathbb{E}_r(X)} \, |I_N[Xf_r, \mathcal{D}_N] - I[Xf_r, \mathcal{D}]| \right).$$

Le lemme 3.1 estime la mesure discrète $I_N[f_r, \mathcal{D}_N]$ et la différence entre les mesures, l'une discrète $I_N[f_r, \mathcal{D}_N]$ et l'autre continue $I[f_r, \mathcal{D}_N]$, et ce, de manière uniforme pour $N \to \infty$ et $r \to -1$,

$$I_N[f_r, \mathcal{D}_N] = \Theta\left(\frac{1}{(r+1)}\right)\left[1 - N^{-(r+1)}\right], \qquad |I[f_r; \mathcal{D}_N] - I[f_r; \mathcal{D}]| = \frac{1}{N^{r+1}} O\left(\frac{1}{r+1}\right).$$

et aboutit donc à l'estimation

$$\frac{|I_N[f_r, \mathcal{D}_N] - I[f_r, \mathcal{D}]|}{I_N[f_r, \mathcal{D}_N]} = O\left(\frac{1}{N^{r+1} - 1}\right).$$

Par ailleurs, le théorème **B**, ainsi que le lemme 3.1 estiment les deux espérances continues au voisinage de $r = 1$,

$$\mathbb{E}_r(X) = \Theta\left(\frac{1}{(r+1)^{e(X)}}\right), \quad I[f_r, \mathcal{D}] = \Theta\left(\frac{1}{r+1}\right),$$

alors que les deux lemmes 3.3 et X permettent à eux deux d'váluer la différence entre les espérances continues et discrètes, de manière uniforme pour $N \to \infty$ et $r \to -1$,

$$|I_N[Xf_r, \mathcal{D}_N] - I[Xf_r, \mathcal{D}]| = \frac{1}{r+1} \frac{(\log N)^{e(X)}}{N^{r+1}} O\left(\max\{1, (r+1)\log N\}\right).$$

On obtient donc en tout une estimation pour le terme

$$\frac{I[f_r,\mathcal{D}]}{\mathbb{E}_r(X)}\ |I_N[Xf_r,\mathcal{D}_N]-I[Xf_r,\mathcal{D}]| = \frac{[(r+1)(\log N)]^{e(X)}}{N^{r+1}} O\left(\max\{1,(r+1)\log N\}\right),$$

et on conclut à l'estimation finale, toujours uniforme pour $N\to\infty$ et $r\to -1$,

$$\mathbb{E}_{r,N}(X) = \mathbb{E}_r(X)\left[1+\frac{[(r+1)(\log N)]^{e(X)}}{N^{r+1}} O\left(\max\{1,(r+1)\log N\}\right) + O\left(\frac{1}{N^{r+1}-1}\right)\right].$$

Si maintenant, on se limite au cas où $(r+1)\log M$ est un $\Omega(1)$, alors on a

$$O\left(\max\{1,(r+1)\log N)\}\right) = O\left((r+1)\log N\right),$$

ce qui permet d'obtenir le résultat (3.6).

Etablissons maintenant les conséquences de (3.6). Supposons d'abord que $r+1=(\log N)^{-\alpha}$, avec $0<\alpha<1$. Puisque $N^{r+1}=\exp[(r+1)\log N]=\exp[(\log N)^{1-\alpha}]$ tend vers l'infini, on a alors

$$\frac{(\log N(r+1))^{e(X)+1}}{N^{r+1}}\to 0, \qquad \text{et} \qquad \left[\frac{1}{N^{r+1}-1}\right]\to 0,$$

et en conséquence

$$\mathbb{E}_{r,N}(X)\sim\mathbb{E}_r(X).$$

Avec cette dernière identité, le théorème **B** montre (3.7). Avec les identités (3.7), on vérifie de plus les équivalents asymptotiques

$$\mathbb{E}_{r,N}(C)\sim 2\frac{\mathbb{E}[c]}{h(\mathcal{E})}(\log N)^{\alpha}, \qquad \mathbb{E}_{r,N}(D)\sim\frac{\mathbb{E}[\ell]}{h(\mathcal{E})}(\log N)^{2\alpha},$$

On peut alors en déduire des informations précises sur la complexité binaire B. L'égalité (2.8), page 38, avec l'encadrement

$$0\le[\underline{D}(u,v)-D(u,v)]\le 2Q(u,v),$$

impliquent la relation $\mathbb{E}_{r,N}(B)=(2\log N)\,\mathbb{E}_{r,N}(Q)+\mathbb{E}_{r,N}(D)+O(\mathbb{E}_{r,N}(Q))$. Et, alors, les équivalents précédents, permettent de conclure à

$$\mathbb{E}_{r,N}(B)\sim 2\frac{\mathbb{E}[\ell]}{h(\mathcal{E})}(\log N)^{1+\alpha} \tag{3.36}$$

Par ailleurs, lorsque $(r+1)\log N=\Theta(1)$, les estimations

$$\frac{(\log N(r+1))^{e(X)+1}}{N^{r+1}}=\Theta(1), \qquad \text{et} \qquad \left[\frac{1}{N^{r+1}-1}\right]=\Theta(1),$$

montrent que

$$\mathbb{E}_{r,N}(X)=\Theta(\mathbb{E}_r(X))=\Theta(\log N)^2)$$

Cela termine la preuve du dernier théorème de ce chapitre.

Nous avons terminé la présentation de nos résultats concernant l'analyse de l'exécution de l'algorithme de Gauss. Nous nous tournons maintenant vers l'étude probabiliste de la configuration de sortie.

Troisième partie

Analyses de l'algorithme de Gauss : Étude probabiliste de la configuration de sortie

Chapitre 1

Étude géométrique de la configuration de sortie.

Sommaire

Ce chapitre, le premier de cette partie III, décrit les ensembles de niveau des trois principaux paramètres qui décrivent la configuration de l'algorithme de Gauss, le premier minimum λ, le deuxième minimum orthogonalisé μ, et le défaut d'Hermite γ. On commence donc par une description précise de ces ensembles de niveau relatifs aux trois variables λ, γ, μ, qui fait intervenir, de manière naturelle, des objets classiques dans la géométrie du demi-plan de Poincaré, comme les suites de Farey et les disques de Ford.

Tout au long de cette partie, les parties réelles et imaginaires du complexe z (resp. $\hat{z}$) seront notées x et y (resp. $\hat{x}$ et $\hat{y}$). Le complexe $\hat{z}$ représentera toujours la sortie de l'algorithme Gauss-positif sous entrée z, soit $\hat{z} = \text{Gauss-positif}(z)$. Les événements, au sens probabiliste du terme, seront notés entre crochets, et les prédicats qui les définissent porteront toujours sur l'élément générique $z = x + iy \in \mathcal{B} \setminus \mathcal{F}$ et/ou ses parties réelles ou imaginaires x et y respectivement. Ainsi,

$$[\mathcal{P}(x, y, z)] = \{z = x + iy \in \mathcal{B} \setminus \mathcal{F} \mid \mathcal{P}(x, y, z)\}.$$

Notre objectif dans cette section est de caractériser les ensembles de niveau reliés aux trois principaux paramètres géométriques λ, μ, γ,

$$G(\rho) = [\gamma(z) \leq \rho], \tag{1.1}$$
$$L(t) = [\lambda(z) \leq t], \tag{1.2}$$
$$M(u) = [\mu(z) \leq u]. \tag{1.3}$$

Cela nous permettra, dans le prochain chapitre de cette partie III, de calculer leur fonction de répartition de γ, λ et μ.

1.1 Caractérisation des ensembles de niveau des trois paramètres.

Nous montrons que chacun des trois domaines de niveau peut s'écrire sous la forme $[\gamma(z) \leq f(z)]$ pour une fonction f, qui dépend de chaque paramètre étudié. Après avoir déterminé précisément la fonction f, on particularise le résultat à chacun des ensembles originaux et on en étudie la géométrie en détail. Nous allons retrouver la géométrie de la partition topologique sous-jacente au système dynamique Gauss-positif, vue dans le chapitre 1 de la partie II.

1.1.1 Expressions complexes des trois paramètres λ, μ, γ.

La proposition suivante donne les expressions des trois paramètres géométriques de sortie $\gamma(z)$, $\lambda(z)$ et $\mu(z)$ en fonction de z. Ces paramètres ont été définis de manière vectorielle dans les équations (2.11), (2.12) (partie I), et le passage du vocabulaire vectoriel au vocabulaire complexe est décrit dans le chapitre 1 de la partie II.

Proposition 1.1. *Considérons l'algorithme* Gauss-positif *avec une entrée $z = x + iy \in \mathcal{B} \setminus \mathcal{F}$ une sortie $\hat{z} = \hat{x} + i\hat{y} \in \mathcal{F}$, vérifiant donc $\hat{z} = \text{Gauss-positif}(z)$. Alors, les trois paramètres $\gamma(z), \lambda(z), \mu(z)$ vérifient*

$$\gamma(z) = \frac{1}{\hat{y}}, \quad \lambda^2(z) = \frac{y}{\hat{y}}, \quad \mu^2(z) = y\hat{y}.$$

De plus, les inclusions suivantes sont satisfaites

$$[\lambda(z) \leq t] \supset \left[\Im(z) \leq \frac{\sqrt{3}}{2} t^2\right], \qquad [\mu(z) \leq u] \subset \left[\Im(z) \leq \frac{2}{\sqrt{3}} u^2\right]. \tag{1.4}$$

Démonstration. Le déterminant du réseau $\mathcal{L}(z)$, engendré par $(1, z)$, et son défaut d'Hermite $\gamma(z)$ vérifient

$$\det \mathcal{L}(z) = \lambda(z)\mu(z) \qquad \text{et} \qquad \gamma(z) = \frac{\lambda^2(z)}{\det \mathcal{L}(z)} = \frac{\lambda(z)}{\mu(z)}. \tag{1.5}$$

Or, le déterminant du réseau $\mathcal{L}(z)$ est égal à l'aire du parallellograme déterminé par 1 et z, dont la base est de longueur 1 et la hauteur vaut y. Ainsi,

$$\lambda(z)\mu(z) = \det \mathcal{L}(z) = y.$$

Par ailleurs, puisque le défaut d'Hermite est invariant par similitude, nous avons

$$\gamma(z) = \gamma(\hat{z}).$$

La base $(1, \hat{z})$ est par définition réduite, son premier minimum $\lambda(\hat{z})$ est égal à 1 et son deuxième minimum orthogonalisé est égal $\mu(\hat{z})\ \hat{y}$. Nous déduisons donc que

$$\gamma(z) = \gamma(\hat{z}) = \frac{\lambda(\hat{z})}{\mu(\hat{z})} = \frac{1}{\hat{y}}. \tag{1.6}$$

Finalement, avec (1.5) et (1.6), nous obtenons

$$\lambda^2(z) = \det \mathcal{L}(z) \cdot \gamma(z) = y \cdot \gamma(z) = \frac{y}{\hat{y}} \qquad \mu^2(z) = \frac{\det \mathcal{L}(z)}{\gamma(z)} = \frac{y}{\gamma(z)} = y\hat{y},$$

comme souhaité. Les inclusions (1.4) se prouvent en utilisant la relation $\hat{y} \geq \sqrt{3}/2$, conséquence directe de $\hat{z} \in \mathcal{F}$ et de la forme de $\mathcal{F}$; on observe alors que

$$[\lambda^2(z) > t^2] = [y/\hat{y} > t^2] = [y > t^2 \cdot \hat{y}] \subset [y > t^2\sqrt{3}/2]$$

$$[\mu^2(z) \leq u^2] = [y\hat{y} \leq u^2] = [y \leq u^2/\hat{y}] \subset [y \leq \frac{2}{\sqrt{3}}u^2].$$

Cela achève la preuve de la proposition. □

À l'aide de cette proposition 1.1, nous montrons qu'il est possible de mener une étude générale, valable pour chacun des trois paramètres.

1.1.2 Principe d'une étude commune.

Cette section montre que l'étude des ensembles $G(\rho)$, $L(t)$ et $M(u)$ se ramène à l'étude de l'événement $[\gamma(z) \leq f(z)]$, pour une fonction f bien choisie.

Proposition 1.2. *Les ensembles de niveau $G(\rho)$, $L(t)$ et $M(u)$ s'écrivent*

$$G(\rho) = [\gamma(z) \leq \rho] \quad = \quad [\gamma(z) \leq f(z,\rho)] \qquad \text{avec } f(z,\rho) = \rho \tag{1.7}$$

$$L(t) = [\lambda(z) \leq t] \quad = \quad [\gamma(z) \leq f(z,t)] \qquad \text{avec } f(z,t) = \frac{t^2}{y} \tag{1.8}$$

$$M(u) = [\mu(z) \leq u] \quad = \quad (\mathcal{B} \setminus \mathcal{F}) \setminus [\gamma(z) < f(z,u)] \qquad \text{avec } f(z) = \frac{y}{u^2}. \tag{1.9}$$

Ainsi, l'étude des ensembles de niveau des deux paramètres λ, μ se ramène à celle de $[\gamma(z) \leq f(z)]$.

Démonstration. La proposition 1.1 entraîne les équivalences

$$\mu(z) > u \Longleftrightarrow \gamma(z) < \frac{y}{u^2}, \qquad \text{et} \qquad \lambda(z) \leq t \Longleftrightarrow \gamma(z) \leq \frac{t^2}{y},$$

ce qui établit les relations (1.7), (1.8) et (1.9). □

Dans ce qui suit, nous étudions l'ensemble $[\gamma(z) \leq f(z)]$, pour $f : \mathcal{B} \setminus \mathcal{F} \to \mathbb{R}$ une fonction quelconque. On l'appelera ensemble de f-niveau pour $\gamma(z)$. La caractérisation de $[\gamma(z) \leq f(z)]$ nous fournira directement celle de $G(\rho)$, $L(t)$ et $M(u)$.

1.1.3 Caractérisation locale commune.

L'étude locale de $[\gamma(z) \leq f(z)]$ consiste à partitionner cet ensemble $[\gamma(z) \leq f(z)]$ et à en étudier chacun des "morceaux" séparément. D'après la proposition 1.2 (partie II), toute homographie $h \in \mathcal{G}$ s'écrit $h = h_{(a,c)} \circ T^q$, où $h_{(a,c)} \in \mathcal{G}$ est l'unique homographie telle que

$$h_{(a,c)}(z) = \frac{az + b_0}{cz + d_0} \quad \text{avec} \quad |b_0| \leq \frac{|a|}{2},\ |d_0| \leq \frac{|c|}{2}, \tag{1.10}$$

avec (a, c) appartenant à l'ensemble

$$\mathcal{C} = \{(a, c) \in \mathbb{Z} \times \mathbb{N} \mid \frac{a}{c} \in [-1/2, 1/2], \text{pgcd}(a, c) = 1\}, \tag{1.11}$$

et T la transformée $T : z \mapsto z + 1$. Comme déjà défini dans la section 1.2.3 (partie II), le feston $\mathcal{F}_{(a,c)}$ est alors la réunion

$$\mathcal{F}_{(a,c)} = h_{(a,c)} \circ \left(\bigcup_{q \in \mathbb{Z}} T^q(\mathcal{F}) \right) \tag{1.12}$$

des transformés de $\mathcal{F}$ par des homographies relatives au même couple (a, c). On obtient alors la décomposition

$$[\gamma(z) \leq f(z)] = \bigcup_{(a,c) \in \mathcal{C}} [\gamma(z) \leq f(z)] \cap \mathcal{F}_{(a,c)}.$$

Nous désignons par $z_{(a,c)}$ le complexe $z_{(a,c)} := h^{-1}_{(a,c)}(z)$; alors, pour une entrée z appartenant à $\mathcal{F}_{(a,c)}$, la sortie $\hat{z}$ vérifie

$$\hat{z} = T^{-q} \circ z_{(a,c)} = z_{(a,c)} - q, \qquad \Im\hat{z} = \Im z_{(a,c)}, \qquad \gamma(z) = \frac{1}{\Im z_{(a,c)}},$$

et nous pouvons résumer ces observations dans la proposition suivante.

Proposition 1.3. *Considérons l'ensemble $\mathcal{C}$ défini en (1.11), l'homographie $h_{(a,c)}$ définie en (1.10), le complexe $z_{(a,c)}$ défini par la relation $z_{(a,c)} := h^{-1}_{(a,c)}(z)$, le feston $\mathcal{F}_{(a,c)}$ défini en (1.12). Alors, l'ensemble de f-niveau associé au défaut d'Hermite γ admet la description locale suivante :*

$$[\gamma(z) \leq f(z)] = \bigcup_{(a,c) \in \mathcal{C}} \left(\left[\frac{1}{\Im z_{(a,c)}} \leq f(z) \right] \cap \mathcal{F}_{(a,c)} \right). \tag{1.13}$$

Cette proposition fournit une caractérisation "locale" de l'ensemble $[\gamma(z) \leq f(z)]$, définie pour chaque feston $\mathcal{F}_{(a,c)}$. Dans la suite nous allons en déduire une caractérisation "globale" indépendante des festons.

1.1.4 Caractérisation globale commune.

Dans cette section, on s'affranchit de la présence des festons $\mathcal{F}_{(a,c)}$ dans (1.13).

Théorème 1.1. *Considérons l'ensemble $\mathcal{C}$ défini en (1.11), l'homographie $h_{(a,c)}$ définie en (1.10), le complexe $z_{(a,c)}$ défini par la relation $z_{(a,c)} := h^{-1}_{(a,c)}(z)$. Alors, pour toute fonction $f : \mathcal{B} \backslash \mathcal{F} \to \mathbb{R}$, l'ensemble de f-niveau du défaut d'Hermite $\gamma(z)$ du réseau engendré par $(1, z)$ vérifie*

$$[\gamma(z) \leq f(z)] = \bigcup_{(a,c) \in \mathcal{C}} \left[\frac{1}{\Im z_{(a,c)}} \leq f(z) \right]. \tag{1.14}$$

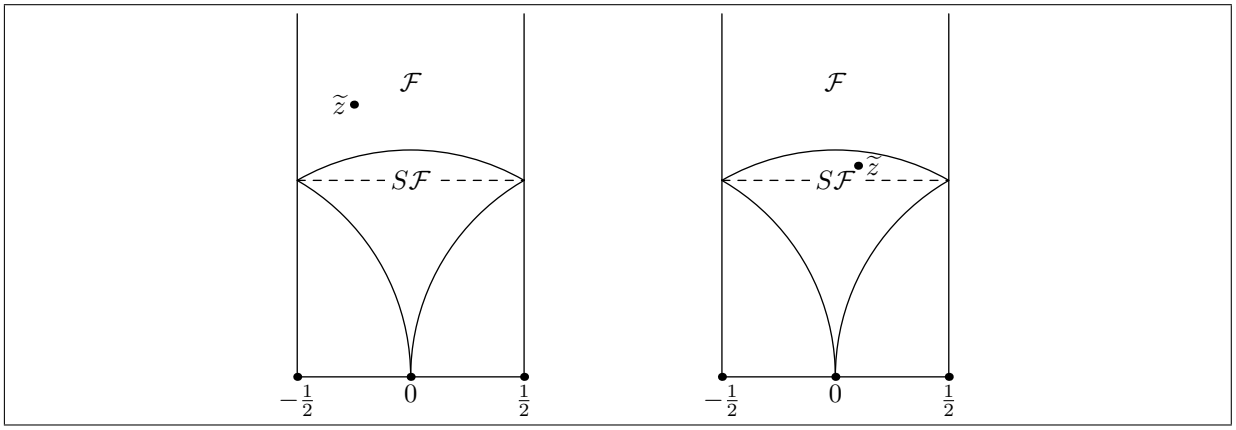

FIGURE 1.1 – Les deux cas possibles pour le complexe $\tilde{z}$ défini dans (1.15) (cf. théorème 1.1).

Démonstration. L'inclusion $\subseteq$ découle immédiatement de (1.13). On démontre $\supseteq$, et donc l'implication logique, pour $z \in \mathcal{B} \setminus \mathcal{F}$, et $(a, c) \in \mathcal{C}$,

$$1/\Im z_{(a,c)} \leq f(z) \Longrightarrow \gamma(z) \leq f(z).$$

Cette implication est triviale quand $f(z) \leq 0$ (antécédant faux), ou quand $f(z) \geq 2/\sqrt{3}$ (conséquent vrai). Lorsque, pour le complexe z, la relation $f(z) \in]0, 2/\sqrt{3}[$ est vérifiée, nous associons au complexe $z_{(a,c)}$ le complexe

$$\tilde{z} = z_{(a,c)} - \lfloor \Re z_{(a,c)} \rceil, \tag{1.15}$$

qui vérifie, lorsque $\Im z_{(a,c)} \geq 1/f(z)$, les relations

$$\Im \tilde{z} = \Im z_{(a,c)} \geq \frac{1}{f(z)} \geq \frac{\sqrt{3}}{2} \quad \text{et} \quad |\Re \tilde{z}| \leq \frac{1}{2}.$$

Deux possibilités se présentent alors, comme l'illustre la figure 1.1. Ou bien

$$\tilde{z} \in \mathcal{F}, \qquad \hat{z} = \tilde{z}, \qquad \gamma(z) = \frac{1}{\Im \tilde{z}} \leq f(z).$$

Ou bien, $\tilde{z} \notin \mathcal{F}$ mais dans ce cas $\tilde{z} \in S\mathcal{F}$ où S est la transformation $z \mapsto -1/z$. Mais dans ce dernier cas, puisque $|\tilde{z}| \leq 1$, on a

$$S(\tilde{z}) \in \mathcal{F}, \qquad \hat{z} = S(\tilde{z}), \qquad \frac{1}{\gamma(z)} = \frac{1}{\gamma(z)} = \Im S(\tilde{z}) = \frac{\Im \tilde{z}}{|\tilde{z}|^2} \geq \Im \tilde{z} \geq \frac{1}{f(z)}.$$

En conclusion, dans les deux cas, l'inégalité $\gamma(z) \leq f(z)$ est vraie, comme il fallait le démontrer. □

Nous exploitons maintenant ce théorème 1.1 pour caractériser les ensembles de niveau asociés à chacun des trois paramètres..

1.1.5 Caractérisations des ensembles de niveau pour chaque paramètre.

On obtient maintenant, pour chacun de ces ensembles de niveau $G(\rho)$, $L(t)$ et $M(u)$, une description géométrique qui fait intervenir des familles classiques de disques (disques de Farey

pour le premier minimum λ, ou disques de Ford pour le défaut d'Hermite γ), ou des secteurs angulaires pour le deuxième minimum orthogonalisé μ.

Une partie des résultats énoncés dans le théorème suivant a déjà été obtenue par Laville et Vallée en 1994 dans [45]. Mais, ces auteurs ne considéraient pas le paramètre μ, et notre preuve pour les paramètres λ, γ est plus complète et plus unifiée que la leur.

Théorème E. *Dans un réseau $\mathcal{L}(1, z)$ déterminé par un complexe $z \in \mathcal{B} \setminus \mathcal{F}$, les ensembles de niveau pour le défaut d'Hermite $\gamma(z)$, le premier minimum $\lambda(z)$ et le deuxième minimum orthogonalisé $\mu(z)$ sont décrits comme suit :*

$$G(\rho) = \left(\bigcup_{(a,c)\in\mathcal{C}} \mathsf{Fo}(a,c,\rho) \right) \cap \mathcal{B} \setminus \mathcal{F}, \qquad L(t) = \left(\bigcup_{(a,c)\in\mathcal{C}} \mathsf{Fa}(a,c,t) \right) \cap \mathcal{B} \setminus \mathcal{F}, \tag{1.16}$$

$$\text{et} \quad M(u) = \left(\bigcap_{(a,c)\in\mathcal{C}} \mathsf{Se}(a,c,u) \right) \cap \mathcal{B} \setminus \mathcal{F}. \tag{1.17}$$

Les domaines $\mathsf{Fo}(a, c, \rho)$ (voir figure 1.2, gauche) sont des disques tangents à l'axe réel en a/c ; ils généralisent les disques de Ford classiques (qu'on retrouve en posant $\rho = 1$) et admettent l'équation suivante

$$\mathsf{Fo}(a,c,\rho) = \left\{ x+iy \in \mathbb{H} \mid \left(x - \frac{a}{c}\right)^2 + \left(y - \frac{\rho}{2c^2}\right)^2 \leq \left(\frac{\rho}{2c^2}\right)^2 \right\}. \tag{1.18}$$

Les domaines $\mathsf{Fa}(a, c, t)$ (voir figure 1.2, centre), appelés t–disques de Farey, sont des demi-disques de centre a/c, liès aux intervalles de Farey ; ils admettent l'équation suivante

$$\mathsf{Fa}(a,c,t) = \left\{ x+iy \in \mathbb{H} \mid \left(x - \frac{a}{c}\right)^2 + y^2 \leq \frac{t^2}{c^2} \right\}. \tag{1.19}$$

Les domaines $\mathsf{Se}(a, c, u)$ (voir figure 1.2, droite) sont des secteurs angulaires, centrés en a/c. Égaux à tout le demi-plan $\mathbb{H}$ si $cu \leq 1$, ils admettent l'équation suivante

$$\mathsf{Se}(a,c,u) = \left\{ x+iy \in \mathbb{H} \mid y \leq \left|x - \frac{a}{c}\right| \frac{cu}{\sqrt{1-(cu)^2}} \right\} \qquad \text{si } cu < 1. \tag{1.20}$$

Démonstration. On applique le théorème 1.1. Tout d'abord, avec l'expression de $h_{(a,c)}$ donnée en (1.10) et la défintion de $z_{(a,c)}$ par l'égalité $z_{(a,c)} := h^{-1}_{(a,c)}(z)$, on observe que $z_{(a,c)}$ s'écrit

$$z_{(a,c)} = \frac{d_0 z - b_0}{-cz + a} \qquad \text{et donc} \quad \Im z_{(a,c)} = \frac{y}{|cz-a|^2}.$$

Ainsi, la conclusion du théorème 1.1 se réécrit en

$$[\gamma(z) \leq f(z)] = \bigcup_{(a,c)\in\mathcal{C}} \left[|cz-a|^2 \leq f(z) y \right]. \tag{1.21}$$

Et, avec (1.7),

$$G(\rho) = \bigcup_{(a,c)\in\mathcal{C}} [|cz-a|^2 \leq \rho y], \qquad L(t) = \bigcup_{(a,c)\in\mathcal{C}} [|cz-a|^2 \leq t^2],$$

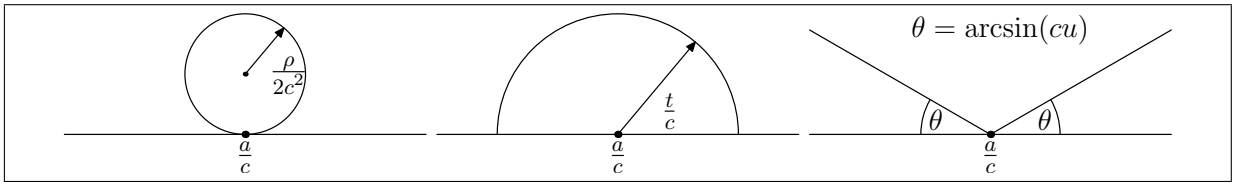

FIGURE 1.2 – Briques de base pour la construction des ensembles de niveau $G(\rho)$, $L(t)$ et $M(u)$: disques de Ford, demi-disques de Farey, et secteurs angulaires.

$$M(u) = (\mathcal{B} \setminus \mathcal{F}) \setminus \left(\bigcup_{(a,c)\in\mathcal{C}} [|cz-a|^2 < \frac{y^2}{u^2}] \right) = \bigcap_{(a,c)\in\mathcal{C}} [|cz-a|^2 \geq \frac{y^2}{u^2}].$$

Maintenant, avec des calculs élémentaires, on obtient l'équivalence

$$|cz-a|^2 \leq \rho y \Leftrightarrow \left(x - \frac{a}{c}\right)^2 + \left(y - \frac{\rho}{2c^2}\right)^2 \leq \left(\frac{\rho}{2c^2}\right)^2,$$

ce qui conduit bien à l'équation (1.18) définissant le disque de Ford $\mathsf{Fo}(a,c,\rho)$. De manière analogue, on obtient l'équivalence

$$|cz-a|^2 \leq t^2 \Leftrightarrow \left(x - \frac{a}{c}\right)^2 + y^2 \leq \frac{t^2}{c^2},$$

ce qui correspond à l'équation (1.19) définissant le disque de Farey $\mathsf{Fa}(a,c,t)$. Pour $\mathsf{Se}(a,c,u)$, on obtient l'équivalence

$$|cz-a|^2 \geq \frac{y^2}{u^2} \Longleftrightarrow (x - \frac{a}{c})^2 \geq y^2 \left(\frac{1}{(cu)^2} - 1\right)$$

et la dernière inégalité est vérifiée pour tout $z \in \mathbb{H}$ lorsque $cu \geq 1$. Si $cu < 1$, on trouve

$$y \leq \left|x - \frac{a}{c}\right| \frac{cu}{\sqrt{1-(cu)^2}},$$

ce qui correspond à l'équation (1.9) définissant le secteur angulaire $\mathsf{Se}(a,c,u)$. Ceci conclut la preuve. □

1.2 Préliminaires pour l'étude de $L(t)$ et $M(u)$

Les caractérisations géométriques de $L(t)$ et $M(u)$ données dans la suite de ce chapitre décrivent précisément l'intersection de chacun de ces ensembles avec des *bandes verticales* bien choisies, que nous définissons maintenant.

Définition 1.1. *Soient a/c et b/d deux rationnels. La* bande verticale *ou* bande *de a/c et b/d est l'ensemble*

$$\left\langle \frac{a}{c}, \frac{b}{d} \right\rangle = \left\{ z \in \mathbb{H} \mid \frac{a}{c} \leq \Re(z) \leq \frac{b}{d} \right\}. \tag{1.22}$$

Les rationnels a/c et b/d sont appelés les extrémités *de la bande.*

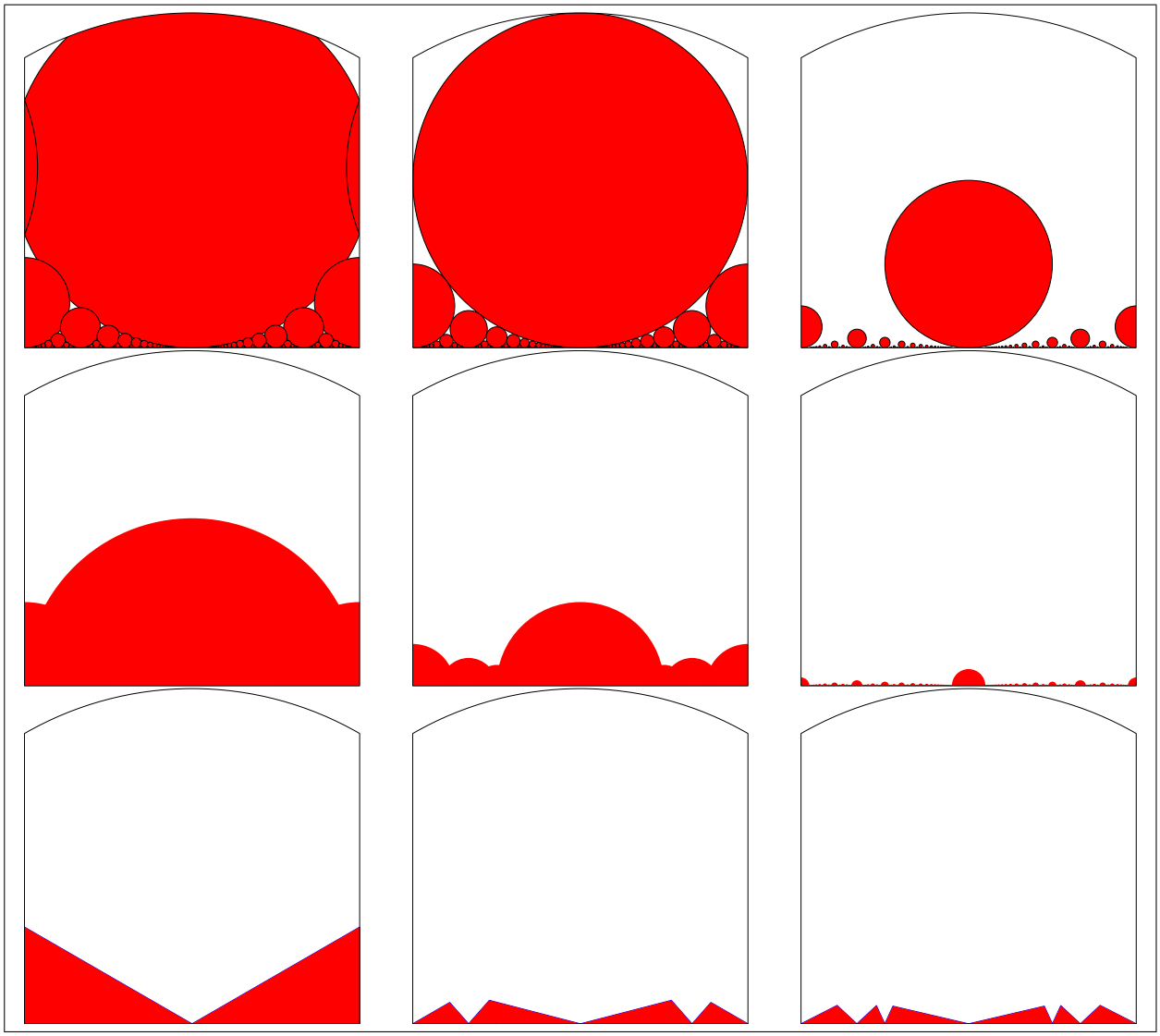

FIGURE 1.3 – Ensembles de niveau pour les paramètres géométriques. Première ligne : ensembles $G(\rho)$ pour $\rho = \frac{1+2/\sqrt{3}}{2}; 1; 1/2$. Deuxième ligne : ensembles $L(t)$ pour $t = 1/2; 1/4; 1/20$. Troisième ligne : ensembles $M(u)$ pour $u = 1/2; 1/4; 5/22$.

La finesse de la caractérisation géométrique est liée à un choix adéquat de ces bandes, et dans notre cas, ce choix sera lié aux suites de Farey. Cette section rappelle la notion de suite de Farey, qui va permettre de définir la suite des bonnes bandes verticales. Nous commençons par des rappels sur les fractions, leur construction, la notion d'adjacence, avant de définir la consécutivité.

Les propriétés suivantes font intervenir des nombres rationnels. A un nombre rationnel, est associé une seule fraction, associée à la représentation *irréductible* de ce rationnel, avec un *dénominateur positif*. Ainsi, quand on parlera de la fraction a/c, on sous-entendra que a et c sont premiers entre eux et $c \geq 1$.

1.2.1 Adjacence

La relation d'adjacence entre fractions a été introduite par L. Ford en 1938, dans [24]. Dans cet article, Ford représente chaque fraction a/c par son disque de Ford $\mathsf{Fo}(a, c, 1)$ (qui a été déjà introduit dans le théorème **E**), et la relation d'adjacence des fractions est définie pour que les fractions adjacentes soient exactement celles dont les disques de Ford sont tangents. Nous reviendrons sur ce point plus tard, dans la section 1.5. Dans ce paragraphe, nous rappelons la définition d'adjacence ainsi que quelques propriétés arithmétiques des fractions adjacentes.

Définition 1.2. *On dit que deux rationnels a/c et b/d sont* adjacents *si $ad - bc = \pm 1$.*

La proposition suivante établit une bijection entre les couples de fractions adjacentes de l'intervalle $[-1/2, 1/2]$ et les couples de dénominateurs de ces fractions. Cette correspondance nous a déjà servi dans la preuve de la proposition 1.2 (partie II) et nous sera utile dans toute la suite, car elle montre le rôle essentiel joué par les dénominateurs dans l'étude géométrique.

Proposition 1.4. *Les couples de fractions adjacentes $(a/c, b/d)$ de l'intervalle $[-1/2, 1/2]$, qui satisfont $a/c < b/d$ sont en correspondance bijective avec les couples (c, d) vérifiant $c, d \geq 1, (c, d) \neq (1, 1)$ et c et d premiers entre eux, en l'occurrence avec leur couple de dénominateurs.*

Démonstration. D'abord, à chaque couple $(a/c, b/d)$ on peut associer le couple de dénominateurs (c, d) sans ambigüité puisque les fractions sont supposées irréductibles et ordonnées. Par définition, l'adjacence implique, grâce au lemme de Bézout, que c et d sont premiers entre eux. De plus, on ne peut avoir $(c, d) = (1, 1)$, car l'intervalle $[-1/2, 1/2]$ ne contient qu'une fraction de dénominateur égal à 1.

Inversement, considérons un couple (c, d) d'entiers premiers entre eux qui satisfait $c \geq 1$, $d \geq 1$ et $(c, d) \neq (1, 1)$, et donc $c + d \geq 3$. Alors, la relation de Bézout "centrée" entraîne l'existence d'un couple (a, b) unique vérifiant

$$ad - bc = -1 \qquad \text{avec} \qquad \frac{-1}{2} \leq \frac{a}{c}, \frac{b}{d} \leq \frac{1}{2}, \tag{1.23}$$

comme nous le montrons maintenant. Le lemme de Bézout classique prouve l'existence d'un couple (a', b') vérifiant

$$a'd - b'c = -1 \qquad \text{avec} \qquad 0 \leq \frac{a'}{c}, \frac{b'}{d} \leq 1, \tag{1.24}$$

Nous montrons que le couple (a, b) de l'équation de Bézout centrée se calcule facilement à partir de ce couple (a', b'). En effet, si a'/c et b'/d appartiennent tous les deux à l'intervalle $[0, 1/2]$, alors le couple (a, b) avec $a = a'$ et $b = b'$ convient. Si a'/c et b'/d appartennent tous les deux à

l'intervalle $[1/2, 1]$, alors le couple (a, b) avec $a = a' - c$ et $b = b' - d$ convient. Il reste donc à traiter le cas où l'on a $a'/c < 1/2 < b'/d$. Ce cas est impossible car l'inégalité $c + d \geq 3$ entraîne alors

$$\frac{1}{cd} = \left| \frac{a'}{c} - \frac{b'}{d} \right| \geq \frac{1}{2c} + \frac{1}{2d} = \frac{c+d}{2cd} \geq \frac{3}{2cd}.$$

Montrons maintenant qu'un tel couple (a, b) est unique. Supposons l'existence de deux tels couples (a_1, b_1) et (a_2, b_2) vérifiant (1.23). Alors, on a nécessairement $|a_1 - a_2| = c, |b_1 - b_2| = d$, ce qui implique $|a_1| = |a_2| = c/2$ et $|b_1| = |b_2| = d/2$. Ce n'est possible que si c et d sont tous les deux pairs, et c'est impossible puisque c et d sont premiers entre eux. □

Repérage par rapport à un couple de fractions adjacentes. Dans la suite, nous utilisons le couple (c, d) de dénominateurs pour indexer les objets liés au couple de rationnels $(a/c, b/d)$. La proposition suivante nous sera de grande utilité lorsque dans nos calculs une paire de fractions aura un rôle prépondérant.

Proposition 1.5. *Soient deux fractions adjacentes a/c et b/d vérifiant $a/c < b/d$. Alors :*

(i) Pour chaque rationnel e/f, il existe un couple unique (m, n) d'entiers premiers entre eux, pour lequel

$$e = ma + nb, \qquad f = mc + nd.$$

La position de e/f par rapport à l'intervalle $[a/c, b/d]$ se lit sur les signes de m et n, et

(i) e/f est plus petit (resp. égal ou plus grand) que a/c si et seulement si n est négatif, (resp. nul ou positif).

(ii) e/f est plus petit (resp. égal ou plus grand) que b/d si et seulement si m est positif, (resp. nul ou négatif).

(ii) Soit e/g une fraction de l'intervalle ouvert $]a/c, b/d[$. Alors : $g \geq (c + d) > \max(c, d)$.

(iii) Soit une autre paire $e/g, f/h$ de fractions adjacentes vérifiant $e/g < f/h$. Alors, les intervalles $]a/c, b/d[$ et $]e/g, f/h[$ sont ou bien disjoints, ou bien emboîtés.

Démonstration. Prouvons (i)Soit e/f un rationnel avec $f \geq 1$. Puisque e et f sont premiers entre eux, il existe un couple (α, β) pour lequel $e\alpha + f\beta = 1$. Alors, la matrice

$$\begin{pmatrix} m & p \\ n & q \end{pmatrix} := \begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} \begin{pmatrix} e & \beta \\ f & -\alpha \end{pmatrix} = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \begin{pmatrix} e & \beta \\ f & -\alpha \end{pmatrix}. \tag{1.25}$$

a ses coefficients dans $\mathbb{Z}$, puisque a/c et b/d sont adjacents. Ceci montre que le couple (m, n) est entier. La relation

$$\begin{pmatrix} e & \beta \\ f & -\alpha \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} m & p \\ n & q \end{pmatrix}.$$

montre les égaités $e = ma+nb$ et $f = mc+nd$. Enfin, en considérant le déterminant des matrices de (1.25), on conclut que $mq - np = 1$, d'où $\operatorname{pgcd}(m, n) = 1$ comme voulu. Les identités

$$\frac{a}{c} - \frac{e}{f} = -\frac{n}{cf} \qquad \frac{b}{d} - \frac{e}{f} = \frac{m}{df}. \tag{1.26}$$

montrent que la position de e/f par rapport à l'intervalle $[a/c, b/d]$ est caractérisée par les signes de m et n.

Pour prouver (ii), on observe que d'après (i), le dénominateur de e/g vérifie $g = mc + nd$ avec $m, n \geq 1$, et donc $g \geq c + d > \max(c, d)$.

Pour (iii),on suppose, à contrario, que $a/c < e/g < b/d < f/h$. Alors, d'après (i), $g > \max(c, d) \geq d$ et $d > \max(g, h) \geq g$, en claire contradiction. □

Parmi les fractions comprises dans un intervalle d'extrémités adjacentes $]a/c, b/d[$, la fraction de plus petit dénominateur joue un rôle souvent distingué. On l'appelle le *médian* de a/c et b/d.

Définition 1.3. *Nous appelons* médian *de deux fractions adjacentes a/c et b/d le rationnel $(a+b)/(c+d)$.*

Voici trois propriétés remarquables du médian :

Proposition 1.6. *Soient a/c, b/d deux rationnels adjacents. Alors :*

(i) Ils sont adjacents à leur médian $(a+b)/(c+d)$.

(ii) La fraction $(a+b)/(c+d)$ est irréductible.

(iii) Le médian est le rationnel de plus petit dénominateur dans $]a/c, b/d[$.

Démonstration. On procède dans l'ordre. Dans le cas de (i), la relation $ad - bc = \pm 1$ entraîne

$$a(c+d) - c(a+b) = ad - bc = \pm 1 \qquad \text{et} \qquad b(c+d) - d(a+b) = bc - ad = \mp 1, \tag{1.27}$$

ce qui montre l'adjacence.

Pour (ii), les relations (1.27) permettent de conclure que $\operatorname{pgcd}(a+b, c+d) = 1$ et donc que la fraction $(a+b)/(c+d)$ est irréductible.

Enfin, pour (iii), d'après la proposition 1.5, tout rationnel de $]a/c, b/d[$ est de la forme $(ma+nb)/(mc+nd)$ avec $m, n \geq 1$, ce qui établit la propriété. □

Les extrémités des bandes verticales qu'on utilisera dans les caractérisations seront données par des fractions *consécutives* dans une *suite de Farey*, que nous présentons maintenant.

1.2.2 Suite de Farey

Dans la littérature (voir [30]), la suite de Farey d'ordre n est la suite finie croissante formée par toutes les fractions irréductibles de l'intervalle $[0, 1]$ dont le dénominateur est au plus n. La suite de Farey d'ordre $n+1$ se construit à partir de la suite d'ordre n en y rajoutant les médians de dénominateur au plus égal à $n+1$ des fractions consécutives. Ici, nous étendons cette notion, en utilisant des bornes réelles, et nous remplaçons aussi l'intervalle $[0, 1]$ par l'intervalle $[-1/2, 1/2]$.

Définition 1.4. *On appelle t-*suite de Farey*, l'ensemble fini de fractions*

$$\mathfrak{F}^{(t)} := \left\{ \frac{a}{c} \in [-1/2, 1/2] \qquad ; \quad 1 \leq c \leq \frac{1}{t} \right\}.$$

Deux rationnels $a/c < b/d$ sont donc consécutifs dans la suite de Farey $\mathfrak{F}^{(t)}$ ssi leur couple de dénominateurs (c, d) appartient à l'ensemble

$$\mathfrak{D}^{(t)} = \left\{ (c,d) \ ; \ c, d \geq 1, \ (c,d) = 1, \ c \leq \frac{1}{t}, \ d \leq \frac{1}{t}, \ (c+d) > \frac{1}{t} \right\}.$$

La définition suivante sera importante dans la suite, en particulier dans les caractérisations géométriques de $L(t)$ et $M(u)$.

Définition 1.5. *On appelle t-*bande de Farey *une bande verticale $\langle a/c, b/d \rangle$ dont les extrémités a/c et b/d sont des rationnels consécutifs dans $\mathfrak{F}^{(t)}$.*

1.2.3 Sommes de Riemann arithmétiques

Nous serons souvent conduits à utiliser des sommes de Riemann pour lesquelles la sommation suit une contrainte de type arithmétique (typiquement, les entiers sur lesquels on somme sont premiers entre eux). Nous trouverons ce type de somme, d'abord dans ce chapitre, pour évaluer des probabilités limite, qui interviendront dans les caractérisations géométriques, comme la probabilité ϱ_{D} du théorème **I** et la probabilité ϱ_{T} du théorème **J**. Elles interviendront aussi dans le chapitre suivant, dans l'estimation des mesures de $L(t)$ et de $M(u)$.

Définition 1.6. *Soit f une fonction positive intégrable sur un sous-ensemble $\Delta \subset [0,1]^2$, et sa somme de Riemann*

$$\overline{F}[f,\Delta](u) := u^2 \sum_{\substack{c,d\geq 1\\(cu,du)\in\Delta}} f(cu,du). \tag{1.28}$$

On appelle somme de Riemann arithmétique, la somme $F[f,\Delta](u)$, contrainte par la condition $(c,d)=1$,

$$F[f,\Delta](u) := u^2 \sum_{\substack{c,d\geq 1\\(cu,du)\in\Delta\\(c,d)=1}} f(cu,du). \tag{1.29}$$

Nous montrons maintenant le résultat suivant :

Théorème 1.2. *Soit f une fonction positive intégrable sur un sous-ensemble $\Delta \subset [0,1]^2$, dont l'intégrale de f sur Δ est désignée par $I[f,\Delta]$. Considérons sa somme de Riemann $\overline{F}[f,\Delta]$ et sa somme arithmétique $F[f,\Delta]$. Supposons que les deux conditions sont vérifiées*

(i) La somme de Riemann $\overline{F}[f,\Delta](u)$ tend vers l'integrale $I[f,\Delta]$ pour $u \to 0$.

(ii) La somme de Riemann arithmétique $F[f,\Delta](u)$ est bornée pour $u \to 0$.

Alors, la somme de Riemann arithmétique $F[f,\Delta](u)$ vérifie

$$\lim_{u\to 0} F[f,\Delta](u) = \frac{1}{\zeta(2)} I[f,\Delta]$$

Démonstration. On fixe f et Δ et on pose $F \equiv F[f,\Delta]$, $\overline{F} \equiv \overline{F}[f,\Delta]$. La preuve comporte deux étapes : la première exprime F en fonction de $\overline{F}$; la seconde calcule la limite de $F(u)$ pour $u \to 0$, en exploitant la relation trouvée.

On exprime d'abord $\overline{F}$ en fonction de F, en regroupant les couples dont le pgcd est g,

$$\frac{\overline{F}(u)}{u^2} = \sum_{g\geq 1} \sum_{\substack{c',d'\geq 1\\(c'gu,d'gu)\in\Delta\\(c',d')=1}} f(c'gu,d'gu) = \sum_{g\geq 1} \frac{1}{g^2u^2} F(gu). \tag{1.30}$$

On remarque en passant que le fait que F soit bornée pour $u \geq 0$ implique, d'après 1.30, que $\overline{F}(u)$ est bornée aussi pour $u \geq 0$.

Pour exprimer F en fonction de $\overline{F}$, on inverse la relation (1.30), en utilisant un lemme qui généralise l'inversion de Moebius, et utilise la fonction de Moebius $m : \mathbb{N}^\star \to \{-1,0,+1\}$, définie par

$$m(n) = \begin{cases} 1 & \text{si } n = 1\\ (-1)^k & \text{si } n \text{ est le produit de } k \text{ nombres premiers distincts,}\\ 0 & \text{dans tout autre cas.}\end{cases} \tag{1.31}$$

Le résultat suivant est adapté à notre contexte , et sa preuve peut être trouvée par exemple dans la section 2.14 de [7].

Lemme 1.1. *A partir d'une une fonction G à valeurs réelles ou complexes, définie sur $]0,+\infty[$ et nulle sur $]1,\infty[$, on définit la fonction $\overline{G}$ par la relation*

$$\overline{G}(u) = \sum_{gu\geq 1} G(gu).$$

Alors, G s'exprime en fonction de $\overline{G}$ via la fonction de Moebius m, sous la forme

$$G(u) = \sum_{gu\geq 1} m(g)\overline{G}(gu).$$

Après avoir remarqué que $\overline{F}(u)$ et $F(u)$ sont nulles pour $u > 1$ (puisque $\Delta \subset [0,1]^2$), on applique le lemme 1.1 aux fonctions $G : x \mapsto x^{-2}F(x)$ et $\overline{G} : x \mapsto x^{-2}\overline{F}(x)$, et on obtient la relation inversée

$$F(u) = \sum_{g\geq 1} \frac{m(g)}{g^2}\overline{F}(gu). \tag{1.32}$$

Ayant exprimé F en fonction de $\overline{F}$, nous passons au calcul de la limite. Nous séparons la somme (1.32) en deux parties, la somme $F^-(u)$, correspondant aux $g \leq 1/\sqrt{u}$, et la somme $F^+(u)$ correspondant au reste de la sommation. Puisque $\overline{F}(u)$ tend vers $I[f,\Delta]$, alors

$$\overline{F}(gu) = I[f,\Delta]| + o(1), \qquad \text{pour tout } g \leq 1/\sqrt{u},$$

avec un $o(1)$ uniforme en g, puisque $gu \leq \sqrt{u}$. Ainsi,

$$F^-(u) := \sum_{g\leq 1/\sqrt{u}} \frac{m(g)}{g^2}\overline{F}(gu) = I[f,\Delta]\left(\sum_{g\leq 1/\sqrt{u}} \frac{m(g)}{g^2}\right) + \left(\sum_{g\leq 1/\sqrt{u}} \frac{m(g)}{g^2}\right)\cdot o(1).$$

Comme la série de terme général $m(g)/g^2$ est convergente (de somme $1/\zeta(2)$), la fonction F^- a une limite pour $u \to 0$ et

$$\lim_{u\to 0} F^-(u) = \lim_{u\to 0} \sum_{g\leq 1/\sqrt{u}} \frac{m(g)}{g^2}\overline{F}(gu) = \frac{1}{\zeta(2)} I[f,\Delta].$$

Par ailleurs, la série de terme général $(m(g)/g^2)\overline{F}(gu)$ est une série normalement convergente, car $m(g)\overline{F}(gu)$ est bornée, et donc $F^+(u)$ tend vers 0 pour $u \to 0$. On en conclut que

$$\lim_{u\to 0} F(u) = \frac{1}{\zeta(2)} I[f,\Delta]$$

comme on voulait prouver. □

Remarque. Dans le cas d'une fonction positive f, l'énoncé du théorème se simplifie, car l'hypothèse (ii) est impliquée par l'hypothèse (i). L'hypothèse (i) devient elle-même superflue quand la fonction f est bornée. L'hypothèse (i) est aussi superflue quand l'intégrale de f est impropre, mais avec une fonction f monotone. Ce sera toujours dans ces contextes-là que le théorème sera utilisé.

Nous abordons maintenant les caractérisations géométriques des ensembles de niveau. Il s'agit d'étudier la géométrie des ensembles de niveau de manière suffisamment précise, pour pouvoir ensuite en calculer la mesure (ce qu'on fera dans le chapitre suivant). Les ensembles $L(t)$ et $M(u)$ se décrivent bien localement, en utilisant des bandes de Farey (cf. section 1.2.2). L'ensemble $G(\rho)$, lui, se décrit bien globalement.

1.3 Géométrie de l'ensemble de niveau du premier minimum.

Dans ce paragraphe nous décrivons en détail la géométrie de l'ensemble $L(t) := [\lambda(z) \leq t]$.

1.3.1 Description de $L(t)$.

Comme on l'a vu dans l'étude locale, la géométrie de $L(t)$ est liée à des t-demi-disques dont les centres sont des éléments de la suite de Farey $\mathfrak{F}^{(t)}$. Ils ont été étudiés par Laville et Vallée [45], et la caractérisation présentée ici leur est due. Laville et Vallée commencent par introduire des t-intervalles de Farey, dont certaines propriétés se généralisent plus tard aux t-demi-disques, Ici nous avons fait le choix de traiter directement les propriétés des disques, et on se permettra de parler de disques de Farey, en sous-entendant toujours qu'il s'agit de demi-disques. Le résultat principal est résumé dans le théorème suivant, qui sera prouvé dans les sections suivantes.

Théorème F. *Soit t un nombre réel de $[0,1]$. Pour un demi-disque de Farey $\mathsf{Fa}(a,c,t)$, on désigne par $\mathsf{Fa}^+(a,c,t)$ le quart de disque droit et par $\mathsf{Fa}^-(a,c,t)$ le quart de disque gauche. Alors, trois cas se présentent pour $L(t)$, selon la valeur de t :*

(i) Si $t \in]1/2, 1/\sqrt{3}[$, alors

$$L(t) = \mathsf{Fa}^+(-1,2,t) \bigcup \left(\mathsf{Fa}(0,1,t) \cap \left\langle \frac{-1}{2}, \frac{1}{2} \right\rangle \right) \bigcup \mathsf{Fa}^-(1,2,t),$$

(ii) Si $t > 1/\sqrt{3}$, alors

$$L(t) = \mathsf{Fa}(0,1,t) \bigcap \left\langle \frac{-1}{2}, \frac{1}{2} \right\rangle .$$

(iii) Si $t \leq 1/2$, si $a/c < b/d$ sont consécutifs dans la suite de Farey $\mathfrak{F}^{(t)}$, la portion de $L(t)$ comprise dans la t-bande $\langle a/c, b/d \rangle$ est égale à la réunion de deux quarts de disque et d'un disque de Farey. Plus précisément,

$$L(t) \bigcap \langle \frac{a}{c}, \frac{b}{d} \rangle = \mathsf{Fa}^+(a,c,t) \bigcup \mathsf{Fa}^-(b,d,t) \bigcup \mathsf{Fa}(a+b,c+d,t)\,. \tag{1.33}$$

Nous avons tout simplement

$$L(t) \bigcap \langle \frac{a}{c}, \frac{b}{d} \rangle = \mathsf{Fa}^+(a,c,t) \bigcup \mathsf{Fa}^-(b,d,t), \tag{1.34}$$

si et seulement si $(c^2+cd+d^2)t^2 \geq 1$, et la proportion de couples de (c,d) vérifiant (1.34) tend vers

$$\varrho_D = 2 - \frac{2}{\sqrt{3}} \cdot \frac{\pi}{3} \approx 0,7908004 \qquad t \to 0.$$

1.3.2 Position des disques de Farey.

La preuve du théorème **F** résulte d'une étude sur la position des t-disques de Farey par rapport aux t-bandes de Farey, et aussi sur la position relative des t-disques de Farey entre eux. Les deux propositions 1.7 et 1.8 qui suivent présentent ces propriétés.

Proposition 1.7 (Intersection disque-bande)**.** *Soit $t \in [0,1]$ et soient $a/c < b/d$ deux rationnels tels que $(c,d) \in \mathfrak{D}^{(t)}$. Les énoncés suivants sont vérifiés par les t-demi-disques et t-bandes de Farey :*

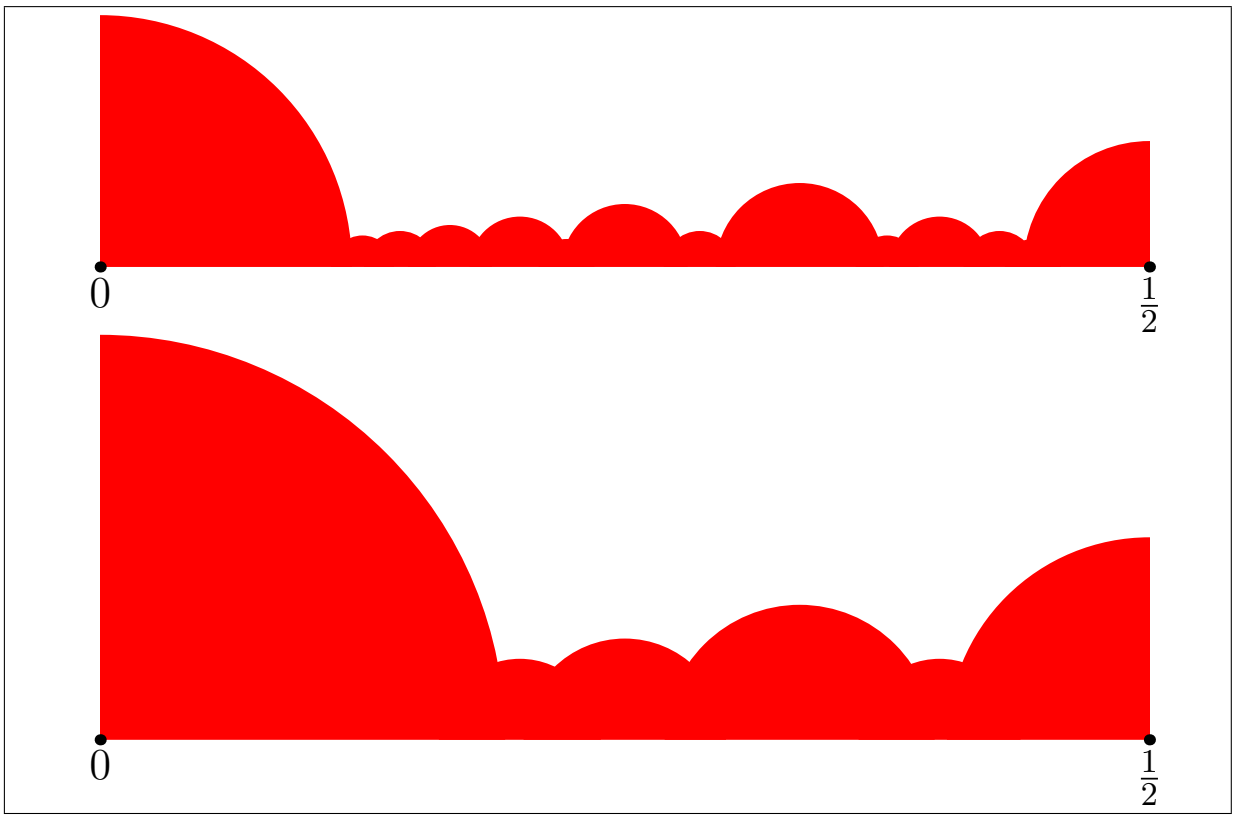

FIGURE 1.4 – Ensemble $L(t)$. Les valeurs illustrées sont $t = 0,12$ et $t = 0,193$.

(i) Tout t-demi-disque de Farey dont le centre n'est pas dans l'intervalle $[a/c, b/d]$ est disjoint avec la bande $\langle a/c, b/d\rangle$.

(ii) Tout t-demi-disque de Farey dont le centre appartient à $]a/c, b/d[$ est inclus dans la bande $\langle a/c, b/d\rangle$. En particulier le disque associé au médian $(a+b)/(c+d)$ l'est.

(iii) Les t-quart-de-disques positifs et négatifs associés respectivement à a/c et à b/d sont tous les deux inclus dans $\langle a/c, b/d\rangle$.

Démonstration. Soit $t \in [0,1]$ et soient $a/c < b/d$ deux rationnels tels que $(c,d) \in \mathfrak{D}^{(t)}$.

Prouvons (i). Soit $(e,f) \in \mathcal{C}$ tel que e/f n'appartient pas à $[a/c, b/d]$. Nous voulons prouver que $\mathsf{Fa}(e,f,t) \cap \langle a/c, b/d\rangle = \emptyset$. En effet, le couple (e,f) s'écrit $(e,f) = (ma+nb, mc+nd)$ pour m, n premiers entre eux (proposition 1.5), avec $|m| \geq 1$, $|n| \geq 1$. Le diamètre de $\mathsf{Fa}(e,f,t)$, c'est-à-dire l'intervalle $[e/f, (e+t)/f]$, vérifie respectivement lorsque $e/f < a/c$ et $e/f > b/d$,

$$\left|\frac{e}{f} - \frac{a}{c}\right| = \frac{|n|}{cf} > \frac{t}{f}|n| \geq \frac{t}{f} \qquad \text{et} \qquad \left|\frac{e}{f} - \frac{b}{d}\right| = \frac{|m|}{df} > \frac{t}{f}|m| \geq \frac{t}{f},$$

puisque $ct < 1$ et $dt < 1$. Cela établit (i).

Prouvons (ii). Soit $(e,f) \in \mathcal{C}$ tel que $e/f \in]a/c, b/d[$. Nous allons montrer que le diamètre du disque $\mathsf{Fa}(e,f,t)$ est inclus dans $]a/c, b/d[$. En effet, le couple (e,f) s'écrit $(ma+nb, mc+nd)$ avec $m \geq 1, n \geq 1$ premiers entre eux, et nous avons

$$\frac{e}{f} - \frac{a}{c} = \frac{n}{cf} > \frac{t}{f}n \geq \frac{t}{f} \qquad \text{et} \qquad \frac{b}{d} - \frac{e}{f} = \frac{m}{df} > \frac{t}{f}m \geq \frac{t}{f},$$

puisque $ct < 1$ et $dt < 1$, ce qui établit (ii).

Enfin, rouvons (iii). Pour montrer que les t-quarts-disques $\mathsf{Fa}^+(a,c,t)$ et $\mathsf{Fa}^-(b,d,t)$ sont inclus dans $\langle a/c, b/d\rangle$, il suffit de vérifier que le rayon de chaque disque est plus court que la longueur de l'intervalle, égale à $1/cd$. Or, cela s'établit immédiatement en divisant les inégalités $ct < 1$ et $dt < 1$ par cd. La preuve est donc achevée. □

La proposition suivante décrit les inclusions entre demi-disques. La quatrième assertion nous sera utile seulement dans le chapitre suivant.

Proposition 1.8 (Inclusions entre demi-disques)**.** *Soient $a/c < b/d$ deux rationnels tels que $(c,d) \in \mathfrak{D}^{(t)}$. Alors :*

(i) Les deux cercles délimitant $\mathsf{Fa}(a,c,t)$ et $\mathsf{Fa}(b,d,t)$ sont toujours sécants et l'abscisse $x_{c,d}$ du point d'intersection des deux cercles est égale à

$$x_{c,d} = \frac{a}{c} + \frac{1+t^2(d^2-c^2)}{2cd} = \frac{b}{d} - \frac{1+t^2(c^2-d^2)}{2cd}.$$

(ii) Si $e/f \in]a/c, b/d[$ n'est pas le médian $(a+b)/(c+d)$, alors le t-disque $\mathsf{Fa}(e,f,t)$ est inclus dans $\mathsf{Fa}^+(a,c,t)$ ou bien dans $\mathsf{Fa}^-(b,d,t)$.

(iii) Le demi-disque associé au médian est conditionnellement inclus dans la réunion des t-quart-de-disques positif et négatif associés respectivement à a/c et à b/d. Plus précisément,

$$\mathsf{Fa}(a+b, c+d, t) \subset \mathsf{Fa}^+(a,c,t) \cup \mathsf{Fa}^-(b,d,t) \iff (c^2+cd+d^2)t^2 \geq 1.$$

(iv) Les t-quart-de-disques associés au médian sont conditionnellement inclus dans les t-quarts-disques associés à a/c et b/d Plus précisément,

$$\begin{aligned} \mathsf{Fa}^-(a+b, c+d, t) \subset \mathsf{Fa}^+(a,c,t) \quad &\iff \quad ((c+d)^2 - c^2)t^2 \geq 1 \\ \mathsf{Fa}^+(a+b, c+d, t) \subset \mathsf{Fa}^-(b,d,t) \quad &\iff \quad ((c+d)^2 - d^2)t^2 \geq 1. \end{aligned}$$

Démonstration. Soit $t \in [0,1]$ et soient $a/c < b/d$ deux rationnels tels que $(c,d) \in \mathfrak{D}^{(t)}$.

Prouvons (ii). Les équations des cercles $\mathsf{Fa}(a,c,t)$ et $\mathsf{Fa}(b,d,t)$ s'écrivent

$$\left(x - \frac{a}{c}\right)^2 + y^2 = \frac{t^2}{c^2} \qquad \text{et} \qquad \left(x - \frac{b}{d}\right)^2 + y^2 = \frac{t^2}{d^2},$$

et, avec les changements de variables

$$x' = x - \frac{a}{c},\ y' = y \qquad \text{ou} \qquad x' = x - \frac{b}{d},\ y' = y,$$

et l'identité $b/d - a/c = 1/(cd)$, on obtient

$$x_{c,d} - \frac{a}{c} = \frac{1 + t^2(d^2 - c^2)}{2cd} \qquad x_{c,d} - \frac{b}{d} = \frac{-1 + t^2(d^2 - c^2)}{2cd}.$$

À présent, on prouve (ii) Supposons que $e/f \in]a/c, (a+b)/(c+d)[$. Nous allons montrer que l'intervalle $](e-t)/f, (e+t)/f[$ est inclus dans $]a/c, (a+t)/c)[$. En effet, on peut écrire $(e,f) = (ma + n(a+b), mc + n(c+d))$, avec $m, n \geq 1$ premiers entre eux (proposition 1.5) et nous avons

$$\frac{e}{f} - \frac{a}{c} = \frac{n}{cf} > \frac{t}{f} n \geq \frac{t}{f}, \qquad \text{et} \qquad \frac{e+t}{f} - \frac{a}{c} = \frac{n + ct}{cf} < \frac{n(c+d)t + mct}{cf} = \frac{t}{c},$$

et donc $\mathsf{Fa}(e,f,t) \subseteq \mathsf{Fa}^+(a,c,t)$. Suivant le même raisonnement on montre que $\mathsf{Fa}(e,f,t) \subseteq \mathsf{Fa}^-(b,d,t)$ lorsque $e/f \in](a+b)/(c+d), b/d[$.

Passons à (iii). Désignons par $[x_c^-, x_c^+]$ le diamètre du disque $\mathsf{Fa}(a,c,t)$, par $[x_d^-, x_d^+]$ le diamètre du disque $\mathsf{Fa}(b,d,t)$ et par $[x_{c+d}^-, x_{c+d}^+]$ le diamètre du disque $\mathsf{Fa}(a+b, c+d, t)$. On a toujours

$$x_c^- < \frac{a}{c} < x_{c+d}^- < x_d^- < x_c^+ < x_{c+d}^+ < \frac{b}{d} < x_d^+. \tag{1.35}$$

Si donc on suit le cercle délimitant $\mathsf{Fa}(a+b, c+d, t)$ en partant de x_c^- et en suivant des abscisses croissantes, on est d'abord dans $\mathsf{Fa}(a,c,t)$ sans être dans $\mathsf{Fa}(b,d,t)$. Si on quitte $\mathsf{Fa}(a,c,t)$ avant d'être entré dans $\mathsf{Fa}(b,d,t)$, alors le cercle délimitant $\mathsf{Fa}(a+b, c+d, t)$ n'est pas inclus dans la réunion $\mathsf{Fa}(a,c,t) \cup \mathsf{Fa}(b,d,t)$. Si, par contre, on entre dans $\mathsf{Fa}(b,d,t)$ avant d'avoir quitté $\mathsf{Fa}(a,c,t)$, alors le cercle délimitant $\mathsf{Fa}(a+b, c+d, t)$ est inclus dans la réunion $\mathsf{Fa}(a,c,t) \cup \mathsf{Fa}(b,d,t)$. L'abscisse $x_{c,c+d}$ du point d'intersection des cercles associés à $\mathsf{Fa}(a,c,t)$ et $\mathsf{Fa}(a+b, c+d, t)$ et l'abscisse $x_{c+d,d}$ du point d'intersection des cercles associés à $\mathsf{Fa}(b,d,t)$ et $\mathsf{Fa}(a+b, c+d, t)$ jouent donc un rôle essentiel (voir figure 1.5). La discussion fait intervenir la position relative de ces abscisses $x_{c,c+d}$ et $x_{c+d,d}$ et on a l'équivalence

$$\mathsf{Fa}(a+b, c+d, t) \subset \mathsf{Fa}^+(a,c,t) \cup \mathsf{Fa}^-(b,d,t) \qquad \text{si et seulement si} \qquad x_{c,c+d} \geq x_{c+d,d}.$$

Avec le calcul mené en (i), on déduit que l'inégalité $x_{c,c+d} \geq x_{c+d,d}$ équivaut à $(c^2 + cd + d^2)t^2 \geq 1$, comme on voulait montrer.

Enfin, prouvons (iv). Compte-tenu de la suite d'inégalités (1.35), il suffit de situer le médian par rapport aux deux abscisses $x_{c,c+d}$ et $x_{c+d,d}$, qui se calculent comme en (i)

$$x_{c,c+d} = \frac{a}{c} + \frac{1 + t^2[(c+d)^2 - c^2]}{2c(c+d)}, \quad x_{c+d,d} = \frac{b}{d} - \frac{1 + t^2[(c+d)^2 - d^2]}{2d(c+d)}.$$

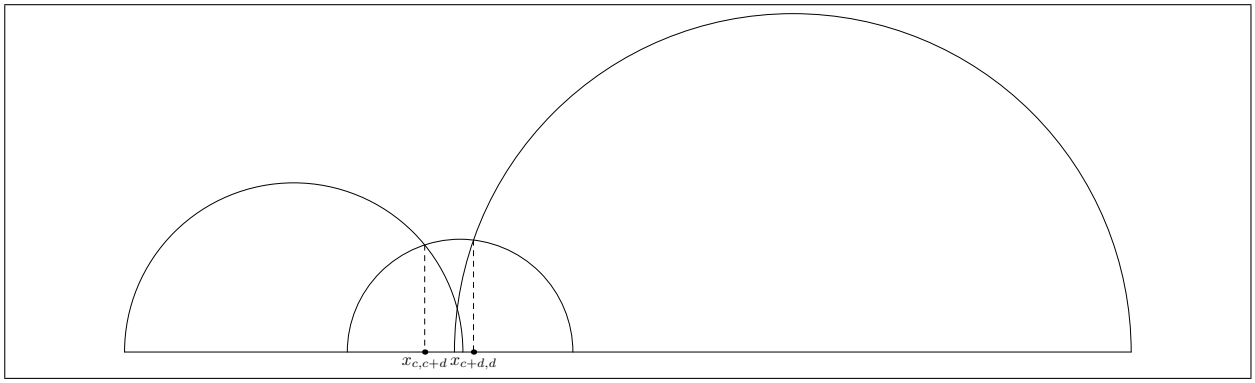

FIGURE 1.5 – La réunion des cercles de Farey centrés en a/c (à gauche) et en b/d (à droite) contient le disque associé au médian $(a+b)/(c+d)$ (au milieu) si et seulement si $x_{c,c+d} \geq x_{c+d,d}$.

Cette position est donc régie par les conditions suivantes :

$$\begin{aligned} \frac{a+b}{c+d} < x_{c,c+d} \quad &\iff \quad ((c+d)^2 - c^2)t^2 > 1 \\ \frac{a+b}{c+d} \in [x_{c,c+d}, x_{c+d,d}] \quad &\iff \quad ((c+d)^2 - c^2)t^2 \leq 1 \text{ et } ((c+d)^2 - d^2)t^2 \leq 1 \\ x_{c+d,d} < \frac{a+b}{c+d} \quad &\iff \quad ((c+d)^2 - d^2)t^2 > 1. \end{aligned}$$

La preuve est donc achevée. □

Maintenant que les propriétés des disques sont établies, nous sommes prêts à prouver le théorème F.

1.3.3 Preuve de la caractérisation géométrique de $L(t)$.

Dans cette preuve, il convient de considérer la suite $\mathfrak{F}^{(t)}$ comme définie dans $[-1, 1]$. Les résultats des propositions 1.7 et 1.8 y sont toujours valables puisqu'elles sont fondées sur la définition de $\mathfrak{D}^{(t)}$, essentiellement indépendante de l'intervalle où sont les fractions.

Soient $a/c < b/d$ deux fractions consécutives dans la suite de Farey $\mathfrak{F}^{(t)}$. Nous allons élaguer l'intersection suivante

$$\left(\bigcup_{(e,f)\in\mathcal{C}} \mathsf{Fa}(e, f, t) \right) \cap \langle \frac{a}{c}, \frac{b}{d} \rangle \tag{1.36}$$

qui est à la base de la caractérisation globale (1.16) de $L(t)$. La proposition 1.7 sur les intersections entre disques et bandes prouve que les couples (e, f) tels que $e/f \notin [a/c, b/d]$ sont redondants dans la réunion 1.36 puisque les disques associés sont disjoints avec $\langle a/c, b/d \rangle$. Par ailleurs, la proposition 1.8 prouve que les disques associés aux couples (e, f) tels que $e/f \in]a/c, b/d[$ avec $e/f \neq (a+b)/(c+d)$ sont inclus soit dans $\mathsf{Fa}^+(a, c, t)$, soit dans $\mathsf{Fa}^-(b, d, t)$. Ainsi, la réunion (1.36) se limite au plus aux disques $\mathsf{Fa}^+(a, c, t)$, $\mathsf{Fa}^-(b, d, t)$ et $\mathsf{Fa}(a+b, c+d, t)$. Par ailleurs, la proposition 1.8 affirme que le demi-disque $\mathsf{Fa}(a+b, c+d, t)$ est inclus dans la réunion des quarts-disques $\mathsf{Fa}(a, c, t)$ et $\mathsf{Fa}(b, d, t)$ ssi $(c^2 + cd + d^2)t^2 \geq 1$. Ainsi, (1.36) s'écrit

$$\mathsf{Fa}^+(a, c, t) \cup \mathsf{Fa}^-(b, d, t) \cup \mathsf{Fa}(a+b, c+d, t)$$

et tout simplement

$$\mathsf{Fa}^+(a,c,t) \cup \mathsf{Fa}^-(b,d,t)$$

si et seulement si $(c^2+cd+d^2)t^2 \geq 1$. À partir de (1.36), on obtient $L(t) \cap \langle a/c, b/d\rangle$ juste en prenant l'intersection avec $\mathcal{B} \setminus \mathcal{F}$.

(i) et (ii). Commençons par traiter le cas où $t > 1/2$, qui se subdivise en deux sous-cas, selon la position de t par rapport à $1/\sqrt{3}$.
Pour $t \in [1/\sqrt{3}, 1]$, on a :

$$L(t) \cap \left\langle \frac{-1}{1}, \frac{0}{1} \right\rangle = (\mathsf{Fa}^+(-1,1,t) \cup \mathsf{Fa}^-(0,1,t) \cap \mathcal{B} \setminus \mathcal{F}$$

$$L(t) \cap \left\langle \frac{0}{1}, \frac{1}{1} \right\rangle = (\mathsf{Fa}^+(0,1,t) \cup \mathsf{Fa}^-(1,1,t) \cup \mathsf{Fa}(1,2,t)) \cap \mathcal{B} \setminus \mathcal{F},$$

Pour $t \in [1/2, 1\sqrt{3}]$, on a :

$$L(t) \cap \left\langle \frac{-1}{1}, \frac{0}{1} \right\rangle = \big(\mathsf{Fa}^+(-1,1,t) \cup \mathsf{Fa}^-(0,1,t) \cup \mathsf{Fa}(-1,2,t)\big) \cap \mathcal{B} \setminus \mathcal{F}$$

et pour $t > 1/2$

$$L(t) \cap \left\langle \frac{0}{1}, \frac{1}{1} \right\rangle = \big(\mathsf{Fa}^+(0,1,t) \cup \mathsf{Fa}^-(1,1,t) \cup \mathsf{Fa}(1,2,t)\big) \cap \mathcal{B} \setminus \mathcal{F}.$$

Sachant que $\mathcal{B} \setminus \mathcal{F} = \mathsf{Fa}(0,1,1) \cap \langle -1/2, 1/2\rangle$, et vue la symétrie des demi-disques $\mathsf{Fa}(-1,1,t)$ et $\mathsf{Fa}(0,1,t)$ par rapport à $x = -1/2$, et des demi-disques $\mathsf{Fa}(0,1,t)$ et $\mathsf{Fa}(1,1,t)$ par rapport à $x = 1/2$, on conclut que $x = -1/2$ et $x = 1/2$ sont leur points d'intersection respectifs, et donc que la partie de $\mathsf{Fa}(-1,1,t)$ (resp. $\mathsf{Fa}(1,1,t)$) qui est à droite (resp. gauche) de $x = -1/2$ (resp. $x = 1/2$), est contenue dans $\mathsf{Fa}(0,1,t)$. Ainsi, en réunissant les égalités obtenues, nous avons finalement

$$L(t) = \mathsf{Fa}^+(-1,2,t) \cup \left(\mathsf{Fa}(0,1,t) \cap \left\langle \frac{-1}{2}, \frac{1}{2} \right\rangle\right) \cup \mathsf{Fa}^-(1,2,t)$$

si $t < 1/\sqrt{3}$ et

$$L(t) = \mathsf{Fa}(0,1,t) \cap \left\langle \frac{-1}{2}, \frac{1}{2} \right\rangle,$$

si $t \geq 1/\sqrt{3}$, comme voulu.

(iii) Lorsque $t \leq 1/2$, nous avons $L(t) \subset [\Im(z) \leq 1/2]$, puisque le t-disque de plus grand rayon est justement de rayon t. Ainsi, en prenant en compte que la bande $\langle -1/2, 1/2\rangle)$ est dans ce cas partitionnée en t-bandes, on conclut que l'intersection avec $\mathcal{B} \setminus \mathcal{F}$ est redondante. Ainsi, nous avons

$$L(t) \cap \langle \frac{a}{c}, \frac{b}{d}\rangle = \mathsf{Fa}^+(a,c,t) \cup \mathsf{Fa}^-(b,d,t) \cup \mathsf{Fa}(a+b,c+d,t)$$

et tout simplement

$$L(t) \cap \langle \frac{a}{c}, \frac{b}{d}\rangle = \mathsf{Fa}^+(a,c,t) \cup \mathsf{Fa}^-(b,d,t)$$

ssi $(c^2+cd+d^2)t^2 \geq 1$, comme souhaité.

La proportion de sections $\langle a/c, b/d\rangle \cap L(t)$ qui vérifient (1.34) correspond exactement à la proportion de couples $(c,d) \in \mathfrak{D}^{(t)}$ telles que $(c^2+cd+d^2)t^2 \geq 1$. Nous allons appliquer le théorème 1.2. Considérons les ensembles Δ et Δ_2 définis comme suit

$$\Delta = \{(x,y) \ : \ 0 < x, y \leq 1, \ x+y > 1 \} \qquad \Delta_2 = \{(x,y) \in \Delta \ : \ x^2+xy+y^2 \geq 1\}.$$

Alors, la proportion $\varrho_{\mathrm{D}}(t)$ de réunions doubles est égale à la proportion des $(ct, dt) \in \Delta$ avec $c, d \geq 1$ et $(c, d) = 1$, qui appartiennent à Δ_2. En reprenant les notations du théorème 1.2, on a

$$\varrho_{\mathrm{D}}(t) = \frac{F[1, \Delta_2](t)}{F[1, \Delta](t)}. \tag{1.37}$$

et le théorème 1.2 prouve que, pour $X \subseteq [0,1]^2$, la somme de Riemann contrainte $F[1, X](t)$ tend vers $(1/\zeta(2))I[1, X]$ quand t tend vers 0, puisque 1 est une fonction proprement Riemann-intégrable sur X. Donc,

$$\lim_{t\to 0} \varrho_{\mathrm{D}}(t) = \frac{I[1, \Delta_2]}{I[1, \Delta]}.$$

Les intégrales en jeu se calculent facilement,

$$I[1, \Delta_2] = 1 - \frac{\pi}{3\sqrt{3}} \qquad \text{et} \qquad I[1, \Delta] = \frac{1}{2},$$

et donc

$$\lim_{t\to 0} \varrho_{\mathrm{D}}(t) = 2 - \frac{2}{\sqrt{3}} \cdot \frac{\pi}{3} \approx 0,7908004.$$

1.3.4 Encadrement de $L(t)$.

À première vue, l'ensemble $L(t)$ peut paraître compliqué à construire, comme le suggère la figure 1.4. Or, regardé de près, il est suffisamment simple pour envisager le calcul de sa mesure. On remarque également que, lorsque t est petit, à peu près 4 t-bandes sur 5 contiennent des réunions doubles. Ce calcul de proportion sera utile notamment lors du calcul des constantes liées à la mesure de $L(t)$, de même que la proposition suivante, qui fournira un encadrement pour la mesure de $L(t)$.

La proposition suivante propose une famille de demi-disques disjoints qui nous sera d'utilité pour trouver un sous-ensemble de $L(t)$ dont la mesure se calcule facilement. Elle est due à Laville et Vallée [45].

Proposition 1.9. *Le domaine $L(t)$ est encadré par les deux domaines suivants, comme suit,*

$$\bigcup_{\frac{a}{c} \in \mathfrak{F}^{(2t)}} \mathsf{Fa}(a, c, t) \subset L(t) \subset \bigcup_{\frac{a}{c} \in \mathfrak{F}^{(\sqrt{3}t/2)}} \mathsf{Fa}(a, c, t) \tag{1.38}$$

la réunion de gauche étant une réunion disjointe.

Démonstration. Nous commençons par un lemme technique.

Lemme 1.2. *Soient $a/c < b/d$ deux rationnels tels que $(c, d) \in \mathfrak{D}^{(2t)}$, $t \in [0, 1]$. Alors, les t-disques (noter bien le t et pas $2t$) de Farey qui leur sont associés sont toujours disjoints.*

Démonstration. Soit $t \in [0, 1]$. Il suffit de vérifier que les intervalles $]a/c, (a+t)/c[$ et $](b-t)/d, b/d[$ sont disjoints. En effet, lorsque $a/c < b/d$ sont tels que $(c, d) \in \mathfrak{D}^{(2t)}$, les deux relations $c(2t) < 1$ et $d(2t) < 1$ prouvent l'inégalité

$$\frac{b-t}{d} - \frac{a+t}{c} = \left(\frac{b}{d} - \frac{a}{c}\right) + \frac{1 - t(c+d)}{cd} \geq 0\,, \tag{1.39}$$

ce qui établit le résultat. □

La suite de Farey $\mathfrak{F}^{(2t)}$ est un sous-ensemble de $\mathfrak{F}^{(t)}$, ce qui entraîne directement l'inclusion de gauche dans (1.38). Grâce au lemme 1.2, les t-disques de Farey centrés sur les termes de la suite $\mathfrak{F}^{(2t)}$ sont disjoints, et donc la réunion de gauche dans (1.39) est bien disjointe.

Pour prouver l'inclusion de droite de (1.38), on montre que la suite de Farey $\mathfrak{F}^{(\sqrt{3}t/2)}$ contient tous les couples dont le cercle de Farey participe non trivialement à une réunion. En effet, soit $(a/c, b/d)$ un couple de fractions consécutives dans $\mathfrak{F}^{(t)}$. Ces fractions sont bien sûr incluses dans $\mathfrak{F}^{(\sqrt{3}t/2)}$ car $1/t < 2/(\sqrt{3}t)$. Maintenant, le cercle associé au médian $(a+b)/(c+d)$ participe à la réunion si $(c^2+cd+d^2)t^2 < 1$. Dans ce cas, en utilisant l'inégalité $4cd \leq (c+d)^2$,

$$(c+d)^2t^2 = (c^2+cd+d^2)t^2 + cdt^2 < 1 + \frac{(c+d)^2}{4}t^2$$

et donc $(c+d) < 2/(\sqrt{3}t)$. Ainsi, la suite de Farey $\mathfrak{F}^{(\sqrt{3}t/2)}$ décrit bien un ensemble de disques dont la réunion est $L(t)$. L'inclusion de droite de (1.38) est donc prouvée. □

1.4 Géométrie de l'ensemble de niveau du second minimum orthogonalisé.

L'étude de la géométrie de $M(u)$ ressemble dans beaucoup d'aspects celle de $L(t)$, comme nous allons le voir ici. Mais, ce sont les secteurs $\mathsf{Se}(a,c,u)$, définis dans le théorème E qui remplacent maintenant les disques de Farey.

1.4.1 Description de $M(u)$.

La caractérisation géométrique de l'ensemble $M(u)$ est donnée par le théorème suivant. Quelques commentaires suivent le théorème.

Théorème G. *Soit $u \in [0,1]$. Pour toute fraction a/c, on définit les deux ensembles*

$$\mathsf{Se}^+(a,c,u) := \mathsf{Se}(a,c,u) \cap \left\langle -\infty, \frac{a}{c} \right\rangle \qquad \mathsf{Se}^-(a,c,u) := \mathsf{Se}(a,c,u) \cap \left\langle \frac{a}{c}, \infty \right\rangle$$

appelés respectivement demi-secteur*positif et* demi-secteur négatif. *Deux cas se présentent pour le domaine $M(u) = [\mu(z) \leq u]$, selon la valeur de u :*

(i) Si $u > 1/2$, alors

$$M(u) = \mathsf{Se}(0,1,u) \cap \mathcal{B} \setminus \mathcal{F}.$$

(ii) Si $u \leq 1/2$, la portion de $M(u)$ comprise dans la u-bande de Farey $\langle a/c, b/d \rangle$ est un domaine, égal à un triangle ou à un quadrilatère convexe, qui est défini comme l'intersection de trois demi-secteurs. Plus précisement,

$$M(u) \cap \langle \frac{a}{c}, \frac{b}{d} \rangle = \mathsf{Se}^+(a,c,u) \cap \mathsf{Se}^-(b,d,u) \cap \mathsf{Se}^\epsilon(\epsilon \cdot (a-b), \epsilon \cdot (c-d), u) \tag{1.40}$$

où ϵ est le signe de $c-d$. Cette intersection (1.40) est un triangle et s'écrit tout simplement

$$M(u) \cap \langle \frac{a}{c}, \frac{b}{d} \rangle = \mathsf{Se}^+(a,c,u) \cap \mathsf{Se}^-(b,d,u)\,, \tag{1.41}$$

si et seulement si

$$(cd)u^2 \leq \frac{1}{2} \quad \vee \quad (c^2 - cd + d^2)u^2 \leq \frac{3}{4}\,, \tag{1.42}$$

et la proportion de couples (c,d) vérifiant (1.42) tend vers

$$\varrho_T = \frac{1}{2} + \frac{\pi\sqrt{3}}{12} \approx 0,95344984, \qquad (u \to 0).$$

Avant de continuer, il convient d'introduire un peu de vocabulaire et de notations accompagnés de remarques.

Définition 1.7.

(i) On appelle demi-secteurs *les ensembles*

$$\mathsf{Se}(a,c,u) \cap \left\langle -\infty, \frac{a}{c} \right\rangle, \qquad \textit{et} \quad \mathsf{Se}(a,c,u) \cap \left\langle \frac{a}{c}, \infty \right\rangle.$$

Comme dans l'énoncé du theorème G, l'ensemble $\mathsf{Se}^+(a,c,u) := \mathsf{Se}(a,c,u) \cap \left\langle -\infty, \frac{a}{c} \right\rangle$ est appelé demi-secteur *positif et $\mathsf{Se}^-(a,c,u) := \mathsf{Se}(a,c,u) \cap \left\langle \frac{a}{c}, \infty \right\rangle$ est appelé* demi-secteur négatif. *Un secteur $\mathsf{Se}(a,c,u)$, ou un demi-secteur est dit* trivial *lorsque $cu \geq 1$, c'est-à-dire lorsque $\mathsf{Se}(a,c,u) = \mathbb{H}$.*

(ii) Pour $cu < 1$, on désigne par $\theta_c(u) = \arcsin(cu)$ l'angle associé au secteur $\mathsf{Se}(a,c,u)$. Si le contexte le permet, on le désigne simplement par θ_c.

(iii) On désigne par $\ell^-_{(a,c)}(u)$ et $\ell^+_{(a,c)}(u)$ les demi-droites *qui déterminent le secteur $\mathsf{Se}(a,c,u)$,*

$$\ell^-_{(a,c)}(u) = \{x + iy \in \mathbb{H} \mid x < a/c,\, y \leq \tan(\theta_c) \cdot |x - a/c|\}$$

$$\ell^+_{(a,c)}(u) = \{x + iy \in \mathbb{H} \mid x > a/c,\, y \leq \tan(\theta_c) \cdot |x - a/c|\}.$$

Si le contexte le permet, on les désigne simplement par $\ell^-_{(a,c)}$ et $\ell^+_{(a,c)}$.

(iv) L'intersection non-vide de deux demi-secteurs non-triviaux (l'un positif, l'autre négatif) définit un triangle *Une intersection entre un triangle et un demi-secteur est un* quadrilatère *lorsqu'elle n'est pas un triangle.*

(v) La hauteur *d'un sous-ensemble $S \in \mathbb{H}$ est le nombre $\sup\{\Im(z) \ : \ z \in S\}$. Pour deux demi-secteurs $\mathsf{Se}^+(a,c,u)$ et $\mathsf{Se}^-(b,d,u)$, on désigne par $h_{c,d}$ la hauteur du triangle $\mathsf{Se}(a,c,u) \cap \mathsf{Se}(b,d,u)$.*

(vi) La base *d'un triangle ou quadrilatère est l'intervalle de l'axe horizontal (ouvert sauf contre-indication) sur lequel ce triangle (ou quadrilatère) repose. C'est l'intersection de l'adhérence du triangle ou quadrilatère avec l'axe horizontal.*

L'ensemble $M(u)$ est défini localement, via les u-bandes de Farey, par des triangles ou par des quadrilatères. Les quadrilatères sont très peu fréquents. Lorsque $u \to 0$, on trouve un quadrilatère dans moins d'un cas sur 20. Les triangles et quadrilatères qui participent à $M(u)$ ont des hauteurs assez homogènes. Elles varient entre u^2 et $2u^2/\sqrt{3}$, comme nous allons voir dans la proposition 1.10. L'ensemble $M(u)$ ne contient pas de bande horizontale au voisinage de l'axe horizontal, à la différence de ce qui se passe pour $L(t)$ (cf. proposition 1.1). Les deux propriétés – contenir une bande horizontale– ou, pour les briques de base (demi-disques, secteurs) –avoir une hauteur du même ordre de grandeur– sont des propriétés "duales" . Elles reposent toutes deux sur l'existence d'une borne inférieure pour les ordonnées des points du domaine fondamental $\mathcal{F}$.

Dans la caractérisation géométrique de $L(t)$, les réunions étaient élaguées en considérant les relations d'inclusion entre disques de Farey. Dans le cas de $M(u)$, l'idée est d'élaguer les

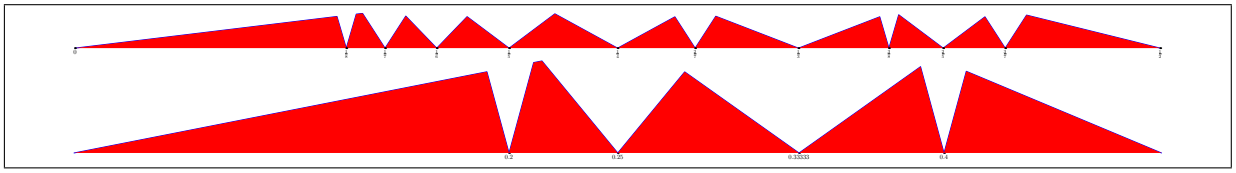

FIGURE 1.6 – Ensemble $M(u) \cap \langle 0, 1/2\rangle$. Les valeurs illustrées sont $u = 0,12$ et $u = 0,193$.

intersections, dans le sens suivant : si S est l'intersection d'une famille d'ensembles $(S_i)_{i\in I}$, et si par aileurs $S \subset \overline{S}$, alors S est l'intersection des $(S_i \cap \overline{S})$ avec $S_i \not\subset \overline{S}$. Si $\overline{S}$ est contenu dans la plupart des S_i, l'intersection définissant S peut être élaguée de façon importante. Plus précisément,

$$S = \bigcap_{i\in I} S_i \quad \text{et} \quad S \subset \overline{S} \quad \Longrightarrow \; S = \bigcap_{i\in I \,:\, \overline{S}\not\subset S_i} (S_i \cap \overline{S}). \tag{1.43}$$

Nous savons que $M(u)$ est inclus dans la bande horizontale $[\Im(z) \le 2u^2/\sqrt{3}]$. On peut donc se servir de (1.43). Le lemme 1.4 déterminera les secteurs qui interviendront effectivement dans l'intersection.

1.4.2 Position des secteurs angulaires.

Cette section étudie les possibles configurations pour l'intersection d'un secteur avec une u-bande de Farey, ainsi qu'entre secteurs. Nous commençons avec un lemme technique.

Lemme 1.3. *Soient $a/c < b/d$ deux rationnels consécutifs dans $\mathfrak{F}^{(u)}$. Alors, la hauteur du triangle $\mathsf{Se}(a,c,u) \cap \mathsf{Se}(b,d,u)$ vérifie*

$$h_{c,d} = \frac{u^2}{\sin(\theta_c + \theta_d)}.$$

Démonstration. La hauteur h d'un triangle de base b et d'angles de base α et β est donnée par

$$h = b \cdot \frac{\sin\alpha \sin\beta}{\sin(\alpha+\beta)}\,. \tag{1.44}$$

En effet, supposons que les points $(0,0)$ et $(b,0)$ sont les extrémités de la base du triangle, et que le troisième sommet a pour coordonnées (x,y), avec $y > 0$. On suppose que les angles intérieurs associés au point $(0,0)$ et $(0,b)$ sont α et β. Nous avons

$$y = x \cdot \tan\alpha = (b-x)\tan\beta\,,$$

ce qui entraîne

$$h = y = b \cdot \frac{\tan\alpha\tan\beta}{\tan\alpha + \tan\beta} = b \cdot \frac{\sin\alpha\sin\beta}{\sin(\alpha+\beta)},$$

comme voulu. Ainsi, si un triangle est formé par l'intersection de deux demi-secteurs $\mathsf{Se}^+(a,c,u)$ et $\mathsf{Se}^-(b,d,u)$ tels que $ad - bc = -1$, alors la longueur de sa base est $1/cd$ et sa hauteur $h_{c,d}$ vérifie

$$h_{c,d} = u^2/\sin(\theta_c + \theta_d), \tag{1.45}$$

comme annoncé. □

Le lemme suivant caractérise les intersections secteur-bande. Il se sert de l'argument (1.43).

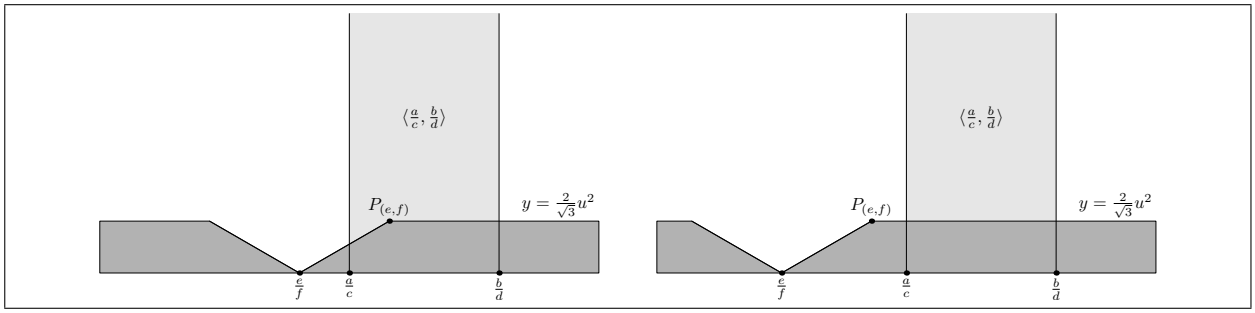

FIGURE 1.7 – La figure illustre le cas où un secteur a une influence potentielle sur $[\mu(z) \leq u] \cap \langle a/c, b/d \rangle$ (à gauche), et le cas où le rectangle est inclus dans le secteur (à droite).

Lemme 1.4. *Soit $u \leq 1/2$ et soient $a/c < b/d$ deux rationnels consécutifs dans $\mathfrak{F}^{(u)}$. Alors, les seuls secteurs dans lesquels le rectangle*

$$\mathcal{R} := \langle \frac{a}{c}, \frac{b}{d} \rangle \cap \left[\Im(z) \leq \frac{2}{\sqrt{3}} u^2 \right]$$

n'est pas inclus sont $\mathsf{Se}^+(a, c, u)$, $\mathsf{Se}^-(b, d, u)$ et peut-être $\mathsf{Se}^+(a-b, c-d, u)$ si $c > d$, ou $\mathsf{Se}^-(b-a, d-c, u)$ si $d > c$.

Démonstration. Par définition, le rectangle $\mathcal{R}$ est inclu dans les secteurs triviaux, et la discussion suivante est restreinte aux secteurs non-triviaux. Soit (e, f) un couple de $\mathcal{C}$ telle que $fu \leq 1$. La figure 1.7 illustre les situations que nous voulons identifier. Soit $P_{(e,f)}$ le point d'abscisse $x_{(e,f)}$ (voir figure 1.7), défini par l'intersection suivante

$$P_{(e,f)} = \begin{cases} \{\Im(z) = 2/\sqrt{3}u^2\} \cap \ell^+_{(e,f)} & \text{si } e/f \leq a/c \\ \{\Im(z) = 2/\sqrt{3}u^2\} \cap \ell^-_{(e,f)} & \text{si } e/f \geq b/d. \end{cases}$$

Le rectangle est inclus dans le secteur si et seulement si

$$\frac{e}{f} \in \left[-\frac{1}{2}, \frac{a}{c} \right] \quad \text{et} \quad \left| \frac{e}{f} - x_{(e,f)} \right| \leq \left| \frac{e}{f} - \frac{a}{c} \right| \tag{1.46}$$

$$\frac{e}{f} \in \left[\frac{b}{d}, \frac{1}{2} \right] \quad \text{et} \quad \left| \frac{e}{f} - x_{(e,f)} \right| \leq \left| \frac{e}{f} - \frac{b}{d} \right| . \tag{1.47}$$

Plaçons-nous dans le cas (1.46) où $e/f \leq a/c$. Dans ce cas, le couple (e, f) s'écrit $(e, f) = (ma + nb, mc + nd)$ avec $m \geq 1$, $n \leq 0$ (cf. proposition 1.5, (i)). Par ailleurs, l'abscisse $x_{(e,f)}$ satisfait

$$\left| \frac{e}{f} - x_{(e,f)} \right| = \frac{2}{\sqrt{3}} u^2 \cdot \frac{\sqrt{1 - (fu)^2}}{fu}.$$

Et la condition (1.46) s'écrit

$$\frac{|n|}{cf} \geq \frac{2}{\sqrt{3}} u^2 \cdot \frac{\sqrt{1 - (fu)^2}}{fu}$$

ou, de façon équivalente,

$$(fu)^2 + \frac{3}{4} \frac{n^2}{(cu)^2} \geq 1 . \tag{1.48}$$

Il nous faut donc montrer que pour tout couple $(e, f) = (ma + nb, mc + nd)$ distinct de (a, c) et de $(a - b, c - d)$, c'est-à-dire associé à un couple (m, n) d'entiers premiers entre eux satisfaisant

$m \geq 1$, $n < 0$, distinct de $(1, -1)$, la condition (1.48) est satisfaite. Elle est clairement satisfaite dès que $|n| \geq 2$. En effet, dans ce cas

$$(fu)^2 + \frac{3}{4}\frac{n^2}{(cu)^2} \geq (fu)^2 + \frac{3}{(cu)^2} \geq 1,$$

puisque $cu < 1$. Supposons maintenant $n = -1$ et $m > 1$, et donc $(e, f) = (ma - b, mc - d)$. Posons $x = cu$, $y = du$. Puisque (c, d) est élément de $\mathfrak{D}^{(u)}$, alors (x, y) est élement $[0, 1]^2$. La condition (1.48) s'ecrit alors

$$(mx - y)^2 + \frac{3}{4x^2} \geq 1,$$

et est clairement satisfaite si $x \leq \sqrt{3}/2$ puisque $3/(4x^2) \geq 1$. Lorsque $x > \sqrt{3}/2$, les conditions $m \geq 2, y \leq 1, x \leq 1$ entraînent l'inégalité

$$(mx - y) \geq (\sqrt{3} - 1) \qquad \text{et donc} \quad (mx - y)^2 + \frac{3}{4x^2} \geq (\sqrt{3} - 1)^2 + \frac{3}{4} \geq 1.$$

Ainsi, lorsque $e/f \leq a/c$ les seuls couples qui peuvent participer à (1.40) sont (a, c) et $(a - b, c - d)$ pourvu que $c > d$. Compte-tenu de la position du secteur $\mathsf{Se}(e, f, u)$ par rapport à la bande, il est clair que ce sont les secteurs positifs qui participent, c'est-à-dire $\mathsf{Se}^+(a, c, u)$ et $\mathsf{Se}^+(a - b, c - d, u)$ si $c > d$.
Maintenant, si $e/f \geq b/d$, le raisonnement est analogue. En effet, la condition (1.47) fournit l'inéquation

$$(fu)^2 + \frac{3}{4}\frac{m^2}{(du)^2} \geq 1,$$

qui devient (1.48) juste en changeant m et n et c et d. Ainsi, dans ce cas les couples pouvant participer à l'intersection sont (b, d) et $(b - a, d - c)$ pourvu que $d > c$. La position des secteurs assure que ce sont les secteurs négatifs qu'on utilise, c'est à dire $\mathsf{Se}^-(b, d, u)$ et $\mathsf{Se}^-(b-a, d-c, u)$ si $d > c$. Cela achève la preuve du lemme. □

Enfin, le lemme ci-dessous va permettre de préciser le résultat du lemme 1.4.

Lemme 1.5. *Soient $a/c < b/d$ deux rationnels consécutifs dans $\mathfrak{F}^{(u)}$, et soit ϵ le signe de $c - d$. On a l'équivalence suivante*

$$\mathsf{Se}^+(a, c, u) \cap \mathsf{Se}^-(b, d, u) \subset \mathsf{Se}^\epsilon(\epsilon(a - b), \epsilon(c - d), u) \Longleftrightarrow (cd)u^2 \leq \frac{1}{2} \text{ ou } (c^2 - cd + d^2)u^2 \leq 3/4.$$

Démonstration. Supposons d'abord que $c \geq d$. Nous allons comparer la hauteur $h_{c,d}$ du triangle detérminé par $\mathsf{Se}^+(a, c, u) \cap \mathsf{Se}^-(b, d, u)$ avec la hauteur $h_{c-d,d}$ du triangle déterminé par $\mathsf{Se}^+(a - b, c - d, u) \cap \mathsf{Se}^-(b, d, u)$. En s'inspirant de la figure 1.8, on conclut que la condition du lemme est vérifiée si et seulement si $h_{c,d} \leq h_{c-d,d}$. Grâce au lemme 1.3, cette condition est équivalente à

$$\sin(\theta_{c-d} + \ \theta_d) \leq \sin(\theta_c + \theta_d),$$

et cela équivaut à

$$2(cu)(du) - 1 \leq 2\sqrt{1 - (cu)^2}\sqrt{1 - (du)^2}\,. \tag{1.49}$$

Ainsi, la condition est immédiatement vérifiée si le côté gauche de (1.49) est négatif, c'est-à-dire, si $(cd)u^2 \leq 1/2$. Si $(cd)u^2 > 1/2$ alors le côté gauche de (1.49) est positif et on peut élever au carré et obtenir la condition équivalente

$$(c^2 - cd + d^2)u^2 \leq \frac{3}{4}\,.$$

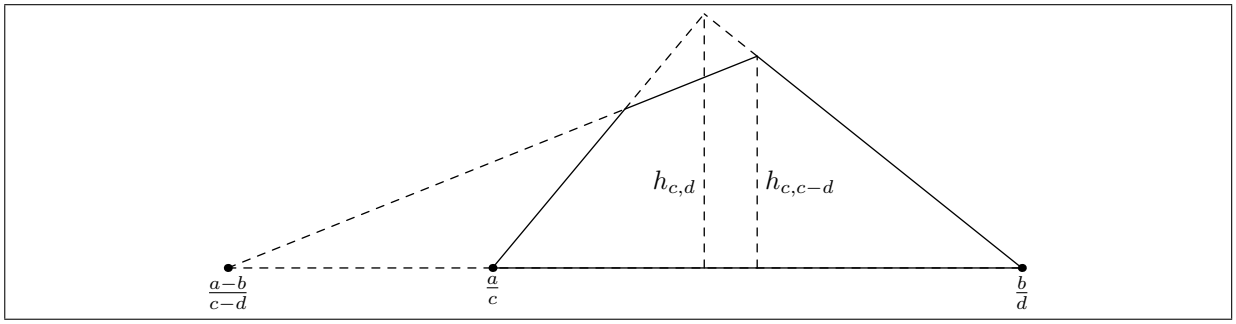

FIGURE 1.8 – Illustration des hauteurs $h_{c,d}$ et $h_{c-d,d}$, dont la comparaison permet de déterminer la nature de $M(u) \cap \langle a/c, b/d \rangle$.

En conclusion, pour $c > d$ nous aurons (1.41) si et seulement si

$$(cd)u^2 \leq \frac{1}{2} \vee (c^2 - cd + d^2)u^2 \leq \frac{3}{4}\,.$$

La condition étant symétrique en c et d, elle est toujours valable pour $c < d$, ce qui établit le résultat. □

1.4.3 Preuve de la caractérisation géométrique de $M(u)$.

Nous passons à la preuve du théorème **G**. La caractérisation globale E de l'ensemble $M(u)$ fournit l'égalité

$$M(u) = \left(\bigcap_{(e,f)\in\mathcal{C}} \mathsf{Se}(e,f,u) \right) \cap \mathcal{B} \setminus \mathcal{F}.$$

On remarque tout d'abord que les secteurs $\mathsf{Se}(e,f,u)$ pour lesquels $fu \geq 1$, sont égaux à $\mathbb{H}$ et qu'ils sont donc superflus dans l'intersection. Dans le cas $u > 1/2$, cela élimine tous les secteurs sauf $\mathsf{Se}(0,1,u)$, ce qui prouve (i). Lorsque $u \leq 1/2$, les rationnels $-1/2$ et $1/2$ appartiennent à la suite $\mathfrak{F}^{(u)}$, et on peut partitionner $M(u)$ en tranches de la forme

$$M(u) \cap \langle \frac{a}{c}, \frac{b}{d} \rangle = \left(\bigcap_{(e,f)\in\mathcal{C}} \mathsf{Se}(e,f,u) \cap \langle \frac{a}{c}, \frac{b}{d} \rangle \right) \cap \mathcal{B} \setminus \mathcal{F},$$

où a/c et b/d sont consécutifs dans la suite de Farey $\mathfrak{F}^{(u)}$. Cette intersection se simplifie en notant que puisque $M(u)$ est contenu dans la bande horizontale $[y \leq 2u^2/\sqrt{3}]$ (cf. prop. 1.1), sa hauteur est, dans le cas actuel où $u \leq 1/2$, plus petite que $\sqrt{3}/2$, ce qui rend superflue l'intersection avec $\mathcal{B} \setminus \mathcal{F}$. Avec ces observations, il ne nous reste qu'à élaguer

$$M_{c,d}(u) := M(u) \cap \langle \frac{a}{c}, \frac{b}{d} \rangle = \bigcap_{(e,f)\in\mathcal{C}} \mathsf{Se}(e,f,u) \cap (\langle \frac{a}{c}, \frac{b}{d} \rangle \cap [y \leq 2u^2/\sqrt{3}]). \tag{1.50}$$

Le lemme 1.4 caractérise les couples (e,f) pour lesquels les secteurs $\mathsf{Se}(e,f,u)$ contiennent le rectangle $(\langle \frac{a}{c}, \frac{b}{d} \rangle \cap [y \leq 2u^2/\sqrt{3}])$. Ces secteurs peuvent donc être éliminés de l'intersection (1.50), qui se réduit alors à

$$M_{c,d}(u) = \mathsf{Se}^{+}(a,c,u) \cap \mathsf{Se}^{-}(b,d,u) \cap \mathsf{Se}^{\epsilon}(\epsilon\cdot(a-b), \epsilon\cdot(c-d), u) \tag{1.51}$$

où ϵ est le signe de $c-d$.

Le lemme 1.5 caractérise les cas où (1.51) est réduite à l'intersection des secteurs $\mathsf{Se}^+(a,c,u)$ et $\mathsf{Se}^-(b,d,u)$, et maintenant on calcule la proportion de ces cas. Le calcul de la proportion de couples $(c,d) \in \mathfrak{D}^{(u)}$ pour lesquels la section $M_{c,d}(u)$ est un triangle repose sur l'application du théorème 1.2. Définissons les ensembles Δ, Δ_T comme suit

$$\Delta := \{(x,y) \in [0,1]^2;\ \ x+y>1\}, \quad \Delta_T = \{(x,y) \in \Delta;\ \ xy \le \frac{1}{2} \quad \text{ou} \quad x^2-xy+y^2 \le \frac{3}{4}\}.$$

Alors, le couple (c,d) appartient à $\mathfrak{D}^{(u)}$ si et seulement si le couple (cu,du) appartient à Δ et $M_{c,d}(u)$ est un triangle si et seulement le couple (cu,du) appartient à Δ_T. Donc, la proportion $\varrho_{\mathrm{T}}(u)$ de paires (c,d) de $\mathfrak{D}^{(u)}$ pour lesquels $M_{c,d}(u)$ est un triangle vérifie, avec les notations du théorème 1.2

$$\varrho_{\mathrm{T}}(u) = \frac{F[1,\Delta_T](u)}{F[1,\Delta](u)}. \tag{1.52}$$

Les sommes $F[1,\Delta](u)$ sont naturellement bornées et la fonction 1 est Riemann-intégrable au sens propre dans Δ. En appliquant le théorème 1.2, on obtient

$$\lim_{u\to 0} \varrho_{\mathrm{T}}(u) = \frac{I[1,\Delta_T]}{I[1,\Delta]}.$$

Les intégrales en jeu se calculent facilement,

$$I[1,\Delta_T] = \frac{1}{4} + \frac{\pi\sqrt{3}}{24} \qquad \text{et} \qquad I[1,\Delta] = \frac{1}{2},$$

et donc

$$\lim_{u\to 0} \varrho_{\mathrm{T}}(u) = \frac{1}{2} + \frac{\pi\sqrt{3}}{12} \approx 0,95344984.$$

Ainsi, nous avons établi (ii) et donc le théorème **G**.

1.4.4 Encadrement de $M(u)$

Tout d'abord, notons que la convexité de $[\mu(z) \le u] \cap \langle a/c, b/d\rangle$ découle automatiquement de (1.40). En effet, un demi-secteur est une intersection de demi-plans, et il est donc convexe[6]. Il s'en suit que $[\mu(z) \le u] \cap \langle a/c, b/d\rangle$ est une intersection d'ensembles convexes, et il est donc lui-même convexe.

Proposition 1.10. *Il existe deux ensembles $\underline{M}(u)$ et $\overline{M}(u)$ encadrant $M(u)$,*

$$\underline{M}(u) \subseteq M(u) \subseteq \overline{M}(u)$$

et vérifiant les propriétés suivantes :

(i) Tous les deux sont des réunions de triangles disjoints basés sur l'axe réel, et les bases de ces triangles forment une partition du segment $[-1/2, 1/2]$,

(ii) La hauteur de $\underline{M}(u)$ est au moins égale à u^2 tandis que la hauteur de $\overline{M}(u)$ est au plus égale à $2u^2/\sqrt{3}$.

6. Une intersection d'ensembles convexes est un ensemble convexe.

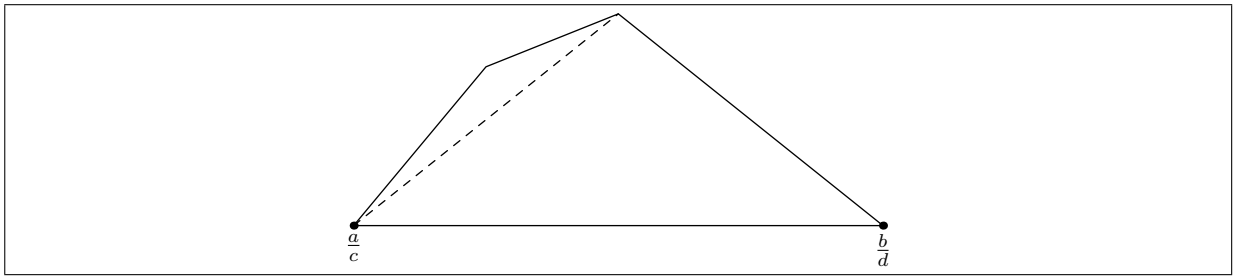

FIGURE 1.9 – Les quadrilatères convexes comme les nôtres contiennent un triangle de même base et de même hauteur.

Démonstration. L' ensemble $\underline{M}(u)$ se construit facilement à partir de $M(u)$: pour chaque couple de fractions $a/c, b/d$ successives dans la suite de Farey $\mathfrak{F}^{(u)}$, nous considérons $M_{c,d}(u)$. Si ce dernier ensemble est un triangle, on le garde tel quel, et si c'est un quadrilatère, on le transforme en un triangle de base $[a/c, b/d]$ obtenu en joignant les points a/c et b/d au plus haut sommet du quadrilatère, comme le montre la figure 1.9. Un tel ensemble $\underline{M}(u)$ est, par construction, formé par des triangles disjoints dont les bases forment une partition de $[-1/2, 1/2]$. La hauteur h de chacun de ces triangles vérifie toujours

$$h = u^2/\sin(\theta_- + \theta_+) \geq u^2,$$

où θ_+ et θ_- sont les angles des secteurs qui déterminent la hauteur. Ainsi, l'ensemble $\underline{M}(u)$ vérifie bien les propriétés annoncées.

L'ensemble $\overline{M}(u)$ se construit à partir des triangles couvrants $\mathsf{TC}(c, d, u)$ définis par

$$\mathsf{TC}(c,d,u) = \begin{cases} \mathsf{Se}^+(a,c,u) \cap \mathsf{Se}^-(b,d,u) & \text{si } M_{c,d}(u) \text{ est un triangle.} \\ \mathsf{Se}^+(a-b,c-d,u) \cap \mathsf{Se}^-(b,d,u) & \text{si } M_{c,d}(u) \text{ est un quadrilatère et } c > d \\ \mathsf{Se}^+(a,c,u) \cap \mathsf{Se}^-(b-a,d-c,u) & \text{si } M_{c,d}(u) \text{ est un quadrilatère et } d > c, \end{cases},$$

et nous montrons que l'ensemble défini par

$$\overline{M}(u) = \bigcup_{(c,d)\in\mathfrak{D}^{(u)}} \mathsf{TC}(c,d,u) \tag{1.53}$$

convient à nos objectifs. Pour cela, nous allons prouver que pour deux couples (c, d) et (f, h) de $\mathfrak{D}^{(u)}$, les triangles $\mathsf{TC}(c, d, u)$ et $\mathsf{TC}(f, h, u)$ sont soit disjoints, soit inclus l'un dans l'autre.

Lemme 1.6. *Les triangles couvrants possèdent les propriétés suivantes :*

(i) Leurs bases sont des intervalles inclus dans $[-1/2, 1/2]$ *et leur sommet (non réel) appartient à* $M(u)$.

(ii) La partie de $M(u)$ *comprise dans la bande verticale définie par la base d'un triangle couvrant est comprise dans le triangle couvrant.*

(iii) Deux triangles couvrants sont soit disjoints, soit inclus l'un dans l'autre. Ils sont disjoints si et seulement si leurs bases sont disjointes, et inclus l'un dans l'autre si et seulement si leurs bases sont incluses l'une dans l'autre.

Preuve. Fixons un triangle couvrant $\mathsf{TC}(c, d, u)$. Tout d'abord, si $M_{c,d}(u)$ est un triangle, (i) et (ii) sont évidentes. On suppose donc que $M_{c,d}(u)$ est un quadrilatère, et on prouve (i) et (ii). Pour (i), le fait que les bases soient des intervalles inclus dans $[-1/2, 1/2]$ vient du fait que les

sommets des secteurs en jeu y sont compris. Par ailleurs, le fait que le sommet (non réel) du triangle couvrant appartienne à $M(u)$ est une conséquence directe de la condition d'existence d'un quadrilatère (cf. lemme 1.5 et figure 1.8). Maintenant, le point (ii) est trivial, car, à l'intérieur de la bande en question, $M(u)$ est par définition l'intersection d'une famille de secteurs, et le triangle couvrant n'est qu'une intersection partielle de ces secteurs.

Maintenant nous prouvons (iii). Soient $\mathsf{TC}(c,d,u)$ et $\mathsf{TC}(e,f,u)$ deux triangles couvrants quelconques. Les bases des ces triangles ont, par définition, des extrémités qui sont des fractions adjacentes. Donc, grâce à la propriété (iii) de la proposition 1.5, les intervalles associés sont soit disjoints, soit l'un inclus dans l'autre. Maintenant, si leurs bases sont disjointes, les triangles sont disjoints puisque leurs angles de base sont aigus ou droits, mais jamais obtus. Par ailleurs, si la base de $\mathsf{TC}(c,d,u)$ contient la base de $\mathsf{TC}(e,f,u)$, alors il contient aussi le sommet de $\mathsf{TC}(e,f,u)$. En effet, par (i), le sommet de $\mathsf{TC}(c,d,u)$ appartient à $M(u)$, et donc, par (ii), le triangle $\mathsf{TC}(c,d,u)$ doit contenir ce sommet car le triangle $\mathsf{TC}(e,f,u)$ est inclus dans la bande verticale définie par sa base et donc par celle définie par la base de $\mathsf{TC}(c,d,u)$. En conclusion, $\mathsf{TC}(c,d,u)$ contient les trois sommets de $\mathsf{TC}(e,f,u)$, et puisqu'il est convexe, il contient l'envolture convexe de ces trois points, qui est égale à $\mathsf{TC}(e,f,u)$. La preuve est donc achevée. □

Ainsi, la réunion (1.53) peut s'élaguer en gardant seulement les triangles couvrants maximaux pour l'inclusion. Les triangles maximaux pour l'inclusion sont disjoints par définition et leur réunion est égale à $\overline{M}(u)$. La réunion de leurs bases est égale à l'intervalle $[-1/2, 1/2]$ (à un ensemble de mesure nulle près), et leur hauteur est au plus l'hauteur de $M(u)$, qui est bornée par $2u^2/\sqrt{3}$, grâce à la proposition 1.1. Cela achève la preuve. □

1.5 Géométrie de l'ensemble de niveau du défaut d'Hermite.

Dans cette section nous décrivons la géométrie de l'ensemble de niveau lié au défaut d'Hermite.

1.5.1 Description de $G(\rho)$.

La géométrie de $G(\rho)$ est liée aux disques de Ford, introduits par L. Ford en 1938 [24], dans le but de visualiser géométriquement des résultats arithmétiques. L'idée lui est venue à partir des études de Bianchi à propos du groupe de Picard, où des familles de sphères invariantes intervenaient. Les disques de Ford correspondent au paramètre $\rho = 1$ dans la notation actuelle.

Les disques de Ford permettent de représenter des fractions. Si a/c est une fraction sous forme irréductible, on lui associe le disque de rayon $1/2c^2$, tangent à l'axe réel en a/c. Ces disques sont disjoints, sauf peut-être pour leur frontière, et on montre que deux disques sont tangents si et seulement si les fractions associées sont adjacentes. Cette propriété s'avère utile pour visualiser des développements en fractions continues en tant que suites de disques adjacents. On peut, par exemple, "calculer" des approximations diophantiennes visuellement. L'article de Ford décrit ces algorithmes en détail, et il présente aussi des *sphères de Ford* pour représenter des fractions complexes.

Le résultat suivant a été fourni par Laville et Vallée [45].

Théorème 1.3. *L'ensemble $G(\rho)$, donné par*

$$G(\rho) = \left(\bigcup_{(a,c)\in\mathcal{C}} \mathsf{Fo}(a,c,\rho) \right) \cap \mathcal{B} \setminus \mathcal{F} \tag{1.54}$$

satisfait les propriétés suivantes :

(i) Lorsque $\rho \leq 1$, il s'agit d'une réunion quasi-disjointe de ρ-disques de Ford, sauf pour les couples $(-1,2)$ et $(1,2)$ où il s'agit de demi-disques :

$$G(\rho) = \mathsf{Fo}^+(-1,2,\rho) \cup \left(\bigcup_{(a,c)\in\mathcal{C},\ c\neq 2} \mathsf{Fo}(a,c,\rho) \right) \cup \mathsf{Fo}^-(1,2,\rho).$$

(ii) Lorsque $\rho > 1$, il s'agit d'une réunion non-disjointe, où les intersections non-vides entre ρ-disques correspondent à une paire de disques adjacents.

Preuve. Tout d'abord, on montre que si $a/c < b/d$ sont adjacents, alors le disque associé au médian $\mathsf{Fo}(a+b, c+d, \rho)$ est inclus dans $\langle a/c, b/d\rangle$, pour tout ρ. En effet, étant donné ρ, en comparant les distances du médian aux extrémités et le rayon du cercle, nous avons

$$\min\left\{ \frac{1}{c(c+d)},\ \frac{1}{(c+d)d} \right\} > \frac{1}{\sqrt{3}(c+d)^2} \geq \frac{\rho}{2(c+d)^2},$$

ce qui montre que les disques sont bien inclus dans la bande. D'après ce fait, on conclut que tous les disques de Ford de (1.54) ; sauf ceux associés à $-1/2$, $1/2$, et celui de $0/1$ si $\rho \geq 1$; sont compris dans $\langle -1/2, 1/2\rangle$. On vérifie également que les ρ-disques toujours inclus dans $\langle -1/2, 1/2\rangle$ sont aussi inclus dans $\mathcal{B} \setminus \mathcal{F}$, car leur hauteur est au plus $\rho/9 < 2/(9\sqrt{3}) < \sqrt{3}/2$. On vérifie immédiatement que le disque $\mathsf{Fo}(0,1,\rho)$ est inclus dans $\mathcal{B} \setminus \mathcal{F}$ ssi $\rho \leq 1$. Pour montrer (i), il ne reste qu'à montrer que la réunion est disjointe.

Considérons encore $a/c < b/d$ adjacents ainsi que deux disques $\mathsf{Fo}(a,c,\rho)$ et $\mathsf{Fo}(b,d,\rho)$. Pour que les disques s'intersectent, la distance entre leurs centres doit être plus petite que la somme de leurs rayons ; cela s'écrit

$$\left(\frac{a}{c} - \frac{b}{d}\right)^2 + \left(\frac{\rho}{2c^2} - \frac{\rho}{2d^2}\right)^2 \leq \left(\frac{\rho}{2c^2} + \frac{\rho}{2d^2}\right)^2,$$

et on peut simplifier pour arriver à

$$|ad - bc| \leq \rho.$$

Ainsi, si a/c et b/d ne sont pas adjacents, les disques sont disjoints, et cela pour tout $0 \leq \rho \leq 2/\sqrt{3}$. Par ailleurs, si $|ad-bc| = 1$, les disques sont disjoints si $\rho < 1$, tangents si $\rho = 1$, et leur intersection est d'intérieur non-vide si $\rho > 1$. Cela achève la preuve. □

1.6 Conclusion

Nous avons étudié les domaines de niveau associés aux paramètres λ, μ et γ, et nous avons montré, en suivant les résultats de Laville et Vallée [45], qu'ils sont liés de façon étroite d'une part à des objets classiques de la géométrie du demi-plan de Poincaré, comme le sont les disques de Ford, et d'autre part à des objets moins classiques mais d'une nature similaire, comme le sont les disques de Farey et les secteurs. Les caractérisations fournies dans les théorèmes **F**, **G** et **1.3** vont nous permettre, dans le chapitre suivant, d'évaluer la distribution de probabilité des paramètres de la base de sortie de l'algorithme de Gauss.

Chapitre 2

Étude probabiliste de la configuration de sortie.

Sommaire

Il s'agit, dans ce chapitre, de répondre à la question : *étant donnée une distribution de probabilité sur l'ensemble des bases d'entrée de l'algorithme, que peut-on dire de la distribution de la géométrie des bases de sortie ?* Ce chapitre offre une réponse en quatre volets. La première section décrit la densité de la variable aléatoire "sortie de l'algorithme de Gauss". Elle établit un lien entre la densité de sortie dans le modèle de valuation r et les séries d'Eisenstein. Le reste du chapitre est consacré à l'étude des fonctions de répartition des trois principaux paramètres de

sortie pour l'algorithme GAUSS-POSITIF : premier minimum λ, second minimum orthogonalisé μ et défaut d'Hermite γ. Nous utilisons ici de manière cruciale les caractérisations géométriques obtenues dans le chapitre précédent. Nous travaillons avec une densité de valuation r, qui nous permet de quantifier précisément l'influence de la qualité des bases d'entrée sur la qualité des bases de sortie. Nous terminons ce chapitre en comparant les résultats et en en décrivant les conséquences.

2.1 Préliminaires pour l'étude probabiliste

2.1.1 Fonctions arithmétiques

Dans la suite, certaines séries de Dirichlet, liées de près à la fonction ζ de Riemann, arrivent naturellement dans notre contexte, et nous aurons besoin d'estimer leurs sommes partielles. La série de Riemann

$$\zeta(s) = \sum_{n\geq 1} \frac{1}{n^s} \qquad \text{lorsque } \Re(s) > 1,$$

et la fonction indicatrice d'Euler $\varphi(n) = \text{card}\{x \in \mathbb{N};\ \ 1 \leq x \leq n, (x,n) = 1\}$. vont jouer un rôle dans la suite, et nous utiliserons la proposition suivante (dont la preuve se trouve en exercice dans [7]).

Proposition 2.1. *On considère, pour $\alpha \in \mathbb{R}$, la série S_α de terme général $\varphi(c)c^{-\alpha}$.*

(i) Pour $\alpha > 2$, la série est convergente et sa somme s'exprime en fonction de la fonction ζ sous la forme suivante

$$\sum_{c\geq 1} \frac{\varphi(c)}{c^\alpha} = \frac{\zeta(\alpha-1)}{\zeta(\alpha)}.$$

(ii) Pour $\alpha \leq 2$, la série est divergente, et les sommes partielles de la série S_α admettent les équivalents suivants

$$\sum_{c\leq x} \frac{\varphi(c)}{c^2} \quad \sim_{x\to\infty} \quad \frac{1}{\zeta(2)} \log x \tag{2.1}$$

$$\sum_{c\leq x} \frac{\varphi(c)}{c^\alpha} \quad \sim_{x\to\infty} \quad \frac{1}{\zeta(2)} \frac{x^{2-\alpha}}{2-\alpha} \qquad \textit{pour } 1 < \alpha < 2. \tag{2.2}$$

2.1.2 Mesure des ensembles de base

Nous travaillons ici, rappelons-le, avec la mesure ν_r associée à la densité standard de valuation r, de la forme $f_r(x,y) = y^r$ avec $r > -1$. Elle donne lieu, sur $\mathcal{B} \setminus \mathcal{F}$, à une probabilité $\mathbb{P}_{(r)}$ une fois normalisée par $A(r) := \nu_r[\mathcal{B} \setminus \mathcal{F}]$. Nous calculons d'abord les mesures par rapport à la la densité non normalisée $f_r(x,y) = y^r$. Nous normaliserons ensuite.

Proposition 2.2. *Dans l'ensemble $\mathbb{H}$, muni de la mesure ν_r de valuation r,*

(i) La mesure d'un demi-disque C_ρ de rayon ρ centré en $x = 0$, et la mesure d'une portion de demi-disque $C_\rho(a)$ limitée par les verticales $x = 0$ et $x = \rho a$ sont respectivement

$$\nu_r[C_\rho] = \frac{1}{2(r+1)}\rho^{r+2}B((r+3)/2, 1/2), \qquad \nu_r[C_\rho(a)] = \frac{1}{2(r+1)}\rho^{r+2}B(a^2; (r+3)/2, 1/2)$$

(ii) la mesure $\nu_r[T_{b,h}]$ d'un triangle $T_{b,h}$ de base b et de hauteur h basé sur l'axe réel est égale à

$$\nu_r[T_{b,h}] = I[f_r, T_{b,h}] = \frac{1}{(r+1)(r+2)} b h^{r+1}$$

(iii) La mesure $\nu_r[D_\rho]$ d'un disque D_ρ de rayon ρ tangent à l'axe réel, la mesure de la portion $D_\rho(a)$ du disque D_ρ en dessous la droite d'équation $y = 2\rho a$ avec $a \in [0,1]$ sont respectivement

$$\nu_r[D_\rho] = 2(2\rho)^{r+2}\, B(r + (3/2), 3/2), \qquad \nu_r[D_\rho(a)] = 2(2\rho)^{r+2}\, B(a; r + (3/2), 3/2).$$

Démonstration. (i). La mesure $\nu_r[C_\rho(a)]$ de la portion $C_\rho(a)$ du demi-disque C_ρ limitée par les verticales $x = 0$ et $x = \rho a$ s'écrit comme

$$\nu_r[C_\rho(a)] = \int_0^{\rho a} dx \int_0^{(\rho^2 - x^2)^{1/2}} y^r dy = \frac{1}{2(r+1)} \rho^{r+2} \int_0^{a^2} (1-t^2)^{(r+1)/2} t^{-1/2} dt$$

et s'exprime finalement en fonction d'une intégrale beta incomplète sous la forme

$$\nu_r[C_\rho(a)] = \frac{1}{2(r+1)} \rho^{r+2} B(a^2; (r+3)/2, 1/2)$$

tandis que $\nu_r[C_\rho]$ s'exprime sus la forme d'une fonction beta

$$\nu_r[C_\rho] = \frac{1}{2(r+1)} \rho^{r+2} B((r+3)/2, 1/2).$$

(ii) Il suffit de calculer la mesure d'un triangle rectangle $T_{b,h}$ de base b et de hauteur h basé sur l'axe réel, qui s'écrit comme

$$\nu_r[T_{b,h}] = \int_0^h y^r \frac{b}{h}(h-y) dy = \frac{b}{h} h^{r+2} \left[\frac{1}{r+1} - \frac{1}{r+2} \right] = \frac{1}{(r+1)(r+2)} b h^{r+1}.$$

(iii) La mesure $\nu_r[D_\rho]$ d'un disque D_ρ tangent à l'axe réel, à l'origine, de rayon ρ s'écrit comme

$$\nu_r[D_\rho] = 2 \int_0^{2\rho} y^r \left(2\rho y - y^2\right)^{1/2} dy,$$

et se ramène, avec le changement de variable $y = 2\rho t$, à une intégrale beta de la forme

$$\nu_r[D_\rho] = 2(2\rho)^{r+2} \int_0^1 t^{r+1/2} (1-t)^{1/2} dt = 2(2\rho)^{r+2} B(r + (3/2), 3/2).$$

On voit donc aussi, en même temps, que la mesure $\nu_r[D_\rho(a)]$ s'exprime à l'aide d'une fonctions beta incomplète,

$$\nu_r[D_\rho(a)] = 2(2\rho)^{r+2} B(a; r + (3/2), 3/2).$$

□

Avec le formulaire sur les fonctions beta, donné dans la section 3.3.1 de la partie I, et le calcul de la constante de normalisation $A(r)$, qui se calcule directement avec le point (i) de la proposition 2.2, on obtient facilement le résultat suivant :

Proposition 2.3. *La constante de normalisation $A(r)$ vérifie*

$$A(r) := I[f_r, \mathcal{B} \setminus \mathcal{F}] = \nu_r[\mathcal{B} \setminus \mathcal{F}] = 2\nu_r[C_1(1/2)] = \frac{1}{r+1} B(1/4; (r+3)/2, 1/2). \tag{2.3}$$

Dans l'espace probabilisé défini par l'ensemble $\mathcal{B} \setminus \mathcal{F}$, la tribu borélienne et la probabilité $\mathbb{P}_{(r)}$ associée à la densité standard de valuation r, les domaines $T_{b,h}, D_\rho, C_\rho$ ont les mesures suivantes

$$\mathbb{P}_{(r)}[C_\rho] = A_1(r)\rho^{r+2} \qquad \mathbb{P}_{(r)}[T_{b,h}] = A_2(r) b h^{r+1}, \qquad \mathbb{P}_{(r)}[D_\rho] = A_3(r)(2\rho)^{r+2}$$

où les constantes $A_1(r), A_2(r), A_3(r)$ s'expriment en fonction de $A(r) = \nu_r[\mathcal{B} \setminus \mathcal{F}]$ sous la forme suivante :

$$A_1(r) = \frac{\sqrt{\pi}}{A(r)(r+1)} \frac{\Gamma((r+3)/2)}{\Gamma((r/2)+2)} \quad A_2(r) = \frac{1}{A(r)(r+1)(r+2)} \quad A_3(r) = \frac{\sqrt{\pi}}{A(r)} \frac{\Gamma(r+(3/2))}{\Gamma(r+3)}. \tag{2.4}$$

Lorsque $r \to -1$ les constantes $A(r), A_1(r), A_2(r)$ et $A_3(r)$ vérifient

$$A(r) \sim \frac{1}{r+1}, \qquad A_1(r) \to 2 \qquad A_2(r) \to 1 \qquad A_3(r) \sim \pi(r+1). \tag{2.5}$$

2.2 Densité de sortie

L'algorithme de Gauss définit naturellement une fonction $\mathfrak{G}$ de son ensemble des entrées $\mathcal{E}$ dans son ensemble des sorties $\mathcal{S}$. Si z est l'entrée de l'algorithme, alors la sortie $\widehat{z}$ s'écrit comme $\widehat{z} := \mathfrak{G}(z)$. Lorsque l'on munit l'ensemble des entrées $\mathcal{E}$ d'une probabilité associée à une densité f, la fonction $\mathfrak{G}$ est alors une variable aléatoire à valeurs dans $\mathcal{S}$, et il s'agit ici de déterminer sa loi. Cette section est consacrée à déterminer la densité de probabilité de $\mathfrak{G}$, dite *densité de sortie* et d'expliquer comment elle est liée à la densité d'entrée.

La section comporte deux principaux résultats. Le premier résultat, établi dans le théorème 2.1, est valide pour chacun des trois algorithmes Gauss-positif, Gauss-aigu, Gauss-interne, et pour une densité d'entrée quelconque. Il exprime la densité de sortie en fonction de la densité d'entrée. Le second résultat, établi dans le théorème **H**, est centré sur le double cas particulier : l'algorithme considéré est l'algorithme Gauss-positif, et la densité d'entrée est une densité standard de valuation r. On montre alors que la densité de sortie fait intervenir explicitement des séries d'Eisenstein, objets classiques en théorie analytique des nombres. Ce résultat démontre aussi qu'il y a une relation forte entre la densité de sortie de Gauss-positif et la mesure de Haar du demi-plan de Poincaré, puisque, lorsque la valuation r tend vers -1, la densité de sortie tend vers la densité de la mesure de Haar.

2.2.1 Expression générale de la densité de sortie.

Le premier théorème donne une expression formelle générale pour la densité de sortie. Le même résultat a déjà été prouvé pour l'algorithme Gauss-interne dans la proposition 2.3 de la partie II, dans le langage des opérateurs de transfert.

Théorème 2.1. *Considérons un des trois algorithmes* Gauss-positif, Gauss-aigu, *ou encore* Gauss-interne, *muni d'une densité d'entrée f sur $\mathcal{B} \setminus \mathcal{F}$ pour* Gauss-positif, *$\widetilde{\mathcal{B}} \setminus \widetilde{\mathcal{F}}$ pour* Gauss-aigu, *ou $\mathcal{D}$ pour* Gauss-interne. *Alors la densité de sortie $\widehat{f}$ sur l'ensemble $\mathcal{F}$ pour*

GAUSS-POSITIF, $\widetilde{\mathcal{F}}$ *pour* GAUSS-AIGU, $\widetilde{\mathcal{B}} \setminus \mathcal{D}$ *pour* GAUSS-INTERNE *s'exprime en fonction de la densité d'entrée* f *sous la forme*

$$\hat{f}(\hat{x},\hat{y}) = \sum_h |h'(\hat{z})|^2 f \circ \underline{h}(\hat{x},\hat{y}) = \sum_h |h'(\hat{z})|^2 f \circ h(\hat{z})\,,$$

où la somme porte sur l'ensemble des branches inverses de l'algorithme : $\mathcal{G}$ *pour l'algorithme* GAUSS-POSITIF, $\widetilde{\mathcal{G}}$ *pour* GAUSS-AIGU, $\mathcal{H}^+$ *pour* GAUSS-INTERNE.

Démonstration. Nous faisons la preuve dans le cas de l'algorithme GAUSS-POSITIF, en suivant les mêmes arguments que dans la proposition 2.3 de la partie II. La preuve est analogue pour les deux autres algorithmes. Considérons une partie mesurable $\mathcal{A}$ de l'ensemble $\mathcal{F}$ de sortie, et calculons la probabilité qu'une exécution de l'algorithme finisse dans $\mathcal{A}$, lorsque les entrées sont choisies selon la densité f. L'égalité suivante a lieu par définition même de l'algorithme et de l'ensemble $\mathcal{G}$ de ses branches inverses

$$[\hat{z} \in \mathcal{A}] = \bigcup_{h\in\mathcal{G}} [z \in h(\mathcal{A})].$$

Toujours par définition de l'algorithme, ces parties $h(\mathcal{A})$ sont disjointes deux à deux, et donc, la mesure de $\mathcal{A}$ est égale à la somme des mesures des ensembles $h(\mathcal{A})$,

$$\iint_{\mathcal{A}} \hat{f}(\hat{x},\hat{y})d\hat{x}d\hat{y} = \sum_{h\in\mathcal{G}} \iint_{\underline{h}(\mathcal{A})} f(x,y)dxdy\,.$$

Alors, le changement de variable $(\hat{x},\hat{y}) = \underline{h}(x,y)$ dont le jacobien est calculé dans le lemme 2.1 de la partie II, et l'interversion de la somme et de l'intégrale conduisent à la relation

$$\sum_{h\in\mathcal{G}} \iint_{\mathcal{A}} |h'(\hat{z})|^2 f \circ \underline{h}(\hat{x},\hat{y})d\hat{x}d\hat{y} = \iint_{\mathcal{A}} \left(\sum_{h\in\mathcal{G}} |h'(\hat{z})|^2 f \circ \underline{h}(\hat{x},\hat{y}) \right) d\hat{x}d\hat{y}\,,$$

qui termine la preuve. □

2.2.2 La mesure de Haar sur $SL_2(\mathbb{R})$ et les réseaux aléatoires.

Dans la section 3.2.2 de la partie I, nous avons décrit une construction de Siegel qui définit une probabilité naturelle sur les réseaux. Dans ce paragraphe, nous rappelons l'expression explicite de cette probabilité, en dimension 2, lorsque chaque réseau est décrit par un élément $x + iy$ de $\mathbb{F}$. Nous suivons l'exposition de [44, p. 41-43].

Le quotient $SL_2(\mathbb{R})/K$ de $SL_2(\mathbb{R})$ par le groupe K des matrices orthogonales 2×2 est en bijection avec le demi-plan $\mathbb{H}$. Considérons en effet l'application Φ définie par

$$\Phi : \sigma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto \sigma[i] = \frac{ai+b}{ci+d}.$$

Comme Φ est clairement surjective, et que le groupe K est l'ensemble des matrices pour lesquelles $\sigma[i] = i$, l'application Φ passe au quotient et définit bien la bijection cherchée. On transporte alors la mesure de Haar de $SL_2(\mathbb{R})$ en une mesure sur $\mathbb{H}$, de la forme

$$\frac{dxdy}{y^2}.$$

On vérifie bien que cette mesure est invariante dans la représentation du demi-plan. En effet, soit h une homographie induite par un élément de $SL_2(\mathbb{R})$, de la forme

$$h(z) = \frac{az+b}{cz+d},$$

et considérons la mesure qui résulte d'un changement de variables $(x,y) := \underline{h}(x,y)$ où $\underline{h}$ est l'interprétation de h comme fonction de $\mathbb{R}^2$ en $\mathbb{R}^2$ (cf. section 2.2.5 de la partie II). Un calcul direct, utilisé à maintes reprises dans la thèse, donne

$$\Im(h(z)) = \frac{y}{|cz+d|^2} = y \cdot |h'(z)|,$$

et le lemme 2.1 de la partie II implique les égalités

$$\mathrm{Jac}(\underline{h})(x,y)\,\frac{dy}{\Im(h(z))^2}dx = |h'(z)|^2\frac{dy}{(y\cdot|h'(z)|)^2}dx = \frac{dxdy}{y^2},$$

ce qui montre l'invariance de la mesure. Par ailleurs, la construction de Siegel identifie l'ensemble des réseaux avec le quotient $SL_2(\mathbb{R})/SL_2(\mathbb{Z})$, qui, dans le demi-plan de Poincaré, s'identifie naturellement au domaine fondamental $\mathcal{F}$. Comme l'intégrale de la mesure invariante sur $\mathcal{F}$ vaut

$$\iint_{\mathcal{F}} \frac{dxdy}{y^2} = \int_{-1/2}^{1/2}\int_{\sqrt{1-x^2}}^{\infty}\frac{dydx}{y^2} = \frac{\pi}{3},$$

la mesure de probabilité sur les réseaux est finalement associée à la densité par rapport à la mesure de Lebesgue suivante, appelée densité de Haar,

$$\eta(x,y) := \frac{3}{\pi}\frac{1}{y^2}. \tag{2.6}$$

Cette densité apparaît naturellement dans le calcul explicite de la densité de sortie de l'algorithme Gauss-positif, que l'on présente maintenant.

2.2.3 Le cas de l'algorithme Gauss-positif et de la densité standard de valuation r : lien avec les séries d'Eisenstein et la mesure de Haar.

Ici, nous considérons le cas de l'algorithme Gauss-positif. Nous montrons que la densité de sortie $\hat{f}_r$ associée à une densité d'entrée standard de valuation r, a de belles propriétés mathématiques. Elle s'exprime en fonction de la densité de Haar η, qu'on vient de définir en (2.6), mais aussi en fonction des séries d'Eisenstein dont on rappelle maintenant la définition. La série d'Eisenstein E_s, de poids s est définie par

$$E_s(x,y) := \frac{1}{2}\sum_{\substack{(c,d)\in\mathbb{Z}^2\\(c,d)\neq(0,0)}}\frac{y^s}{|cz+d|^{2s}}. \tag{2.7}$$

Théorème H. *On suppose que l'ensemble $\mathcal{B}\setminus\mathcal{F}$ des entrées de l'algorithme* Gauss-positif *est muni d'une densité d'entée f_r standard de valuation $r > -1$. Alors,*

(*i*) *Pour tout $r > -1$, la densité de sortie $\hat{f}_r$ s'exprime sous la forme*

$$\hat{f}_r = \frac{\pi}{3A(r)}\,F_{2+r}\cdot\eta, \tag{2.8}$$

Ici $A(r)$ est le coefficient de normalisation défini en (2.3), η est la densité de Haar définie en (2.6) et F_s est étroitement liée à la série d'Eisenstein analytique réelle E_s (définie en (2.7)) par la relation

$$F_s(x,y) = \sum_{\substack{(c,d)=1 \\ c\geq 1}} \frac{y^s}{|cz+d|^{2s}} = \frac{E_s(x,y)}{\zeta(2s)} - y^s \tag{2.9}$$

(*ii*) *Lorsque la valuation r de la densité d'entrée tend vers -1, alors la densité de sortie $\hat{f}_r$ converge ponctuellement vers la densité de Haar, qui est la densité des "réseaux aléatoires".*

Démonstration. Il s'agit d'appliquer le résultat général fourni par le théorème précédent 2.1,

$$\hat{f}_r(\hat{x},\hat{y}) = \sum_{h\in\mathcal{G}} |h'(\hat{z})|^2 f_r \circ h(\hat{z}) \tag{2.10}$$

au cas particulier envisagé. La preuve comporte deux étapes : nous utilisons d'abord la description explicite des homographies de $\mathcal{G}$, et nous cherchons ensuite à faire apparaître les séries d'Eisenstein et la mesure de Haar.

La proposition 1.2 de la partie II fournit une caractérisation de l'ensemble $\mathcal{G}$ des homographies utilisées par GAUSS-POSITIF : elle détermine l'ensemble $\mathcal{G}$ et montre qu'il est en bijection avec l'ensemble des couples (c,d) tels que $c \geq 1$ et $(c,d) = 1$. Elle fournit de plus une description effective de cette bijection : à chaque couple (c,d), on associe l'unique homographie de $\mathcal{G}$ vérifiant $h(\hat{z}) = (a\hat{z}+b)/(c\hat{z}+d)$ pour une paire (a,b) unique bien choisie. Une telle homographie vérifie

$$h'(\hat{z}) = \frac{1}{(c\hat{z}+d)^2} \qquad \text{et} \qquad \Im(h(\hat{z})) = \frac{\hat{y}}{|c\hat{z}+d|^2},$$

et donc aussi

$$|h'(\hat{z})|^2 = \frac{1}{|c\hat{z}+d|^4} \qquad \text{et} \qquad f_r \circ h(\hat{z}) = \frac{1}{A(r)} \frac{\hat{y}^r}{|c\hat{z}+d|^{2r}}.$$

L'égalité (2.10) se traduit alors en

$$\hat{f}_r(\hat{x},\hat{y}) = \frac{\pi}{3A(r)} \left(\sum_{\substack{(c,d)=1 \\ c\geq 1}} \frac{\hat{y}^{2+r}}{|cz+d|^{2(2+r)}} \right) \left(\frac{3}{\pi} \frac{1}{\hat{y}^2} \right).$$

En utilisant les notations de l'énoncé et la définition (2.6) de la mesure de Haar, on obtient donc la première relation (2.9).

Nous relions maintenant F_s à la série d'Eisenstein, en traitant d'abord le cas des couples pour lesquels $c = 0$. Observons que

$$F_s(x,y) := \sum_{\substack{(c,d)=1 \\ c\geq 1}} \frac{y^s}{|cz+d|^{2s}} = \frac{1}{2} \sum_{\substack{(c,d)=1 \\ c\neq 0}} \frac{y^s}{|cz+d|^{2s}},$$

satisfait

$$F_s(x,y) + y^s = \frac{1}{2} \left(\sum_{\substack{(c,d)=1 \\ c\neq 0}} \frac{y^s}{|cz+d|^{2s}} + 2y^s \right) = \frac{1}{2} \sum_{(c,d)=1} \frac{y^s}{|cz+d|^{2s}}.$$

La condition $(c,d)=1$ s'élimine aisément en remarquant que la fonction $(c,d)\mapsto|cz+d|^{-2s}$ est homogène de degré $-2s$ et en rajoutant un facteur $\zeta(2s)$. Il s'ensuit que

$$\zeta(2s)\cdot(F_s(x,y)+y^s)=\frac{1}{2}\sum_{\substack{(c,d)\in\mathbb{Z}^2\\(c,d)\neq(0,0)}}\frac{y^s}{|cz+d|^{2s}}=E_s(x,y),$$

comme on voulait le montrer.

Le développement en série de Fourier de $E_{2+r}(x,y)$ (voir par exemple [13]) permet de montrer que, en tout point (x,y),

$$E_{2+r}(x,y)\sim_{r\to-1}\frac{1}{2(r+1)}.$$

Comme $(r+1)A(r)$ tend vers 1 quand r tend vers -1, on obtient finalement, en tout point (x,y),

$$\lim_{r\to-1}\frac{\pi}{3A(r)}F_{2+r}(x,y)=\lim_{r\to-1}\frac{\pi}{3A(r)}\left(\frac{1}{2\zeta(2)(r+1)}-y\right)=\frac{1}{\pi},$$

qui est bien le résultat cherché. □

La convergence de $\hat{f}_r$ vers η n'est pas *uniforme* sur $\mathcal{F}$ car le terme constant dans le développement en série de Laurent de E_{2+r} autour de $r=-1$ est une fonction non-bornée de y (donnée par la première formule de Kronecker).

2.3 Distribution du premier minimum λ

Dans le chapitre 1, nous avons étudié la géométrie de l'ensemble de niveau $L(t)$ et vu que c'était une réunion de demi-disques, avec des réunions doubles ou triples entre demi-disques. Nous en avons fourni une description locale, dans les bandes de Farey, et caractérisé les cas où l'on a des réunions doubles ou triples entre demi-disques. Comme cette description locale est précise, elle va permettre maintenant l'estimation fine de la mesure de $L(t)$, dans le modèle de valuation r. Nous pourrons ainsi évaluer la fonction de répartition du paramètre λ. La preuve manipule des sommes de Riemann arithmétiques, et le théorème 1.2 permettra de d'évaluer leur comportement limite pour $t\to 0$.

2.3.1 Enoncé du résultat principal.

Théorème I. *Considérons l'algorithme* Gauss-positif *où l'ensemble des entrées est muni de la probabilité $\mathbb{P}_{(r)}$ de valuation r. Alors, la fonction de répartition du premier minimum λ vérifie, lorsque $t\to 0$ et $r>-1$ est fixe,*

$$\mathbb{P}_{(r)}[\lambda(z)\leq t]\sim_{t\to0}\quad A_1(r)\frac{\zeta(r+1)}{\zeta(r+2)}\cdot t^{r+2}\quad \textit{pour}\quad r>0\,,$$

$$\mathbb{P}_{(r)}[\lambda(z)\leq t]\sim_{t\to0}\quad \frac{A_1(0)}{\zeta(2)}\cdot t^2|\log t|\quad \textit{pour}\quad r=0\,,$$

$$\mathbb{P}_{(r)}[\lambda(z)\leq t]\sim_{t\to0}\quad \frac{A_4(r)}{\zeta(2)}\cdot t^{2r+2}\quad \textit{pour}\quad r<0\,.$$

De plus, pour toute valuation $r > -1$, l'inégalité suivante est vérifiée :

$$\mathbb{P}_{(r)}[\lambda(z) \leq t] \geq \frac{1}{A(r)} \frac{1}{r+1} \left(\frac{\sqrt{3}}{2}\right)^{r+1} t^{2r+2}. \tag{2.11}$$

Ici, les constantes $A(r)$ et $A_1(r)$ sont définies en (2.3) et (2.4) et $A_4(r) := I[D_r, \Delta]$ est définie par l'intégrale de la fonction D_r sur le domaine Δ, où D_r et Δ définis dans le lemme 2.1.

2.3.2 Preuve de l'encadrement

L'inclusion prouvée dans la proposition 1.1

$$[\lambda(z) \leq t] \supset \left[\Im(z) \leq \frac{\sqrt{3}}{2} t^2\right],$$

induit naturellement une inégalité sur les mesures, à savoir

$$\mathbb{P}_{(r)}[\lambda(z) \leq t] \geq \frac{1}{A(r)} \frac{1}{r+1} \left(\frac{\sqrt{3}}{2}\right)^{r+1} t^{2r+2},$$

ce qui est exactement (2.11).

2.3.3 Comportement quand $t \to 0$: cas d'une valuation $r \geq 0$.

La proposition 1.9 fournit l'encadrement suivant

$$\bigcup_{\frac{a}{c} \in \mathfrak{F}^{(2t)}} \mathsf{Fa}(a, c, t) \subset L(t) \subset \bigcup_{\frac{a}{c} \in \mathfrak{F}^{(\sqrt{3}t/2)}} \mathsf{Fa}(a, c, t)$$

où la réunion de gauche est disjointe. Elle induit naturellement un encadrement pour la mesure de $L(t)$, à savoir

$$\sum_{c \leq 1/t} \frac{\varphi(c)}{c^{r+2}} \leq \frac{\mathbb{P}_{(r)}[\lambda(z) \leq t]}{A_1(r) t^{r+2}} \leq \sum_{c \leq 2/(\sqrt{3}t)} \frac{\varphi(c)}{c^{r+2}}. \tag{2.12}$$

La proposition 2.1 fournit des équivalents pour les bornes de l'encadrement. On en conclut que

$$\mathbb{P}_{(r)}[\lambda(z) \leq t] \sim_{t \to 0} A_1(r) \frac{\zeta(r+1)}{\zeta(r+2)} \cdot t^{r+2} \qquad \text{pour} \quad r > 0\,,$$

et puis

$$\mathbb{P}_{(r)}[\lambda(z) \leq t] \sim_{t \to 0} A_1(0) \frac{1}{\zeta(2)} \log(1/t) \cdot t^{r+2} \qquad \text{pour} \quad r = 0\,,$$

comme voulu.

2.3.4 Comportement quand $t \to 0$: cas d'une valuation $r < 0$.

Dans le cas $r < 0$, il faut raisonner plus finement en utilisant les deux lemmes suivants. Nous commençons par calculer la mesure $\mathbb{P}_{(r)}[L_{c,d}(t)]$, puis nous remarquons que la mesure $\mathbb{P}_{(r)}[L(t)]$ s'exprime comme une somme de Riemann arithmétique.

Lemme 2.1. *Soit $(c,d) \in \mathfrak{D}^{(t)}$. Alors, la mesure $\mathbb{P}_{(r)}[L_{c,d}(t)]$ est égale à*

$$\mathbb{P}_{(r)}[L_{c,d}(t)] = t^{2r+4} D_r(ct, dt)$$

où la fonction D_r est définie à partir de la fonction β_r, elle-même définie dans le domaine Δ par une fonction beta incomplète,

$$\beta_r(x,y) = \frac{1}{2x^{r+2}} \frac{B([1-(x^2-y^2)]^2(2y)^{-2}; (r+3)/2, 1/2)}{B(1; (r+3)/2, 1/2)},$$

et des domaines $\Delta_2, \Delta_3^{(g)}, \Delta_3^{(c)}, \Delta_3^{(d)}$

$$\begin{aligned}
\Delta_2 &= \{(x,y) \in \Delta \ : \ x^2 + xy + y^2 \geq 1\} \\
\Delta_3 &= \Delta \setminus \Delta_2 \\
\Delta_3^{(g)} &= \{(x,y) \in \Delta_3 \ : \ (x+y)^2 - x^2 > 1\} \\
\Delta_3^{(c)} &= \{(x,y) \in \Delta_3 \ : \ (x+y)^2 - x^2 \leq 1 \text{ et } (x+y)^2 - y^2 \leq 1\} \\
\Delta_3^{(d)} &= \{(x,y) \in \Delta_3 \ : \ (x+y)^2 - y^2 > 1\},
\end{aligned}$$

comme suit :

$$D_r(x,y) = \begin{cases}
\beta_r(x,y) + \beta_r(y,x) & \text{si } (x,y) \in \Delta_2, \\
\beta_r(x,x+y) - \beta_r(x+y,x) + \beta_r(x+y,y) + \beta_r(y,x+y) & \text{si } (x,y) \in \Delta_3^{(g)}, \\
\beta_r(x,x+y) + \beta_r(x+y,x) + \beta_r(x+y,y) + \beta_r(y,x+y) & \text{si } (x,y) \in \Delta_3^{(c)}, \\
\beta_r(x,x+y) + \beta_r(x+y,x) - \beta_r(x+y,y) + \beta_r(y,x+y) & \text{si } (x,y) \in \Delta_3^{(d)},
\end{cases}$$

Démonstration. La proposition 2.2 a évalué la probabilité d'une portion $C_\rho(a)$ d'un demi-disque centré sur l'axe, de rayon ρ, délimitée par les verticales $x = 0$ et $x = \rho a$

$$\mathbb{P}_{(r)}[C_\rho(a)] = \frac{1}{2}\, \rho^{r+2} \frac{B(a^2; (r+3)/2, 1/2)}{B(1/4; (r+3)/2, 1/2)}.$$

Nous considérons plusieurs cas, correspondant aux différentes positions relatives du triplet formé par deux disques de Farey et le disques du médian, décrits par la proposition 1.8.

Commençons par le cas plus simple, où $L_{c,d}(t)$ est une réunion double. La proposition 1.8, montre que le point d'intersection des circonférences délimitant les quarts-disques $\mathsf{Fa}^+(a,c,t)$ et $\mathsf{Fa}^-(b,d,t)$ est

$$x_{c,d} = \frac{a}{c} + \frac{1+t^2(d^2-c^2)}{2cd} = \frac{b}{d} - \frac{1+t^2(c^2-d^2)}{2cd}.$$

Ainsi, la mesure de la réunion $L_{c,d}(t)$ est la somme des mesures des portions délimitées par les abcisses a/c et $x_{c,d}$ dans $\mathsf{Fa}^+(a,c,t)$, et par les abcisses $x_{c,d}$ et b/d dans $\mathsf{Fa}^-(b,d,t)$. Plus précisément,

$$\mathbb{P}_{(r)}[L_{c,d}(t)] = \mathbb{P}_{(r)}\left[C_{t/c}\left(\frac{1+t^2(d^2-c^2)}{2dt}\right)\right] + \mathbb{P}_{(r)}\left[C_{t/d}\left(\frac{1-t^2(d^2-c^2)}{2ct}\right)\right],$$

ce qui se réécrit en

$$\mathbb{P}_{(r)}[L_{c,d}(t)] = t^{2r+4}(\beta_r(ct,dt) + \beta_r(dt,ct)),$$

et qui établit le résultat dans le cas d'une réunion double.

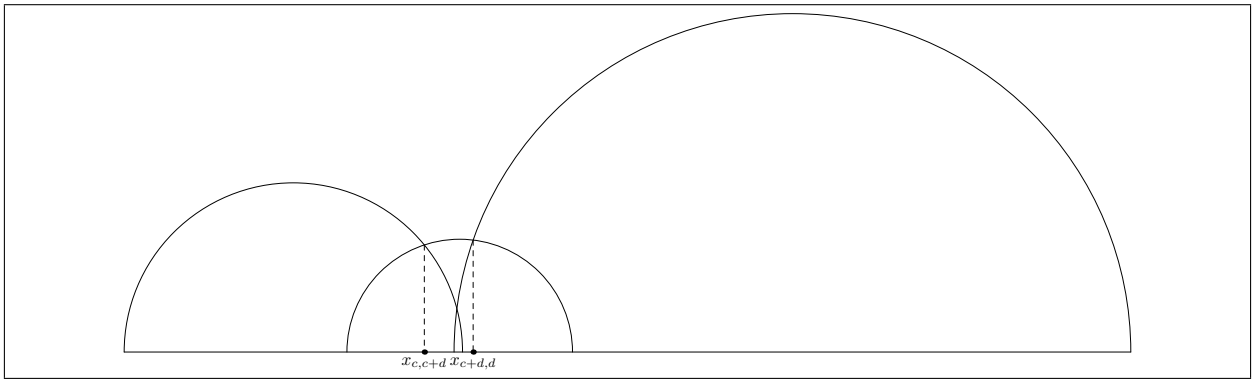

FIGURE 2.1 – Les disques de Farey centrés en a/c (à gauche) et en b/d (à droite), et le disque centré en $(a+b)/(c+d)$ (au milieu) : définition des abscisses $x_{c,c+d}$ et $x_{c+d,d}$.

Supposons à présent que $L_{c,d}(t)$ est une réunion triple. Comme l'a décrit la proposition 1.8, trois sous-cas se présentent, selon que les conditions

$$\mathsf{Fa}^-(a+b,c+d,t) \subset \mathsf{Fa}^+(a,c,t) \quad \text{et} \quad \mathsf{Fa}^+(a+b,c+d,t) \subset \mathsf{Fa}^-(b,d,t)$$

sont toutes les deux fausses, ou que l'une d'entre elles est vérifiée (si les deux étaient vérifiées, on serait dans le cas d'une intersection double). Si aucune de ces conditions n'est vérifiée, on peut se ramener au cas d'une réunion double. En effet, il suffit de répéter le calcul précédent pour $\mathsf{Fa}^+(a,c,t)$ et $\mathsf{Fa}^-(a+b,c+d,t)$, puis pour $\mathsf{Fa}^+(a+b,c+d,t)$ et $\mathsf{Fa}^-(b,d,t)$, pour obtenir

$$\mathbb{P}_{(r)}[L_{c,d}(t)] = t^{2r+4}(\beta_r(ct,(c+d)t) + \beta_r((c+d)t,ct) + \beta_r((c+d)t,dt) + \beta_r(dt,(c+d)t)).$$

Supposons maintenant que l'on soit dans le cas de l'inclusion $\mathsf{Fa}^-(a+b,c+d,t) \subset \mathsf{Fa}^+(a,c,t)$. Dans ce cas, on ne peut se ramener directement à une réunion double. On procède autrement, conformément à la figure 1.5, déjà faite au chapitre précédent, qu'on recopie ici (voir Figure 2.1). On considère les disques $\mathsf{Fa}^+(a,c,t)$ et $\mathsf{Fa}^-(a+b,c+d,t)$ et on calcule la mesure de la section déterminée par a/c et l'abcisse $x_{c,c+d}$, puis on *soustrait* la mesure de la portion du disque $\mathsf{Fa}^-(a+b,c+d,t)$ déterminée par $x_{c,c+d}$ et par son centre $(a+b)/(c+d) < x_{c,c+d}$. On conclut en rajoutant la mesure de la réunion de $\mathsf{Fa}^+(a+b,c+d,t)$ et $\mathsf{Fa}^-(b,d,t)$. On obtient

$$\mathbb{P}_{(r)}[L_{c,d}(t)] = t^{2r+4}(\beta_r(ct,(c+d)t) - \beta_r((c+d)t,ct) + \beta_r((c+d)t,dt) + \beta_r(dt,(c+d)t)).$$

Bien entendu, le cas restant est analogue au dernier. Les domaines $\Delta_2, \Delta_3^{(g)}, \Delta_3^{(c)}, \Delta_3^{(d)}$ sont déterminés par les conditions qui déterminent la nature de $L_{c,d}(t)$, et qu'on a vues dans la proposition 1.8. La preuve est ainsi achevée. □

Lemme 2.2. *Soit $r < 0$. Alors la fonction D_r est intégrable sur Δ et, quand t tend vers 0, la mesure $\mathbb{P}_{(r)}[L(t)]$ vérifie*

$$\mathbb{P}_{(r)}[L(t)] \sim t^{2r+2}\frac{I[D_r,\Delta]}{\zeta(2)}.$$

Démonstration. La mesure $\mathbb{P}_{(r)}[L(t])$ s'exprime comme une somme de Riemann arithmétique de la fonction D_r, au sens de la definition 1.6, sous la forme

$$\mathbb{P}_{(r)}[L(t)] = t^{2r+2}F[D_r,\Delta](t).$$

La fonction D_r est Riemann-intégrable, car elle est définie et continue en tout point de $\overline{\Delta}$ sauf en $(0,1)$ et en $(1,0)$. Mais, elle est intégrable dans Δ, comme nous le montrons maintenant : la symétrie de D_r permet de se restreindre à faire l'étude autour de $(0,1)$. et, lorsque $(x,y) \to (0,1)$, $D_r(x,y)$ est équivalente à C/x^{r+2}, qui est intégrable dans le voisinage $V_\epsilon = \{(x,y) \in \Delta \mid 0 \leq x \leq \epsilon\}$ pour ϵ assez petit, pour $r < 0$. En effet pour $r < 0$,

$$\iint_{V_\epsilon} \frac{1}{x^{r+2}} dydx = \int_0^\epsilon \int_{1-x}^1 \frac{1}{x^{r+2}} dydx = C\frac{\epsilon^{|r|}}{|r|}$$

Et donc, l'intégrale de D_r sur Δ est convergente pour $r < 0$ et D_r est intégrable sur Δ. De plus, il est facile de voir que $x \mapsto D_r(x,y)$ est décroissante sur V_ϵ si ϵ est suffisamment petit. Ainsi, le thèorème 1.2 s'applique, avec sa remarque, et la somme de Riemann arithmétique $F[D_r, \Delta](t)$ converge vers $(1/\zeta(2))I[D_r, \Delta]$. □

2.3.5 Commentaires.

Le régime de la fonction de répartition du paramètre λ change donc lorsque le signe de r change. Il y a deux parties dans le domaine $L(t)$: la partie inférieure, constituée de l'intersection complète du rectangle $[0,1] \times [0, (2/\sqrt{3})t^2]$ avec $\mathcal{B} \backslash \mathcal{F}$, et la partie supérieure, où $L(t)$ est beaucoup lacunaire. Quand la valuation r est négative, c'est la mesure de la partie inférieure qui est dominante, alors que, quand r est positive, c'est la partie supérieure qui est dominante. Il y a une transition de phase entre les deux régimes en $r = 0$; c'est ce qui arrive en particulier avec une densité uniforme.

2.4 Distribution du second minimum orthogonalisé μ

Dans le chapitre 1, nous étudié la géométrie de l'ensemble de niveau $M(u)$ et vu que c' était une réunion de triangles et quadrilatères, avec une proportion très majoritaire de triangles, au moins lorsque $u \to 0$. Nous en avons fourni une description locale, dans les bandes de Farey. Comme la description des conditions sous lesquelles on trouve un triangle ou un quadrilatère est précise, elle va permettre maintenant l'estimation fine de la mesure de $M(u)$, dans le modèle de valuation r. Nous pourrons ainsi évaluer la fonction de répartition du paramètre μ. Comme dans le cas de $L(t)$, la preuve manipule des sommes de Riemann arithmétiques arithmétiques, et le théorème 1.2 l permettra de d'évaluer leur comportement limite pour $u \to 0$.

2.4.1 Enoncé du résultat principal.

Théorème J. *Considérons l'algorithme* Gauss-positif *où l'ensemble des entrées est muni de la probabilité $\mathbb{P}_{(r)}$ de valuation r. Alors, la fonction de répartition du second minimum orhogonalisé μ vérifie, lorsque $u \to 0$ et $r > -1$ est fixe,*

$$\mathbb{P}_{(r)}[\mu(z) \leq u] \sim_{u \to 0} \frac{1}{\zeta(2)} A_5(r) \cdot u^{2r+2}.$$

Pour toute valuation $r > -1$, nous avons,

$$A_2(r)u^{2r+2} \leq \mathbb{P}_{(r)}[\mu(z) \leq u] \leq A_2(r) \left(\frac{2}{\sqrt{3}}\right)^{r+1} u^{2r+2},$$

et donc,

$$\mathbb{P}_{(r)}[\mu(z) \leq u] \sim_{r \to -1} A_2(r)u^{2r+2}.$$

Ici, la constante $A_2(r)$ est définie en (2.4) et $A_5(r) := I[T_r, \Delta]$ est l'intégrale de la fonction T_r sur le domaine Δ, où T_r et Δ sont définis dans le lemme 2.3.

2.4.2 Preuve de l'encadrement

L'encadrement de $\mathbb{P}_{(r)}[M(u)]$, est obtenu en appliquant la proposition 1.10. Il suffit de calculer l'aire des deux familles encadrantes. Comme la somme des bases des triangles vaut 1, et comme les hauteurs sont bornées inférieurement par u^2 pour la première famille et supérieurement par $2/\sqrt{3}u^2$ pour la deuxième famille, nous obtenons,

$$A_2(r)u^{2r+2} \leq \mathbb{P}_{(r)}[M(u)] \leq A_2(r)\left(\frac{2}{\sqrt{3}}\right)^{r+1} u^{2r+2},$$

qui est bien l'inégalité annoncée.

2.4.3 Comportement quand $u \to 0$.

Nous commençons par calculer la mesure $\mathbb{P}_{(r)}[M_{c,d}(u)]$, puis nous remarquons que la mesure $\mathbb{P}_{(r)}[M(u)]$ s'exprime comme une somme de Riemann arithmétique, à laquelle on peut appliquer le théorème 1.2.

Lemme 2.3. *Considérons un couple $(c,d) \in \mathfrak{D}^{(u)}$. Alors, La mesure $\mathbb{P}_{(r)}[M_{c,d}(u)]$ vérfie*

$$\mathbb{P}_{(r)}[M_{c,d}(u)] = u^{2r+4} T_r(cu, du)$$

où la fonction $T_r(x,y)$ est définie à partir de la fonction

$$\gamma_r(x,y) := \frac{A_2(r)}{xy\sin^{r+1}(\arcsin x + \arcsin y)}$$

et des ensembles $\Delta := \{(x,y) \ : \ 0 < x, y \leq 1, \ x + y > 1 \}$, $\Delta_T, \Delta_Q, \Delta_Q^{\pm}$

$$\begin{array}{rcl} \Delta_T & = & \{(x,y) \in \Delta; \quad xy \leq \frac{1}{2}\} \cup \{(x,y) \in \Delta; \quad x^2 - xy + y^2 \leq \frac{3}{4}\}, \\ \Delta_Q & = & \Delta \setminus \Delta_T \\ \Delta_Q^+ & = & \{(x,y) \in \Delta_Q; \quad x > y\} \\ \Delta_Q^- & = & \{(x,y) \in \Delta_Q; \quad x < y\} \end{array}$$

comme suit

$$T_r(x,y) = \left\{ \begin{array}{cc} \gamma_r(x,y) & si\ (x,y) \in \Delta_T, \\ \gamma_r(x-y,y) + \gamma_r(x-y,-x) & si\ (x,y) \in \Delta_Q^+, \\ \gamma_r(y-x,x) + \gamma_r(y-x,-y) & si\ (x,y) \in \Delta_Q^- \end{array} \right.$$

Démonstration. Nous avons calculé la mesure d'un triangle basé sur l'axe réel dans la proposition 2.2. La mesure de $M_{c,d}(u)$, lorsque $M_{c,d}(u)$ est un triangle vérifie

$$\frac{1}{A_2(r)}\mathbb{P}_{(r)}[M_{c,d}(u)] = (h_{c,d})^{r+1}(a/c - b/d) = u^{2r+2}\frac{1}{cd\sin^{r+1}(\theta_c + \theta_d)}.$$

La mesure de $M_{c,d}(u)$, lorsque $M_{c,d}(u)$ est un quadrilatère est la différence de la mesure de deux triangles, l'un d'entre eux étant le triangle couvrant défini dans la preuve de la proposition 1.10. Ainsi, si $c > d$, elle vérifie

$$\frac{1}{A_2(r)}\mathbb{P}_{(r)}[M_{c,d}(u)] = (h_{(c-d),d})^{r+1} \cdot \left(\frac{b}{d} - \frac{a-b}{c-d}\right) - (h_{(c-d),c})^{r+1} \cdot \left(\frac{a}{c} - \frac{a-b}{c-d}\right),$$

et donc, pour $c > d$,

$$\frac{1}{A_2(r)}\mathbb{P}_{(r)}[M_{c,d}(u)] = u^{2r+2}\frac{1}{(c-d)}\left(\frac{1}{d\sin^{r+1}(\theta_{c-d}+\theta_d)} - \frac{1}{c\sin^{r+1}(\theta_c - \theta_{c-d})}\right).$$

de manière analogue, pour $c < d$,

$$\frac{1}{A_2(r)}\mathbb{P}_{(r)}[M_{c,d}(u)] = u^{2r+2}\frac{1}{(d-c)}\left(\frac{1}{c\sin^{r+1}(\theta_{d-c}+\theta_c)} - \frac{1}{d\sin^{r+1}(\theta_d - \theta_{d-c})}\right).$$

Ceci prouve le lemme. □

Lemme 2.4. *La fonction T_r est intégrable sur Δ et, quand u tend vers 0, la mesure $\mathbb{P}_{(r)}[M(u)]$ satisfait*

$$\mathbb{P}_{(r)}[M(u)] \sim u^{2r+2}\frac{I[T_r,\Delta]}{\zeta(2)}.$$

Démonstration. La mesure $\mathbb{P}_{(r)}[M(u)]$ est la somme des termes $\mathbb{P}_{(r)}[M_{c,d}(u)]$ lorsque (c,d) décrit $\mathfrak{D}^{(u)}$. Par ailleurs, nous remarquons, d'après ce qui précède que chaque terme $\mathbb{P}_{(r)}[M_{c,d}(u)]$ est de la forme

$$\mathbb{P}_{(r)}[M_{c,d}(u)] = u^{2r+2}u^2T_r(cu,du)$$

On en déduit donc que $\mathbb{P}_{(r)}[M(u)]$ s'exprime donc comme une somme de Riemann arithmétique. En utilisant les notations du théorème 1.2, elle se met sous la forme

$$\mathbb{P}_{(r)}[M(u)] = u^{2r+2}F[T_r,\Delta](u).$$

On peut alors appliquer le théorème 1.2, dans les mêmes conditions que dans le lemme 2.2. Cela prouve alors le résultat. □

2.4.4 Commentaires.

A la différence de ce qui se passe pour le paramètre λ, la fonction de répartition du paramètre μ a toujours le même régime. En particulier, pour des valeurs négatives de la valuation r, les fonctions de répartition des paramètres λ et μ sont de même ordre.

2.5 Distribution du défaut d'Hermite γ

Comme dans le cas de $L(t)$ et de $M(u)$, l'ensemble de niveau $G(\rho)$ associé au défaut d'Hermite a une caractérisation géométrique, cette fois-ci via des disques de Ford (généralisés). Cela va permettre de calculer la mesure de l'ensemble $G(\rho)$, dans le modèle de valuation r, facilement dans le cas $\rho \leq 1$.

2.5.1 Enoncé du résultat principal.

Théorème K. *Considérons l'algorithme* Gauss-positif *où l'ensemble des entrées est muni de la probabilité $\mathbb{P}_{(r)}$ de valuation r. Alors, la fonction de répartition du défaut d'Hermite γ vérifie,*

$$\mathbb{P}_{(r)}[\gamma(z) \leq \rho] = A_3(r) \cdot \frac{\zeta(2r+3)}{\zeta(2r+4)} \cdot \rho^{r+2} \qquad \textit{pour} \quad \rho \leq 1,$$

et fait intervenir la constante $A_3(r)$ définie en (2.4).

Démonstration. Si $\rho \leq 1$, alors $G(\rho)$ est formé par des disques de Ford disjoints deux à deux, et la mesure de $G(\rho)$ est donc la somme des mesures des disques de Ford qui composent $G(\rho)$. Un ρ–disque de Ford a pour diamètre ρ/c^2, et sa mesure est donc, d'après (2.2),

$$\nu_r[\mathsf{Fo}(a,c,\rho)] = A_3(r)\frac{\rho^{r+2}}{c^{2r+4}}.$$

Ainsi

$$\mathbb{P}_{(r)}[\gamma(z) \leq \rho] = \sum_{(a,c)\in\mathcal{C}} \nu_r[\mathsf{Fo}(a,c,\rho)] = A_3(r)\rho^{r+2} \sum_{(a,c)\in\mathcal{C}} \frac{1}{c^{2r+4}}.$$

Comme, pour c fixé, le nombre de rationnels de la forme a/c avec $a \in [0,c]$ et a premier à c est égal à $\phi(c)$ où ϕ est la fonction d'Euler, on obtient l'égalité,

$$A_3(r)\rho^{r+2} \sum_{(a,c)\in\mathcal{C}} \frac{1}{c^{2r+4}} = A_3(r)\rho^{r+2} \sum_{c\geq 1} \frac{\varphi(c)}{c^{2r+4}} = A_3(r)\rho^{r+2}\frac{\zeta(2r+3)}{\zeta(2r+4)},$$

la deuxième égalité découlant de la proposition 2.1. on obtient donc bien le résultat annoncé pour $\rho \leq 1$.

□

2.5.2 Commentaires.

Est-il possible de décrire plus précisément la fonction de répartition du paramètre γ pour $\rho > 1$? La figure 1.3 montre que ce régime change quand $\rho = 1$. Ceci va être important pour obtenir une estimation précise de la valeur moyenne $\mathbb{E}_{(r)}[\gamma]$ comme fonction de r et pour comparer cette valeur aux expériences décrites dans la section 3.4 de la partie I.

2.5.3 Relation avec la densité de sortie. Les coins du domaine fondamental

Pour chaque $y_0 \geq 1$, l'évenement $[\Im\hat{z} \geq y_0]$ coincide avec l'évènement $[\gamma(z) \leq \frac{1}{y_0}]$, et donc, d'après ce qui précède,

$$\mathbb{P}_{(r)}[\hat{y} \geq y_0] = \mathbb{P}_{(r)}[\gamma(z) \leq \frac{1}{y_0}] = A_3(r)\frac{\zeta(2r+3)}{\zeta(2r+4)}\frac{1}{y_0^{r+2}}.$$

La fonction $y_0 \mapsto \mathbb{P}_{(r)}[\gamma(z) \leq \frac{1}{y_0}]$ définit une fonction π de la variable y_0, dont la dérivée est reliée clairement à la densité de sortie $\hat{f}_r$ du théorème **H**, par l'égalité

$$\pi_r'(y_0) := \int_{-1/2}^{+1/2} \hat{f}_r(x,y_0)dx.$$

Pour $r \to -1$, la fonction $\pi'_r(y)$ possède une limite qui est exactement la densité η, définie en (2.6), associée à la mesure de Haar de $SL_2(\mathbb{R})$ sur le demi-plan $\mathbb{H}$, définie dans la section 2.2.2 de ce chapitre.

Le théorème **K** permet aussi de calculer la probabilité qu'une base de sortie appartienne aux "coins" du domaine fondamental, et d'observer son évolution en fonction de la valuation r. C'est un premier pas vers la compréhension de la figure 3.4 (droite) (partie I).

Proposition 2.4. *Quand la densité d'entrée dans $\mathcal{B} \setminus \mathcal{F}$ est la densité standard de valuation r, la probabilité pour qu'une base de sortie appartienne aux coins $\{z \in \mathcal{F} \ : \ \Im z \leq 1\}$ est*

$$C(r) := 1 - A_3(r) \cdot \frac{\zeta(2r+3)}{\zeta(2r+4)}.$$

Il y a trois cas d'intérêt pour $1 - C(r)$:

$$[r \to -1] : \ \ 1 - \frac{3}{\pi} \approx 0.045 \qquad [r = 0] : \ \ 1 - \frac{3\pi}{2\pi + 3\sqrt{3}} \frac{\zeta(3)}{\zeta(4)} \approx 0.088 \qquad [r \to \infty] : \ \ 1 - \sqrt{\frac{\pi}{r}} e^{-3/2}.$$

On a par exemple : $C_{20} \approx 0.911, \quad C_{100} \approx 0.960.$

2.6 Conclusion du chapitre

Dans ce chapitre, nous avons effectué une analyse probabiliste de la géométrie de la base de sortie de l'algorithme. Nous avons commencé par étudier la densité de sortie, induite par une densité d'entrée générale, puis nous avons concentré notre étude sur la densité standard de valuation r. Dans ce cas, la densité de sortie $\hat{f}_r$ associée à la densité standard de valuation r est fortement liée aux séries d'Eisenstein et à la densité de Haar de $SL_2(\mathbb{R})$. En particulier, lorsque $r \to -1$, la densité de sortie $\hat{f}_r$ converge ponctuellement vers cette densité invariante, qui définit une probabilité naturelle sur les réseaux, comme on l' a vu en 3.2.2 (partie I).

Dans ce chapitre, nous avons aussi déterminé les fonctions de répartition des principales variables λ, μ, γ qui permettent de décrire la géométrie de la base de sortie. Nous avons fourni des estimations précises des distributions de λ et μ, en particulier pour $r \to -1$. Dans le cas du défaut d'Hermite γ, il y a aussi une formule exacte pour la distribution, mais pas sur toute la portée du paramètre.

Nous avons déjà expliqué en quoi les informations sur la distribution de la variable γ peuvent apporter des premiers éléments de réponse aux interrogations suscitées par les expérimentations présentées en 3.4 (partie I). Dans la prochaine partie IV, nous expliquerons le rôle que peut jouer la distribution des variables λ et μ pour débuter l'analyse de l'algorithme LLL-Impair-Pair, présenté dans la section 2.3.6 (partie I).

Quatrième partie

Conclusions.

Chapitre 1

Retour à l'analyse de l'algorithme LLL

Sommaire

Dans ce chapitre nous voulons donner quelques directions futures pour l'analyse de l'algorithme LLL, qui s'inspirent des résultats présentés dans cette thèse.

1.1 L'algorithme Pair-Impair

L'algorithme LLL cherche à réduire les bases locales U_k (cf. paragraphe 2.3.3, page 42) dans le sens de Gauss. Pour obtenir la densité de sortie à la fin de l'algorithme, il est intéressant de décrire l'évolution de la distribution des bases locales tout au long de l'exécution de l'algorithme. La variante LLL-Impair-Pair décrite dans la section 2.3.6 de la partie I est bien adaptée à ce propos. Nous la recopions ici

1.1.1 Le rapport de Siegel au début de la deuxième phase.

Dans la première phase, l'algorithme LLL-Impair-Pair traite les bases dont l'indice est impair. Considérons deux bases successives U_k et U_{k+2} respectivement munies des densités d'entrée F_k et F_{k+2}. Notons z_k et z_{k+2} les nombres complexes associés aux bases locales (u_k, v_k) et (u_{k+2}, v_{k+2}) via la relations $z_i = v_i/u_i$, $i \in \{k, k+2\}$. Alors, l'algorithme LLL-Impair-Pair réduit ces deux bases locales (dans le sens de Gauss) et calcule deux bases locales réduites notées $(\hat{u}_k, \hat{v}_k)$ et $(\hat{u}_{k+2}, \hat{v}_{k+2})$, que satisfont en particulier

$$|\hat{v}_k^\star| = |\hat{u}_k| \cdot \mu(z_k), \qquad |\hat{u}_{k+2}| = |u_{k+2}| \cdot \lambda(z_{k+2}).$$

Alors, les théorèmes **J**, et **I** fournissent des pistes sur la distribution de $\mu(z_k), \lambda(z_{k+2})$. Comme dans notre modèle les variables aléatoires $|u_k|$ et z_k (resp. $|u_{k+2}|$ et z_{k+2}) sont indépendantes (voir

LLL Pair–Impair (t) $\quad [t > 1]$

Input. Une base B d'un réseau $\mathcal{L}$ de dimension p.
Output. Une base réduite $\hat{B}$ de $\mathcal{L}$.

`Gram` : calculer la base orthogonale $B^\star$ et la matrice $\mathcal{P}$.
`Tant que` B n'est pas réduite `faire`
  `Phase Impaire` (B) :
    `Pour` $i = 1$ à $\lfloor n/2 \rfloor$ `faire`
      `Réduction-de-taille-principale` (b_{2i}) ;
      $\mathcal{M}_i := t$–GAUSS-AIGU (U_{2i-1}) ;
      $(b_{2i-1}, b_{2i}) := (b_{2i-1}, b_{2i})^t \mathcal{M}_i$;
    `Pour` $i = 1$ à n `faire` `Réduction-de-taille-secondaire` (b_i) ;
    `Recalculer` $B^\star, \mathcal{P}$;
  `Phase Paire` (B) :
    `Pour` $i = 1$ à $\lfloor (n-1)/2 \rfloor$ `faire`
      `Réduction-de-taille-principale` (b_{2i+1}) ;
      $\mathcal{M}_i := t$–GAUSS-AIGU (U_{2i}) ;
      $(b_{2i}, b_{2i+1}) := (b_{2i}, b_{2i+1})^t \mathcal{M}_i$;
    `Pour` $i = 1$ à n `faire` `Réduction-de-taille-secondaire` (b_i) ;
    `Recalculer` $B^\star, \mathcal{P}$;

FIGURE 1.1 – La variante PAIR-IMPAIR de l'algorithme LLL.

section 1.3.1, partie II), nous obtenons une information précise sur la distribution des normes $|\hat{v}_k^\star|, |\hat{u}_{k+2}|$. Dans la seconde phase, l'algorithme considère les bases locales avec un indice pair. Or, la base U_{k+1} est formée (à une similitude près) à partir des deux bases de sortie précédantes, de la manière suivante :

$$u_{k+1} = |\hat{v}_k^\star|, \qquad v_{k+1} = \nu |\hat{v}_k^\star| + i|\hat{u}_{k+2}|,$$

où ν suit une loi qui peut être supposée uniforme dans l'intervalle $[-1/2, 1/2]$. En plus, au moins au début de l'algorithme, les deux variables $|\hat{v}_k^\star|, |\hat{u}_{k+2}|$ sont indépendantes. Tout ceci nous permet d'obtenir des informations précises sur la nouvelle densité d'entrée F_{k+1} dans la base locale U_{k+1}.

1.1.2 La suite de l'évolution du rapport de Siegel.

Nous aimerions pouvoir "suivre" l'évolution des densités des bases locales tout au long de l'exécution de l'algorithme LLL-IMPAIR-PAIR. Beaucoup de questions se posent à ce sujet

Cette approche est-elle suffisamment robuste pour pouvoir s'appliquer à toute l'exécution de l'algorithme LLL-IMPAIR-PAIR ? Bien entendu, au milieu de l'algorithme, les deux variables $\hat{v}_k^\star, \hat{u}_{k+2}$ ne sont plus indépendantes. Sont-elles "très" dépendantes ? Peut-on réutiliser l'argument pour les itérations suivantes ?

Est-ce vrai que les variables ν au *début* de la phase ont une distribution proche de la distribution uniforme sur $[-1/2, 1/2]$? Les résultats expérimentaux de Nguyen et Stehlé [60] montrent que ce n'est pas le cas, de même que les expérimentations (communication personnelle) de Lhote, mais que cette distribution est sans doute quasi-uniforme....

1.2 Modélisation par des tas de sable.

Dans cette section, nous présentons un modèle simplifié pour l'algorithme LLL, proposé très récemment par Madritsch et Vallée [51].

1.2.1 Algorithme LLL avec version de Siegel.

L'évolution de l'algorithme LLL, appliqué sur une base $B = (b_1, \ldots, b_n)$ est principalement décrit par les rapports de Siegel $r_i = \ell_{i+1}/\ell_i$, pour $i \in \{1, \ldots, n-1\}$, où $\ell_i = |b_i^*|$ est la norme de l'i-ème orthogonalisé de la base orthonormée de B.

Considérons, comme dans le chapitre 2 de la partie I, deux réels $t > 1$ et s liés par la relation

$$\frac{1}{t^2} = \frac{1}{s^2} + \frac{1}{4}, \qquad \text{de sorte que} \quad s > \frac{2}{\sqrt{3}}.$$

Nous avons montré (chapitre 2 de la partie I) que la condition de s-Siegel

$$r_i \geq \frac{1}{s}, \tag{1.1}$$

entraînait alors la condition de t-Lovász

$$\ell_{i+1}^2 + m_{i+1}^2 \ell_i^2 \geq \frac{1}{t^2} \ell_i^2.$$

Par ailleurs, les coefficients $m_{i+1,i}$ jouent un rôle secondaire par rapport aus longueurs de Siegel ℓ_i et au rapport de Siegel r_i. Il est donc usuel de considérer la variante dite de Siegel de l'algorithme LLL, où les opérations sont les mêmes que dans l'algorithme usuel, mais où le test de sortie, celui de Siegel, est légèrement plus faible que le test usuel, celui de Lovász. La base de sortie sera de qualité un peu moindre. Remarquons cependant que les propriétés qu'on sait prouver sur une base réduite au sens de Siegel sont les mêmes que celles qu'on sait prouver sur la base de sortie de l'algorithme LLL usuel.

L'algorithme est décrit dans la figure 1.2.

LLL (s) $\quad [s > 2/(\sqrt{3})]$

Entrée. Une base B d'un réseau L de dimension n.
Sortie. Une base $\hat{B}$ de L, s-Siegel réduite

Calculer la base $B^\star$ et la matrice $\mathcal{P}$.
$i := 1$;
`Tant que` $i < n$ `faire`
  1– `Translation` $(i+1, i)$;
  2– **Si** $\ell_{i+1} \geq (1/s)\,\ell_i$, **alors** $i := i + 1$
    sinon `Échange` $(i+1, i)$
      `Récalculer` $(B^\star, \mathcal{P})$;
      $i := \max(i-1, 1)$;
`Translations finales`.

FIGURE 1.2 – Description de l'algorithme LLL avec des conditions de Siegel.

On rappelle que l'exécution de l'algorithme LLL effectue des translations et des échanges, de façon à assurer que tous les rapports de Siegel r_i soient minorés par $1/s$. Alors que les

translations ne changent pas la longueur des orthogonalisés ℓ_i, les échanges modifient la longueur de ces orthogonalisés. Après un échange entre les vecteurs b_i et b_{i+1}, quand la condition de Siegel n'est pas vérifiée, les nouvelles valeurs $\check{\ell}_i$ et $\check{\ell}_{i+1}$ des orthogonalisés vérifient

$$\check{\ell}_i^2 := \ell_{i+1}^2 + m_{i+1}^2 \ell_i^2, \quad \text{de sorte que} \quad \check{\ell}_i = \rho \ell_i \quad \text{avec} \quad \rho^2 = \frac{\ell_{i+1}^2}{\ell_i^2} + m_{i+1,i}^2, \tag{1.2}$$

alors que l'invariance du déterminant implique la relation $\check{\ell}_i \check{\ell}_{i+1} = \ell_i \ell_{i+1}$, et donc l'égalité $\check{\ell}_{i+1} = (1/\rho)\ell_{i+1}$. Par ailleurs, la condition de propreté $|m_{i+1,i}| \leq 1/2$, et le fait que la condition de Siegel n'était pas vérifiée auparavant impliquent que

$$\frac{\ell_{i+1}}{\ell_i} < \frac{1}{s} \leq \frac{\sqrt{3}}{2} \text{et donc} \quad \rho \leq \rho_0(s) \quad \text{avec} \quad \rho_0(s) = \frac{1}{s^2} + \frac{1}{4} < 1. \tag{1.3}$$

Pour mieux voir ce qui se passe, on adopte un point de vue additif, en posant

$$q_i := \log_s \ell_i.$$

La condition de Siegel devient alors $q_i \leq q_{i+1} + 1$, et l'échange dans l'algorithme LLL se récrit comme

$$\textbf{If} \quad q_i > q_{i+1} + 1, \quad \textbf{then} \quad [\check{q}_i = q_i + \log_s \rho, \quad \check{q}_{i+1} = q_{i+1} - \log_s \rho].$$

1.2.2 Hypothèse simplificatrice de régularité.

La simplification principale consiste à supposer que l'exécution de l'algorithme est régulière et que ρ est constant tout au long de l'exécution de l'algoritthme. On pose alors

$$\alpha := -\log_s \rho,$$

de sorte que $\alpha \geq -\log_s \rho_0(s) > 0$. L'algorithme LLL devient alors un modèle de tas de sable,

$$\textbf{If} \quad q_i > q_{i+1} + 1, \quad \textbf{then} \quad [\check{q}_i = q_i - \alpha, \quad \check{q}_{i+1} = q_{i+1} + \alpha],$$

ou, si on travaille avec les rapports de Siegel et qu'on pose

$$c_i := -\log_s r_i = q_{i+1} - q_i,$$

il devient un jeu de tir [7]

$$\textbf{If} \quad c_i > 1, \quad \textbf{then} \quad [\check{c}_i = c_i - 2\alpha, \quad \check{c}_{i+1} = c_{i+1} + \alpha].$$

Les versions régularisées de LLL, en version "multiplicative" et "additive" (pour le tas de sable) sont décrites dans la figure 1.3.

1.2.3 Arguments en faveur de l'hypothèse de régularité

Evidemment, les exécutions de l'algorithme LLL ne peuvent pas être exactement régulières, et l'hypothèse d'une constante α universelle qui apparaîtrait dans toutes les exécutions de l'algorithme sur tous les modèles d'entrées possible est bien trop forte. Ici, nous cherchons à répondre aux questions suivantes :

7. chip-firing game en anglais.

RLLL (ρ, s)
with $s > 2/\sqrt{3}$, $\rho \leq \rho_0(s) < 1$

Entrée. Une suite $(\ell_1, \ell_2, \dots \ell_n)$
Sortie. Une suite $(\hat{\ell}_1, \hat{\ell}_2, \dots \hat{\ell}_n)$
avec $\hat{\ell}_{i+1} \geq (1/s)\hat{\ell}_i$.

$i := 1$;
`Tant que` $i < n$ `faire`
Si $\ell_{i+1} \geq (1/s)\ell_i$, **alors** $i := i+1$
sinon $\ell_i := \rho\, \ell_i$;
$\ell_{i+1} := (1/\rho)\, \ell_{i+1}$;
$i := \max(i-1, 1)$;

ARLLL (α) avec $\alpha > \alpha_0(s)$.

Entrée. Une suite $(q_1, q_2, \dots q_n)$
Output. Une suite $(\hat{q}_1, \hat{q}_2, \dots \hat{q}_n)$
with $\hat{q}_i - \hat{q}_{i+1} \leq 1$.

$i := 1$;
`Tant que` $i < n$ `faire`
Si $\hat{q}_i - \hat{q}_{i+1} \leq 1$, **alors** $i := i+1$
sinon $q_i := q_i - \alpha$;
$q_{i+1} := q_{i+1} + \alpha$;
$i := \max(i-1, 1)$;

FIGURE 1.3 – Les versions régularisées de l'algorithme LLL. À gauche, la version multiplicative, qui dépend des paramètres s, ρ, avec $\rho_0(s)$ défini en (1.3). À droite, la version additive, qui dépend du paramètre $\alpha := \log_s \rho$, avec $\alpha_0 := -\log_s \rho_0(s)$

(*i*) Pourquoi est-ce est plausible de considérer que α est fixe durant une exécution de l'algorithme ?

(*ii*) La valeur de α dépend-elle du modèle probabiliste d'entrée ? Si oui, comment ?

Nous donnons ici un certain nombre d'arguments avancés par les auteurs.

Si les coefficients m_{i+1}, i sont presque uniformes. Tout d'abord, dans l'expression de ρ donnée dans (1.2), les coefficients $m_{i+1,i}$ jouent un rôle très secondaire par rapport aus longueurs de Siegel ℓ_i et au rapport de Siegel r_i. C'est pour cela que Madritsch et Vallée supposent que ces coefficients $m_{i+1,i}$ distribués uniformément en $[-1/2, 1/2]$, et indépendants des rayons de Siegel r_i. Dans un tel cas, la valeur moyenne de $m_{i+1,i}^2$ est $1/12$. On peut donc fixer $m_{i+1,i}^2$ à $1/12$ en tant qu'hypothèse simplificatrice. Si on choisit pour s la valeur maximale $s = s_0 = 2/\sqrt{3}$ (correspodant à $t = 1$), alors le paramètre α vérifie

$$-\frac{1}{2}\log_{s_0}\left(\frac{3}{4}+\frac{1}{12}\right) \leq \alpha := -\frac{1}{2}\log_{s_0}\left(r^2+\frac{1}{12}\right) \leq -\frac{1}{2}\log_{s_0}\left(\frac{1}{12}\right),$$

et α varie dans l'intervalle $[0.63; 8.64]$.

Le cas de la dimension 2 et la modèle à valuation. Dans le cas de la dimension 2 et de la distribution avec valuation θ (il y a un conflit de notation avec le rapport de Siegel, et la valuation, notée préceédemment r, est maintenant notée θ), nous avons montré dans le théorème 3.4 que le nombre d'itérations P de l'algorithme de Gauss suit asymptotiquement une loi géométrique de rapport $\lambda(2+\theta)$, où $\lambda(s)$ est la valeur propre dominante de l'opérateur $\underline{\mathbf{H}}_s$. On a donc

$$-\log_s \mathbb{P}[K \geq k] \sim k \log \lambda(2+\theta)$$

Par ailleurs, en dimension 2, et dans le vocabulaire complexe, le rapport de Siegel n'est autre que la partie imaginaire $y = \Im z$, et dans le modèle à valuation $\theta > -1$, on a par définition

$$\mathbb{P}[y \leq x] = x^{\theta+1} \qquad \text{pour } x \in [0,1]$$

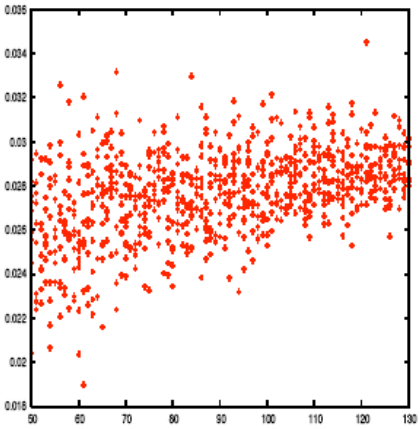

FIGURE 1.4 – La distribution du paramètre $(1/n)\log_2\gamma(\hat{B})$ en fonction de la dimension n. Les valeurs expérimentales de $(1/n)\log_2\gamma(\hat{B})$ paraissent appartenir à un petit intervalle centré en 0.03.

Ce qui, dans le jeu de tir s'écrit additivement en

$$\mathbb{P}\left[c \geq kH\right] = \exp[-(\theta+1)kH]$$

Donc, si α est supposé constant, on a

$$-\log_s \mathbb{P}[P \geq k] \sim -\log_s \mathbb{P}[c \geq k\mathbb{E}[\alpha]] \sim \mathbb{E}[\alpha](\theta+1)k$$

et donc, la valeur moyenne de α peut se choisir comme

$$\mathbb{E}[\alpha] \sim \frac{-1}{1+\theta}\log_s(\lambda(2+\theta)).$$

On utilise alors deux propriétés importantes de la valeur propre dominante,

$$\frac{-1}{1+\theta}\log\lambda(2+\theta) \to |\lambda'(1)| \quad (\theta \to -1), \qquad \frac{-1}{1+\theta}\log\lambda(2+\theta) \to 2\log(1+\sqrt{2}) \quad (\theta \to \infty),$$

où $|\lambda'(1)| \approx 3,41$ est l'entropie de l'algorithme d'Euclide. Sous l'hypothèse de régularité, on en déduit que la valeur moyenne de α vérifie

$$\mathbb{E}[\alpha] \sim \frac{|\lambda'(1)|}{\log s} \quad (\theta \to -1), \qquad \mathbb{E}[\alpha] \sim 2\frac{\log(1+\sqrt{2})}{\log s} \quad (\theta \to \infty).$$

Cela entraîne que, en dimension 2, et pour $s = s_0$ (valeur maximale de s), la valeur moyenne $\mathbb{E}[\alpha]$ varie dans l'intervalle $[14, 23]$ selon la valuation du modèle dans lequel on se place.

Un argument supplémentaire en dimension 2. Dans ce cas, on peut travailler avec deux éléments de $\mathbb{C}$ (voir chapitre 1 de la partie II), et supposer que $b_1 := \ell_1$ et $b_2 := m_{2,1}\ell_1 + i\ell_2$. Dans ce cas, le rapport ρ défini en (1.2) est exactement $\rho = |b_2/b_1|^2$, et dans le cas de la distribution d'Ajtai de paramètre θ, le paramètre α suit approximativement une loi exponentielle, de la forme $\mathbb{P}[\alpha \leq u] = e^{-u(\theta+1)}$, qui prouve que α est assez concentré.

1.2.4 Résultats dans le modèle simplifié.

Les résultats dans le modèle simplifié décrivent ce qui se passe pour des exécutions de l'algorithme LLL pendant lesquelles le facteur de décroissance est constant et égal à ρ. Des telles exécutions sont appellées ρ–régulières.

Nous rappelons que, dans ce cas, l'exécution de l'algorithme en dimension n peut être vu comme un modèle de tas de sable $\mathcal{Q}_n(\mathbf{q}, 1, \alpha)$ avec paramètre $\alpha := -\log_s \rho$, et configuration initiale $\mathbf{q} := (\log_s \ell_1, \ldots, \log_s \ell_n)$ dont les composants $q_i = \log_s \ell_i$ sont reliés aux longueurs ℓ_i de la base orthogonalisée $B^\star$ de la base d'entrée B. Les objets principaux du modèle de tas de sable, à savoir l'énergie $E(\mathbf{q})$ ou la masse totale $M(\mathbf{q})$ sont liés de façon proche aux principales caractéristiques de la base d'entrée, à savoir le potentiel $D(B)$ ou le déterminant $\det(B)$, puisque l'on a

$$E(\mathbf{q}) = \log_s D(B), \qquad M(\mathbf{q}) = \log_s \det(B).$$

Dans le modèle simplifié, Madritsch et Vallée étudient le nombre d'itérations pour des bases totalement-non-réduites, pour lesquelles la i-ème condition de s-Siegel (1.1) n'est satisfaite en entrée pour aucun i. Dans ce cas, le jeu de tir satisfait $c_i > 1$, et le tas de sable est strictement croissant.

Ils obtiennent deux principaux résultats dans ce modèle, l'un sur le nombre d'itérations de l'algorithme, l'autre sur un paramètre de la géométrie de sortie.

Nombre d'itérations. Les auteurs analysent le cas où la base d'entrée est distribuée selon une distribution de valuation θ.

Théorème 1.1. *Considérons une base d'entrée B, qui suit une distribution d'Ajtai de paramètre θ. Si l'exécution de l'algorithme LLL en dimension n est ρ–régulière sur la base B, alors le nombre d'itérations K_n de l'algorithme LLL sur la base B satisfait*

$$K_n(\rho, \theta) \sim \frac{n^3}{12} \left(\frac{\rho^{\theta+1}}{1 - \rho^{\theta+1}} \right). \tag{1.4}$$

L'équivalent (1.4) peut se comparer à la borne supérieure que Daudé et Vallée ont obtenue dans le modèle sphérique. comme nous l'avons rappelé dans le théorème 3.3 de la partie I. Cette borne supérieure est en en $O(n^2 \log n)$, et l'équivalent (1.4) donne un nombre d'itérations asymptotiquement plus grand que la borne de Daudé et Vallée. Cela est raisonnable, puisque le modèle sphérique donne lieu à des modèles avec des valuations très grandes, dès qu'on s'éloigne des bases de la fin (voir chapitre 3 de la partie I). Si, dans le modèle d'Ajtai, on choisit une valuation θ de la forme $\theta = -1 + n^{-a}$, avec a positif, comme l'a fait historiquement Ajtai, le nombre d'itérations K_n est en $\Theta(n^{3+a})$, ce qui est en accord avec les expérimentations de Nguyen et Stehlé [60].

Géométrie de la sortie. A la fin de l'algorithme, par définition, tous les rapports de Siegel $\hat{r}_i$ ont une valeur minorée par $1/s$. Mais, les auteurs se posent aussi la question que nous nous étions déjà posée à la fin du chapitre 3 de la partie I : Que peut-on dire de la valeur moyenne de ces rapports de Siegel $\hat{r}_i$? Est-elle proche de $1/s$? Ils étudient en particulier le paramètre γ dont nous avions parlé dans le chapitre 3 de la partie I, défini par

$$\gamma(\hat{B}) := \frac{||\hat{b}_1||}{(\det L)^{1/n}} = \left[\prod_{i=1}^{n} \left(\frac{1}{\hat{r}_i} \right)^{n-i} \right]^{1/n}.$$

La distribution du paramètre γ dépend-elle de la distribution d'entrée, et en particulier de la valuation de la distribution d'entrée ? Nous savons que la distribution du défaut d'Hermite dépend de la valuation d'entrée en dimension 2. Mais est-ce généralement le cas ? Nous avons aussi évoqué cette question dans le chapitre 3 de la partie I.

Dans le cas d'une exécution ρ–régulière, Madritsch et Vallée montrent que, pour chaque indice i pour lequel l'entrée ne satisfait pas la condition $\mathcal{S}_s(i)$, le rapport de Siegel $\hat{r}_i$ de sortie satisfait

$$\rho s \leq \frac{1}{\hat{r}_i} = \frac{\hat{\ell}_i}{\hat{\ell}_{i+1}} \leq s. \tag{1.5}$$

Quand la base d'entrée est donc totalement non-réduite, le premier vecteur $\hat{b}_1$ de la base de sortie $\hat{B}$ vérifie donc, avec (1.5),

$$\rho(s \cdot \rho)^{(n-1)/2} \leq \gamma(\hat{B}) := \frac{||\hat{b}_1||}{(\det L)^{1/n}} \leq s^{(n-1)/2}.$$

Les auteurs ont donc prouvé :

Théorème 1.2. *Considérons une base totalement non réduite sur laquelle l'exécution de l'algorithme LLL–Siegel (pour le paramètre s) est ρ-régulière. Alors, le paramètre $\gamma(\hat{B})$ défini dans l'équation (2.19) de la partie I satisfait*

$$\frac{2}{n-1} \log \gamma(\hat{B}) \in [\log s + \log \rho,\ \log s]. \tag{1.6}$$

Nous voyons que ce résultat est, lui, et contrairement à celui qui précède, indépendant de la distribution d'entrée. Il dépend seulement du degré de régularité de l'exécution. Ce résultat est compatible avec les expérimentations faites par Nguyen et Stehlé [60]. La figure 1.4, dont nous avions déjà parlé au chapitre 3 de la partie I, montre qu'il y a une valeur moyenne $\beta \sim 1.04$, telle que, pour la plupart des bases de sortie $\hat{B}$, le rapport $\gamma(\hat{B})$ est proche de $\beta^{(n-1)/2}$. La relation $\beta \sim s\sqrt{\rho}$ est donc plausible, ce qui montrerait (indirectement) que la valeur ρ "usuelle" serait proche de 0.81.

Conclusion

Nous résumons maintenant les contributions de cette thèse.

Tout d'abord, nous avons introduit un modèle réaliste, celui d'une densité avec valuation, qui permet de donner un cadre unificateur qui rassemble des instances de difficulté variable par rapport au problème de la réduction. Dans ce modèle, nous avons mené une étude probabiliste complète et précise de la *complexité* et de la *qualité de sortie* de l'algorithme de Gauss. Nous avons décrit la transition entre l'algorithme de Gauss et l'algorithme d'Euclide. Dans notre étude de la complexité, nous avons repris le cadre de l'analyse dynamique, que nous avons adapté à l'étude de ce modèle à valuation.

Dans le cas de l'algorithme de Gauss, nous montrons que la distribution de toute une classe de coûts naturels, dits additifs, est asymptotiquement géométrique. Ce résultat généralise le résultat de Daudé, Flajolet, et Vallée, qui avaient démontré ce résultat, seulement pour le nombre d'itérations, et dans le cas d'une densité uniforme. Nous prouvons que la raison de cette loi géométrique est reliée aux propriétés spectrales de l'opérateur de transfert associé. Ce résultat exhibe une différence de comportement très importante entre les deux algorithmes (Gauss et Euclide), puisque les mêmes coûts ont une distribution asymptotiquement gaussienne, dans le cas de l'algorithme d'Euclide. Nous étudions aussi la complexité binaire des deux algorithmes. Elle est linéaire dans le cas de l'algorithme de Gauss, et quadratique dans le cas de l'algorithme d'Euclide. Lorsqu'on biaise la distribution d'entrée de l'algorithme de Gauss en donnant plus de poids aux entrées colinéaires, on s'approche alors de l'ordre quadratique de la complexité en bits de l'algorithme d'Euclide.

L'étude de la géométrie de sortie a repris les travaux de Laville et Vallée [45], qui avaient déjà étudié deux paramètres, premier minimum et défaut d'Hermite, mais uniquement dans le cas d'une densite uniforme (cas de la valuation nulle). Nous leur avons donné à la fois un cadre unificateur et plus général. Nous avons introduit un paramètre supplémentaire, le deuxième minimum orthogonalisé, que Laville et Vallée n'avaient pas envisagé, et qui peut jouer un rôle très important dans l'analyse de l'algorithme LLL. Nous avons conduit cette analyse dans le cadre d'une valuation quelconque.

Ces résultats permettent d'élaborer une première stratégie pour l'analyse de l'algorithme LLL. Dans la variante LLL-Impair-Pair, la sortie d'une phase de l'algorithme est l'entrée de phase suivante, nous pouvons "réinjecter" à chaque phase les résultats de notre analyse dans chaque base locale. Nous avons ainsi un premier accès à l'évolution de la distribution sur les bases locales. Cette approche peut être assemblée avec l'analyse par modèle de tas de sable proposée très récemment par Madritsch et Vallée [51].

Pour l'instant, néanmoins, ces approches en sont à leur tout début. Une analyse de l'algorithme LLL présente certainement de très grosses difficultés techniques. Cette thèse vise juste à apporter une pierre à l'édifice qu'il faut construire pour espérer conduire l'analyse de LLL à son terme.

Bibliographie

[1] M. Ajtai : Generating hard instances of lattice problems (extended abstract). *In Proceedings of STOC'96*, p. 99–108. ACM Press, 1996.

[2] M. Ajtai : The shortest vector problem in L_2 is NP-hard for randomized reductions (extended abstract). *In Proceedings of STOC'98*, p. 10–19. ACM Press, 1998.

[3] A. Akhavi : *Analyse comparative d'algorithmes de réduction dans les réseaux aléatoires.* Thèse de doctorat, Université de Caen, 1999.

[4] A. Akhavi : Random lattices, threshold phenomena and efficient reduction algorithms. *Theoretical Computer Science*, 287(2):359–385, 2002.

[5] A. Akhavi, J.-F. Marckert et A. Rouault : On the reduction of a random basis. *In Proceedings of SIAM ALENEX/ANALCO '07*, 2007.

[6] A. Akhavi et B. Vallée : Average bit-complexity of Euclidean algorithms. *In ICALP'2000 - Genève*, vol. 1853 de *Lecture Notes in Computer Science*, p. 373–387. Springer, 2000.

[7] T. M. Apostol : *Introduction to Analytic Number Theory.* Springer, 1976.

[8] K. Babenko : On a problem of Gauss. *Soviet Mathematical Doklady*, 1(19):136–140, 1978.

[9] V. Baladi et B. Vallée : Euclidean algorithms are Gaussian. *Journal of Number Theory*, 2(110):331–386, 2005.

[10] J. Binet : Note sur le nombre des divisions à effectuer pour obtenir le plus grand diviseur commun de deux nombres entiers ; suivie d'une remarque sur une classe de séries récurrentes. *C.R. Acad. Sci. Paris*, 19:867–870, Novembre 1844.

[11] J. Bourdon, B. Daireaux et B. Vallée : Dynamical analysis of alpha-Euclidean algorithms. *J. Algorithms*, 44(1):246–285, 2002.

[12] R. P. Brent : Analysis of the binary Euclidean algorithm. *In* J. F. Traub, éd. : *Algorithms and Complexity*, p. 321–355. Academic Press, 1976.

[13] D. J. Bump : *Automorphic Forms and Representations.* Cambridge Studies in Advanced Mathematics. Cambridge University Press, 1998.

[14] E. Cesaratto, J. Clément, B. Daireaux, L. Lhote, V. Maume-Deschamps et B. Vallée : Analysis of fast versions of the Euclid algorithm. *In Proceedings of SIAM ANALCO'07*, 2007.

[15] F. Chazal, V. Maume-Deschamps et B. Vallée : Erratum to "Dynamical sources in information theory : fundamental intervals and word prefixes". *Algorithmica*, (38):591–596, 2004.

[16] J. H. Conway et N. Sloane : *Sphere Packings, Lattices and Groups.* Springer-Verlag NY, third édn, 1998.

[17] D. Coppersmith : Small solutions to polynomial equations, and low exponent rsa vulnerabilities. *Journal of Cryptology*, (10):233–260, 1997.

[18] H. Daudé, P. Flajolet et B. Vallée : An average-case analysis of the Gaussian algorithm for lattice reduction. *In Combinatorics, Probability and Computing*, vol. 6, p. 397–433. Cambridge University Press, 1997.

[19] H. Daudé et B. Vallée : An upper bound on the average number of iterations of the LLL algorithm. *Theoretical Computer Science*, 123:95–115, 1994.

[20] G. L. Dirichlet : Verallgemeinerung eines satzes aus der lehre von den kettenbrüchen nebst einigen anwendungen auf die theorie der zahlen. *Bericht über die Verhandlungen der Königlich Preussischen Akademie der Wissenschaften*, p. 93–95, 1842.

[21] J. Dixon : The number of steps in the Euclidean algorithm. *Journal of Number Theory*, (2):414–422, 1970.

[22] A. Dupré : Sur le nombre de divisions à effectuer pour obtenir le plus grand commun diviseur entre deux nombres entiers. *Journal de Mathématiques Pures et Appliquées*, 11:41–74, 1846.

[23] P. Flajolet et R. Sedgewick : *Analytic Combinatorics*. Cambridge University Press, 2008.

[24] L. Ford : Fractions. *American Mathematical Monthly*, (9):586–601, 1938.

[25] A. M. Frieze, J. Hastad, R. Kannan, J. C. Lagarias et A. Shamir : Reconstructing truncated integer variables satisfying linear congruences. *SIAM Journal on Computing*, 17(2):262–280, 1988.

[26] C. F. Gauss : *Disquisitiones Arithmeticae*. Leipzig, 1801.

[27] D. Goldstein et A. Mayer : On the equidistribution of Hecke points. *Forum Mathematicum*, 15(2):165–189, 2003.

[28] A. Grothendieck : Résumé des résultats essentiels dans la théorie des produits tensoriels topologiques et des espaces nucléaires. *Annales de l'institut Fourier*, (4):73–112, 1952.

[29] A. Grothendieck : La théorie de Fredholm. *Bulletin de la Société Mathématique de France*, (84):319–384, 1956.

[30] G. Hardy et E. Wright : *An Introduction to the Theory of Numbers*. Oxford University Press, 5ème édn, 1988.

[31] H. Heilbronn : On the average length of a class of continued fractions. *In* P. Turan, éd. : *Number Theory and Analysis*, p. 87–96, 1969.

[32] D. Hensley : The number of steps in the Euclidean algorithm. *Journal of Number Theory*, 2(49):149–182, 1994.

[33] R. Kannan, A. Lenstra et L. Lovász : Polynomial factorization and non-randomness of bits of algebraic and some transcendental numbers. *In Proceedings of STOC'84*, p. 191–200. ACM Press, 1984.

[34] T. Kato : *Perturbation Theory for Linear Operators*. Springer-Verlag, 1980.

[35] D. E. Knuth : *Fundamental Algorithms*, vol. 2 de *The Art of Computer Programming*. Addison-Wesley, 3ème édn, 1997.

[36] D. E. Knuth : *Seminumerical Algorithms*, vol. 2 de *The Art of Computer Programming*. Addison-Wesley, 3ème édn, 1998.

[37] D. E. Knuth : *Sorting and Searching*, vol. 3 de *The Art of Computer Programming*. Addison-Wesley, 2ème édn, 1998.

[38] M. Krasnoselskii : *Positive Solutions of Operator Equations*. P. Noordhoff, Groningen, 1964.

[39] J. C. Lagarias : The computational complexity of simultaneous Diophantine approximation problems. *SIAM Journal on Computing*, 14(1):196–209, 1985.

[40] J. C. Lagarias et A. M. Odlyzko : Solving low-density subset sum problems. *Journal of the ACM*, 32(1):229–246, jan. 1985.

[41] T. Lagny : Analyse générale ou méthodes nouvelles pour résoudre les problèmes de tous les genres et de tous les degrés à l'infini. *Mém. Acad. Sci. Paris*, 11:363–364, 1733.

[42] J.-L. Lagrange : Recherches d'arithmétique. *Nouveaux mémoires de l'Académie royale des sciences et belles-lettres de Berlin, années 1773 et 1775*.

[43] G. Lamé : Note sur la limite du nombre des divisions dans la recherche du plus grand commun diviseur entre deux nombres entiers. *C.R. Acad. Sci. Paris*, 19:867–870, Octobre 1844.

[44] S. Lang : $SL_2(\mathbb{R})$. Springer, 1985.

[45] H. Laville et B. Vallée : Distribution de la constante d'Hermite et du plus court vecteur dans les réseaux de dimension deux. *J. Théor. Nombres Bordeaux*, 6(1), 1994.

[46] A. Lenstra, H. Lenstra et L. Lovász : Factoring polynomials with rational coefficients. *Mathematische Annalen*, 261(4):515–534, 1982.

[47] L. Lhote : Computation of a class of continued fraction constants. *In Proceedings of SIAM ALENEX/ANALCO '04*, p. 199–210, 2004.

[48] L. Lhote et B. Vallée : Sharp estimates for the main parameters of the Euclid algorithm. *In LATIN*, p. 689–702, 2006.

[49] L. Lhote et B. Vallée : Gaussian laws for the main parameters of the Euclid algorithms. *Algorithmica*, 50(4):497–554, 2008.

[50] C. Ludwig : A faster lattice reduction method using quantum search. *In* T. Ibaraki, N. Katoh et H. Ono, éds : *ISAAC*, vol. 2906 de *Lecture Notes in Computer Science*, p. 199–208. Springer, 2003.

[51] M. Madritsch et B. Vallée : Modelling the LLL algorithm via sandpiles. 2009. (submitted).

[52] J. Martinet : *Perfect Lattices in Euclidean Spaces*. Springer-Verlag, 2003.

[53] D. H. Mayer : Continued fractions and related transformations. *In Ergodic Theory, Symbolic Dynamics and Hyperbolic Spaces,*, p. 175–222. Oxford University Press, 1991.

[54] R. Merkle et M. Hellman : Hiding information and signatures in trapdoor knapsacks. *IEEE Transactions in Information Theory*, IT-24:525–530, 1978.

[55] D. Micciancio et S. Goldwasser : *Complexity of Lattice Problems : a cryptographic perspective*, vol. 671 de *The Kluwer International Series in Engineering and Computer Science*. Kluwer Academic Publishers, Boston, Massachusetts, mars 2002.

[56] D. Micciancio et O. Regev : Lattice-based cryptography. *In* D. J. Bernstein et J. Buchmann, éds : *Post-quantum Cryptography*. Springer, 2008.

[57] H. Minkowski : *Geometrie der Zahlen.* B. G. Teubner, Leipzig, 1896.

[58] P. Nguyen et D. Stehlé : Floating-point LLL revisited. *In Proceedings of Eurocrypt 2005*, vol. 3494 de *Lecture Notes in Computer Science*, p. 215–233. Springer-Verlag, 2005.

[59] P. Nguyen et J. Stern : The two faces of lattices in cryptology. *In Proceedings of CALC '01*, vol. 2146 de *Lecture Notes in Computer Science*, p. 146–180. Springer, 2001.

[60] P. Q. Nguyen et D. Stehlé : LLL on the average. *In* F. Hess, S. Pauli et M. E. Pohst, éds : *ANTS*, vol. 4076 de *Lecture Notes in Computer Science*, p. 238–256. Springer, 2006.

[61] NTRU Cryptolabs : Recovery challenge. `http://www.ntru.com/cryptolab/challenges_background.htm#recovery`.

[62] A. M. Odlyzko et H. te Riele : Disproof of the Mertens conjecture. *J. Reine Angew. Math.*, (357):138–160, 1985.

[63] D. Ruelle : *Thermodynamic formalism : the mathematical structures of classical equilibrium statistical mechanics*. Addison-Wesley, 1978.

[64] R. Sedgewick et P. Flajolet : *An Introduction to the Analysis of Algorithms*. Addison-Wesley, 1996.

[65] J.-P. Serre : *A course in arithmetic*. Graduate Texts in Mathematics. Springer-Verlag, 1973.

[66] A. Shamir : A polynomial-time algorithm for breaking the basic Merkle-Hellman cryptosystem. *IEEE Transactions on Information Theory*, 30(5):699–704, 1984.

[67] V. Shoup : NTL : A library for doing number theory. `http://www.shoup.net/ntl/doc/LLL.txt`.

[68] C. L. Siegel : A mean value theorem in geometry of numbers. *Annals of Mathematics*, 46(2), 1945.

[69] C. L. Siegel : *Lectures on the Geometry of Numbers*. Springer, 1989.

[70] D. Stehlé : Floating-point LLL : theoretical and practical aspects. *In LLL+25*, Information Security and Cryptography Series. Springer-Verlag, 2009.

[71] D. Stehlé : `fplll-3.0`. `http://perso.ens-lyon.fr/damien.stehle/#software`.

[72] J. Stern : Secret linear congruential generators are not cryptographically secure. *In Proceedings of FOCS'87*, p. 421–426. IEEE Computer Society, 1987.

[73] B. Vallée et P. Flajolet : The lattice reduction algorithm of Gauss : An average case analysis. *In Proceedings of FOCS'90*, vol. II, p. 830–839. IEEE Computer Society, 1990.

[74] B. Vallée et A. Vera : Lattice reduction in two dimensions : analyses under realistic probabilistic models. *In Proceedings of AofA'07*. DMTCS, 2007.

[75] B. Vallée et A. Vera : Probabilistic analyses of lattice reduction algorithms. *In LLL+25*, Information Security and Cryptography Series. Springer-Verlag, 2009.

[76] B. Vallée : Gauss' algorithm revisited. *Journal of Algorithms*, 12:556–572, 1991.

[77] B. Vallée : Opérateurs de Ruelle-Mayer généralisés et analyse en moyenne des algorithmes de Gauss et d'Euclide. *Acta Arithmetica*, 2(81):101–144, 1997.

[78] B. Vallée : Dynamics of the binary Euclidean algorithm : Functional analysis and operators. *Algorithmica*, 22:660–685, 1998.

[79] B. Vallée : Dynamical sources in information theory : fundamental intervals and word prefixes. *Algorithmica*, (29):262–306, 2001.

[80] B. Vallée : Euclidean dynamics. *Discrete and Continuous Dynamical Systems*, 1(15):281–352, 2006.

[81] P. van Emde Boas : Another NP-complete partition problem and the complexity of computing short vectors in a lattice. Rap. tech. 81-04, Mathematisch Instituut, Amsterdam, 1981.

[82] E. Wirsing : On the theorem of Gauss-Kuzmin-Lévy and a Frobenius-type theorem for function spaces. *Acta Arithmetica*, (24):507–528, 1974.