ASTÉRISQUE 276

SÉMINAIRE BOURBAKI
VOLUME 1999/2000
EXPOSÉS 865-879

Société Mathématique de France 2002
Publié avec le concours du Centre National de la Recherche Scientifique

Association des collaborateurs de Nicolas Bourbaki, École normale supérieure,
45, rue d'Ulm, F-75230 Paris Cedex 05.
Url : http://www.bourbaki.ens.fr/

***Mots clefs et classification mathématique par sujets* (2000)**

Exposé n° 865. — Groupes de tresses, groupes ordonnés, groupes de difféotopies, problème des mots, grands cardinaux, autodistributivité. — 20F36, 20F60, 20F10, 57M07, 06F15, 03E55, 08A50.
Exposé n° 866. — Mouvement brownien, propriété trajectorielle, marche aléatoire à boucles effacées, exposant d'intersection, exposant de croissance, invariance conforme, pavage par dominos, arbre couvrant, fonction de couplage. — 60J65, 60J15, 05C70.
Exposé n° 867. — Group action, K-theory, derived category, quotient variety, resolution of singularity, motivic integration, McKay correspondence, Hilbert schemes of G-orbits. — 14xx.
Exposé n° 868. — Vortex, supraconductivity, Ginzburg-Landau equations, Abrikosov lattices, abelian gauge theory, Yang-Mills Higgs equations, elliptic PDE, variational methods, free boundary problems. — 35Jxx, 35Qxx, 49Jxx, 49Rxx, 58E15, 81T13.
Exposé n° 869. — Conjecture de Baum-Connes, K-théorie, réseaux dans les groupes de Lie, propriété T. — 19K99, 22E50, 58G12.
Exposé n° 870. — Théorie des modèles, géométrie diophantienne, corps de différence, conjecture de Manin-Mumford. — 03C60, 14K15, 11G10, 03C45.
Exposé n° 871. — Elliptic curves, modular curves, Galois representations, automorphic representations. — 11F80, 11G18, 14G35.
Exposé n° 872. — Variété algébrique réelle, variété symplectique, courbes rationnelles, variété de Fano, flot hamiltonien, trajectoire de Floer. — 53D12, 14P25.
Exposé n° 873. — Correspondance de Langlands, chtoucas de Drinfeld, corps de fonctions. — 11Fxx, 14Fxx, 22Exx.
Exposé n° 874. — Motivic integration, motivic zeta function, motivic convolution, McKay correspondence. — 14Exx, 14F42.
Exposé n° 875. — Vertex algebra, Virasoro algebra, conformal field theory, moduli space of curves, chiral algebra. — 17B67, 17B68, 17B90, 81T40, 14H10, 14H60.
Exposé n° 876. — Eisenstein series, Shimura curves, height pairings, arithmetic cycles. — 14G40, 11G15, 11G18, 11F27, 11F30, 14G35, 11G50, 11F46.
Exposé n° 877. — Polynomial functors, Steenrod algebra, Schur algebra, polynomial representations, general linear group, stable K-theory. — 19D55, 55S10, 16G10.
Exposé n° 878. — Braid groups, Burau representations, BMW-algebras. — 20F36, 57M99.
Exposé n° 879. — Random matrices, random permutations, integrable systems and Painlevé equations. — 15A52, 37K10.

SÉMINAIRE BOURBAKI
VOLUME 1999/2000
EXPOSÉS 865-879

Résumé. — Comme les précédents volumes de ce séminaire, celui-ci contient quinze exposés de synthèse sur des sujets d'actualité : deux exposés de théorie des groupes, un sur les algèbres de dimension infinie, trois de géométrie algébrique, deux de géométrie arithmétique, un sur la correspondance de Langlands, deux de probabilités, un d'équations aux dérivées partielles, un d'algèbres d'opérateurs, un de théorie des modèles et un sur les foncteurs polynomiaux.

***Abstract* (Séminaire Bourbaki, volume 1999/2000, exposés 865-879)**

As in the preceding volumes of this seminar, one finds here fifteen survey lectures on topics of current interest: two lectures on group theory, one on infinite dimensional algebras, three of algebraic geometry, two of arithmetic geometry, one on the Langlands correspondence, two on probabilities, one on partial differential equations, one on operator algebras, one on model theory, and one on polynomial functors.

Christian KASSEL – *L'ordre de Dehornoy sur les tresses*

Au début des années 1990 Dehornoy a déduit un ordre total sur le groupe des tresses d'Artin à partir de l'étude générale des systèmes autodistributifs, définis comme des ensembles munis d'une loi de composition vérifiant l'identité $x(yz) = (xy)(xz)$. Cette étude avait été motivée par un axiome indémontrable de théorie des ensembles impliquant l'existence d'un système autodistributif remarquable. Dans ce texte on présente les travaux de Dehornoy ainsi que leur lien inattendu avec la théorie des ensembles. On expose aussi deux constructions géométriques récentes de l'ordre de Dehornoy.

Jean-François LE GALL – *Exposants critiques pour le mouvement brownien et les marches aléatoires [d'après Kenyon, Lawler et Werner]*

Nous discutons certains travaux récents qui étudient les exposants critiques pour le mouvement brownien et les marches aléatoires. Dans la première partie, nous décrivons les résultats de Lawler et Werner donnant des informations précises sur les exposants d'intersection browniens. Dans la seconde, nous présentons les travaux de Kenyon qui conduisent au calcul exact de l'exposant de croissance pour la marche aléatoire à boucles effacées dans le plan.

Miles REID – *La correspondance de McKay*

Let M be a quasiprojective algebraic manifold with $K_M = 0$ and G a finite automorphism group of M acting trivially on the canonical class K_M; for example, a subgroup G of $\mathrm{SL}(n, \mathbb{C})$ acting on $\mathbb{C}^n$ in the obvious way. We aim to study the quotient variety $X = M/G$ and its resolutions $Y \to X$ (especially under the assumption that Y has $K_Y = 0$) in terms of G-equivariant geometry of M. At present we know 4 or 5 quite different methods of doing this, taken from string theory, algebraic geometry, motives, moduli, derived categories, etc.

For G in $\mathrm{SL}(n, \mathbb{C})$ with $n = 2$ or 3, we obtain several methods of cobbling together a basis of the homology of Y consisting of algebraic cycles in one-to-one correspondence with the conjugacy classes or the irreducible representations of G.

Tristan RIVIÈRE – *Ginzburg-Landau Vortices: the static model*

We study the formation of vortices (zero sets having a "topological charge") for critical points of the Ginzburg-Landau Functional arising in the physics of supraconductivity and more generally in abelian gauge theory. We present results regarding the existence or non existence of vortices in the fundamental state, their location in space and the eventual formation of Abrikosov lattices. We will establish a link between the position of the vortices and the total space of solutions in the strongly repulsive limit.

Georges SKANDALIS – *Progrès récents sur la conjecture de Baum–Connes. Contribution de Vincent Lafforgue*

Cet exposé, à la suite de celui de Pierre Julg en 1998, fait un nouveau point sur la conjecture de Baum-Connes. Une nouvelle barrière vient en effet juste d'être franchie par V. Lafforgue : celle de la propriété T de Kazhdan. Le but ici est d'exposer les diverses étapes du travail de Lafforgue.

Élisabeth BOUSCAREN – *Théorie des modèles et conjecture de Manin-Mumford* [*d'après Ehud Hrushovski*]

Nous présenterons des applications récentes de la théorie des modèles à des questions de géométrie diophantienne sur les corps de nombres. Nous indiquerons en particulier comment E. Hrushovski, en utilisant la théorie des corps algébriquement clos munis d'un automorphisme, donne une nouvelle démonstration de la conjecture de Manin-Mumford, démonstration qui produit de bonnes bornes effectives.

Bas EDIXHOVEN – *Rational elliptic curves are modular* [*after Breuil, Conrad, Diamond and Taylor*]

In 1994, Wiles and Taylor-Wiles proved that every semistable elliptic curve over $\mathbf{Q}$ is modular, in the sense that it is a quotient of the jacobian of some modular curve. This result has since then been generalized by an increasing sequence of groups of authors, culminating in the proof by Breuil, Conrad, Diamond and Taylor that all elliptic curves over $\mathbf{Q}$ are modular. Their strategy is essentially that of Wiles, but they have to overcome many technical problems. Especially the deformation theory of local Galois representations becomes much harder.

Viatcheslav KHARLAMOV – *Variétés de Fano réelles* [*d'après C. Viterbo*]

Les variétés de Fano sont les variétés à fibré anticanonique ample. Elles sont dominées par des familles de courbes rationnelles. Un cas particulier d'une conjecture de J. Kollàr affirme que la partie réelle d'une variété de Fano définie sur $\mathbb{R}$ et de dimension supérieure à trois n'admet aucune métrique à courbure strictement négative. Utilisant des méthodes de géométrie symplectique, C. Viterbo a montré cette conjecture dans le cas particulier où $b_2 = 1$ et la variété est dominée par des courbes rationnelles d'aire minimale : en étudiant le flot de Floer, il montre comment l'existence de courbes rationnelles entraîne celle de géodésiques fermées d'indice non nul sur la partie réelle.

Gérard LAUMON – *La correspondance de Langlands sur les corps de fonctions* [*d'après Laurent Lafforgue*]

Laurent Lafforgue a établi la correspondance de Langlands pour GL_r sur un corps de fonctions. Sa preuve suit la stratégie introduite, il y a plus de 25 ans, par V. Drinfeld pour traiter le cas $r = 2$. Une des innovations principales est la construction d'une compactification toroïdale de $\mathrm{PGL}_r^{n+1} / \mathrm{PGL}_r$ pour n arbitraire qui généralise la compactification de C. De Concini et C. Procesi ($n = 1$).

Dans l'exposé oral nous présenterons le théorème de L. Lafforgue et ses conséquences, ainsi que quelques aspects de sa très longue démonstration (près de 700 pages).

Eduard LOOIJENGA – *Motivic measures*

Kontsevich proposed a Haar measure on $k[[t]]$, k a field of characteristic zero, with values in a universal ring that is obtained in a simple manner from the varieties over k: it assigns to the ideal $t^n k[[t]]$ the value $\mathbb{L}^{-n}$, where $\mathbb{L}$ stands for the affine line over k. This leads to a measure with the same value ring for schemes over $k[[t]]$. Denef & Loeser and Batyrev have shown that this measure gives rise to invariants with remarkable properties. These are new even in case the scheme is obtained from a k-variety by base change.

Edward FRENKEL – *Vertex algebras and algebraic curves*

Vertex algebras are algebraic objects that formalize the concepts of vertex operators and operator product expansion which originated from physics. They were defined by Borcherds and studied extensively in the last fifteen years. After giving the definition of vertex algebra and describing a few examples, we show how to attach to a vertex algebra a geometric object: a vector bundle on a formal disc equipped with a flat connection and a canonical horizontal section. We then associate to a vertex algebra and an algebraic curve an invariant called the space of conformal blocks. These spaces reflect the geometric structure of various moduli spaces associated to algebraic curves.

Stephen S. KUDLA – *Derivatives of Eisenstein series and generating functions for arithmetic cycles*

In their classic work, Hirzebruch and Zagier showed that certain generating functions whose coefficients are the cohomology classes of curves on Hilbert modular surfaces are the q–expansions of elliptic modular forms of weight 2. In this talk I will describe an analogous family of generating functions whose coefficients arise from arithmetical algebraic geometry, e.g., from 0–cycles on the arithmetic surfaces associated to Shimura curves. The identification of such a function with the derivative of a Siegel–Eisenstein series at its center of symmetry provides a kind of arithmetic analogue of the Siegel–Weil formula.

Teimuraz PIRASHVILI – *Polynomial functors over finite fields* [*after Franjou, Friedlander, Henn, Lannes, Schwartz, Suslin*]

For finite fields there is an essential difference between polynomial maps and polynomials. This yields two different versions of polynomial endofunctors of finite vector spaces. The first one is closely related to unstable modules over Steenrod algebra, while the second one is related to representations of algebraic groups. Recent works by Betley, Franjou, Friedlander, Suslin and others lead to a comparaison theorem between associated Ext-groups. These results are a key for the solutions of Betley-Pirashvili conjecture involving the stable K-theory of Waldhausen and for computation of cohomology of the general linear group over finite fields with twisted coefficients.

Vladimir TURAEV – *Faithful linear representations of the braid groups*

Recently D. Krammer and S. Bigelow showed that a certain homomorphism of the group of braids on n strings B_n to a group of real matrices is injective for all $n \geq 1$. This answers in the positive the long-standing question of linearity of B_n. We shall discuss these results of D. Krammer and S. Bigelow as well as earlier results of J. Moody (1991) and others on the non-faithfulness of the classical Burau representation of B_n.

Pierre VAN MOERBEKE – *Random matrices and permutations, matrix integrals and integrable systems*

This lecture present a survey of recent developments in the area of random matrices (finite and infinite) and random permutations. These probabilistic problems suggest matrix integrals (or Fredholm determinants), which arise very naturally as integrals over the tangent space to symmetric spaces, as integrals over groups and finally as integrals over symmetric spaces. Upon apropriately adding time-parameters, these matrix integrals are natural tau-functions for integrable lattices, like the Toda, Pfaff and Toeplitz lattices, but also for integrable PDE's, like the Korteweg-de Vries equation. These matrix integrals or Fredholm determinants also satisfy Virasoro constraints, which combined with the integrable equations lead to (partial) differential equations for the original probabilities.

Séminaire BOURBAKI Novembre 1999
52e année, 1999-2000, n° 865, p. 7 à 28

L'ORDRE DE DEHORNOY SUR LES TRESSES

par **Christian KASSEL**

Une des découvertes récentes les plus intéressantes et inattendues concernant le groupe des tresses B_n d'Artin est due à Dehornoy qui a démontré fin 1991 que B_n a un ordre total invariant par multiplication à gauche. Il en résulte notamment que l'anneau du groupe B_n n'a pas de diviseurs de zéro. Dehornoy déduit cet ordre de l'étude générale des systèmes autodistributifs, définis comme des ensembles munis d'une loi de composition vérifiant l'identité $x(yz) = (xy)(xz)$. Des travaux sur un axiome indémontrable de théorie des ensembles avaient mis en évidence un système autodistributif remarquable et suggéré que l'étude de tels systèmes devait donner des résultats non triviaux.

Fenn, Greene, Rolfsen, Rourke et Wiest ont donné en 1997 une construction géométrique de l'ordre de Dehornoy. Leurs techniques ont permis également de mettre des structures d'ordre sur des groupes de difféotopies de surfaces. Plus récemment, Thurston a indiqué comment construire un ordre total sur B_n en utilisant les travaux de Nielsen sur les surfaces.

Le but de cet exposé est de présenter ces travaux ainsi que le lien inattendu avec la théorie des ensembles. Au n° 1 nous énonçons le résultat principal sur l'ordre des tresses et quelques propriétés de cet ordre. Les deux numéros suivants sont consacrés à une présentation des techniques de Dehornoy et à sa démonstration du résultat principal ; le lien entre tresses et systèmes autodistributifs est exposé au n° 2 tandis qu'au n° 3 nous étudions deux systèmes autodistributifs particulièrement importants. Au n° 4 nous indiquons comment l'étude des grands cardinaux en théorie des ensembles a été à l'origine de ces travaux. La construction géométrique de Fenn *et al.* ainsi que l'approche à la Nielsen seront présentées au n° 5. Pour plus de détails, le lecteur pourra consulter les références citées, et notamment la monographie [D00].

Thomas Delzant m'a aidé à comprendre l'idée de Thurston et à mettre le n° 5.2 en forme. Je le remercie très chaleureusement ainsi que Patrick Dehornoy, Dale Rolfsen et Bertold Wiest qui m'ont fait profiter de leurs commentaires.

1. LES RÉSULTATS PRINCIPAUX

Dans ce numéro nous rappelons la définition des groupes de tresses et celle des groupes ordonnables avant d'énoncer le théorème principal de Dehornoy impliquant l'existence d'un ordre sur les tresses.

1.1. Le groupe des tresses d'Artin

Soit n un entier ≥ 2. Définissons B_n comme le groupe engendré par $n-1$ générateurs $\sigma_1, \ldots, \sigma_{n-1}$ soumis aux relations

$$\sigma_i \sigma_j = \sigma_j \sigma_i \tag{1.1}$$

si $|i-j| > 1$ et

$$\sigma_i \sigma_j \sigma_i = \sigma_j \sigma_i \sigma_j \tag{1.2}$$

si $|i-j| = 1$. On utilisera souvent le mot *tresse* pour désigner un élément de B_n.

E. Artin a introduit le groupe B_n dans les années 1920 et a résolu le problème des mots[(1)] pour ce groupe (voir [A26], [A47]). En 1967 Garside [Ga] résout à la fois ce problème et le problème de conjugaison[(1)] (voir aussi [Mk]). D'autres solutions ont été proposées, fondées notamment sur une forme normale due à Adyan [Ad], et redécouverte un peu plus tard par ElRifai et Morton [EM] et par Thurston, ce dernier élaborant à cette occasion la théorie des groupes automatiques et établissant l'automaticité de B_n (les résultats de Thurston sont exposés au chapitre 9 de [E+]). Rappelons aussi que les groupes de tresses jouent un rôle important dans des branches des mathématiques aussi diverses que la théorie des nœuds, la théorie de l'homotopie, la géométrie algébrique, la théorie des groupes, celle des représentations, *etc.* (on trouvera un reflet de cette diversité dans [B+] et dans [Ca]).

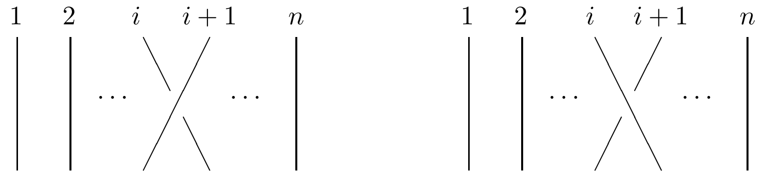
Un *mot de tresse* w est un mot formé à partir des générateurs $\sigma_1, \ldots, \sigma_{n-1}$ et de leurs inverses. Il est commode de représenter w par une configuration plane formée par n intervalles, appelés les brins de w, plongés de manière lisse dans une bande verticale bornée du plan. Une telle configuration s'obtient comme suit.

(a) On représente le générateur σ_i et son inverse σ_i^{-1} par les configurations de la figure 1. Le croisement de la configuration de σ_i (resp. de σ_i^{-1}) est dit *positif* (resp. *négatif*).

(b) Si D est la configuration associée au mot w et D' celle du mot w', alors la configuration associée à ww' est obtenue en plaçant D au-dessus de D' et en joignant les extrémités inférieures des brins de D aux extrémités supérieures de D'.

---

(1) Résoudre le problème des mots (resp. le problème de conjugaison) dans un groupe, c'est trouver un algorithme permettant de reconnaître ou de décider si deux mots donnés représentent ou non le même élément dans le groupe (resp. la même classe de conjugaison).



σ_i $\qquad$ σ_i^{-1}

FIGURE 1.

1.2. Groupes ordonnables

Un groupe G est dit *ordonnable* s'il possède un ordre total $<$ invariant par multiplication à gauche, c'est-à-dire vérifiant $a < b \Rightarrow ca < cb$ pour tout a, b, $c \in G$. Si, de plus, l'ordre est invariant par multiplication à droite (c'est-à-dire $a < b \Rightarrow ac < bc$), on dit que le groupe est *biordonnable*.

Pour tout groupe ordonnable G, notons P l'ensemble $\{x \in G \mid x > 1\}$ de ses éléments positifs. Le caractère total de l'ordre se traduit par la partition $G = P \amalg \{1\} \amalg P^{-1}$ où $P^{-1} = \{x \in G \mid x^{-1} \in P\}$. L'invariance par multiplication à gauche se traduit par $P^2 \subset P$, où P^2 est formé des produits de deux éléments de P dans G. Réciproquement, s'il existe un sous-ensemble P de G vérifiant

$$G = P \amalg \{1\} \amalg P^{-1} \quad \text{et} \quad P^2 \subset P,$$

alors G est ordonnable pour l'ordre défini par $x < y$ si $x^{-1}y \in P$. L'ordre en question fait de G un groupe biordonnable si et seulement si $xPx^{-1} \subset P$ pour tout $x \in G$.

L'ordonnabilité d'un groupe a de nombreuses conséquences (voir [MR], chap. 7 ou [Pa], chap. 13). D'abord, tout groupe ordonnable G est sans torsion. Si, de plus, R est un anneau sans diviseurs de zéro, alors l'algèbre du groupe $R[G]$ n'a pas de diviseurs de zéro (elle vérifie donc une fameuse conjecture de Kaplansky). Il en résulte que les seuls idempotents de $R[G]$ sont triviaux, à savoir 0 ou 1. En outre, les seuls éléments inversibles de $R[G]$ sont triviaux, c'est-à-dire de la forme rx où r est un élément inversible de R et $x \in G$.

Les groupes libres de type fini, les groupes fondamentaux de surfaces fermées, les groupes abéliens libres de type fini sont biordonnables. Pour ce qui concerne les tresses, les groupes B_n ne sont pas biordonnables alors que les groupes de tresses pures le sont (*cf.* [KR] et [RZ]).

1.3. L'ordre de Dehornoy sur B_n

Pour $i \in \{1, \ldots, n-1\}$, on dit qu'un mot de la forme $w_0\sigma_i w_1 \sigma_i \ldots \sigma_i w_r$ est σ_i-positif si les sous-mots $w_0, w_1, \ldots, w_r$ sont des mots dans les lettres $\sigma_j^{\pm 1}$ avec $j > i$. En d'autres termes, dans un mot σ_i-positif, le générateur σ_i d'indice minimal n'apparaît qu'avec des puissances positives. S'il n'apparaît qu'avec des puissances négatives, on dit que le mot est σ_i-négatif (donc un mot est σ_i-négatif si son inverse est σ_i-positif).

Une tresse est dite σ_i-positive (resp. σ_i-négative) si elle peut être représentée par un mot σ_i-positif (resp. σ_i-négatif). Par exemple, le mot $\sigma_1\sigma_2\sigma_1^{-N}$ n'est pas σ_1-positif si N est un entier ≥ 1, mais la tresse qu'il représente l'est (le montrer !).

Nous dirons qu'une tresse est *σ-positive* (resp. *σ-négative*) s'il existe un entier i tel qu'elle est σ_i-positive (resp. σ_i-négative). On peut alors énoncer le théorème de Dehornoy.

THÉORÈME 1.1. — *Toute tresse différente de* 1 *est soit σ-positive, soit σ-négative.*

La démonstration de ce théorème occupera les numéros 2 et 3. Celui-ci a pour conséquence les résultats nouveaux suivants non connus antérieurement.

COROLLAIRE. — (a) *Pour tout n, le groupe B_n est ordonnable.*

(b) *Pour tout n et tout anneau R sans diviseurs de zéro, l'algèbre $R[B_n]$ du groupe B_n n'a pas de diviseurs de zéro, ni d'unités ou d'idempotents non triviaux.*

Démonstration. — (a) Soit P l'ensemble des tresses σ-positives de B_n. Il est clair que P^2 est inclus dans P. Le théorème 1.1 implique $B_n = P \amalg \{1\} \amalg P^{-1}$. Il en résulte que B_n est ordonnable.

(b) C'est une conséquence de (a) et des remarques faites au n° 1.2. $\square$

1.4. Propriétés

L'ordre de Dehornoy ainsi défini est l'unique ordre total de B_n invariant par multiplication à gauche tel que $\beta\,\sigma_1\,\beta' > 1$ pour tous les éléments β, β' du sous-groupe de B_n engendré par $\sigma_2, \ldots, \sigma_{n-1}$. Dans [D97b] Dehornoy a construit un algorithme de résolution du problème des mots fondé sur l'ordre total des tresses. En pratique, cet algorithme est beaucoup plus efficace que les algorithmes antérieurs (mais le problème reste ouvert de montrer qu'il est quadratique).

Soit B_n^+ le sous-monoïde de B_n constitué des tresses représentées par des produits des générateurs $\sigma_1, \ldots, \sigma_{n-1}$, mais non de leurs inverses. Laver [L96] a montré que l'ordre de Dehornoy prolonge l'ordre partiel considéré par ElRifai et Morton dans [EM] ; il en déduit le résultat remarquable suivant, qui distingue probablement cet ordre parmi les ordres de B_n mentionnés au n° 5.2.

THÉORÈME 1.2. — *L'ordre de Dehornoy restreint à B_n^+ est un bon ordre.*

Rappelons qu'un ordre total sur un ensemble E est un *bon ordre* si toute partie de E a un élément minimal. Dans un bon ordre, chaque élément a un rang ; par exemple, le générateur σ_{n-1} de B_n^+ est le plus petit élément > 1. Comme tout élément de B_n possède une écriture canonique comme fraction de deux éléments de B_n^+ (voir [E+]), le théorème de Laver permet de spécifier une tresse à l'aide d'un couple d'ordinaux. Burckel [Bu] a donné une autre preuve du théorème 1.2 et montré que le type de l'ordre de B_n^+ est $\omega^{\omega^{n-2}}$, où ω désigne le plus petit ordinal infini. Voir aussi l'article de Wiest [W] basé sur la définition géométrique de l'ordre de B_n présentée au n° 5.1.

2. COLORIAGE DES TRESSES ET SYSTÈMES AUTODISTRIBUTIFS

Le but du n° 2 est d'indiquer comment Dehornoy démontre le théorème 1.1.

2.1. Comment colorier les tresses ?

L'idée n'est pas nouvelle ; elle a notamment été utilisée par Joyce [Joy], Matveev [Mt] et Brieskorn [Br] dans les années 1980. Partons d'un mot de tresse w représentant un élément de B_n. Colorions de gauche à droite les extrémités supérieures des brins de la configuration plane associée à w avec les éléments $a_1, a_2, \ldots, a_n$ d'un ensemble S. Si nous propageons ces couleurs le long des brins, nous lirons au bas de la configuration une suite de couleurs qui est une permutation de la suite $a_1, a_2, \ldots, a_n$ de départ : c'est évidemment la permutation sous-jacente à la tresse.

Les choses deviennent plus intéressantes si nous autorisons les couleurs à interagir à chaque croisement des brins. Supposons que les croisements soient exclusivement positifs, c'est-à-dire que le mot w soit un produit des générateurs σ_i et non de leurs inverses. On dira d'un tel mot qu'il est *positif*. Décidons qu'à chaque croisement la couleur a du brin inférieur reste inchangée tandis que la couleur b du brin supérieur s'amalgame avec la couleur a de l'autre brin pour former une nouvelle couleur $a*b$ qui dépend de a et de b (voir la figure 2). Ceci revient à munir S d'une loi de composition

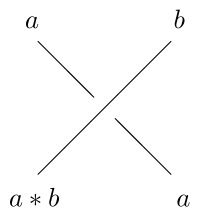


FIGURE 2.

$(a, b) \mapsto a*b$. Procédant de cette manière pour tous les croisements, nous construisons ainsi une action (à droite) du monoïde des mots de tresse positifs à n brins sur la puissance n-ième S^n de l'ensemble S. Si on note ε le mot vide, alors cette action est définie par récurrence sur la longueur des mots par $(a_1, \ldots, a_n)\,\varepsilon = (a_1, \ldots, a_n)$ et

$$(a_1, \ldots, a_n)\,\sigma_i w = (a_1, \ldots, a_{i-1}, a_i * a_{i+1}, a_i, a_{i+2}, \ldots, a_n)\,w \tag{2.1}$$

où $(a_1, a_2, \ldots, a_n) \in S^n$, w est un mot de tresse et $1 \leq i < n$.

Au n° 1.4 nous avons défini B_n^+ comme le sous-monoïde du groupe B_n formé des tresses représentées par des mots positifs. On sait que le monoïde B_n^+ a une présentation donnée par les relations (1.1) et (1.2).

Lemme 2.1. — *La formule (2.1) définit une action du monoïde B_n^+ sur S^n si et seulement si, pour tout a, b, $c \in S$, on a*

$$a * (b * c) = (a * b) * (a * c). \tag{AD}$$

Démonstration. — On se ramène aux calculs suivants où a, b, c, d désignent des éléments de l'ensemble S. Le premier calcul reflète la relation (1.1) :

$$(a, b, c, d)\, \sigma_1\sigma_3 = (a * b, a, c * d, c) = (a, b, c, d)\, \sigma_3\sigma_1.$$

La relation non triviale (1.2) montre la nécessité de l'identité (AD) :

$$(a, b, c)\, \sigma_1\sigma_2\sigma_1 = \big((a * b) * (a * c), a * b, a\big),$$
$$(a, b, c)\, \sigma_2\sigma_1\sigma_2 = \big(a * (b * c), a * b, a\big).$$

Définition. — *Un système autodistributif (ou encore un système AD) est un ensemble muni d'une loi de composition vérifiant la relation (AD) du lemme 2.1.*

Par la suite, on utilisera les notions évidentes de morphisme de systèmes AD, de sous-système AD, de partie génératrice d'un système AD.

2.2. Ensembles automorphes

Dans ce numéro, nous donnons des exemples de systèmes autodistributifs. Il s'agit ici de ce que Brieskorn [Br] a appelé des *ensembles automorphes*, à savoir des systèmes autodistributifs pour lesquels les multiplications à gauche $b \mapsto a * b$ sont bijectives (ces systèmes particuliers ont également été utilisés sous d'autres noms, comme « wrack » chez Conway et Wraith, « rack » chez Fenn et Rourke [FR], « crystal » chez Kauffman [Kau] ; on trouvera dans [FR] une introduction historique aux ensembles automorphes).

Si S est un ensemble automorphe, alors l'action de B_n^+ sur S^n, qui est à valeurs dans un ensemble de bijections, s'étend en une action du groupe B_n tout entier. On obtient ainsi une méthode systématique de construction de représentations de B_n, méthode dont les exemples (a) et (b) ci-dessous montrent l'intérêt.

(a) *(Conjugaison)* Tout groupe G est muni d'une structure de système AD dont la loi de composition $*$ est définie par $a * b = aba^{-1}$ où a, $b \in G$.

Si G est le groupe libre F_n sur n générateurs $x_1, \ldots, x_n$, posons $(y_1, \ldots, y_n) = (x_1, \ldots, x_n)\, \beta$ pour tout $\beta \in B_n$ La correspondance $(x_1, \ldots, x_n) \mapsto (y_1, \ldots, y_n)$ définit un automorphisme $\widehat{\beta}$ de F_n. L'application $\beta \mapsto \widehat{\beta}^{-1}$ est l'homomorphisme injectif de groupes $B_n \to \mathrm{Aut}(F_n)$ bien connu (voir par exemple [BZ], chap. 10).

(b) *(Barycentre)* L'anneau $\mathbf{Z}[t, t^{-1}]$ des polynômes de Laurent à coefficients entiers a une structure de système AD avec $a * b = (1 - t)a + tb$ où a, $b \in \mathbf{Z}[t, t^{-1}]$. L'action linéaire de B_n^+ sur $\mathbf{Z}[t, t^{-1}]^n$ s'étend en un homomorphisme de groupes $B_n \to \mathrm{GL}_n(\mathbf{Z}[t, t^{-1}])$: c'est la *représentation de Burau*.

Les ensembles automorphes pour lesquels $a * a = a$ pour tout a, comme dans les exemples (a) et (b), ont été appelés « quandles » par Joyce [Joy] et « distributive groupoids » par Matveev [Mt] (tous deux ont utilisé ce type de systèmes AD pour classifier les nœuds dans $\mathbf{R}^3$). Voici deux autres exemples d'ensembles automorphes.

(c) *(Systèmes de racines)* Les vecteurs non nuls d'un espace vectoriel euclidien forment un ensemble automorphe : si a et b sont non nuls, on définit $a * b$ comme l'image de b par la symétrie orthogonale par rapport à l'hyperplan orthogonal à a. En particulier, tout système de racines est un ensemble automorphe.

(d) Si l'on note F_X le groupe libre engendré par les éléments d'un ensemble X, on peut munir l'ensemble-produit $B_X = F_X \times X$ d'une loi autodistributive par

$$(w_1, x_1) * (w_2, x_2) = (w_1 x_1 w_1^{-1} w_2, x_2)$$

où $w_1, w_2 \in F_X$ et $x_1, x_2 \in X$. Le système B_X est un ensemble automorphe libre ; tout ensemble automorphe S est quotient de B_X où X est une partie génératrice de S.

2.3. Cycles et acyclicité

Si a et b sont des éléments d'un système autodistributif S, nous dirons que a divise b (à gauche) s'il existe $c \in S$ tel que $b = a * c$. Un *cycle* est une suite $(a_1, \dots, a_k)$ d'éléments de S tels que a_i divise a_{i+1} pour tout $i < k$ et a_k divise a_1. On dit qu'un système AD est *acyclique* s'il ne possède pas de cycle. Un ensemble automorphe n'est pas acyclique puisque tout élément se divise lui-même.

L'observation suivante montre l'intérêt des systèmes acycliques. Sur un système autodistributif S on peut définir une relation binaire $\prec$ comme suit : $a \prec b$ s'il existe un entier $k \geq 1$ et des éléments $a_1, \dots, a_k$ de S tels que $a_1 = a$, $a_k = b$ et a_i divise a_{i+1} pour tout $i < k$. La relation $\prec$ est transitive, et c'est une relation d'ordre partiel si S est acyclique. Avant les travaux de théorie des ensembles exposés au n° 4, on ne connaissait pas d'exemple de système acyclique.

Proposition 2.2. — *Il existe des systèmes autodistributifs acycliques.*

Démonstration. — Il suffit d'exhiber un exemple. Voici celui de Larue [La] qui utilise l'image dans les automomorphismes d'un groupe libre de l'opération sur les tresses que nous décrirons au n° 3.3. Soit F_∞ le groupe libre sur une infinité dénombrable de lettres $x_1, x_2, \dots$ et $\mathrm{Aut}(F_\infty)$ le groupe de ses automorphismes de groupe. Nous allons munir $\mathrm{Aut}(F_\infty)$ d'une loi autodistributive $*$. Soit α l'automorphisme de F_∞ déterminé par

$$\alpha(x_i) = \begin{cases} x_1 x_2 x_1^{-1} & \text{si } i = 1, \\ x_1 & \text{si } i = 2, \\ x_i & \text{si } i > 2, \end{cases}$$

et T_0 l'endomorphisme de F_∞ donné par $T_0(x_i) = x_{i+1}$ pour tout $i \geq 0$. Ce dernier permet de définir l'endomorphisme T de $\mathrm{Aut}(F_\infty)$ par les formules $T(\varphi)(x_1) = x_1$ et $T(\varphi)(x_i) = T_0(\varphi(x_{i-1}))$ si $i > 1$ et $\varphi \in \mathrm{Aut}(F_\infty)$. Pour φ et $\psi \in \mathrm{Aut}(F_\infty)$, posons

$$\varphi * \psi = \varphi \circ T(\psi) \circ \alpha \circ T(\varphi^{-1}). \qquad (2.2)$$

On vérifie par un calcul direct que la loi de composition $*$ est autodistributive.

Nous affirmons que le système $(\mathrm{Aut}(F_\infty), *)$ est acyclique. En effet, soit E l'ensemble des éléments de F_∞ dont l'expression réduite en les générateurs $x_1, x_2, \ldots$ et leurs inverses se termine par x_1^{-1}. On observe d'abord que α, ainsi que tout élément $\varphi \in \mathrm{Aut}(F_\infty)$ qui est dans l'image du décalage T, préserve le sous-ensemble E. Supposons maintenant que le système $\mathrm{Aut}(F_\infty)$ possède un cycle. Cette hypothèse se traduit par une égalité de la forme $\varphi = \varphi \circ T(\varphi_0) \circ \alpha \circ T(\varphi_1) \circ \cdots \circ T(\varphi_{k-1}) \circ \alpha \circ T(\varphi_k)$, où $k \geq 1$ et $\varphi, \varphi_0, \ldots, \varphi_{k-1}, \varphi_k$ sont des automorphismes de F_∞. De manière équivalente, on a

$$\mathrm{id} = T(\varphi_0) \circ \alpha \circ T(\varphi_1) \circ \cdots \circ T(\varphi_{k-1}) \circ \alpha \circ T(\varphi_k) \tag{2.3}$$

où id représente l'automorphisme identité de F_∞. Appliquons les deux membres de (2.3) à l'élément x_1 de F_∞. À gauche, on obtient x_1 qui n'appartient pas au sous-ensemble E. Par contre, comme $(\alpha \circ T(\varphi_k))(x_1) = \alpha(x_1) = x_1 x_2 x_1^{-1}$ appartient à E, il résulte de l'observation ci-dessus que l'élément de F_∞ obtenu en appliquant le membre de droite de (2.3) à x_1 appartient à E. D'où une contradiction. $\square$

2.4. Retournement des mots

Nous avons vu au n° 2.1 comment définir une action du monoïde B_n^+ sur la puissance n-ième d'un système AD S et au n° 2.2 que cette action s'étend au groupe B_n tout entier si (et seulement si) S est un ensemble automorphe. Or, à la vue des remarques faites au n° 2.3 et des travaux dont nous parlerons au n° 4, la question se pose d'étendre l'action du monoïde B_n^+ à B_n lorsque S est acyclique. En développant une méthode spécifique, dite de retournement des mots, Dehornoy a réussi à définir une action partielle de B_n sur S^n lorsque S est un système AD *simplifiable* (à gauche), c'est-à-dire un système pour lequel les multiplications à gauche sont injectives.

On dit qu'un mot de tresse w *se transforme par retournement* (à gauche) en le mot w' si w' s'obtient en remplaçant dans w des facteurs du type $\sigma_i \sigma_j^{-1}$ par le mot vide si $i = j$, par $\sigma_j^{-1}\sigma_i$ si $|i - j| > 1$, et par $\sigma_j^{-1}\sigma_i^{-1}\sigma_j\sigma_i$ si $|i - j| = 1$. Par définition du retournement, w et w' représentent le même élément dans B_n.

Garside [Ga] avait démontré que toute tresse est le quotient de deux tresses représentées par des mots de tresse positifs. Dehornoy reprend les ingrédients de l'approche de Garside, mais travaille sur les mots de tresse plutôt que sur les tresses, ce qui lui permet d'établir dans [D97a] le résultat plus fin suivant.

LEMME 2.3. — *Tout mot de tresse se transforme par retournement en un mot de la forme $u^{-1}v$ où u et v sont des mots de tresse positifs.*

Remarquons qu'il n'est pas évident *a priori* que l'algorithme de retournement converge en un nombre fini d'étapes car la longueur des mots peut augmenter à chaque retournement ; on pourra observer ce phénomène sur le mot $\sigma_2\sigma_4\sigma_6\sigma_5^{-1}\sigma_3^{-1}\sigma_1^{-1}$. Signalons que des techniques analogues de retournement des mots permettent la construction de formes normales et de structures automatiques pour d'autres groupes dont les groupes d'Artin de groupe de Coxeter fini (voir [D98], [DP]).

Soit maintenant S un système AD simplifiable et n un entier ≥ 2. L'action à droite $(\vec{a}, w) \mapsto \vec{a}w$ du monoïde B_n^+ sur le produit S^n, donnée par la formule (2.1), se prolonge à certains éléments $(\vec{a}, w)$ de $S^n \times B_n$ de façon que les règles de coloriage restent vérifiées. C'est par exemple le cas de $(\vec{b}u, u^{-1})$ avec $b \in S^n$ et $u \in B_n^+$. Dans ce cas, on dira que l'élément $\vec{a}w$ est défini.

LEMME 2.4. — *Soit S un système AD simplifiable et $\vec{a} \in S^n$. Si le mot de tresse w se transforme par retournement en w' et si $\vec{a}w'$ est défini, alors $\vec{a}w$ est défini.*

Les lemmes 2.3 et 2.4 impliquent la proposition suivante selon laquelle on peut toujours trouver un ensemble de couleurs dans un système AD simplifiable pour colorier une tresse donnée en suivant la règle énoncée au n° 2.1.

PROPOSITION 2.5. — *Soit S un système AD simplifiable. Pour tout mot de tresse w, il existe $\vec{a} \in S^n$ tel que $\vec{a}w$ soit défini. Si w et w' représentent la même tresse et que $\vec{a}w$ et $\vec{a}w'$ soient définis, alors $\vec{a}w = \vec{a}w'$.*

Démonstration. — Contentons-nous d'établir la première assertion. D'après le lemme 2.3, w se transforme par retournement en $w' = u^{-1}v$ où u et v sont des mots positifs. Posons $\vec{a} = \vec{b}u$ où $\vec{b}$ est n'importe quel élément de S^n. Alors $\vec{a}w' = \vec{b}v$ est défini. On termine en appliquant le lemme 2.4.

Pour établir la seconde assertion, on utilise une version à droite du retournement des mots de tresse et l'hypothèse que le système est simplifiable. □

2.5. Démonstration du théorème 1.1

Supposons que nous disposions d'un système autodistributif S acyclique et monogène (c'est-à-dire engendré par un seul élément) tel que l'ordre $\prec$ associé (défini au n° 2.3) soit total. On vérifie que $a \prec b$ implique $c * a \prec c * b$, d'où l'on déduit que S est simplifiable.

Puisque S possède un ordre total, on peut munir chaque puissance S^n de S de l'ordre total lexicographique $\prec_\ell$ défini par $(a_1, \ldots, a_n) \prec_\ell (b_1, \ldots, b_n)$ s'il existe $i \geq 1$ tel que $a_j = b_j$ pour tout $j < i$ et $a_i \prec b_i$. Nous utilisons l'ordre total $\prec_\ell$ sur S^n pour définir une partie « positive » P de B_n comme suit. Si w est un mot de tresse, alors, d'après la proposition 2.5, il existe $\vec{a} \in S^n$ tel que $\vec{a}w$ soit défini. Nous dirons que la tresse représentée par w est dans P si $\vec{a} \prec_\ell \vec{a}w$.

PROPOSITION 2.6. — *Une tresse est dans P si et seulement si elle est σ-positive. La relation $\beta < \beta'$ définie sur B_n par $\beta^{-1}\beta' \in P$ est une relation d'ordre total invariant par multiplication à gauche, qui coïncide avec l'ordre de Dehornoy du n° 1.3.*

Démonstration. — Montrons que, si un mot de tresse est σ-positif, alors il représente une tresse du sous-ensemble P défini ci-dessus. Un mot σ_i-positif w est de la forme $w = w_0\sigma_i w_1\sigma_i \dots \sigma_i w_k$ où $w_0, w_1, \dots, w_k$ sont des mots dans les lettres σ_j et σ_j^{-1} avec $j > i$. Soit $\vec{a} = (a_1, \dots, a_n) \in S^n$ tel que $\vec{a}w = (b_1, \dots, b_n) \in S^n$ soit défini. La formule (2.1) implique que $b_j = a_j$ pour tout $j < i$ et que b_i est de la forme $b_i = (\cdots((a_1 * c_1) * c_2) * \cdots) * c_k$ où $c_1, c_2, \dots, c_k \in S$ et k est le nombre d'occurrences de σ_i dans w. Par définition de l'ordre $\prec$ de S, il en résulte que $a_i \prec b_i$. Par conséquent, $\vec{a} \prec_\ell \vec{a}w$ et $w \in P$.

La réciproque, à savoir qu'une tresse dans P est σ-positive, utilise la notion de tresse spéciale que nous introduirons au n° 3.3 et sera démontrée au n° 3.4.

Une fois l'équivalence précédente établie, il est clair que la relation $<$ est bien définie et que c'est l'ordre total du n° 1.3, dont l'existence est ainsi démontrée. □

3. SYSTÈME AUTODISTRIBUTIF LIBRE ET TRESSES SPÉCIALES

Nous étudions maintenant deux systèmes autodistributifs particuliers. Le premier est libre et fournit le type de systèmes nécessaires à la démonstration du théorème 1.1, telle qu'elle est donnée au n° 2.5. Le second est constitué de tresses.

3.1. Le système autodistributif libre sur un générateur

Il est caractérisé par la propriété universelle suivante.

PROPOSITION 3.1. — *Il existe un système autodistributif D_1 engendré par un singleton $\{x\}$ tel que, pour tout système autodistributif S et tout élément s de S, il existe un unique morphisme $f : D_1 \to S$ de systèmes AD vérifiant $f(x) = s$. Le système D_1 est unique à isomorphisme près.*

Démonstration. — L'unicité est claire. Pour l'existence, on considère le magma libre M sur un générateur x (voir [Bo], I, § 7, n° 1). Les éléments de M sont des parenthésages complets de puissances strictement positives de la lettre x. Notons $*$ la loi de composition du magma. Sur M on considère la relation d'équivalence $\equiv_{\mathrm{AD}}$ invariante par composition et engendrée par les paires $(t_1 * (t_2 * t_3), (t_1 * t_2) * (t_1 * t_3))$. On définit D_1 comme l'ensemble des classes d'équivalence $M/{\equiv_{\mathrm{AD}}}$. On vérifie aisément que D_1 est un système AD satisfaisant à la propriété universelle ci-dessus. □

Rappelons la relation binaire $\prec$ introduite au n° 2.3. Le théorème suivant, établi par Dehornoy dans [D94], fournit l'exemple qui nous manquait au n° 2.5 pour construire un ordre total sur les groupes de tresses.

THÉORÈME 3.2. — *Le système autodistributif D_1 est acyclique et la relation $\prec$ le munit d'un ordre total.*

Le système libre D_1 est un ensemble de classes d'équivalence de parenthésages pour la relation $\equiv_{\mathrm{AD}}$ définie dans la preuve de la proposition 3.1. Avant de démontrer le théorème 3.2, établissons quelques propriétés de ces parenthésages.

Pour tout entier $n \geq 1$ définissons les puissances droites $x^{[k]}$ de x par $x^{[1]} = x$ et $x^{[k]} = x * x^{[k-1]}$ si $k > 1$. La relation d'équivalence $\equiv_{\mathrm{AD}}$ vérifie la propriété fondamentale d'absorption suivante.

LEMME 3.3. — *Pour tout parenthésage t, on a $x^{[k+1]} \equiv_{\mathrm{AD}} t * x^{[k]}$ pour tout k suffisamment grand.*

Démonstration. — On procède par récurrence sur la longueur de t. Si $t = x$, alors $x^{[k+1]} = t * x^{[k]}$ par définition des puissances droites. Si $t = t_1 * t_2$, et si $x^{[k+1]} \equiv_{\mathrm{AD}} t_1 * x^{[k]}$ et $x^{[k+1]} \equiv_{\mathrm{AD}} t_2 * x^{[k]}$ pour tout k suffisamment grand, alors

$$\begin{aligned} x^{[k+2]} &\equiv_{\mathrm{AD}} t_1 * x^{[k+1]} \equiv_{\mathrm{AD}} t_1 * (t_2 * x^{[k]}) \\ &\equiv_{\mathrm{AD}} (t_1 * t_2) * (t_1 * x^{[k]}) \equiv_{\mathrm{AD}} t * x^{[k+1]}. \end{aligned}$$

La troisième équivalence est une conséquence de la définition de $\equiv_{\mathrm{AD}}$. □

Nous dirons qu'un parenthésage t' est une *expansion* d'un parenthésage t si on l'obtient en remplaçant à partir de t des sous-expressions de la forme $t_1 * (t_2 * t_3)$ par $(t_1 * t_2) * (t_1 * t_3)$, c'est-à-dire en appliquant l'identité (AD), exclusivement de la gauche vers la droite. Il est clair que $t' \equiv_{\mathrm{AD}} t$ si t' est une expansion de t. On a la réciproque suivante.

LEMME 3.4. — *Si t_1 et t_2 sont des parenthésages tels que $t_1 \equiv_{\mathrm{AD}} t_2$, alors ils ont une expansion commune.*

Définissons maintenant les sous-termes gauches $G_k(t)$ d'un parenthésage t. Si $t = x$, seul $G_0(t) = t$ est défini. En général, posons $G_0(t) = t$ et, si $t = t_1 * t_2$ et $k \geq 1$, posons $G_k(t) = G_{k-1}(t_1)$ si ce dernier est défini. On a $G_\ell(G_k(t)) = G_{k+\ell}(t)$ lorsque ces expressions sont définies. Notons que l'élément de D_1 représenté par un sous-terme gauche $G_k(t)$, où $k \geq 1$, est plus petit pour l'ordre $\prec$ que l'élément représenté par t.

LEMME 3.5. — *Si t' est une expansion de t, alors, pour tout k tel que $G_k(t)$ soit défini, il existe un entier $k' \geq k$ tel que $G_{k'}(t')$ soit une expansion de $G_k(t)$.*

3.2. Démonstration du théorème 3.2

Si le système libre D_1 n'était pas acyclique, il en serait de même de tout système AD, ce qui contredirait la proposition 2.2. Montrons maintenant que deux éléments quelconques de D_1 sont comparables pour l'ordre $\prec$. Plus précisément, soient y et z deux éléments de D_1 que nous représentons respectivement par des parenthésages t_1 et t_2. Il s'agit d'établir que soit $y = z$, soit $y \prec z$, soit $z \prec y$. D'après le lemme 3.3, il existe un entier $k \geq 1$ tel que $t_1 * x^{[k]} \equiv_{\mathrm{AD}} t_2 * x^{[k]}$. Le lemme 3.4 nous permet d'affirmer que les parenthésages $t_1 * x^{[k]}$ et $t_2 * x^{[k]}$ ont une expansion commune t'.

Le lemme 3.5 implique l'existence d'entiers naturels p et q tels que $G_p(t')$ soit une expansion de $G_1(t_1 * x^{[k]}) = t_1$ et que $G_q(t')$ soit une expansion de $G_1(t_2 * x^{[k]}) = t_2$. Les parenthésages $G_p(t')$ et $G_q(t')$ représentent respectivement les éléments y et z de D_1. Trois cas se présentent à nous :

(i) si $p = q$, alors $G_p(t') = G_q(t')$, et donc $y = z$;

(ii) si $p > q$, alors $G_p(t') = G_{p-q}(G_q(t'))$, et donc $y \prec z$.

(iii) si $p < q$, alors $G_q(t') = G_{q-p}(G_p(t'))$, et donc $z \prec y$. □

3.3. Une loi autodistributive sur les tresses

Pour achever la démonstration de la proposition 2.6 du n° 2.5, nous allons introduire ce que Dehornoy appelle des tresses spéciales. Dans ce but, considérons la limite inductive B_∞ des groupes de tresses B_n pour les inclusions évidentes ; une présentation du groupe B_∞ est obtenue en prenant une famille infinie de générateurs σ_i indexée par les entiers ≥ 1 et les relations (1.1) et (1.2). Soit T l'endomorphisme injectif de B_∞ défini par $T(\sigma_i) = \sigma_{i+1}$ pour tout $i \geq 1$. Posons

$$\alpha * \beta = \alpha\, T(\beta)\, \sigma_1\, T(\alpha^{-1}). \tag{3.1}$$

Proposition 3.6. — *La formule (3.1) munit B_∞ d'une structure de système AD simplifiable et acyclique.*

Démonstration. — Nous laissons au lecteur le soin de vérifier que B_∞ est un système AD simplifiable. Pour établir son acyclicité, on observe que l'homomorphisme de groupes $B_\infty \to \mathrm{Aut}(F_\infty)$ qui est la limite inductive des homomorphismes $B_n \to \mathrm{Aut}(F_n)$ mentionnés au n° 2.2 (a) transporte la loi (3.1) sur la loi (2.2). Comme le système $(\mathrm{Aut}(F_\infty), *)$ est acyclique, il en est de même de $(B_\infty, *)$. □

Dans le système autodistributif B_∞, considérons le sous-système B_{sp} engendré par la tresse unité 1. Les éléments de B_{sp} sont appelés *tresses spéciales*. Exemples de tresses spéciales : $1{*}1 = \sigma_1$, $1{*}(1{*}1) = \sigma_2\sigma_1$, $(1{*}1){*}1 = \sigma_1^2\sigma_2^{-1}$, $1{*}(1{*}(1{*}1)) = \sigma_3\sigma_2\sigma_1$.

Proposition 3.7. — *Le système autodistributif B_{sp} est acyclique et la relation $\prec$ le munit d'un ordre total.*

Démonstration. — Le système B_{sp} hérite de l'acyclicité de B_∞. La relation $\prec$ est donc une relation d'ordre sur B_{sp}. Il reste à montrer que deux éléments quelconques de B_{sp} sont comparables pour $\prec$. Comme B_{sp} est engendré par le singleton $\{1\}$, c'est un quotient du système libre D_1. Par le théorème 3.2, deux éléments quelconques de D_1 sont comparables pour $\prec$, donc, par projection, il en est de même dans B_{sp}. □

L'argument précédent montre que, dans tout système AD acyclique et monogène, la relation $\prec$ est une relation d'ordre total. En fait, comme l'a montré Laver [L92], tout système AD acyclique monogène est isomorphe à D_1. En particulier, $B_{\mathrm{sp}} \cong D_1$.

L'intérêt des tresses spéciales vient de la possibilité d'exprimer toute tresse en termes de produit de tresses spéciales.

LEMME 3.8. — *Pour tout $\beta \in B_n$, il existe $2n$ tresses spéciales α_1, …, α_n, γ_1,…, γ_n telles que*

$$\beta = T^{n-1}(\alpha_n)^{-1} \dots T(\alpha_2)^{-1}\, \alpha_1^{-1}\, \gamma_1\, T(\gamma_2) \dots T^{n-1}(\gamma_n). \tag{3.2}$$

Démonstration. — Comme B_{sp} est un système AD simplifiable, nous pouvons appliquer la proposition 2.5, qui fournit une action partielle de B_n sur B_{sp}^n. En particulier, pour toute tresse $\beta \in B_n$, il existe des n-uples $(\alpha_1, \dots, \alpha_n)$ et $(\gamma_1, \dots, \gamma_n) \in B_{\mathrm{sp}}^n$ tels que $(\alpha_1, \dots, \alpha_n)\, \beta = (\gamma_1, \dots, \gamma_n)$. On vérifie alors facilement l'égalité des produits

$$\alpha_1\, T(\alpha_2) \dots T^{n-1}(\alpha_n)\, \beta = \gamma_1\, T(\gamma_2) \dots T^{n-1}(\gamma_n)$$

dans le groupe B_∞, d'où la formule (3.2). □

3.4. Fin de la preuve de la proposition 2.6

Supposons que α et γ soient deux tresses spéciales. On a ou bien $\alpha \prec \gamma$, ou bien $\alpha = \gamma$, ou bien $\gamma \prec \alpha$. Revenant à la définition de la relation $\prec$ et de la loi (3.1), on voit que les trois relations précédentes entraînent respectivement que la tresse $\alpha^{-1}\gamma$ est σ_1-positive, égale à 1, ou σ_1-négative. Si maintenant β est une tresse quelconque, on considère une décomposition de type (3.2) de β : si i est le plus petit indice tel que $\alpha_i \neq \gamma_i$, alors β est σ_i-positive ou σ_i-négative suivant que $\alpha_i \prec \gamma_i$ ou $\gamma_i \prec \alpha_i$, ce qui achève la preuve de la proposition 2.6.

4. RAPPORT AVEC LA THÉORIE DES ENSEMBLES

Dans ce numéro nous expliquons comment des développements en théorie des ensembles ont mené à la construction d'un ordre total sur les tresses.

4.1. Grands cardinaux et extensions de ZF

On sait depuis le théorème d'incomplétude de Gödel que le système ZF d'axiomes de Zermelo-Fraenkel est incomplet, c'est-à-dire qu'il existe des énoncés qui sont indécidables dans ZF. C'est le cas de l'axiome du choix et de l'hypothèse du continu. Dès lors, le but principal de la théorie des ensembles consiste à étudier, voire classifier, les extensions du système ZF dans lesquels les énoncés ouverts deviennent décidables. Gödel a émis l'idée que ces extensions pourraient être classifiées à l'aide d'axiomes de grands cardinaux comme celui qui postule l'existence de $\aleph_1$. L'un des premiers axiomes considérés a été celui qui affirme l'existence d'un cardinal inaccessible.[2] Non seulement cet axiome est indémontrable dans ZF, mais encore — à la différence de l'axiome du choix ou de l'hypothèse du continu — sa cohérence, c'est-à-dire le fait qu'il ne soit pas contradictoire, ne peut y être établie.

[2] α est inaccessible si tout ensemble de parties d'un ensemble de cardinalité $< \alpha$ et toute union d'ensembles de cardinalité $< \alpha$ sont de cardinalité $< \alpha$.

Dans les années 1960 une hiérarchie totalement ordonnée d'axiomes de grands cardinaux est apparue en théorie des ensembles, s'imposant comme les bonnes extensions de ZF (voir la monographie [Kan]). Nous allons nous intéresser au n° 4.2 à un axiome particulier de cette hiérarchie.

4.2. Plongements élémentaires

Un ensemble infini se caractérise par l'existence d'une injection non bijective. Une telle injection ne préserve pas nécessairement toutes les structures de l'ensemble considéré : ainsi l'application $n \mapsto n+1$ de l'ensemble des entiers naturels dans lui-même préserve l'ordre naturel, mais non l'addition. Pour définir des notions d'infini plus fortes, il est naturel de postuler l'existence d'une injection non bijective d'un ensemble dans lui-même, préservant *toutes* les propriétés définissables par les opérations ensemblistes de base. Appelons *plongement élémentaire* une telle injection et *autosimilaire* un ensemble pour lequel il existe un plongement élémentaire. On démontre assez facilement qu'un ensemble autosimilaire est forcément très grand : son cardinal doit être inaccessible, ce qui entraîne que l'existence d'un tel ensemble ne peut être démontrée dans ZF.

Il est usuel en théorie des ensembles de considérer des ensembles particuliers appelés rangs, qui ont la propriété technique que toute fonction d'un rang R dans lui-même est (essentiellement) un élément de R. Les axiomes affirmant l'existence de plongements élémentaires d'un rang dans un autre ou d'un rang dans lui-même ont été au cœur de la théorie des ensembles dans les années 1980. Un des résultats majeurs obtenus à l'aide de tels axiomes a été la démonstration en 1984 par Martin et Steel de la propriété de détermination pour les ensembles projectifs de réels, un énoncé technique qui, d'une certaine façon, décrit complètement la structure fine de la droite réelle (voir [D89]). L'axiome qui nous intéresse est le suivant.

AXIOME A. — *Il existe un rang autosimilaire.*

4.3. Le système autodistributif d'un rang autosimilaire

Notons E_R l'ensemble des plongements élémentaires du rang autosimilaire R dont nous venons de postuler l'existence. L'ensemble E_R est trivialement muni d'une structure de monoïde pour la composition des plongements. Mais, en vertu de l'axiome A et des propriétés spécifiques des rangs, il est aussi muni d'une autre opération binaire qu'on peut appeler *l'application.* En effet, nous avons dit qu'une fonction définie sur un rang R peut être vue comme un élément de R. Si donc i et j sont des plongements élémentaires de R, i s'appliquant par hypothèse aux éléments de R et j pouvant être vu comme un tel élément, nous pouvons appliquer i à j. On montre que le nouvel objet $i(j)$ ainsi obtenu, héritant de toutes les propriétés définissables de j, est aussi un plongement élémentaire de R. La loi de composition $(i,j) \mapsto i(j)$ ainsi définie sur E_R est autodistributive. L'argument est le suivant : si y est l'image de x par l'application f, alors $i(y)$ est l'image de $i(x)$ par $i(f)$ chaque fois que i est un plongement élémentaire ; en d'autres termes, on a $i(f(x)) = i(f)(i(x))$, identité valide, en particulier, lorsque f et x sont des plongements élémentaires.

Laver [L92] a établi en 1989 le résultat suivant.

PROPOSITION 4.1. — *Si j est un plongement élémentaire du rang R, alors le sous-système autodistributif de E_R engendré par j est acyclique.*

Avec ce résultat, Laver a construit le premier exemple d'un système AD acyclique, mais un exemple dont l'existence dépend d'un axiome indémontrable de théorie des ensembles. La situation ainsi créée était étrange : dans des travaux indépendants, Dehornoy et Laver avaient, à cette époque, partiellement résolu le problème des mots de l'identité (AD), en construisant, sous l'hypothèse de l'existence d'un système AD acyclique, un algorithme permettant de reconnaître si deux parenthésages donnés sont ou non équivalents modulo $\equiv_{\mathrm{AD}}$. La proposition 4.1 prouvait donc la décidabilité du problème des mots en question à partir de l'axiome indémontrable A, établissant un lien paradoxal entre le problème des mots de (AD), qui ne met en jeu que des manipulations de mots finis, et un rang autosimilaire, qui est un objet gigantesque.

Une solution a été proposée vers la fin 1991 par Dehornoy qui, à partir d'une étude fine de l'autodistributivité dont nous donnerons une idée au numéro suivant, a montré par une démonstration directe qu'un système AD libre est acyclique, et a déduit de son analyse les liens avec les tresses (voir [D92a], [D92b], [D94]).

4.4. Le groupe de l'autodistributivité

L'étude algébrique des systèmes AD libres, en particulier celle du système D_1 à un générateur, ont fait l'objet de nombreux travaux de Laver et Dehornoy. Nous ne donnerons ici qu'un aperçu de l'approche de Dehornoy et de la façon dont elle mène aux tresses.

Pour étudier D_1, il est commode de représenter les parenthésages par des arbres binaires avec racine. On représente la variable x par l'arbre réduit à sa racine ; si t_1 et t_2 sont des parenthésages, on représente le parenthésage $t_1 * t_2$ par l'arbre obtenu en greffant l'arbre associé à t_1 à gauche d'une nouvelle racine, et l'arbre associé à t_2 à droite. Tout sommet d'un arbre binaire a une *adresse* qui est une suite finie de 0 et de 1 : l'adresse de la racine est la suite vide $\varnothing$, et, en s'éloignant de la racine, on passe de l'adresse d'un sommet à celui de son « successeur » gauche en ajoutant 0, et à celle de son « successeur » droit en ajoutant 1.

Au n$^{\mathrm{o}}$ 3.1 nous avons défini la notion d'expansion d'un parenthésage. Cette définition se transpose facilement aux arbres binaires. Nous pouvons décrire plus finement la situation en considérant le cas où l'identité (AD) est appliquée une seule fois, toujours de la gauche vers la droite, et en prenant en compte l'adresse où elle l'est dans l'arbre correspondant. Nous noterons ∇_α l'opération consistant à appliquer (AD) à l'adresse α (l'opération ∇_α n'est pas partout définie sur M). Il est facile de vérifier les relations suivantes :

$$\nabla_{\alpha 0\beta}\,\nabla_{\alpha 1\gamma} = \nabla_{\alpha 1\gamma}\,\nabla_{\alpha 0\beta}, \tag{4.1}$$

$$\nabla_{\alpha 0\beta}\,\nabla_{\alpha} = \nabla_{\alpha}\,\nabla_{\alpha 10\beta}\,\nabla_{\alpha 00\beta}, \tag{4.2}$$

$$\nabla_{\alpha 10\beta}\,\nabla_{\alpha} = \nabla_{\alpha}\,\nabla_{\alpha 01\beta}, \tag{4.3}$$

$$\nabla_{\alpha 11\beta}\,\nabla_{\alpha} = \nabla_{\alpha}\,\nabla_{\alpha 11\beta}, \tag{4.4}$$

$$\nabla_{\alpha 1}\,\nabla_{\alpha}\,\nabla_{\alpha 1}\,\nabla_{\alpha 0} = \nabla_{\alpha}\,\nabla_{\alpha 1}\,\nabla_{\alpha}, \tag{4.5}$$

où α, β et γ sont des adresses quelconques.

Dans [D94] Dehornoy introduit alors le groupe G_{AD} engendré par les générateurs ∇_α (indexés par les adresses α) et les relations (4.1)–(4.5), et montre que ce groupe opère sur le magma libre M de telle sorte que les orbites de l'action sont exactement les classes d'équivalence d'éléments de M pour la relation $\equiv_{\mathrm{AD}}$. Il établit que le système D_1 est acyclique en étudiant le groupe G_{AD} à l'aide d'une méthode de retournement des mots semblable à celle décrite au n$^{\mathrm{o}}$ 2.4 pour les tresses, et en construisant un préordre sur G_{AD}. Une étape essentielle consiste à reprendre la démonstration du lemme d'absorption 3.3, qui se lit, dans le groupe G_{AD}, comme l'existence d'une loi binaire $*$ donnée par la formule

$$a * b = a\,T(b)\,\nabla_{\varnothing}\,T(a)^{-1}, \tag{4.6}$$

où T est cette fois-ci l'endomorphisme qui envoie ∇_α sur $\nabla_{1\alpha}$ pour tout α.

Le lien avec les tresses résulte de l'observation suivante : l'application π définie par $\pi(\nabla_\alpha) = 1$ si l'adresse α contient au moins un 0, et $\pi(\nabla_\alpha) = \sigma_{i+1}$ si α est une suite constituée de i fois le chiffre 1, induit un homomorphisme surjectif de groupes de G_{AD} sur le groupe de tresses B_∞. En outre, π transporte la loi (4.6) de G_{AD} sur la loi autodistributive (3.1) de B_∞.

Signalons ici la parenté du groupe G_{AD} avec le groupe F de Thompson, qui en est l'exacte contrepartie lorsqu'on remplace l'identité AD par l'associativité. Dans cette correspondance, la relation (4.5) est l'analogue de la relation du pentagone de Mac Lane–Stasheff. On trouvera un exposé de synthèse sur le groupe F dans [CPF].

4.5. Systèmes autodistributifs finis

En étudiant les quotients finis du système AD associé à un plongement élémentaire, Laver [L95] a montré que, pour tout entier $n \geq 1$, il existe une unique structure autodistributive A_n sur l'ensemble fini $\{1, 2, \ldots, 2^n\}$ tel que $k * 1 \equiv k + 1$ modulo 2^n (on trouvera plus bas les tables de multiplication de A_1, A_2 et A_3). Les ensembles A_n forment un système projectif de systèmes AD et jouent un rôle-clé parmi les systèmes autodistributifs. En particulier, Drápal [Dr] a montré que tout système AD monogène fini se construit simplement à partir de systèmes A_n. Chaque ligne de la table de multiplication de A_n est la répétition périodique d'une suite croissante de nombres allant jusqu'à 2^n. De par l'existence du système projectif mentionné ci-dessus, la période p_n de la première ligne de la table de A_n est une suite croissante au sens large. En traduisant une propriété non triviale des ordinaux et des plongements élémentaires, Laver [L95] a démontré la proposition suivante, qui implique notamment que le sous-système AD engendré par 1 dans la limite projective des A_n est libre.

A_1	1	2
1	2	2
2	1	2

A_2	1	2	3	4
1	2	4	2	4
2	3	4	3	4
3	4	4	4	4
4	1	2	3	4

A_3	1	2	3	4	5	6	7	8
1	2	4	6	8	2	4	6	8
2	3	4	7	8	3	4	7	8
3	4	8	4	8	4	8	4	8
4	5	6	7	8	5	6	7	8
5	6	8	6	8	6	8	6	8
6	7	8	7	8	7	8	7	8
7	8	8	8	8	8	8	8	8
8	1	2	3	4	5	6	7	8

TABLES DE MULTIPLICATION DE A_1, A_2 ET A_3.

PROPOSITION 4.2. — *S'il existe un rang autosimilaire, alors la suite* $(p_n)_n$ *tend vers l'infini.*

Pour l'instant aucune tentative pour démontrer ce résultat finitiste sans faire appel à la théorie des ensembles n'a abouti. On sait seulement que la première valeur de n pour laquelle p_n dépasse 32 est gigantesque [Do], [DJ]. Les systèmes A_n sont des objets d'une très grande complexité combinatoire, et restent encore très mal connus. Une question évidente, mais complètement ouverte, est de se demander s'ils sont susceptibles d'applications topologiques.

5. APPROCHES GÉOMÉTRIQUES

Récemment, deux constructions géométriques de l'ordre de Dehornoy ont été proposées par des topologues.

5.1. La définition de Fenn, Greene, Rolfsen, Rourke et Wiest

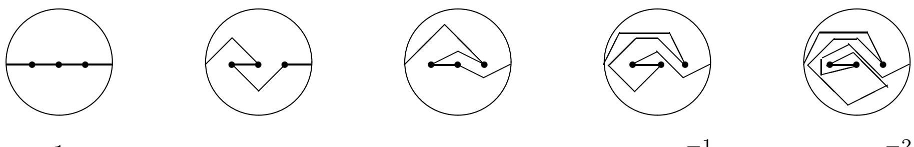
Notons D_n le disque unité fermé du corps $\mathbf{C}$ des nombres complexes, muni de n points marqués $P_1, \ldots, P_n$ intérieurs au disque et situés sur le diamètre réel $\Gamma = [-1, 1]$. Ordonnons les points marqués de sorte que le diamètre Γ soit la juxtaposition des $n+1$ segments $[P_0, P_1]$, $[P_1, P_2]$, …, $[P_{n-1}, P_n]$, $[P_n, P_{n+1}]$ que nous numéroterons de 1 à $n+1$ (pour simplifier, on a posé $P_0 = -1$ et $P_{n+1} = 1$).

On sait (voir [BZ], chap. 10) que le groupe de tresses B_n est isomorphe au groupe des classes d'isotopie des homéomorphismes de D_n qui fixent le bord du disque point par point et permutent les points marqués. Considérons l'image de Γ par un tel homéomorphisme φ : la courbe simple $\varphi(\Gamma)$ sépare D_n en deux composantes connexes et est l'union de $n+1$ « petites courbes » $\varphi([P_i, P_{i+1}])$ dont les extrémités sont les points $P_0, P_1, \ldots, P_{n+1}$. On appelle $\varphi(\Gamma)$ le *diagramme de courbes* de φ. Suivant un procédé bien connu et utilisé par exemple dans [Mo], on rectifie $\varphi(\Gamma)$ en raccourcissant les petites courbes de manière à faire coïncider le maximum d'entre elles avec des segments $[P_i, P_{i+1}]$ et à éliminer les intersections inutiles avec Γ. Il est montré dans [F+] que deux tels homéomorphismes de D_n représentent la même tresse dans B_n si et seulement si leurs diagrammes de courbes rectifiés sont isotopes *via* une isotopie qui fixe les petites courbes coïncidant avec des segments $[P_i, P_{i+1}]$. Pour $i = 1, \ldots, n-1$, on peut donc considérer l'ensemble Π_i des éléments de B_n correspondant aux diagrammes de courbes rectifiés pour lesquels $\varphi([P_{j-1}, P_j]) = [P_{j-1}, P_j]$ pour tout $j < i$ et $\varphi([P_{i-1}, P_i])$ monte vers le demi-plan supérieur de $\mathbf{C}$. On pose $\Pi = \bigcup_{i=1}^{n-1} \Pi_i$.

Fenn *et al.* [F+] ont établi en 1997 que la relation $\beta <_\Pi \beta'$ définie par $\beta^{-1}\beta' \in \Pi$ est une relation d'ordre total invariant par multiplication à gauche sur le groupe B_n. Ils ont aussi montré qu'une tresse est dans Π_i si et seulement si elle est σ_i-positive, ce qui entraîne que l'ordre $<_\Pi$ coïncide avec celui défini au n° 1.3. On obtient ainsi une définition géométrique de l'ordre de Dehornoy.

Utilisant les mêmes techniques, Rourke et Wiest [RW] (voir aussi [SW]) ont montré que le groupe des difféotopies (« mapping class group » en anglais) d'une surface compacte (non nécessairement orientable) à bord non vide est un groupe ordonnable. Ils ont aussi construit un algorithme qui décide en temps quadratique si un élément différent de l'identité est positif ou négatif pour l'ordre construit.

La figure 3 montre les diagrammes de courbes de cinq tresses. On voit, par exemple, sur ces diagrammes que $\sigma_1\sigma_2\sigma_1^{-1}$ et $\sigma_1\sigma_2\sigma_1^{-2}$ sont σ_1-positifs.



1 $\quad$ σ_1 $\quad$ $\sigma_1\sigma_2$ $\quad$ $\sigma_1\sigma_2\sigma_1^{-1}$ $\quad$ $\sigma_1\sigma_2\sigma_1^{-2}$

FIGURE 3.

5.2. L'approche à la Nielsen

En septembre 1998, W. P. Thurston [Th] a expliqué comment les travaux classiques de Nielsen sur les surfaces permettaient de construire un ordre invariant sur B_n à partir de la proposition suivante.

PROPOSITION 5.1. — *Le groupe B_n opère sur la droite réelle par des homéomorphismes croissants de sorte que l'action soit libre en au moins un point réel.*

Avant d'établir la proposition 5.1, montrons comment on en déduit un ordre total sur B_n. Soit $x \in \mathbf{R}$ un point pour lequel l'action de la proposition 5.1 est libre (c'est-à-dire un point dont le stabilisateur est trivial). Pour β, $\beta' \in B_n$, posons $\beta <_x \beta'$ si $\beta(x) < \beta'(x)$ pour l'ordre naturel de $\mathbf{R}$. La liberté de l'action en x implique que la relation $<_x$ est un ordre, et c'est alors clairement un ordre total. Comme B_n opère par des homéomorphismes croissants, l'ordre $<_x$ est invariant par multiplication à gauche.

Short et Wiest [SW] ont classifié les ordres $<_x$ à conjugaison près (il y en a une infinité non dénombrable) et montré que l'on retrouve ainsi l'ordre de Dehornoy.

5.3. Démonstration de la proposition 5.1

Soit S une surface orientée fermée de genre un marquée en n points $P_1, \ldots, P_n$ et C une courbe simple séparant S en une surface A de genre un et une surface de genre zéro contenant les points marqués. Le groupe de tresses B_n s'identifie aussi au groupe des classes d'isotopie de difféomorphismes de S préservant l'orientation, induisant l'identité sur A et permutant les points marqués.

Munissons $S - \{P_1, \ldots, P_n\}$ d'une métrique hyperbolique complète pour laquelle C est une géodésique et les points marqués sont des « cusps ». Fixons un point-base x_0 sur C et le revêtement universel $(D, 0)$ de $(S - \{P_1, \ldots, P_n\}, x_0)$. La métrique hyperbolique identifie $(D, 0)$ avec le disque unité de $\mathbf{C}$. Tout difféomorphisme φ de S qui fixe x_0 et permute les points marqués se relève en un difféomorphisme $\widetilde{\varphi}$ de D fixant 0. Dans [Ni], § 10, Nielsen a montré que $\widetilde{\varphi}$ se prolonge en un homéomorphisme $\partial\widetilde{\varphi}$ du bord ∂D de D, c'est-à-dire du cercle unité, que $\partial\widetilde{\varphi}$ conserve l'orientation si φ la conserve, et que $\partial\widetilde{\varphi}$ ne dépend que de la classe d'isotopie de φ. On obtient ainsi une action de B_n sur le cercle, préservant l'orientation. Pour obtenir une action sur $\mathbf{R}$, il suffit de remarquer que l'action précédente fixe un point, par exemple l'une des extrémités d'un relèvement de C passant par 0.

L'existence d'un point de $\mathbf{R}$ pour lequel l'action est libre se déduit également d'un résultat de Nielsen qui dit que, si φ n'est pas isotope à l'identité, alors l'ensemble des points fixes de $\partial\widetilde{\varphi}$ est un fermé d'intérieur vide (voir [Ni], § 14). Comme une union dénombrable de fermés d'intérieur vide est d'intérieur vide (théorème de Baire), il résulte que l'action de B_n sur $\mathbf{R}$ est libre sur une partie G_δ-dense, donc en au moins un point. □

Remarque : La lecture de [Ni] montre que Nielsen avait déjà établi l'existence d'une action de B_n sur le cercle, préservant l'ordre cyclique, fixant un point et ayant une orbite libre.

RÉFÉRENCES

[Ad] S. I. ADYAN – *Fragments of the word Δ in a braid group*, Mat. Zametki **36** (1984), 25–34 (= Math. Notes **36** (1984), 505–510).

[A26] E. ARTIN – *Theorie der Zöpfe*, Abh. Math. Sem. Univ. Hamburg **4** (1926), 47–72.

[A47] E. ARTIN – *Theory of braids*, Ann. of Math. **48** (1947), 101–126.

[Bo] N. BOURBAKI – *Algèbre*, chapitres 1 à 3, Hermann, Paris, 1970.

[B+] *Braids*, Actes Conf. Santa Cruz (1986) publiés par Joan S. Birman et Anatoly Libgober, Contemp. Math. **78**, Amer. Math. Soc., Providence, RI, 1988.

[Br] E. BRIESKORN – *Automorphic sets and braids and singularities*, publié dans [B+], 45–115.

[Bu] S. BURCKEL – *The wellordering on positive braids*, J. Pure Appl. Algebra **120** (1997), 1–17.

[BZ] G. BURDE, H. ZIESCHANG – *Knots*, Studies in Math. **5**, Walter de Gruyter, Berlin-New York, 1985.

[CFP] J. W. CANNON, W. J. FLOYD, W. R. PARRY – *Introductory notes on Richard Thompson's groups*, Enseign. Math. (2) **42** (1996), 215–256.

[Ca] P. CARTIER – *Développements récents sur les groupes de tresses ; applications à la topologie et à l'algèbre*, Séminaire Bourbaki, exposé **716** (novembre 1989), Astérisque, vol. 189–190, Soc. Math. France, Paris (1991), 17–67.

[D89] P. DEHORNOY – *La détermination projective (d'après Martin, Steel et Woodin)*, Séminaire Bourbaki, exposé **710** (juin 1989), Astérisque, vol. 177–178, Soc. Math. France, Paris (1989), 261–276.

[D92a] P. DEHORNOY – *Preuve de la conjecture d'irréflexivité pour les structures distributives libres*, C. R. Acad. Sci. Paris Sér. I Math. **314** (1992), 333–336.

[D92b] P. DEHORNOY – *Deux propriétés des groupes de tresses*, C. R. Acad. Sci. Paris Sér. I Math. **315** (1992), 633–638.

[D94] P. DEHORNOY – *Braid groups and left distributive operations*, Trans. Amer. Math. Soc. **345** (1994), 115–150.

[D95] P. DEHORNOY – *From large cardinals to braids via distributive algebra*, J. Knot Theory Ramifications **4** (1995), 33–79.

[D96] P. DEHORNOY – *Une autre application de la théorie des ensembles*, Gaz. Math. **69** (1996), 3–20.

[D97a] P. DEHORNOY – *Groups with a complemented presentation*, J. Pure Appl. Algebra **116** (1997), 115–137.

[D97b] P. DEHORNOY – *A fast method for comparing braids*, Adv. Math. **125** (1997), 200–235.

[D98] P. DEHORNOY – *Gaussian groups are torsion free*, J. of Algebra **210** (1998), 291–297.

[D00] P. DEHORNOY – *Braids and self-distributivity*, Progress in Math. **192**, Birkhäuser, Basel, Boston, Berlin, 2000.

[DP] P. DEHORNOY, L.P. PARIS – *Gaussian groups and Garside groups, two generalizations of Artin groups*, Proc. London Math. Soc. **79** (1999), 569–604.

[Do] R. DOUGHERTY – *Critical points in an algebra of elementary embeddings*, Ann. Pure Appl. Logic **65** (1993), 211–241.

[DJ] R. DOUGHERTY, T. JECH – *Finite left-distributive algebras and embedding algebras*, Adv. Math. **130** (1997), 201–241.

[Dr] A. DRÁPAL – *Finite left distributive algebras with one generator*, J. Pure Appl. Algebra **121** (1997), 233–251.

[EM] E.A. ELRIFAI, H.R. MORTON – *Algorithms for positive braids*, Quart. J. Math. Oxford Ser. (2) **45** (1994), 479–497.

[E+] D. B. A. EPSTEIN, J. W. CANNON, D. F. HOLT, S. V. F. LEVY, M. S. PATERSON, W. P. THURSTON – *Word processing in groups*, Jones and Bartlett Publ., Boston, Mass., 1992.

[F+] R. FENN, M. T. GREENE, D. ROLFSEN, C. ROURKE, B. WIEST – *Ordering the braid groups*, Pacific J. Math. **191** (1999), 49–74.

[FR] R. FENN, C. ROURKE – *Racks and links in codimension two*, J. Knot Theory Ramifications **1** (1992), 343–406.

[Ga] F. A. GARSIDE – *The braid group and other groups*, Quart. J. Math. Oxford **20** (1969), 235–254.

[Joy] D. JOYCE – *A classifying invariant of knots, the knot quandle*, J. Pure Appl. Algebra **23** (1982), 37–65.

[Kan] A. KANAMORI – *The higher infinite. Large cardinals in set theory from their beginnings.* Perspectives in Mathematical Logic, Springer-Verlag, Berlin, 1994.

[Kau] L. H. KAUFFMAN – *Knots and physics*, Series on Knots and Everything, vol. 1, World Scientific Publishing Co., Inc., River Edge, NJ, 1991.

[KR] D. M. KIM, D. ROLFSEN – *Ordering groups of pure braids and hyperplane arrangements*, prépublication Univ. British Columbia, 1998.

[La] D. M. LARUE – *On braid words and irreflexivity,* Algebra Universalis **31** (1994), 104–112.

[L92] R. LAVER – *The left distributive law and the freeness of an algebra of elementary embeddings*, Adv. Math. **91** (1992), 209–231.

[L95] R. LAVER – *On the algebra of elementary embeddings of a rank into itself,* Adv. Math. **110** (1995), 334–346.

[L96] R. LAVER – *Braid group actions on left distributive structures, and well orderings in the braid groups*, J. Pure Appl. Algebra **108** (1996), 81–98.

[Mk] G. S. MAKANIN – *The conjugacy problem in the braid group*, Dokl. Akad. Nauk SSSR **182** (1968), 495–496 (= Soviet Math. Dokl. **9** (1968), 1156–1157).

[Mt] S. V. Matveev – *Distributive groupoids in knot theory*, Mat. Sb. (N.S.) **119 (161)** (1982), 78–88 (= Math. USSR Sbornik **47** (1984), 73–83).

[Mo] L. Mosher – *Mapping class groups are automatic*, Ann. of Math. (2) **142** (1995), 303–384.

[MR] R. Botto Mura, A. H. Rhemtulla – *Orderable groups*, Lecture Notes in Pure and Applied Mathematics, vol. **27**, Marcel Dekker, Inc., New York, Basel, 1977.

[Ni] J. Nielsen – *Untersuchungen zur Topologie der geschlossenen zweiseitigen Flächen*, Acta Math. **50** (1927), 189–358 ; traduit en anglais par John Stillwell dans Jakob Nielsen, *Collected mathematical papers*, Birkhäuser, Boston, Basel, Stuttgart, 1986.

[Pa] D. S. Passman – *The algebraic structure of group rings*, Pure and Applied Mathematics, Wiley-Interscience, New York, London, Sydney, 1977.

[RZ] D. Rolfsen, J. Zhu – *Braids, orderings and zero divisors*, J. Knot Theory Ramifications **7** (1998), 837–841.

[RW] C. Rourke, B. Wiest – *Order automatic mapping class groups*, prépublication University of Warwick 1998, à paraître dans Pacific J. Math.

[SW] H. Short, B. Wiest – *Orderings of mapping class groups after Thurston*, prépublication CMI, Université de Provence 1999, math.GT/9907104.

[Th] W. P. Thurston – Courrier électronique à Bertold Wiest, 4 septembre 1998.

[W] B. Wiest – *Dehornoy's ordering of the braid groups extends the subword ordering*, Pacific J. Math. **191** (1999), 183–188.

Christian KASSEL

Institut de Recherche Mathématique Avancée
Université Louis Pasteur - C.N.R.S.
7 rue René Descartes
F-67084 Strasbourg Cedex, France
E-mail : kassel@math.u-strasbg.fr

Séminaire BOURBAKI Novembre 1999
52^{e} année, 1999-2000, n^{o} 866, p. 29 à 51

EXPOSANTS CRITIQUES POUR LE MOUVEMENT BROWNIEN ET LES MARCHES ALÉATOIRES [d'après Kenyon, Lawler et Werner]

par **Jean-François LE GALL**

INTRODUCTION

Cet exposé présente plusieurs travaux récents qui étudient certains exposants asymptotiques pour le mouvement brownien ou les marches aléatoires dans le plan. La première partie, issue des articles de Lawler et Werner, est consacrée aux exposants d'intersection, qui donnent l'ordre de grandeur de la probabilité pour deux ou plusieurs trajectoires de mouvements browniens, ou de marches aléatoires, de ne pas se rencontrer sur un intervalle de temps long. Si l'existence de ces exposants et parfois leurs valeurs exactes ont été prédites depuis longtemps par les physiciens théoriciens, leur construction rigoureuse ne date que de quelques années, et c'est seulement avec les travaux récents de Lawler et Werner que leur étude mathématique a connu des progrès vraiment significatifs. Un pas décisif a été l'obtention de relations « en cascade » entre les différents exposants (Théorème 1.3). Ces relations permettent ensuite de montrer que la classe importante des exposants dans un demi-espace peut s'écrire en termes d'une seule fonction réelle U pour laquelle on dispose d'une conjecture très plausible. La fonction U possède une propriété d'universalité remarquable puisqu'elle apparaît dès que l'on cherche à définir des exposants asymptotiques pour des mesures définies sur une classe de sous-ensembles compacts du plan et vérifiant une propriété convenable d'invariance conforme (Théorème 1.7). Après la soumission de la première version de cet exposé, nous avons reçu un travail tout récent de Lawler, Schramm et Werner [26] qui résout certaines des conjectures relatives aux exposants d'intersection (une brève discussion de ces derniers résultats est donnée à la fin de la première partie).

La deuxième partie a pour objectif principal de présenter un résultat de Kenyon (Théorème 2.1) donnant la valeur exacte de l'exposant de croissance de la marche aléatoire à boucles effacées dans le plan. La marche aléatoire à boucles effacées est un chemin aléatoire auto-évitant (i.e. sans auto-intersection) et est à ce titre un objet d'intérêt pour les physiciens théoriciens qui étudient les modèles de polymères. Le calcul de l'exposant de croissance applique des méthodes que Kenyon a développées pour l'étude asymptotique du nombre de pavages par dominos de certains sous-ensembles du plan, lorsque le pas du réseau devient petit. Un résultat significatif de Kenyon dans cette direction figure aussi à la fin de cet exposé (Théorème 2.7). Ce résultat est susceptible de plusieurs interprétations, puisque les pavages par dominos des régions considérées sont en bijection avec les arbres couvrants du graphe associé, et qu'on sait depuis Kirchhoff que le nombre d'arbres couvrants est égal au produit des valeurs propres non nulles du Laplacien discret sur le graphe.

Les deux parties sont indépendantes, et les techniques impliquées sont très différentes. Cependant il existe des liens étroits entre les objets étudiés : l'article de Lawler et Werner qui fournit l'essentiel de la matière de la première partie donne aussi des conjectures précises sur le comportement asymptotique de la marche aléatoire à boucles effacées. En outre, les propriétés d'invariance conforme, pour les trajectoires browniennes dans la première partie ou la fonction de couplage asymptotique dans la seconde, jouent un rôle fondamental dans tout l'exposé.

1. EXPOSANTS D'INTERSECTION DU MOUVEMENT BROWNIEN

1.1. La définition des exposants

Soient des entiers $k \geq 2$ et $n_1, \ldots, n_k \geq 1$ et soit une famille $(B^i_j; 1 \leq i \leq k, 1 \leq j \leq n_i)$ de $n_1 + \cdots + n_k$ mouvements browniens plans indépendants, dont le point de départ est uniformément distribué sur le cercle unité. Pour tout $R > 1$, soit

$$T^i_j(R) = \inf\{t \geq 0 : |B^i_j(t)| = R\}$$

le temps de sortie du mouvement brownien B^i_j hors du disque de rayon R centré en l'origine. Pour $a \geq 0$ on note $B^i_j[0, a] = \{B^i_j(t); 0 \leq t \leq a\}$ l'ensemble des points visités par B^i_j sur l'intervalle de temps $[0, a]$.

Proposition 1.1. — *Il existe un réel $\zeta(n_1, \ldots, n_k) > 0$, tel que, pour $R \to \infty$,*

$$P\Big[\Big(\bigcup_{j=1}^{n_i} B^i_j[0, T^i_j(R)]\Big) \cap \Big(\bigcup_{j=1}^{n_{i'}} B^{i'}_j[0, T^{i'}_j(R)]\Big) = \varnothing, \text{ pour tous } i \neq i'\Big] \approx R^{-\zeta(n_1,\ldots,n_k)}. \tag{1}$$

La notation $f(R) \approx g(R)$ signifie que $\frac{\log f(R)}{\log g(R)} \longrightarrow 1$ quand $R \to \infty$.

Le nombre $\zeta(n_1, \ldots, n_k)$ est appelé *exposant d'intersection* pour k paquets de respectivement $n_1, \ldots, n_k$ mouvements browniens plans. La première construction mathématique de ces exposants est due à Burdzy et Lawler [2] (voir aussi [6]).

Esquissons la preuve de la proposition. Soient z_j^i, $1 \le i \le k, 1 \le j \le n_i$ des points du cercle unité. Écrivons $P^{(z_j^i)}$ pour la probabilité sous laquelle les processus B_j^i sont des mouvements browniens plans indépendants et B_j^i part de z_j^i. Soit $\mathcal{A}_R$ l'événement

$$\mathcal{A}_R = \Big\{ \Big(\bigcup_{j=1}^{n_i} B_j^i[0, T_j^i(R)] \Big) \cap \Big(\bigcup_{j=1}^{n_{i'}} B_j^{i'}[0, T_j^{i'}(R)] \Big) = \varnothing, \text{ pour tous } i \neq i' \Big\}.$$

Posons

$$f(R) = \sup_{(z_j^i)} P^{(z_j^i)}(\mathcal{A}_R),$$

où le supremum porte sur tous les choix possibles de familles (z_j^i) de points sur le cercle unité. Une application de la propriété de Markov forte du mouvement brownien, puis de la propriété d'invariance par changement d'échelle, montre que, pour tous $R, R' > 1$,

$$f(RR') \le f(R) f(R').$$

Donc $\log f$ est fonction sous-additive de $\log R$, et la limite

$$\lim_{R\to\infty} \frac{\log f(R)}{\log R} = -\zeta(n_1, \ldots, n_k)$$

existe. Des estimations simples montrent que $0 < \zeta(n_1, \ldots, n_k) < \infty$ (pour voir que $\zeta(n_1, \ldots, n_k)$ est fini, on observe que $\mathcal{A}_R$ est réalisé dès que les k paquets de mouvements browniens sont contenus dans des cônes disjoints basés en l'origine).

Il est ensuite facile d'obtenir l'énoncé de la proposition. La quantité $P(\mathcal{A}_R)$ qui figure à gauche dans (1) est évidemment majorée par $f(R)$. Inversement, si on observe que la loi du point d'atteinte par B_j^i du cercle de rayon 2 centré en 0 est majorée par une constante fois la probabilité uniforme sur ce cercle, on obtient aisément l'estimation $f(2R) \le C\, P(\mathcal{A}_R)$, C étant une constante indépendante de R.

Remarque 1.2. — Modulo un certain travail technique, on peut remplacer les temps de sortie de grands disques centrés en l'origine par des instants déterministes qu'on fait tendre vers l'infini. Par exemple, on peut définir l'exposant $\zeta(1,1)$ en disant que si B et B' sont deux mouvements browniens plans indépendants issus de points distincts, on a

$$P[B[0,t] \cap B'[0,t] = \varnothing] \approx t^{-\zeta(1,1)/2}$$

quand $t \to \infty$. On peut aussi [2] remplacer les mouvements browniens par des marches aléatoires : si S et S' sont deux marches aléatoires simples (i.e. à plus proches voisins) indépendantes dans $\mathbb{Z}^2$, issues de points distincts, on a

$$P[\{S_k; 0 \le k \le n\} \cap \{S'_k; 0 \le k \le n\} = \varnothing] \approx n^{-\zeta(1,1)/2}$$

quand $n \to \infty$. La définition des exposants qui résulte de la Proposition 1.1 est cependant la plus facile à justifier.

Dans toute la classe des exposants $\zeta(n_1,\ldots,n_k)$ un seul a pu être calculé exactement : $\zeta(2,1)=2$. Cependant, Lawler et Werner [24] ont montré qu'il existe des relations remarquables et inattendues entre ces exposants. L'objectif principal des paragraphes qui suivent est d'expliquer la démarche qui conduit à ces relations. Parallèlement, nous donnerons aussi certains résultats qui justifient l'importance du calcul des exposants.

Un rôle crucial dans l'approche de [24] est joué par une seconde classe d'exposants, les exposants d'intersection dans un demi-plan. Soit $H=\{z:\mathrm{Im}z>0\}$ le demi-plan supérieur. Des arguments très semblables à ceux de la preuve précédente montrent que

$$P[\mathcal{A}_R\cap\{B^i_j[0,T^i_j(R)]\subset H \text{ pour tous } i,j\}]\approx R^{-\xi(n_1,\ldots,n_k)}$$

avec un exposant $\xi(n_1,\ldots,n_k)\geq\zeta(n_1,\ldots,n_k)$. Remarquons que cela a un sens de définir l'exposant ξ lorsque $k=1$ (alors $\mathcal{A}_R$ n'intervient pas), et des estimations faciles montrent que $\xi(n_1)=n_1$.

Bien qu'il n'y ait jusqu'à présent pas de calcul mathématique rigoureux des exposants, les physiciens théoriciens ont proposé des valeurs exactes pour certains d'entre eux. Sur la base de simulations et de considérations non-rigoureuses de théorie des champs conforme, Duplantier et Kwon [11] (voir aussi [12]) ont conjecturé que

$$\xi(1^{\otimes k})=\frac{k(2k+1)}{3}\,,\quad \zeta(1^{\otimes k})=\frac{4k^2-1}{12}\,,$$

où $\alpha^{\otimes k}$ désigne le k-uplet dont toutes les composantes sont égales à α. Bien que la valeur donnée ci-dessus pour ζ lorsque $k=1$ ($\zeta(1)=\frac{1}{4}$) n'ait a priori pas de sens, Duplantier a suggéré que cette valeur correspond à l'exposant de disconnexion que nous introduirons plus loin. Après avoir eu connaissance du travail de Lawler et Werner [24], Duplantier [8],[9] a justifié et étendu les conjectures précédentes dans le cadre (non rigoureux pour les mathématiciens) de la théorie de la gravité quantique.

1.2. L'extension aux valeurs réelles des paramètres

Dans ce paragraphe, nous montrons comment on étend la définition des exposants ζ ou ξ au cas où certains des paramètres $n_1,\ldots,n_k$ prennent des valeurs réelles. Par souci de simplicité, nous définissons d'abord $\zeta(1,\lambda)$ pour tout $\lambda\geq 0$. Nous considérons deux mouvements browniens plans B et W partant comme ci-dessus d'un point uniformément réparti sur le cercle unité, définis respectivement sous les probabilités P et P'. Notons T_R, resp. S_R, pour le temps d'atteinte par B, resp. W, du cercle de rayon R. Par définition,

$$P\otimes P'[B[0,T_R]\cap W[0,S_R]=\varnothing]\approx R^{-\zeta(1,1)}.$$

En intégrant d'abord par rapport à P', on peut aussi écrire

$$P\otimes P'[B[0,T_R]\cap W[0,S_R]=\varnothing]=E[Z_R(B)]$$

où $Z_R(B) = P'[B[0, T_R] \cap W[0, S_R] = \varnothing]$ est une fonction du mouvement brownien B (en fait de l'ensemble aléatoire $B[0, T_R]$). Si W est un autre mouvement brownien plan défini sous P' (indépendant de W et avec la même loi de départ), le même argument montre aussi que

$$E[Z_R(B)^2] = P \otimes P'[B[0, T_R] \cap \big(W[0, S_R] \cup \tilde{W}[0, \tilde{S}_R]\big) = \varnothing] \approx R^{-\zeta(1,2)}$$

et plus généralement, pour tout entier $n \geq 1$,

$$E[Z_R(B)^n] \approx R^{-\zeta(1,n)}.$$

Ceci suggère de définir pour tout *réel* $\lambda \geq 0$ l'exposant $\zeta(1, \lambda)$ par :

$$E[Z_R(B)^\lambda] \approx R^{-\zeta(1,\lambda)},$$

avec la convention inhabituelle que lorsque $\lambda = 0$, $0^0 = 0$. Un argument de sous-additivité montre à nouveau qu'il existe un tel exposant $\zeta(1, \lambda)$.

La même idée permet pour tout entier $k \geq 1$, pour tous les entiers $n_1, \ldots, n_k \geq 1$ et tous les réels $\lambda_1, \ldots, \lambda_k \geq 0$, de définir un exposant $\zeta(n_1, \lambda_1, \ldots, n_k, \lambda_k)$ qui étend la définition précédente. Remarquons qu'il est (pour l'instant) nécessaire d'alterner entre valeurs réelles et entières des paramètres. En effet, on conditionne d'abord par rapport aux k paquets de respectivement $n_1, \ldots, n_k$ mouvements browniens, puis on considère k variables $Z_R^1, \ldots, Z_R^k$ qui sont (toutes) non nulles seulement sur l'ensemble où les k paquets sont d'intersection vide deux à deux, et de plus sont ordonnés dans le sens trigonométrique autour de l'origine. La variable Z_R^i est alors la probabilité conditionnelle qu'un mouvement brownien indépendant reste « entre le i-ième et le $i+1$-ième paquet » jusqu'au temps d'atteinte du rayon R. Finalement, l'exposant $\zeta(n_1, \lambda_1, \ldots, n_k, \lambda_k)$ est défini par

$$E[(Z_R^1)^{\lambda_1} \cdots (Z_R^k)^{\lambda_k}] \approx R^{-\zeta(n_1,\lambda_1,\ldots,n_k,\lambda_k)}.$$

Une extension analogue vaut pour le cas des exposants ξ dans un demi-espace. En fait pour ces exposants il est naturel de définir $\xi(\lambda_0, n_1, \lambda_1, \ldots, n_k, \lambda_k)$ pour $n_1, \ldots, n_k$ entiers (non-nuls) et $\lambda_0, \lambda_1, \ldots, \lambda_k$ réels positifs. La méthode est exactement celle décrite ci-dessus, mais on impose aux paquets de mouvements browniens de rester dans le demi-plan supérieur H et en plus de $Z_R^1, \ldots, Z_R^k$, on introduit aussi Z_R^0 qui est la probabilité pour un mouvement brownien indépendant de rester (jusqu'au temps d'atteinte du rayon R) entre la demi-droite réelle positive et le premier paquet de n_1 mouvements browniens. Par convention on prend $\xi(\lambda) = \lambda$ pour tout $\lambda \geq 0$, ce qui est cohérent avec le cas où λ est entier.

On notera dans la suite $\xi(1, \lambda) = \xi(0, 1, \lambda)$.

1.3. Propriétés trajectorielles du mouvement brownien

Avant d'aller plus loin, expliquons pourquoi le calcul des exposants est important en liaison avec l'étude de la trajectoire brownienne. Remarquons d'abord que les exposants $\zeta(p, 0)$, pour p entier strictement positif, ont une interprétation remarquable.

En effet, avec les notations ci-dessus, on a

$$P[Z_R(B) > 0] \approx R^{-\zeta(1,0)}.$$

Or l'événement $\mathcal{B}_R = \{Z_R(B) > 0\}$ est réalisé ssi la trajectoire $B[0, T_R]$ ne disconnecte pas le cercle unité de l'infini. On appelle donc $\zeta(1,0)$ l'exposant de disconnexion pour une trajectoire brownienne. On a une interprétation analogue pour $\zeta(p,0)$, qui est l'exposant de disconnexion pour p trajectoires browniennes indépendantes. Lawler a montré que ces exposants sont liés à une conjecture fameuse de Mandelbrot [28] sur la dimension de Hausdorff du contour de la courbe brownienne. Le contour, noté $\mathcal{C}$, de $B[0,1]$ est par définition la frontière de la composante connexe non bornée de $\mathbb{C}\backslash B[0,1]$. La conjecture de Mandelbrot est $\dim \mathcal{C} = \frac{4}{3}$, et Lawler [22] a montré que

$$\dim \mathcal{C} = 2 - \zeta(2,0)$$

(un premier résultat dans cette direction avait été obtenu dans [3]). L'idée de la preuve est relativement simple : si $z = B_t$, avec $t \in]0,1[$, est un point de $\mathcal{C}$, la courbe de B considérée avant et après l'instant t fournit localement deux trajectoires browniennes indépendantes issues de z qui ne disconnectent pas le point z de l'infini.

Un second lien entre exposants d'intersection et propriétés trajectorielles de la courbe brownienne concerne les points de coupure. Un point $z \in \mathbb{C}$ est appelé point de coupure de B si $z = B_s$ pour une valeur de $s \in]0,1[$ telle que $B[0,s] \cap B[s,1] = \{B_s\}$. Si $\mathcal{D}$ désigne l'ensemble des points de coupure, Lawler [21] a montré que

$$\dim \mathcal{D} = 2 - \zeta(1,1).$$

Voir encore [4] pour d'autres applications des exposants d'intersection.

1.4. Les relations en cascade

Théorème 1.3 ([24]). — *Soient $n_1, \ldots, n_k \geq 1$ des entiers et $\lambda_0, \lambda_1, \ldots, \lambda_k$ des réels positifs. Alors, pour tout entier $j \in \{1, \ldots, k-1\}$,*

$$\xi(\lambda_0, n_1, \lambda_1, \ldots, n_k, \lambda_k) = \xi(\lambda_0, n_1, \lambda_1, \ldots, n_j, \xi(\lambda_j, n_{j+1}, \lambda_{j+1}, \ldots, n_k, \lambda_k))$$

et

$$\zeta(n_1, \lambda_1, \ldots, n_k, \lambda_k) = \zeta(n_1, \lambda_1, \ldots, n_j, \xi(\lambda_j, n_{j+1}, \lambda_{j+1}, \ldots, n_k, \lambda_k)).$$

Remarque 1.4. — La première relation ne met en jeu que les exposants ξ, alors que la seconde mélange ξ et ζ. D'une certaine manière, les exposants dans un demi-espace sont plus fondamentaux.

Preuve (esquisse) — Nous allons traiter en détail un cas particulier du théorème, à savoir la relation $\xi(1,1,1) = \xi(1, \xi(1,1))$. Les mêmes arguments montreraient aussi bien l'égalité $\zeta(1,1,1) = \zeta(1, \xi(1,1))$, et la preuve de la forme générale des relations en cascade repose sur des idées très semblables.

Nous considérons trois mouvements browniens plans indépendants W, B, B' issus respectivement de y, z_1, z_2 avec $y = e^{i\gamma}$, $z_1 = e^{i\theta_1}$, $z_2 = e^{i\theta_2}$, où $0 < \gamma < \theta_1 < \theta_2 < \pi$.

Par commodité, nous supposerons que W, B, B' sont en fait des mouvements browniens conditionnés à atteindre le cercle de rayon R avant de pénétrer dans le disque unité ouvert (bien qu'il s'agisse d'un conditionnement par un ensemble de probabilité nulle, on vérifie assez facilement que les exposants peuvent être définis aussi bien en termes de ces processus conditionnés). On a alors, avec des notations évidentes

$$P\big[W[0,S_R]\cap B[0,T_R]=\varnothing,\quad B[0,T_R]\cap B'[0,T'_R]=\varnothing,\\ W[0,S_R]\subset H,\quad B[0,T_R]\subset H,\quad B'[0,T'_R]\subset H\big]\approx R^{-\xi(1,1,1)}.$$

Observer que l'événement $\{W[0,S_R]\cap B'[0,T'_R]=\varnothing\}$ est automatiquement réalisé sur l'ensemble que nous considérons.

Nous allons maintenant conditionner par rapport à la trajectoire $W[0,S_R]$ (en supposant que $W[0,S_R]\subset H$). Notons $H_R=\{z\in H : 1<|z|<R\}$ et U_R la composante connexe de $H_R\backslash W[0,S_R]$ dont la frontière contient z_1 et z_2. Alors,

$$\begin{aligned}&P\big[W[0,S_R]\cap B[0,T_R]=\varnothing,\quad B[0,T_R]\cap B'[0,T'_R]=\varnothing,\\ &\qquad W[0,S_R]\subset H,\quad B[0,T_R]\subset H,\quad B'[0,T'_R]\subset H\big]\\ &=E\left[1_{\{W[0,S_R]\subset H\}}\,P\big[B[0,T_R]\cap B'[0,T_R]=\varnothing, B[0,T_R]\subset U_R, B'[0,T'_R]\subset U_R\mid W\big]\right].\end{aligned}$$

On peut trouver un réel $L_R>1$ (dépendant de U_R et donc de W) et une bijection conforme φ de U_R sur H_{L_R}, qui se prolonge par continuité à $\bar U_R$, de telle manière que :

– φ envoie $\partial^1U_R:=\{z\in\partial U_R : |z|=1\}$ sur $\partial^1H_{L_R}:=\{z\in\partial H_{L_R} : |z|=1\}$;
– φ envoie $\partial^2U_R:=\{z\in\partial U_R : |z|=R\}$ sur $\partial^2H_{L_R}:=\{z\in\partial H_{L_R} : |z|=L_R\}$.

Le nombre L_R s'interprète en termes de la distance extrémale entre ∂^1U_R et ∂^2U_R dans U_R, voir [1].

L'invariance conforme du mouvement brownien plan montre alors que la probabilité conditionnelle

$$P\big[B[0,T_R]\cap B'[0,T_R]=\varnothing, B[0,T_R]\subset U_R, B'[0,T'_R]\subset U_R\mid W\big]$$

est égale à la probabilité que deux mouvements browniens indépendants issus de $\partial^1H_{L_R}$ ne se rencontrent pas et ne quittent pas H_{L_R} avant d'atteindre le cercle de rayon L_R. Par définition des exposants dans un demi-espace, nous savons que cette probabilité se comporte comme $L_R^{-\xi(1,1)}$.

De même par invariance conforme, la quantité

$$P\big[B[0,T_R]\subset U_R\mid W\big]$$

est égale à la probabilité qu'un mouvement brownien issu de $\partial^1H_{L_R}$ ne quitte pas H_{L_R} avant d'atteindre le cercle de rayon L_R. Une estimation facile montre que cette probabilité se comporte comme L_R^{-1}.

En mettant bout à bout les considérations précédentes, on obtient

$$\begin{aligned}
R^{-\xi(1,1,1)} &\approx E\big[1_{\{W[0,S_R]\subset H\}} L_R^{-\xi(1,1)}\big] \\
&\approx E\big[1_{\{W[0,S_R]\subset H\}} P\big[B[0,T_R]\subset U_R \mid W\big]^{\xi(1,1)}\big] \\
&\approx R^{-\xi(1,\xi(1,1))}
\end{aligned}$$

où la dernière ligne découle directement de la définition des exposants $\xi(1,\lambda) = \xi(0,1,\lambda)$ pour λ réel. On obtient la relation $\xi(1,1,1) = \xi(1,\xi(1,1))$ comme annoncé. $\square$

En utilisant aussi un argument de symétrie, on déduit assez facilement du Théorème 1.3 des *relations de commutation* pour les exposants d'intersection. Par exemple, pour toutes permutations ρ et σ de $\{1,\ldots,k\}$, on a

$$\zeta(n_1,\lambda_1,\ldots,n_k,\lambda_k) = \zeta(n_{\rho(1)},\lambda_{\sigma(1)},\ldots,n_{\rho(k)},\lambda_{\sigma(k)}),$$

et on a un énoncé analogue pour les exposants ξ.

1.5. Le théorème principal

Les relations en cascade permettent d'étendre la définition des exposants à toutes les valeurs réelles des exposants. Commençons par le cas des exposants dans un demi-espace. On montre que la fonction $\lambda\to\xi(1,\lambda)$ est strictement croissante et continue de $[0,\infty[$ sur $[1,\infty[$ (la continuité sur $]0,\infty[$ découle du fait que cette fonction est concave). Alors, si α et α' sont deux réels positifs quelconques, on définit $\xi(\alpha,\alpha')$ comme étant l'unique réel positif tel que $\xi(1,\xi(\alpha,\alpha')) = \xi(\alpha,1,\alpha')$ (ce choix est forcé si l'on veut conserver les relations de commutation et les relations en cascade décrites ci-dessus). Puis on définit simplement par récurrence

$$\xi(\alpha_1,\ldots,\alpha_k) = \xi(\alpha_1,\xi(\alpha_2,\ldots,\alpha_k))$$

pour tous *réels* $\alpha_1,\ldots,\alpha_k \geq 0$. Il est ensuite possible de vérifier que cette définition est cohérente avec le cas où les α_j sont entiers, puis que les relations en cascade $(\xi(\alpha_1,\ldots,\alpha_k) = \xi(\alpha_1,\ldots,\alpha_j,\xi(\alpha_{j+1},\ldots,\alpha_k)))$ ainsi que les relations de commutation $(\xi(\alpha_1,\ldots,\alpha_k) = \xi(\alpha_{\sigma(1)},\ldots,\alpha_{\sigma(k)}))$ sont valides.

On procède (presque) de même pour les exposants ζ. Il y a cependant une difficulté. On ne définit $\zeta(\alpha_1,\ldots,\alpha_k)$, pour $k\geq 2$ et $\alpha_1,\ldots,\alpha_k$ réels positifs, que si deux au moins des nombres α_i sont supérieurs ou égaux à 1. Si $\alpha_j\geq 1$, on définit β_j par la relation $\alpha_j = \xi(1,\beta_j)$ puis on pose

$$\zeta(\alpha_1,\ldots,\alpha_k) = \zeta(1,\xi(\alpha_1,\ldots,\alpha_{j-1},\beta_j,\alpha_{j+1},\ldots,\alpha_k)).$$

La condition que deux des nombres α_i sont supérieurs ou égaux à 1 intervient pour montrer que cette définition de $\zeta(\alpha_1,\ldots,\alpha_k)$ est bien cohérente avec celle introduite précédemment. Les relations en cascade

$$\zeta(\alpha_1,\ldots,\alpha_k) = \zeta(\alpha_1,\ldots,\alpha_j,\xi(\alpha_{j+1},\ldots,\alpha_k))$$

et les relations de commutation

$$\zeta(\alpha_1,\ldots,\alpha_k)=\zeta(\alpha_{\sigma(1)},\ldots,\alpha_{\sigma(k)})$$

restent vraies.

Théorème 1.5 ([24]). — *Il existe une fonction $U:[0,\infty[\longrightarrow[0,\infty[$ continue et strictement croissante, avec $U(0)=0$, $U(\infty)=\infty$, telle que, pour tous $k\geq 1$ et $\alpha_1,\ldots,\alpha_k\geq 0$,*

$$\xi(\alpha_1,\ldots,\alpha_k)=U^{-1}\big(U(\alpha_1)+\cdots+U(\alpha_k)\big).$$

De plus, il existe une fonction croissante continue $\eta:[\xi(1,1),\infty[\longrightarrow[0,\infty[$ telle que, pour tout entier $k\geq 2$ et tous $\alpha_1,\ldots,\alpha_k\geq 0$ avec $\alpha_1\geq 1$ et $\alpha_2\geq 1$, on a

$$\zeta(\alpha_1,\ldots,\alpha_k)=\eta(\xi(\alpha_1,\ldots,\alpha_k)).$$

La fonction U peut être obtenue par la formule

$$U^2(\alpha)=\lim_{N\to\infty}\frac{1}{N^2}\xi(\alpha^{\otimes N}).$$

La preuve de la première partie du théorème utilise fortement les relations en cascade entre les exposants. Un autre ingrédient important est le comportement asymptotique de l'exposant d'intersection pour k paquets de mouvements browniens, lorsque le nombre de mouvements browniens à l'intérieur de chaque paquet tend vers l'infini [33].

L'expression de ξ en fonction de ζ est une conséquence presque directe des relations en cascade. Rappelons que la fonction $\kappa(\lambda)=\xi(1,\lambda)$ est bijective de $[0,\infty[$ sur $[1,\infty[$ et, en supposant $\alpha_1\geq 1$ notons $\beta_1=\kappa^{-1}(\alpha_1)$. Alors, en utilisant d'abord la construction donnée ci-dessus de $\zeta(\alpha_1,\ldots,\alpha_k)$, puis en appliquant deux fois les relations en cascade, on a

$$\begin{aligned}\zeta(\alpha_1,\ldots,\alpha_k)&=\zeta(1,\xi(\beta_1,\alpha_2,\ldots,\alpha_k))\\&=\zeta(1,\kappa^{-1}(\xi(1,\beta_1,\alpha_2,\ldots,\alpha_k)))\\&=\zeta(1,\kappa^{-1}(\xi(\alpha_1,\ldots,\alpha_k)))\,,\end{aligned}$$

ce qui donne la dernière formule du théorème avec $\eta(\lambda)=\zeta(1,\kappa^{-1}(\lambda)))$.

Remarque. Du point de vue de la physique théorique [8],[9], les formules du Théorème 1.5 trouvent une interprétation dans le cadre de la théorie (non rigoureuse) de la gravité quantique. À titre d'exemple, la fonction U^{-1} est la transformation qui fait passer les exposants du demi-plan « quantique » (où les exposants ont une structure additive) vers le demi-plan « standard ».

1.6. Conjectures

Les prédictions de [11] pour les valeurs des exposants $\zeta(1^{\otimes N})$ et $\xi(1^{\otimes N})$ conduisent à conjecturer les formes explicites suivantes pour les fonctions U et η qui interviennent dans le Théorème 1.5 :

$$U(x) = \frac{\sqrt{24x+1}-1}{\sqrt{24}}\ , \qquad \eta(\alpha) = \frac{(\sqrt{24\alpha+1}-1)^2-4}{48}\ .$$

Il est très plausible que la dernière formule du Théorème 1.5 reste vraie sous la seule condition que $\max(\alpha_i) \geq 1$ (qui est nécessaire pour définir $\zeta(\alpha_1,\ldots,\alpha_p)$). En particulier on devrait avoir, pour tout entier $p \geq 1$, $\zeta(p,0) = \eta(\xi(p,0)) = \eta(p)$, et la fonction η s'interprèterait comme un exposant de disconnexion généralisé. La valeur $\eta(1) = \frac{1}{4}$ est bien cohérente avec les prédictions de Duplantier et Kwon [11]. De même la valeur $\eta(2) = \frac{2}{3}$ est cohérente avec la conjecture de Mandelbrot sur la dimension de Hausdorff du contour $\mathcal{C}$ de la courbe brownienne.

Les conjectures pour U et η et le Théorème 1.5 conduisent aux valeurs suivantes pour les exposants ζ et ξ :

$$\zeta(\alpha_1,\ldots,\alpha_k) = \frac{(\sqrt{24\alpha_1+1}+\cdots+\sqrt{24\alpha_k+1}-k)^2-4}{48}$$
$$\xi(\alpha_1,\ldots,\alpha_k) = \frac{(\sqrt{24\alpha_1+1}+\cdots+\sqrt{24\alpha_k+1}-(k-1))^2-1}{24}.$$

Il est rassurant de voir qu'on retrouve la seule valeur connue $\zeta(2,1) = 2$!

1.7. Prolongements

Très récemment, Lawler et Werner [25] ont montré une propriété d'universalité remarquable des exposants ξ, ou de manière équivalente de la fonction U du Théorème 1.5. Ces exposants interviennent non seulement dans l'étude des trajectoires browniennes mais bien plus généralement dès que l'on s'intéresse à des mesures définies sur une certaine classe de sous-ensembles compacts du disque unité et vérifiant une propriété d'invariance conforme. Dans ce paragraphe, nous donnons une présentation succincte d'un résultat majeur de [25].

On note D le disque unité ouvert dans le plan. Un *pont* est par définition un sous-ensemble compact K simplement connexe de $\bar{D}$ tel que $\partial K \cap \partial D$ a exactement deux composantes connexes. Le pont est *orienté* lorsque l'on a choisi l'une de ces deux composantes connexes, qui sera notée $d_1 = d_1(K)$, l'autre étant $d_2 = d_2(K)$. On note $\mathcal{K}$ l'ensemble des ponts orientés. Si A_1 et A_2 sont deux arcs (fermés) disjoints sur ∂D, on note $\mathcal{K}(A_1, A_2)$ l'ensemble des ponts orientés tels que $d_1 \subset A_1$ et $d_2 \subset A_2$. Lorsque $A_2 = V_\delta^+ := \{e^{i\theta}; -\delta \leq \theta \leq \delta\}$ et $A_1 = V_\delta^- := -A_2$, on note simplement $\mathcal{K}_\delta = \mathcal{K}(V_\delta^-, V_\delta^+)$. Enfin, si $D' \subset D$ et $A_1 \cup A_2 \subset \partial D'$, on note $\mathcal{K}(A_1, A_2; D')$ l'ensemble des ponts orientés de $\mathcal{K}(A_1, A_2)$ qui ont la propriété additionnelle que $K \cap D \subset D'$.

Définition 1.6. — *Une mesure μ sur $\mathcal{K}$ est dite complètement invariante conforme (CIC) si elle possède les deux propriétés suivantes.*

- *Il existe un exposant $\alpha = \alpha(\mu) > 0$ tel que, quand $\delta \to 0$,*

$$\mu(\mathcal{K}_\delta) \approx \delta^{2\alpha}.$$

- *Soit D' un sous-ensemble ouvert simplement connexe de D et soient A_1, A_2 deux arcs (fermés non vides) disjoints contenus dans $\partial D \cap \partial D'$. Il existe une unique valeur $\delta = \delta(A_1, A_2; D')$ et une bijection conforme $\Phi : D' \longrightarrow D$ telles que $\Phi(A_1) = V_\delta^-$ et $\Phi(A_2) = V_\delta^+$. Alors, si $\mu_{A_1,A_2}^{D'}$ désigne la restriction de μ à $\mathcal{K}(A_1, A_2; D')$, on a*

$$\Phi\big(\mu_{A_1,A_2}^{D'}\big) = \mu_{V_\delta^-,V_\delta^+}^{D}.$$

Dans la dernière formule, $\Phi\big(\mu_{A_1,A_2}^{D'}\big)$ désigne la mesure image de $\mu_{A_1,A_2}^{D'}$ par l'application Φ, qui agit de manière évidente sur $\mathcal{K}(A_1, A_2; D')$.

La propriété la plus importante est la deuxième. L'exposant $\alpha(\mu)$ qui intervient dans la première est appelé exposant de croisement de μ. Il mesure la difficulté pour un pont distribué selon μ de passer par des « petites » portes d'entrée et de sortie. On vérifie aisément qu'une mesure CIC est nécessairement de masse totale infinie.

Exemple. — La mesure d'excursion du mouvement brownien plan à l'intérieur du cercle unité permet de construire un premier exemple de mesure CIC, que nous noterons μ_0 et qui fait le lien avec les paragraphes précédents. Le compact K est obtenu sous μ_0 comme complémentaire de la composante connexe non bornée du complémentaire de la trajectoire sous la mesure d'excursion. Dans ce cas, les arcs d_1 et d_2 sont réduits à des points, et on voit facilement que l'exposant de croisement vaut 1.

Nous allons maintenant définir un exposant d'intersection pour une famille finie de mesures CIC. Nous devons auparavant introduire une notation supplémentaire. Si K est un pont orienté, on note $D^- = D^-(K)$ et $D^+ = D^+(K)$ les composantes connexes de $D\backslash K$ numérotées de sorte que les arcs d_1, $\partial D^- \cap \partial D$, d_2 et $\partial D^+ \cap \partial D$ sont dans l'ordre trigonométrique. Soient alors $\mu_1, \ldots, \mu_p$ des mesures CIC, et supposons que les ponts orientés $K_1, \ldots, K_p$ sont choisis selon la mesure produit $\mu_1 \otimes \cdots \otimes \mu_p$. On note $\mathcal{O}$ l'événement sur lequel les ponts (orientés) $K_1, \ldots, K_p$ sont ordonnés, au sens où

$$K_j \cap D \subset D^-(K_{j+1})\ , \qquad j = 1, \ldots, p-1$$

(de manière imagée, le pont K_1 est situé « en dessous » du pont K_2, lequel est lui-même en-dessous de K_3, etc.). L'exposant d'intersection $\alpha = \alpha(\mu_1, \ldots, \mu_p)$ est alors défini par

$$\mu_1 \otimes \cdots \otimes \mu_p\big((\mathcal{K}_\delta)^p \cap \mathcal{O}\big) \approx \delta^{2\alpha}$$

quand $\delta \to 0$, lorsqu'un tel exposant α existe.

Autrement dit, on évalue la « probabilité » que les ponts « indépendants » $K_1, \ldots, K_p$ aient des arcs d'entrée et de sortie proches (puisque contenus dans V_δ^-, respectivement V_δ^+) mais néanmoins ne se rencontrent pas (en fait on leur demande même d'être dans un ordre bien défini, mais cela ne fait pas de différence pour le comportement logarithmique de la probabilité). Lorsque $\mu_1, \dots, \mu_p$ sont toutes égales à la mesure μ_0 construite à partir de la mesure d'excursion brownienne, on vérifie que $\alpha(\mu_0, \dots, \mu_0)$ existe, et

$$\alpha(\mu_0, \dots, \mu_0) = \xi(1, \dots, 1).$$

Le théorème qui suit généralise considérablement cette observation.

THÉORÈME 1.7 ([25]). — *Soient $\mu_1, \dots, \mu_p$ des mesures CIC. Alors l'exposant d'intersection $\alpha(\mu_1, \dots, \mu_p)$ existe, et*

$$\alpha(\mu_1, \dots, \mu_p) = \xi(\alpha(\mu_1), \dots, \alpha(\mu_p)).$$

L'énoncé du Théorème 1.7 ne fait plus intervenir de mouvement brownien, si on oublie la manière dont les exposants ξ ont été introduits ci-dessus. Cependant la preuve de ce théorème fait une utilisation intensive de la mesure d'excursion brownienne. L'intérêt du Théorème 1.7 est qu'il est susceptible de s'appliquer aux mesures CIC autres que la mesure μ_0, que l'on s'attend à obtenir par des passages à la limite convenables à partir de modèles probabilistes comme les marches aléatoires auto-évitantes ou les amas de percolation (tout récemment, Schramm [30] a discuté l'existence de modèles continus pour la marche aléatoire à boucles effacées qui sera étudiée dans la partie suivante). L'article [25] contient une série de conjectures dans cette direction, dont plusieurs rejoignent les prédictions des physiciens théoriciens.

Après la rédaction de la première version de cet exposé, nous avons eu connaissance d'un travail tout récent de Lawler, Schramm et Werner [26] qui démontre certaines des conjectures mentionnées dans le paragraphe 1.6 ci-dessus. L'idée est d'utiliser le processus de croissance introduit par Schramm [30] pour lequel des calculs explicites des exposants sont possibles. Grâce aux résultats d'universalité décrits dans le paragraphe 1.7, cela permet aussi de calculer au moins certaines valeurs des exposants d'intersection browniens, en accord avec les conjectures du paragraphe 1.6. Des résultats même plus complets que ceux de [26], incluant une preuve de la conjecture de Mandelbrot $\dim \mathcal{C} = \frac{4}{3}$, ont été annoncés par Lawler, Schramm et Werner.

2. MARCHES ALÉATOIRES À BOUCLES EFFACÉES ET PAVAGES PAR DOMINOS

2.1. Marches aléatoires à boucles effacées et arbres couvrants

La marche aléatoire à boucles effacées a été introduite par Lawler [19] en 1980. Commençons par rappeler sa définition. Soit $\gamma = (\gamma(0), \gamma(1), \dots, \gamma(N))$ un chemin déterministe dans $\mathbb{Z}^d$ ou sur un graphe quelconque. Si ce chemin est auto-évitant, c'est-à-dire si les $\gamma(i)$ sont distincts, on arrête là la construction. Sinon, on choisit j minimal tel qu'il existe $i < j$ avec $\gamma(j) = \gamma(i)$, on remplace γ par $\gamma' = (\gamma(0), \ldots, \gamma(i), \gamma(j+1), \ldots, \gamma(N))$ et on recommence. La construction s'arrête au bout d'un nombre fini d'étapes et donne alors un chemin auto-évitant ayant mêmes points de départ et d'arrivée que γ.

Si on considère maintenant un chemin infini $\gamma = (\gamma(0), \gamma(1), \ldots)$ ayant la propriété de ne passer qu'un nombre fini de fois en chaque sommet du graphe, la même méthode fournit un chemin infini auto-évitant. Ceci s'applique à une trajectoire de marche aléatoire simple dans $\mathbb{Z}^d$ pour $d \geq 3$ et le chemin aléatoire ainsi obtenu est appelé marche aléatoire à boucles effacées dans $\mathbb{Z}^d$. En revanche, dans le graphe $\mathbb{Z}^2$ (ou dans $\mathbb{N} \times \mathbb{Z}$), la construction précédente ne marche pas directement à cause de la récurrence de la marche aléatoire simple. Néanmoins, on peut l'appliquer d'abord à la marche aléatoire issue de 0 arrêtée au premier temps de sortie d'une grande boîte centrée en 0, puis faire tendre la taille de la boîte vers l'infini pour obtenir encore la marche aléatoire à boucles effacées dans $\mathbb{Z}^2$, ou dans $\mathbb{N} \times \mathbb{Z}$ (voir le dernier chapitre de [20]). L'article récent de Lawler [23] donne un aperçu très complet des problèmes ouverts et des conjectures sur les marches aléatoires à boucles effacées.

Le théorème suivant, dû à Kenyon [17], résout l'une des plus importantes de ces conjectures.

Théorème 2.1. — *Pour la marche aléatoire à boucles effacées dans $\mathbb{N} \times \mathbb{Z}$ issue de l'origine, le nombre moyen de points visités qui sont à distance inférieure à N de l'origine est $N^{\frac{5}{4}+o(1)}$. Plus précisément, la probabilité qu'un point de la forme $(x, y) = re^{i\theta}$ soit visité se comporte quand $r \to \infty$ comme*

$$r^{-\frac{3}{4}(1+o(1))}\,((\cos\theta)^{\frac{1}{4}} + o(1)),$$

où le terme $o(1)$ dans l'exposant ne dépend pas de θ.

On s'attend à ce qu'un résultat analogue soit vrai pour la marche aléatoire à boucles effacées dans $\mathbb{Z}^2$, même si cette extension présente probablement des difficultés techniques. Le théorème suggère fortement que la moyenne à l'instant n du carré de la norme d'une marche aléatoire à boucles effacées dans $\mathbb{Z}^2$ se comporte asymptotiquement comme $n^{\frac{8}{5}}$. Cette valeur de l'exposant $\gamma = \frac{8}{5}$ a été prédite par les physiciens théoriciens (Majumdar [27], voir aussi [7]). Lawler [20] donne la minoration $\gamma \geq \frac{3}{2}$.

La formulation du théorème précédent qui figure dans [17] est différente et utilise la notion d'arbre couvrant uniforme. Si G est un graphe connexe quelconque, on appelle arbre couvrant de G un sous-graphe connexe de G qui est un arbre et qui contient pour chaque sommet de G au moins une arête incidente à ce sommet. Lorsque G est fini, on définit un arbre couvrant uniforme comme étant un sous-graphe aléatoire de G dont la loi est la probabilité uniforme sur l'ensemble (fini) de tous les arbres couvrants de G.

Cette définition peut être étendue à des graphes infinis. Considérons le cas où $G = \mathbb{Z}^d$. Pour tout entier $n \geq 1$, notons D_n le cube de coté $2n$ centré en l'origine dans

$\mathbb{Z}^d$. Soit μ_n la probabilité uniforme sur l'ensemble des arbres couvrants de D_n, vu ici comme contenu dans l'ensemble des sous-graphes de $\mathbb{Z}^d$, qui est muni de la topologie-produit (on identifie un sous-graphe à un élément de $\{0,1\}^{\mathcal{A}}$, si $\mathcal{A}$ est l'ensemble des arêtes de $\mathbb{Z}^d$). On montre alors [29] que μ_n converge étroitement vers une mesure de probabilité μ portée par les sous-graphes de $\mathbb{Z}^d$. De plus, lorsque $d \leq 4$, la mesure μ est portée par les arbres couvrants de $\mathbb{Z}^d$, et on appelle arbre couvrant uniforme sur $\mathbb{Z}^d$ un arbre aléatoire dont la loi est μ. La même construction s'applique au graphe $\mathbb{N} \times \mathbb{Z}$.

On montre aussi [29], toujours en dimension $d \leq 4$, que la mesure μ est portée par les arbres qui ont un seul « bout », c'est-à-dire tels que la suppression d'un sommet divise l'arbre en deux composantes dont une seule est infinie. Il est alors immédiat qu'il existe dans l'arbre couvrant uniforme un seul chemin (sans recoupement) joignant 0 à ∞.

Proposition 2.2 ([29]). — *Pour l'arbre aléatoire couvrant uniforme dans $\mathbb{Z}^d$, $d \leq 4$ ou dans $\mathbb{N} \times \mathbb{Z}$, l'arc joignant 0 à l'infini a la loi d'une marche aléatoire à boucles effacées issue de 0.*

2.2. Arbres couvrants et pavages par dominos

Grâce à la Proposition 2.2, on peut ramener la preuve du Théorème 2.1 à celle d'un énoncé équivalent sur les arbres couvrants. Nous allons voir qu'on peut introduire encore une autre formulation mettant en jeu les pavages par dominos de certaines régions du plan.

Soit T le pavage « en échiquier » de $\mathbb{R}^2$ par des carrés unité, chaque carré étant centré en un point du réseau $\mathbb{Z}^2$. Nous supposons que le carré centré en l'origine est blanc, et nous notons W_0 l'ensemble des carrés (blancs) dont les coordonnées du centre sont paires, W_1 l'ensemble des autres carrés blancs. Nous notons aussi B_0 l'ensemble des carrés noirs dont les coordonnées du centre sont égales à $(1,0)$ modulo 2, B_1 l'ensemble des autres carrés noirs.

Un *polyomino* est une réunion finie de carrés de T dont la frontière est une courbe fermée simple (pour simplifier, nous ne considérons ici que des polyominos simplement connexes). Un polyomino est dit *pair* si tous les carrés de coin, que ce coin soit convexe ou concave, sont de type B_1. Enfin un polyomino *temperlien* est obtenu à partir d'un polyomino pair en enlevant un (seul) carré noir d adjacent à la frontière de ce dernier (d sera appelé le *point de base* du polyomino). À cause de la définition d'un polyomino pair, d est forcément de type B_1 (dans un polyomino pair les seuls carrés noirs adjacents à la frontière sont de type B_1). Remarquons qu'un polyomino temperlien contient le même nombre de carrés blancs et noirs.

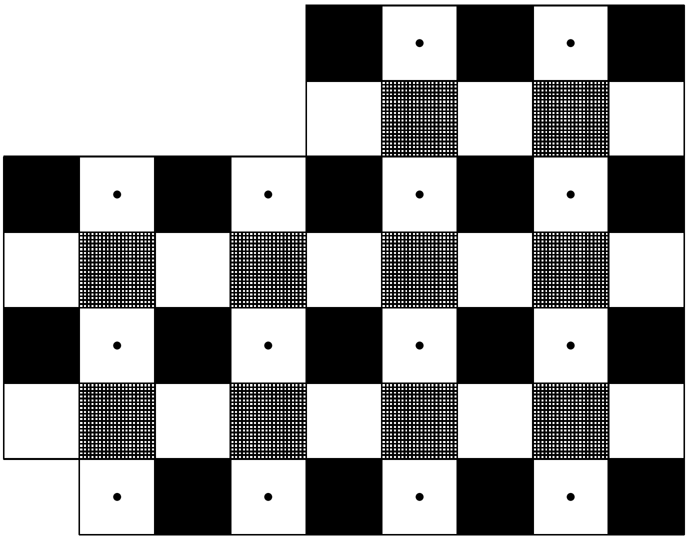


FIGURE 1.

La figure 1 donne un exemple de polyomino temperlien, pour lequel le carré de base (qui ne fait pas partie du polyomino) est en bas à gauche. Sur cette figure, les carrés noirs sont de type B_1, les carrés gris de type B_0, les carrés blancs pointés de type W_1 et enfin les carrés blancs de type W_0.

Soit P un polyomino temperlien, et P' le polyomino pair dont est issu P. On note $\mathbf{B}_1(P)$ le graphe dont les sommets sont les carrés de type B_1 contenus dans P', deux sommets étant reliés par une arête si les centres des carrés sont à distance 2. Remarquons que d est un sommet du graphe $\mathbf{B}_1(P)$, et qu'à chaque arête de $\mathbf{B}_1(P)$ est associé un unique carré blanc de P. Le dual du graphe $\mathbf{B}_1(P)$ est le graphe $\mathbf{B}_0(P)$ dont les sommets sont les carrés de type B_0 dans P.

On appelle *domino* un rectangle 1×2 ou 2×1 dont les sommets sont des points de $\mathbb{Z}^2$. Il est facile de vérifier qu'un polyomino temperlien est toujours pavable par dominos. Le théorème suivant est dû à Temperley [31] dans le cas d'un rectangle (voir [5] et [18] pour des généralisations incluant celle-ci).

THÉORÈME 2.3. — *Les pavages par dominos de P sont en bijection avec les arbres couvrants du graphe $\mathbf{B}_1(P)$.*

De manière intuitive, si l'on part d'un arbre couvrant dont on oriente les arêtes en partant du point de base d, on construit le pavage par dominos correspondant de la manière suivante. On « jette » pour chaque arête de l'arbre un domino dont le carré noir est le sommet terminal de l'arête et le carré blanc est le carré associé à l'arête. Pour compléter le pavage, il suffit ensuite de « boucher les trous », ou, ce qui revient au même, de jeter pareillement les dominos le long de l'arbre dual. La figure 2 donne un exemple de pavage par dominos du polyomino de la figure 1 et de l'arbre associé dont les sommets sont les points de $\mathbf{B}_1(P)$.

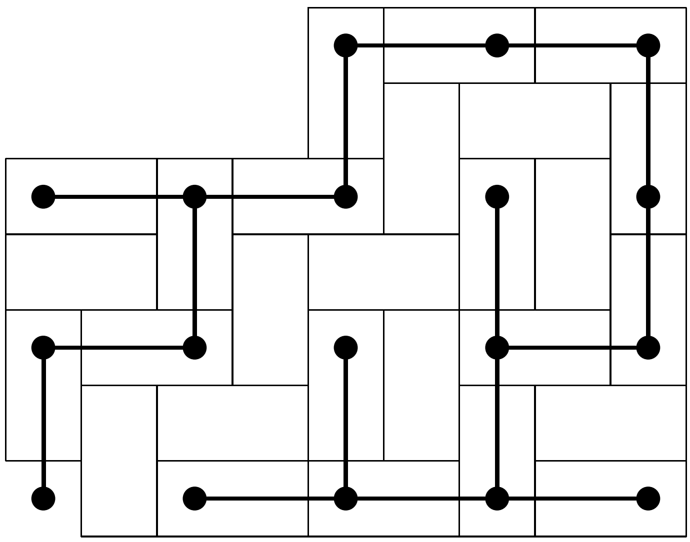


FIGURE 2.

Considérons maintenant, pour un entier pair n, le polyomino temperlien P rectangulaire défini comme la réunion des carrés unité centrés aux points de $\{1, 2, \ldots, 2n+1\} \times \{-n, -n+1, \ldots, n-1, n\}$, à l'exception du carré d centré en $(1, 0)$. Soit Q l'ensemble obtenu à partir de P en enlevant un carré noir b adjacent à la frontière de P (donc de type B_1) ainsi qu'un carré blanc w (qu'on peut voir comme une arête de $\mathbf{B}_1(P)$).

LEMME 2.4 ([17]). — *Les pavages par dominos de Q sont en bijection avec les arbres couvrants de $\mathbf{B}_1(P)$ pour lesquels l'arc de b à d contient l'arête w.*

Ce lemme nous permet d'esquisser la méthode d'approche du Théorème 2.1 [17]. Nous voulons estimer la probabilité que dans l'arbre couvrant uniforme sur $\mathbb{N} \times \mathbb{Z}$, l'arc joignant 0 à ∞ contienne l'arête w fixée. Le lemme précédent nous permettra d'estimer la probabilité que pour l'arbre couvrant uniforme dans une boîte rectangulaire arbitrairement grande, l'arc joignant le point de base à un autre point de la frontière contienne l'arête w. Cette probabilité s'exprimera d'après le Lemme 2.4 et le Théorème 2.3 comme le rapport du nombre de pavages par dominos de Q sur le nombre de pavages par dominos de P. Dans les sections suivantes, nous expliquons comment on estime le nombre de pavages par dominos d'un polyomino temperlien. Ceci ne s'applique pas directement à Q qui n'est pas temperlien, mais la méthode peut être adaptée.

2.3. Matrice de Kasteleyn et fonction de couplage

Dans ce paragraphe, nous considérons un polyomino temperlien P et nous notons M l'ensemble des carrés de P. L'ensemble M est muni d'une structure de graphe évidente, où deux carrés sont adjacents ssi ils ont un côté en commun. La *matrice de Kasteleyn* de P est la matrice symétrique $K = (K(v,v'), v, v' \in M)$ telle que $K(v,v') = 0$ si v et v' ne sont pas adjacents et, pour $v \in M \cap (W_0 \cup W_1)$ et $v' \in M$,

$$K(v,v') = \begin{cases} 1 & \text{si } v' = v + (1,0) \\ i & \text{si } v' = v + (0,1) \\ -1 & \text{si } v' = v + (-1,0) \\ -i & \text{si } v' = v + (0,-1). \end{cases}$$

Kasteleyn [14] a montré (avec une définition différente de K) que le nombre de pavages par dominos de P est la racine carrée du module du déterminant de K. En général, sauf dans le cas du rectangle où une formule explicite existe, il est cependant difficile d'utiliser ce résultat pour estimer le nombre de pavages par dominos de P.

La *fonction de couplage* de P est par définition la matrice inverse $C = K^{-1}$.

Proposition 2.5 ([15]). — *Si u, v sont deux carrés adjacents de P, la probabilité qu'un pavage aléatoire uniforme de P contienne le domino $\{u, v\}$ est égale à $|C(u,v)|$.*

Plus généralement [15], la probabilité qu'un pavage aléatoire uniforme contienne une famille fixée de dominos disjoints s'exprime comme la valeur absolue du déterminant d'une sous-matrice de C.

Nous verrons dans la partie suivante que si l'on se donne une suite de polyominos temperliens P_ε de $\varepsilon\mathbb{Z}^2$ qui approchent une région du plan, il est possible d'obtenir des informations précises sur le comportement asymptotique des fonctions de couplage C_ε associées. Afin de mieux comprendre l'origine de ces informations, donnons quelques propriétés simples de la fonction de couplage.

Si on énumère les éléments de M en écrivant successivement les carrés de type W_0, W_1, B_0 et enfin B_1, la matrice K est de la forme

$$K = \begin{pmatrix} 0 & 0 & K_1 & iK_2 \\ 0 & 0 & iK_3 & K_4 \\ K_1^t & iK_3^t & 0 & 0 \\ iK_2^t & K_4^t & 0 & 0 \end{pmatrix}$$

où les matrices K_1, K_2, K_3, K_4 sont réelles. Il est facile de voir que l'inverse de K doit être aussi de cette forme. En particulier, $C(v,v') = 0$ si v et v' sont tous les deux blancs, ou tous les deux noirs, $C(v,v')$ est réel quand $v' = v + (1,0)$ modulo 2 et imaginaire pur quand $v' = v + (0,1)$ modulo 2.

Par définition de C, on a $KC(v,v') = \delta_v(v')$, où δ_v est la fonction indicatrice du singleton $\{v\}$. Notons K^* la matrice conjuguée de K. On vérifie facilement que la matrice K^*K agit sur les fonctions définies sur $\mathbf{B}_0(P)$, et que pour une telle fonction f on a pour tout $v \in \mathbf{B}_0(P)$,

$$(2)\quad K^*Kf(v) = 4f(v) - f(v+(2,0)) - f(v+(-2,0)) - f(v+(0,2)) - f(v+(0,-2)).$$

Cette formule est vraie même si les quatre voisins (dans B_0) de v ne sont pas tous dans $\mathbf{B}_0(P)$, auquel cas il faut affecter la valeur 0 à la fonction f aux points extérieurs à $\mathbf{B}_0(P)$. La formule précédente s'interprète en disant que $K^*K = \Delta$ est (4 fois) le Laplacien discret sur $\mathbf{B}_0(P)$, avec conditions de Dirichlet au bord. Ce qui précède reste vrai si on remplace $\mathbf{B}_0(P)$ par $\tilde{\mathbf{B}}_1(P) = \mathbf{B}_1(P)\backslash\{d\}$, mais la formule (2) doit être modifiée si v est adjacent à la frontière de P. Dans ce cas, le terme de droite de (2) est $f(v)$ multiplié par le nombre de voisins de v dans $\mathbf{B}_1(P)$ moins la somme des valeurs de f en ces voisins (en prenant $f(d) = 0$).

En appliquant K^* à l'égalité $KC(v,v') = \delta_v(v')$, on trouve pour $v \in W_0 \cup W_1$,

$$\Delta C(v,\cdot) = \delta_{v+(1,0)} - \delta_{v+(-1,0)} - i\delta_{v+(0,1)} + i\delta_{v+(0,-1)}.$$

Fixons $v \in W_0 \cap M$. La restriction, notée f, de $C(v,\cdot)$ à $\mathbf{B}_0(P)$ est à valeurs réelles, et l'égalité précédente montre que cette fonction est harmonique (au sens discret) sur $\mathbf{B}_0(P)\backslash\{v-(1,0), v+(1,0)\}$, avec conditions frontière de type Dirichlet. En revanche la restriction, notée ig, de $C(v,\cdot)$ à $\tilde{\mathbf{B}}_1(P)$ est à valeurs imaginaires pures et est harmonique sur $\tilde{\mathbf{B}}_1(P)\backslash\{v+(0,1), v+(0,-1)\}$. Les conditions frontière sont maintenant de type Neumann : la valeur de g en un point frontière v' est la moyenne de ses valeurs aux voisins de v' dans $\mathbf{B}_1(P)$, en prenant $g(d) = 0$.

De plus, on peut dire que g est la conjuguée harmonique de f, au sens où on a les équations de Cauchy-Riemann discrètes

$$\begin{aligned} \partial_x f(v') &= \partial_y g(v') && \text{pour } v' \in (M \cap W_0)\backslash\{v\} \\ \partial_y f(v') &= -\partial_x g(v') && \text{pour } v' \in M \cap W_1, \end{aligned}$$

avec la notation $\partial_x f(v) = f(v+(1,0)) - f(v+(-1,0))$, $\partial_y f(v) = f(v+(0,1)) - f(v+(0,-1))$.

On a des résultats analogues pour la fonction $C(v,\cdot)$ lorsque $v \in W_1$.

2.4. Comportement asymptotique de la fonction de couplage

Soit U un polygone rectiligne de $\mathbb{R}^2$ (U est un domaine simplement connexe et sa frontière est composée de segments parallèles aux axes), et soit $d_0 \in \partial U$. Considérons pour chaque $\varepsilon > 0$ un polyomino temperlien P_ε de $\varepsilon\mathbb{Z}^2$, et supposons que ces polyominos convergent vers U de la manière évidente (en particulier, P_ε a autant de coins que U, et chaque coin de P_ε converge vers le coin correspondant de U). On suppose aussi que les points de base d_ε de P_ε convergent vers le point d_0.

Notons C_ε la fonction de couplage de P_ε et M_ε l'ensemble des carrés de P_ε.

THÉORÈME 2.6 ([16]). — *Soient, pour chaque* $\varepsilon > 0$, u_ε, v_ε, w_ε, x_ε *des éléments de* M_ε *appartenant respectivement à* W_0, W_1, B_0, B_1, *et convergeant respectivement vers des points distincts* u, v, w, $x \in U$. *Alors,*

$$\lim_{\varepsilon\to 0} \frac{1}{\varepsilon} C_\varepsilon(u_\varepsilon, w_\varepsilon) = \operatorname{Re} F_0(u,w)$$
$$\lim_{\varepsilon\to 0} \frac{1}{\varepsilon} C_\varepsilon(u_\varepsilon, x_\varepsilon) = i \operatorname{Im} F_0(u,x)$$
$$\lim_{\varepsilon\to 0} \frac{1}{\varepsilon} C_\varepsilon(v_\varepsilon, w_\varepsilon) = \operatorname{Re} F_1(v,w)$$
$$\lim_{\varepsilon\to 0} \frac{1}{\varepsilon} C_\varepsilon(v_\varepsilon, x_\varepsilon) = i \operatorname{Im} F_1(v,x)$$

où les fonctions F_0 *et* F_1 *sont caractérisées par les propriétés suivantes.*

Pour $v \in U$, *la fonction* $F_0(v,\cdot)$ *est méromorphe sur* U, *sa partie réelle s'annule sur* ∂U, *elle a un zéro en* $z = d_0$, *un pôle simple de résidu* $\frac{1}{\pi}$ *en* $z = v$ *et pas d'autre pôle sur* $\bar{U}$.

De même, la fonction $F_1(v,\cdot)$ *est méromorphe sur* U, *sa partie imaginaire s'annule sur* ∂U, *elle a un zéro en* $z = d_0$, *un pôle simple de résidu* $\frac{1}{\pi}$ *en* $z = v$ *et pas d'autre pôle sur* $\bar{U}$.

REMARQUE — Un résultat analogue est vrai pour des ouverts U plus généraux : voir [16].

Les propriétés des fonctions F_0 et F_1, et le fait qu'on obtienne tantôt $\operatorname{Re} F$ tantôt $\operatorname{Im} F$, viennent directement des propriétés des fonctions de couplage décrites dans la partie précédente. Donnons rapidement l'idée de la preuve du théorème. On note $\mathbf{B}_0'(P_\varepsilon)$ le graphe obtenu en ajoutant à $\mathbf{B}_0(P_\varepsilon)$ tous les carrés noirs (de type B_0) qui sont adjacents à un carré blanc à la frontière de P_ε, et les arêtes reliant deux carrés noirs si le carré blanc situé entre eux est dans P_ε. Soit alors $G_\varepsilon(w,w')$ la fonction de Green de $\mathbf{B}_0'(P_\varepsilon)$: pour $w \in \mathbf{B}_0(P_\varepsilon)$, la fonction $G_\varepsilon(w,\cdot)$ vérifie $\Delta G_\varepsilon(w,\cdot) = \delta_w$ sur $\mathbf{B}_0(P_\varepsilon)$, et $G_\varepsilon(w,w') = 0$ pour $w' \in \mathbf{B}_0'(P_\varepsilon)\backslash\mathbf{B}_0(P_\varepsilon)$. Par abus de notation, écrivons $\operatorname{Re} C_\varepsilon(u_\varepsilon,\cdot)$ pour la restriction de $C_\varepsilon(u_\varepsilon,\cdot)$ à $\mathbf{B}_0(P_\varepsilon)$ (qui est à valeurs réelles) et prolongeons-la à $\mathbf{B}_0'(P_\varepsilon)$ en lui donnant la valeur 0 sur $\mathbf{B}_0'(P_\varepsilon)\backslash\mathbf{B}_0(P_\varepsilon)$. Nous avons vu que, sur $\mathbf{B}_0(P_\varepsilon)$,

$$\Delta \operatorname{Re} C_\varepsilon(u_\varepsilon,\cdot) = \delta_{u_\varepsilon+\varepsilon} - \delta_{u_\varepsilon-\varepsilon}.$$

Puisque $\operatorname{Re} C_\varepsilon(u_\varepsilon,\cdot)$ s'annule à la frontière, il en découle que

$$\operatorname{Re} C_\varepsilon(u_\varepsilon,w) = G_\varepsilon(u_\varepsilon+\varepsilon,w) - G_\varepsilon(u_\varepsilon-\varepsilon,w).$$

On peut alors déduire la première convergence du théorème de résultats sur le comportement asymptotique des fonctions de Green discrètes.

Les fonctions de couplage asymptotiques F_0 et F_1 possèdent des propriétés d'invariance conforme qui sont très importantes pour les applications. Posons $F_+^U = F_0 + F_1$

et $F_-^U = F_0 - F_1$. Alors, $F_+^U(v,z)$ est méromorphe en v et en z, et $F_-^U(v,z)$ est méromorphe en z et anti-méromorphe en v. De plus, si $f : V \longrightarrow U$ est une bijection conforme envoyant le point de base de V sur le point de base de U, on a :

$$F_+^V(v,z) = f'(v)\, F_+^U(f(v), f(z)),$$
$$F_-^V(v,z) = \overline{f'(v)}\, F_-^U(f(v), f(z)).$$

2.5. L'asymptotique du nombre de pavages par dominos

Dans ce paragraphe, nous décrivons brièvement un résultat important de [17] donnant l'asymptotique du logarithme du nombre $N(P_\varepsilon)$ de pavages par dominos de P_ε (sous les hypothèses de la partie précédente). Ce résultat ne peut pas être appliqué directement à la preuve du Théorème 2.1, qui nécessite d'estimer le rapport $N(Q_\varepsilon)/N(P_\varepsilon)$ où Q_ε n'est pas temperlien. Cependant, les techniques impliquées dans cette preuve sont proches de celles que nous décrivons très succinctement ci-dessous.

Nous notons A_ε l'aire de P_ε, $\mathrm{perim}(P_\varepsilon)$ son périmètre, et V le nombre de sommets du polygone V.

Théorème 2.7 ([17]). — *Soit $N(P_\varepsilon)$ le nombre de pavages par dominos de P_ε. Alors,*

$$\log N(P_\varepsilon) = \frac{c_0 A_\varepsilon}{\varepsilon^2} + \frac{c_1\,\mathrm{perim}(P_\varepsilon)}{\varepsilon} - \frac{1}{2}\log\frac{1}{\varepsilon} - \frac{V-4}{36} h(\varepsilon) + c_2(U) + o(1),$$

où $c_0 = \frac{G}{\pi}$, si $G = 1 - \frac{1}{3^2} + \frac{1}{5^2} - \dots$ est la constante de Catalan, $c_1 = \frac{G}{2\pi} + \frac{\log(\sqrt{2}-1)}{4}$, $c_2(U)$ est une constante dépendant seulement de U, et la fonction h est indépendante de U et vérifie $h(\varepsilon) = \log\frac{1}{\varepsilon} + o(\log\frac{1}{\varepsilon})$.

Remarque. — Le premier terme du développement ci-dessus est essentiellement contenu dans le travail de Burton et Pemantle [5].

Dans le cas particulier où U est un rectangle, le développement du théorème peut être obtenu à partir des expressions explicites dues à Kasteleyn [13] et Temperley et Fischer [32] (voir [10] pour une dérivation du développement dans ce cas). Pour se ramener au cas du rectangle, on utilise une technique de découpage. Considérons un segment (vertical ou horizontal) γ' qui traverse U, touche ∂U seulement en ses points de départ et d'arrivée, et évite le point de base d_0 de U. Pour chaque $\varepsilon > 0$, considérons aussi une bande γ_ε de largeur ε formée de carrés de P_ε, qui traverse P_ε en restant à une distance $O(\varepsilon)$ de γ'. On peut imposer que les carrés noirs contenus dans γ_ε soient de type B_0, et que si l'une ou l'autre extrémité de γ' est à un coin concave de U, l'extrémité correspondante de γ_ε soit au coin correspondant de P_ε.

La longueur de γ_ε doit alors être un multiple impair de ε. Si nous retirons la bande γ_ε de P_ε, il reste deux polyominos disjoints. Celui, noté P_1, qui contient le carré de base de P_ε est temperlien. L'autre, noté P_2, devient temperlien si on lui enlève un seul carré s de type B_1 adjacent à l'un des points terminaux de γ_ε. Remarquons que la réunion $\gamma_\varepsilon \cup s$ a un seul pavage par dominos. Le nombre de pavages par dominos de P_ε est égal au produit du nombre de pavages de P_1 et du nombre de pavages de P_2, divisé par la probabilité que le pavage de $\gamma_\varepsilon \cup s$ apparaisse dans un pavage de P_ε tiré au hasard uniformément.

Nous pouvons répéter cette procédure à partir de P_1 et P_2 de manière à découper le polyomino initial en des morceaux de plus en plus simples. Au bout d'un nombre fini d'itérations on arrive à des rectangles temperliens pour lesquels on peut utiliser la formule exacte. Tout le problème est donc d'estimer la probabilité que dans un pavage aléatoire uniforme de P_ε, apparaisse le pavage de $\gamma_\varepsilon \cup s$. Notons $a_1, \dots, a_N$ les dominos qui forment ce pavage. Les a_i sont mis bout à bout (à partir de la frontière) sauf pour a_N qui est perpendiculaire aux autres. La probabilité que tous les a_i soient présents s'écrit comme le produit

$$\prod_{j=1}^{N} P(a_j \mid a_1, \dots, a_{j-1})$$

des probabilités conditionnelles que a_j soit présent sachant que $a_1, \dots, a_{j-1}$ le sont. Pour j fixé, cette dernière probabilité conditionnelle est la probabilité que a_j soit présent dans un pavage aléatoire de la région $P_\varepsilon^{(j)} := P_\varepsilon \backslash \{a_1, \dots, a_{j-1}\}$, qui est elle-même un polyomino temperlien. Soit $C_\varepsilon^{(j)}$ la fonction de couplage de $P_\varepsilon^{(j)}$, et soient $u_\varepsilon^{(j)}$, $v_\varepsilon^{(j)}$ les deux carrés qui composent le domino a_j. Alors, la probabilité que a_j soit présent dans un pavage aléatoire de $P_\varepsilon^{(j)}$ est d'après la Proposition 2.5 égale à $|C_\varepsilon^{(j)}(u_\varepsilon^{(j)}, v_\varepsilon^{(j)})|$. Le comportement asymptotique de ces quantités peut être décrit en termes des fonctions de couplage asymptotiques F_0, F_1 de la partie 4, correspondant à la région U privée d'une coupure le long (d'une partie) du segment γ'. À l'aide d'une analyse détaillée de ces asymptotiques et d'une utilisation intensive des propriétés d'invariance conforme, cela conduit à des estimations suffisamment précises pour établir le Théorème 2.7.

RÉFÉRENCES

[1] L. V. Ahlfors – *Conformal Invariants, Topics in Geometric Function Theory.* McGraw-Hill, New York 1973

[2] K. Burdzy, G. F. Lawler – *Non-intersection exponents for random walk and Brownian motion. Part I : Existence and an invariance principle.* Probab. Th. Rel. Fields **84** (1990), 393-410.

[3] K. Burdzy, G. F. Lawler – *Non-intersection exponents for random walk and Brownian motion. Part II : Estimates and applications to a random fractal.* Ann. Probab. **18** (1990), 981-1009.

[4] K. Burdzy, W. Werner – *No triple point of planar Brownian motion is accessible.* Ann. Probab. **24** (1996), 125-147.

[5] R. Burton, R. Pemantle – *Local characteristics, entropy and limit theorems for spanning trees and domino tilings via transfer impedances.* Ann. Probab. **21** (1993), 1329-1371.

[6] M. Cranston, T. Mountford – *An extension of a result by Burdzy and Lawler.* Probab. Th. Rel. Fields **89** (1991), 487-502.

[7] B. Duplantier – *Loop-erased self-avoiding walks in two dimensions : exact critical exponents and winding numbers.* Physica A **191** (1992), 516-522.

[8] B. Duplantier – *Random walks and quantum gravity in two dimensions.* Phys. Rev. Lett. **81** (1998), 5489-5492.

[9] B. Duplantier – *Two-dimensional copolymers and exact conformal multifractality.* Phys. Rev. Lett. **82** (1999), 880-883.

[10] B. Duplantier, F. David – *Exact partition functions and correlation functions of multiple Hamiltonian walks on the Manhattan lattice.* J. Stat. Phys. **51** (1988), 327-434.

[11] B. Duplantier, K.-H. Kwon – *Conformal invariance and intersections of random walks.* Phys. Rev. Lett. **61** (1988), 2514-2517.

[12] B. Duplantier, G. F. Lawler, J. F. Le Gall, T. J. Lyons – *The geometry of the Brownian curve.* Bull. Sci. math. **117** (1993), 91-106.

[13] P. Kasteleyn – *The statistics of dimers on a lattice. I. The number of dimer arrangements on a quadratic lattice.* Physica **27** (1961), 1209-1225.

[14] P. Kasteleyn – *Graph theory and crystal physics.* In : *Graph Theory and Theoretical Physics.* Academic Press, London 1967

[15] R. Kenyon – *Local statistics of lattice dimers.* Ann. Inst. Henri Poincaré Probab. Stat. **33** (1997), 591-618.

[16] R. Kenyon – *Conformal invariance of domino tilings.* Preprint.

[17] R. Kenyon – *The asymptotic determinant of the discrete Laplacian.* Preprint.

[18] R. Kenyon, J. Propp, D. Wilson – *Trees and matchings.* Preprint.

[19] G. F. Lawler – *A self-avoiding random walk.* Duke Math. J. **47** (1980), 655-693.

[20] G. F. Lawler – *Intersections of Random Walks.* Birkhäuser, Boston 1991

[21] G. F. Lawler – *Hausdorff dimension of cut points for Brownian motion.* Electron. J. Probab. **1** (1996), no 2.

[22] G. F. Lawler – *The dimension of the frontier of planar Brownian motion.* Electronic Commm. Probab. **1** (1996), no 5.

[23] G. F. Lawler – *Loop-erased random walk.* In : *Perplexing Problems in Probability.* Birkhäuser, Boston 1999

[24] G. F. Lawler, W. Werner – *Intersection exponents for planar Brownian motion.* Preprint.

[25] G. F. Lawler, W. Werner – *Universality for conformally invariant intersection exponents.* Preprint.

[26] G. F. Lawler, O. Schramm, W. Werner – *Values of Brownian intersection exponents I : Half-plane exponents.* Preprint.

[27] S. N. Majumdar – *Exact fractal dimension of the loop-erased self-avoiding walk in two dimensions.* Phys. Rev. Lett. **68** (1992), 2329-2331.

[28] B. B. Mandelbrot – *The Fractal Geometry of Nature.* Freeman 1982.

[29] R. Pemantle – *Choosing a spanning tree for the integer lattice uniformly.* Ann. Probab. **19** (1991), 1559-1574.

[30] O. Schramm – *Scaling limits of loop-erased random walks and uniform spanning trees.* Preprint.

[31] H. Temperley – *Combinatorics : Proceedings of the British Combinatorial Conference 1973.* London Math. Soc. Lecture Notes Series **13** (1974), 202-204.

[32] H. Temperley, M. Fisher – *The dimer problem in statistical mechanics – an exact result.* Phil. Mag. **6** (1961), 1061-1063.

[33] W. Werner – *Asymptotic behaviour of disconnection and intersection exponents.* Probab. Th. Rel. Fields **108** (1997), 131-152.

Jean-François LE GALL

DMA – École normale supérieure
45, rue d'Ulm
F-75230 PARIS Cedex 05
E-mail : `legall@dmi.ens.fr`

Séminaire BOURBAKI Novembre 1999
52e année, 1999-2000, nO 867, p. 53 à 72

LA CORRESPONDANCE DE McKAY

by **Miles REID**

1. COMMENT C'EST

1.1. Model case: the binary dihedral group BD_{4n}

For $G \subset \mathrm{SL}(2, \mathbb{C})$ a finite group, the quotient variety $X = \mathbb{C}^2/G$ is called a *Klein quotient singularity.* I draw the quotient map $\pi\colon \mathbb{C}^2 \to X$ and the minimal resolution of singularities $Y \to X$ together in the diagram:

$$\begin{array}{ccc} & & \mathbb{C}^2 \\ & & \downarrow \pi \\ Y & \xrightarrow{\varphi} & X \end{array}$$

This situation has been well studied, since Klein around 1870 and Coxeter and Du Val in the 1930s: the subgroup G is classified as cyclic, binary dihedral or a binary group corresponding to one of the Platonic solids; the quotient singularity is a hypersurface $X \subset \mathbb{C}^3$ with defining equation one of a list of simple functions. The resolution Y is a surface with $K_Y = \varphi^* K_X$, and the exceptional locus $\varphi^{-1}(0) \subset Y$ of the resolution consists of a bunch of -2-curves E_i (that is, $E_i \cong \mathbb{P}^1_{\mathbb{C}}$ and E_i has self-intersection $E_i^2 = -2$), and the intersection $E_i E_j$ is given by one of the Dynkin diagrams A_n, D_n, E_6, E_7, E_8. To avoid writing out lists, let us simply discuss the binary dihedral group

$$G = \mathrm{BD}_{4n} = \langle \alpha, \beta \rangle\,, \quad \text{where} \quad \alpha = \begin{pmatrix} \varepsilon & 0 \\ 0 & \varepsilon^{-1} \end{pmatrix}, \quad \beta = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

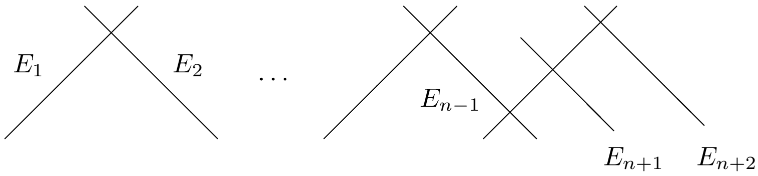
where $\varepsilon = \exp \frac{2\pi i}{2n}$. If u, v are coordinates on $\mathbb{C}^2$, the G-invariant polynomials are

$$\mathbb{C}[x, y, z]/(z^2 - yx^2 + 4y^{n+1}), \quad \text{where} \quad x = u^{2n} + v^{2n}, y = u^2v^2, z = uv(u^{2n} - v^{2n});$$

thus the quotient variety is the singularity $X : (z^2 = yx^2 - 4y^{n+1}) \subset \mathbb{C}^3$ of type D_{n+2}, and the quotient morphism $\pi\colon (u, v) \mapsto (x, y, z)$. The resolution of singularities

$Y \to X$ has exceptional locus consisting of -2-curves $E_1, \ldots, E_{n+2}$ forming the D_{n+2} configuration:

(1.1)

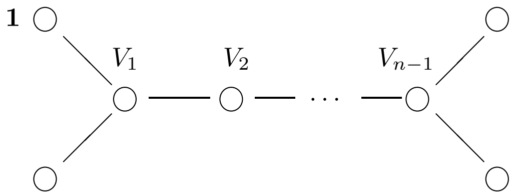


The classical McKay correspondence begins in the late 1970s with the observation that the same graph arises in connection with the representation theory of G. For a group G and a given representation Q, the *McKay graph* (or McKay quiver) has a node for each irreducible representation, and an edge $V \to V'$ whenever V' is a direct summand of $V \otimes Q$. In our case, BD_{4n} has the 2-dimensional representations

$$V_i \cong \mathbb{C}^2, \quad \text{with action} \quad \alpha = \begin{pmatrix} \varepsilon^i & 0 \\ 0 & \varepsilon^{-i} \end{pmatrix}, \quad \beta = \begin{pmatrix} 0 & 1 \\ (-1)^i & 0 \end{pmatrix} \quad \text{for } i = 0, \ldots, n.$$

This is irreducible for $0 < i < n$, and splits into 2 eigenlines when $i = 0$ or n. The inclusion $G \subset \mathrm{SL}(2, \mathbb{C})$ provides the *given* representation $Q = V_1$. It is a simple exercise [Homework] to write down the action of G on a basis $\{e_i \otimes e'_j\}$ of $Q \otimes V_i$ to get $V_i \otimes Q = V_{i-1} \oplus V_{i+1}$ for $0 < i < n$, so that the McKay graph of BD_{4n} is the extended Dynkin diagram $\widetilde{D}_{n+2}$:

(1.2) $\widetilde{D}_{n+2}$



Here **1** is the trivial 1-dimensional representation.

This example, and the other $\mathrm{SL}(2, \mathbb{C})$ cases observed by McKay, suggest that there is a one-to-one correspondence between the components of the exceptional locus of $Y \to X$ in (1.1) and the nontrivial irreducible representations of $G \subset \mathrm{SL}(2, \mathbb{C})$ in (1.2). This talk explains this coincidence in several different ways, and discusses higher dimensional generalisations.

1.2. General assumption

I use the following diagram throughout:

$$\begin{array}{ccc} & & M \\ & & \downarrow{\scriptstyle \pi} \\ Y & \xrightarrow{\varphi} & X \quad = M/G \end{array} \tag{1.3}$$

Here M is a quasiprojective algebraic manifold with $K_M = 0$ and G a finite automorphism group of M that acts trivially on a global basis $s_M \in H^0(K_M)$. The object of study is the quotient variety $X = M/G$ and its resolutions $Y \to X$, sometimes assumed to have $K_Y = 0$. An important motivating case is a finite subgroup $G \subset \mathrm{SL}(3,\mathbb{C})$ acting on $M = \mathbb{C}^3$.

1.3. Definition–Reassurance

The quotient varieties $X = M/G$ occuring here are singular. The theory of minimal models of higher dimensional algebraic varieties (Mori theory) has a whole battery of definitions that deal systematically with singular varieties; here I only need one small item: the orbifolds X here have trivial canonical class $K_X = 0$ (or trivial Serre–Grothendieck dualising sheaf $\omega_X = \mathcal{O}_X$). In concrete terms, this means the following: X is a complex n-fold (algebraic or analytic variety), nonsingular in codimension 1, and its nonsingular locus $\operatorname{NonSing} X$ has an everywhere nondegenerate holomorphic n-form s_X (deduced from s_M). So s_X is a complex volume element at every nonsingular point of X, or in other words, it is a global basis of $\Omega^n_{\operatorname{NonSing} X}$. A resolution of singularities $\varphi\colon Y \to X$ is *crepant* if $K_Y = \varphi^* K_X$ or $\omega_Y = \varphi^*\omega_X$, which simply means that Y is a nonsingular n-fold with $K_Y = 0$ or $\omega_Y = \Omega^n_Y = \mathcal{O}_Y \cdot s_Y$, where $s_Y = \varphi^* s_X$. More generally, an arbitrary proper birational map $\varphi\colon V \to X$ has a *discrepancy divisor* $\Delta_\varphi = \sum a_i E_i$ defined by $K_V = \varphi^* K_X + \sum a_i E_i$ with $a_i \geq 0$; a divisor E_i is *crepant* if $a_i = 0$. The discrepancy Δ_φ is the divisor of zeros on V of the basic n-form s_X on X, generalising the divisor of zeros of the Jacobian determinant; in Mori theory, it measures how far V is from minimal.

1.4. Summary and slogan

I start with a preview of different approaches to the McKay correspondence, which are treated in more detail in later sections. Each of these approaches gives a result in the case of a finite subgroup $G \subset \mathrm{SL}(3,\mathbb{C})$ acting on $M = \mathbb{C}^3$.

(1) Gonzalez-Sprinberg and Verdier sheaves: the first direct link from the representation theory of G to the geometry of the resolution $Y \to X$ was the work of Gonzalez-Sprinberg and Verdier [GSpV]: for a Kleinian subgroup $G \subset \mathrm{SL}(2,\mathbb{C})$, they constructed *sheaves $\mathcal{F}_\rho$ on Y, indexed by the irreducible representations of G, whose first Chern classes base the cohomology of Y.*

(2) String theory: the first hint of a McKay correspondence in higher dimensions comes from work of the string theorists Dixon, Harvey, Vafa and Witten [DHVW] around 1985: if $G \subset \mathrm{SL}(3,\mathbb{C})$ and $Y \to X = \mathbb{C}^3/G$ is a crepant resolution of the quotient $\mathbb{C}^3/G$, *the Euler number of Y equals the number of conjugacy classes of G (or the number of its irreducible representations).*

(3) Explicit methods: the finite subgroups $G \subset \mathrm{SL}(3,\mathbb{C})$ are classified, and work in the early 1990s of Roan, Ito, Markushevich and others proved case-by-case *the*

existence of crepant resolutions, and the validity of the formula of [DHVW] for the Betti numbers of Y.

(4) Valuation theory: for a finite subgroup $G \subset \mathrm{SL}(n, \mathbb{C})$, the paper [IR] shows that G has a grading by *age*, analogous to the weight grading in Hodge theory, and proves that *the conjugacy classes of* junior *elements $g \in G$ (elements of age 1) correspond one-to-one with the crepant divisors of a resolution* (more precisely, their discrete valuations). This result holds for any $G \subset \mathrm{SL}(n, \mathbb{C})$ and is intrinsic, classification-free; but for $n \geq 4$ it only gives a small part of a McKay correspondence (so far).

(5) Nakamura's G-Hilbert scheme: a resolution of singularities $Y \to X$, even if it is a Mori minimal model theory, is not at all unique. Moreover, if $X = M/G$, the construction of a resolution Y need not have much to do with G. In 1995, Nakamura made the revolutionary suggestion that in many interesting cases, *the G-Hilbert scheme is a preferred resolution Y of X* (see [IN2], [N], [R]). When this holds, Y is a "very good" moduli space over M, and the general yoga of moduli suggests that there should be a "tautological" treatment of the geometry of Y (comparable to the cohomology of Grassmann varieties).

(6) Fourier–Mukai transform: the derived category $\mathrm{D}(V)$ of coherent sheaves on a variety V (considered up to isomorphism of triangulated categories) can be used as a geometric characteristic of V, in place of K theory or cohomology. The Fourier–Mukai transform is a general method for constructing isomorphisms of derived categories (see [Mu], [O], [BO1], [Br], [BrM]). Bridgeland and others [BKR] have recently used this technique to prove that, if $Y = G\text{-Hilb}\, M$ is a crepant resolution, then $\mathrm{D}^G(M) = \mathrm{D}(Y)$. This implies the corresponding result in K theory.

(7) Motivic integration: the motivic integration of Batyrev, Denef and Loeser, and Kontsevich is a rigorous and comparatively simple mathematical trick that mimics some aspects of the path integrals of QFT. Very roughly, if $\varphi \colon Y \to X$ is a resolution of singularities, possibly far from minimal, with discrepancy divisor $K_Y - \varphi^* K_X = \sum a_i E_i$, the calculation amounts to defining the *stringy homology* of X by picking only $\frac{1}{a_i+1}$th of the homology of E_i. Quite remarkably, *this is well defined, agrees with the predictions of* [DHVW] *mentioned in (2) above, and provides an exact form of the homological McKay correspondence for finite subgroups $G \subset \mathrm{SL}(n, \mathbb{C})$.*

(8) Explicit methods (bis): for a finite group $G \subset \mathrm{SL}(3, \mathbb{C})$, the results of (6) (maybe also (7)) imply that Gonzalez-Sprinberg–Verdier sheaves $\mathcal{F}_\rho$ base the K theory of the resolution $Y \to X$, so that their Chern classes or Chern characters base the cohomology. Reworking this in explicit terms presents a treasure chest of delightful computational problems – already the Abelian cases lead to lovely pictures (compare [R], [CR], [C2]).

I believe that many other approaches to the McKay correspondence remain to be discovered, and many interrelations between the different approaches; this problem area is recommended to afficionados of noncommutative geometry, perverse sheaves, Gromov–Witten invariants, elliptic cohomology, Chow groups, etc. Here is an attempt to describe the subject in a single statement:

PRINCIPLE 1.1. — *Let M be an algebraic manifold, G a group of automorphisms of M, and $Y \to X$ a resolution of singularities of X. Then the answer to any well posed question about the geometry of Y is the G-equivariant geometry of M.*

I give two illustrations

I. If $G \subset \mathrm{SL}(n, \mathbb{C})$ acts on $\mathbb{C}^n$ and the quotient $X = \mathbb{C}^n/G$ has a crepant resolution $Y \to X$, the homology or K theory of Y is expected (or known) to be independent of Y. In this case, the principle says that the homology or K theory of Y is the representation theory of G (equal to the G-equivariant geometry of $\mathbb{C}^n$ because $\mathbb{C}^n$ is contractible).

II. Let M be a Calabi–Yau n-fold and G a group of automorphisms of M that acts trivially on Ω^n_M. The stringy homology of $X = M/G$ (see Sections 3 and 4) is well defined by [DL1]. The principle says that it must agree with the G-equivariant homology of M. (I expand on what this means in Section 4.)

Viewed as an orbifold or stack, $X = M/G$ contains M and the G action, and you can of course derive tautological question-and-answer pairs from this (this is often popular as a source of questions after the talk). The content of my slogan is that *the equivariant geometry of M already knows about the crepant resolution $Y \to X$.* Minimal models exist for surfaces by classical work, and for 3-folds by Mori theory (or by explicit methods). Minimal models of orbifolds by finite subgroups $G \subset \mathrm{SL}(3, \mathbb{C})$ provide infinitely many examples of local models of Calabi–Yau 3-folds; calculating their Betti numbers or K theory in a priori terms is in no sense a tautology. If you prefer to think of the singular X as the fundamental object, and not resolve it (a perfectly sensible alternative), the content is that X *has invariants that can be described from the orbifold M/G, but are birationally invariant* under appropriate conventions about resolutions.

2. AGE AND DISCREPANCY

Let $G \subset \mathrm{SL}(n, \mathbb{C})$ be a finite group; any element $g \in G$ has finite order r, say. For any such r, I choose at the outset a primitive rth root of 1, say $\exp \frac{2\pi i}{r}$. A choice of eigenbasis diagonalises the action of $g \in G$ on $M = \mathbb{C}^n$, giving

$$g = \mathrm{diag}(\varepsilon^{a_1}, \dots, \varepsilon^{a_n}) \quad \text{with } 0 \le a_i < r. \tag{2.1}$$

I write $g = \frac{1}{r}(a_1, \dots, a_n)$, possibly depending on the choices made. Now $\sum a_i \equiv 0 \mod r$ because $g \in \mathrm{SL}(n, \mathbb{C})$. Following [IR], define the *age* of g by $\operatorname{age} g = \frac{1}{r}\sum a_i$. As we will see, this is an analog of *weight* in Hodge theory; Denef and Loeser [DL2] refer to it by the long-winded but not inappropriate term *valuation theoretic weight*.

Clearly, $\operatorname{age} g$ is an integer in the range $[0, \ldots, n-1]$, and only the identity has age 0. The elements of age 1 are *junior*.

Junior elements of G give rise to crepant divisors of a resolution $V \to M/G$ by the following toric mechanism (for more details, and a picture, see [IR], 2.6–7). For $g \in G$ (not the identity), consider the a_i of (2.1), and suppose $(a_1, \ldots, a_n) \in \mathbb{Z}^n$ is primitive. The coordinate subspace corresponding to the x_i with $a_i = 0$ is the fixed locus $\operatorname{Fix} g$; it splits off as a direct product, and I assume that all $a_i > 0$ to short-cut some simple arguments. A useful example to bear in mind is when all the $a_i = 1$ (compare Example 4.1).

I view the integers $(a_1, \ldots, a_n)$ as *weights*. They define the grading $\operatorname{wt} x_i = a_i$ on the coordinate ring $\mathbb{C}[x_1, \ldots, x_n]$, or equivalently, the action $x_i \mapsto \lambda^{a_i} x_i$ of $\mathbb{C}^*$ on $M = \mathbb{C}^n$ that defines the weighted projective space

$$\mathbb{P}(a_1, \ldots, a_n) = (\mathbb{C}^n \setminus 0)/\mathbb{C}^*.$$

We obtain the weighted blowup $B_g \to M$ as the closed graph of the quotient map $M \dashrightarrow \mathbb{P}(a_1, \ldots, a_n)$; it has the exceptional divisor $B_g \supset E_g = \mathbb{P}(a_1, \ldots, a_n)$. Obviously g acts on B_g, and fixes E_g pointwise (because g acts on M as $\varepsilon \in \mathbb{C}^*$). Therefore $B_g \to B_g/\langle g\rangle$ is totally ramified along E_g.

THEOREM 2.1 ([IR], 2.6–7). — *Suppose that $V \to X$ is any resolution of singularities of the quotient $X = M/G$. Then V contains a divisor F_g rationally dominated by E_g under the rational map $B_g \to M \dashrightarrow V$. The discrepancy of F_g is given by $a_{F_g} = \operatorname{age} g - 1$, and in particular*

$$F_g \text{ is crepant} \iff g \text{ is junior.}$$

Every crepant divisor of any resolution V occurs thus.

Discussion of proof. — Write $X_g = M/\langle g\rangle$ for the partial quotient. Then $B_g/\langle g\rangle \to X_g$ is a partial resolution, with the single exceptional divisor E_g. An easy toric calculation gives the discrepancy of $E_g \subset B_g$ or $E_g \subset B_g/\langle g\rangle$ (compare [YPG], 4.8): the standard basis of Ω^n_M is $s_M = \mathrm{d}x_1 \wedge \cdots \wedge \mathrm{d}x_n$. For K_{B_g}, choose a Laurent monomial $y_1 = x^{\mathbf{m}}$ of weight 1 (recall that the a_i were coprime). Then y_1 is the defining equation of $E_g \subset B_g$ at a general point of E_g (away from all the coordinate hyperplanes), and y_1^r that of $E_g \subset B_g/\langle g\rangle$. Choosing Laurent monomials $y_2, \ldots, y_n$ forming a basis of the lattice of monomials of weight 0, we get that

$$s_{B_g} = \mathrm{d}y_1 \wedge \cdots \wedge \mathrm{d}y_n \in \Omega^n_{B_g}$$

is the required basis. The discrepancy is the exponent a in $s_M = (\text{unit}) \cdot y_1^a s_{B_g}$, and is determined by weighty considerations: s_M has weight $\sum a_i$ and s_{B_g} weight 1, so $a = \sum a_i - 1$. On the quotient $B_g/\langle g\rangle$ we only have y_1^r, so we get the stated discrepancy $\frac{1}{r}\bigl(\sum a_i - 1 - (r-1)\bigr) = \operatorname{age} g - 1$.

The quotient morphism $M \to X$ is a Galois cover with group G; a cyclic subgroup $\langle g \rangle$ corresponds to an intermediate cover $M \to M/\langle g \rangle = X_g \to X$. The reduction to a cyclic group is in terms of ramification theory; see [IR], 2.6–7. Roughly, over the general point of any exceptional divisor F of $V \to X$, the Galois extension of function fields $k(X) \subset k(M)$ forms a tower, starting with a cyclic ramified cover. For a crepant exceptional divisor, the cyclic ramification can be chased back up to a conjugacy class of junior elements $g \in G$.

Remark 2.2. — This argument works in all dimensions, but it only identifies the *divisors* of a crepant resolution Y, and thus only gives a basis of $H^2(Y, \mathbb{Q})$ or $H_{2n-2}(Y, \mathbb{Z})$ corresponding in McKay style to junior conjugacy classes of G. In 3 dimensions, we used Poincaré duality to bootstrap ourselves up to a basis of $H^*(Y, \mathbb{Q})$ in [IR]. Historically, this was the first intrinsic proof of the conjectured formula of [DHVW] for the Betti numbers of a crepant resolution.

As Brylinski [B] remarks (following Mumford), if $V \to X$ is any resolution, the group G can be viewed as the fundamental group of V minus the branch locus, so that an exceptional divisor F of a resolution V corresponds directly to a conjugacy class of G as a little anticlockwise loop around F; for crepant divisors, this is of course the same relation as in [IR]. But I don't know how to use this idea to get a well defined relation between, say, codimension 2 cycles of Y and age 2 conjugacy classes of G.

3. L'INNOMMABLE

This section is mainly for sociological and historical interest, but some harmless hilarity may derive from my garrulous display of incompetence and ignorance in physics.

A theoretical prediction of string theory: Fermionic strings propagate in 10-dimensional space-time. Indeed, a universe of any other dimension would have particles moving faster than the speed of light. Since this prediction, on the face of it, contradicts the empirically observed 4-dimensions of space-time, string theorists want 6 of the dimensions to be filled up with tiny Calabi–Yau 3-folds. (This means (i) a 6-dimensional Riemannian manifold with SU(3) holonomy, or (ii) a complex manifold V with a Ricci flat Kähler metric and $H^1(V, \mathbb{R}) = 0$, or (iii) an algebraic manifold V with $K_V = 0$ and $H^1(V, \mathcal{O}_V) = 0$. It seems that the holonomy or Kähler conditions on V, together with some finite volume, are required by the physics, whereas making V nonsingular, compact, and a constant fibre over macroscopic space-time are just convenient choices when you try to guess a model.)

The two papers [DHVW] were concerned with trying to calculate string theory on examples of Calabi–Yau varieties obtained by dividing a 3-dimensional complex torus M by a finite group G preserving a basic holomorphic 3-form, so that the stabiliser subgroup at any point is a subgroup of $\mathrm{SL}(3, \mathbb{C})$. A closed string on the quotient may lift either to a closed string on the cover, or to a path that goes from x to $g \cdot x$. The latter are called *twisted sectors*. The physicists need to take care of these in order to relate $\int_X$ to G-equivariant $\int_M$, and they are the key to the form of the McKay correspondence in Theorem 4.4, (4).

Taking limits is a tradition in physics, where the old is frequently the limit of the new: Newtonian mechanics is the limit of special relativity as $c \to \infty$, classical mechanics the limit of quantum physics as $\hbar \to 0$, groups and their Hopf algebras the limit of quantum groups as $q \to 1$. In string theory, if the scale (or radius of curvature) of the tiny Calabi–Yau tends to zero, the theory should approximate ordinary Lorentz 4-dimensional space-time, whereas letting it tend to macroscopic proportions would approximate flat Lorentz 10-dimensional space-time. In this context, the twisted sector near a point $x \in M^H$ plays the role of strings that are topologically nontrivial, but are allowed to remain of finite length (and so contribute to path integrals) as the scale becomes large. To calculate something called the 1-*loop partition function*, DHVW considered mapping the elliptic curve $S^1 \times S^1$ (with parameters σ and τ along the copies of S^1) into X, or the σ, τ square into M with equivariant boundary conditions depending on g, h. Thinking about twisted sectors and limits led DHVW (I confess that their logic eludes me somewhat) to the formula

$$e_{\text{string}}(X) = e(M,G) := \frac{1}{|G|} \sum_{\substack{g,h \in G \\ \text{commuting}}} e(M^{\langle g,h \rangle}). \tag{3.1}$$

Here $e(M,G)$ on the left-hand side is the *G-equivariant Euler number of M*; on the right-hand side, the sum runs over all commuting pairs of elements of G, $\langle g,h \rangle$ is the Abelian group they generate, $M^{\langle g,h \rangle}$ its fixed locus in M, and e is the usual Euler number. The formula is a replacement for the Euler number of the singular orbifold X. The papers [DHVW] contain more-or-less explicitly the conjecture that this number is the Euler number of a minimal resolution of singularities.

It is not hard (see [HH], [Roan] and [Homework]) to rearrange the sums in (3.1) to give

$$e_{\text{string}}(X) = e(M,G) = \sum_{[H] \subset G} e(X^H) \times \operatorname{card}\{[h] \in H\}, \tag{3.2}$$

where (i) the first sum runs over conjugacy classes of subgroups $H \subset G$; (ii) the stratum X^H is the set of $x \in X$ such that $\operatorname{Stab} y$ is conjugate to H for any point $y \in M$ over x; (iii) the second factor is the number of conjugacy classes in H. This means that $X^H \subset X$ contributes to $e(M,G)$ with multiplicity the representation theory of H.

Remark 3.1. — The physicists want to do path integrals, that is, they want to integrate some "Action Man functional" over the space of all paths or loops $\gamma \colon [0,1] \to Y$. This impossibly large integral is one of the major schisms between math and fizz.

The physicists learn a number of computations in finite terms that approximate their path integrals, and when sufficiently skilled and imaginative, can use these to derive marvellous consequences; whereas the mathematicians give up on making sense of the space of paths, and not infrequently derive satisfaction or a misplaced sense of superiority from pointing out that the physicists' calculations can equally well be used (or abused!) to prove $0 = 1$. Maybe it's time some of us also evolved some skill and imagination. The motivic integration treated in the next section builds a miniature model of the physicists' path integral, by restricting first to germs of holomorphic paths $\gamma\colon U \to Y$, where $0 \in U \subset \mathbb{C}$ is a neighbourhood of 0, then to formal power series $\gamma\colon \operatorname{Spec}\mathbb{C}[\![z]\!] \to Y$.

4. MOTIVIC INTEGRATION

The material in this section is due to Batyrev [Ba1], [Ba2], Denef and Loeser [DL1], [DL2] and Kontsevich [K]. I recommend Craw [C1] as a readable first introduction to these ideas.

Rather than trying to restrict to crepant resolutions, take an arbitrary normal crossing resolution $\varphi\colon Y \to X$, marked by the discrepancy divisor $D = \Delta_\varphi = \sum_{i\in I} a_i D_i$ (here I is the indexing set of the components D_i). The normal crossing divisor D defines a stratification of Y, with

$$\textit{closed strata } D_J = \bigcap_{j\in J} D_j, \quad \text{and} \quad \textit{open strata } D_J^\circ = D_J \setminus \bigcup_{J' \supsetneq J} D_{J'}$$

for $J \subset I$ (including, of course, $Y = D_\varnothing$ and $Y \setminus D = D_\varnothing^\circ$).

Motivic integration is discussed and defined below, but it is convenient to start from the answer: the *stringy motive* of (Y, D), or of X itself, turns out to be

$$h_{\text{string}}(X) = h(Y, D) = \sum_{J\subset I} [D_J^\circ] \cdot \prod_{j\in J} \frac{\mathbb{L} - 1}{\mathbb{L}^{a_j+1} - 1}\,. \tag{4.1}$$

Here $\mathbb{L} = [\mathbb{A}^1_{\mathbb{C}}] = [\mathbb{C}]$ is the Tate motive, and the formula takes place in a certain ring of motives with formal power series in $\mathbb{L}^{-1}$ adjoined. We will worry about the coefficient ring later, but in lucky cases it will happen that the cyclotomic polynomials in the denominators cancel out, leaving an integral motive (see Example 4.1 and [Homework] for examples). It follows from Theorem 4.4, (2) and (3) that $h(Y, D)$ is independent of the choice of the normal crossing resolution Y, so depends only on X. In the case when $D = aE$ has a single component with discrepancy a, it boils down to

$$[Y - E] + \frac{[E]}{1 + \mathbb{L} + \mathbb{L}^2 + \cdots + \mathbb{L}^a} = [Y - E] + \frac{[E]}{[\mathbb{P}^a]}. \tag{4.2}$$

Example 4.1. — Let $n = ab$, and consider the n-fold quotient singularity X of type $\frac{1}{b}(1,\dots,1)$, that is, the quotient $\mathbb{C}^n/(\mathbb{Z}/b)$, with the diagonal action of $\varepsilon = \exp\frac{2\pi i}{b}$. It is the cone over the bth Veronese embedding of $\mathbb{P}^{n-1}$, so that its resolution $Y \to X$ has exceptional divisor $E = \mathbb{P}^{n-1}$ with $\mathcal{O}_E(E) = \mathcal{O}_{\mathbb{P}^{n-1}}(-b)$. The discrepancy is $a-1$, to fit the adjunction formula, with $K_Y = (a-1)E$, and $K_E = \mathcal{O}_E(aE) = \mathcal{O}_{\mathbb{P}^{n-1}}(-n)$.

Now whereas Y is homotopy equivalent to $\mathbb{P}^{n-1}$, so has n homology classes, one in each dimension $0, 2, \dots, 2(n-1)$, the effect of dividing by $[\mathbb{P}^{a-1}]$ in (4.2) is to throw away most of these, leaving only the b stringy homology classes in dimension $0, 2a, 4a, 2a(b-1)$. This is exactly what we need for the McKay correspondence: the b elements of $\mathbb{Z}/b$ have age $0, a, 2a, \dots, a(b-1)$ and correspond to the stringy classes in dimension $2ia$.

Example 4.2. — Consider the blowup $\sigma\colon Y_1 \to Y$ of a subvariety $C \subset Y$ that intersects all the strata of D transversally, and set $D_1 = \sigma^*D + (c-1)E$, where $E = \sigma^{-1}C$ is the exceptional divisor of the blowup and $c = \operatorname{codim} C$. The coefficient is the discrepancy of E, so that $K_{Y_1} - D_1 = \sigma^*(K_Y - D)$. It is an exercise to see that

$$h(Y, D) = h(Y_1, D_1).$$

(This is rather trivial if $C \cap D = \varnothing$ in view of Grothendieck's *formule clef* for the motive of a blowup; see [Homework] for more hints.) This is good evidence for the birational invariance of $h(Y, D)$.

I now describe briefly the mechanics of motivic integration, following [C1]. Start from the Grothendieck ring $K_0(\mathcal{V})$ of classes of varieties under the equivalence relation $[V] = [V \setminus W] + [W]$. Addition and multiplication are quite harmless. The Tate motive is $\mathbb{L} = [\mathbb{A}^1_{\mathbb{C}}] = [\mathbb{C}]$. We formally adjoin $\mathbb{L}^{-1}$ to $K_0(\mathcal{V})$, and make a fairly mild $(\mathbb{L}^{-1})$-adic completion to give the value ring $R = \widehat{K}_0(\mathcal{V})[\mathbb{L}^{-1}]$. This value ring is the really clever thing about the whole construction. (Exercise: $(\mathbb{L}^a - 1)^{-1}$ can be written as a formal power series in $\mathbb{L}^{-1}$, so all the terms on the right-hand side of (4.1) are in R.)

Motivic integration takes place over the infinite jet space $J_\infty Y$, which coincides with the set $Y(\mathbb{C}[\![z]\!])$ of points of Y with values in the formal power series ring $\mathbb{C}[\![z]\!]$. An element $\gamma \in Y(\mathbb{C}[\![z]\!])$ is a point $y = \gamma(0) \in Y$ together with a formal arc $\gamma\colon \operatorname{Spec} \mathbb{C}[\![z]\!] \to Y$ starting at y; if convergent, γ is the Taylor series of a holomorphic germ $\widetilde{\gamma}\colon (\mathbb{C}, 0) \to Y$. The infinite jet space $J_\infty Y$ is the profinite limit $\varprojlim_k J_k Y$ of the finite jet spaces $J_k Y$; recall that $J_0 Y = Y$, $J_1 Y$ is the total space of the tangent bundle T_Y, and $J_{k+1} \to J_k$ is a $\mathbb{C}^n$-fibre bundle.

The projection maps $\pi_k\colon J_\infty Y \to J_k$ of the profinite limit allow us to define a *cylinder set* in $J_\infty Y$ to be $\pi_k^{-1}(B_k)$ for a constructible set $B_k \subset J_k$. The measure on

$J_\infty Y$ is initially defined on these, by setting[(1)]

$$\mu(\pi_k^{-1}(B_k)) := [B_k] \cdot \mathbb{L}^{-nk} \in R. \tag{4.3}$$

It is straightforward to see that this is independent of k, and is a "finitely additive measure".

As our measurable functions, consider an effective divisor D on Y, and define a function $F_D \colon J_\infty Y \to \mathbb{Z}_{\geq} 0$ by $F_D(\gamma) = D \cdot \gamma$ (intersection number). In other words, suppose $\gamma(0) = P \in Y$ and let g_D be the local defining equation of D at P; then $F_D(\gamma)$ is the order of $\gamma^*(g_D) \in \mathbb{C}[\![z]\!]$. Since the first s coefficients of $\gamma^*(g_D)$ clearly only depend on $\pi_s(\gamma) \in J_s$, it is obvious that $F_D^{-1}(s)$ is a cylinder set.

The grand definition is now: for Y a nonsingular variety and D a normal crossing divisor, the motivic integral is

$$h(Y, D) = \int_{J_\infty Y} \mathbb{L}^{-F_D} := \sum_{s \in \mathbb{Z}_{\geq 0}} \mu\big(F_D^{-1}(s)\big) \cdot \mathbb{L}^{-s} \in R. \tag{4.4}$$

Remark 4.3. — I omit some tricky details on convergence required to get a genuine measure (involving the $[\mathbb{L}^{-1}]$-adic completion). To tell the truth, I don't know if they are at all essential. A basic point for applications is that the measure of $F_D^{-1}(s)$ tends to 0 as $s \to \infty$; this is plausible enough (because arcs γ with $\gamma \cdot D \geq s$ have codimension $\geq s$ in $J_\infty Y$), and is an intuitive reason behind birational invariance: the arcs in a Zariski closed subset of Y have measure zero.

Theorem 4.4. — *$h(Y, D)$ of (4.4) has the following properties:*

(1) *If $D = 0$ then $h(Y, D) = [Y]$.*

(2) *$h(Y, D)$ is calculated by the right-hand side of (4.1).*

(3) *Birational invariance: let Y', D' and Y, D be pairs, and $\varphi \colon Y' \to Y$ a birational morphism such that $K_{Y'} - D' = \varphi^*(K_Y - D)$; then*

$$h(Y', D') = h(Y, D).$$

(4) *If $X = M/G$ is as in Assumption 1.2, $Y \to X$ a normal crossing resolution, and D the discrepancy, then*

$$h_{\text{string}}(X) = h(Y, D) = \sum_{[H] \subset G} \big[X^H\big] \cdot \sum_{[g] \in H} \mathbb{L}^{\operatorname{age} g}, \tag{4.5}$$

where the range of summation is as in (3.2), and the second sum is over conjugacy classes in H.

[(1)]The papers [DL1] and [C1] have the exponent $\mathbb{L}^{-n(k+1)}$. This is just a normalising convention, giving $h(Y, D) = [Y] \cdot \mathbb{L}^{-n}$ in Theorem 4.4, (1), and making the motive of Y 0-dimensional. I prefer my version.

Discussion of proof. — I give some indications, leaving most of the proof as references to [DL1] and [DL2]. Alternatively, do them as exercises (see [Homework] for more hints). The key point of the proof is that, whatever its substance, (4.4) has the formal properties of an integral, and is subject to the same kind of *change of variables* formula. In the words of the Master:

"La théorie consiste pour l'essentiel dans des questions de variance"

([H], Introduction). Note first that the condition in (3) says that $D' - D = \operatorname{div}(\operatorname{Jac}\varphi)$ is the divisor of zeros of the Jacobian determinant of φ (I omit φ^* from now on). Composition defines a map $j_\varphi\colon J_\infty Y' \to J_\infty Y$, and, unless it falls entirely in the locus of indeterminacy of φ^{-1}, an arc in Y has a birational transform as an arc in Y'; in other words, away from subsets of measure zero, j_φ is a bijection on the infinite jet spaces. For (3), it remains only to stratify the finite jet spaces J_kY' and J_kY so that the corresponding morphism $j_k\colon J_kY' \to J_kY$ is a $\mathbb{C}^t$-bundle on each stratum with $F_{D'-D}(\gamma) = \operatorname{div}(\operatorname{Jac}\varphi)\cdot\gamma = t$ (see [DL1], Lemma 3.4 and [Homework]).

(2) is proved in [DL1], Proposition 6.3.2, [Ba2], Theorem 6.28, and worked out in detail in [C1], Theorem 1.16. The proof of (4) consists of two steps, relating to the two morphisms $\pi\colon M \to X$ and $\varphi\colon Y \to X$ of Assumption 1.2.

Step I. — We translate the twisted sectors of [DHVW] into the language of formal arcs, obtaining the stratification (4.6) below.

Let $y \in M^H$ be a point with $\operatorname{Stab} y = H$ and $x = \pi(y) \in X^H$. As at the start of Section 2, suppose that r is an integer divisible by the order of each $g \in H$, and choose an rth root ε of 1 and an rth root $\zeta = z^{1/r}$ of the parameter used for formal arc, so that a formal arc γ at $x \in X$ parametrised by z lifts to a formal arc at $y \in M$ parametrised by ζ. Unless γ falls entirely in the branch locus of $\pi\colon M \to X$, there is a unique conjugacy class $g \in H$ defined by $\gamma(\varepsilon\zeta) = g\gamma(\zeta)$. Here g is the *twisted sector*, the conjugacy class of γ in the local fundamental group H (where γ is viewed as a little loop in X minus the branch locus).

This argument shows that, after we delete the subset of arcs falling entirely in the branch locus (which has infinite codimension, so measure zero) the infinite jet space $J_\infty X$ is a disjoint union

$$J_\infty X = \coprod_{[H]\subset G} \ \coprod_{[g]\in H} J_\infty^{H,g} Y, \tag{4.6}$$

where H, g are as in (3.2), and $J_\infty^{H,g}Y$ is the set of arcs with $\gamma(0) \in X^H$ in the twisted sector g.

Step II. — Using change of variables as in the proof of (3), one calculates that $J_\infty^{H,g}Y$ contributes $X^H \cdot L^{g^{-1}}$ to $h(Y,D)$ ([DL2], Lemma 4.3). The difference in appearance of the formulas here and in [DL2] is explained by two trivial shifts of notation: as explained in the footnote on page 63, my measure is $\mathbb{L}^n$ times theirs; and they diagonalise g as ε^{e_i} with $1 \le e_i \le r$, defining $w(g) = \frac{1}{r}\sum e_i = n - \operatorname{age}(g^{-1})$.

Remark 4.5. — Statement (4) is an exact analogue of the [DHVW] formula (3.2), saying that the stratum X^H appears in the stringy homology of Y multiplied by the set of conjugacy classes in H.

As discussed in Definition 1.3, the discrepancy $D = \operatorname{div} s_X$ is the divisor of zeros of s_X, the global basis of $\Omega^n_{\mathrm{NonSing}\,X}$. In the normal course of events, integrating *functions* on Y requires a volume form; here we take s_X as a holomorphic volume form, viewing its zeros on D as scaling down the contribution from neighbourhood of the discrepant exceptional divisors. This is what produces a birationally invariant answer.

5. HILBERT SCHEMES OF G-ORBITS

This section explains the definition of the G-orbit Hilbert scheme G-Hilb M, and Nakamura's idea of using it to resolve certain quotient singularities. We know by general results (especially Hironaka's theorems) that the singularities of a quotient variety $X = M/G$ can be resolved somehow-or-other, but the construction of an actual resolution is messy, involves lots of choices, and will probably have almost nothing to do with the group action. Around 1995, Ito and Nakamura observed that in the case of $G \subset \mathrm{SL}(2,\mathbb{C})$, *the Hilbert scheme* G-Hilb $\mathbb{C}^2$ *of G-clusters is a crepant resolution of the quotient* $\mathbb{C}^2/G$. Nakamura conjectured that this continues to hold for $G \subset \mathrm{SL}(3,\mathbb{C})$, and this has since been confirmed and extended to some other cases by work of Bridgeland and others (see [BKR] and Theorem 6.1).

First, a *cluster* in a variety M (say, quasiprojective and nonsingular) is a 0-dimensional subscheme $Z \subset M$, defined by an ideal $\mathcal{I}_Z \subset \mathcal{O}_M$, so that the cokernel $\mathcal{O}_Z = \mathcal{O}_M/\mathcal{I}_Z$ is a finite dimensional $\mathbb{C}$-vector space. The *degree* of Z is the dimension of $\mathcal{O}_Z$. Like the intersection of two plane curves in Bezout's theorem, a cluster Z may consist of reduced points $Z = P_1 + \cdots + P_N$, or may have a nonreduced structure; in the latter case, we *keep track of the ideal* $\mathcal{I}_Z \subset \mathcal{O}_M$, as a way of using algebraic equations to keep information about the relative positions when some of the points P_i come together. For example,

$$(x^2, xy, y^2) \quad \text{and} \quad (x - ay - by^2, y^3) \quad \text{for any } a, b \in \mathbb{C}$$

are clusters of degree 3 supported at $0 \in \mathbb{C}^2$.

Lemma 5.1. — *All clusters $Z \subset M$ of given degree N in M are parametrised by a quasiprojective scheme* $\mathrm{Hilb}^N M$*, which is a fine moduli space.*

Proof. — The assertion is quite elementary. M is quasiprojective; choose an embedding $M \subset \mathbb{P}^s$. Every ideal $\mathcal{I}_Z \subset \mathcal{O}_M$ of codimension N defines and is defined by a codimension N vector subspace

$$H^0(\mathbb{P}^s, \mathcal{I}_Z(N)) \subset H^0(\mathbb{P}^s, \mathcal{O}_{\mathbb{P}^s}(N)),$$

the forms of degree N vanishing on Z (same N). Subspaces of given codimension are parametrised by a Grassmann variety, and the condition that a space of forms defines a cluster of degree N in M is a locally closed condition. (It can be written in terms of rank of a matrix $= N$.) $\square$

Remark 5.2. — The map $\operatorname{Hilb}^N M \to S^N M$ to the symmetric product, defined at the level of sets by $Z \mapsto \operatorname{Supp} Z$, is a morphism of schemes, the Hilbert–Chow morphism (see [GIT], Chapter 5, §4). For a curve, $\operatorname{Hilb}^N C$ is just the symmetric product $S^N C$, which is itself already nonsingular. For a surface, the symmetric product $S^N S$ is singular at the diagonals, and $\operatorname{Hilb}^N S \to S^N S$ is a crepant resolution, in fact, a symplectic resolution; see [IN2], §6. But $\operatorname{Hilb}^N M$ is singular as soon as $\dim M \geq 3$ and $N = \deg Z \geq 4$, and usually even has components of excess dimension.

PROPOSITION–DEFINITION 5.3 (Ito and Nakamura). — *Let G be a finite group of order N acting faithfully on an algebraic manifold M; consider the action of G on* $\operatorname{Hilb}^N M$ *and its fixed locus* $(\operatorname{Hilb}^N M)^G$. *This has a unique irreducible component that contains a general orbit $G \cdot y$ of G on M. This component is defined to be the G-*Hilbert scheme*, and denoted by* $G\text{-Hilb}\, M$. *The composite* $G\text{-Hilb}\, M \hookrightarrow \operatorname{Hilb}^N M \to S^N M$ *induces a Hilbert–Chow morphism* $G\text{-Hilb}\, M \to M/G$ *which is proper and birational.*

A cluster $Z \in G\text{-Hilb}$ *is G- invariant, and is called a G-*cluster*; its defining ideal $\mathcal{I}_Z$ is G-invariant, and as a representation of G, the quotient $\mathcal{O}_Z = \mathcal{O}_M/\mathcal{I}_Z$ is the regular representation* $\mathbb{C}[G]$.

See also [CR], 4.1 for a rival definition and a comparison between the two.

Proof. — The general orbit $G \cdot y$ consists of N points permuted simply transitively by G, so is a G-invariant cluster in $(\operatorname{Hilb}^N M)^G$. These orbits fill out an irreducible open set in $(\operatorname{Hilb}^N M)^G$, because a small G-invariant deformation of $G \cdot y$ is clearly still a set of N distinct points permuted by G and disjoint from any fixed locus. The closure of this component is $G\text{-Hilb}\, M$ by definition. The composite $G\text{-Hilb}\, M \hookrightarrow \operatorname{Hilb}^N M \to S^N M$ is a morphism; by definition, a dense open set of $G\text{-Hilb}\, M$ consists of general orbits $G \cdot y$, and these maps to orbits in $S^N M$, that is, to M/G.

Finally, the quotient sheaves $\mathcal{O}_Z$ for $Z \in G\text{-Hilb}\, M$ fit together as a locally free sheaf $\mathcal{O}_{\mathcal{Z}}$ over $G\text{-Hilb}\, M$, with a G-action that makes it the regular representation on a dense open set. Its isotypical decomposition under the idempotents of $\mathbb{C}[G]$ is a direct sum, so each component must also vary as a locally free sheaf, therefore $\mathcal{O}_Z \cong \mathbb{C}[G]$ for every $Z \in G\text{-Hilb}\, M$ (since $G\text{-Hilb}\, M$ is defined to be irreducible). $\square$

The G-Hilbert scheme is a crepant resolution for finite groups $G \subset \mathrm{SL}(3, \mathbb{C})$. The general case of this is proved by Bridgeland and others [BKR] using derived category methods and a homological characterisation of regularity. For a diagonal Abelian group, $A\text{-Hilb}\, \mathbb{C}^3$ is a completely explicit construction of Nakamura (see [N] and [CR]):

the monomial xyz is A-invariant, and every G-cluster Z is defined by 7 (possibly redundant) equations of the form

$$\begin{array}{ll} x^{a+1} = \lambda y^d z^g & y^{d+1} z^{g+1} = \alpha x^a \\ y^{b+1} = \mu z^e x^h & z^{e+1} x^{h+1} = \beta y^b \qquad \text{and} \quad xyz = \xi \\ z^{c+1} = \nu x^f y^i & x^{f+1} y^{i+1} = \gamma z^c \end{array}$$

for appropriate exponents $a, \dots, i$ and coefficients $\alpha, \dots, \xi$ satisfying $\alpha\lambda = \beta\mu = \gamma\nu = \xi$. The monomial basis of $\mathcal{O}_Z$ forms a tripod shaped Newton polygon in the plane lattice $\mathbb{Z}^2$ of Laurent monomials modulo xyz; this lattice is naturally the universal cover of the McKay quiver and the tripod is a choice of fundamental domain for the covering group (see [N] and [R] for pictures). The explicit calculations remain an interesting challenge in the non-Abelian cases, e.g., in the trihedral case.

Example 5.4. — These results are known to fail for finite $G \subset \mathrm{SL}(4, \mathbb{C})$. In the first place, most quotient singularities $X = \mathbb{C}^4/G$ do not have any crepant resolution. For example, the series of cyclic quotient singularities $\mathbb{C}^4/(\mathbb{Z}/r)$ of type $\frac{1}{r}(1, r-1, i, r-i)$ have no junior elements, so are terminal; compare Example 4.1. These examples motivated the initial exploration of stringy homology in [BD].

Next, even when a crepant resolution exists, the G-Hilbert scheme may be singular or discrepant or both. A simple example is the quotient singularity $\mathbb{C}^4/G$ by the maximal diagonal subgroup $(\mathbb{Z}/2)^{\oplus 3} \subset \mathrm{SL}(4, \mathbb{C})$ of exponent 2. The junior simplex Δ has all the midpoints of the edges $\frac{1}{2}(1, 1, 0, 0)$ etc., as lattice points. This has several subdivisions into basic simplexes, giving crepant resolutions, but none that is symmetric under permuting the coordinates – the only symmetric thing you can do is chop off the 4 basic simplexes at the corners, leaving a terminal simplex of volume 2. On the other hand, $G\text{-Hilb}\,\mathbb{C}^4$ is obviously symmetric.

6. COHERENT DERIVED CATEGORY

Grothendieck and Verdier introduced the derived category $\mathrm{D}(X)$ of coherent sheaves on a variety X in the 1960s as a technical convenience in homological algebra; it has enjoyed an unfortunate reputation for technicality and abstraction ever since then. Recently, however, it has been increasingly used as a geometric characteristic of X similar to K theory: whereas K theory works with the group of bundles or sheaves modulo the relation $F = F' + F''$ for every short exact sequence $0 \to F' \to F \to F'' \to 0$, the derived category $\mathrm{D}(X)$ consists of complexes $F^\bullet$ modulo the relation of quasi-isomorphism (defined at the start of the theory, and thankfully never referred to again). Following Mukai's pioneering work [Mu] for Abelian varieties, Orlov and Bondal [O], [BO1] have advocated the idea of considering the derived category $\mathrm{D}(X)$ (up to isomorphism of triangulated categories) as a geometric characteristic of X. From this point of view, $\mathrm{D}(X)$ behaves like an enriched version of K theory.

A variety X with $\pm K_X$ ample can be reconstructed from its derived category $\mathrm{D}(X)$ (as a triangulated category) [BO1], but if $K_X = 0$ (notably for an Abelian variety or a K3 surface), the same triangulated category may occur as $\mathrm{D}(X)$ for different X, or there may be infinitely many symmetries of $\mathrm{D}(X)$ not arising from automorphisms of X. Isomorphisms $\mathrm{D}(X) \cong \mathrm{D}(Y)$ arise as *Fourier–Mukai transforms* $\Phi^{\mathcal{F}}_{X\to Y}$ corresponding to a sheaf $\mathcal{F}$ on $X \times Y$, defined as the composite of the functors p_X^*, $\otimes\mathcal{F}$ and q_{Y*} (more precisely, their derived functors); for an up-to-date treatment, see [Br] and the references given there. In practice, Y is most frequently a moduli space of coherent sheaves on X and $\mathcal{F}$ the universal sheaf over $X \times Y$, so that Y parametrises sheaves F_y on X; in very good cases, the apparatus of moduli functors, stable bundles, and deformation theory gives essentially for free that the F_y have orthonormality properties under Ext functors (formally analogous to those of trig functions in the theory of Fourier transform).

Let M be a nonsingular quasiprojective n-fold with $K_M = 0$, and G a finite group acting on M, with trivial action on K_M. Set $Y = G\text{-Hilb}\, M$. Since Y is a fine moduli space for G-clusters $Z \subset M$, there is a universal G-cluster $\mathcal{Z} \subset Y \times M$, fitting in a diagram

$$
\begin{array}{ccc}
\mathcal{Z} & \xrightarrow{q} & M \\
{\scriptstyle p}\downarrow & & \downarrow{\scriptstyle \pi} \\
Y & \xrightarrow{\varphi} & X
\end{array}
\tag{6.1}
$$

Bridgeland and others [BKR] prove the following theorem.

THEOREM 6.1. — *Suppose that the inverse image of the diagonal* $(\varphi \times \varphi)^{-1}(\Delta_X)$ *has dimension* $\leq n+1$ *(automatic for* $n=3$*). Then* Y *is a crepant resolution of* X *and the Fourier–Mukai functor* $\Phi = \mathbb{R}q_* \circ p^* \colon \mathrm{D}(Y) \to \mathrm{D}^G(M)$ *is an equivalence of categories.*

Once we know that Y is a crepant resolution, ω_M is trivial as a G-sheaf and ω_Y is trivial, so that both the derived categories $\mathrm{D}^G(M)$ and $\mathrm{D}(Y)$ have Serre duality functors; the remainder of the proof is then standard Fourier–Mukai technology. However, the surprising thing here is Bridgeland's derivation of the nonsingularity of Y from the famous theorem of commutative algebra known for a long time as Serre's "Intersection conjecture".

7. FIN DE PARTIE

Samuel Beckett's play of the same title has the wonderful line:

"Personne au monde n'a jamais pensé aussi tordu que nous."

This seems to reflect a truth about math research: progress beyond the obvious takes really twisted thinking. In this spirit, let me raise all the open questions I can think of.

There are two basic flavours of McKay correspondence:

(1) conjugacy classes of $G \leftrightarrow$ homology of Y (or stringy homology); and

(2) representations of $G \leftrightarrow$ derived category $\mathrm{D}(Y)$ or K theory of Y.

Is there a "bivariant" version of the correspondence containing both (1) and (2) at the same time? For example, in some contexts, $\mathcal{D}$-modules or perverse sheaves manage to accommodate both coherent and topological cohomology. Note that (1) and (2) achieve a well posed question in completely different ways: (1) takes accounts of discrepancy systematically, whereas (2) currently only works under the very strict condition that $Y = G$-Hilb is a crepant resolution.

The representation theory of finite groups has two ingredients, conjugacy classes and irreducible representations, and a character table, which is a nonsingular matrix making them "dual" (I apologise to group theorists for this gratuitous vulgarity). Although in substance very different, the homology and K theory of a variety Y could be described in similar terms. In cases when McKay holds, is there any direct relation?

All the different approaches to McKay described here have one thing in common: none of them seems to say anything very useful about multiplicative structures. The following questions seem most likely to be approachable: can tensor product of G-modules and tensor product in K theory of Y be related? Can you reconstruct the McKay quiver in $\mathrm{D}(Y)$ or $K_0 Y$?

Motivic integration takes a fraction of the homology of a discrepant exceptional divisor, say, half the homology of the exceptional $\mathbb{P}^3$ for the quotient singularity $\mathbb{C}^4/(\mathbb{Z}/2)$ (the cone on the second Veronese embedding $v_2(\mathbb{P}^3)$). In contrast, half of a derived category is something no-one has ever seen. In the case of $v_2(\mathbb{P}^3)$, the Gonzalez-Sprinberg–Verdier sheaves corresponding to the characters ± 1 are $\mathcal{O}_Y$ and $\mathcal{O}_Y(1)$. Breaking up the derived category $\mathrm{D}(Y)$ into two bits, one of which will correspond to the representations of $\mathbb{Z}/2$, doesn't seem to make any sense. On the other hand, in this case we can extend the action of $\mathbb{Z}/2$ to the action $\frac{1}{4}(1,1,1,1)$ of $\mathbb{Z}/4$, whose quotient does have a crepant resolution.

Another general problem area: resolutions of Gorenstein quotient singularities give a collection of examples of Calabi–Yau 3-folds with very nice properties: the homology of the resolution is well defined (independent of the choice of resolution), and the homology and K theory are closely related by something like a duality. Do these properties hold for Calabi–Yau 3-folds more generally? It seems very likely that birational Calabi–Yau 3-folds have isomorphic derived categories, but so far this only seems to be established when they are related by classic flops [BO2].

Part of motivic integration is the simple idea of using $\varphi^* s_X$ as the volume form, even though it vanishes along the discrepancy divisor D (compare Remark 4.5). Maybe this idea can be used with differentials on X itself (not passing to $J_\infty X$) to get birationally invariant de Rham and Hodge cohomology?

Elliptic cohomology is another area of geometry with an alleged stringy interpretation – as the index of the Dirac operator on the space of loops. Could part of this theory have a rigorous treatment in terms of spaces of formal arcs, like motivic integration in Section 4? If we believe that the elliptic cohomology of M/G has a well defined answer (see Totaro [T] for some evidence) then Principle 1.1 predicts what the answer must look like in a whole pile of substantial cases.

Which Gorenstein quotient singularities admit crepant resolutions? Since 4-fold singularities usually do not have crepant resolutions, those that do are of particular interest; see [DHZ] for examples. How does this relate to complex symplectic geometry? The papers of Verbitsky [Vb] and Kaledin [Ka1], [Ka2] study crepant resolutions and related issues for symplectic quotient singularities. When crepant resolutions exist they are symplectic [Vb], therefore "semismall", giving a complete and elegant solution to the homological form (1) of the McKay correspondence [Ka2]. Is it possible that there is a "special" geometry in 3 complex dimensions (such as complexified imaginary quaternions), like symplectic or hyper-Kähler geometry for complex surfaces or 4-folds, that explain why crepant resolutions exist for 3-folds?

How should we interpret Nakamura's results and conjectures on G-Hilb? If a crepant resolution exists, it would be exceedingly convenient to be able to describe it as a fine moduli space of something; G- clusters have no especially privileged role, but the requirement that the space be birational to M/G seems to impose some relation with the moduli space of group orbits. Nakamura and Nakajima have raised the question of whether the *other* crepant resolutions (after a flop) can also be interpreted as moduli, for example as Quot schemes; a single convincing example of this would add weight to their suggestion. Do the crepant resolutions in Example 5.4 have interpretations as moduli?

REFERENCES

[Ba1] V. Batyrev, Birational Calabi–Yau n-folds have equal Betti numbers, in New trends in algebraic geometry, Klaus Hulek and others (eds.), CUP, 1999, pp. 1–11.

[Ba2] V. Batyrev, Stringy Hodge numbers of varieties with Gorenstein canonical singularities, in Integrable systems and algebraic geometry (Kobe/Kyoto, 1997), 1–32, World Sci. Publishing, River Edge, NJ, 1998.

[Ba3] V. Batyrev, Non-Archimedean integrals and stringy Euler numbers of log-terminal pairs, J. Eur. Math. Soc. **1** (1999), 5–33.

[BD] V. Batyrev and D. Dais, Strong McKay correspondence, string-theoretic Hodge numbers and mirror symmetry, Topology **35** (1996), 901–929.

[BO1] A. Bondal and D. Orlov, Reconstruction of a variety from the derived category and groups of autoequivalences, Max Planck Inst. Bonn preprint MPI-97-36, `math.AG/9712029`, 20 pp.

[BO2] A. Bondal and D. Orlov, Semi-orthogonal decomposition for algebraic varieties, preprint `alg-geom/9506012`.

[Br] T. Bridgeland, Equivalences of triangulated categories and Fourier–Mukai transforms, Bull. London Math. Soc. **31** (1999), 25–34.

[BrM] T. Bridgeland and A. Maciocia, Fourier–Mukai transforms for K3 fibrations, preprint `math/9908022`, 18 pp.

[B] J.-L. Brylinski, A correspondence dual to McKay's, preprint `alg-geom/9612003`.

[BKR] T. Bridgeland, A. King and M. Reid, Mukai implies McKay, preprint `math/9908027`, 17 pp.

[C1] A. Craw, An introduction to motivic integration, preliminary draft available from `www.maths.warwick.ac.uk/~craw`, 23 pp.

[C2] A. Craw, A-Hilb $\mathbb{C}^3$ and McKay correspondence, work in progress.

[CR] A. Craw and M. Reid, How to calculate A-Hilb $\mathbb{C}^3$, preprint `math/9909085`, 29 pp.

[DHVW] L. Dixon, J. Harvey, C. Vafa and E. Witten, Strings on orbifolds. I, Nuclear Phys. B **261** (1985), 678–686. II, same J. **274** (1986), 285–314.

[DHZ] D. I. Dais, M. Henk and G. M. Ziegler, All Abelian quotient c.i. singularities admit crepant resolutions, preprint `alg-geom/9704007`, 35 pp.

[DL1] J. Denef and F. Loeser, Germs of arcs on singular algebraic varieties and motivic integration, Invent. Math. **135** (1999), 201–232.

[DL2] J. Denef and F. Loeser, Motivic integration, quotient singularities and the McKay correspondence, preprint `math/9903187`, 20 pp., to appear in Compositio Math.

[GSpV] G. Gonzalez-Sprinberg and J.-L. Verdier, Construction géométrique de la correspondance de McKay, Ann. Sci. École Norm. Sup. (4) **16** (1983), 409–449.

[H] R. Hartshorne, Residues and duality, L.N.M. **20** Springer, 1966.

[HH] F. Hirzebruch and T. Höfer, On the Euler number of an orbifold, Math. Ann. **286** (1990), 255–260.

[IN] Y. Ito and H. Nakajima, McKay correspondence and Hilbert schemes in dimension three, preprint `math.AG/9803120`, 35 pp., to appear in Topology.

[IN1] Y. Ito and I. Nakamura, McKay correspondence and Hilbert schemes, Proc. Japan Acad. Ser. A Math. Sci. **72** (1996), 135–138.

[IN2] Y. Ito and I. Nakamura, Hilbert schemes and simple singularities, in New trends in algebraic geometry, Hulek and others (eds.), CUP 1999, pp. 155–233.

[IR] Y. Ito and M. Reid, The McKay correspondence for finite subgroups of $\mathrm{SL}(3,\mathbb{C})$, in Higher-dimensional complex varieties (Trento, 1994), 221–240, de Gruyter, Berlin, 1996.

[K] M. Kontsevich, Motivic integration, Legendary lecture at Orsay, Thu 7th Dec 1995.

[Ka1] D. Kaledin, McKay correspondence for symplectic quotient singularities, preprint `math/9907087`, 28 pp.

[Ka2] D. Kaledin, Dynkin diagrams and crepant resolutions of quotient singularities, preprint `math/9903157`, 30 pp.

[Kr] P. B. Kronheimer, The construction of ALE spaces as hyper-Kähler quotients, J. Diff. Geom. **29** (1989), 665–683.

[GIT] D. Mumford, J. Fogarty and F. Kirwan, Geometric invariant theory (3rd edn.), Springer, 1994.

[Mu] S. Mukai, Duality between $\mathrm{D}(X)$ and $\mathrm{D}(\hat{X})$ with its application to Picard sheaves, Nagoya Math. J. **81** (1981), 153–175.

[N] I. Nakamura, Hilbert schemes of Abelian group orbits, to appear in J. Alg. Geom.

[O] D.O. Orlov, Equivalences of derived categories and K3 surfaces, Algebraic geometry, **7** J. Math. Sci. (New York) **84** (1997), 1361–1381, preprint `alg-geom/9606006`, 28 pp.

[YPG] M. Reid, Young person's guide to canonical singularities, in Algebraic Geometry, Bowdoin 1985, ed. S. Bloch, Proc. of Symposia in Pure Math. **46**, A.M.S. (1987), vol. 1, 345–414.

[R] M. Reid, McKay correspondence, in Proc. of algebraic geometry symposium (Kinosaki, Nov 1996), T. Katsura (ed.), 14–41, preprint alg-geom/9702016, 30 pp.

[Homework] Homework sheets will be on my website `www.maths.warwick.ac.uk/~miles`, including examples, exercises, more hints, and errata to this lecture.

[Roan] S.-S. Roan, Orbifold Euler characteristic, in Mirror symmetry, II, AMS 1997, pp. 129–140.

[T] B. Totaro, Chern numbers for singular varieties and elliptic homology, Ann. of Math. **151** (2000), no. 2, 757–791. Preprint available from `www.dpmms.cam.ac.uk/~bt219`.

[V] C. Vafa, String vacua and orbifoldized LG models, Modern Phys. Lett. A **4** (1989), 1169–1185.

[Vb] M. Verbitsky, Holomorphic symplectic geometry and orbifold singularities, preprint `math.AG/9903175`, 17 pp.

[Z] E. Zaslow, Topological orbifold models and quantum cohomology rings, Comm. Math. Phys. **156** (1993), 301–331.

Miles REID

Mathematical Institute
University of Warwick
GB-Coventry CV4 7AL
E-mail : `Miles@Maths.Warwick.Ac.UK`

Séminaire BOURBAKI Novembre 1999
52^e année, 1999-2000, n^O 868, p. 73 à 103

GINZBURG-LANDAU VORTICES: THE STATIC MODEL

by **Tristan RIVIÈRE**

1. INTRODUCTION

1.1. Physical origin of the problem

One of the first explanatory models for superconductivity (which refers to the existence of permanent currents in certain substances, with no energy dissipation) has been proposed during the fifties by V. Ginzburg and L. Landau, from the Landau theory of phase transitions. Following this model, the degree of superconductivity of a body occupying a domain Ω of $\mathbb{R}^3$, is characterized by a "wave function" $\Omega \to \mathbb{C}$ referred to as the order parameter. In the quantum theory of J. Bardeen, L.N. Cooper and J. Schrieffer (BCS theory), which came in 1957 to justify the Ginzburg-Landau phenomenological model, the square of the modulus of this order parameter $|u|^2$ represents the local electron pair (Cooper pairs) density, responsible for the superconductivity. For $|u| = 1$ this density is maximum and minimum for $|u| = 0$.

The energy functional for a superconductor proposed by Ginzburg and Landau is

$$\mathcal{J}(u, A) = \int_\Omega |\kappa^{-1} du - iAu|^2 + \frac{1}{2}|1 - |u|^2|^2 + |dA|^2 - 2\int_\Omega dA.h_e$$

where A is the 1-form vector potential associated to the induced field dA in the superconductor ($du - iAu$ is thus a 1-form taking its values in $\mathbb{C}$). h_e is the 2-form representing the external field applied to the superconductor. This is one of the parameters of the problem together with the constant κ, known as the coupling constant, which depends on the sample considered, and which plays an essential part in the theory, as we shall see in the following. As a ratio of two lengths, $\kappa = \frac{\lambda}{\xi}$, where λ is the penetration depth of the external field h_e in the sample (see the following) and ξ is the characteristic size of a vortex (see section 2), is a dimensional constant. Note that this functional is also the Yang-Mills-Higgs action in the abelian gauge theory modeling the interaction between a classical magnetic field and a Higgs particle.

Schematically, the observed phenomena are as follows. When the applied field is zero, the superconductor is said to be in the *pure state*:

$$\begin{cases} |u| = 1 & \text{in } \Omega \\ dA = 0 & \text{in } \Omega. \end{cases}$$

The density of Cooper pairs is maximum and the induced field is zero. When the applied field is sufficiently strong (sample dependent) the superconductivity disappears:

$$\begin{cases} |u| = 0 & \text{in } \Omega \\ dA = h_e & \text{in } \Omega. \end{cases}$$

The density of Cooper pairs is then minimal and the induced field coincides with the applied field. The superconductor is in the *normal state*.

The nature of the transition from the *pure state* to the *normal state* depends on the composite one and in particular on the value of κ. One observes that for $\kappa < 1/\sqrt{2}$ (type I superconductor), this transition is sharp and happens for a certain strength of the applied field which is independent of κ. Instead, for $\kappa > 1/\sqrt{2}$ (type II superconductor), as the external field increases, to go from the *pure state* to the *normal state*, we pass through a different phase known as a *mixed state*, where more and more regions of normal state contained in tubes (vorticity filaments) around which the phase of u makes one or several circular turns, appears. When the sample is homogeneous and the external field uniform, these tubes line up in the direction of the field, to form periodic Abrikosov lattices, named after the physician who first showed their existence. It is observed that this lattice is triangular in the fundamental state. We pass from the *pure state* to the *mixed state*, for an applied field known as the "first critical field" $H_{c_1} \simeq O\left(\frac{\log\kappa}{\kappa}\right)$, and we leave the *mixed state* to go into the *normal state* for an applied field known as the "second critical field" $H_{c_2} \simeq O(\kappa)$. The phase diagram (figure 1) summarizes the observations mentioned above. For a more complete account of the physics of superconductors the reader can refer to: [dG], [SST], [Ti]...

1.2. The mathematical questions underlying superconductivity

There are numerous difficulties that arise when one wants to give a mathematically rigorous sense to the previous observations, starting from the Ginzburg-Landau model. A first reduction is to consider a 1 or 2-dimensional version of the model ($\Omega \subset \mathbb{R}^3$ and h_e then have the symmetries corresponding to those reductions: space comprised between two parallel planes or infinite cylinder, in uniform magnetic fields...). In this talk, we shall not consider the studies in 1-dimension which are however extensive and which enable very often a more refined analysis of the phase diagram ([BH1], [BH2], [Af]... for a complete presentation of these results, see [AT]). We shall consider here only the 2-dimensional case, which is the minimal dimension to observe vortices (dimension 3 and higher dimensions are treated in [Ri2], [LR], [LR2] and also in [BBM]).

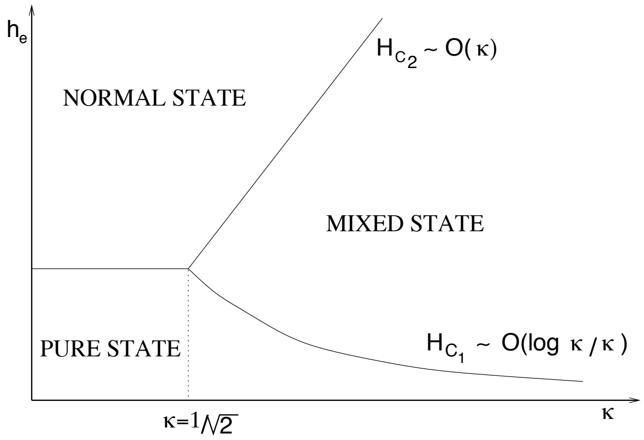


FIGURE 1. Phases diagram

Ω is thus an infinite cylinder and h_e is a uniform field parallel to the direction of the axis of the cylinder. Ω then denotes the 2-dimensional section of this cylinder and h_e being a 2-form which is constant on this section is often confused with the number giving its intensity. The aim is thus to understand the nature of fundamental states of the functional $\mathcal{J}$, and also of the critical points in general, as a function of the different values of (κ, h_e) in the phase diagram represented in figure 1. By "understanding the nature of the fundamental states of $\mathcal{J}$", we mean essentially identifying the zero set of the order parameter u of a solution minimizing $\mathcal{J}$, which corresponds to the 2-dimensional section of the vortex lattice expected in the mixed phase.

To simplify the analysis we consider the change of variables $A \longrightarrow \kappa A$ in the original model, which then leads us to the functional

$$\mathcal{G}_\kappa(u, A) = \int_\Omega |du - iAu|^2 + \frac{\kappa^2}{2}|1 - |u|^2|^2 + |dA|^2 + 2h_e \int_\Omega dA.$$

This new functional verifies the gauge invariance $\mathcal{G}_\kappa(u, A) = \mathcal{G}_\kappa(e^{i\phi}u, A + d\phi)$ for any function ϕ on Ω. It is then possible to extend the model to any domain Ω which is any 2-dimensional manifold. (u, A) are then respectively the sections and connections of a complex line bundle E on Ω on which we fix a hermitian product whose real part is $(\ ,\ )$ or $|\ |^2$ for the quadratic form. $du - iAu$ is replaced by the covariant derivative $d_A u$ of u with respect to A and dA is the curvature of the connection A. In the following, we note $h = *dA$.

Section 2 is devoted to the study of the Ginzburg-Landau free energy $\mathcal{F}$ without interaction with the external field (i.e $h_e = 0$). In 2.1, we present the work of Jaffe and Taubes on the integrable or non-interacting case $\kappa = 1/\sqrt{2}$. We state their conjectures for the cases $\kappa < 1/\sqrt{2}$ and $\kappa > 1/\sqrt{2}$. In 2.2, we describe the BBH asymptotic analysis (F. Bethuel, H. Brezis and F. Hélein [BBH]) in the London limit: $\kappa \to +\infty$ which corresponds to the strongly repulsive case. In this limit, to which we shall restrict in the following, the vorticity phenomena appear more clearly; this case is also close to many of the type II superconductors we have in practice, for which the a-dimensional parameter κ is very large. In section 2.3, we revisit the part played by the renormalized energy $\mathcal{W}$ coming from the BBH asymptotic analysis used to describe the critical points of $\mathcal{F}$. Finally, we give answers to the conjectures of Jaffe and Taubes in the London limit and we extend them to more general cases. The third part is devoted to the complete study of the functional $\mathcal{G}$ comprising the interaction term with the external field. The vorticity is then no longer a fixed parameter as in the previous section but becomes a variable of the problem. The contents of this section covers part of S. Serfaty's PhD thesis, and the work she did in collaboration with E. Sandier.

2. STUDY OF THE FREE ENERGY FUNCTIONAL $\mathcal{F}$

2.1. The integrable or non-interacting case $\kappa = 1/\sqrt{2}$

In [JT], A. Jaffe and C. Taubes study the critical points on $\mathbb{R}^2$ of the free energy functional

$$\mathcal{F}_\kappa(u,A) = \int_{\mathbb{R}^2} |d_A u|^2 + \frac{\kappa^2}{2}|1-|u|^2|^2 + |dA|^2$$

and which are solutions to the Euler equations

$$\text{(1)} \qquad \left\{ \begin{array}{l} d_A^* d_A u = \kappa^2 u(1-|u|^2) \\ \\ d^* dA = (iu, d_A u) \end{array} \right.$$

where d_A^* is the operator acting on the 1-forms η such as $d_A^*\eta = d^*\eta + iA\wedge\eta$. Supposing that the intrinsic quantities $|d_A u|$, $|1-|u||$ and dA are decreasing (polynomially), the renormalized magnetic field is an integer N

$$N = \frac{1}{2\pi}\int_{\mathbb{R}^2} dA$$

which corresponds to the degree of $u/|u|$ on circles of sufficiently large radii. This is known as the homotopy class of the couple (u,A). For a given N, say $N \geq 0$, it has been observed by E.B. Bogomol'nyi [Bog] that, for the particular value $\kappa = 1/\sqrt{2}$, the functional $\mathcal{F}$ can be rewritten under the following form

$$\begin{aligned} \mathcal{F}_\kappa(u,A) \int_{\mathbb{R}^2} \quad & |\Re(d_{A_1}u) - \Im(d_{A_2}u)|^2 + |\Re(d_{A_2}u) - \Im(d_{A_1}u)|^2 \\ & + |*dA + \tfrac{1}{2}(|u|^2-1)|^2 + 2\pi N \end{aligned}$$

($\Re$ denotes the real part and $\Im$ the imaginary part of a complex number). One of the principal results of [JT] is then the following:

THEOREM 2.1 ([JT]). — *Any critical point (u, A) of $\mathcal{F}_{1/\sqrt{2}}$ of finite energy has a defined homotopy class N and verifies*

$$\mathcal{F}_{1/\sqrt{2}}(u, A) = 2\pi\ |N|. \tag{2}$$

In particular, it minimizes $\mathcal{F}_{1/\sqrt{2}}$ in its homotopy class.

The proof of this result can be understood as follows. Consider a critical point (u, A) of $\mathcal{F}_k$. From the Euler equations (1), we can deduce the following elliptic equations verified by the intrinsic quantities $1 - |u|^2$ and $h = *dA$

$$\begin{cases} -\Delta \dfrac{(1-|u|^2)}{2} + 2\kappa^2 |u|^2 \dfrac{(1-|u|^2)}{2} = |d_A u|^2 \\ \\ -\Delta h + |u|^2 h = (d_A u; i * d_A u) \end{cases} \tag{3}$$

where $-\Delta = d^* d$ and

$$(d_A u; i * d_A u) = -<\Re(d_A u), \Im(*d_A u)> + <\Im(d_A u), \Re(*d_A u)>$$

($< , >$ is the scalar product on the 1-forms). Using (3) we get from the maximum principle, on the one hand that $|u| < 1$ (unless $|u| \equiv 1$ on $\mathbb{R}^2$), and on the other hand that the intrinsic quantities $|h|$, $|d_A u|$ and $|1 - |u||$ decrease exponentially fast to infinity, so that one has in particular a well defined homotopy class for (u, A). These exponential decreases have an important physical interpretation linked to the mass of the Higgs particle.

Another important ingredient to prove theorem 2.1 is the conservation law of the energy-momentum tensor, which we shall use extensively throughout this talk. The energy-momentum tensor $(T_{ij})_{i,j \in \{1,2\}}$ is given by

$$T_{i,j} = 2\delta_{ij} |dA|^2 + 2(d_{A_i} u, d_{A_j} u) - \delta_{ij} f_\kappa(u, A), \tag{4}$$

where $f_k(u, A)$ is the free energy density $f_\kappa(u, A) = |d_A u|^2 + \frac{\kappa^2}{2} |1 - |u|^2|^2 + |dA|^2$. The conservation law of this tensor results from the fact that (u, A) is on the one hand a critical point of $\mathcal{F}_k$, and on the other hand C^∞, which can be deduced from the Euler equations (1); it is thus also a critical point for variations of the domain. Using Noether theorem, the translational invariance of the domain then gives rise to divergence free quantities, which constitute this conservation law

$$\forall j = 1, 2 \qquad \frac{\partial}{\partial x_i} T_{ij} = 0. \tag{5}$$

From this law, we can deduce the following Pohozaev identity, which is a consequence of the fact that (u, A) is a critical point of the infinitesimal action of dilations $r \frac{\partial}{\partial r}$ (i.e we multiply (5) by x_j, sum on j and integrate over $\mathbb{R}^2$)

$$\int_{\mathbb{R}^2} \frac{\kappa^2}{2}|1-|u|^2|^2 = \int_{\mathbb{R}^2} |dA|^2. \tag{6}$$

Let us restrict ourselves to the case $\kappa = 1/\sqrt{2}$. One can easily see that for that particular value of the parameter, by summing or subtracting the two equations of (3), the maximum principle leads to

$$|dA| = |h| \leq \frac{|1-|u|^2|}{2}. \tag{7}$$

Moreover, for $\kappa = 1/\sqrt{2}$, the Pohozaev identity becomes

$$\int_{\mathbb{R}^2} \left(\frac{1-|u|^2}{2} - *dA\right)\left(\frac{1-|u|^2}{2} + *dA\right) = 0. \tag{8}$$

Excluding the simple case where $|u| \equiv 1$ ($N = 0$) and thus $dA \equiv 0$ from (6), we have $|u| < 1$ and combining (7) and (8) gives either $*dA = \frac{1-|u|^2}{2}$ (for $N > 0$), or $*dA = -\frac{1-|u|^2}{2}$ (for $N < 0$). Bootstrapping for example $*dA = \frac{1-|u|^2}{2}$ in the equations, one easily obtains that $\Re(d_{A_1}u) = \Im(d_{A_2}u)$ and that $\Re(d_{A_2}u) = \Im(d_{A_1}u)$, which together with the observation of Bogomol'nyi proves the theorem.

Theorem 1 tells us that solving the solutions of the second order equations, comes up in the case $\kappa = 1/\sqrt{2}$ and for the class of configurations (u, A) of finite energy, to the study of the solutions of the first order equations (for $N > 0$)

$$*dA = \frac{1-|u|^2}{2} \tag{9}$$

$$\Re(d_{A_2}u) = \Im(d_{A_1}u) \tag{10}$$

$$\Re(d_{A_1}u) = \Im(d_{A_2}u), \tag{11}$$

which justifies the fact that this case is called integrable. The qualitative study of the solutions to these equations gives quite easily that the set of zeros of $|u|$ is constituted of a finite number of points where the index of u is strictly positive. We can thus represent it by exactly N points, not necessarily distinct from one another, $\{x_1 ... x_N\}$ each of multiplicity 1. We can then verify that $v = \log|u|^2$ is a solution of

$$-\Delta v + e^v - 1 = -4\pi \sum_{k=1}^{N} \delta_{x_k} \qquad \text{in } \mathcal{D}'(\mathbb{R}^2). \tag{12}$$

A convexity argument shows that the solution v to equation (12) is unique for any configuration of points $\{x_1 ... x_N\}$. We easily bootstrap this uniqueness of $|u|$ in equations (9), (10) and (11) to deduce the uniqueness (up to the gauge action) of the configuration (u, A). We can then go the other way round, and taking N arbitrary points in the plane, exhibit a solution to (9)...(11). We have then proven the theorem enunciated below.

Theorem 2.2 ([JT]). — *The space of the solutions of finite energy of the abelian Yang-Mills-Higgs (1) in the plane, in the integrable case $\kappa = 1/\sqrt{2}$, is nothing but (up to the action of the gauge group) the space of configurations of points in the plane, having integer multiplicities, all of the same sign.*

One of the striking points of the integrable case $\kappa = 1/\sqrt{2}$, is that not only can the vortices (zeros of u) be anywhere in the plane, but that the energy of the solutions is independent of the relative positions of the vortices, and is equal to 2π times the number of vortices. This is why the integrable case is also known as the non-interacting case. By handwaving arguments, A. Jaffe and C. Taubes conjecture that $\kappa = 1/\sqrt{2}$ is the limiting case between two opposite behaviors of vortices amongst themselves; for $\kappa < 1/\sqrt{2}$, whatever be their signs, the vortices have a tendency to attract one another, which justifies in some way the absence of a mixed phase, while in the case $\kappa > 1/\sqrt{2}$ the vortices with same signs should repel one another, but since their presence is imposed by the energy input due to the external field (see section 3), this leads to the possibility of a mixed state (see figure 1). Precisely the conjecture of Jaffe and Taubes for the repulsive case ($\kappa > 1/\sqrt{2}$) is as follows.

Conjecture 2.3 ([JT]). — *For $\kappa > 1/\sqrt{2}$, there exist stable solutions of YMH (1) on all $\mathbb{R}^2$ if and only if $|N| = 1, 0$; moreover they have an axial symmetry (up to the gauge action).*

In both of the next sub-sections, we shall rigorously account for these expected behaviors of the vortices amongst themselves in the strongly repulsive case ($\kappa \to +\infty$) and give a partial answer to the conjecture 2.3 (see theorem 2.10).

2.2. The strongly repulsive case or "London limit" $\kappa \to +\infty$: the BBH asymptotic analysis

From now on, we shall study the behavior of the vortices in the strongly repulsive case ($\kappa \to +\infty$). To prevent the vortices from separating from one another to infinity, we study $\mathcal{F}$ on a compact 2-dimensional manifold M without boundary, in order not to have to take artificial boundary conditions. For clarity, we shall restrict ourselves to the case where M is a flat torus. In fact the metric on M, as well as its topology, does not modify the qualitative aspects of the results. We also take a hermitian complex line bundle on M whose Euler class $e(E)$ verifies $\int_M e(E) = N > 0$, which implies that any section u intersecting transversely the zero section, does it algebraically N times. We thus fix the total vorticity of the problem, which becomes a parameter of the problem. In section 3 of this talk, we shall take into account the influence of the external field, so that the total vorticity will be once again a variable of the problem as in the original model. The existence of a couple section-connection (u_k, A_k) minimizing $\mathcal{F}_k$ in the Sobolev spaces $W^{1,2}$ of sections and connections (the expression in a trivialization of the bundle of (u, A) gives the functions and 1-forms $W^{1,2}$) is now a classical problem which requires the use of Coulomb gauges in order to render the functional coercitive (see [Uh]). We thus propose to study the behavior of such minimizing couples (u_k, A_k) when κ tends to infinity. The difficulty of such an analysis comes from the fact that we do not dispose of estimates *a priori* sufficient, independent of κ, in any functional space, to prove any weak convergence towards something. In particular, we can verify that $\mathcal{F}_\kappa(u_k, A_\kappa) \to +\infty$ and more precisely we have

$$\mathcal{F}_\kappa(u_\kappa, A_\kappa) \simeq 2\pi N \log \kappa. \tag{13}$$

F. Bethuel, H. Brezis and F. Hélein give a complete description of this asymptotic in [BBH] for the case $A = 0$ on a domain of $\mathbb{R}^2$ (this is completed in [St], adapted to the gauge invariant model in [BR1] and later in [Qin], and followed the study of the symmetric case on $\mathbb{R}^2$ in [BC]). They then establish the following result:

THEOREM 2.4 ([BBH]). — *Given a sequence of configurations (u_κ, A_κ) minimizing $\mathcal{F}_\kappa$ for a sequence of κ tending towards infinity, there exist a sub-sequence (u'_κ, A'_κ) and N distinct points $\{p_1...p_N\}$ of M such that*

$$(u_{\kappa'}, A_{\kappa'}) \longrightarrow (u_\star, A_\star) \qquad \text{in } C^k_{loc}(\tilde{M}) \tag{14}$$

where $\tilde{M} = M \setminus \{p_1...p_N\}$, $(u_\star, A_\star)$ is a couple unitary section-connection of E over $\tilde{M}$, which is a critical point of the functional

$$\mathcal{F}_\star(u_\star, A_\star) = \int_{\tilde{M}} |d_{A_\star} u_\star|^2 + |dA_\star|^2. \tag{15}$$

Moreover the index of the singular $A_\star$-harmonic section $u_\star$ at each p_j is $+1$, which gives in particular that the limiting curvature $h_\star = *dA_\star$ verifies the "London equation"

$$d^* dh_\star + h_\star = 2\pi \sum_{k=1}^{N} \delta_{p_k} \qquad \text{in } \mathcal{D}'(M). \tag{16}$$

REMARK 1. — *The positions of the limiting vortices $p_1...p_N$ determine uniquely the couple $(u_\star, A_\star)$ (up to the gauge invariance).*

We sketch the main points of the proof below—we omit the index κ.

Again, the maximum principle applied to the first equation of (3) gives $|u| \leq 1$. Combining this L^∞ bounding of $|u|$ and the bounding of the configuration energy (u, A) given by (13), by means of classical elliptic estimates (in the spirit of interpolation inequalities of the Gagliardo-Nirenberg type [BBH0]), we get the following control over the L^∞ norm of the covariant derivative of u (which is actually optimal)

$$\|d|u|\|_{L^\infty(M)} \leq \|d_A u\|_{L^\infty(M)} = O(\kappa). \tag{17}$$

The general strategy will consist in identifying and covering in the best possible way the zero set of $|u|$, which coincides with the region of loss of compactness of the sequence of configurations (u, A) in $W^{1,2}$. This region is called "bad set"; more precisely, it refers to the following set:

$$\mathcal{M} = \{x \in M \quad \text{such that} \quad |u|(x) < 1/2\} \tag{18}$$

(the constant 1/2 is in fact chosen arbitrarily between 0 and 1). We then have the following "quantization" result.

There exists $\delta > 0$ independent of κ such that, for all x_0 in M,

$$\kappa^2 \int_{B_{\kappa^{-1}}(x_0)} |1 - |u|^2|^2 \leq \delta \qquad \Longrightarrow \qquad x_0 \notin \mathcal{M}. \tag{19}$$

This result is an immediate consequence of (17). It implies that a point of $\mathcal{M}$ contributes, on a ball of radius κ^{-1}, to a "finite" part of the Higgs energy $\kappa^2 \int_{B_{\kappa^{-1}}(x_0)} |1 - |u|^2|^2$ (i.e larger than a value $\delta > 0$ independent of κ).

The joint use of the upper bounding of the energy and the conservation law of the energy-momentum tensor (5), from which we deduce the Pohozaev identity on any geodesic ball, gives in particular (cf [BR1])

$$\forall 0 < \alpha < 1 \quad \forall x \in M \qquad \kappa^2 \int_{B_{\kappa^{-\alpha}}(x_0)} |1 - |u|^2|^2 \leq C_\alpha, \tag{20}$$

where C_α is independent of κ. The use of the intermediate scales $\kappa^{-\alpha}$ in [BR1] between the natural scales of the problem (1 and κ^{-1}) enables us to get rid of the magnetic field in the Pohozaev identities and to obtain (20). These scales (which are not really necessary in the case $A = 0$ in dimension 2 [BBH]) are extensively used in the gauge invariant model ([BR1], [Ri1], [Ri2]...). Combining (19) and (20), we get the following covering of $\mathcal{M}$ on a whole ball of radius $\kappa^{-\alpha}$.

$$\forall x \in M \qquad \mathcal{M} \cap B_{\kappa^{-\alpha}}(x) \subset \cup_{j=1}^{N_\alpha} B_{\kappa^{-1}}(x_j) \quad \text{and } N_\alpha = O(1) \tag{21}$$

where $(x_j)_{j\in\{1\ldots N_\alpha\}}$ is a family of points in $B_{\kappa^{-\alpha}}(x)$ et N_α and N_α is bounded independently of κ.

Once again, the use of the stress-energy tensor (5) enables us to prove the quantization result below:

Lemma 2.5 (eta-compactness). — *There exists $\eta > 0$ independent of κ such that for any radius $1 \geq \rho \geq \kappa^{-1}$ and any x in M, one has*

$$\begin{aligned} &\int_{B_\rho(x)} f_\kappa(u, A) \leq \eta \ \log(\rho\kappa) \\ &\Longrightarrow B_{\rho/2}(x) \cap \mathcal{M} = \varnothing \end{aligned} \tag{22}$$

where $f_\kappa(u, A)$ is the free energy density $f(u, A) = |d_A u|^2 + \frac{\kappa^2}{2}|1 - |u|^2|^2 + |dA|^2$.

This lemma tells us that the contribution of the bad set to the total energy on a ball of radius ρ is at least greater than $\eta \ \log(\rho\kappa)$. The proof in 2 dimensions is quite straightforward (see [Ri1]), while in larger dimensions, it becomes much more technical (see [Ri2] [LR2]).

The combination of the bounding of the energy (13), (21) and the lemma of eta-compactness enables us to conclude easily that the set $\mathcal{M}$ is contained in a uniform number of bounded balls of radius κ^{-1}:

$$\mathcal{M} \subset \cup_{j=1}^{n} B_{\kappa^{-1}}(x_j) \qquad n = O(1). \tag{23}$$

We can then extract a sub-sequence of the original sequence $\kappa \to +\infty$ such that the family $(x_j)_{j=1..n}$ converges in M.

Outside $\mathcal{M}$, $h = *dA$ verifies the following elliptic equation deduced from the second Ginzburg-Landau equation (1)

$$d^*\left[\frac{1}{|u|^2}dh\right] + h = 0 \qquad \text{in } M \setminus \cup_{j=1}^{n} B_{\kappa^{-1}}(x_j). \tag{24}$$

What prevents this equation from being verified on all M is the index of u around the zero set. Then if $d_j = \deg(\frac{u}{|u|}; \partial B_{\kappa^{-1}}(x_j))$, one verifies that

$$\int_{\partial B_{\kappa^{-1}}(x_j)} \frac{dh}{|u|^2} + \int_{B_{\kappa^{-1}}(x_j)} h = 2\pi d_j. \tag{25}$$

From (24) and (25), we then have that h is a κ^{-1}-approximation of the following linear problem

$$d^* dk + k = 2\pi \sum_{j=1}^{n} \delta_{x_j}. \tag{26}$$

From (17), we can deduce that the d_j are uniformly bounded and thus that the $W^{1,p}$ norms ($p < 2$) of k are bounded independently of κ. These boundings can be transmitted easily to $h = *dA$ and by bootstrapping then in the Ginzburg-Landau equations (1), we prove theorem 2.4 (moreover to show that $n = N$, we have used the minimality of the solution).

As opposed to the non-interacting case $\kappa = 1/\sqrt{2}$, the vortices x_j and their limits $(p_1...p_N)$ cannot be anywhere in the domain M, the configuration $(p_1...p_N)$ minimizes a certain energy W of $M^N \setminus \Delta$ in $\mathbb{R}$, known as the renormalized energy (Δ denotes the diagonal of M^N).

THEOREM 2.6 ([BBH]). — *The configuration of limiting vortices $(p_1...p_N)$ given by theorem 2.4 is a fundamental state of the following function defined on $M^N \setminus \Delta$*

$$\begin{aligned} W(z_1...z_N) = \quad & 2\sum_{i\neq j} \int_M dG_i \cdot dG_j + 2\sum_{j=1}^{N} \int_M dR \cdot dG_j \\ & + \int_M |dR|^2 + \int_M |k|^2. \end{aligned} \tag{27}$$

The fact that the vortices strongly repel one another can be clearly seen in W. Indeed, we can verify that if two vortices z_i and z_j come close to each other, the others remaining fix, then we have $W \simeq 2\pi \log|z_i - z_j|$.

The proof of theorem 2.6 relies on the following decomposition. We take some $\delta > 0$ independent of κ. We divide M into two disjoint parts $M = \cup_{k=1}^{N} B_\delta(p_k) \cup M_\delta$. We decompose the total energy as a sum of the energies on each ball $B_\delta(p_k)$ and on M_δ. According to theorem 2.4, $|u|$ converges uniformly towards 1 while $*dA$ converges uniformly towards $h_\star$ which is solution of (16) on M_δ. Using then the second Ginzburg-Landau equation, we get

$$\lim_{\kappa\to+\infty} \int_{M_\delta} f_\kappa(u, A) = \int_{M_\delta} |dh_\star|^2 + |h_\star|^2. \tag{28}$$

Moreover, an explicit calculation enables us to verify that

$$\lim_{\delta\to 0} \int_{M_\delta} |dh_\star|^2 + |h_\star|^2 - 2\pi N \log \frac{1}{\delta} = W(p_1 ... p_N) + C, \tag{29}$$

where C is independent of the positions of the vortices. Using a covering of $\mathcal{M}$ by a finite number of balls of radii κ^{-1}, we can show by means of the convergence given by theorem 2.4 that the principal part of the energy around a limiting vortex is independent of its position and of the existence of the other vortices

$$\int_{B_\delta(p_k)} f_\kappa(u, A) = 2\pi \log(\delta\kappa) + C_0 + o_\delta(1), \tag{30}$$

where C_0 is a universal constant. By combining (28), (29) and (30) we can easily verify that to optimize $\mathcal{F}_\kappa$, we have to optimize the configuration of vortices $(p_1 ... p_N)$ with respect to W. This proves theorem 4.

Moreover, we have shown that $\mathcal{F}_\kappa(u_\kappa, A_\kappa)$ has the following asymptotic expansion.

$$\mathcal{F}_\kappa(u_\kappa, A_\kappa) = 2\pi N \log \kappa + W(p_1 ... p_N) + C_0 N + o(1). \tag{31}$$

This asymptotic expansion can be interpreted as follows: each vortex interacts with itself with an energy with principal part $2\pi \log \kappa$, W is the interaction energy amongst vortices, and finally C_0 is the renormalized energy of an isolated particle.

If we now follow a sequence (u_κ, A_κ) of Ginzburg-Landau solutions (1) in the "London limit" ($\kappa \to +\infty$), which is now not necessarily a sequence of minima, we prove results corresponding to theorems 2.4 and 2.6, the condition being to remain in "reasonable" energy levels $\mathcal{F}_\kappa(u_\kappa, A_\kappa) = O(\log \kappa)$. The number of limiting vortices $p_1 .. p_Q$ is now not necessarily N. To each vortex is associated an integer multiplicity d_j which is the index of the limiting singular section $u_\star$ around p_j. The equation verified by the limiting field $h_\star$ becomes

$$d^* dh_\star + h_\star = 2\pi \sum_{j=1}^{Q} d_j \delta_{p_j} \qquad \text{in } \mathcal{D}'(M) \tag{32}$$

while the positions of the p_j is a critical point of the function W given by (27) where G_j is replaced by $d_j G_j$.

2.3. The space of solutions in the "London limit"

Study of the zero set in the London limit. — We saw that in the London limit ($\kappa \to +\infty$), the zero set of the minimizing section u_k is forced to converge towards a very precise locus of points of the domain: $p_1...p_N$ which is a critical point of the function W. We now want to have a more complete description of this zero set, and eventually, to see in which limit the property obtained in the non-interactive case, saying that two solutions having the same zeros are gauge equivalent, remains valid in the strongly repulsive case. This would then bring the study of the space of solutions to that of the eventual zero set and would justify the tendency in physics to call Ginzburg-Landau vortices the solutions to the Ginzburg-Landau equations themselves.

Consider a sequence of minima (u_κ, A_κ) of $\mathcal{F}_k$ converging as in theorem 2.2 towards a couple section-singular connections $(u_\star, A_\star)$. We first look for the zero set of $|u_\kappa|$— we shall omit the index κ except where necessary. The limiting section being of index 1 to p_1, the sum of the indices of u on the bad balls $B_{\kappa^{-1}}(x_i)$ converging towards p_1 is also $+1$. We can always suppose that $|u|(x_j) < 1/2$.

If ever the centers x_l and x_k of two bad balls separate faster that $O(\kappa^{-1})$, we obtain a contradiction for the following reason. Suppose that this actually occurs, then up to the extraction of a sub-sequence, we can suppose that

$$\kappa|x_k - x_l| \longrightarrow +\infty. \tag{33}$$

Consider then the dilation of the section around the point x_k given by (in a local Coulomb gauge): $\hat{u}_\kappa(z) = u(\kappa^{-1}z + x_k)$. The convergence of the field dA established in the proof of theorem 2.4 as well as in the L^∞ bound of the covariant derivative (17) enable to conclude that, up to an extraction of a sub-sequence, $\hat{u}_\kappa$ converges in $C^1_{loc}(\mathbb{R}^2)$ towards a solution of

$$\begin{cases} \Delta\hat{u} + \hat{u}(1 - |\hat{u}|^2) = 0 \qquad \text{on } \mathbb{R}^2 \\[2ex] |\hat{u}|(0) < 1 \\[2ex] \displaystyle\int_{\mathbb{R}^2} |1 - |\hat{u}|^2|^2 < +\infty. \end{cases} \tag{34}$$

It is shown in [BMR] that the Higgs energy of the solutions of (34) is quantized:

$$\int_{\mathbb{R}^2} |1 - |\hat{u}|^2|^2 \;\in\; 2\pi\mathbb{N}^*. \tag{35}$$

Thus, if we go back to the original scale, for a sufficiently large radius R, independent of κ, with κ sufficiently large, then at the limit the contribution of the ball $B_{\kappa^{-1}R}(x_k)$ to the Higgs energy is at least 2π. This also holds for x_l and thus, since both points separate faster than $O(\kappa)$, from (33) we can deduce that for all $\delta > 0$ independent of

κ

$$\liminf_{\kappa\to+\infty} \kappa^2 \int_{B_\delta(p_1)} |1-|u_\kappa|^2|^2 \geq 4\pi. \tag{36}$$

Moreover, the Pohozaev identity on the ball $B_\delta(p_1)$ deduced from the energy-momentum conservation law, combined with the convergences of theorem 2.4, easily gives

$$\lim_{\kappa\to+\infty} \kappa^2 \int_{B_\delta(p_1)} |1-|u_\kappa|^2|^2 = 2\pi + o_\delta(1). \tag{37}$$

(36) and (37) are then in contradiction, and so (33) cannot occur even for a sub-sequence. Thus, there exists $\lambda > 0$ independent of κ such that the part of the bad set $\mathcal{M}$ converging towards p_1 can be covered by exactly one ball of radius $\lambda\kappa^{-1}$. So let x_1 be the center of this ball and more generally x_i be the center of the ball converging towards p_i and containing $\mathcal{M}$ in the vicinity of p_i. We can always suppose that $u(x_i) = 0$ since the limiting index of $u_\star$ is non-zero at p_i, but equal to 1, and that u must become 0 somewhere in the part of $\mathcal{M}$ which converges towards p_i which is contained in $B_{\lambda\kappa^{-1}}(x_i)$. The dilation argument previously applied from x_1 then says that, in a local Coulomb gauge, $\hat{u}_\kappa(z) = u_\kappa(\kappa z + x_1)$ converges in $C^1_{loc}(\mathbb{R}^2)$ up to the extraction of a sub-sequence towards a solution $\hat{u}$ of the following problem.

$$\begin{cases} \Delta\hat{u} + \hat{u}(1-|\hat{u}|^2) = 0 \qquad \text{in } \mathbb{R}^2 \\[2ex] |\hat{u}(0)| = 0 \\[2ex] \displaystyle\int_{\mathbb{R}^2} |1-|\hat{u}|^2|^2 < +\infty \\[2ex] \operatorname{ind}(\hat{u}, +\infty) = +1. \end{cases} \tag{38}$$

(It is shown in [BMR] that from the first three lines of (38), one can show that $|\hat{u}|$ converges uniformly towards 1 at infinity and thus that the index $\hat{u}$, $\operatorname{ind}(\hat{u}, +\infty)$ is well-defined.) Equation (38) is known as the “profile” equation of the Ginzburg-Landau vortices. The problem of the multiplicity of the solutions of (38) is not a standard problem in non-linear elliptic equations; indeed $\hat{u}$ does take real but complex values and the classical approaches using the maximum principle to show an eventual symmetry of the solution cannot apply here.

P. Mironescu in [Mi] gave the following proof of the uniqueness (up to rotations) of the “profile”.

It is quite standard to verify that there exists a unique solution of (38) of the form $\rho(r)e^{i\theta}$ ((r,θ) being the polar co-ordinates on $\mathbb{R}^2$). Let us divide $\hat{u}$, which is any solution of (38), by this axially symmetric solution. The ratio $w = \hat{u}/\rho(r)e^{i\theta}$ is a critical point on $\mathbb{R}^2 \setminus \{0\}$ of the functional

$$\mathcal{E}(w) = \int \rho^2 |\nabla w|^2 + 2\frac{\rho^2}{r}\left(iw, \frac{\partial w}{\partial \theta}\right) + \rho^4 |1-|w|^2|^2.$$

w is the critical point for the infinitesimal action of dilations $r\frac{\partial}{\partial r}$ on $\mathbb{R}^2 \setminus \{0\}$. This then gives the Pohozaev identity

$$0 = \int_{\mathbb{R}^2} r\frac{\rho'}{\rho}\left|\frac{\partial w}{\partial r}\right|^2 + \frac{r\rho\rho' + \rho^2}{2}|1-|w|^2|^2 \tag{39}$$

(the index 1 being used at infinity to get rid of the boundary terms). A simple study of the modulus of the radial solution ρ enables to conclude that $\rho' > 0$. The identity (39) then tells us that $|w| = 1$, $\frac{\partial w}{\partial r} = 0$ and thus $\hat{u} = \rho(r)e^{i(\theta+\alpha)}$ where α is a constant. This proves the uniqueness of the profile up to rotations.

Since $\hat{u}_\kappa$ converges (up to the extraction of a sub-sequence) in $C^1_{loc}(\mathbb{R}^2)$ towards this profile (unique up to rotations), also since $\hat{u}$ becomes zero exactly at 0 and since $\nabla^2\hat{u}(0)$ has rank 2, one can then verify that on $B_\lambda(0)$ $\hat{u}_k^{-1}(\{0\}) = \{0\}$. Thus on the ball $B_\delta(p_1)$, u is equal to zero at a unique point.

We have up to now used the fact that the couple (u_κ, A_κ) is minimal in order to get the convergence of theorem 2.4 and the index 1 of $u_\star$ at each p_j. From the extension of the results 2.4 and 2.6 to any of the critical configurations $(u_\kappa, A_{,\kappa})$ under the energy levels $O(\log(\kappa))$ described at the end of section 2.2, we have shown the following proposition.

Proposition 2.7. — *Let (u_κ, A_k) be a sequence of configurations ($\kappa \to +\infty$), solutions of the Ginzburg-Landau equation* (1), *verifying the energy bounding $\mathcal{F}_\kappa(u_\kappa, A_\kappa) = O(\log \kappa)$ and converging towards a couple section-connection $(u_\star, A_\star)$ singular at $p_1 ... p_Q$. If the index of $u_\star$ at p_j is ± 1 for all j, then there exists κ_0 such that for $\kappa > \kappa_0$, we have*

$$|u_\kappa|^{-1}(\{0\}) = \{x_1 ... x_Q\} \qquad \textit{and} \qquad x_j - p_j \to 0 \;\; \textit{for all } j.$$

The role of the renormalized energy W. — Let us again place ourselves in the hypotheses of proposition 2.7. The section u_κ for κ sufficiently large thus intersects exactly Q times the zero section with each time an intersection index ± 1. Each of those zero points of $|u_k|$ converges towards a limiting vortex $p_1 ... p_Q$. Considering how important the part played by the position of the zero is in P. Mironescu's uniqueness argument, a natural question is whether, more generally, two solutions of (1) having the same zero sets in the London limit, are equal (up to the gauge invariance). This proposition would generalize to the strongly repulsive case the same proposition proved above by Jaffe and Taubes for the non-interacting case. The answer to this question is yes under the same hypotheses as in proposition 2.7.

PROPOSITION 2.8 ([PR]). — *Let* (u_κ, A_κ) *and* (v_κ, B_κ) *be two sequences of solutions of the Ginzburg-Landau equations* (1) *in the London limit* ($\kappa \to +\infty$) *verifying the common energy bounding* $\mathcal{F}_\kappa(u_\kappa, A_k) = O(\log\kappa)$, $\mathcal{F}_\kappa(v_k, B_k) = O(\log\kappa)$, *and both converging towards the same singular configuration* $(u_\star, A_\star)$. *We suppose that the index of* $u_\star$ *at each singularity is* ± 1. *Then there exists* κ_0 *such that for* $\kappa > \kappa_0$

$$|u_\kappa|^{-1}(\{0\}) = |v_\kappa|^{-1}(\{0\}) \qquad \Longrightarrow \qquad (u_\kappa, A_\kappa) \simeq (v_\kappa, B_\kappa),$$

where $\simeq$ *denotes the gauge equivalence.*

The proof of the preceding proposition relies amongst other things on the following generalization of P. Mironesu's argument. In a Coulomb gauge in the vicinity of any limiting vortex p_k, we reconsider under the hypotheses of the proposition, the ratio of the two solutions $w = \frac{u}{v}$. Instead of making dilations with respect to the conformal field $X = r\frac{\partial}{\partial r}$ centered at the common zero of u and v, which do not give any results in that case, we dilate with respect to the following field, constructed from the second solution of v,

$$Y = |v|^2 \frac{(iv, *dv)}{|(iv, *dv)|^2}$$

where we identify the field with the dual 1-form given by the scalar product. We can observe that Y coincides with the usual conformal dilation field $X = r\frac{\partial}{\partial r}$ in the case where v has a radial symmetry. The action of this field on the functional for which (w, A) is a critical point gives a Pohozaev identity in the vicinity of p_k, which all put together enable us to conclude that for sufficiently large κ $(u_\kappa, A_\kappa) \simeq (v_\kappa, B_\kappa)$.

The use of Y rather that the usual field $X = r\frac{\partial}{\partial r}$ is *a posteriori* natural in view of the functional for which (w, A) is a critical point and of the property of $|v|$-conformality of Y (see [PR]).

The proposition 2.8 is the first step to describe the space of solutions below the $O(\log\kappa)$ energy levels in the London limit. The second step consists in constructing a sequence (v_κ, B_κ) converging towards a $(u_\star, A_\star)$ whose configuration of associated vortices $p_1...p_Q$ is any critical point of W (non-degenerate up to the actions of isometries of M) and whose multiplicities d_j are ± 1. Such a construction has been done in [LL] for $d_j = +1$ by means of variational arguments, by establishing a link between the level sets of the function W and the functional $\mathcal{F}_\kappa$. In [PR] this construction is done for $d_j = \pm 1$ by means of the local inversion theorem which also has the advantage to bring the local uniqueness. The argument of linearization around a solution is rendered complex by the existence of a "kernel at infinity": the inverse of the linearized operator of (1) around an accumulation of vortices which converges towards $(u_\star, A_\star)$, blows-up in the standard norms $W^{2,2} - L^2$ when κ tends to infinity. This is due to the action of the invariance group of the limiting profile equation (38) (here the isometries of the plane). We are thus led to develop the Lyapunov-Schmidt reduction argument which first consists in working in a direction perpendicular to this kernel at infinity, which we reintroduce in the final non-linear argument. This technique applied to the elliptic P.D.E. has often been used in a number of constructive problems of differential geometry like the existence of minimal surfaces, the existence of surfaces of constant mean curvature, the Yamabe problem, the Yang-Mills monopoles... by N. Kapouleas, R. Mazzeo, F. Pacard, R. Schoen, K. Uhlenbeck, S.T. Yau... with each time new difficulties. (v_κ, B_κ) being thus constructed benefits from a local uniqueness in an adapted functional space. If (u_κ, A_κ) is another solution of (1), converging towards the same limiting configuration whose singularities are $p_1...p_Q$, we know (proposition 2.7) that the zeros of $|u_\kappa|$ converge towards those singularities and are thus close to those of $|v_\kappa|$. The argument of the proof of proposition 2.8 is then converted in order to obtain, no longer directly that $(u_\kappa, A_\kappa) \simeq (v_\kappa, B_\kappa)$, a sufficiently small bound of the "separation" of (u_κ, A_κ) and (v_κ, B_κ) which, by means of the local uniqueness established previously, enables to conclude that $(u_\kappa, A_\kappa) \simeq (v_\kappa, B_\kappa)$. We thus have the following theorem.

THEOREM 2.9 ([PR]). — *Let $(p_1...p_Q)$ be a critical point of the renormalized energy W for the multiplicities $(d_1...d_Q)$ in $\{-1, +1\}$. Suppose that this critical point is non-degenerate up to the action of isometries of M. Let $(u_\star, A_\star)$ be the singular couple section-connection associated to (p_j, d_j). Then let (u_κ, A_κ) and (v_κ, B_κ) be two sequences of Ginzburg-Landau equations* (1) *verifying the common energy bounding $\mathcal{F}_\kappa(u_\kappa, A_\kappa) = O(\log\kappa)$ and $\mathcal{F}_\kappa(v_\kappa, B_\kappa) = O(\log\kappa)$, if their singular closure set is the configuration $(u_\star, A_\star)$, then for sufficiently large κ*

$$(u_\kappa, A_\kappa) \simeq (v_\kappa, B_\kappa)$$

up to the action of isometries of M ($\simeq$ is the gauge equivalence).

Comments. — Theorem 2.9 tells us that under certain hypotheses, the study of the critical points of the functional $\mathcal{F}_\kappa$ in the London limit—*infinite dimensional* problem—can be brought to that of the critical points of the function W—*finite dimensional* problem. The limits of our description of solutions to the Ginzburg-Landau equations in the strongly repulsive case are the following. First, in order to be in the context of the BBH asymptotic, we have restricted the critical points to energies below $C\log\kappa$. How about other eventual solutions to (1)? In the case $A = 0$ on a star-shaped domain, it has been shown that *all* the Ginzburg-Landau solutions are under a certain energy level $C\log\kappa$. The second restriction in the hypotheses of theorem 2.9 is the constraint on the limiting multiplicities d_j who have to belong to $\{+1, -1\}$. How about the branches of solutions linking a critical point of W where some of $|d_j|$'s are greater than 1? This question is completely open. Neither the variational methods, in the spirit of [LL], or [AB1], [AB2], nor critical points methods, seen above, have enabled up to now the construction of solutions (other than the axially symmetric solutions) converging towards a critical point of W with multiplicities $d_j \neq \pm 1$. We do not even have a precise idea of the aspect of such solutions near the vortices. The understanding of the possible limiting "profiles" is also lacking in that case (see conjectures on that theme in [OS]). Nevertheless, it seems that the case $d_j = \pm 1$ is "generic" in the following sense: for a "generic" perturbation of the metric, a critical point of W having multiplicity indices greater than 1 should be transformed into a critical point whose indices are ± 1.

To end up the study of the free energy, we should mention that the techniques presented above help to bring a partial answer to the Jaffe and Taubes conjectures on all $\mathbb{R}^2$ in the strongly repulsive case.

THEOREM 2.10 ([Ri3]). — *The conjecture* 2.3 *is true in the London limit (sufficiently large κ) if we replace "stable critical point" by "minimum in the homotopy class".*

3. INTERACTION WITH AN EXTERNAL MAGNETIC FIELD; TOWARDS THE ABRIKOSOV LATTICES

3.1. Introduction

Up to now, we have studied the Ginzburg-Landau free energy $\mathcal{F}_\kappa$ without the interaction term with the magnetic field $2h_e \int dA$. Moreover, we have placed ourselves on a plane torus in order to ignore in a first approach, the border effects which are however very important in phenomenological studies, and about which it would be very interesting to recover the conclusions by the mathematical study itself. Finally, in the absence of external magnetic field, in order not to work with a trivial fundamental state, we have imposed a global vorticity through the choice of a certain bundle on our torus of non zero Euler class. The analysis of our model problem has enabled us to isolate and understand the mechanism

$$\text{vorticity} \qquad \Longrightarrow \qquad \text{vortex formation}$$

and to bring the study of the space of solutions to that of the positions of the vortices, governed by the renormalized energy.

In her PhD thesis [Se1], then in her works in collaboration with E. Sandier [SS1] [SS2], S. Serfaty considers the complete functional $\mathcal{G}_\kappa$ including the interaction term with the external magnetic field, on Ω a simply connected bounded domain of $\mathbb{R}^2$. The vorticity is thus free and should spontaneously appear by increasing the external magnetic field and thus the total free energy of the system as it has been observed for type II superconductors ($\kappa > 1/\sqrt{2}$).

We still place ourselves in the London limit ($\kappa \longrightarrow +\infty$) and the magnetic field h_e is assumed to be uniform (we will confuse the 2-form and the corresponding constant).

3.2. The Meissner solution

When the applied field $h_e = 0$, the fundamental state of $\mathcal{G}_\kappa$ is clearly reached by the *pure state* solution $h = *dA = 0$. When we increase the field h_e, the superconducting character of the sample changes a little without however forming any vortex ($|u_\kappa| \geq 1/2$ for sufficiently large κ). We expect this configuration which is vortex-free, stable, and absolute minimum for not too strong external fields, to be unique. This is what S. Serfaty shows by combining energy estimates and convexity arguments.

THEOREM 3.1 ([Se3]). — *In the London limit (κ sufficiently large), there exists a unique stable, vortex free ($|u_\kappa| \geq 1/2$) critical point of $\mathcal{G}_\kappa$ which minimizes $\mathcal{G}_\kappa$ amongst the vortex-free critical points, with the condition $h_e = O(\kappa^\alpha)$ where α is a positive constant. This solution known as the "Meissner solution" verifies in particular*

$$\begin{cases} -div\left(\dfrac{\nabla h}{|u|^2}\right) + h = 0 & \text{in } \Omega \\ h = h_e & \text{on } \partial\Omega. \end{cases} \tag{40}$$

The constant $1/2$ is of course a constant chosen at random between 0 and 1. In fact the Meissner solution remains stable up to a critical value of the applied field $h_e = H_{sh} \simeq C\kappa$ known as the "super-heating field", and above which it is no longer stable. This separation has been studied in detail by H. Berestycki, A. Bonnet and J. Chapman in [BBC]. Moreover, the uniqueness of the Meissner solution in the vicinity of the field H_{sh} is proven by A. Bonnet, J. Chapman and R. Monneau in [BCM].

What is then natural to ask oneself is for which value H_{c_1} of the external field h_e, the Meissner solution ceases to be a minimum of $\mathcal{G}_\kappa$. The calculations of Abrikosov on all $\mathbb{R}^2$ predict $H_{c_1} \simeq \frac{\log \kappa}{2}$. As we shall see, the work of S. Serfaty has enabled to show, that following a conjecture in [BR2], H_{c_1} is in fact larger on a bounded domain. This is a consequence of border effects and is more precisely as follows: in the principal term of H_{c_1}, the factor $1/2$ in front of $\log \kappa$ has to be replaced by $k_1 = 1/2 \max |\xi_0|$ where ξ_0 is the solution of

$$\begin{cases} \Delta \xi_0 = \xi_0 + 1 & \text{in } \Omega \\ \xi_0 = 0 & \text{on } \partial\Omega. \end{cases} \tag{41}$$

ξ_0 is also solution of $-\Delta^2 \xi_0 + \Delta \xi_0 = 0$. We can easily verify using the maximum principle that $\Delta \xi_0 > 1$ and $\xi_0 < 0$ inside Ω and thus that $|\xi_0| < 1$. E. Sandier and S. Serfaty then show the following result.

THEOREM 3.2 ([SS1]). — *There exists a constant $C_1 > 0$ such that if*

$$h_e \leq k_1 \log \kappa - C_1 \log \log \kappa$$

the Meissner solution is an absolute solution of $\mathcal{G}_\kappa$ and there exists C_2 such that if

$$h_e \geq k_1 \log \kappa + C_2$$

the Meissner solution is no longer an absolute minimum of $\mathcal{G}_\kappa$.

It is then tempting to place oneself in the vicinity of $k_1 \log \kappa$ and to try to observe the vortices appearing one after the other as the external field is increased. Technically, this implies that we should work with the absolute minimum of $\mathcal{G}_\kappa$ for which we would try to cover precisely the "bad set" $\mathcal{M}$, in the spirit of the BBH asymptotic, by means of uniformly bounded number of balls of radii κ^{-1} of which we would then study the minimal value (eventually 0) and the energetically most advantageous respective positions.

With the above scheme, we unfortunately come across a major difficulty which has not yet been solved and which is exposed in the next section. The BBH analysis was built on the initial estimate of the free energy $\mathcal{F}_\kappa(u_\kappa, A_k) = O(\log \kappa)$ which can be expressed by a bound of the number of vortices *a priori* independent of κ (knowing that the cost for one "isolated" vortex is about $2\pi \log \kappa$). In the case we are actually considering, where the total vorticity is moreover not under control and has become a variable of the problem, the bound of the energy is *a priori*

$$|\mathcal{G}_k(u_\kappa, A_\kappa)| = O(\log^2 \kappa). \tag{42}$$

From that we cannot deduce a better estimate than $\mathcal{F}_\kappa(u_\kappa, A_\kappa) = O(\log^2 \kappa)$ for the free energy part. This thus tells us that the number of vortices that have to be considered *a priori* is $O(\log \kappa)$ (which must be true for a field $h_e = 2k_1 \log \kappa$, but not for $h_e = k_1 \log \kappa + O(1)$ where we expect rather a uniformly bounded number of vortices). The difficulty of working with an *a priori* so large number of vortices is responsible for the fact that we have to get out the BBH asymptotic scheme and that the lack of precision is larger as can be seen between $k_1 \log \kappa - C \log \log \kappa$ and $k_1 \log \kappa + C_2$ in theorem 3.2 which is yet a very nice achievement.

The proof of theorem 3.2 has its roots in the asymptotic arguments of the previous section. The point is to realize the best possible lower bound of the total energy of a critical configuration with vortices and to show that for a sufficiently low external field, the Meissner solution is energetically preferable. In order to obtain such a lower bound, we begin by renormalizing the induced field by "extracting" the external field described following [BR2]. Let ξ be the solution $\Delta \xi = h$ on Ω, which is equal to 0 on $\partial\Omega$ ($A = *d\xi$ is the Coulomb gauge of h on Ω), we note ζ the difference $\zeta = \xi - h_e \xi_0$ (where ξ_0 is given by (41)). $*d\zeta$ is the Coulomb connection on all Ω, from which we have removed the influence of the external field. We show without much difficulty that the energy of the minimal configuration (u_κ, A_κ) can be decomposed as follows:

$$\begin{aligned}\mathcal{G}_\kappa(u,A) \quad &= \mathcal{G}_\kappa(1, h_e * d\xi_0) + \mathcal{F}_\kappa(u, *d\zeta) + \int_\Omega |\Delta\zeta|^2 \\ &-2h_e \int_\Omega (iu, du) \wedge d\xi_0 + o(1).\end{aligned} \tag{43}$$

We suppose then that $h_e = O(\log\kappa)$ and we proceed to a decomposition of the domain, in the same spirit as that used to prove theorem 2.6, but more complex in the present case where the "bad set" is *a priori* far much bigger. We shall have $\Omega = \Omega_s \cup \Omega_v$ where Ω_s and Ω_v are disjoint and Ω_v will be a union of disjoint balls $\{B_{r_i}(a_i)\}_{i\in I}$ covering $\mathcal{M}$, the "bad set", and containing as much self energy as possible, without taking into account the interaction energy of the vortices amongst themselves as in the proof of theorem 2.6. Precisely, by means of a technique introduced by R. Jerrard [J], we show (see [SS1]) the existence of a family of balls $\mathcal{B} = \{B_{r_i}(a_i)\}_{i\in I}$ such that

– $\mathcal{M} \subset \cup_{i\in I} B_{r_i}(a_i)$
– $\displaystyle\int_{B_{r_i}(a_i)} f_\kappa(u, *d\zeta) \geq 2\pi|d_i|\,(\log\kappa - O(\log\log\kappa))$ where $d_i = \deg\Big(\frac{u}{|u|}; \partial B_{r_i}(a_i)\Big)$
– $r_i = O(\log^{-6}\kappa)$
– $\operatorname{Card} I = O(\log^2\kappa)$.

We then choose $\Omega_v = \cup_{i\in I} B_{r_i}(a_i)$. The $O(\log^{-6}\kappa)$ radii replace the radii $\delta = O(1)$ of the proof of theorem 2.6 since the number of vortices tends towards infinity *a priori* as $O(\log\kappa)$. The role of the a_i is a bit like intermediate vortices of degree d_i and this choice of covering gives in particular

$$\int_\Omega (iu, du) \wedge d\xi_0 = 2\pi \sum_{i\in I} d_i \xi_0(a_i) + o(1). \tag{44}$$

Combining (43), the choice of Ω_v, (44) and the minimizing character of (u, A) for $\mathcal{G}_\kappa$, we have

$$0 \geq 2\pi \sum_{i\in I} |d_i|(\log\kappa + O(\log\log\kappa)) - 4\pi\, h_e\, \max|\xi_0| \left(\sum_{i\in I} |d_i|\right) \tag{45}$$

and so if $\sum_{i\in I} |d_i| \neq 0$, we must get $h_e \geq k_1 \log\kappa + O(\log\log\kappa)$, which establishes the first part of theorem 3.2. The second part will be a consequence of theorem 3.3.

This technique of extracting the external magnetic field and then decomposing the domain in order to optimize the lower bound of the energy is a recurrent theme in the work of E. Sandier and S. Serfaty and appears through different forms to prove the theorems below.

3.3. Family of stable solutions having $1, 2, 3$...vortices; the blooming of the lattice

For a while, let us give up the idea of considering the absolute minimum of $\mathcal{G}_\kappa$, but yet keeping an external field in the vicinity of $k_1 \log\kappa$.

In her first work [Se1], S. Serfaty constructs a family of stable solutions which have a finite number of vortices for the fields $1 \ll h_e \ll \kappa^\alpha$ (α being positive non explicit constant). To counteract the difficulty of showing evidence of a finite number of vortices for the fundamental states themselves, S. Serfaty makes an ansatz which consists in considering, instead of the absolute minimum (u_κ, A_κ) of $\mathcal{G}_\kappa$ in all $W^{1,2}(\Omega,\mathbb{C}) \times W^{1,2}(\Omega,\mathbb{R}^2)$, the (u_κ, A_κ) contained in the following set.

$$\mathcal{D}_n = \left\{u \in W^{1,2}(\Omega,\mathbb{C}) \quad \text{such that} \quad 2\pi n \log\kappa + B \le F_\kappa(u_T) \le (2\pi n+1)\log\kappa\right\}$$

where

$$F_\kappa(u_T) = \int_\Omega |du_T|^2 + \frac{\kappa^2}{2}|1-|u_T|^2|^2.$$

u_T is the projection of the function u on the unitary disc ($u_T = u$ if $|u| \le 1$ and $u_T = u/|u|$ otherwise) and B is a negative constant which remains fixed throughout the proof. The idea is that, in the Coulomb gauge on Ω, the cost for a vortex is $(F_\kappa(u_T) \simeq 2\pi\log\kappa)$. Thus, restraining to $\mathcal{D}_n$, we *a priori* constrain the configuration (u, A) not to have more than n vortices so that we can get back to the BBH asymptotic in order to establish a complete description of (u, A). The main difficulty of this approach comes from the fact that it is not at all evident that a minimum of $\mathcal{G}_\kappa$ in $\mathcal{D}_n$, once we have established its existence, should be in the interior of $\mathcal{D}_n$ and thus be a solution of the Ginzburg-Landau equation (1). Suppose that we have proven the existence of a minimum (u_κ, A_κ) of $\mathcal{G}_\kappa$ in $\mathcal{D}_n$. We are tempted to use once more the arguments of section 2 to cover the "bad set" $\mathcal{M}$... in order to establish an asymptotic expansion of $\mathcal{F}_\kappa(u_\kappa, A_\kappa)$ as in (31) and verify that (u_κ, A_k) is well in the interior of $\mathcal{D}_n$ (i.e. $2\pi n\log\kappa + B < F(u_T) < (2\pi n+1)\log\kappa$) and thus solution of (1). The difficulty of this approach resides in the fact that as long as we do not know that (u, A) is solution of (1), we do not have an estimate of the form $\|d|u|\|_{L^\infty} = O(\kappa)$, no longer dispose of the quantization result (19) and the set $\mathcal{M}$ can be very diffuse, etc. We also no longer have any control of its size, which was given in section 2 by Pohozaev identities. The idea is then, not to work any longer with (u, A) itself, but with its parabolic regularization v which minimizes

$$\min_{v\in H^1(\Omega,\mathbb{C})} \int_\Omega |\nabla v|^2 + \frac{\kappa^2}{2}|1-|v|^2|^2 + \kappa^{2\gamma}\int_\Omega |u_\kappa - v|^2$$

for some $\gamma > 0$. The advantage of replacing u by its parabolic regularisation v is that (v, A) will both be energetically very close to (u, A) and benefit from the properties of the minimum of $\mathcal{F}_\kappa$ in section 2 (such as $|v| \le 1$, $|d|v|| = O(\kappa)$...) which come from the elliptic equation verified by v. This idea of the parabolic regularization has been introduced in the context of Ginzburg-Landau equations by L. Almeida and F. Bethuel in [AB2], in order to define an approximated configuration of vortices for any function u in the energy zone $F(u) = O(\log\kappa)$. We then establish an asymptotic expansion for v in the spirit of section 2.2, which enables to deduce that (v, A) (and thus (u, A) too) which is energetically close to it, is in the interior of $\mathcal{D}_n$. The existence of a minimum $\mathcal{G}_k$ in $\mathcal{D}_n$ also rests on the use of the parabolic regularization. We then prove the following theorem which we present in the case where Ω is a disc centered at 0: $\Omega = B_R(0)$.

THEOREM 3.3 ([Se1]). — *Let D be an arbitrary positive constant, $h_e(\kappa)$ a function of κ tending towards $+\infty$ at infinity and verifying $h_e(\kappa) \leq \kappa^\alpha$, where α is a positive constant. Then there exists a positive constant $\kappa_0(D)$ such that for $\kappa \geq \kappa_0(D)$ and for all $n \in \mathbb{N}^*$, $n < \frac{D}{\pi}$, there exists a family of stable critical points (u_κ, A_k) of $\mathcal{G}_\kappa$, solution of the Ginzburg-Landau equations* (1), *verifying*

– $|u_\kappa|^{-1}(\{0\}) = \{a_1^\kappa ... a_n^\kappa\}$ *where* $a_1^\kappa ... a_n^\kappa$ *are isolated points of* $B_R(0)$.

– *For all* $j = 1...n$, *ind*$(u_\kappa, a_j^\kappa) = +1$.

– *The configuration* $\tilde{a}_j = a_j\sqrt{h_e(\kappa)}$ *converges, up to the extraction of a subsequence, towards a configuration of n points of $\mathbb{R}^2$ minimizing*

$$W(x_1...x_n) = -2\pi \sum_{i \neq j} \log|x_i - x_j| + 2\pi\ddot{\xi}_0(0) \sum_{i=1}^n |x_i|^2. \tag{46}$$

– *The asymptotic expansion of* $\mathcal{G}_k(u_\kappa, A_\kappa)$ *is*

$$\begin{aligned} \mathcal{G}_\kappa(u_\kappa, A_\kappa) = \mathcal{G}_\kappa(1, h_e * d\xi_0) \quad & +2\pi n \left(|\log\kappa - \frac{h_e}{k_1}| \right) + \pi(n^2 - n)\log h_e \\ & +W(\tilde{a}_1...\tilde{a}_n) + Q_n + o(1) \end{aligned} \tag{47}$$

where Q_n only depends on n.

This theorem is very rich in information which have to be extracted one by one, according to the questions posed in the introduction of this talk.

The information on the location of the zeros is not as precise in Serfaty's original work but can be deduced from the arguments in part 2.3 of the present survey.

As opposed to the case without border of section 2, the vortices here tend towards a same point which coincides with the point where $|\xi_0|$ is maximum. Had there been several points where $|\xi_0|$ is maximal, the vortices would have separated to form groups of n_j vortices, in an optimal way amongst those points (we optimize $\sum_j(n_j^2 - n_j)$).

We can calculate, as a function of h_e, the number of vortices n which optimize the energy and find through the Serfaty ansatz $F(u_T) \leq 2\pi D$ a very precise estimate of H_{c_1}. We can clearly see that the optimal number of vortices is an increasing function of h_e for fixed κ (large κ). We can easily verify the existence of $k_2 > 0$ such that, for $h_e > k_1 \log\kappa + k_2$, the Meissner solution ceases to be minimizing amongst those solutions with a finite number of vortices, which proves the second part of theorem 2.2. More generally, denoting by H_n the value of the field where the solution with n vortices is energetically preferable, we calculate $(n > 1)$

$$H_n - k_1 \ \log\kappa \simeq (n-1)k_1 \ \log\log\kappa. \tag{48}$$

Figure 2 graphically represents those results.

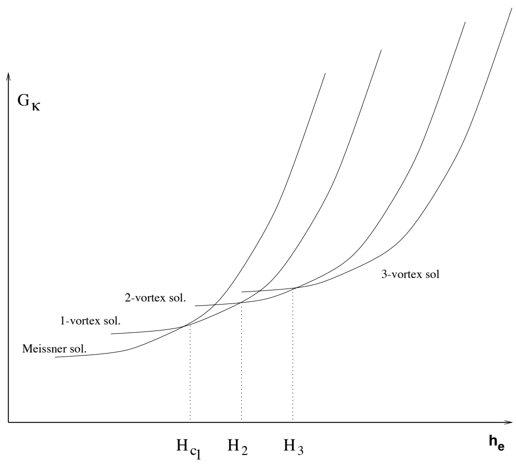


FIGURE 2. Solutions with 1,2,3... vortex

To know if the solution with n vortices for $H_n \leq h_e \leq H_{n+1}$ is a fundamental state on the whole space is an open seemingly difficult problem discussed in the next subsection. An important point is also the stable character of these solutions proven for $1 \ll h_e \ll \kappa^{\alpha}$ and which we have seen above for the Meissner solution and the "super-heating" field. This stability is responsible for hysteric phenomena observed in experiments. The stable character of the solutions having a finite number of vortices for fields smaller than H_{c_1} has been observed by Q. Du and F.H. Lin in [DL].

Although the expression for the renormalized energy W is relatively simple, the eventual symmetries of its fundamental states, which constitute the bloomings of the Abrikosov lattices, are still to be completely explored. In [GS], S. Gueron et I. Shafrir have done stability analyses of the symmetric configurations as well as numerical studies. They have made the following observations on the probable fundamental states (see figure 3)

– $n \leq 3$: these are regular polygons centered in 0

– $7 \leq n \leq 10$: these are regular, star-shaped (regular polygon + origin)

– $4 \leq n \leq 6$: the two previous types of configurations are locally minimizing and are observed

– $n \geq 11$: these are clusters of regular concentric polygons which "converge" towards a triangular lattice centered in 0 when n becomes large.

3.4. An interesting open problem: remove S. Serfaty's Ansatz

It would be nice to really characterize the phase change at H_{c_1} through a rigorous proof of the spontaneous creation of vorticity for absolute minimizers. In other words an interesting question would be to remove S. Serfaty ansatz and to replace the

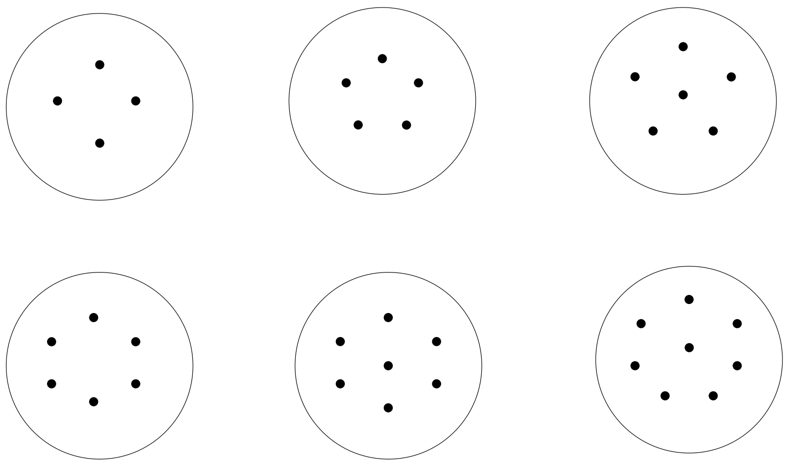


FIGURE 3. Blooming of Abrikosov Lattices.

constrained minimization of $\mathcal{G}_\kappa$ inside $\mathcal{D}_n$ by the minimization of $\mathcal{G}_\kappa$ inside the whole space $W^{1,2}(\Omega,\mathbb{C}) \times W^{1,2}(\Omega,\mathbb{R}^2)$. The hope is to prove that the solutions obtained in theorem 3.3 are the absolute minimizers of $\mathcal{G}_\kappa$ for the various ranges of external fields given by (48).

There is an interesting discrete model derived from the original one in the large $\kappa-$limit. Assume each vortex has zero size and $|u| \equiv 1$ out of the vortices located at $x_1 ... x_N$ with vorticity $d_1 ... d_N$. Following [BR2] one can extract the main terms in the energy (43):

$$\mathcal{H}(x_i,d_i) = -\sum_{i\neq j} d_i d_j \ \log|x_i - x_j| + \sum_i d_i^2 \ \log\kappa - 2h_e \sum_i d_i \ \xi_0(x_i). \tag{49}$$

One of the major step in understanding fully the creation of vortices and the phase transition at H_{c_1} would be to show that a minimizing configuration $(x_i,d_i)_{i=1...N}$ of $\mathcal{H}$ for an external field of the order $h_e = H_{c_1} + O(\log\log\kappa) = k_1 \log\kappa + O(\log\log\kappa)$ verifies

$$N = O_\kappa(1) \qquad \text{as} \qquad \kappa \to +\infty \tag{50}$$

or at least

$$\sum_{i=1}^N d_i = O(1). \tag{51}$$

This question is the main issue in describing the formation of vortices. If one assumes that all the d_i's have the same sign, answering to this question seems reasonably easy. The difficulty here comes from possible mixing of $+$ and $-$ creating negative clusters which are not *a-priori* energetically less favorable.

3.5. The second renormalization: the free boundary problem

In this last section, come back on the study of the fundamental states of $\mathcal{G}_\kappa$ but for much more intense fields that H_{c_1}, inside the mixed phase (see figure 1) and still in the London limit ($\kappa \to +\infty$). In that case, the number of vortices will *a priori* tend towards infinity with κ. One must then develop adapted methods to account for the limiting set occupied by the vortices in Ω. The first renormalization method consists in subtracting the interaction of each vortex on itself ($2\pi log\kappa$), that is $2\pi N \log\kappa$, to the minimal energy. For reasons seen above, because the *a priori* estimate (42) is insufficient, the precision on N to order $O(1)$ is very difficult to reach for the fundamental state for fields larger than $k_1 \log\kappa + O(\log\log\kappa)$. The second renormalization proposed by E. Sandier and S. Serfaty in [SS2] consists in dividing $\mathcal{G}_\kappa$ by h_e^2 and in studying the Γ-limit of the ratio, for at least the fundamental states. The result is then as follows.

THEOREM 3.4 ([SS2]). — *Let $h_e(\kappa)$ be a positive function so that $\lambda = \lim_{\kappa\to+\infty} \frac{\log\kappa}{h_e}$ exists. If $\lambda = 0$, we suppose $h_e(\kappa) = o(\kappa^2)$. Let $k_\star$ be a solution of the following problem*

$$\min \left\{ \begin{array}{l} E(k) = \lambda|\mu|(\Omega) + \displaystyle\int_\Omega |\nabla k|^2 + |k-1|^2 \\ \\ k \in W^{1,2}(\Omega) \quad k = 1 \quad \textit{on } \partial\Omega \\ \\ \mu = -\Delta k + k \quad \textit{ is a Radon measure} \end{array} \right\}$$

Then $k_\star$ is unique, $\mu = -\Delta h_\star + h_\star$ is positive and coincides with the characteristic function $\mathbf{1}_\omega$ of the locus of points ω where $k_\star$ is minimal and equal to $1 - \frac{\lambda}{2}$

$$\mu = \left(1 - \frac{\lambda}{2}\right) \mathbf{1}_\omega \tag{52}$$

and

$$\frac{h}{h_e} \rightharpoonup k_\star \qquad \textit{weakly in } W^{1,2}(\Omega). \tag{53}$$

Moreover, the lack of strong convergence is exactly given by the measure of the following defect

$$\lambda\mu = \lim_{\kappa\to+\infty} |\nabla(h/h_e)|^2 - |\nabla k|^2. \tag{54}$$

Finally we have the following expansion of the renormalized fundamental energy

$$\lim_{\kappa\to+\infty} \frac{\mathcal{G}_\kappa(u_\kappa, A_\kappa)}{h_e^2} = E(k_\star) \tag{55}$$

where (u_κ, A_κ) is a minimal configuration.

REMARK 2. — *When λ is equal to 0, (i.e. $h_e \gg \log\kappa$) μ is the characteristic function of all Ω and h/h_e converges strongly towards 1.*

The absence of strong convergence for the case $\lambda > 0$ can be understood in the following sense. Apart from the bad set which is covered by balls $\mathcal{M} \subset \Omega_v = \cup_{i\in I} B_{\kappa^{-1}}(x_i)$ (supposedly disjoint), $h = *dA$ verifies

$$-\mathrm{div}\left(\frac{\nabla h}{|u|^2}\right) + h = 0 \qquad \text{in } \Omega \setminus \Omega_v$$

h/h_e is then very close to the solution of

$$-\Delta k + k = 2\pi \sum_{i\in I} d_i \frac{\delta_{x_i}}{h_e}$$

where d_i is the degree of $u/|u|$ on the boundary $B_{\kappa^{-1}}(x_i)$. We recall that the second Ginzburg-Landau equation (1) gives in particular

$$-\int_{\partial B_{\kappa^{-1}}(x_i)} \frac{1}{|u|^2}\frac{\partial h}{\partial \nu} + \int_{B_{\kappa^{-1}}(x_i)} h = 2\pi d_i$$

and that, moreover, again by means of this equation,

$$\frac{\mathcal{G}_\kappa(u,A)}{h_e^2} \simeq \int_{\Omega\setminus\Omega_v} \frac{1}{|u|^2}\left|d\frac{h}{h_e}\right|^2 + \left|\frac{h}{h_e} - 1\right|^2 \simeq \int_{\Omega\setminus\Omega_v} \left|d\frac{h}{h_e}\right|^2 + \left|\frac{h}{h_e} - 1\right|^2.$$

Let μ be the limit of the vortex density per unit external field.

$$\mu = \lim_{\kappa\to+\infty} 2\pi \sum_{i\in I} d_i \frac{\delta_{x_i}}{h_e}.$$

If we consider the energy of the limit $k_\star$, solution of $-\Delta k_\star + k_\star = \mu$ instead of the limit of the energy

$$\int_{\Omega\setminus\Omega_v} |dk|^2 + |k-1|^2,$$

we forget that μ has been obtained by means of Dirac sums and we miss the part of the energy coming from the interaction of each Dirac on itself divided by h_e^2 which is

$$\frac{2\pi N \log\kappa}{h_e^2} \longrightarrow \mu(\Omega) \lim_{\kappa\to+\infty} \frac{\log\kappa}{h_e} = \lambda\mu(\Omega).$$

The difficulty of the analysis of this second renormalization is not only due to the above handwaving understanding of the mechanism of the energy partitioning, but essentially in the application of rigorous arguments which enable to establish theorem 3.4. The proof is somewhat in the same spirit as that of theorem 3.2, where, in order to obtain an optimal lower bound of the energy, we proceed by decomposing the domain which separates the self energy of the vortices from the rest of the energy. This decomposition of $\mathcal{M}$ relies on a more refined covering of the "bad set" than that of theorem 3.2, where the method of enlargement of balls introduced in [Sa] is used.

As described by the above simple reasoning, μ is the limiting density of vortices per unit applied field. It is either maximal and equal to $1-\lambda/2$ in the sub-domain ω of Ω where $k_\star = 1 - \frac{\lambda}{2}$ also, or equal to zero in its complementary (see figure 4). The problem verified by $k_\star$ is a free boundary problem in the sense that the knowledge of

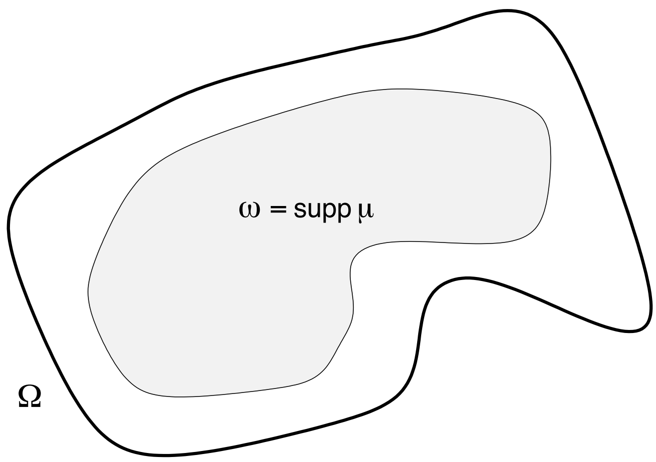


FIGURE 4. Free-boundary problem

ω determines $k_\star$ in a unique way. This is an obstacle-problem, now quite classical, considered in particular in [R]. It has been shown by A. Bonnet and R. Monneau in [BM] that $\partial\omega$ is regular for almost all the values of λ and that, whenever it is the case, ω is determined by the existence of the solution (...) of the following problem.

$$\begin{cases} -\Delta k_\star + k_\star = 0 & \text{in } \Omega \setminus \omega \\ k_\star = 1 - \dfrac{\lambda}{2} & \text{in } \omega \\ \dfrac{\partial k_\star}{\partial \nu} = 0 & \text{on } \partial\omega \\ k_\star = 1 & \text{on } \partial\Omega. \end{cases} \tag{56}$$

3.6. Conclusion

What about the Abrikosov lattices? It would be interesting, in the previous approach, to place oneself at a scale $1/\sqrt{h_e}$ -the minimal mean relative distance between two vortices- and to try then to understand the limit $\kappa \to +\infty$. We can hope in this asymptotic, to show the existence of an infinite renormalized energy on countable configurations of points of $\mathbb{R}^2$, which govern the set up of the vortices amongst themselves and try to understand the fundamental states of this energy by restraining ourselves, in a first approach, to periodic lattices. The first difficulty of this analysis is to understand the terms of lower order in the expansion of the energy (55). The first term only gives the mean density of vortices and does not see their relative positions to $O(1/\sqrt{h_e})$.

In his PhD thesis, M. Dutour [D] adopts a different approach which we shall not develop more extensively here since that would lead us beyond the scope of this talk. One places oneself on any flat torus T, which can be seen as a unit cell of a lattice of $\mathbb{R}^2$, and whose size and geometry are variables of the problem. On that cell T which is considered occupied by only one vortex (one fixes a complex line bundle E on T of Euler class equal to 1), one studies the minima of $\mathcal{G}_\kappa$ the Ginzburg-Landau functional under the action of an external uniform field h_e. When the external field increases, the increase in the vortex density, in that model, gives rise to a decrease in the size of the cell T, etc. M. Dutour thus gives a very complete description of the phase diagram (figure 1) namely in the vicinity of H_{c_2}, where he manages to show the existence of exactly two solutions; the first with vortices (close to one another by $\simeq \kappa^{-1}$) and the second in the normal state $dA = h_e$. He establishes whether each of these solutions is minimizing or not with respect to h_e, and accounts for the large precision of the Abrikosov bifurcation from the critical value 1/2. The optimality of such or such lattice is only partially discussed, but this approach seems to open a new direction towards a rigorous understanding of the relative positions of the vortices in the fundamental state (very far from the boundary which has disappeared from this model) and eventually of the energetically favorable nature of the Abrikosov triangular lattices.

REFERENCES

[Ab] A. Abrikosov – *On the magnetic properties of superconductors of the second type*, Soviet Phys. JETP 5, 1174-1182 (1957).

[Af] A. Aftalion – *On the minimizers of the Ginzburg-Landau energy for high kappa: the one-dimensional case*, European J. Appl. Math. **8**, 331-345 (1997).

[AT] A. Aftalion and C. Troy – *On the solutions of the one dimensional Ginzburg-Landau Equations*, Preprint LMENS (1998).

[AB1] L. Almeida and F. Bethuel – *Multiplicity results for the Ginzburg-Landau equation in presence of symmetries*, Houston J. Math. **23**, 733-764 (1997).

[AB2] L. Almeida and F. Bethuel – *Topological methods for the Ginzburg-Landau equations*, J. Math. Pures Appl. **77**, 1-49 (1998).

[BBC] H. Berestycki, A. Bonnet and J. Chapman – *A semi-elliptic system arising in the theory of type-II superconductivity*, Comm. Appl. Non-linear Anal. **1**, 1-21 (1994).

[BC] M. S. Berger and Y. Y. Chen – *Symmetric vortices for the Ginzburg-Landau equations of superconductivity and the nonlinear desingularization phenomenon*, Journal of Funct. Anal. **82**, 259-295 (1989).

[BBH0] F. Bethuel, H. Brezis, and F. Hélein – *Asymptotics for the minimization of a Ginzburg-Landau functional*, Calc. Var. Partial Differ. Equ. **1**, 2, 123-148 (1993).

[BBH] F. Bethuel, H. Brezis and F. Hélein – *Ginzburg-Landau vortices*, Birkhäuser (1994).

[BR1] F. Bethuel and T. Rivière – *Vortices for a variational problem related to supraconductivity*, Ann. Inst. Henri Poincaré, Anal. Non Linéaire **12**, 3, 243-303 (1995).

[BR2] F. Bethuel and T. Rivière – *Vorticité dans les modèles de Ginzburg-Landau pour la supraconductivité*, Seminaire EDP de l'École Polytechnique, exposé XVI (1994).

[Bog] B. Bogomol'nyi – Soviet J. Nucl. Phys. **24**, 449 (1976).

[BH1] C. Bolley and B. Helffer – *Rigorous results on Ginzburg-Landau models in a film submitted to an exterior parallel magnetic field I and II*, Non Linear Stud. **3**, 1-29 and 121-152 (1996).

[BH2] C. Bolley and B. Helffer – *The ginzburg-Landau equations in a semi-infinite superconducting film in the large κ limit*, European J. Appl. Math. 8, 347-367 (1997).

[BCM] A. Bonnet, J. Chapman and R. Monneau – *Convergence of Meissner minimizers of the Ginzburg-Landau energy of superconductivity as $\kappa \to +\infty$*, preprint (1999).

[BM] A. Bonnet and R. Monneau – *Existence of a smooth free-boundary in a superconductor with a Nash-Moser inverse function theorem argument*, preprint (1999).

[BBM] J. Bourgain, H. Brezis and P. Mironescu – *Lifting properties of Sobolev maps*, in preparation.

[BMR] H. Brezis, F. Merle and T. Rivière – *Quantization effects for $-\Delta u = u(1-|u|^2)$ in $\mathbb{R}^2$*, Arch. Rat. Mech. Analysis. **126**, 123-145 (1994).

[DL] Q. Du and H. Lin – *Ginzburg-Landau vortices: dynamics, pinning and hysteresis*, Siam J. Math. Anal. **28**, 1265-1293 (1997).

[D] M. Dutour – *Bifurcation vers l'état d'Abrikosov et diagramme de phase*, thèse de l'Université Paris-Sud, Orsay (1999).

[dG] G. De Gennes – *Superconductivity of Metal and Alloys*, Benjamin, New-York and Amsterdam (1966).

[GS] S. Gueron and I. Shafrir – *On a discrete variational problem involving interacting particles*, to appear in SIAM J. Appl. Math. (1999).

[JT] A. Jaffe and C. Taubes – *Vortices and Monopoles*, Birkhäuser (1980).

[J] R. Jerrard – *Lower bounds for generalized Ginzburg-Landau functionals*, SIAM J. Math. Anal. **30**, 721-746 (1999).

[LL] H. LIN and C. LIN – *Minimax solutions of the Ginzburg-Landau equations,* Selecta Mathematica, New Series **3**, 99-113 (1997).

[LR] H. LIN and T. RIVIÈRE – *Complex Ginzburg-Landau equations and codimension 2 minimal surfaces*, J.E.M.S. **1**, 3 (1999).

[LR2] H. LIN and T. RIVIÈRE – *A quantization property for static Ginzburg-Landau vortices*, Comm. Pure and App. Math. (2000).

[Mi] P. MIRONESCU – *Les minimiseurs locaux pour l'équation de Ginzburg-Landau sont à symétrie radiale*, C.R. Acad. Sci. Paris, Ser. 1, **323**, 593-598 (1996).

[OS] N. OVCHINNIKOV and M. SIGAL – *The Ginzburg-Landau equation I. Static vortices*, Partial differential equations and their applications (Toronto, ON, 1995), 199-220, CRM Proc. Lecture Notes **12**, Amer. Math. Soc., Providence, RI (1997).

[PR] F. PACARD and T. RIVIÈRE – *Linear and non-linear aspects of vortices*, Birkhäuser (2000).

[Qin] J. QING – *Renormalized energy for Ginzburg-Landau vortices on closed surfaces*, Math. Z. **225**, 1-34 (1997).

[Ri1] T. RIVIÈRE – *Asymptotic analysis for the Ginzburg-Landau equations,* ETH-Zürich course, January 1997, Boll U.M.I. (1999).

[Ri2] T. RIVIÈRE – *Line vortices in the $U(1)-$Higgs Model*, COCV **1**, 77-167 (1995).

[Ri3] T. RIVIÈRE – *Some progress towards Jaffe and Taubes Conjectures*, preprint (1999).

[R] F. RODRIGUES – *Obstacle problems in mathematical physics*, Mathematical Studies, North Holland (1987).

[SST] D. SAINT-JAMES, G. SARMA and J. THOMAS – *Type-II Superconductivity*, Pergamon Press (1969).

[Sa] E. SANDIER – *Lower bounds for the energy of unit vector fields and applications*, J. Funct. Anal. **152**, 379-403 (1998).

[SS1] E. SANDIER and S. SERFATY – *Global minimizers for the Ginzburg-Landau functional below the first critical magnetic field*, to appear in Annales de l'Inst. Henri Poincaré, Analyse non-linéaire (1999).

[SS2] E. SANDIER and S. SERFATY – *A rigourous derivation of a free-boundary problem arising in superconductivity*, Ann. Sci. Éc. Norm. Sup. **33** (2000), 561-592.

[Se1] S. SERFATY – *Étude mathématique de l'équation de Ginzburg-Landau de la supraconductivité*, Thèse de l'Université Paris-Sud, Orsay (1999).

[Se2] S. SERFATY – *Local minimizers for the Ginzburg-Landau energy near critical magnetic field I and II*, Comm. Contemporary Mathematics **1**, n° 2 et 3, 213-254 (1999).

[Se3] S. Serfaty – *Stable configurations in superconductivity: uniqueness, multiplicity and vortex-nucleation*, to appear in : Arch. Rat. Mech. Anal. (1999).

[St] M. Struwe – *On the asymptotic behavior of minimizers of the Ginzburg-Landau model in* 2 *dimensions,* J. Diff. Int. Equations **7** (1994), 1613-1624 and Erratum in J. Diff. Int.

[Ti] M. Tinkham – *Introduction to Superconductivity*, 2nd edition, McGraw-Hill (1996).

[Uh] K. Uhlenbeck – *Connections with L^p bounds on curvature*, Comm. Math. Phys. **83**, 31-42 (1982).

Tristan RIVIÈRE

Centre de Mathématiques
École polytechnique
F-91128 Palaiseau Cedex
E-mail : riviere@math.polytechnique.fr

Séminaire BOURBAKI Novembre 1999
52e année, 1999-2000, n° 869, p. 105 à 135

PROGRÈS RÉCENTS SUR LA CONJECTURE DE BAUM–CONNES. CONTRIBUTION DE VINCENT LAFFORGUE

par **Georges SKANDALIS**

Cet exposé fait de nouveau le point sur la conjecture de Baum-Connes après celui de Pierre Julg ([25]) qui rendait compte d'une avancée spectaculaire obtenue par Higson et Kasparov ([19]). Une nouvelle barrière vient juste d'être franchie par V. Lafforgue : celle de la propriété T de Kazhdan. Le but ici est d'exposer les diverses étapes du travail de Lafforgue. En fin d'exposé, nous donnerons quelques autres résultats en relation avec la conjecture de Baum-Connes et celle de Novikov obtenus récemment par G. Yu ([51]), N. Higson ([18]) et quelques autres.

La conjecture de Baum-Connes prédit la K-théorie de la C^*-algèbre (réduite) d'un groupe discret G et, plus généralement, d'un groupe localement compact, et même d'un groupoïde localement compact (par exemple de la C^*-algèbre associée à un feuilletage) *etc.* Cette conjecture a plusieurs conséquences : l'injectivité du morphisme de Baum-Connes a des implications géométriques importantes — la plus fameuse étant la conjecture de Novikov sur les « hautes signatures » ; sa surjectivité a des applications fondamentales dans la théorie des C^*-algèbres (*e.g.* une conjecture de Kadison, déjà ancienne, sur les idempotents).

Malgré les efforts de plusieurs mathématiciens, une barrière subsistait pour la preuve de cette conjecture : la propriété T de Kazhdan (*cf.* [32], [12]), qui empêche certaines idées de démonstration d'aboutir et rend fausses certaines généralisations : en effet, cette conjecture est alors fausse si la C^*-algèbre pleine remplace la réduite (*cf.* [24], voir aussi [10]) ; la généralisation de cette conjecture au bifoncteur de Kasparov (qui généralise la K-théorie) peut elle-même être fausse (*cf.* [44]). Au-delà de cette barrière, les seuls cas connus avant le travail de Lafforgue étaient les groupes de Lie réels réductifs (A. Wassermann [50]) ou p-adiques (Baum-Higson-Plymen [6]) utilisant la description complète de la C^*-algèbre de ces groupes, mais aucun groupe discret infini possédant la propriété T n'avait pu être obtenu.

Lafforgue établit la conjecture de Baum-Connes pour beaucoup de groupes localement compacts ayant la propriété T : tout groupe de Lie semi-simple réel ou réductif p-adique ET les sous-groupes discrets cocompacts des groupes de Lie de rang réel 1 ou de SL_3 d'un corps local.

Dans tous ces cas, l'injectivité du morphisme de Baum-Connes était connue (travaux de Miščenko, Solovjev, Kasparov...). En fait, il y a un élément noté γ par Kasparov dans l'anneau de représentations (défini par Kasparov) $KK_G(\mathbf{C},\mathbf{C})$ dont la différence avec 1, mesure la non surjectivité du morphisme de Baum-Connes. On doit montrer que $\gamma = 1$ en tant qu'endomorphisme de la K-théorie de la C^*-algèbre réduite $C_r^*(G)$ de G. Remarquons que, par la propriété T, γ lui-même n'est pas égal à 1 dans l'anneau $KK_G(\mathbf{C},\mathbf{C})$ et même dans l'anneau $KK(C_r^*(G), C_r^*(G))$ (du moins si G est un sous-groupe discret cocompact de $Sp(n,1)$ — *cf.* [44]).

La démonstration de Lafforgue a trois étapes. La première est sa principale contribution conceptuelle ; la seconde est — de loin — la plus difficile techniquement ; dans la troisième, il utilise des travaux de Jolissaint dans le cas $Sp(n,1)$ et de Ramagge-Robertson-Steger dans le cas de SL_3 d'un corps local non archimédien ; enfin, il adapte la méthode de Ramagge-Robertson-Steger pour obtenir aussi le cas des sous-groupes discrets cocompacts de $SL_3(\mathbf{R})$ et $SL_3(\mathbf{C})$.

Plus explicitement :

1. Dans la première étape, Lafforgue construit un bifoncteur $KK^{\mathrm{ban}}(A,B)$ pour des algèbres de Banach, en remplacement de la théorie bivariante de Kasparov. Il s'agit là d'une construction nouvelle et très intéressante. Elle suit d'assez près celle de Kasparov. Cependant, les C^*-modules hilbertiens qui sont à la base du travail de Kasparov sont ici remplacés par des paires de modules banachiques en dualité (les modules hilbertiens des C^*-algèbres sont en ce sens autoduaux). Malheureusement, le produit de Kasparov ne se généralise pas directement à ce cadre. Cependant, Lafforgue parvient à étendre plusieurs constructions de Kasparov dans cette « KK-théorie banachique » : il démontre qu'elle opère sur la K- théorie, construit une KK-théorie banachique équivariante par un groupe localement compact, ainsi qu'un analogue de l'homomorphisme de descente $j_G : KK_G(A,B) \to KK(A \rtimes G, B \rtimes G)$ de Kasparov.

2. Dans la deuxième étape il établit l'égalité $\gamma = 1$ dans sa KK-théorie banachique équivariante pour tout groupe localement compact agissant de façon continue, propre et isométrique sur :

– une variété riemannienne complète, connexe, simplement connexe à courbure sectionnelle négative ou nulle (cette classe contient tous les groupes de Lie réels et tous leurs sous-groupes fermés) ;

– sur un immeuble affine (cette classe contient tous les groupes de Lie p-adiques et tous leurs sous-groupes fermés).

C'est la partie la plus technique. Il construit une homotopie qui utilise :

– dans le cas réel des espaces de Sobolev de formes L^p sur des variétés de courbure sectionnelle négative ;

– dans le cas p-adique une combinatoire assez élaborée (ainsi que des espaces ℓ^p).

À l'aide des parties 1 et 2, Lafforgue établit la conjecture de Baum-Connes à valeurs dans la sous-algèbre $L^1(G)$ de $C_r^*(G)$ (ce qui prouve une conjecture de Bost), ainsi que toutes les complétions de $C_c(G)$ pour des « normes inconditionnelles » (dans le langage de Lafforgue) *i.e.* une norme d'algèbre N satisfaisant $N(f) \leq N(g)$ si $f, g \in C_c(G)$ satisfont, pour tout $x \in G$, $|f(x)| \leq |g(x)|$ (notons cependant que la norme de C^*_r n'est malheureusement pas « inconditionnelle »).

3. Pour prouver la conjecture de Baum-Connes, il suffit de construire une « complétion inconditionnelle » de $C_c(G)$, qui soit une sous-algèbre A de $C^*_r(G)$ avec la même K-théorie que $C^*_r(G)$. Pour cela, il suffit que A soit stable par calcul fonctionnel holomorphe dans $C^*_r(G)$. Une telle sous-algèbre A existe :

– si G est un groupe de Lie semi-simple, Lafforgue montre qu'une variante de l'algèbre de Harish-Chandra convient (*cf.* [36]) ;

– si G est un groupe hyperbolique *e.g.* un sous-groupe discret cocompact dans $Sp(n, 1)$ (travaux de Haagerup pour les groupes libres, de Jolissaint pour les sous-groupes cocompacts des groupes de Lie de rang 1, et de de la Harpe pour un groupe hyperbolique quelconque [17]) ;

– si G est discret cocompact dans SL_3 d'un corps local non archimédien (ce résultat est dû à Ramagge, Robertson et Steger [41]) ;

– si G est discret cocompact dans $SL_3(\mathbf{R})$ ou $SL_3(\mathbf{C})$: ce résultat est dû à Lafforgue [34].

Dans ces trois derniers cas, l'algèbre A est donnée par des conditions de décroissance rapide au sens ℓ^2.

Nous commencerons cet exposé par quelques rappels, notamment sur les C^*-algèbres de groupes et la théorie de Kasparov. Ensuite, nous rappellerons la conjecture de Baum-Connes et ses généralisations. Nous renvoyons à ce sujet à l'excellent exposé de Pierre Julg [25]. Les trois paragraphes suivants seront consacrés aux trois étapes du travail de Lafforgue citées ci-dessus. Dans le dernier paragraphe, nous décrivons de très récents résultats de Yu ([51]) et Higson ([18]) reliés à la conjecture de Baum-Connes.

Ajout (mars 2000). — Depuis que cet exposé a été écrit, des contre-exemples à la conjecture de Baum-Connes pour les feuilletages, ainsi que ses analogues pour les « structures à l'infini » ont été donnés (*cf.* [20]). Ceux-ci sont basés sur des idées de M. Gromov (*cf.* [15]).

Je tiens à remercier Vincent Lafforgue pour de nombreuses explications sur ses travaux. Je le remercie aussi, ainsi que Saad Baaj, Jean Renault et Alain Valette, pour une lecture attentive de ce manuscrit. Un grand merci enfin à Alberto Arabia pour son aide amicale.

1. RAPPELS ET NOTATIONS : C*-ALGÈBRES DE GROUPES, THÉORIE DE KASPAROV

1.1. Algèbres de Banach ; C*-algèbres

Toutes les algèbres de Banach que nous considérons dans cet exposé sont complexes. Si A est une algèbre de Banach, on supposera toujours que sa norme vérifie $\|ab\| \leq \|a\|\|b\|$, pour tout $a, b \in A$. On ne suppose pas en général que nos algèbres de Banach possèdent un élément unité. Si A est une algèbre de Banach, on note $\widetilde{A}$ l'algèbre de Banach unifère construite de la façon suivante : en tant qu'espace vectoriel $\widetilde{A}$ est isomorphe au produit $A \times \mathbf{C}$; sa norme est donnée par $\|(a, \lambda)\| = \|a\| + |\lambda|$ (pour $a \in A$ et $\lambda \in \mathbf{C}$) et son produit est donné par $(a, \lambda)(b, \mu) = (ab + \lambda b + \mu a, \lambda\mu)$ (pour $a, b \in A$ et $\lambda, \mu \in \mathbf{C}$). On identifie A avec son image dans $\widetilde{A}$ par l'homomorphisme $a \mapsto (a, 0)$; l'élément $(0, 1)$ est l'unité de $\widetilde{A}$.

Un A-module banachique à droite (*resp.* à gauche) est un espace de Banach E muni d'une action à droite (*resp.* à gauche) de A telle que, pour tout $x \in E$ et tout $a \in A$, on ait $\|xa\| \leq \|x\|\|a\|$ (*resp.* $\|ax\| \leq \|a\|\|x\|$). Un A-module banachique à droite (*resp.* à gauche) est un $\widetilde{A}$-module banachique à droite (*resp.* à gauche) : il suffit de poser $x(a, \lambda) = xa + \lambda x$ (*resp.* $(a, \lambda)x = ax + \lambda x$) pour $a \in A,\ x \in E$, et $\lambda \in \mathbf{C}$.

Rappelons qu'une *inolution* sur une algèbre de Banach est une application antilinéaire, antimultiplicative, involutive et isométrique notée $x \mapsto x^*$. Une *algèbre de Banach involutive* est une algèbre de Banach munie d'une involution.

Une C^*-algèbre est une algèbre de Banach involutive A telle que, pour tout $x \in A$, on ait $\|x^*x\| = \|x\|^2$. Si A est une C^*-algèbre, l'algèbre de Banach obtenue à partir de A en adjoignant une unité est involutive (on pose $(a, \lambda)^* = (a^*, \overline{\lambda})$) ; ce n'est cependant pas une C^*-algèbre. Il existe néanmoins une (unique) norme de C^*-algèbre sur $\widetilde{A}$.

1.2. C^*-algèbres de groupes

Soit G un groupe localement compact. Notons dx une mesure de Haar à gauche sur G et $C_c(G)$ l'algèbre de convolution des fonctions continues à support compact sur G. Rappelons que la convolée $f \star g$ de $f, g \in C_c(G)$ est donnée par la formule

$$f \star g(x) = \int_G f(y) g(y^{-1}x)\, dy$$

pour tout $x \in G$.

Par ailleurs, on définit une involution $f \mapsto f^*$ sur $C_c(G)$ en posant $f^*(x) = \Delta(x^{-1})\overline{f(x^{-1})}$ où Δ désigne la fonction module sur G (telle que, pour tout $f \in C_c(G)$, on ait $\int f(x^{-1})\, dx = \int f(x)\Delta(x)\, dx$).

Il y a plusieurs façons de compléter $C_c(G)$ en une algèbre de Banach, c'est-à-dire des normes sur $C_c(G)$ telles que, pour tout $f, g \in C_c(G)$ on ait $\|f \star g\| \leq \|f\|\|g\|$. Nous en utiliserons (au moins) 3 :

1. La plus simple est de prendre la norme $\| \ \|_1$. On obtient l'algèbre de Banach $L^1(G)$.

2. Une autre façon est de considérer $C_c(G)$ comme opérant sur $L^2(G)$ par convolution : pour $f \in C_c(G)$, on note $\lambda(f)$ l'application de $L^2(G)$ dans lui-même donnée par $g \mapsto \lambda(f)(g) = f \star g$. On note $\|f\|_r$ la norme d'opérateur de $\lambda(f)$ (dans $L^2(G)$). Le complété de $C_c(G)$ pour cette norme est une C^*-algèbre appelée la *C^*-algèbre réduite de G* et notée $C^*_r(G)$. De façon équivalente, $C^*_r(G)$ s'identifie avec l'adhérence normique de $\lambda(C_c(G))$ dans l'algèbre des opérateurs de l'espace hilbertien $L^2(G)$. On peut remarquer que :

– pour $f \in C_c(G)$, on a $\|\lambda(f)\| \leq \|f\|_1$, donc la représentation λ s'étend à $L^1(G)$, et que

– la représentation λ est *involutive* : pour tout $f \in C_c(G)$, on a $\lambda(f^*) = \lambda(f)^*$ où, pour un opérateur T d'un espace hilbertien, T^* désigne l'adjoint de T.

3. Considérons toutes les représentations involutives σ de $C_c(G)$ dans un espace hilbertien H_σ telles que, pour tout $f \in C_c(G)$, on ait $\|\sigma(f)\| \leq \|f\|_1$. Posons $\|f\|_p = \sup_\sigma \|\sigma(f)\|$. On obtient encore une norme ; le complété de $C_c(G)$ pour cette norme est une C^*-algèbre notée $C^*(G)$ et appelée la C^*-algèbre de G ou la C^*-algèbre pleine de G. L'identité de $C_c(G)$ se prolonge évidemment en un morphisme $C^*(G) \to C^*_r(G)$. Ce morphisme est toujours surjectif. Il est injectif si et seulement si le groupe G est *moyennable.*

1.3. Produits croisés

Soient G un groupe localement compact et A une algèbre de Banach. Une *action de G dans A* est un homomorphisme α de G dans le groupe des automorphismes isométriques de A, tel que, pour tout $a \in A$, l'application $x \mapsto (\alpha(x))(a)$ soit continue. Le plus souvent, pour $x \in G$ et $a \in A$ l'élément $(\alpha(x))(a) \in A$ s'écrira juste $x.a$.

Soit α une action d'un groupe G dans une algèbre de Banach A. L'espace de Banach $L^1(G; A)$ est muni du produit de convolution donné par la formule

$$f \star g(x) = \int_G f(y)\, y.(g(y^{-1}x))\, dy$$

pour tout $f, g \in C_c(G; A) \subset L^1(G; A)$ et $x \in G$.

Si A est une C^*-algèbre et α respecte l'involution ($x.(a^*) = (x.a)^*$ pour tout $a \in A$ et tout $x \in G$), alors l'algèbre de Banach $L^1(G; A)$ est naturellement munie d'une involution par la formule $(f^*)(x) = \Delta(x^{-1})x.(f(x^{-1}))^*$. Généralisant la construction de $C^*_r(G)$ et $C^*(G)$, on obtient par complétion de $L^1(G; A)$ deux C^*-algèbres $A \rtimes_r G$ et $A \rtimes G$ appelées respectivement produit croisé réduit et plein de A par G.

1.4. Théorie de Kasparov

Nous rappelons ici, très sommairement, certaines constructions et propriétés du bifoncteur de Kasparov (*cf.* [28], [29] ; voir aussi les exposés de T. Fack [13] et P. Julg [25]).

1.4.a. *Modules hilbertiens*

Les A-*modules hilbertiens* sont les espaces hilbertiens avec les scalaires remplacés par A.

Rappelons qu'un élément a d'une C^*-algèbre A est appelé *positif* s'il est de la forme x^*x ($x \in A$). L'ensemble A_+ des éléments positifs joue le rôle de $\mathbf{R}_+$ dans $\mathbf{C}$.

On appelle A-*module préhilbertien* sur A un A-module à droite E muni d'un « produit scalaire » *i.e.* d'une application sesquilinéaire $E \times E \ni (x, y) \mapsto \langle x, y\rangle \in A$, telle que, pour tout $x \in E$, l'application $y \mapsto \langle x, y\rangle$ est A-linéaire (à droite), satisfaisant $\langle x, y\rangle^* = \langle y, x\rangle$ et $\langle x, x\rangle \in A_+$ pour tout $x \in E$.

Si E est un A-module préhilbertien, l'application $x \mapsto \sqrt{\|\langle x, x\rangle\|}$ est une semi-norme sur E. Si E est séparé et complet pour cette norme on dit que c'est un A-*module hilbertien.*

1.4.b. *Opérateurs ; opérateurs compacts*

Soient E, F des A-modules hilbertiens. L'ensemble des opérateurs de E dans F que nous considérons sont les opérateurs $T : E \to F$ possédant un adjoint, *i.e.* une application $T^* : F \to E$ satisfaisant $\langle T(x), y\rangle = \langle x, T^*(y)\rangle$ pour tout $x \in E$ et $y \in F$. Un tel opérateur est automatiquement linéaire, A- linéaire et continu. L'ensemble des opérateurs de E dans F est noté $\mathcal{L}(E, F)$. L'ensemble $\mathcal{L}(E) = \mathcal{L}(E, E)$, muni de la norme des applications linéaires continues, est une C^*-algèbre.

En particulier, pour tout $x \in E$ et $y \in F$, posons $\theta_{y,x} : z \mapsto y\langle x, z\rangle$. C'est un élément de $\mathcal{L}(E, F)$ et $\theta_{y,x}^* = \theta_{x,y}$. On note $\mathcal{K}(E, F)$ le sous-espace vectoriel fermé de $\mathcal{L}(E, F)$ engendré par les $\theta_{y,x}$. L'ensemble $\mathcal{K}(E) = \mathcal{K}(E, E)$ est un idéal bilatère fermé dans $\mathcal{L}(E)$.

Il sera utile de considérer les modules hilbertiens $\mathbf{Z}/2\mathbf{Z}$-gradués, c'est-à-dire les modules hilbertiens $E = E^{(0)} \oplus E^{(1)}$, où $E^{(0)}$ et $E^{(1)}$ sont orthogonaux pour $\langle\ ,\ \rangle$. Pour $i \in \mathbf{Z}/2\mathbf{Z}$, on note $\mathcal{L}(E)^{(i)}$ l'ensemble des $T \in \mathcal{L}(E)$ tels que $T(x) \in E^{(i+j)}$ pour tout $j \in \mathbf{Z}/2\mathbf{Z}$ et tout $x \in E^{(j)}$.

1.4.c. *Cycles en KK-théorie*

Soient A et B des C^*-algèbres. Le groupe de Kasparov $KK(A, B)$ est donné par des cycles et une relation d'équivalence sur ces cycles. Un cycle est un triplet (E, π, F) où E est un B-module hilbertien $\mathbf{Z}/2\mathbf{Z}$-gradué, $\pi : A \to \mathcal{L}(E)^{(0)}$ est un homomorphisme involutif (tel que $\pi(a^*) = \pi(a)^*$) et $F \in \mathcal{L}(E)^{(1)}$ est tel que $F = F^*$ et, pour tout $a \in A$, les opérateurs $F\pi(a) - \pi(a)F$ et $(F^2 - 1)\pi(a)$ sont dans $\mathcal{K}(E)$.

La somme des cycles (E, π, F) et (E', π', F') est, avec des notations évidentes, $(E \oplus E', \pi \oplus \pi', F \oplus F')$.

1.4.d. *Homotopie*

Notons $B[0,1]$ la C^*-algèbre des applications continues de l'intervalle $[0,1]$ dans B. Un cycle pour $KK(A, B[0,1])$ est une famille (E_t, π_t, F_t) de cycles pour $KK(A,B)$: c'est par définition une homotopie entre (E_0, π_0, F_0) et (E_1, π_1, F_1).

Le groupe de Kasparov $KK(A,B)$ est l'ensemble des classes d'homotopie de cycles. Ce groupe est fonctoriel en A et en B : pour tout A on a un foncteur covariant $B \to KK(A,B)$ et pour tout B on a un foncteur contravariant $A \to KK(A,B)$.

1.4.e. *Produit de Kasparov*

Kasparov a défini un produit $KK(A,B) \times KK(B,C) \to KK(A,C)$ (où A, B, C sont des C^*-algèbres) qui est bilinéaire, naturel et associatif. Le fait que ce produit soit hautement non trivial donne toute sa force à la théorie. Muni de ce produit, $KK(A,A)$ est un anneau. Son élément unité noté 1_A est représenté par $(A, \pi, 0)$ où A est gradué par $A^{(0)} = A$ et $\pi(a)b = ab$.

1.4.f. *Action sur la* K*-théorie*

Le groupe $KK(\mathbf{C}, A)$ a été identifié par Kasparov avec le groupe de K-théorie $K_0(A)$. Le produit de Kasparov donne en particulier un morphisme de groupes $KK(A,B) \to \mathrm{Hom}(K_0(A), K_0(B))$.

1.4.g. *Le cas équivariant*

On suppose à présent qu'un groupe localement compact G opère sur les C^*-algèbres A et B. On définit un groupe $KK_G(A,B)$ en demandant que G opère aussi sur les cycles : un cycle pour $KK_G(A,B)$ est encore donné par un triple (E, π, F), qui est un cycle pour $KK(A,B)$, tel que G opère dans E de façon continue (*i.e.* pour tout $x \in E$ l'application $g \mapsto g.x$ est continue) ; on suppose de plus que, pour tout $x, y \in E$, $a \in A$, $b \in B$ et $g \in G$, on a

$$g.(xb) = (g.x)(g.b)\,,\ \langle g.x, g.y\rangle = g.(\langle x, y\rangle)\,,\ g.(\pi(a)(x)) = \pi(g.a)(g.x)\,;$$

enfin, pour tout $a \in A$ et tout $g \in G$ on a $\pi(a)(g.F - F) \in \mathcal{K}(E)$ et $g \mapsto \pi(a)(g.F)$ est continue en norme — où $g.F$ est donné par $(g.F)(x) = g.(F(g^{-1}.x))$. Le groupe $KK_G(A,B)$ est l'ensemble des classes d'homotopie de cycles. Toutes les propriétés se généralisent au cas équivariant ; en particulier, le produit de Kasparov se généralise à ce cadre.

Remarque. — Soient G un groupe discret et (H, π, F) un cycle pour $KK_G(A, \mathbf{C})$. L'action de G dans H et la représentation π de A donnent une représentation du produit croisé $A \rtimes G$ dans H. On en déduit que $KK_G(A, \mathbf{C}) = KK(A \rtimes G, \mathbf{C})$. Plus généralement si l'action de G dans B est triviale $KK_G(A,B) = KK(A \rtimes G, B)$.

1.4.h. ***Le morphisme j_G***

Si (E, π, F) est un cycle pour $KK_G(A, B)$, on peut former le produit croisé $E \rtimes_r G$ qui est un $B \rtimes_r G$-module hilbertien. On obtient aussi une représentation naturelle $\widetilde{\pi} : A \rtimes_r G \to \mathcal{L}(E \rtimes_r G)$; de plus, on associe à F un opérateur $\widetilde{F} \in \mathcal{L}(E \rtimes_r G)$. Le triplet $(E \rtimes_r G, \widetilde{\pi}, \widetilde{F})$ est un cycle pour $KK(A \rtimes_r G, B \rtimes_r G)$. On obtient ainsi un homomorphisme $j_G^r : KK_G(A, B) \to KK(A \rtimes_r G, B \rtimes_r G)$, compatible avec les produits de Kasparov.

On construit de même un homomorphisme pour les produits croisés pleins $j_G : KK_G(A, B) \to KK(A \rtimes G, B \rtimes G)$.

2. LA CONJECTURE DE BAUM-CONNES

La conjecture de Baum-Connes (*cf.* [5]) prédit la K-théorie de $C_r^*(G)$.

Si G est un groupe de Lie connexe semisimple, elle prédit que l'induction de Dirac est un isomorphisme de la K-théorie de $C^*(K)$, *i.e.* l'anneau des représentations de K (1) — où K est un sous-groupe compact maximal de G, sur celle de $C_r^*(G)$.

Si G est un groupe discret sans torsion, la conjecture prédit que la K-théorie de $C_r^*(G)$ est décrite par les G-indices sur des revêtements $\widetilde{M}$ de groupe G d'une variété compacte M (au sens de [3]).

Pour l'énoncer correctement dans le cas général, nous aurons besoin de revenir sur la notion d'espace classifiant pour un groupe.

2.1. Espaces classifiants

Soit G un groupe discret. L'espace classifiant BG est un espace muni d'un fibré principal ξ_0 de fibre G. De plus, pour tout espace topologique X (2) et tout fibré principal ξ de fibre G, il existe une application continue $f : X \to BG$, unique à homotopie près, telle que ξ soit isomorphe à $f^*(\xi_0)$. Ces propriétés déterminent BG à équivalence d'homotopie près.

Le groupe G opère sur l'espace total du fibré ξ_0 — traditionnellement noté EG et BG est le quotient EG/G par l'action de G. L'espace EG et l'espace BG sont en général construits comme complexes cellulaires infinis. On peut cependant les supposer localement compacts (à l'aide d'une construction de télescope infini).

L'action de G sur EG est *libre* et *propre*(3). De plus, supposons que G opère librement et proprement sur un espace localement compact $\widetilde{X}$. En considérant le fibré principal sur $\widetilde{X}/G$, on déduit qu'il existe une application continue G-équivariante

(1) à des problèmes mineurs de « K-orientation » près.

(2) On suppose X suffisamment régulier *e.g.* localement compact σ-compact, ou un complexe cellulaire.

(3) Autrement dit, l'application $(x, g) \mapsto (x, gx)$ est injective et propre de $X \times G$ dans $X \times X$.

$g : \widetilde{X} \to EG$, unique à homotopie G-équivariante près. En ce sens, EG est un G-espace libre et propre final dans cette catégorie.

La conjecture de Baum-Connes utilise un G-espace propre final dans la catégorie des G-espaces propres. Un tel espace existe et est unique à G-homotopie près (*cf.* [5], [31]). On le note $\underline{E}G$. La construction de cet espace est en fait plus simple que celle de EG :

– un modèle pour $\underline{E}G$ est l'ensemble des mesures de probabilité sur G ;

– un modèle localement compact pour $\underline{E}G$ est l'ensemble des mesures positives sur G de masse totale dans $]\frac{1}{2}, 1]$, muni de la topologie de dualité avec l'algèbre $C_0(G)$ des fonctions continues nulles à l'infini sur G ;

– si G est compact, on peut prendre pour $\underline{E}G$ un espace à un point !

– si G est un groupe de Lie réel connexe, on peut prendre $\underline{E}G = G/K$ où K est un sous-groupe compact maximal ;

– si G est un groupe de Lie réductif p-adique, on prendra pour $\underline{E}G$ la réalisation géométrique de l'immeuble affine associé ;

– si G est discret, une présentation de $\underline{E}G$ *via* un complexe simplicial localement fini rend possible l'utilisation d'arguments de type Mayer-Vietoris.

2.2. Le groupe de K-théorie topologique de Baum-Connes

Le groupe de K-théorie topologique $K_{0,top}(G)$, défini d'abord dans certains cas particuliers (groupes discrets, groupes de Lie, feuilletages... - *cf.* [4]), n'a été bien compris en général que récemment (*cf.* [5], [47]) :

Soit G un groupe localement compact. Le groupe de K-théorie topologique est

$$K_{0,top}(G) = \varinjlim KK_G(C_0(Y), \mathbf{C}),$$

où la limite inductive est prise sur les parties fermées G-invariantes Y de $\underline{E}G$ telles que Y/G soit compact(4).

Remarque. — Supposons que le groupe G soit discret. Dans ce cas, $K_{0,top}(G) = \varinjlim KK(C_0(Y) \rtimes G, \mathbf{C})$ (voir remarque en 1.4.g).

Si de plus G est sans torsion, toute action propre de G est libre. Alors, la C^*-algèbre $C_0(Y) \rtimes G$ est équivalente au sens de Morita à $C(Y/G)$ (en fait si G est infini, $C_0(Y) \rtimes G$ est isomorphe à $C(Y/G) \otimes \mathcal{K}$, où $\mathcal{K}$ désigne l'algèbre des opérateurs compacts sur un espace hilbertien). Comme la KK-théorie est invariante par équivalence de Morita, on en déduit que $K_{0,top}(G) = \varinjlim KK(C(X), \mathbf{C})$ la limite inductive étant prise sur les parties compactes X de BG ; en d'autres termes, $K_{0,top}(G)$ est $K_{0,c}(BG)$, la K- homologie à support compact de BG.

(4) À une action continue à gauche de G sur un espace localement compact Y, correspond une action de G sur la C^*-algèbre $C_0(Y)$ par la formule $(g.f)(y) = f(g^{-1}y)$, pour $g \in G,\ f \in C_0(Y),\ y \in Y$.

2.3. L'homomorphisme de Baum-Connes

Soit Y un espace topologique localement compact muni d'une action propre de G ; on suppose de plus que Y/G est compact.

Si l'action de G est de plus libre, l'équivalence de Morita entre $C_0(Y) \rtimes G$ et $C(Y/G)$ est donnée par un $C_0(Y) \rtimes G$-module hilbertien canonique E, tel que $\mathcal{K}(E)$ soit isomorphe à $C(Y/G)$ et qui est plein, c'est-à-dire que le sous-espace vectoriel fermé de $C_0(Y) \rtimes G$ engendré par les produits scalaires $\langle x, y \rangle$, $x, y \in E$ est $C_0(Y) \rtimes G$.

Si l'action est propre sans être libre, on fait exactement la même construction. La seule chose qui change est que le $C_0(Y) \rtimes G$-module hilbertien canonique E n'est plus plein. Cependant, $\mathcal{K}(E)$ est encore égal à $C(Y/G)$; en particulier, comme $\mathcal{K}(E)$ est unifère, $(E, \pi, 0)$ définit un élément Λ_Y de $KK(\mathbf{C}, C_0(Y) \rtimes G)$.

Une façon équivalente plus élémentaire de définir Λ_Y est la suivante : comme l'action de G dans Y est propre à quotient compact, il existe une fonction continue, positive, à support compact $c : Y \to \mathbf{R}$ telle que, pour tout $y \in Y$, $\int_G c(gy)\, dg = 1$. On note $p \in C_c(Y \times G) \subset C_c(G; C_0(Y)) \subset C_0(Y) \rtimes G$ la fonction $(y, g) \mapsto \Delta(g)^{-1/2} c(y)^{1/2} c(g^{-1}y)^{1/2}$. C'est un projecteur dans $C_0(Y) \rtimes G$ (*i.e.* $p = p^2 = p^*$) ; il définit donc un élément de $K_0(C_0(Y) \rtimes G) = KK(\mathbf{C}, C_0(Y) \rtimes G)$.

La classe de p ne dépend pas de la fonction c. En fait, le module $p(C_0(Y) \rtimes G)$ est isomorphe à E, ce qui fait le lien avec l'autre définition.

Soit maintenant $x \in K_{0,top}(G)$; il est représenté par une partie $Y \subset \underline{E}G$ fermée G-invariante telle que Y/G soit compact et un $y \in KK_G(C_0(Y), \mathbf{C})$. L'élément $\mu_G(x)$ est le produit de Kasparov de $\Lambda_Y \in KK(\mathbf{C}, C_0(Y) \rtimes G)$ et de $j_G(y) \in KK(C_0(Y) \rtimes G, C^*(G))$.

On vérifie que cette construction passe à la limite inductive et qu'on obtient ainsi un homomorphisme $\mu_G : K_{0,top}(G) \to K_0(C^*(G))$. L'homomorphisme de Baum-Connes, ou homomorphisme d'assemblage en K-théorie est l'homomorphisme $\mu_{G,r} : K_{0,top}(G) \to K_0(C_r^*(G))$ obtenu par composition de μ_G avec l'homomorphisme de K-théorie associé à l'homorphisme de C^*-algèbres $\lambda : C^*(G) \to C_r^*(G)$.

Conjecture de Baum-Connes. — *L'homomorphisme $\mu_{G,r}$ est un isomorphisme.*

2.4. Lien avec la conjecture de Novikov

Soit M une variété lisse compacte orientée. Notons Γ son groupe fondamental, $\widetilde{M}$ son revêtement universel. L'opérateur de signature de M peut être considéré comme un opérateur Γ-invariant sur $\widetilde{M}$; il définit ainsi un élément de K-homologie équivariante $KK_\Gamma(C_0(\widetilde{M}), \mathbf{C})$, donc de $K_{0,top}(\Gamma)$ (comme l'action de Γ sur $\widetilde{M}$ est propre).

La conjecture de Novikov peut s'énoncer de la façon suivante : l'élément de $K_{0,top}(\Gamma)$ associé à l'opérateur de signature de M est rationnellement invariant par homotopie. Autrement dit, si $f : M \to N$ est une équivalence d'homotopie entre variétés lisses compactes orientées, la différence entre les éléments de $K_{0,top}(\Gamma)$ associés aux opérateurs de signature de M et de N est de torsion.

Miščenko a démontré que l'image par la flèche de Baum-Connes μ_Γ de cet élément est invariant par homotopie (*cf.* [39]). Donc l'injectivité rationnelle de l'homomorphisme de Baum-Connes implique la conjecture de Novikov.

Pour plus de détails sur ce sujet, on peut voir aussi les exposés [13], [45].

2.5. Généralisations et variantes

Conjecture de Baum-Connes à coefficients

Soient G un groupe localement compact et A une C^*-algèbre munie d'une action de G. On construit de même un groupe de K-théorie topologique à coefficients $K_{0,top}(G;A) = \varinjlim KK_G(C_0(Y), A)$ et des homomorphismes de Baum-Connes $\mu_G^A : K_{0,top}(G;A) \to K_0(A \rtimes G)$ et $\mu_{G,r}^A : K_{0,top}(G;A) \to K_0(A \rtimes_r G)$. On conjecture alors que $\mu_{G,r}^A$ est un isomorphisme.

Conjecture de Baum-Connes pour les groupoïdes

Soit G un groupoïde localement compact. On peut encore (*cf.* [42]) lui associer une C^*-algèbre et une C^*-algèbre réduite (c'est de cette manière que Connes construit la C^*-algèbre d'un feuilletage — *cf.* [9]). À l'aide d'une généralisation aux groupoïdes de la K-théorie équivariante due à Le Gall (*cf.* [38]), on peut encore définir la K-théorie topologique et construire l'homomorphisme de Baum-Connes (*cf.* [4], [47]). On peut de plus définir un homomorphisme de Baum-Connes pour un groupoïde G à coefficients dans une G-algèbre.

Conjecture de Baum-Connes à valeurs dans L^1 (due à J.-B. Bost).

En regardant d'un peu plus près la construction de l'homomorphisme de Baum-Connes, on se rend compte que l'on peut construire en fait un homomorphisme μ_{G,L^1} : $K_{0,top}(G) \to K_0(L^1(G))$ et tel que μ_G est la composée de μ_{G,L^1} avec l'homomorphisme de K-théorie associé à l'homomorphisme $L^1(G) \to C^*(G)$. C'est en fait V. Lafforgue qui, grâce à sa KK-théorie banachique, a défini proprement l'homomorphisme μ_{G,L^1}.

2.6. Le point sur la conjecture

– La conjecture de Baum-Connes à coefficients dans n'importe quelle C^*-algèbre est connue pour tous les groupes (ou groupoïdes) localement compacts moyennables et, plus généralement, ceux opérant proprement par isométries affines sur un espace hilbertien ([19] — *cf.* [48] pour le cas des groupoïdes — voir à ce sujet l'exposé de P. Julg [25]).

Par ailleurs, Oyono ([40]) et Chabert ([8]) ont démontré des résultats de stabilité de la conjecture à coefficients par certaines opérations sur les groupes.

– L'injectivité de l'homomorphisme de Baum-Connes (à coefficients) est de plus connue

a) pour un groupe opérant proprement par isométries sur une variété riemannienne complète à courbure sectionnelle négative ou nulle ([29]), et en particulier pour un sous-groupe fermé d'un groupe de Lie réel ;

b) pour un groupe opérant proprement par isométries sur un immeuble affine et en particulier pour un sous-groupe fermé d'un groupe de Lie p-adique ([30]) ;

c) pour un groupe opérant proprement par isométries sur un espace métrique discret ayant un « bon » comportement à l'infini (faiblement géodésique, à géométrie bornée et « bolique »— [31]) ;

d) pour un groupe admettant une action moyennable sur un espace compact ([18]).

– Lafforgue démontre la conjecture de Baum-Connes :

e) pour tout groupe de Lie semi-simple réel ou p-adique. Ce résultat était connu dans le cas des groupes de Lie réels linéaires connexes réductifs (*cf.* [50]) et les groupes GL_n p-adiques (*cf.* [6]) ;

f) pour les sous-groupes discrets cocompacts des groupes de Lie de rang réel 1 ou de $SL_3(K)$ — avec $K = \mathbf{R},\ \mathbf{C},\ \mathbf{Q}_p$. Dans ce cas on peut aussi admettre des coefficients dans une algèbre $C(X)$, où X est une variété riemannienne compacte munie d'une action isométrique de G. Dans le cas de rang 1, on n'a pas besoin de supposer que l'action de G dans X est isométrique.

C'est le résultat f) le plus spectaculaire de son travail, car il franchit la barrière de la propriété T pour un groupe discret, pour lequel il n'y a donc aucune description possible de $C^*_r(G)$.

2.7. L'élément γ de Kasparov

Dans tous les cas cités ci-dessus où l'on connaît l'injectivité de l'homomorphisme μ_G, on a en fait construit un homomorphsime inverse. Dans les cas a), b) et c) ci-dessus on sait décrire l'image de μ_G : on construit (*cf.* [29] dans le cas a), [26] dans le cas b), [31] dans le cas c)) un élément $\gamma \in KK_G(\mathbf{C}, \mathbf{C})$, qui est idempotent pour le produit de Kasparov et satisfait :
L'image de $\mu_{G,r}$ coïncide avec celle de l'endomorphisme idempotent de $K_0(C^*_r(G))$ associé à $j^r_G(\gamma) \in KK(C^*_r(G), C^*_r(G))$.

En d'autres termes, la conjecture de Baum-Connes se ramène à la formulation suivante : l'endomorphisme de $K_0(C^*_r(G))$ associé à $j^r_G(\gamma)$ est l'identité.

2.8. L'obstruction de la propriété T de Kazhdan

Rappelons (*cf.* [12]) qu'un groupe localement compact G possède la propriété T si la représentation triviale est isolée dans son dual. Du point de vue des C^*-algèbres, cela signifie qu'il existe un projecteur $p \in C^*(G)$ tel que, pour toute représentation π de $C^*(G)$ dans un espace hilbertien, $\pi(p)$ soit le projecteur orthogonal sur les vecteurs invariants par G.

Soit G un groupe localement compact non compact possédant la propriété T. Notons $\lambda : C^*(G) \to C^*_r(G)$ l'homomorphisme canonique. Comme la représentation régulière de G ne possède pas de vecteurs G-invariants (car G n'est pas compact), on a $\lambda(p) = 0$. Or la classe de p est un élément non trivial de la K-théorie de $C^*(G)$, puisque son image par la représentation triviale $\varepsilon : C^*(G) \to \mathbf{C}$ n'est pas nulle dans $K_0(\mathbf{C}) = \mathbf{Z}$. Il s'ensuit que λ n'induit pas un isomorphisme en K-théorie, donc que μ_G et $\mu_{G,r}$ ne sont pas tous deux bijectifs. Une démonstration éventuelle de la conjecture de Baum-Connes doit donc nécessairement distinguer entre $C^*(G)$ et $C^*_r(G)$.

Supposons de plus que G est un groupe comme dans a), b) ou c) ci-dessus. En particulier, il possède un élément γ. Par ce qui précède, l'endomorphisme de $K_0(C^*(G))$ associé à $j_G(\gamma)$ s'annule sur la classe de p ; il en résulte que $\gamma \neq 1$.

Par ailleurs, l'élément γ est représenté par un couple (H, F) où la représentation de G dans l'espace hilbertien H est un sous-multiple de la régulière ; comme la représentation triviale est isolée, il est de toute façon impossible d'espérer une homotopie entre γ et 1.

3. *KK*-THÉORIE BANACHIQUE

On a vu qu'en présence de la propriété T, il n'est pas possible de faire une homotopie de γ à 1 dans $KK_G(\mathbf{C}, \mathbf{C})$.

Cependant, la représentation triviale est isolée parmi les représentations *unitaires* de G, c'est-à-dire des représentations dans des espaces hilbertiens ; elle ne l'est pas dans les représentations banachiques. Par exemple la famille des représentations naturelles de G dans $L^p(G)$ contient faiblement, quand p tend vers $+\infty$, la représentation triviale.

Il est donc possible d'espérer construire une homotopie entre γ et 1 si on s'autorise des représentations banachiques.

Lafforgue est ainsi amené à construire une KK-théorie, où les espaces hilbertiens sont remplacés par des espaces de Banach - et donc les C^*-algèbres par des algèbres de Banach. C'est la KK-théorie banachique.

3.1. *B*-paires

Contrairement aux espaces hilbertiens, les espaces de Banach ne sont pas autoduaux. Pour cette raison, Lafforgue remplace les modules hilbertiens par des paires de modules banachiques en dualité.

Soit B une algèbre de Banach. Une *B-paire* est un couple de modules banachiques en dualité. Autrement dit, une B-paire $(E^<, E^>)$ est la donnée d'un B-module banachique à gauche (*resp.* à droite) $E^<$ (*resp.* $E^>$) et d'une application $\langle\ ,\ \rangle : E^< \times E^> \to B$ satisfaisant : $\forall x \in E^>,\ \xi \in E^<$, l'application $\eta \mapsto \langle \eta, x\rangle$ (*resp.* $y \mapsto \langle \xi, y\rangle$) est $\mathbf{C}$-linéaire, B-linéaire à gauche (*resp.* à droite) et $\|\langle \xi, x\rangle\| \leq \|\xi\|\|x\|$.

Soient $E = (E^<, E^>)$ et $F = (F^<, F^>)$ des B-paires. Un *morphisme de B-paires* de E dans F est un couple $f = (f^<, f^>)$ où $f^< : F^< \to E^<$ (*resp.* $f^> : E^> \to F^>$) est une application $\mathbf{C}$-linéaire, B- linéaire à gauche (*resp.* à droite), continue et satisfait : $\forall x \in E^>,\ \eta \in F^<$, on a $\langle \eta, f^>(x)\rangle = \langle f^<(\eta), x\rangle$. On note $\mathcal{L}(E,F)$ l'espace de Banach des morphismes de B-paires de E dans F (muni de la norme $(f^<, f^>) \mapsto \sup(\|f^<\|, \|f^>\|)$).

Si E, F, G sont des B-paires, $f \in \mathcal{L}(E,F)$ et $g \in \mathcal{L}(F,G)$ on définit $gf \in \mathcal{L}(E,F)$ en posant $(gf)^> = g^> \circ f^>$ et $(gf)^< = f^< \circ g^<$. Munie de ces opérations $\mathcal{L}(E) = \mathcal{L}(E,E)$ est une algèbre de Banach.

Soient $y \in F^>$ et $\xi \in E^<$. On note $\theta_{y,\xi} \in \mathcal{L}(E,F)$ (ou $|y\rangle\langle\xi|$) le morphisme donné par le couple d'applications $F^< \ni \eta \mapsto \langle \eta, y\rangle\xi \in E^<$ et $E^> \ni x \mapsto y\langle \xi, x\rangle \in F^>$. On note $\mathcal{K}(E,F)$ le sous-espace vectoriel fermé de $\mathcal{L}(E,F)$ engendré par les morphismes $\theta_{y,\xi}$.

3.2. Définition de la *KK*-théorie banachique

Soient A et B des algèbres de Banach. Un cycle pour $KK^{\mathrm{ban}}(A,B)$ est encore un triplet (E, π, f) où E est une B-paire $\mathbf{Z}/2\mathbf{Z}$-graduée, $\pi : A \to \mathcal{L}(E)^{(0)}$ est un homomorphisme et $f \in \mathcal{L}(E)^{(1)}$ est tel que, pour tout $a \in A$, les morphismes $f\pi(a) - \pi(a)f$ et $(f^2 - 1)\pi(a)$ sont dans $\mathcal{K}(E)$.

La somme des cycles et l'homotopie sont définies comme dans le cas des modules hilbertiens. Il est facile de voir que l'ensemble des classes d'homotopie de cycles (muni de l'addition des cycles) est un groupe abélien.

Le groupe $KK^{\mathrm{ban}}(A,B)$ est l'ensemble des classes d'homotopie de cycles.

La fonctorialité en A ne pose pas de problème. Soit $\varphi : B_1 \to B_2$ un homomorphisme d'algèbres de Banach. Notons encore $\varphi : \widetilde{B}_1 \to \widetilde{B}_2$ son extension unitale. Soit E_1 une B_1-paire. Considérons-la comme une $\widetilde{B}_1$ paire. On définit une B_2-paire $E_2 = \varphi_*(E_1)$ qui est le quotient du produit tensoriel $E_1 \otimes_\pi \widetilde{B}_2$ par le sous-espace fermé engendré par $xb_1 \otimes b_2 - x \otimes \varphi(b_1)b_2$ avec $x \in E_1,\ b_1 \in \widetilde{B}_1$ et $b_2 \in \widetilde{B}_2$. On construit de plus un homomorphisme $\varphi_* : f \mapsto (f \otimes 1)$ de $\mathcal{L}(E_1)$ dans $\mathcal{L}(E_2)$ qui envoie $\mathcal{K}(E_1)$ dans $\mathcal{K}(E_2)$. On en déduit un homomorphisme $\varphi_* : KK^{\mathrm{ban}}(A, B_1) \to KK^{\mathrm{ban}}(A, B_2)$.

3.3. Le cas équivariant

On suppose à présent qu'un groupe localement compact G opère sur les algèbres A et B. On définit le groupe $KK_G^{\mathrm{ban}}(A,B)$ exactement de la même façon que le groupe de Kasparov correspondant.

Si A et B sont des C^*-algèbres, on a évidemment un homomorphisme naturel $KK_G(A,B) \to KK_G^{\mathrm{ban}}(A,B)$.

Dans le paragraphe suivant, on esquissera le principal résultat technique de Lafforgue montrant (presque) que l'image de γ est égale à celle de 1 dans $KK_G^{\mathrm{ban}}(\mathbf{C}, \mathbf{C})$.

Pour en déduire la conjecture de Baum-Connes, il reste à construire les équivalents en KK-théorie banachique

– du morphisme $j_G : KK_G(\mathbf{C}, \mathbf{C}) \to KK(C_r^*(G), C_r^*(G))$,

– de l'action de la KK-théorie sur la K-théorie, autrement dit un homomorphisme $KK^{\mathrm{ban}}(A, B) \to \mathrm{Hom}(K_0(A), K_0(B))$.

3.4. Action sur la K-théorie

Rappelons que si A est une algèbre de Banach unifère, le groupe $K_0(A)$ est le groupe de Grothendieck du monoïde des classes d'isomorphisme de A-modules projectifs de type fini. Un tel module est isomorphe à un pA^n où $p \in M_n(A)$ est un idempotent.

Remarquons que $KK^{\mathrm{ban}}(M_n(A), B)$ est isomorphe à $KK^{\mathrm{ban}}(A, B)$. Par ailleurs, tout idempotent $p \in A$ détermine un homomorphisme $i_p : \lambda \mapsto \lambda p$ de $\mathbf{C} \to A$. De ces faits, on déduit une application bilinéaire $\varphi : (p, x) \mapsto (i_p)^*(x)$ de $K_0(A) \times KK^{\mathrm{ban}}(A, B)$ dans $KK^{\mathrm{ban}}(\mathbf{C}, B)$.

Cet homomorphisme subsiste si A n'est plus supposée unifère.

Lafforgue démontre de plus que, comme dans le cas de la théorie de Kasparov, l'application naturelle $(x \mapsto \varphi(x, 1_A))$ de $K_0(A)$ dans $KK^{\mathrm{ban}}(\mathbf{C}, A)$ est un isomorphisme.

On en déduit donc que la KK-théorie banachique opère encore sur la K-théorie.

3.5. Complétions « inconditionnelles » et le morphisme j_G

La construction de l'homomorphisme $j_G^r : KK_G(A, B) \to KK(A \rtimes_r G, B \rtimes_r G)$ n'est malheureusement pas possible telle quelle : pour cela, il faudrait un analogue du produit croisé pour les algèbres de Banach, qui pour une C^*-algèbre donnerait le produit croisé réduit.

Cependant, on peut construire dans le cadre des algèbres de Banach le produit croisé $L^1(G; A)$. On en déduit un homomorphisme

$$j_G^{L^1} : KK_G^{\mathrm{ban}}(A, B) \to KK^{\mathrm{ban}}(L^1(G; A), L^1(G; B)).$$

À l'aide de cet homomorphisme $j_G^{L^1}$, l'égalité $\gamma = 1$ dans $KK_G^{\mathrm{ban}}(\mathbf{C}, \mathbf{C})$ implique la conjecture de Baum-Connes dans $L^1(G)$. Pour conclure, il faudrait savoir que l'inclusion $L^1(G) \to C_r^*(G)$ induit un isomorphisme en K-théorie (la surjectivité suffit). Cependant, on n'a pas encore d'outil permettant de démontrer cet isomorphisme (*cf.* la discussion dans [7] où cet isomorphisme est établi dans quelques cas).

Ici intervient une très jolie et simple idée de Lafforgue : celle de « complétion inconditionnelle » : on remarque que pour certaines normes d'algèbre sur $C_c(G)$, en général plus petites que la norme L^1, on peut quand même former le produit croisé banachique :

DÉFINITION 1. — *Une norme N d'algèbre sur $C_c(G)$ sera dite une norme inconditionnelle si, pour tout $f, g \in C_c(G)$, on a $|f(x)| \leq |g(x)|$, $\forall x \in G \Rightarrow N(f) \leq N(g)$.*

On peut construire des produits croisés relatifs à toute norme inconditionnelle. Soient en effet G un groupe localement compact, N une norme inconditionnelle sur $C_c(G)$ et A une algèbre de Banach munie d'une action isométrique de G. Munissons $C_c(G;A)$ du produit de convolution tordu par l'action de G dans A (voir 1.3).

Notons $p_N : C_c(G;A) \to \mathbf{R}_+$ l'application qui à f associe $N(\varphi)$ où φ est l'application $x \mapsto \|f(x)\|$ de $G \to \mathbf{R}$. L'application p_N est une norme d'algèbre sur $C_c(G;A)$.

Notons $A \rtimes_N G$ le complété de $C_c(G;A)$ pour la norme p_N. Si E est un espace de Banach muni d'une action isométrique de G, on peut de même construire l'espace de Banach $E \rtimes_N G$. Si $(E^<, E^>)$ est une B-paire, alors $(E^< \rtimes_N G, E^> \rtimes_N G)$ est une $B \rtimes_N G$-paire.

On obtient un homomorphisme $j_G^N : KK_G^{\mathrm{ban}}(A,B) \to KK^{\mathrm{ban}}(A \rtimes_N G, B \rtimes_N G)$.

Soit G un groupe dans les cas a), b) ou c) de 2.6. Pour montrer la conjecture de Baum-Connes pour G, il suffit donc de

1. Démontrer que $\gamma = 1$ dans $KK_G^{\mathrm{ban}}(\mathbf{C},\mathbf{C})$.
2. Construire une complétion inconditionnelle de $C_c(G)$ qui a la même K-théorie que $C_r^*(G)$.

3.6. Représentations non isométriques de G

En fait, l'homotopie entre γ et 1 n'aura pas lieu dans le groupe $KK_G^{\mathrm{ban}}(\mathbf{C},\mathbf{C})$, mais dans une généralisation.

Soit ℓ une fonction longueur sur G, c'est-à-dire une application continue $G \to \mathbf{R}_+$ telle que $\ell(1) = 0$ et $\ell(xy) \leq \ell(x) + \ell(y)$ pour tout $x, y \in G$ (la plupart des fonctions longueur rencontrées, vérifient de plus $\ell(x^{-1}) = \ell(x)$).

On définit un groupe $KK_{G,\ell}^{\mathrm{ban}}(A,B)$ [(5)] exactement comme $KK_G^{\mathrm{ban}}(A,B)$ sauf qu'ici on suppose que l'action de G dans les B-paires n'est plus isométrique mais elle est *contrôlée par* ℓ c'est-à-dire qu'elle est continue et que la norme de l'action de $x \in G$ est $\leq \exp(\ell(x))$.

Soient ℓ une fonction longueur sur G, N une norme inconditionnelle sur $C_c(G)$ et A une G-algèbre de Banach. Notons N' la norme inconditionnelle $f \mapsto N(e^\ell f)$. Il est facile de construire une application $j_G^{N,\ell} : KK_{G,\ell}^{\mathrm{ban}}(A,B) \to KK^{\mathrm{ban}}(A \rtimes_{N'} G, B \rtimes_N G)$.

De plus soit ℓ une fonction longueur positive. Pour $s \in \mathbf{R}_+^*$, notons N_s la norme inconditionnelle $f \mapsto N(e^{s\ell} f)$. Pour toute G-algèbre de Banach A, la sous-algèbre $\bigcup_{s \in \mathbf{R}_+^*} A \rtimes_{N_s} G$ de $A \rtimes_N G$ a la même K-théorie que $A \rtimes_N G$.

Pour démontrer que γ agit comme l'identité dans $K_0(A \rtimes_N G)$, il suffit de démontrer l'égalité $\gamma = 1$ dans chacun des groupes $KK_{G,s\ell}^{\mathrm{ban}}(\mathbf{C},\mathbf{C})$.

(5) Lafforgue note ce groupe $KK_{G,e^\ell}^{\mathrm{ban}}(A,B)$

4. HOMOTOPIE DE γ À 1

Il y a en fait ici deux cas :

– le cas « géométrique » (groupe opérant proprement et isométriquement sur une variété riemannienne complète de courbure sectionnelle négative ou nulle) ; ce cas contient les groupes de Lie semi-simples réels et leurs sous-groupes fermés.

– le cas « combinatoire » (groupe opérant proprement et isométriquement sur un espace métrique « fortement bolique ») ; ce cas contient les groupes de Lie réductifs p-adiques et leurs sous-groupes fermés ainsi que les sous-groupes discrets cocompacts des groupes de Lie semi-simples réels (et leurs sous-groupes).

Nous nous contenterons ici de donner la construction dans un cas combinatoire : uniquement dans le cas des immeubles de type $\widetilde{A_2}$.

4.1. Le cas des immeubles de type $\widetilde{\mathbf{A_2}}$

Soient X un immeuble de type $\widetilde{A_2}$ (*e.g.* l'immeuble associé à $SL_3(\mathbf{Q}_p)$) et G un groupe localement compact opérant par isométries sur X. Dans ce cas l'élément γ a été construit par P. Julg et A. Valette ([26]).

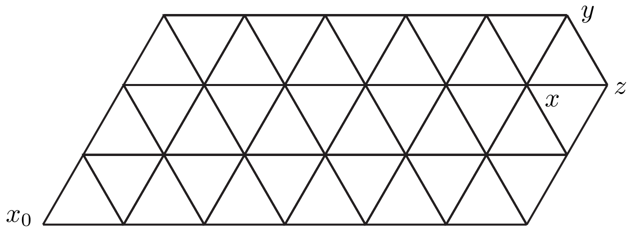
4.1.a. *L'élément γ de Julg-Valette*

On note $X^{(i)}$ $(i = 0, 1, 2)$ l'ensemble des faces de dimension i de X. Notons $(e_x)_{x\in X^{(0)}}$ la base hilbertienne canonique de $\ell^2(X^{(0)})$. L'élément γ est représenté par un couple (H, F) :

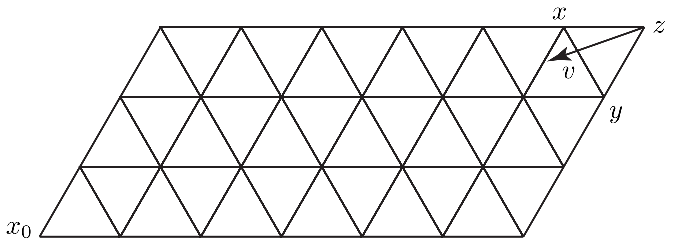
– L'espace hilbertien $H = \displaystyle\bigoplus_{i=0}^{2} H_i$ où $H_0 = \ell^2(X^{(0)})$ et H_1 (*resp.* H_2) est le sous-espace de la deuxième (*resp.* troisième) puissance extérieure de $\ell^2(X^{(0)})$ engendré par les $e_x \wedge e_y$ (*resp.* $e_x \wedge e_y \wedge e_z$) où $\{x, y\} \in X^{(1)}$ (*resp.* $\{x, y, z\} \in X^{(2)}$). Il est gradué sur $\mathbf{Z}/2\mathbf{Z}$ par $H^{(0)} = H_0 \oplus H_2$ et $H^{(1)} = H_1$.

– L'opérateur F s'écrit $F_1 + F_1^*$, où F_1 est défini de la manière suivante : on choisit une origine $x_0 \in X^{(0)}$. Soit $\sigma = \{x, y, z\}$ une chambre et A un appartement contenant x_0 et A. On écrit x_0 dans l'espace affine A à l'aide des coordonnées barycentriques $x_0 = \lambda x + \mu y + \nu z$, ($\lambda, \mu, \nu \in \mathbf{R}$ de somme 1). Quitte à permuter x, y, z, on peut supposer que $\lambda \geq \mu \geq \nu$.

– Si $\lambda > 0 > \mu \geq \nu$, on prendra $F_1(e_y \wedge e_z) = e_x \wedge e_y \wedge e_z$.



– Supposons $\mu \geq 0 \geq \nu$; soit (λ', μ') dans l'espace euclidien $\mathbf{R}^2$, positivement proportionnel à (λ, μ) et de norme 1. On pose $v = \lambda' e_x + \mu' e_y$ $F_1(e_z) = v \wedge e_z$, $F_1(e_y \wedge e_z) = v \wedge (e_y \wedge e_z) = \lambda'(e_x \wedge e_y \wedge e_z)$ et $F_1(e_x \wedge e_z) = \lambda' e_x + \mu' e_y \wedge (e_x \wedge e_z) = -\mu'(e_x \wedge e_y \wedge e_z)$.



On vérifie sans peine que F_1 est ainsi bien défini, qu'on a $F = F^*$, que $1 - F^2$ est le projecteur orthogonal sur e_{x_0}. L'action de G ne fait que changer l'origine x_0. Mais si x_0 et y_0 sont deux points de $X^{(0)}$, la différence entre les coordonnées barycentriques de x_0 et de y_0 sont bornées sur tout l'immeuble. On en déduit que $F_{x_0} - F_{y_0}$ est compact, donc que (H, F) est un élément de $KK_G(\mathbf{C}, \mathbf{C})$.

Il résulte de [30] que sa classe est l'élément γ au sens de 2.7.

4.1.b. *L'homotopie*

Dans la construction qui suit, nous ne considérons pas des $\mathbf{C}$-paires mais des espaces de Banach. Cependant, à un espace de Banach E correspond une $\mathbf{C}$-paire (E^*, E).

Notons $\varphi : X \to \mathbf{R}_+$ la fonction qui à une face f associe la distance à x_0 de son point le plus éloigné et $\ell : G \to \mathbf{R}_+$ la fonction $g \mapsto \varphi(g(x_0))$. C'est une fonction longueur puisque $\ell(gh) = d(gh(x_0), x_0) \leq d(gh(x_0), g(x_0)) + d(g(x_0), x_0)$; or, comme l'action de G est isométrique, $d(gh(x_0), g(x_0)) = d(h(x_0), x_0)$.

Le principal résultat technique de Lafforgue se lit dans notre cas :

Théorème 1. — *Pour tout $s > 0$, les classes de γ et de 1 coïncident dans* $KK^{\mathrm{ban}}_{G,s\ell}(\mathbf{C}, \mathbf{C})$.

Pour $p \in [1, +\infty]$ notons E_p l'espace de Banach, gradué par $0, 1, 2$ dont la composante de degré i est $\ell^p(X^{(i)})$ (vu comme partie de la puissance extérieure $i^{\text{ème}}$ de $\ell^p(X^{(0)})$).

Considérons l'opérateur $\partial : E_p \to E_p$ donné par $\partial(e_x) = 0$, $\partial(e_x \wedge e_y) = e_y - e_x$ et $\partial(e_x \wedge e_y \wedge e_z) = (e_y \wedge e_z) - (e_x \wedge e_z) + (e_x \wedge e_y)$ (pour tout $(x, y, z) \in X^{(2)}$).

Soit enfin $t \in \mathbf{R}_+$; notons A_t l'opérateur non borné de multiplication par $e^{t\varphi}$ et posons $\partial_t = A_t \circ \partial \circ A_t^{-1}$. C'est encore un opérateur borné. De plus, pour tout $g \in G$, $\partial_t - g.\partial_t$ est compact dans tout E_p. En effet, $g.\partial_t$ est construit comme ∂_t, mais avec l'origine x_0 remplacée par gx_0. Mais dans une chambre située loin de x_0, les fonctions distance à x_0 et à gx_0 diffèrent d'une constante additive.

Le résultat crucial est le suivant :

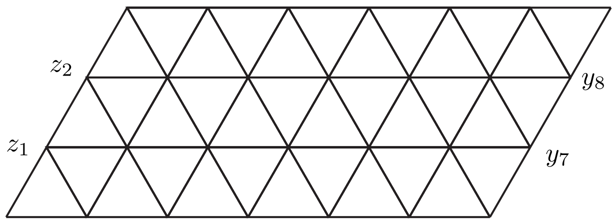
Proposition 1. — *Soit $s > 0$. Il existe $T_s : E_1 \to E_1$ tel que $\mathrm{id}_{E_1} - (T_s\partial_s + \partial_s T_s)$ soit un idempotent de rang 1 et tel que, pour tout $g \in G$, $T_s - g.T_s$ soit compact. De plus, il existe $t > 0$ tel que $T_t = A_{t-s} \circ T_s \circ A_{t-s}^{-1}$ soit continu de E_p dans E_p pour tout $p \in [1, 2]$, et vérifie :*

a) $\mathrm{id}_{E_p} - (T_t\partial_t + \partial_t T_t)$ *est un idempotent de rang* 1 *;*

b) *pour tout $g \in G$, $T_t - g.T_t$ est compact dans tout E_p.*

L'opérateur T_s est défini (par exemple) de la manière suivante :

Soit $x \in X^{(0)}$. Les points x et x_0 déterminent un parallélogramme x_0, y, x, z dans l'immeuble X. Notons j et $n - j$ les distances respectives de x à y et z. Notons $z_0 = x_0, z_1, \ldots, z_j = z, \ldots, z_n = x$ et $y_0 = x_0, y_1, \ldots, y_{n-j} = y, \ldots, y_n = x$ les points sur les deux trajets de x_0 à x par z et y respectivement.



On pose

$$T_0(e_x) = (1 - \frac{j}{n}) \sum_{k=1}^{n} e_{z_{k-1}} \wedge e_{z_k} + \frac{j}{n} \sum_{k=1}^{n} e_{y_{k-1}} \wedge e_{y_k}$$

Il est clair que $\partial \circ T_0(e_x) = e_x - e_0$.

Pour définir $T_0(e_x \wedge e_y)$ on utilise le lemme suivant :

Lemme 1. — *$e_x \wedge e_y - T_0\partial(e_x \wedge e_y)$ est dans l'image de ∂.*

Il suffit pour démontrer ce lemme de se restreindre à un parallélogramme contenant $\{x_0, x, y\}$ et de remarquer que la restriction de ∂ à ce parallélogramme est exacte en dimension 1 et 2.

On définit $T_0(e_x \wedge e_y)$ comme l'élément ξ tel que $\partial(\xi) = e_x \wedge e_y - T_0\partial(e_x \wedge e_y)$ décrit par le lemme 1 (ou plutôt sa démonstration).

Pour achever la démonstration de la proposition 1, il suffit de poser $T_s = A_s \circ T_0 \circ A_s^{-1}$. On montre que :

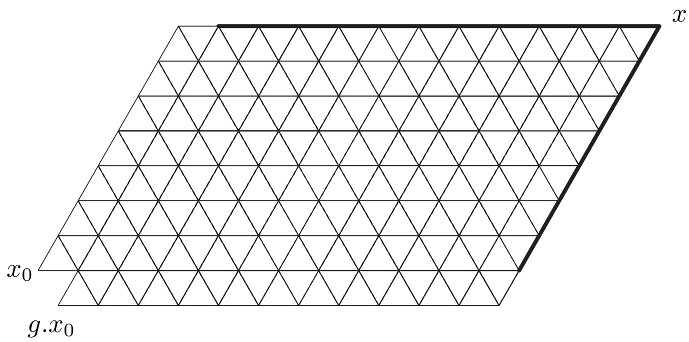
1. *Pour tout $s > 0$, T_s est continu de E_1 dans E_1.*

Pour cela, il suffit de montrer que $\|T_s(e_x)\|$ et $\|T_s(e_x \wedge e_y)\|$ sont bornées indépendamment de $\{x, y\} \in X^{(1)}$: en effet, les faces que font intervenir $T_0(e_x)$ et $T_0(e_x \wedge e_y)$ sont situées dans le parallélogramme considéré ci-dessus ; leur nombre est polynomial en la distance de x à x_0 ; les coefficients apparaissant sont bornés. En conjugant par

A_s, on multiplie ces coefficients par un terme qui décroît exponentiellement, d'où le résultat.

2. *Pour tout $s > 0$ et tout $g \in G$, $T_s - g.T_s \in \mathcal{K}(E_1)$.*

Il suffit pour cela de vérifier que, quand $\{x, y\}$ tend vers l'infini, $\|(T_s - g.T_s)(e_x)\|$ et $\|(T_s - g.T_s)(e_x \wedge e_y)\|$ tendent vers 0. Pour cela, remarquons que $g.T_0$ est construit comme T_0 mais avec l'origine x_0 remplacée par gx_0. Si x est loin de x_0, les trajets de x à x_0 et de x à $g.x_0$ utilisés dans la construction de T_0 coïncident près de x. Par ailleurs, ce sont les points près de x qui « comptent » à cause de la conjugaison par A_s.



3. *Il existe $t > 0$ tel que, pour tout $p \in [1, 2]$, T_t soit continu de E_p dans E_p et, pour tout $g \in G$, $T_t - g.T_t \in \mathcal{K}(E_p)$.*

La difficulté ici vient de ce que $T_t^*(e_x \wedge e_y)$ fait intervenir tous les points z tels que x, y soit sur le trajet de z à x_0. Le coefficient apparaissant décroît en $\exp(-t\varphi(z))$; le nombre de tels z croît exponentiellement. Cependant, en prenant t suffisamment grand, on peut contrôler la norme ℓ^1 de $T_t^*(e_x \wedge e_y)$ et celle de $T_t^*(e_x \wedge e_y \wedge e_z)$. Il en résulte que T_t^* est continu de E_1 dans E_1, donc que T_t est continu de E_∞ dans E_∞. Comme il est continu de E_1 dans E_1, il est continu de E_p dans E_p pour tout p (par interpolation). Un raisonnement analogue montre que $T_t - g.T_t \in \mathcal{K}(E_p)$ pour tout $g \in G$.

En faisant varier $u \in [s, t]$ puis $p \in [1, 2]$ on dispose à présent d'une homotopie entre $(E_1, \partial_s + T_s)$ et $(E_2, \partial_t + T_t)$.

Maintenant, pour achever la démonstration du théorème, on va utiliser les deux faits suivants : pour s tendant vers 0, ∂_s devient de plus en plus G-invariant « donc » $(E_1, \partial_s + T_s)$ tend vers 1 ; pour t tendant vers $+\infty$, $(E_2, \partial_t + T_t)$ ressemble de plus en plus à γ.

Pour mettre en œuvre cette idée, Lafforgue procède de la manière suivante :

On peut construire une homotopie entre $(E_1, \partial_s + T_s)$ et $(E_1^s, \partial + T_0)$ où E_1^s est l'espace de Banach, gradué par $0, 1, 2$ dont la composante de degré i est l'espace ℓ^1 de $X^{(i)}$ à poids $e^{s\varphi}$. Notons que l'action de G dans E_1^s n'est plus isométrique ; c'est pour cela qu'on doit ici quitter $KK_G^{\mathrm{ban}}(\mathbf{C}, \mathbf{C})$ et le remplacer par $KK_{G,s\ell}^{\mathrm{ban}}(\mathbf{C}, \mathbf{C})$.

Pour montrer que la classe de $(E_1^s, \partial + T_0)$ est 1 et celle de $(E_2, \partial_t + T_t)$ est γ, on utilise quelques lemmes généraux de KK-théorie banachique.

Soient G un groupe localement compact muni d'une fonction longueur ℓ, A, B des algèbres de Banach, E une B paire $\mathbf{Z}/2\mathbf{Z}$-graduée munie d'actions de A et de G « contrôlées par ℓ » (au sens de 3.6). Notons D l'ensemble des morphismes S de E tels que $[S, a]$ (pour tout $a \in A$) et $g.S - S$ (pour tout $g \in G$) soient dans $\mathcal{K}(E)$ et $g \mapsto g.S$ soit normiquement continu. Remarquons que tout $F \in D$ de degré 1 tel que $\mathrm{id}_E - F^2 \in \mathcal{K}(E)$ définit un élément de $KK_{G,\ell}^{\mathrm{ban}}(A, B)$.

LEMME 2. — *Soit $S \in D$ de degré 1 tel que $S^2 \in \mathcal{K}(E)$.*

a) *Supposons qu'il existe un morphisme $T \in D$ de degré 1 tel que $\mathrm{id}_E - (TS + ST)$ soit compact. Alors, il existe un tel T avec en plus T^2 compact. Dans ce cas $(E, S+T)$ est un élément de $KK_{G,\ell}^{\mathrm{ban}}(A, B)$.*

b) *Dans ce cas, la classe de $(E, S + T)$ dans $KK_{G,\ell}^{\mathrm{ban}}(A, B)$ ne dépend pas de T.*

En effet, dans a) il suffit de remplacer T par TST, en remarquant que, modulo $\mathcal{K}(E)$, $ST = (\mathrm{id}_E - TS)$, donc TS et ST commutent, donc $STTS \in \mathcal{K}(E)$; on en déduit aussi que $ST - STST$ qui est égal, modulo $\mathcal{K}(E)$ à $STTS$ est dans $\mathcal{K}(E)$; de même $TS - TSTS \in \mathcal{K}(E)$.

Pour montrer b), il suffit de remarquer que l'ensemble des $T \in D$ tels que $ST + TS - \mathrm{id}_E \in \mathcal{K}(E)$ est un sous-espace affine de D.

LEMME 3. — *Soient $S, S' \in D$ de degré 1 tels que $S^2 \in \mathcal{K}(E)$ et $(S')^2 \in \mathcal{K}(E)$. Supposons que le spectre dans l'algèbre de Banach quotient $D/\mathcal{K}(E)$ de $SS' + S'S$ ne rencontre pas $\mathbf{R}_-$. Alors :*

a) *Il existe $T, T' \in D$ tels que $T^2, (T')^2, \mathrm{id}_E - (ST+TS), \mathrm{id}_E - (S'T'+T'S') \in \mathcal{K}(E)$.*

b) *Les classes de $(E, S + T)$ et $(E, S' + T')$ dans $KK_{G,\ell}^{\mathrm{ban}}(A, B)$ coïncident.*

Raisonnons dans l'algèbre de Banach quotient $D/\mathcal{K}(E)$. On a $(SS' + S'S)S = SS'S = S(SS' + S'S)$; de même, $S'(SS' + S'S) = (SS' + S'S)S'$. Donc $(SS' + S'S)^{-1}$ commute avec S et S'. On pose $T = S'(SS'+S'S)^{-1}$ et $T' = S(SS'+S'S)^{-1}$. De plus, comme le spectre de $SS' + S'S$ ne rencontre pas $\mathbf{R}_-$, on peut définir un logarithme de $SS' + S'S$. On obtient une homotopie $S(SS' + S'S)^{-t} + S'(SS' + S'S)^{t-1}$ entre $S + T$ et $S' + T'$.

LEMME 4. — *Soient $S, T \in D$ de degré 1. On suppose $T^2 \in \mathcal{K}(E)$, que S commute exactement à A et à G, que $S^2 = 0$ et , $ST + TS = \mathrm{id}_E$. Alors la classe de $(E, S+T)$ dans $KK_{G,\ell}^{\mathrm{ban}}(A, B)$ est nulle.*

On peut remplacer T par TST. Dans ce cas on a $T^2 = 0$. Comme S commute avec les éléments de A et de G, le sous-espace $S(E)$ de E est invariant par ces éléments. Dans la décomposition $E = S(E) \oplus T(E)$, la matrice de ces éléments est de la forme $\begin{pmatrix} c_{1,1} & c_{1,2} \\ 0 & c_{2,2} \end{pmatrix}$. Notons que $c_{1,2}$ est dans $\mathcal{K}(E)$ car $T \in D$. Il suffit de changer ces actions en $\begin{pmatrix} c_{1,1} & 0 \\ 0 & c_{2,2} \end{pmatrix}$ à travers une homotopie $\begin{pmatrix} c_{1,1} & tc_{1,2} \\ 0 & c_{2,2} \end{pmatrix}$ ($t \in [0,1]$). Au bout de cette homotopie, $S+T$ commute exactement avec A et G et $(S+T)^2 = \mathrm{id}_E$. Le lemme en résulte facilement.

Pour finir la démonstration du théorème, notons u la classe dans $KK^{\mathrm{ban}}_{G,s\ell}(\mathbf{C},\mathbf{C})$ de $(E_1^s, \partial + T_0)$; posons $F = E_1^s \oplus \mathbf{C}$ où $\mathbf{C}$ est muni de l'action triviale de G et est de degré 1 pour la graduation. Prolongeons ∂ de F dans F en posant $\partial(e_x) = 1 \in \mathbf{C}$ pour tout $x \in X^{(0)}$ et posons $T_0(1) = e_{x_0}$. Il est facile de voir que la classe de cet élément est $u - 1$ (ou plutôt u — l'image de 1). Par le lemme 4 appliqué à $S = \partial$, on trouve donc que u et l'image de 1 dans $KK^{\mathrm{ban}}_{G,s\ell}(\mathbf{C},\mathbf{C})$ coïncident.

Pour montrer que la classe de $(E_2, \partial_t + T_t)$ est l'image de γ, on commence d'abord par modifier très légèrement F_1 en un opérateur F_2 : le vecteur $\lambda' e_y + \mu' e_z$ par lequel on prend le produit extérieur est normalisé en norme 1 au lieu de l'être en norme 2. Il suffit alors, par le lemme 3, de démontrer que le spectre de $\partial_t F_2 + F_2 \partial_t$ est disjoint de $\mathbf{R}_-$. Pour cela, on remarque que l'image par $(\partial_t F_2 + F_2 \partial_t) - \mathrm{id}_{E_2}$ d'une face f est portée par des faces voisines, plus proches de x_0 que f. Donc, pour t assez grand, le rayon spectral de $(\partial_t F_2 + F_2 \partial_t) - \mathrm{id}_{E_2}$ est < 1.

4.2. Les autres cas

4.2.a. *Le cas « combinatoire »*

Dans ce cas, la construction est exactement la même. La seule chose qui devient réellement plus difficile ici est la démonstration du fait que le complexe ∂ est acyclique.

4.2.b. *Le cas « géométrique »*

Soient M une variété riemannienne complète, connexe, simplement connexe, de courbure sectionnelle négative ou nulle (bornée inférieurement) et G un groupe localement compact agissant sur M par isométries.

Soit x_0 un point de M et notons $\varphi : M \to \mathbf{R}$ l'application donnée par $\varphi(x) = \exp(\sqrt{d(x,x_0)^2 + 1})$.

Dans ce cas, on remplace le bord de la cohomologie simpliciale par la cohomologie de de Rham. On est amené à travailler avec des espaces $W(p,k,s-k,\varphi^t)$: espace de Sobolev L^p, des k formes $s-k$ fois différentiables à poids φ^t (s est, par exemple, la dimension de l'espace M).

Une difficulté qui apparaît ici est que le calcul pseudodifféreniel ne marche bien que pour $p \in]0,+\infty[$. Cependant, $t > 0$ étant choisi, le complexe de de Rham $(W(p,k,s-k,\varphi^t), d)$ est acyclique pour p assez proche de $+\infty$.

Le reste de la construction est essentiellement le même que dans le cas « combinatoire ».

Je dois cependant dire que, bien que je sois absolument convaincu que la construction marche aussi bien que dans le cas combinatoire, je n'ai pas encore vérifié tous les détails du cas géométrique.

5. FIN DE LA DÉMONSTRATION

Soit G un groupe localement compact opérant proprement et isométriquement sur une variété riemannienne complète de courbure sectionnelle négative ou nulle ou sur un immeuble affine. D'après ce qui précède, G possède un élément γ et $\gamma = 1$ dans la KK-théorie banachique. Donc, pour établir la conjecture de Baum-Connes, il suffit de construire une complétion inconditionnelle A de $C_c(G)$ telle que le morphisme $A \to C_r^*(G)$ induise une surjection au niveau des groupes K_0.

5.1. Cas des groupes de Lie

Si G est un groupe de Lie semi-simple, réel ou p-adique, Lafforgue ([36]) montre que certaines variantes de l'espace de Schwarz (ou algèbre de Harish-Chandra) de G fournissent des complétions inconditionnelles de $C_c(G)$ qui permettent de terminer la démonstration. On obtient ainsi une preuve plus directe de résultats de Wassermann dans le cas réel ([50]) et une généralisation de Baum, Higson et Plymen dans le cas p-adique ([6]).

5.2. Propriété (RD) de Haagerup-Jolissaint

Soient G un groupe discret, ℓ une fonction longueur sur G (*i.e.* une application $G \to \mathbf{R}_+$ telle que $\ell(e) = 0$, $\ell(xy) \leq \ell(x) + \ell(y)$ pour tout $x, y \in G$). Notons $H^\infty(G, \ell)$ l'espace vectoriel des fonctions $f : G \to \mathbf{C}$ telles que, pour tout $p \in \mathbf{R}_+$, $\sum\limits_{x \in G} \ell(x)^p |f(x)|^2 < \infty$. Si G est finiment engendré, on note $H^\infty(G)$ l'espace $H^\infty(G, \ell)$ où ℓ est la fonction longueur associée à un système fini de générateurs (l'espace $H^\infty(G)$ n'en dépend pas).

Dans [16], Haagerup a démontré :

Proposition 1. — *Si G est un groupe libre à un nombre fini de générateurs, $H^\infty(G) \subset C_r^*(G)$.*

En effet :

Pour $p \in \mathbf{N}$, notons χ_p la fonction caractéristique des mots de longueur p (pour un système libre de générateurs). Soient $f, g : G \to \mathbf{C}$ des fonctions. On utilise le lemme suivant :

Lemme 1. — *Pour tout $p, q, n \in \mathbf{N}$, on a $\|((f\chi_p) \star (g\chi_q))\chi_n\|_2 \leq \|f\chi_p\|_2 \|g\chi_q\|_2$. De plus si $((f\chi_p) \star (g\chi_q))\chi_n \neq 0$, alors n est compris entre $|p - q|$ et $p + q$ et a la même parité que $p + q$.*

On peut évidemment supposer que $f = f\chi_p$ et $g = g\chi_q$.

La contrainte sur n est claire. Écrivons $p = p_1 + r,\ q = q_1 + r$ et $n = p_1 + q_1$. Notons C_k l'ensemble des mots de longueur k.

Soient $x, y, z \in G$ de longueur respective p, q, n tels que $xy = z$. Écrivons $x = x_1 x_2$ où $x_1 \in C_{p_1}$ et $x_2 \in C_r$ et $y = y_2 y_1$ où $y_1 \in C_{q_1}$ et $y_2 \in C_r$ (ces éléments sont uniquement déterminés par x et y). Comme xy est de longueur n on voit que nécessairement $z = x_1 y_1$ et $y_2 = x_2^{-1}$. Notons que x_1 et y_1 sont entièrement déterminés par z.

Il s'ensuit que

$$|f \star g(z)| = \Big|\sum_{u \in C_r} f(x_1 u) g(u^{-1} y_1)\Big| \leq (\sum_{u \in C_r} |f(x_1 u)|^2)^{1/2} (\sum_{u \in C_r} |g(u y_1)|^2)^{1/2}$$

par l'inégalité de Cauchy-Schwarz.

On a alors

$$\begin{aligned}
\|(f \star g)\chi_n\|_2^2 &= \sum_{x_1 \in C_{p_1},\, y_1 \in C_{q_1},\, x_1 y_1 \in C_n} |f \star g(x_1 y_1)|^2 \\
&\leq \sum_{x_1 \in C_{p_1},\ y_1 \in C_{q_1},\ x_1 y_1 \in C_n} \Big(\sum_{u \in C_r} |f(x_1 u)|^2\Big)\Big(\sum_{u \in C_r} |g(u y_1)|^2\Big) \\
&\leq \sum_{x_1 \in C_{p_1},\, y_1 \in C_{q_1}} \Big(\sum_{u \in C_r} |f(x_1 u)|^2\Big)\Big(\sum_{u \in C_r} |g(u y_1)|^2\Big) \\
&= \|f\|_2^2 \|g\|_2^2 .
\end{aligned}$$

La proposition 1 se déduit très facilement du lemme 1 :

On écrit $\ell^2(G)$ comme somme hilbertienne de $\ell^2(C_r)$.

Soient $f \in C_c(G)$ et $p \in \mathbf{N}$. L'opérateur de convolution à gauche par $(f\chi_p)$ peut être vu dans la décomposition ci-dessus comme une matrice $T_{r,s}$ où $T_{r,s} : \ell^2(C_s) \to \ell^2(C_r)$ est un opérateur de norme $\leq \|(f\chi_p)\|_2$, nul si $|r - s| \geq p$, ou si $r - s$ n'est pas de la même parité que p. On peut écrire T comme somme des $p + 1$ opérateurs S_k où S_k est l'opérateur de matrice $T_{r,s}$ avec $r - s = -p + 2k$ $(0 \leq k \leq p)$. Il est clair que $\|\lambda(f\chi_p)\| \leq \sum_k \|S_k\| \leq (p+1)\|(f\chi_p)\|_2$. Il vient

$$\|\lambda(f)\| \leq \sum (p+1)\|f\chi_p\|_2 \leq (\sum (p+1)^{-2})^{1/2} \sum_p ((p+1)^4 \|f\chi_p\|_2^2)^{1/2}$$

par l'inégalité de Cauchy-Schwartz, d'où la proposition.

On peut en fait, à partir du lemme 1, démontrer :

Proposition 2. — *Si G est un groupe libre à un nombre fini de générateurs, $H^\infty(G)$ est une sous-algèbre de $C_r^*(G)$ stable par calcul fonctionnel holomorphe. En particulier, l'inclusion $H^\infty(G) \to C_r^*(G)$ induit un isomorphisme au niveau de la K-théorie.*

Cette idée est reprise par Jolissaint ([22]). Il donne la définition suivante :

Définition 2. — *On dit qu'un groupe finiment engendré G possède la propriété (RD) si $H^\infty(G) \subset C_r^*(G)$.*

Il démontre que si G a la propriété (RD), alors $H^\infty(G)$ est une sous-algèbre de $C_r^*(G)$ stable par calcul fonctionnel. En fait (*cf.* [23]), il existe $r \in \mathbf{N}$ tel que l'ensemble $A_r = \{f \in \mathbf{C}^G, \ \sum_{x\in G} \ell(x)^r |f(x)|^2 < \infty\}$ est une sous-algèbre de $C_r^*(G)$ stable par calcul fonctionnel holomorphe.

Remarquons que A_r est une algèbre de Banach et une complétion inconditionnelle.

5.3. Groupes possédant la propriété (RD)

Jolissaint ([22]) étend le travail de Haagerup en démontrant que plusieurs groupes se comportant comme les groupes libres, en particulier les sous-groupes cocompacts des groupes de Lie réels simples de rang 1, ont la propriété (RD).

Ce résultat est lui-même étendu par de la Harpe ([17]) aux groupes hyperboliques de Gromov.

Très récemment, Ramagge, Robertson et Steger ([41]) ont établi la propriété (RD) pour des groupes discrets opérant proprement avec quotient compact sur un immeuble $\widetilde{A}_2$. Leur résultat contient en particulier les réseaux cocompacts de $SL_3(\mathbf{F})$ où $\mathbf{F}$ est un corps local non archimédien.

Comme dans la démonstration de Haagerup, on doit analyser les couples x, y ayant un produit donné. On est amené à étudier les triangles dans l'immeuble. La clé de la démonstration est que tout triangle se contracte uniquement sur un triangle équilatéral dans un appartement.

Lafforgue enfin, en remplaçant certaines égalités par des estimations, a adapté la démonstration de Ramagge, Robertson et Steger au cas des sous-groupes cocompacts de $SL_3(\mathbf{R})$ et $SL_3(\mathbf{C})$ (*cf.* [34]).

5.4. Limite de la méthode

D'après une conjecture de Valette ([49]), tout sous-groupe cocompact d'un groupe de Lie simple devrait posséder la propriété (RD).

Cependant, la méthode a ses limites : plusieurs sous-groupes discrets (non cocompacts) de groupes de Lie ne possèdent pas la propriété (RD) du fait de l'existence d'éléments paraboliques. Par exemple, les seuls groupes moyennables possédant la propriété (RD) sont les groupes à croissance polynomiale (*cf.* [22]). De plus cette méthode marche très mal pour les produits croisés — et ne permet de démontrer la conjecture de Baum-Connes à coefficients que dans des cas très particuliers (*cf.* [37]).

6. QUELQUES AUTRES AVANCÉES RÉCENTES

En plus du résultat de Lafforgue que nous venons de discuter, il y a eu récemment quelques autres avancées significatives sur la conjecture de Baum-Connes et des sujets connexes.

6.1. Actions moyennables ([18])

N. Higson vient de faire une remarque très simple et astucieuse qui permet d'établir l'injectivité de l'homomorphisme de Baum-Connes (à coefficients) pour les groupes discrets possédant une action moyennable sur un espace compact.

La moyennabilité d'un groupe est une notion dégagée depuis plusieurs décennies. On dispose à présent de plusieurs propriétés équivalentes. De plus, cette notion a été étendue aux actions de groupes et aux groupoïdes. Nous renvoyons à l'excellente monographie [2] pour une discussion très actuelle sur toutes les notions de moyennabilité.

Du point de vue des C^*-algèbres, un groupe localement compact G est moyennable si $C^*(G) = C^*_r(G)$. Une action de G sur un espace compact X est moyennable si $C(X) \rtimes G = C(X) \rtimes_r G$.

Rappelons que Higson et Kasparov ont démontré la conjecture de Baum-Connes pour tout groupe moyennable, et que Tu a étendu ce résultat au cas des groupoïdes moyennables.

Toute action d'un groupe moyennable est moyennable. Cependant, plusieurs groupes, bien que non moyennables, possèdent des actions moyennables sur des espaces compacts. Par exemple, $G = SL_n(\mathbf{R})$ n'est pas moyennable pour $n \geq 2$, mais son action dans l'espace compact des drapeaux G/P est moyennable où P désigne le sous-groupe de $SL_n(\mathbf{R})$ formé des matrices triangulaires supérieures (à coefficients diagonaux positifs).

La démonstration de Higson est (un peu schématiquement) la suivante :

Si l'action de G dans un espace compact X est moyennable, son action dans l'espace M des mesures de probabilité sur X, muni de la topologie de la convergence vague est moyennable.

Soit A une C^*-algèbre munie d'une action de G. Notons B la C^*-algèbre $C(M; A)$ des fonctions continues de M dans A. L'homomorphisme $A \to B$ qui à $a \in A$ associe la fonction constante égale à a étant équivariant, on obtient un homomorphisme $A \rtimes_r G \to B \rtimes_r G$, d'où un homomorphisme de groupes de K-théorie. De même, on a un homomorphisme $K_{0,top}(G; A) \to K_{0,top}(G; B)$. On obtient un diagramme commutatif

$$\begin{array}{ccc} K_{0,top}(G;A) & \xrightarrow{\mu_G^A} & A \rtimes_r G \\ \downarrow & & \downarrow \\ K_{0,top}(G;B) & \xrightarrow{\mu_G^B} & B \rtimes_r G \end{array}$$

Or, l'action de G dans M étant moyennable, par la généralisation par Tu ([48]) du théorème de Higson-Kasparov, l'homomorphisme de Baum-Connes μ_G^B est un isomorphisme. De plus, comme M est contractile, l'application $A \to B$ est une équivalence d'homotopie, d'où l'on déduit que l'homomorphisme $K_{0,top}(G;A) \to K_{0,top}(G;B)$ est aussi un isomorphisme. En fait pour avoir l'isomorphisme ici, Higson doit supposer que le groupe G est discret, il utilise un modèle simplicial pour $\underline{E}G$ et une suite exacte de Mayer-Vietoris. Il s'ensuit que l'homomorphisme μ_G^A est injectif (et scindé).

On peut facilement obtenir un résultat légèrement plus fort que celui de Higson en remplaçant la condition de moyennabilité de l'action de G par l'existence d'un plongement uniforme de G dans un espace hilbertien (au sens ci-dessous, *cf.* [46]).

6.2. Conjecture de Baum-Connes et géométrie « à l'infini » ([43], [51])

Le résultat de [18] cité ci-dessus (ainsi que sa généralisation [46]) avait d'abord été démontré par Yu dans [51] sous l'hypothèse que G était sans torsion et BG un complexe cellulaire fini(6).

Le démonstration de Yu est tout à fait différente de celle de Higson : elle utilise le langage de la géométrie « grossière » (coarse), je dirais plutôt géométrie à l'infini.

Il s'agit là de tout un monde — et il n'est pas possible de citer ici toutes les contributions importantes dans le sujet. Nous renvoyons à [43] et [51] pour une discussion détaillée de toutes ces notions.

Il est apparu (*cf.* [14]) que la conjecture de Novikov pour un groupe discret G est en fait assez peu dépendante de la structure de groupe. La seule chose qui compte est la structure à l'infini de G.

Soit (X,d) un espace métrique. De même que pour une structure uniforme, une structure à l'infini de X est basée sur les ensembles de la forme $\Delta_r = \{(x,y) \in X \times X,\ d(x,y) < r\}$ (pour $r \in \mathbf{R}_+$). Cependant on s'intéresse à r grand.

On dit qu'une partie $U \subset X \times X$ est un *entourage* pour la structure à l'infini *s'il est contenu* dans un ensemble Δ_r.

Le rôle des applications uniformément continues est joué ici par les applications $f : X \to Y$ telles que l'image directe de tout entourage soit un entourage (autrement dit $\forall r \in \mathbf{R}_+,\ \exists R \in \mathbf{R}_+,\ d(x,x') \leq r \Rightarrow d(f(x), f(x')) \leq R$).

Une application $f : X \to Y$ est un *plongement uniforme* si, pour toute partie U de $X \times X$, U est un entourage si et seulement si $f(U)$ est un entourage.

À un espace métrique, on associe une C^*-algèbre construite de la manière suivante : on fixe un espace hilbertien H et une représentation $\pi : C_0(X) \to \mathcal{L}(H)$ telle que $\pi^{-1}(\mathcal{K}(H)) = \{0\}$ et $\pi(C_0(X))H = H$. Le support d'un élément $T \in \mathcal{L}(H)$ est le plus petit fermé F de $X \times X$ tel que si $f, g \in C_0(X)$ sont telles que $(x,y) \mapsto f(x)g(y)$ est nulle sur F, alors $\pi(f)T\pi(g)$. La C^*-algèbre de X est l'adhérence de l'ensemble

(6) C'est Higson et Roe dans [21] qui ont reformulé le théorème de [51] en termes de moyennabilité.

des opérateurs $T \in \mathcal{L}(H)$ dont le support est un entourage et tels que, pour tout $f \in C_0(X)$, $T\pi(f)$ et $\pi(f)T$ sont compacts.

Il y a aussi une conjecture de Baum-Connes qui prédit la K-théorie de cette C^*-algèbre (*cf.* [43]), du moins si X est à géométrie à l'infini bornée. Rappelons qu'un espace métrique X est dit *à géométrie à l'infini bornée* si, pour tout $r \in \mathbf{R}_+$, il existe $N \in \mathbf{N}$ tel que le nombre d'éléments de toute partie de X de diamètre $\leq r$ est borné par N.

Le lien avec la conjecture de Baum-Connes est le suivant : soit G un groupe finiment engendré. Munissons-le de la distance invariante à gauche associée à la longueur de mots (par rapport à un système fini de générateurs). Si G n'a pas de torsion et BG est un complexe cellulaire fini, la conjecture de Baum-Connes à l'infini pour l'espace métrique G implique l'injectivité de $\mu_{G,r}$.

En utilisant une variante de la méthode de Higson-Kasparov, Yu ([51], voir aussi [46] pour une autre démonstration de ce résultat) montre :

Théorème 1. — *Tout espace métrique à géométrie à l'infini bornée qui admet un plongement uniforme dans un espace hilbertien satisfait la « conjecture de Baum-Connes à l'infini ».*

RÉFÉRENCES

[1] C. Anantharaman-Delaroche – *Classification des C^*-algèbres purement infinies nucléaires (d'après E. Kirchberg)*, Sém. Bourbaki 1995/96, exp. n° 805, Astérisque **241** (1997), 7-27.

[2] C. Anantharaman-Delaroche et J. Renault *Amenable groupoids* (preprint 1999), à paraître à Ens. Math.

[3] M.F. Atiyah – *Elliptic operators discrete groups and von Neumann algebras*, Astérisque **32-33** (1976), 43-72.

[4] P. Baum and A. Connes – *K-theory for Lie groups and foliations* (preprint 1982), à paraître à Ens. Math.

[5] P. Baum, A. Connes and N. Higson – *Classifying space for proper actions and K-theory of group C^*-algebras*, Contemporary Math. **167** (1994), 241-291.

[6] P. Baum, N. Higson and R. P Lymen – *A proof of the Baum-Connes conjecture for p-adic $GL(n)$*, C. R. Acad. Sci. Paris, Sér. I Math. **325**, n° 2 (1997), 171-176.

[7] J.-B. Bost – *Principe d'Oka, K-théorie et systèmes dynamiques non commutatifs*, Inv. Math. **101** (1990), 261-333.

[8] J. Chabert – *Stabilité de la conjecture de Baum-Connes pour certains produits semi-directs de groupes*, C. R. Acad. Sci. Paris, Sér. I Math. **328**, n° 12 (1999), 1129–1132.

[9] A. Connes – *Sur la théorie non commutative de l'intégration*, Lect. Notes in Math. **725** (1979), 19-143.

[10] J. Cuntz – *K-theoretic amenability for discrete groups*, J. Reine ang. Math. **344** (1983), 180-195.

[11] J. Cuntz – *Bivariante K-theorie für lokalconvexe Algebren und der Chern-Connes Character*, Doc. Math. **2** (1997), 139-182.

[12] C. Delaroche et A.A. Kirillov – *Sur les relations entre l'espace dual d'un groupe et la structure de ses sous-groupes fermés (d'après D. A. Kazhdan)*, Sém. Bourbaki 1967/68, exp. n 343, Soc. Math. France **10** (1995), 507-528.

[13] T. Fack – *K-théorie bivariante de Kasparov*, Sém. Bourbaki 1982/83, Astérisque **105-106**, (1983), 149-166.

[14] M. Gromov – *Geometric reflections on the Novikov conjecture*, Novikov conjectures, index theorems and rigidity, Vol. 1 (Oberwolfach, 1993), London Math. Soc. Lecture Note Ser. **226** Cambridge Univ. Press (1995), 164–173.

[15] M. Gromov – *Spaces and questions* (1999).

[16] U. Haagerup – *An example of a nonnuclear C^*-algebra which has the metric approximation property*, Inv. Math. **50** (1979), 279-293.

[17] P. de la Harpe – *Groupes hyperboliques, algèbres d'opérateurs et un théorème de Jolissaint*, C. R. Acad. Sci. Paris, Sér. I Math. **307** (1988), 771-774.

[18] N. Higson – *Bivariant K-theory and the Novikov conjecture* (1999).

[19] N. Higson and G. Kasparov – *Operator K-theory for groups which act properly and isometrically on Hilbert space*, Electronic Research Anouncements, AMS **3** (1997), 131-141.

[20] N. Higson, V. Lafforgue and G. Skandalis – *Counterexamples to the Baum-Connes Conjectures*, en préparation.

[21] N. Higson and J. Roe – *Amenable group actions and the Novikov conjecture* (1999).

[22] P. Jolissaint – *Rapidly decreasing functions in reduced C^*-algebras of groups*, Trans. Amer. Math. Soc. **317** (1990), 167-196.

[23] P. Jolissaint – *K-theory of reduced C^*-algebras and rapidly decreasing functions on groups*, K- theory **2** (1989), 723-735.

[24] P. Julg – *Remarks on the Baum-Connes conjecture and Kaszdan's property T*, Fields Inst. Comm. **13** (1997), 145-153.

[25] P. Julg – *Travaux de N. Higson et G. Kasparov sur la conjecture de Baum-Connes*, Sém. Bourbaki, vol. 1997/98, exp. n 841, Astérisque **252** (1998), 151-183.

[26] P. Julg et A. Valette – *Fredholm modules associated to Bruhat-Tits Buildings*, Proc. of the Center for Math. Analysis, Australian National University **16** (1988), 143-155.

[27] G.G. KASPAROV – *Hilbert C^*-modules : theorems of Stinespring and Voiculescu*, J. Operator Theory **4** (1980), 133-150.

[28] G.G. KASPAROV – *The operator K-functor and extensions of C^*-algebras*, Math. USSR Izv. **16** (1981), n 3, 513-572. *Translated from :* Izv. Akad. Nauk. S.S.S.R. Ser. Mat. **44** (1980), 571-636.

[29] G.G. KASPAROV – *Equivariant KK-theory and the Novikov conjecture*, Inv. Math. **91** (1988), 147-201.

[30] G.G. KASPAROV and G. SKANDALIS – *Groups acting on buidings, Operator K-theory and Novikov's conjecture*, K-theory **4** (1991), 303-337.

[31] G.G. KASPAROV and G. SKANDALIS – *Groupes « boliques » et conjecture de Novikov*, Note C.R.A.S. **319**, Sér. I (1994), 815-820.

[32] D. KAZHDAN – *Connection of the dual space of a group with the structure of its closed subgroups*, Funct. Anal. and its Appl. **1** (1967), 63-65.

[33] V. LAFFORGUE – *Une démonstration de la conjecture de Baum-Connes pour les groupes réductifs sur un corps p-adique et pour certains groupes discrets possédant la propriété (T)*, C. R. Acad. Sci. Paris, Sér. I Math. **327**, n° 5 (1998), 439-444.

[34] V. LAFFORGUE – *A proof of property (RD) for discrete cocompact subgroups of* $SL_3(\mathbf{R})$ (preprint 1998), à paraître à Journal of Lie theory.

[35] V. LAFFORGUE – *K-théorie bivariante pour les algèbres de Banach et conjecture de Baum-Connes* (preprint 1998).

[36] V. LAFFORGUE – *Espaces de Schwartz* (preprint 1998).

[37] V. LAFFORGUE – *K-théorie bivariante pour les algèbres de Banach, groupoïdes et conjecture de Baum-Connes* (preprint en préparation 1999).

[38] P.-Y. LE GALL – *Théorie de Kasparov équivariante et groupoïdes*, C. R. Acad. Sci. Paris, Sér. I Math. **324**, n 6 (1997), 695-698.

[39] A.S. MIŠČENKO – *Homotopy invariance of non simply connected manifolds, I : Rational invariance*, Math. USSR Izv. **4** (1970), 509-519, transl from *Izv. Akad. Nauk. SSSR Ser. Math.* **34**, (1970), 501-514.

[40] H. OYONO-OYONO – *La conjecture de Baum-Connes pour les groupes agissant sur les arbres*, C. R. Acad. Sci. Paris, Sér. I Math. **326**, n° 7 (1998), 799-904.

[41] J. RAMAGGE, G. ROBERTSON and T. STEGER – *A Haagerup inequality for $\widetilde{A}_1 \times \widetilde{A}_1$ and $\widetilde{A}_2$ buildings*, Geometric and Funct. Anal. **8**, n 4 (1998), 702-731.

[42] J.N. RENAULT – *A groupoid approach to C^*-algebras*, Lecture Notes in Math. **793**, Springer-Verlag, New York (1980).

[43] J. ROE – *Index Theory, Coarse Geometry, and Topology of Manifolds*, CBMS Regional Conf. Series in Math. **90**, AMS (1996).

[44] G. SKANDALIS – *Une notion de nucléarité en K-théorie (d'après J.Cuntz)*, K-theory **1** (1988), 549-573.

[45] G. Skandalis – *Approche de la conjecture de Novikov par la cohomologie cyclique. D'après Connes-Gromov-Moscovici*, Sém. Bourbaki, vol. 1990/91, exposé n^{o} 739, Astérisque **201-202-203** (1992), 299-316.

[46] G. Skandalis, J.L. Tu and G. Yu – *Coarse Baum-Connes conjecture and Groupoids* (preprint 1999).

[47] J.L. Tu – *La conjecture de Novikov pour les feuilletages hyperboliques*, K-theory **16**, n 2 (1999), 129-184.

[48] J.L. Tu – *La conjecture de Novikov pour les feuilletages moyennables*, K-theory **17**, n 3 (1999), 215-264.

[49] A. Valette – *Questions*, Novikov conjectures, index theorems and rigidity, Vol. 1 (Oberwolfach, 1993), London Math. Soc. Lecture Note Ser. **226**, Cambridge Univ. Press (1995), p. 74.

[50] A. Wassermann – *Une démonstration de la conjecture de Connes-Kasparov pour les groupes de Lie linéaires connexes réductifs*, C. R. Acad. Sci. Paris, Sér. I Math. **304** (1987), 559-562.

[51] G. Yu – *The coarse Baum-Connes conjecture for spaces which admit uniform embeddings into Hilbert space* (preprint 1999), à paraître à Inv. Math.

Georges SKANDALIS

UMR 7586 du C.N.R.S.
Université de Paris VII
Case Postale 7012
2, place Jussieu
F-75251 PARIS Cedex 05
E-mail : skandal@math.jussieu.fr

Séminaire BOURBAKI Mars 2000
52e année, 1999-2000, n° 870, p. 137 à 159

THÉORIE DES MODÈLES ET CONJECTURE DE MANIN-MUMFORD [d'après Ehud Hrushovski]

par **Élisabeth BOUSCAREN**

1. INTRODUCTION

Je vais présenter ici l'une des applications récentes de la théorie des modèles à la géométrie diophantienne, la nouvelle démonstration, due à E. Hrushovski ([Hr2]), de la conjecture de Manin-Mumford (généralisée) qui est basée sur l'étude, du point de vue de la théorie des modèles, des corps algébriquement clos munis d'un automorphisme.

THÉORÈME 1.1 (Conjecture de Manin-Mumford). — *Soit G un groupe algébrique commutatif connexe défini sur un corps de nombres k et X une sous-variété de G. Alors il y a un nombre fini de points de torsion de G, $a_1, \ldots, a_M$, et de sous-groupes algébriques de G, $G_1, \ldots, G_M$ tels que pour chaque i, $a_i + G_i$ est contenu dans X et*

$$X(k^{alg}) \cap Tor(G) = \bigcup_{i=1}^{M} a_i + Tor(G_i).$$

De plus on peut borner de façon effective le nombre M de translatés : il existe c, e deux constantes ne dépendant que d'invariants liés à G (et non à X) et du choix de deux premiers de bonne réduction pour G, tels que $M \leq c \deg(X)^e$.

En utilisant des méthodes de calcul assez grossières, on obtient que les constantes c et e sont doublement exponentielles.

La démonstration de Hrushovski passe par une première étape où il prend un premier de bonne réduction pour G, de caractéristique résiduelle p, et montre le résultat pour la torsion première à p, c'est-à-dire pour le sous-groupe $Tor_{p'}(G)$ des éléments de torsion de G dont l'ordre n'est pas divisible par p. On obtient dans ce cas que

$$X(k^{alg}) \cap Tor_{p'}(G) = \bigcup_{i=1}^{N} a_i + Tor_{p'}(G_i),$$

où il est facile de montrer directement que $N \leq a\ \deg(X)^b$, avec a doublement exponentiel (en la dimension de A, le degré de l'addition dans A et la cardinalité du corps résiduel) et b exponentiel en la dimension de A.

Le but de cet exposé, en présentant la démonstration de Hrushovski, est d'essayer d'expliquer pourquoi la théorie des modèles a quelque chose à dire sur ce type de questions et aussi de donner une idée des résultats qu'on montre et des outils qu'on développe en théorie des modèles.

Nous passerons donc rapidement sur les parties purement « géométrie algébrique » de la démonstration.

De même, la très brève présentation historique et bibliographique qui suit n'a, cela sera visible tout de suite, aucune prétention d'exhaustivité.

1.1. Petit historique et conjecture de Lang

La conjecture de Manin-Mumford originelle porte sur l'intersection d'une courbe avec les points de torsion d'une variété abélienne. Elle a été tout d'abord démontrée par M. Raynaud ([Ra1]), puis R. Coleman ([Col]) en a donné une autre démonstration par des méthodes très différentes; Raynaud a ensuite ([Ra2]) démontré le résultat pour une sous-variété quelconque, puis M. Hindry ([Hi1]), en s'inspirant d'une idée de S. Lang ([La1]) pour le cas des courbes, a démontré le résultat général pour un groupe algébrique commutatif.

On trouve une autre approche de ces questions dans les travaux sur la conjecture de Bogomolov, on se contentera ici de renvoyer à l'exposé de A. Abbes sur le sujet dans le cadre de ce séminaire ([Ab]) et aux travaux récents de S. David et P. Philippon ([DaPh]).

En revanche, nous allons dire quelques mots de la conjecture de Lang (dite aussi conjecture de Mordell-Lang), énoncée par Lang dans [La1] et qui regroupe dans un cadre général la conjecture de Manin-Mumford et la conjecture de Mordell (voir aussi par exemple [La2]).

Nous l'énonçons ici pour la caractéristique zéro :

Conjecture de Lang (absolue). — *Soient K un corps algébriquement clos de caractéristique zéro, A une variété abélienne définie sur K, X une sous-variété de A définie sur K et Γ un sous-groupe de $A(K)$ de rang fini. Alors il existe $\gamma_1, \ldots, \gamma_n \in \Gamma$ et $B_1, \ldots, B_n$ sous-variétés abéliennes de A tels que $\gamma_i + B_i \subset X$ et*

$$X(K) \cap \Gamma = \cup_{i=1}^n \gamma_i + (B_i(K) \cap \Gamma).$$

Cette conjecture a finalement été démontrée par G. Faltings ([Fa]) après de nombreux travaux entre autres de Manin, Vojta, Laurent, Hindry et Faltings lui-même. On peut trouver des bibliographies récentes, commentées et complètes, dans [Hi2] ou [Maz]. L'extension au cas des variétés semi-abéliennes est due à Vojta([Voj]) et McQuillan ([McQ]).

Si on veut étendre cette conjecture au cas de caractéristique non nulle, il faut se contenter d'une version relative, dite aussi conjecture de Lang pour les corps de fonctions. En voici un énoncé un peu simplifié :

Conjecture de Lang pour les corps de fonctions. — *Soient $K_0 < K$ deux corps algébriquement clos, A une variété abélienne définie sur K de K/K_0-trace zéro, et X une sous-variété de A définie sur K. Soit Γ un sous-groupe de rang fini de $A(K)$. Alors il existe $\gamma_1, \dots, \gamma_n \in \Gamma$ et $B_1, \dots, B_n$ sous-variétés abéliennes de A tels que $\gamma_i + B_i \subseteq X$ et*

$$X(K) \cap \Gamma = \cup_{i=1}^{n} \gamma_i + (B_i(K) \cap \Gamma).$$

On rappelle que Γ est de *rang fini* si il existe un groupe Γ_0 finiment engendré tel que, pour chaque $\gamma \in \Gamma$, il y a un entier $n \geq 1$ tel que $n\gamma \in \Gamma_0$. Si on est en caractéristique $p > 0$, on demande que l'entier n ne soit pas divisible par p. Dire que A est de K/K_0-trace zéro est équivalent à dire que A n'a aucune sous-variété abélienne homomorphe à une variété abélienne définie sur K_0.

Dans [Hr1], Hrushovski a donné une démonstration modèle-théorique de la conjecture de Lang pour les corps de fonctions en toute caractéristique (on peut trouver une présentation de ces résultats dans l'exposé [Go] de ce séminaire, ou dans [Pi3], et une exposition plus détaillée dans le livre [Bo]). Pour la caractéristique $p > 0$, il s'agissait de la première démonstration du cas général, les résultats précédents (voir [AbrVo]) nécessitant des hypothèses supplémentaires. Un peu auparavant, A. Buium ([Bu]) avait donné une nouvelle démonstration du cas de caractéristique zéro, dont l'idée de départ était de passer à un corps différentiellement clos et de remplacer le groupe Γ par un groupe Δ-fermé (ou différentiellement algébrique) de dimension finie. C'est cette idée qui est l'inspiration des démonstrations de Hrushovski, pour les deux résultats, Lang pour les corps de fonctions et Manin-Mumford.

1.2. Le rapport avec la théorie des modèles

L'idée qui est à la base de ces deux démonstrations est donc la suivante : on veut remplacer le sous-groupe Γ ($Tor(G)$ dans notre cas) par un sous-groupe « fermé » ou « algébrique », pour lequel on montrera le résultat. On ne peut pas le faire sans, d'une manière ou d'une autre, rajouter des « fermés » supplémentaires (c'est-à-dire de la structure supplémentaire), puisque, quand G par exemple est une variété abélienne, la torsion est dense dans G. C'est exactement ce que fait Buium dans [Bu], quand il passe dans un corps différentiellement clos, remplace le groupe Γ par un groupe Δ-fermé de dimension finie et utilise ensuite l'arsenal des outils développés dans ce cadre. Or c'est là une méthode qui rentre tout à fait dans la problématique développée en théorie des modèles et pour laquelle nous disposons d'un cadre systématique, d'outils et de résultats déjà existants. En effet, puisant son inspiration dans la géométrie algébrique et les géométries combinatoires, la théorie des modèles développe dans un cadre abstrait des notions d'indépendance, de dimension et de géométrie. Cela a permis d'isoler une notion qui va se retrouver au centre de toutes ces applications récentes : la notion de *modularité* qui caractérise, à partir du comportement de la relation d'indépendance, les groupes dans lesquels les seuls sous-ensembles qu'on peut définir sont les (combinaisons booléennes finies de) translatés de sous-groupes. Montrer que le groupe $Tor(G)$ satisfait l'énoncé de Manin-Mumford, c'est exactement, une fois qu'on s'est placé dans le bon cadre, montrer que c'est un groupe modulaire, nous le verrons dans les sections 2.2 et 2.3 ; et l'on a des outils pour cela.

1.3. Le cadre dans lequel on travaille

On a donc un groupe algébrique, G, défini sur un corps de nombres k. On va considérer un automorphisme σ dans $Gal(k^{alg}/k)$, judicieusement choisi, de manière à ce que la torsion soit contenue dans le sous-groupe S_σ des solutions, dans $G(k^{alg})$, d'une équation faisant intervenir σ (ce qu'on appelle traditionnellement une équation de différence) qu'on peut explicitement calculer. On plonge alors (k^{alg}, σ) dans un gros corps K sur lequel σ se prolonge et qui est clos en ce sens qu'il contient déjà des solutions pour toute équation de différence qui a une solution dans une extension. C'est ce qu'on appellera un corps de différence générique. Et, de même qu'on a la topologie de Zariski sur un corps algébriquement clos ou la Δ-topologie sur un corps différentiellement clos, on définit une σ-topologie sur K. La théorie des modèles permet d'analyser la structure de K avec ces nouveaux ensembles géométriques. Cette analyse a été faite par Z. Chatzidakis et E. Hrushovski dans [ChHr] qui montrent qu'on peut y définir une bonne notion d'indépendance et de dimension. En particulier, ils démontrent en caractéristique zéro un théorème de dichotomie, qui dit à peu près que les seuls ensembles de dimension finie qui ne sont pas « modulaires » se trouvent dans le corps fixé par σ. À partir de cette analyse et de ce théorème de dichotomie, Hrushovski trouve un critère « effectif » pour distinguer, parmi les sous-groupes de $G(K)$ définis par des équations de différence, ceux qui sont modulaires. Il ne reste plus qu'à vérifier que le sous-groupe S_σ judicieusement choisi plus haut est du bon type, puis à borner le nombre de translatés qui vont intervenir par des calculs simples à partir de l'équation qui définit S_σ.

C'est exactement comme cela que cela se passe dans le cas de la torsion première à p. Pour la torsion totale, c'est un peu plus compliqué, car, s'il est facile, en travaillant à partir de deux premiers de bonne réduction distincts, de montrer qu'il existe une σ-équation qui s'annule sur la torsion toute entière et est du bon type, on ne sait pas le faire de manière effective. Donc, si on veut garder des bornes sur le nombre de translatés, il faut travailler un peu plus, en gardant deux équations distinctes, l'une pour la torsion première à p, l'autre pour la p-torsion puis combiner les deux.

Dans une première partie, nous allons présenter le cadre modèle théorique, c'est-à-dire les corps de différence génériques et les groupes que l'on peut y définir avec des équations de différence (section 2). Ensuite nous expliquerons comment utiliser les résultats ainsi obtenus pour démontrer la conjecture de Manin-Mumford (section

3). En fait, nous traiterons le cas de la torsion première à p et ne dirons que quelques mots sur le passage à la torsion totale. Enfin nous terminerons en mentionnant d'autres applications de la même méthode (section 4).

Merci à celles et ceux qui ont bien voulu relire tout ou partie de ce texte, ou répondre aux nombreuses questions auxquelles je ne pouvais manquer d'être confrontée, puisqu'il me faut parler ici non seulement de théorie des modèles mais aussi de géométrie algébrique, domaine qui n'est pas le mien. En particulier, merci à Z. Chatzidakis, F. Delon, D. Bertrand, M. Hindry et aussi à E. Hrushovski. Je suis évidemment seule responsable des erreurs qui se seraient ici glissées dans un énoncé de théorème ou dans une esquisse de preuve, à la suite d'une tentative indue de simplification.

2. LES CORPS DE DIFFÉRENCE GÉNÉRIQUES

2.1. Définition et existence

Nous appellerons *corps de différence* un corps K muni d'un automorphisme distingué σ. On peut par exemple considérer $K = k^{alg}$, la clôture algébrique d'un corps k, muni d'un automorphisme $\sigma \in Gal(k^{alg}/k)$ ou encore, si K est un corps parfait de caractéristique $p > 0$ et $n \geq 1$, K muni de l'automorphisme de Frobenius $x \mapsto x^{p^n}$.

Les corps de différence ont été à l'origine étudiés par Ritt dans les années 30 ; on peut trouver tous les résultats algébriques de base dans le livre de R. Cohn [Coh] (dans ce livre, un corps de différence est un corps avec un monomorphisme distingué ; quand il s'agit d'un automorphisme, le corps est appelé corps de différence inversif).

Il était naturel pour des théoriciens des modèles de s'intéresser à la classe des corps de différence *existentiellement clos* c'est-à-dire tels que tout système fini d'équations de différence ayant une solution dans une extension du corps en a déjà une dans le corps. C'est l'analogue, pour les corps de différence, des corps algébriquement clos pour les corps, des corps différentiellement clos pour les corps différentiels, ou bien encore des corps réels clos pour les corps ordonnés. Les premières propriétés (axiomatisation, décidabilité etc) ont été étudiées par A. Macintyre, L. van den Dries et C. Wood (voir [Ma1]). Ensuite, une étude plus complète de la structure de ces corps du point de vue de la théorie des modèles a été faite par Z. Chatzidakis et E. Hrushovski dans un premier temps ([ChHr]), puis par les mêmes avec un troisième auteur, K. Peterzil ([ChHrPe]). C'est dans le premier de ces deux articles que se trouvent les résultats qui sont utilisés pour la démonstration de Manin-Mumford, notamment le résultat de dichotomie en caractéristique zéro qui est crucial. Enfin, dans l'article où il donne sa démonstration de la conjecture de Manin-Mumford ([Hr2]), Hrushovski commence par pousser plus loin l'étude des groupes commutatifs définissables dans ces corps de différence. Dès la prochaine section, pour simplifier les énoncés, nous nous limiterons à la caractéristique zéro et aux résultats ayant un rapport direct avec les applications dont nous voulons parler. Pour en savoir plus sans se plonger dans les articles eux-mêmes, on pourra consulter [Ch2].

Commençons par décrire ces corps, que nous appellerons ici pour simplifier *corps de différence génériques* (au départ, dans [ChHr], ils apparaissent comme modèles d'une théorie appelée ACFA, « algebraically closed fields with an automorphism » ; on les appelle aussi souvent corps algébriquement clos avec un automorphisme générique).

Convention et notation. — Nous utiliserons le terme de variété uniquement dans le cas irréductible. Si X est une variété affine définie sur K, X^σ est la variété affine définie en appliquant σ aux coefficients des équations qui définissent X.

DÉFINITION 2.1. — *Soit* (K, σ) *un corps de différence. On dit que* (K, σ) *est un* corps de différence générique *si* K *est un corps algébriquement clos vérifiant la propriété :*

(*) *Soient* X *une variété affine définie sur* K *et* Y *une sous-variété de* $X \times X^\sigma$ *définie sur* K, *dont les projections sur les facteurs* X *et* X^σ *sont denses. Alors il existe* $a \in X(K)$ *tel que* $(a, \sigma(a)) \in Y(K)$.

Les corps algébriquement clos satisfaisant (*) sont bien exactement les corps de différence existentiellement clos et tout corps de différence se plonge dans un corps de différence générique.

Ces corps, comme on va le voir dans la suite, sont munis d'une structure plus riche que celle induite uniquement par la topologie de Zariski mais que l'on sait quand même analyser !

Commençons par quelques propriétés élémentaires :

Premières propriétés. — Soit (K, σ) un corps de différence générique. Le corps K est de degré de transcendance infini sur le corps premier. Le corps fixé par σ dans K, $Fix(\sigma)$, est un corps pseudo-fini. Le seul invariant nécessaire pour obtenir les complétions (au sens de la théorie des modèles) de la théorie des corps algébriquement clos est la caractéristique ; pour les corps de différence génériques, il faut préciser en plus la classe de conjugaison de σ dans le groupe des automorphismes de la clôture algébrique du corps premier.

Enfin, mentionnons un résultat frappant qui vient confirmer l'importance de cette classe de corps. Cela faisait plusieurs années que se posait la question de ce que pouvait bien être la théorie d'un « Frobenius non-standard », par exemple la limite des corps de différence $(\mathbb{F}_p^{alg}, x \mapsto x^p)$, pour p tendant vers l'infini. Il s'agit de l'analogue, pour les automorphismes de Frobenius, de la question résolue par Ax quand il a montré que la théorie élémentaire des corps finis était exactement la théorie des corps pseudo-finis [Ax]. La question précise ici était de savoir s'il s'agissait des corps de différence génériques. Une réponse positive a été indépendamment donnée par Hrushovski ([Hr3]) et Macintyre ([Ma2]) : la théorie des corps de différence génériques est exactement la théorie de tous les ultraproduits non principaux de $({\mathbb{F}_q}^{alg}, x \mapsto x^q)$, quand q varie sur l'ensemble des puissances de nombre premiers.

2.2. Ensembles définissables, indépendance

À partir de maintenant, (K, σ) *sera un corps de différence générique de caractéristique zéro.* Les résultats des sections 2.2 et 2.3 ont été démontrés dans [ChHr].

Comme d'habitude en Théorie des modèles (voir par exemple le chapitre d'introduction de [Bo]), on considère un corps de différence générique (K, σ) comme une structure du premier ordre. Cela veut dire que l'on étudie les ensembles que l'on peut définir à partir des opérations et des fonctions de base, dans ce cas $\{+, -, \cdot, 0, 1, \sigma\}$, en prenant la clôture par intersection et réunion finies, par complémentaire et par projection.

Dans un « pur » corps algébriquement clos K, il est naturel de commencer par considérer les sous-ensembles de K^n définis par des équations polynomiales, les fermés de Zariski. Ici il est naturel de considérer les σ- polynômes : $K[X_1, \ldots, X_n]_\sigma$, *l'anneau des polynômes de différence* (ou σ- polynômes) est l'anneau des polynômes sur K en une infinité de variables, $X_1, \ldots, X_n$, $\sigma(X_1), \ldots, \sigma(X_n)$, $\sigma^2(X_1), \ldots, \sigma^2(X_n)$, ... Un σ*-fermé* de K^n est l'ensemble des zéros d'un nombre fini de σ- polynômes de $K[X_1, \ldots, X_n]_\sigma$. Un σ*-idéal* I de $K[X_1, \ldots, X_n]_\sigma$ est un idéal clos par σ et on dit que I est *parfait* si chaque fois que $a^j \sigma^i(a) \in I$, $i, j \in \mathbb{N}$, alors a appartient à I. Les σ-fermés correspondent exactement aux σ-idéaux parfaits et comme on a la condition de chaîne ascendante sur les σ-idéaux parfaits (voir [Coh]), on définit ainsi une topologie noetherienne contenant strictement la topologie de Zariski.

On peut alors considérer les σ*-constructibles*, les combinaisons booléennes finies de σ-fermés. Mais là, contrairement à ce qui se passe avec la topologie de Zariski sur un corps algébriquement clos ou bien avec la Δ-topologie sur un corps différentiellement clos, la classe des σ-constructibles n'est pas close par projection. On clôt donc aussi par projection et on obtient ce qu'on appelle la classe des ensembles σ- *définissables.* Si V est une variété affine définie sur K, l'ensemble $V(K)$ des points K-rationnels de V est un σ-fermé dans un produit K^n. Pour une variété V définie sur K et donnée par un recouvrement fini de cartes affines, $V(K)$ est un ensemble σ-définissable. Le corps $Fix(\sigma)$ (défini par l'équation $\sigma(x) = x$) est un exemple de σ-fermé qui n'est pas de la forme $V(K)$. Nous dirons qu'une application est σ-définissable si son graphe est σ-définissable.

Bien que cela ne soit pas évident a priori, la classe des σ-définissables est également close, à bijection σ- définissable près, par quotient. Cela veut dire précisément que si $F \subseteq K^n$ est σ-définissable et que $R \subseteq K^n \times K^n$ est une relation d'équivalence σ-définissable sur K^n, alors il existe une bijection σ-définissable entre l'ensemble quotient F/R et un sous-ensemble σ-définissable de K^m pour un certain m. On dit des structures qui ont cette propriété qu'elles *éliminent les imaginaires.*

On sera amené à s'intéresser à la *structure induite* par (K, σ) sur les sous-ensembles σ- définissables : si $D \subset K^n$ est σ-définissable, la structure induite sur D est D, muni, pour tout $m > 0$, de la trace sur D^m de tous les sous-ensembles σ-définissables de K^{nm}.

On montre que la structure induite sur le corps $Fix(\sigma)$ se réduit (une fois qu'on a fixé un certain nombre de constantes dans $Fix(\sigma)$) à la pure structure de corps : si D est un sous-ensemble σ-définissable de K^n, alors $D \cap Fix(\sigma)^n$ est un sous-ensemble de $Fix(\sigma)^n$ définissable dans le pur langage des corps (c'est-à-dire constructible au sens habituel du terme, combinaison booléenne de fermés de Zariski).

Il est classique maintenant en théorie des modèles de regarder s'il est possible de définir une bonne notion d'indépendance dans les structures qu'on étudie, du type de l'indépendance algébrique dans les corps algébriquement clos. C'est bien le cas ici :

DÉFINITION 2.2. — *1. Si A est un sous-ensemble de K, la* σ-clôture algébrique *de A, $acl_\sigma(A)$, est égale à la clôture algébrique (au sens habituel des corps) du sous-corps de différence de (K, σ) engendré par A, c'est-à-dire à la clôture algébrique de $\mathbb{Q}(\{\sigma^i(a); a \in A, i \in \mathbb{Z}\})$.*

2. Si $A, B, C \subset K$, on dit que A et B sont indépendants *au-dessus de C si $acl_\sigma(AC)$ et $acl_\sigma(BC)$ sont linéairement disjoints au-dessus de $acl_\sigma(C)$.*

3. Si E est un sous-corps de différence de K et si a est une suite finie d'éléments de K, on définit le σ-degré *de a au-dessus de E, $d_\sigma(a/E)$ comme étant le degré de transcendance de $E(a)_\sigma$, le corps de différence engendré par $E(a)$, au-dessus de E. Si $D \subset K^n$ est un sous-ensemble σ-définissable, on définit le σ-degré de D comme étant le maximum des degrés des éléments de D.*

Le corps $Fix(\sigma)$ est de σ-degré égal à un ; en revanche si V est une variété (de dimension positive) définie sur K, $V(K)$ est de σ-degré infini. Quand il est fini, le σ-degré est une bonne notion de dimension, en particulier il est clair que si $E = acl_\sigma(E)$ et si $d_\sigma(a/E)$ est fini, alors a et b sont σ-indépendants au-dessus de E si et seulement si $d_\sigma(a/E) = d_\sigma(a/acl_\sigma(E(b))$. Il est possible (et indispensable) de définir d'autres notions de dimension ou de rang dans (K, σ) pouvant prendre des valeurs ordinales non finies, mais nous n'en parlerons pas ici. Ce qui est important pour nous, c'est que pour un a donné, ces différentes dimensions sont simultanément finies.

2.3. Modularité, le théorème de dichotomie

Comme nous l'avons dit dans l'introduction, cela fait plusieurs années que la pertinence de la notion abstraite d'ensemble « modulaire » (cette notion apparaît hélas dans la littérature sous plusieurs noms : localement modulaire, un-basé (one-based) etc.) s'est imposée à travers les conclusions très précises que l'on peut déduire sur les propriétés algébriques des structures modulaires.

Nous allons donner les définitions et les résultats sous une forme adaptée à notre contexte précis. Mais il s'agit de cas particuliers ou d'adaptations de résultats valables dans un contexte beaucoup plus général.

Si $E = acl_\sigma(E) \subset K$ et si a, a' sont deux éléments de K^n, on dit que a et a' *ont même type (d'isomorphisme)* au-dessus de E si il existe un isomorphisme ϕ : $acl_\sigma(E,a)) \mapsto acl_\sigma(E,a')$ qui fixe E, envoie a sur a' et commute avec σ.

Définition 2.3. — *Un sous-ensemble σ-définissable D de K^n est*

– stable *si, pour tout $E = acl_\sigma(E) \subset K$, pour toutes suites finies a, a' d'éléments de D ayant même type au-dessus de E, pour tout $F = acl_\sigma(F), E \subset F \subset K$, si a et F sont indépendants au-dessus de E et si a' et F sont indépendants au-dessus de E, alors a et a' ont aussi même type au-dessus de F ;*

– modulaire *si pour tout $E = acl_\sigma(E) \subset K$, pour toutes suites finies a, b d'éléments de D, $acl_\sigma(Ea)$ et $acl_\sigma(Eb)$ sont indépendants au-dessus de $acl_\sigma(Ea) \cap acl_\sigma(Eb)$.*

La stabilité veut donc dire que, si $E \subset K$ est σ-algébriquement clos, et si $a, b \in K$ alors il n'y a (à isomorphisme près) qu'une seule façon pour a et b d'être indépendants au-dessus de E.

Il y a des corps stables : les (purs) corps algébriquement clos, les corps séparablement clos ou encore les corps différentiellement clos. Mais les corps pseudo-finis ([Du]), donc en particulier $Fix(\sigma)$, ne sont pas stables. On le voit ici facilement : on peut trouver $E = acl_\sigma(E) \subset K$ et $a, b, c \in Fix(\sigma) \setminus E$, tels que a et c d'une part, b et c d'autre part sont indépendants au-dessus de E, mais tels que $\sqrt{a-c} \in Fix(\sigma)$ et $\sqrt{b-c} \notin Fix(\sigma)$.

Le corps $Fix(\sigma)$ n'est pas modulaire : on prend trois éléments de $Fix(\sigma)$, a, b, c transcendants sur $\mathbb{Q}$ et algébriquement indépendants. Alors $acl_\sigma(a,b) = \mathbb{Q}(a,b)^{alg}$ et $acl_\sigma(c, ac+b) = \mathbb{Q}(c, ac+b)^{alg}$ s'intersectent en $\mathbb{Q}^{alg}$, et pourtant ils ne sont pas algébriquement indépendants au-dessus de $\mathbb{Q}^{alg}$. En fait, la non-modularité est la traduction « abstraite », dans le cas d'un ensemble de dimension un, de l'existence d'une famille de courbes planes (ici la famille des courbes $y = ax + b$) de dimension deux. En particulier, si un ensemble D est modulaire, alors cela entraîne que *dans la structure induite par (K,σ) sur D*, on ne peut pas définir de corps infini. On va voir un peu plus loin (2.6) qu'avec l'hypothèse de stabilité, cela entraîne beaucoup plus.

Mais ce qui est particulièrement intéressant dans le cas des corps de différence génériques, c'est que toute la non-modularité et la non-stabilité sont concentrées dans le corps fixé. C'est ce que nous dit le théorème de dichotomie en caractéristique zéro.

On dit que deux ensembles σ-définissables D et F de (K,σ) sont *orthogonaux* si pour toute suite finie d d'éléments de D, pour toute suite finie b d'éléments de F, pour tout sous-corps $E = acl_\sigma(E)$ de K, d et b sont indépendantes au-dessus de E.

Théorème 2.4. — Théorème de dichotomie : *Soit $D \subseteq K^n$ un sous-ensemble σ-définissable de σ-degré fini, alors D est stable et modulaire si et seulement si D et le corps fixé $Fix(\sigma)$ sont orthogonaux.*

La démonstration de ce théorème passe par une analyse des ensembles de dimension finie en termes de sous-ensembles de dimension un. Le même type de résultat avait été montré précédemment pour les corps différentiellement clos (avec le sous-corps des constantes, [HrSo]) et pour les corps séparablement clos de caractéristique $p > 0$ (avec le sous-corps des éléments infiniment p-divisibles, [Hr1]) mais par des méthodes différentes. Dans ces deux autre cas, la preuve utilise la théorie des géométries de Zariski, version abstraite de la topologie de Zariski introduite par Hrushovski et Zil'ber ([HrZi], voir aussi [Mr]).

De façon générale, la dichotomie modulaire/non-modulaire est particulièrement utile dans le cas des groupes. En effet la modularité caractérise les structures de groupes stables de manière très précise. À nouveau nous énonçons dans 2.6 un résultat qui est vrai dans un contexte beaucoup plus général.

Définition 2.5. — *Soit $G \subseteq K^n$ un groupe σ-définissable. On dit que G est* de type abélien *si pour tout m et pour tout sous-ensemble σ-définissable X de K^{nm}, $X \cap G^m$ est une combinaison booléenne finie de translatés de sous-groupes connexes (σ-définissables) de G^m.*

Il s'ensuit que le groupe G a un sous-groupe (σ-définissable) abélien d'indice fini et donc en fait que la structure induite par (K, σ) sur G se réduit à une structure de type « module généralisé ».

Proposition 2.6 ([HrPi1]). — *Soit $G \subset K^n$ un groupe σ-définissable stable. Alors G est modulaire si et seulement si G est de type abélien*

On voit maintenant le rapport entre la modularité et les questions de type conjecture de Manin-Mumford ou plus généralement conjecture de Lang. On a d'ailleurs bien une équivalence formelle de la conjecture de Manin-Mumford avec l'énoncé de théorie des modèles suivant : pour tout groupe algébrique commutatif connexe G défini sur $\mathbb{Q}^{alg}$, la structure $(\mathbb{Q}^{alg}, Tor(G))$, dans le langage des anneaux avec un prédicat pour le groupe $Tor(G)$, est stable et le groupe $Tor(G)$ (qui est maintenant définissable dans cette structure) est modulaire (voir [Pi1]). Mais savoir cela ne nous donne pas pour autant une méthode pour montrer que cela est vrai. L'intérêt de passer par les corps de différence génériques c'est que là, on a un vrai critère utilisable pour reconnaître quand un groupe est modulaire grâce à la dichotomie.

2.4. Les sous-groupes σ-définissables des variétés abéliennes

Nous résumons ici les conséquences de l'étude des groupes commutatifs σ-définissables et en particulier des sous-groupes σ-définissables des points K-rationnels des variétés abéliennes faite par Hrushovski dans [Hr2]. On peut aussi en trouver une exposition détaillée dans [Ch1].

L'une des étapes essentielles de cette étude est l'analyse de l'anneau des *endomorphismes σ-définissables* de $A(K)$ pour A une variété abélienne définie sur K.

DÉFINITION 2.7. — *Soit G un groupe algébrique défini sur K. Un* endomorphisme σ-définissable *de $G(K)$ est une application σ-définissable de $G(K)$ dans $G(K)$ qui est un homomorphisme du groupe $G(K)$.*

Les endomorphismes du groupe algébrique G qui sont définis sur K induisent des endomorphismes de $G(K)$ au sens de la définition ci-dessus. Il y en a en général beaucoup d'autres mais Hrushovski montre par exemple :

PROPOSITION 2.8. — *Soient A une variété abélienne simple définie sur K, $End_\sigma(A(K))$ l'anneau des endomorphismes σ-définissables de $A(K)$ et $End_{alg}(A(K))$ le sous-anneau des endomorphismes qui sont induits par les endomorphismes (algébriques) de la variété abélienne A.*

(1) *Si A n'est pas isogène à A^{σ^n}, alors*

$$\mathbb{Q} \otimes End_\sigma(A(K)) = \mathbb{Q} \otimes End_{alg}(A(K)).$$

(2) *Sinon, $\mathbb{Q} \otimes End_\sigma(A(K))$ est isomorphe à un anneau de polynômes tordu au-dessus de $\mathbb{Q} \otimes End_{alg}(A(K))$ (que nous ne décrirons pas ici précisément). Il s'ensuit en particulier que $End_\sigma(A(K))$ est dénombrable, comme $End_{alg}(A(K))$.*

(3) *Pour chaque sous-groupe σ-définissable G de $A(K)$, il existe $f \in End_\sigma(A(K))$ tel que G est un sous-groupe d'indice fini de $Ker(f)$. Il s'ensuit qu'il n'y a qu'un nombre dénombrable de sous-groupes σ-définissables de $A(K)$.*

Des résultats similaires sont montrés pour $A = \mathbb{G}_m$, le groupe multiplicatif. On a alors $\mathbb{Q} \otimes End_\sigma(\mathbb{G}_m(K)) \simeq \mathbb{Q}[\sigma, \sigma^{-1}]$.

Soit A une variété semi-abélienne définie sur $Fix(\sigma)$ et $F[T] \in \mathbb{Z}[T]$ un polynôme à coefficients entiers. Alors $F(\sigma)$ induit naturellement un endomorphisme (σ-définissable) de $A(K)$: si $F[T] = \sum_{i=0}^{r} m_i T^i$, alors $F(\sigma)(a) = m_0 a + m_1 \sigma(a) + \cdots + m_r \sigma^r(a)$, où $+$ est l'addition dans A et $ma = [m]a$ la multiplication par l'entier m dans A.

DÉFINITION 2.9. — *Les groupes B et C sont* commensurables *si $B \cap C$ est d'indice fini dans B et dans C. Un groupe σ-définissable B est* c-minimal *si tout sous-groupe σ-définissable infini de B est d'indice fini dans B.*

PROPOSITION 2.10. — *Supposons que A est une variété abélienne simple définie sur $Fix(\sigma)$ ou bien que A est le groupe multiplicatif $\mathbb{G}_m$. Soit B un sous-groupe σ-définissable de $A(K)$ de σ-degré fini et c-minimal. Alors B n'est pas de type abélien si et seulement si il existe n tel que $B \subseteq Ker(\sigma^n - 1)$.*

Esquisse de démonstration de 2.10. — C'est bien sûr le Théorème de dichotomie (2.4) et l'équivalence (2.6) entre groupe stable modulaire et groupe de type abélien que nous allons utiliser.

Si B est inclus dans $Ker(\sigma^n - 1)$, c'est-à-dire $B \subseteq A(Fix(\sigma^n))$, alors comme $Fix(\sigma^n)$ est une extension finie de $Fix(\sigma)$, il existe une application σ-définissable à fibres finies de $Fix(\sigma)^n$ sur B, qui n'est donc pas orthogonal à $Fix(\sigma)$ et n'est donc pas de type abélien.

Réciproquement, supposons que B n'est pas de type abélien. Par 2.6 et 2.4, B et $Fix(\sigma)$ ne sont pas orthogonaux.

FAIT 2.11. — *Soit B un sous-groupe σ-définissable de σ-degré fini de $H(K)$, où H est un groupe algébrique défini sur K. Si B n'est pas orthogonal à $Fix(\sigma)$, alors il existe un sous-groupe normal D σ- définissable d'indice infini dans B, un entier $m \geq 1$ et une surjection σ-définissable (du produit cartésien) $(Fix(\sigma))^m$ sur B/D.*

Dans notre cas, la c-minimalité de B entraîne que D est fini. On en déduit l'existence d'une application injective σ-définissable h de B/D dans un produit de $Fix(\sigma)$. On transporte la loi de groupe de B/D par h sur $C = h(B/D)$. Le groupe C est alors σ-définissable, mais on sait que tout sous-ensemble σ-définissable dans $Fix(\sigma)$ est définissable dans le pur langage des corps. On peut donc utiliser des résultats antérieurs sur les groupes définissables dans les corps pseudo-finis ([HrPi2]) qui permettent de conclure qu'il existe un homomorphisme σ-définissable, de noyau fini, de C sur un sous-groupe d'indice fini de $G(Fix(\sigma))$, pour G un groupe algébrique défini sur $Fix(\sigma)$. On en déduit l'existence d'un homomorphisme σ-définissable g de B sur un sous-groupe d'indice fini de $G(Fix(\sigma))$, g de noyau fini contenant D. On peut supposer (en remplaçant B par un sous-groupe d'indice fini) que G est connexe, que $g(B) = G(Fix(\sigma))$ et donc que G est un groupe algébrique commutatif simple. Le graphe de g, qui est un sous-groupe σ-définissable de $(H \times G)(K)$ est commensurable avec le noyau d'un endomorphisme σ-définissable (par 2.8). Il n'y a qu'un nombre dénombrable de tels endomorphismes, et il suit que g est défini sur une extension finie k_1 de $Fix(\sigma)$. On peut donc bien trouver un n tel que σ^n fixe B point par point.

On peut maintenant en déduire le corollaire qui va être au centre de la démonstration de la conjecture de Manin-Mumford, en utilisant les lemmes suivants qui permettent en particulier de se ramener au cas d'une variété abélienne simple ou du groupe multiplicatif :

LEMME 2.12. — *a) Si on a une suite exacte d'homomorphismes de groupe σ-définissables*

$$0 \to A_1 \to A_2 \to A_3 \to 0,$$

où A_1, A_2, A_3 sont des groupes σ-définissables, alors A_2 est stable modulaire si et seulement si A_1 et A_2 sont stables modulaires.

b) Si A est une variété abélienne simple définie sur K et si H est un sous-groupe σ-définissable de $A(K)$, alors H est c-minimal si et seulement si il existe $f \in End_\sigma A(K)$, f irréductible dans $\mathbb{Q} \otimes End_\sigma A(K)$, tel que H et $Ker(f)$ sont commensurables.

c) Si B est un sous-groupe σ-définissable de $A(K)$, pour A une variété abélienne définie sur K, et si B est commensurable avec $Ker(f_1 \cdots f_m)$ pour $f_1, \ldots, f_m \in End_\sigma(A(K))$, alors B est stable modulaire si et seulement si tous les $Ker(f_i)$ sont stables modulaires.

Si $F[T] \in \mathbb{Z}[T]$ est un polynôme dont aucune racine n'est racine de l'unité, nous dirons que $F(T)$ est un polynôme sans facteur cyclotomique.

Corollaire 2.13. — *Soit A une variété semi-abélienne définie sur $Fix(\sigma)$ et $F(T) \in \mathbb{Z}[T]$. Soit $H_\sigma = Ker(F(\sigma)) = \{a \in A(K); \sum_{i=0}^{r} m_i\sigma^i(a) = 0\}$. Alors le groupe H_σ est de type abélien si et seulement si $F(T)$ est sans facteur cyclotomique.*

Esquisse de démonstration de 2.13. — Tout d'abord H_σ est bien de σ-degré fini. Ensuite, en utilisant 2.12, on peut se ramener aux deux cas où A est une variété abélienne simple ou bien où $A = \mathbb{G}_m$, le groupe multiplicatif.

Si pour un $N > 0$ $(T^N - 1)$ et $F(T)$ ont un facteur commun $P(T)$, alors $B = Ker(P(\sigma)) \subseteq Ker(\sigma^N - 1)$ n'est pas modulaire. Mais $B \subseteq H_\sigma$ qui n'est donc pas modulaire non plus.

Supposons maintenant que H_σ n'est pas de type abélien et, pour simplifier (en fait il faut utiliser b) et c) du Lemme 2.12), qu'il est c-minimal et que $F(\sigma) = f$ est un élément irréductible de $\mathbb{Q} \otimes End_\sigma(A)$. Par 2.10 à nouveau, il existe un N tel que $Ker(f) \subseteq C = Ker(\sigma^N - 1)$. Mais en choisissant N suffisamment grand on peut supposer que f agit sur C et $F(T)$ étant sans facteur cyclotomique, f devrait être inversible sur C.

3. APPLICATION À LA CONJECTURE DE MANIN-MUMFORD

Nous sommes maintenant en mesure d' expliquer comment appliquer le théorème de dichotomie et son corollaire sur les groupes définis à partir d'un polynôme sans facteur cyclotomique (2.13) pour obtenir une nouvelle démonstration de la conjecture de Manin-Mumford. En fait, comme nous l'avons déjà dit dans l'introduction, nous n'allons vraiment traiter précisément que le cas de la torsion première à p, cas dans lequel l'application de 2.13 donne immédiatement le résultat et qui permet de comprendre pourquoi ce type de preuve fournit naturellement des bornes.

3.1. Préliminaires sans théorie des modèles

Comme on l'a déjà expliqué, la théorie des modèles va nous permettre de montrer que les ensembles qui nous intéressent sont des combinaisons booléennes finies de translatés de sous-groupes et l'on voudra en conclure qu'en fait ce sont des réunions finies de translatés. Remarquons donc tout de suite une fois pour toutes l'équivalence des différentes versions que nous pourrons rencontrer :

LEMME 3.1. — *Soit G un groupe algébrique commutatif connexe défini sur un corps algébriquement clos L, X une sous-variété de G également définie sur L et Γ un sous-groupe de $G(L)$. Les énoncés suivants sont équivalents :*

(i) $X \cap \Gamma$ est une combinaison booléenne finie de translatés de sous-groupes de Γ,

(ii) la clôture de Zariski de $X \cap \Gamma$ est une réunion finie de translatés de sous-groupes algébriques connexes de G,

(iii) $X \cap \Gamma$ est contenu dans une réunion finie de translatés $C_1, \ldots, C_n$ de sous-groupes algébriques connexes de G, chaque C_i étant contenu dans X,

(iv) $X \cap \Gamma$ est une réunion finie de translatés de sous-groupes de Γ.

Maintenant voici le principal fait « algébrique » que l'on va utiliser et sur lesquels nous ne nous étendrons pas ici.

Soit A une variété abélienne de dimension d définie sur un corps de nombres k et $\mathfrak{p}$ un premier de l'anneau $\mathfrak{R}$ des entiers de k, de corps résiduel $k_{\mathfrak{p}}$ de cardinalité q et de caractéristique p. On suppose que $\mathfrak{p}$ est un *premier de bonne réduction pour A*. On a donc que $A_{\mathfrak{p}}$, la réduction de A modulo $\mathfrak{p}$, est une variété abélienne (définie sur $k_{\mathfrak{p}}$) de même dimension que A. On rappelle que $Tor_{p'}(A) = \{a \in Tor(A);\ p \text{ ne divise pas l'ordre de } a\}$ et que $Tor_p(A) = \{a \in Tor(A); p^n a = 0 \text{ pour un } n \geq 1\}$.

FAIT 3.2. — *Il existe un automorphisme $\sigma \in Gal(\mathbb{Q}^{alg}/k)$ et un polynôme à coefficients entiers, $F[T] \in \mathbb{Z}[T]$, sans facteur cyclotomique, tel que l'endomorphisme $F(\sigma)$ de $A(\mathbb{Q}^{alg})$ s'annule sur $Tor_{p'}(A)$. De plus le degré de $F(T)$ est inférieur ou égal à $2d$ et la somme des valeurs absolues de ses coefficients est bornée par $(1+q^{1/2})^{2d}$.*

Des résultats classiques de Weil (voir [We]) sur les endomorphismes des variétés abéliennes définies sur les corps finis et le polynôme caractéristique de l'automorphisme de Frobenius entraînent l'existence d'un tel polynôme $F[T] \in \mathbb{Z}[T]$ tel que $F(\Phi_q)$ s'annule sur $A_{\mathfrak{p}}(k^{alg})$, où Φ_q est l'automorphisme de Frobenius $\Phi_q : x \mapsto x^q$.

On peut alors relever cette équation fonctionnelle de la façon suivante : on prend pour σ un relèvement de Φ_q et, $\mathfrak{p}$ étant un premier de bonne réduction, on déduit du lemme de Hensel que la réduction modulo $\mathfrak{p}$ induit un isomorphisme de $Tor_{p'}(A)$ sur $Tor_{p'}(A_{\mathfrak{p}})$ et donc que $F(\sigma)$ s'annule sur $Tor_{p'}(A)$.

Remarque. — Il est possible de faire un peu mieux en travaillant à partir des facteurs simples de A et en ne comptant qu'une seule fois les facteurs isogènes. On peut ainsi borner le degré du polynôme $F[T]$ et la valeur absolue de ses coefficients en remplaçant d, la dimension de A, par un invariant de A, $d_r(A)$ qui est inférieur ou égal à la dimension de A et tel que, pour tout $n \geq 1$, $d_r(A) = d_r(A^n)$.

3.2. La torsion première à p

Nous allons donner la démonstration dans le cas des variétés abéliennes. Le cas des variétés semi-abéliennes est identique, il suffit de vérifier qu'on peut généraliser sans peine l'existence du polynôme adéquat sans facteur cyclotomique (3.2). En revanche, il faut encore travailler et utiliser de la théorie des modèles pour passer au cas d'un groupe commutatif arbitraire, nous l'expliquerons dans la section 3.2.2.

3.2.1. Les variétés abéliennes. — Soient A une variété abélienne de dimension d définie sur un corps de nombres k et X une sous-variété de A. On fixe un plongement de A dans un espace projectif $\mathbb{P}_n$ et on peut ainsi considérer le degré de A et de ses sous-ensembles algébriques.

On fixe un premier $\mathfrak{p}$ de bonne réduction pour A dont le corps résiduel est de cardinalité q et de caractéristique p.

PROPOSITION 3.3. — *Alors*

$$X \cap Tor_{p'}(A) = \bigcup_{i=1}^{M} a_i + Tor_{p'}(B_i),$$

où chaque B_i est une sous-variété abélienne de A et

$$M \leq c\ (\deg(X)^{(2d+1)(2^{d\dim(X)})})$$

où la constante c ne dépend que d'invariants liés à A (et non à X et à son corps de définition) et de q, et est deux fois exponentielle en d.

Démonstration. —

1. Pour obtenir la réunion finie. — On considère l'automorphisme σ de $Gal(\mathbb{Q}^{alg}/k)$ et le polynôme $F(T) \in \mathbb{Z}[T]$, $F(T) = \sum_{i=0}^{2d} c_i T^i$, donnés par 3.2. Soit (K, σ) un corps de différence générique extension du corps de différence $(\mathbb{Q}^{alg}, \sigma)$. Alors

$$H_\sigma := Ker(F(\sigma) \restriction_{A(K)}) = \{a \in A(K); \sum_{i=0}^{2d} c_i \sigma^i(a) = 0\}$$

est un sous-groupe de $A(K)$ défini par une σ-équation.

Puisque le polynôme $F(T)$ est sans facteur cyclotomique, 2.13 nous assure que H_σ est un sous-groupe de type abélien et donc que tous ses sous-ensembles σ-définissables sont des combinaisons booléennes finies de translatés de sous-groupes σ-définissables. C'est donc le cas en particulier pour $X \cap H_\sigma$ qui lui-même contient $X \cap Tor_{p'}(A)$.

Comme nous l'avons remarqué plus haut, cela entraîne que la clôture de Zariski de $X \cap H_\sigma$, que nous appellerons Z, est réunion finie de translatées de sous-variétés abéliennes de A, chacune contenue dans X, $Z = \bigcup_{i=1}^{M} b_i + B_i$.

On en déduit que $X \cap Tor_{p'}(A)$ est réunion d'au plus M translatés de la forme $a + Tor_{p'}(B)$ avec B sous-variété abélienne de A.

2. Pour borner le nombre M de translatées. — On va en fait borner le nombre de composantes irréductibles de Z, la clôture de Zariski de $X \cap H_\sigma$, en utilisant la définition explicite de H_σ à partir du polynôme $F[T]$.

On considère

$$S = \{(a_0, \dots, a_{2d}) \in A^{2d+1}; \sum_{i=0}^{2d} c_i a_i = 0\}.$$

Donc $H_\sigma = \{a \in A(K); (a, \sigma(a), \dots, \sigma^{2d}(a)) \in S\}$. Soit $U = S \cap (X \times X^\sigma \times \dots \times X^{\sigma^{2d}})$ (chaque X^{σ^i} est une sous-variété de $A^{\sigma^i} = A$), alors

$$X \cap H_\sigma = \{a \in A(L); (a, \sigma(a), \dots, \sigma^{2d}(a)) \in U\}.$$

Pour faire le calcul, on va utiliser une définition un peu modifiée du degré pour les sous-variétés de $(\mathbb{P}_n)^m$, pour $m > 1$, voir [Fu], exemple 8.4.4. On définit aussi le degré d'un fermé non irréductible comme la somme des degrés de (toutes) ses composantes irréductibles. Ce qui est important pour nous ici, c'est que le degré borne bien le nombre de composantes irréductibles.

Les propriétés du degré entraînent que

$$\deg(U) \leq \deg(S)(\deg(X)^{2d+1}) \leq \deg(S)(\deg(A)^{2d+1})$$

et on a que $\dim(U) \leq min\{\dim(S), (2d+1)\dim(X)\} \leq d(2d+1)$.

En utilisant le fait que le corps (K, σ) est un corps de différence générique, et donc satisfait la condition (*) de 2.1, on montre que :

Lemme 3.4. — *Soit $r > 0$, E un sous-ensemble algébrique de $\mathbb{P}_n^{r+1}$,*

$$E_\sigma := \{a \in \mathbb{P}_n(K); (a, \sigma(a), \dots, \sigma^r(a)) \in E\}$$

et V la clôture de Zariski de E_σ. Alors $\deg(V) \leq (\deg(E))^{2^{\dim(E)}}$.

Dans notre cas on obtient donc que $M \leq \deg(Z) \leq \deg(U)^{2^{\dim(U)}}$, ce qui donne bien une borne du type annoncé, une fois qu'on a calculé le degré de S.

3.2.2. Groupes algébriques commutatifs. — Il n'est pas très difficile de généraliser à tous les groupes algébriques commutatifs le Fait 3.2 et donc l'existence du polynôme $F(T)$ sans facteur cyclotomique et de l'automorphisme σ de $\mathbb{Q}^{alg}$ adéquats. On considère comme plus haut un corps de différence générique (K,σ) étendant $(\mathbb{Q}^{alg},\sigma)$.

La généralisation de la Proposition 3.3 aux variétés semi-abéliennes est alors immédiate puisque le corollaire central de la dichotomie (2.13) est vrai pour les variétés semi-abéliennes.

Mais dans le cas d'un groupe commutatif arbitraire 2.13 n'est plus vrai. En effet, si V est un sous-groupe vectoriel du groupe algébrique G, alors $KerF(\sigma)\cap V(K)$ ne peut pas être de type abélien : tout sous-groupe σ-définissable de degré fini de $V(K)$ est un espace vectoriel σ-définissable de dimension finie sur le corps $Fix(\sigma)$ et ne lui est donc pas orthogonal.

Mais on peut séparer la partie de type abélien et la partie espace vectoriel sur $Fix(\sigma)$ et cette dernière n'intervient pas dans le cas de la torsion. Plus précisément :

Définition 3.5. — *Soient G un groupe algébrique commutatif connexe défini sur $Fix(\sigma)$ et V le sous-groupe vectoriel maximal de G. Un sous-ensemble σ-définissable D de $G(K)$ est* spécial *s'il est de la forme $D = C + W$, avec C un translaté d'un sous-groupe σ-définissable de $G(K)$ et W un sous-ensemble σ-définissable de $V(K)$.*

Proposition 3.6. — *Soient G un groupe algébrique commutatif connexe défini sur $Fix(\sigma)$ et $H_\sigma = \{g \in G(K); F(\sigma)(g) = 0\}$ pour un polynôme sans facteur cyclotomique $F(T) \in \mathbb{Z}[T]$. Alors tout sous-ensemble σ-définissable de H_σ est une combinaison booléenne finie de sous-ensembles spéciaux de $G(K)$.*

Les ingrédients de la démonstration de 3.6. — On considère la suite exacte $0 \to V \to G \to B \to 0$, où V est le sous-groupe vectoriel maximal de G et B est donc une variété semi-abélienne. Maintenant par 2.13, $H_B := H_\sigma/V(K)\cap H_\sigma$ est de type abélien mais $H_V := V(K)\cap H_\sigma$ est, lui, un espace vectoriel sur $Fix(\sigma)$. Le théorème de dichotomie (2.4) entraîne que, dans la structure (K,σ), un groupe de type abélien de σ-degré fini et un groupe vectoriel σ- définissable ne peuvent qu'être orthogonaux. Il n'y a donc aucune relation possible entre éléments de H_B et de H_V. On peut alors montrer que tout sous-ensemble σ-définissable de $H_B \times H_V$ est une réunion finie de rectangles $B_i \times V_i$ avec $B_i \subseteq H_B$ et $V_i \subseteq H_V$ et dans notre situation particulière, en déduire que tout sous-ensemble σ-définissable de H_σ est une combinaison booléenne finie de sous-ensembles σ-définissables spéciaux de la forme : $B + W$, où B est un translaté d'un sous-groupe σ-définissable de $G(K)$ et W est un sous-ensemble σ-définissable de $V(K)$.

Corollaire 3.7. — *Soient G un groupe algébrique commutatif connexe défini sur $Fix(\sigma)$ et X une sous-variété de G. Alors la clôture de Zariski de $X \cap T_{p'}(G)$ est une réunion finie de translatés de sous-groupes algébriques de G.*

Démonstration de 3.7. — On considère $H_\sigma = \{g \in G(K); F(\sigma)(g) = 0\}$ pour le polynôme $F(T) \in \mathbb{Z}[T]$ donné par 3.2 (généralisé aux groupes commutatifs arbitraires). Par 3.6 $X \cap H_\sigma$ est combinaison booléenne de sous-ensembles spéciaux. Par passage à la clôture de Zariski, on obtient une réunion finie de sous-variétés spéciales de G, c'est-à-dire de sous-variétés de la forme $B + W$ où B est un translaté d'un sous-groupe algébrique connexe de G et W est une sous-variété de V, le sous-groupe vectoriel maximal de G. Il suffit maintenant de montrer que, si $C = B + W$ est une sous-variété spéciale de G d'intersection non vide avec $Tor_{p'}(G)$, alors la clôture de Zariski de $Tor_{p'}(G) \cap C$ est un translaté d'un sous-groupe algébrique de G. Par hypothèse, $B = g + E$ où E est un sous-groupe algébrique connexe de G. On peut se ramener au cas où $E \cap V = 0$ et on vérifie alors que $Tor_{p'}(G) \cap C$ est un translaté de $Tor_{p'}(E)$.

3.3. La torsion totale

3.3.1. Le cas des variétés semi-abéliennes. — On a vu dans la section précédente comment, pour une variété semi-abélienne, utiliser un premier de bonne réduction $\mathfrak{p}$ et puis un corps de différence générique $(L_\mathfrak{p}, \sigma_\mathfrak{p})$ pour « enfermer » le groupe des éléments de torsion première à p dans un groupe de type abélien en obtenant des bornes raisonnables. Puisque $Tor(A) = Tor_{p'}(A) \oplus Tor_p(A)$, l'idée est d'utiliser un second premier de bonne réduction $\mathfrak{t}$, de caractéristique résiduelle t différente de p, pour « enfermer » la torsion première à t et donc en particulier la p-torsion. C'est effectivement ce que fait Hrushovski, mais le problème va être de garder des bornes effectives. Si on cherche seulement le résultat qualitatif, il y a plusieurs preuves rapides utilisant cette méthode que l'on peut trouver par exemple dans [Pi2], [Ch2] ou [Hr2]. Nous allons en indiquer une dans la section suivante où l'on se trouve confronté à la question du calcul de $[L : k]$ où $L = k(Tor_{p'}(A)) \cap k(Tor_p(A))$.

Pour maîtriser les bornes, Hrushovski met en évidence des familles algébriques de composantes irréductibles, pour lesquelles il faut montrer une version uniforme de 3.4. Nous n'allons pas le faire ici. On peut trouver des énoncés et des calculs précis dans [Hr2] bien sûr mais aussi dans [Ch2]. On peut également en trouver une version « uniforme mais sans les calculs de bornes » dans [Pi2].

Il est important de signaler que, même si c'est moins frappant dans ce cas que dans le cas de la torsion première à p, l'existence de ces bornes effectives découle encore naturellement de la nature même de la démonstration. L'ordre de grandeur indiqué est ici aussi obtenu en effectuant les calculs avec des méthodes assez grossières.

3.3.2. Le cas des groupes commutatifs. — Pour ce qui est du passage à un groupe commutatif arbitraire, on commence par remarquer que, une fois le résultat « qualitatif » connu, pour ce qui concerne le calcul des bornes, le cas général se réduit au cas des variétés semi-abéliennes : soient G un groupe algébrique commutatif défini sur un corps de nombres k et X une sous-variété de G. Soient V le sous-groupe vectoriel maximal de V, $A := G/V$ la variété semi-abélienne correspondante et Y l'image de

X dans A. Alors le nombre de composantes irréductibles de la clôture de Zariski de $X \cap Tor(G)$ est borné par le nombre de composantes irréductibles de la clôture de Zariski de $Y \cap Tor(A)$. En effet, la projection de Y sur A est injective sur la torsion de G et induit donc une bijection entre les composantes irréductibles de la clôture de Zariski de $X \cap Tor(G)$ (dont on sait que ce sont des translatés de sous-groupes) et celles de $Y \cap Tor(A)$, (dont on sait aussi que ce sont des translatés de sous-groupes).

Il suffit donc de montrer, sans avoir à se préoccuper des bornes, que pour un groupe commutatif arbitraire il est bien vrai que l'intersection de X avec la torsion est réunion finie de translatés de sous-groupes algébriques. Comme nous l'avons dit plus haut, il y a plusieurs démonstrations possibles utilisant deux premiers de bonne réduction distincts.

Nous allons en donner une ici qui, en admettant un résultat de Jean-Pierre Serre sur l'indépendance des groupes de Galois l-adiques des variétés semi-abéliennes, permet de conclure très rapidement (voir [Se], pages 33-34 et 56-59 pour le cas des variétés abéliennes, merci à Jean-Pierre Serre de m'avoir confirmé la généralisation au cas des variétés semi-abéliennes). On a donc un groupe algébrique connexe commutatif défini sur un corps de nombres k et X une sous-variété de G. On se donne deux premiers de bonne réduction $\mathfrak{p}$ et $\mathfrak{t}$, de caractéristiques résiduelles distinctes p et t, les polynômes $F_p(T)$ et $F_t(T)$ et les automorphismes $\sigma_p, \sigma_t \in Gal(\mathbb{Q}^{alg}/k)$ correspondants donnés par 3.2. Le résultat de Serre entraîne que $L = k(Tor_{p'}(G)) \cap k(Tor_p(G))$ est une extension galoisienne finie de k, au-dessus de laquelle $k(Tor_{p'}(G))$ et $k(Tor_p(G))$ sont linéairement disjoints. Donc si $m = [L : k]$, il existe $\tau \in Gal(\mathbb{Q}^{alg}/L)$ qui étend σ_p^m sur la torsion première à p et qui étend σ_t^m sur la p-torsion. Si $f_1, \ldots, f_{2d}$ sont les racines de $F_p(T)$ et $g_1, \ldots, g_{2d}$ sont les racines de $F_t(T)$, on pose $J(T) = \prod_{i=1}^{2d}(T - f_i^m)(T - g_i^m)$. Alors $J(\tau)$ s'annule sur la torsion de G. On se place dans un corps de différence générique (K, τ) qui étend $(\mathbb{Q}^{alg}, \tau)$; par 3.6, $X \cap Ker(J(\tau))$ est une combinaison booléenne finie de sous-ensembles spéciaux de $G(K)$. La même démonstration que pour 3.7 montre qu'alors la clôture de Zariski de $X \cap Tor(G)$ est bien une réunion finie de translatés de sous-groupes algébriques de G. Mais, pour ce qui concerne le calcul du nombre de translatés, la définition explicite de $J(T)$ en fonction des deux polynômes F_p et F_t dépend de $[L : k]$.

4. AUTRES APPLICATIONS

Nous terminons en mentionnant deux autres applications des critères de modularité dans les corps de différence génériques. La première est due à Hrushovski pour un premier cas, puis à T. Scanlon, et la seconde, qui elle utilise des résultats en caractéristique p démontrés dans [ChHrPe], à T. Scanlon.

4.1. La conjecture de Tate-Voloch

J. Tate et J.F. Voloch ont conjecturé l'existence d'une borne pour la distance des points de torsion d'une variété semi-abélienne à une sous-variété ([TaVo]) :

CONJECTURE 4.1. — *Soit A une variété semi-abélienne définie sur $\mathbb{C}_p$ et soit X une sous-variété de A. Il existe une constante c telle que si $a \in Tor(G)$, et que la distance p-adique de a à X est inférieure à c, alors $a \in X(\mathbb{C}_p)$.*

$\mathbb{C}_p$ est la complétion de $\mathbb{Q}_p^{alg}$ pour la norme p-adique, qu'on note $|-|_p$. Si X est une sous-variété de l'espace affine $\mathbb{A}^n$ définie sur $\mathbb{C}_p$, on choisit un système générateur $\{f_1,\ldots,f_n\}$ de l'idéal de définition de X et on définit alors la distance p-adique d'un point a de $\mathbb{A}^n(\mathbb{C}_p)$ à X, $d_p(a,X)$ comme étant le maximum des $|f_i(a)|_p$.

Dans [TaVo] la conjecture est montrée dans le cas des tores.

E. Hrushovski, dans un premier temps, a utilisé la même méthode que pour Manin-Mumford pour établir le résultat pour A définie sur $\mathbb{Q}_p^{alg}$ et pour la torsion première à p, avec l'hypothèse que p est de bonne réduction pour A (communication à Voloch). Cela a ensuite été généralisé par T. Scanlon, toujours par les mêmes méthodes, qui montre la conjecture pour toutes les variétés semi-abéliennes définies sur $\mathbb{Q}_p^{alg}$ dans [Sc1] et [Sc2].

Nous allons indiquer ici brièvement comment on peut déduire du corollaire 2.13 le résultat dans le cas d'une variété semi-abélienne A définie sur $\mathbb{Q}_p$, avec la condition que p est de bonne réduction pour A.

Par 3.2 on a un automorphisme σ tel que la torsion première à p, $Tor_{p'}(A)$ est contenu dans $\{a \in A(\mathbb{Q}_p^{alg}); F(\sigma(a)) = 0\}$, pour $F[T] \in \mathbb{Z}[T]$ sans facteur cyclotomique.

Considérons une suite $(a_i)_{i\in\mathbb{N}}$ d'éléments de $Tor_{p'}(A)$ tels que $d_p(a_i, X)$ tend vers zéro. On va montrer que, pour presque tout i, $a_i \in X$.

Sinon il existe un ultrafiltre (non principal) $\mathfrak{U}$ sur $\mathbb{N}$ tel que $\{i \in \mathbb{N}; a_i \in X\} \notin \mathfrak{U}$. Soit R l'anneau des suites de norme p-adique bornée dans $\mathbb{C}_p$, et l'idéal I de R défini comme l'intersection des I_n, où $I_n = \{r \in R; \{i \in \mathbb{N}; |r(i)|_p \leq p^{-n}\} \in \mathfrak{U}\}$. Alors on peut montrer que $D = R/I$ est un corps, on a un plongement (diagonal) naturel $j : \mathbb{C}_p \mapsto D$ et σ induit un automorphisme de D qu'on appellera aussi σ. Soit a^* l'image de la suite $(a_0,\ldots,a_m,\ldots)$ dans D ; on voit que $a^* \in A(D)$, que $F(\sigma))(a^*) = 0$ et que $a^* \in X(D)$.

On passe dans un gros corps de différence générique (L,σ) étendant (D,σ). Le corollaire 2.13 et le lemme 3.1 nous disent que $X(L) \cap Ker(F(\sigma))$ est contenu dans une réunion finie de translatées de sous-variétés semi-abéliennes de A, chaque translatée étant incluse dans X. En particulier $a^* \in c + B \subset X$, B sous-variété semi-abélienne de A. Soit $\pi : A \mapsto A/B$. Alors la suite des $\pi(a_i)$ doit se rapprocher de $\pi(a^*)$. On a donc que, pour presque tout i, les $\pi(a_i)$ ont même image dans le corps résiduel. Par hypothèse de bonne réduction (appliquée à A/B), on a que la réduction est injective sur la p'-torsion de A/B. Donc la suite des $\pi(a_i)$ devient constante et égale à $\pi(a^*)$. Cela veut dire que pour presque tout i, $a_i \in c+B$, c'est-à-dire que pour presque tout i, $a_i \in X$.

4.2. Modules de Drinfeld

Enfin, plus récemment, T. Scanlon, en utilisant l'analyse modèle-théorique des corps de différence génériques de caractéristique p et le théorème de dichotomie (du type de 2.4) qui est montré dans [ChHrPe], a démontré l'équivalent de la conjecture de Manin-Mumford pour les modules de Drinfeld ([Sc3]).

RÉFÉRENCES

[Ab] A. Abbes – *Hauteurs et discrétude [d'après L. Szpiro, E. Ullmo et S. Zhang]*, Sém. Bourbaki, vol. 1996/97, exp. n° 825, *Astérisque* **245** (1997), 141-166.

[AbrVo] D. Abramovich and F. Voloch – *Towards a proof of the Mordell-Lang conjecture in characteristic p*, Intern. Math. Research Notices (IMRN) No.2 (1992), 103-115.

[Ax] J. Ax – *The elementary theory of finite fields*, Annals of Math. 88 (1968), 239-271.

[Bo] É. Bouscaren ed. – *Model Theory and Algebraic Geometry*, Lecture Notes in Mathematics 1696, Springer 1998.

[Bu] A. Buium – *Intersections in jet spaces and a conjecture of Serge Lang*, Annals of Math. 136 (1992), 583-593.

[Ch1] Z. Chatzidakis – *Groups definable in ACFA*, in Algebraic Model Theory, B. Hart, A. Lachlan and M. Valeriote eds., NATO ASI Series, Kluwer Academic Publishers 1997.

[Ch2] Z. Chatzidakis – *A survey on the model theory of difference fields*, in Model Theory, Algebra and Geometry, D. Haskell and C. Steinhorn ed., MSRI Publications 2000, 65-96.

[ChHr] Z. Chatzidakis and E. Hrushovski – *The model theory of difference fields*, Transactions of the A.M.S, Vol. 351 (1999), 2997-3071.

[ChHrPe] Z. Chatzidakis, E. Hrushovski and Y. Peterzil – *The model theory of difference fields II*, preprint 1999.

[Coh] R.M. Cohn – *Difference algebra*, Tracts in Mathematics 17, Interscience Pub. 1965.

[Col] R. Coleman – *p-adic integrals and torsion points on curves*, Annals of Math. 121 (1985), 111-168.

[DaPh] S. David et P. Philippon – *Minorations des hauteurs normalisées des sous-variétés de variétés abéliennes*, in International Conference On Discrete Mathematics and Number Theory (Tiruchiparelli, 1996), K. Murty and M. Waldschmidt eds., Contemp. Math., 1998, 3-17.

[Du] J.L. Duret – *Les corps faiblement algébriquement clos non séparablement clos ont la propriété d'indépendance*, in Model Theory of Algebra and Arithmetic, L. Pacholski et al. ed., Lecture Notes in mathematics 834, Springer 1980, 135-157.

[Fa] G. Faltings – *The general case of Lang's conjecture*, in Symposium in Algebraic Geometry, V. Christante and W. Messing eds., Perspectives in Math. 15, Academic Press , 1994, 175-182.

[Fu] W. Fulton – *Intersection Theory*, Ergebnisse 2, Springer 1984.

[Go] J.B. Goode – *H.L.M. (Hrushovski-Lang-Mordell)*, Sém. Bourbaki, vol. 1995/96, exp. n° 811, *Astérisque* **241** (1997), 179-194.

[Hi1] M. Hindry – *Autour d'une conjecture de Serge Lang*, Invent. Math. 94 (1988), 575-603.

[Hi2] M. Hindry – *Introduction to abelian varieties and the Mordell-Lang conjecture*, in Model Theory and Algebraic Geometry, E. Bouscaren ed., Lecture Notes in Mathematics 1696, Springer 1998.

[Hr1] E. Hrushovski – *The Mordell-Lang conjecture for function fields*, Journal of the AMS 9 (1996), 667-690.

[Hr2] E. Hrushovski – *The Manin-Mumford conjecture and the model theory of difference fields*, preprint 1995, à paraître dans Annals of Pure and Applied Logic.

[Hr3] E. Hrushovski – *The first-order theory of the Frobenius*, preprint 1995.

[HrPi1] E. Hrushovski and A. Pillay – *Weakly normal groups*, in Logic Colloquium '85, North Holland 1987, 233-244.

[HrPi2] E. Hrushovski and A. Pillay – *Groups definable in local fields and pseudo-finite fields*, Israel J. Math. 85 (1994), 203-262.

[HrSo] E. Hrushovski and Ž. Sokolović – *Minimal subsets of differentially closed fields*, à paraître dans les Transactions of the AMS.

[HrZi] E. Hrushovski and B. Zil'ber – *Zariski Geometries*, Journal of the A.M.S. 9 (1996), 1-56.

[La1] S. Lang – *Division points on curves*, Ann. Mat. Pura Appl. (4) 70 (1965), 229-234.

[La2] S. Lang – *Number Theory III, Diophantine Geometry*, volume 60, Encyclopaedia of Mathematical Sciences, Springer 1991.

[Ma1] A. Macintyre – *Generic automorphisms of fields*, Annals of Pure and Applied Logic 88, vol.2-3 (1997), 165-180.

[Ma2] A. Macintyre – *Non-standard Frobenius*, en préparation.

[Maz] B. Mazur – *Abelian varieties and the Mordell-Lang Conjecture*, in Model Theory, Algebra and Geometry, D. Haskell and C. Steinhorn ed., MSRI Publications 2000.

[McQ] M. McQuillan – *Division points on semi-abelian varieties*, Invent. Math. 120 (1995), 143-159.

[Mr] D. Marker – *Zariski geometries*, in Model Theory and Algebraic Geometry, É. Bouscaren ed., Lecture Notes in Mathematics 1696, Springer 1998.

[Mu] D. MUMFORD – *Abelian varieties*, Oxford University Press, Oxford 1985.

[Pi1] A. PILLAY – *The model-theoretic content of Lang's conjecture* in Model Theory and Algebraic Geometry, É. Bouscaren ed., Lecture Notes in Mathematics 1696, Springer 1998.

[Pi2] A. PILLAY – *ACFA and the Manin-Mumford conjecture*, in Algebraic Model Theory, B. Hart, A. Lachlan and M. Valeriote eds., NATO ASI Series, Kluwer Academic Publishers 1997.

[Pi3] A. PILLAY – *Model Theory and diophantine geometry*, Bull. Am. Math. Soc. 34 (1997), 405-422.

[Ra1] M. RAYNAUD – *Courbes sur une variété abélienne et points de torsion*, Invent. Math. 71 (1983), 207-233.

[Ra2] M. RAYNAUD – *Sous-variétés d'une variété abélienne et points de torsion*, in Arithmetic and Geometry, vol.I, M. Artin and J. Tate eds., Birkhäuser 1983, 327-352.

[Sc1] T. SCANLON – *p-adic distance from torsion points of semi-abelian varieties*, Journal für dir Reine und Angewandte Mathematik 499 (1998), 225-236.

[Sc2] T. SCANLON – *The conjecture of Tate and Voloch on p-adic proximity to torsion*, Intern. Math. Research Notices (IMRN) No. 17 (1999), 909-914.

[Sc3] T. SCANLON – *Diophantine geometry of the torsion of a Drinfeld module*, preprint 1999.

[Se] J.-P. SERRE – *Oeuvres, Collected papers, Volume IV, 1985-1998*, Springer 2000.

[TaVo] J. TATE and J.F. VOLOCH – *Linear forms in p-adic roots of unity*, International Mathematics Research Notices (IMRN) No.12 (1996), 589-601.

[Voj] P. VOJTA – *Integral points on subvarieties of semi-abelian varieties*, Invent. Math. 126 (1996), 133-181.

[We] A. WEIL – *Courbes algébriques et variétés abéliennes*, Hermann 1971.

Élisabeth BOUSCAREN

Université Denis-Diderot Paris 7
Équipe de Logique Mathématique
UFR de Mathématiques (case 7012)
2 place Jussieu
F-75251 Paris Cedex 05
E-mail : `elibou@logique.jussieu.fr`

Séminaire BOURBAKI Mars 2000
52e année, 1999-2000, n$^{\text{o}}$ 871, p. 161 à 188

RATIONAL ELLIPTIC CURVES ARE MODULAR
[after Breuil, Conrad, Diamond and Taylor]

by **Bas EDIXHOVEN**

1. INTRODUCTION

In 1994, Wiles and Taylor-Wiles proved that every semistable elliptic curve over $\mathbb{Q}$ is modular, in the sense that it is a quotient of the jacobian of some modular curve (see [64], [60]). This work has been reported upon in this seminar in [50] and [41]; see especially [50, §1.2] for a historical account. As a consequence, Fermat's Last Theorem, known to be a consequence of this modularity result since work of Ribet based on a conjecture of Serre (see [40]), was finally proved. For a more detailed account of all this, see the book [15], and also [17]. Since 1994, this modularity result has been generalized by an increasing sequence of groups of authors: [24], [14], and [4].

Theorem 1.1 (Diamond). — *Every elliptic curve over $\mathbb{Q}$ that is semistable at 3 and 5 is modular.*

Theorem 1.2 (Conrad, Diamond, Taylor). — *Every elliptic curve over $\mathbb{Q}$ that acquires semistable reduction over a tame extension of $\mathbb{Q}_3$ is modular.*

Theorem 1.3 (Breuil, Conrad, Diamond, Taylor). — *Every elliptic curve over $\mathbb{Q}$ is modular.*

The method of the proofs is basically that of Wiles, i.e., for a given elliptic curve E over $\mathbb{Q}$ one tries to prove that the mod ℓ Galois representation $\overline{\rho}_{E,\ell}$ on $E(\overline{\mathbb{Q}})[\ell]$ is modular for some prime number ℓ, and then that all lifts of $\overline{\rho}_{E,\ell}$ to ℓ-adic representations of a suitable type are modular. The second step involves studying deformations of Galois representations, the systematic theory of which was initiated by Mazur, triggered by work of Hida. The key result for the first step is the celebrated theorem of Langlands [36] and Tunnell [61] that says that $\overline{\rho}_{E,3}$ is modular, as $\mathrm{GL}_2(\mathbb{F}_3)$ is solvable and has

Author partially supported by the Institut Universitaire de France, and by the European TMR Network Contract ERB FMRX 960006 "arithmetic algebraic geometry".

a faithful two-dimensional complex representation. The complications that arise in the proofs of the theorems above simply come from having to prove results as in Wiles and Taylor–Wiles, but with fewer hypotheses. In particular, choosing the right deformations of the restriction of $\overline{\rho}_{E,\ell}$ to $G_\ell := \mathrm{Gal}(\overline{\mathbb{Q}}_\ell/\mathbb{Q}_\ell)$ becomes much more complicated if E does not have semistable reduction at ℓ.

The aim of this report is to give a reasonable sketch of the proofs of the theorems above, to describe the relation to some conjectures by Fontaine and Mazur and by Langlands, and to mention some related results. For some applications of the modularity results above, we refer to [16]. The author of this report does not claim to have checked the computations in [4], but he has studied [4] quite seriously and has not encountered any real problem. Let us also state the following theorem (Theorem B of [4]), whose proof is intricately linked to that of Theorem 1.3 above.

THEOREM 1.4 (Breuil, Conrad, Diamond, Taylor). — *Every irreducible continuous representation $\overline{\rho}\colon G_{\mathbb{Q}} \to \mathrm{GL}_2(\mathbb{F}_5)$ with cyclotomic determinant is modular.*

2. RELATION WITH CONJECTURES BY LANGLANDS, FONTAINE AND MAZUR

The Langlands program predicts, among many other things, that all L-functions coming from algebraic geometry are in fact automorphic, i.e., arise from automorphic representations. More precisely, every absolutely irreducible motive of rank n over a number field F and with coefficients in a subfield E of $\overline{\mathbb{Q}}$ should correspond to a cuspidal algebraic automorphic representation of $\mathrm{GL}_n(\mathbb{A}_F)$ with coefficients in E: see [10, Question 4.16], and the paragraph after that.

In that paragraph, Clozel explains how this conjecture relates to the conjecture of Hasse–Weil type that says that the L-function of such a motive extends meromorphically to all of $\mathbb{C}$ and satisfies a certain functional equation. He finishes by remarking that the only cases for which the Hasse–Weil conjecture has been proved are cases where one actually proves the stronger conjecture, i.e., the existence of an automorphic representation; this remains true after the work of Wiles and its generalizations.

Of course, if E is an elliptic curve over $\mathbb{Q}$, the Langlands program predicts that E is modular. Hence the modularity theorem for elliptic curves over $\mathbb{Q}$ is just a tiny part of the Langlands program.

Fontaine and Mazur stated the following conjecture (Conjecture 1 of [32]).

CONJECTURE 2.1. — *Let ℓ be prime, $n \geq 0$, and let $\rho\colon \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{GL}_n(\mathbb{Q}_\ell)$ be an irreducible continuous representation. Then ρ is isomorphic to a subquotient of some etale cohomology group $\mathrm{H}^i(X_{\overline{\mathbb{Q}}}, \mathbb{Q}_\ell(r))$ with X a smooth projective variety over $\mathbb{Q}$, if and only if ρ satisfies the following two conditions:*

(1) *ρ is ramified at only finitely many primes;*

(2) *the restriction* $\rho|_{G_\ell}$ *to a decomposition group at* ℓ *is potentially semistable (see* [31] *for this notion).*

In one direction, this conjecture has been proved: the $\mathrm{H}^i(X_{\overline{\mathbb{Q}}}, \mathbb{Q}_\ell(r))$ are known to be unramified at almost all primes, and the restriction to G_ℓ is known to be potentially semistable by work of Tsuji and de Jong (see [1, §6.3.3]). It is the other direction that is even more spectacular: it is amazing that just these two conditions should imply, for example, that the Frobenius elements at almost all primes have eigenvalues that are algebraic numbers, and even Weil numbers, and that ρ should be part of a compatible system of ℓ-adic representations. The evidence that one has today for this direction of the conjecture consists of the potentially abelian cases (treated in [32, §6]; ρ occurs in the tensor category generated by representations with finite image and representations which arise from potentially CM abelian varieties), and the cases treated by Wiles' method. However, see [39] for a representation that does satisfy the two conditions above, but for which one does not know if it satisfies Conjecture 2.1.

Combined with the Langlands program, Conjecture 2.1 implies (Conjecture 3c of [32]) that every 2-dimensional ρ satisfying the two conditions, up to Tate twist, either has a finite image, or arises from a modular form of weight at least two.

Since the space of modular forms of a given weight and level is finite dimensional, one also expects certain finiteness results concerning ρ as in Conjecture 2.1, which become semistable over a given extension of $\mathbb{Q}$, and are of fixed Hodge-Tate type: see [32, §3]. Most of [32] is in fact concerned with a deformation theoretic study of these finiteness conjectures.

Suppose now that $\ell > 2$.

For a given absolutely irreducible continuous $\overline{\rho}\colon G_{\mathbb{Q}} \to \mathrm{GL}_2(\mathbb{F}_\ell)$ one considers all lifts $\rho\colon G_{\mathbb{Q}} \to \mathrm{GL}_2(\mathbb{Z}_\ell)$ that are unramified outside a fixed set of primes. The ℓ-adic variety (over $\mathbb{Q}_\ell$) of such lifts is conjecturally three dimensional. Now suppose moreover that $\overline{\rho}|_{G_\ell}$ is absolutely irreducible. Then the variety of lifts of $\overline{\rho}|_{G_\ell}$ is smooth and of dimension five by [42, Thm 4.1]. Since one expects the locus of global lifts that are potentially semistable and of a given type (i.e., Hodge-Tate type at ℓ, and semistable over a fixed extension of $\mathbb{Q}$) to be zero dimensional, one expects that the locus of such local lifts is of codimension three in the five dimensional variety. Indeed, in the crystalline case with Hodge-Tate weights in the interval $[0, \ell-1]$, this was proved in [42] (moreover, the two-dimensional space is smooth). We note that, by [3], "potentially Barsotti-Tate" is equivalent to "potentially crystalline with Hodge-Tate weights in $[0,1]$" (we recall that $\ell > 2$).

Of course, the computations done by Ramakrishna and by Fontaine and Mazur are not directly in terms of representations of G_ℓ. Ramakrishna uses the results of Fontaine and Laffaille, and Fontaine and Mazur work with filtered (ϕ, N)-modules. We note that by recent work of Colmez and Fontaine, [11], one actually has an equivalence of tensor categories between semistable ℓ-adic representations of G_ℓ and weakly admissible filtered (ϕ, N)-modules, which makes it possible to translate problems on the Galois side into problems in linear algebra, even more than before the equivalence between "weakly admissible" and "admissible" was known. On the other hand, what is still not available in this generality is a theory that works for $\mathbb{Z}_\ell$-lattices instead of $\mathbb{Q}_\ell$-vector spaces.

3. REVIEW OF WILES' METHOD

Before turning to the work of Breuil, Conrad, Diamond and Taylor, let us review Wiles' method. Good references for this part are [17], [15], [50], [41], and of course [64] (the introduction of which gives the story of the proof) and [60]. For simplicity, we only discuss this method in a case that suffices for modularity of semistable elliptic curves.

Let E be a semistable elliptic curve over $\mathbb{Q}$. The first observation is that there are many elliptic curves E' over $\mathbb{Q}$ such that $E[5]$ and $E'[5]$ are symplectically isomorphic; this is due to the fact that the modular curve that parameterizes such E' (over $\mathbb{Q}$-schemes) is a non-empty open part of $\mathbb{P}^1_{\mathbb{Q}}$. One proves that there is such an E', semistable, and such that the representation $\overline{\rho}_{E',3}\colon G_{\mathbb{Q}} \to \mathrm{Aut}(E'[3](\overline{\mathbb{Q}})) \cong \mathrm{GL}_2(\mathbb{F}_3)$ is surjective (see [50, §3]). By Langlands and Tunnell, the representation $\overline{\rho}_{E',3}$ is modular. The aim is now to show that $\rho_{E',3}\colon G_{\mathbb{Q}} \to \mathrm{Aut}(E'(\overline{\mathbb{Q}})[3^\infty]) \cong \mathrm{GL}_2(\mathbb{Z}_3)$ is modular, by showing that all 3-adic lifts of $\overline{\rho}_{E',3}$ with reasonable properties are modular, and hence so is E'. Before we discuss how that works, let us see how one then establishes the modularity of E itself.

Of course, if $\overline{\rho}_{E,3}$ is surjective, then we could have taken $E' = E$, so let us assume that $\overline{\rho}_{E,3}$ is not surjective. Then $E[3]$ is in fact reducible (this uses the semistability at all primes; see [50, Proposition 1]). But then $\overline{\rho}_{E,5}$ is irreducible, or $E_{\overline{\mathbb{Q}}}$ is isogeneous to the elliptic curve E_1 over $\overline{\mathbb{Q}}$ that has j-invariant $-5{\cdot}29^3/2^5$, as one sees by looking at the modular curve $X_0(15)$, which has genus one and exactly eight rational points, four of which are cusps (see [46, §2.1]). The elliptic curve E_1 has a model over $\mathbb{Q}$ with conductor 50, which can be checked to be modular. Since modularity is invariant under isogeny and twisting, we may now assume that $\overline{\rho}_{E,5}$ is irreducible, and hence surjective ([50, Proposition 1]). In this case, we already know that $\overline{\rho}_{E,5}$ is modular, because E' is, and one proves the same type of result for modularity of 5-adic liftings as in the 3-adic case.

Let us now give a precise statement of these lifting results. We need some terminology and notation, adapted to the type of representations that we are interested in, i.e., those coming from modular forms of weight two. For each prime p, we choose an embedding $\overline{\mathbb{Q}} \to \overline{\mathbb{Q}}_p$, and we let G_p and I_p denote the corresponding decomposition and inertia subgroups of $G_{\mathbb{Q}}$. We let $\varepsilon\colon G_{\mathbb{Q}} \to \mathbb{Z}_\ell^*$ denote the ℓ-cyclotomic character, given by the action on the elements of ℓ-power order in $\overline{\mathbb{Q}}^*$.

Definition 3.1. — *Let ℓ be a prime number, and k a finite field of characteristic ℓ. Let R be a complete local noetherian ring with residue field k, and let M be a free R-module of rank 2 with a continuous action by $G_{\mathbb{Q}}$; a choice of basis then gives a continuous representation $\rho\colon G_{\mathbb{Q}} \to \mathrm{GL}_2(R)$. For p prime and different from ℓ, M is called semistable at p if, with respect to a suitable basis, $\rho|_{I_p}$ is of the form $\left(\begin{smallmatrix} 1 & * \\ 0 & 1 \end{smallmatrix}\right)$. The representation M is called Barsotti-Tate (at ℓ) if for each finite quotient $\overline{M}$ of M there exists a finite group scheme $\overline{\mathcal{M}}$ over $\mathbb{Z}_\ell$ such that M and $\overline{\mathcal{M}}(\overline{\mathbb{Q}}_\ell)$ are isomorphic as $\mathbb{Z}_\ell[G_\ell]$-modules. The representation M is called semistable at ℓ if it is Barsotti-Tate or if, with respect to a suitable basis, $\rho|_{I_p}$ is of the form $\left(\begin{smallmatrix} \varepsilon & * \\ 0 & 1 \end{smallmatrix}\right)$.*

Theorem 3.2 (Wiles, Taylor-Wiles). — *Let $\ell \neq 2$ be a prime number. Let K be a finite extension of $\mathbb{Q}_\ell$, O its ring of integers, and k its residue field. Let $\rho\colon G_{\mathbb{Q}} \to \mathrm{GL}_2(O)$ be an odd continuous representation such that:*

(1) *its reduction $\overline{\rho}\colon G_{\mathbb{Q}} \to \mathrm{GL}_2(k)$ is modular and its restriction to the quadratic subfield of $\mathbb{Q}(\mu_\ell)$ is absolutely irreducible;*

(2) *$\rho|_{G_\ell}$ is semistable;*

(3) *ρ is ramified at only finitely many primes;*

(4) $\det(\rho) = \varepsilon$*;*

(5) *for every $p \equiv -1 \bmod \ell$ such that $\overline{\rho}|_{I_p}$ is reducible, $\overline{\rho}|_{G_p}$ is reducible too.*

Then ρ is modular.

In view of what has been said above, this result implies that all semistable elliptic curves over $\mathbb{Q}$ are modular. Wiles' strategy to prove Theorem 3.2 is to compare systematically all deformations of $\overline{\rho}$ with certain properties when restricted to decomposition groups to those coming from modular forms of a given level. For simplicity, we will now assume that ρ is semistable at all primes, and follow the exposition in [41], with some modifications, anticipating our discussion of [24], [14] and [4].

So suppose that $\overline{\rho}$ is as in Theorem 3.2, and moreover that ρ is semistable at all primes. We will now forget about ρ, for the moment, but keep $\overline{\rho}$. So $\overline{\rho}$ is a continuous representation of $G_{\mathbb{Q}}$ on a 2-dimensional k-vector space, with k a finite extension of $\mathbb{F}_\ell$ with $\ell \neq 2$, and has the following properties: it is modular, absolutely irreducible after restriction to the quadratic subfield of $\mathbb{Q}(\mu_\ell)$, semistable at all primes, and $\det(\overline{\rho}) = \overline{\varepsilon}$. As nothing about these hypotheses changes if we replace k by a finite extension of it, we may suppose, for example, that the characteristic polynomials of the $\overline{\rho}(\sigma)$, σ in $G_{\mathbb{Q}}$, are all split. Let O be the ring of integers in a finite extension K of $\mathbb{Q}_\ell$, with residue field k. (Later in the proof, we need a modular form of "minimal level" giving rise to $\overline{\rho}$, and with coefficients in O.) For any finite set Σ of primes we define two O-algebras $R_{O,\Sigma}$ and $\mathbb{T}_{O,\Sigma}$, as follows.

Definition 3.3. — *Let R be a complete local noetherian O-algebra with residue field k. A deformation of $\overline{\rho}$ to R is a free R-module M of rank two, with a continuous*

$G_{\mathbb{Q}}$ action, such that $k \otimes_R M$ is isomorphic to $\overline{\rho}$. A deformation ρ is said to be of type Σ if $\det(\rho) = \varepsilon$, and ρ is semistable at ℓ and "minimally ramified" outside Σ:

(1) *if $\ell \notin \Sigma$ and $\overline{\rho}$ is Barsotti-Tate, then ρ is Barsotti-Tate;*

(2) *if $p \notin \Sigma \cup \{\ell\}$ and $\overline{\rho}$ is unramified at p, then ρ is unramified at p;*

(3) *if $p \notin \Sigma \cup \{\ell\}$ and $\overline{\rho}$ is ramified at p (and hence semistable, with our hypotheses), then ρ is semistable at p.*

With these definitions, there is, for each Σ, a universal deformation ring $R_{O,\Sigma}$ that represents the functor that sends R to the set of isomorphism classes of deformations of type Σ over R. A very good reference for this is [20]. If $K \to K'$ is a finite extension, then $R_{O',\Sigma} = O' \otimes_O R_{O,\Sigma}$.

Let us now turn to the definition of $\mathbb{T}_{O,\Sigma}$. The reader is referred to Appendix A for certain properties of the Galois representation ρ_f associated to a modular form f of weight two with coefficients in $\overline{\mathbb{Q}}_\ell$. We define $\mathcal{N}_\Sigma$ to be the set of weight two newforms f with coefficients in $\overline{\mathbb{Q}}_\ell$ such that $\rho_f \colon G_{\mathbb{Q}} \to \mathrm{GL}_2(O_f)$ is of type Σ, where O_f is the sub-O-algebra of $\overline{\mathbb{Q}}_\ell$ generated by the coefficients of f. The results in Appendix A imply that there is an integer N_Σ, such that for each f in $\mathcal{N}_\Sigma$, the level of f divides N_Σ. This implies that $\mathcal{N}_\Sigma$ is a finite set. One can take N_Σ as follows:

$$N_\Sigma := \ell^\delta \cdot \prod_{p \mid N(\overline{\rho})} p \cdot \prod_{p \in \Sigma - \{\ell\}} p^2,$$

where δ is 0 if $\overline{\rho}$ is Barsotti-Tate and ℓ not in Σ, and δ is 1 otherwise, and where $N(\overline{\rho})$ is the level associated to $\overline{\rho}$ by Serre in [49] (i.e., $N(\overline{\rho})$ is given by the usual formula for the Artin conductor of a representation, in terms of the ramification subgroups at all $p \neq \ell$). The reason that such a N_Σ suffices is that the wild parts of the conductors of ρ_f and $\overline{\rho}_f$ are equal. For simplicity, we will now only consider Σ that do not contain primes dividing $N(\overline{\rho})$ (this suffices for the application to semistable lifts of $\overline{\rho}$). For each f in $\mathcal{N}_\Sigma$, we have a morphism $R_{O,\Sigma} \to O_f$, and we define $\mathbb{T}_{O,\Sigma}$ to be the image of $R_{O,\Sigma}$ in the product of the O_f. Since R_Σ is generated, as O-algebra, by the traces of elements in the universal representation, $\mathbb{T}_{O,\Sigma}$ is generated by the elements $a_p \colon f \mapsto a_p(f)$ for p not dividing ℓN_Σ.

The method of Wiles is now to show that the surjections $R_{O,\Sigma} \to \mathbb{T}_{O,\Sigma}$ are isomorphisms, by studying how they change as Σ varies. The first step in this is to use what has been proved about Serre's conjectures on modularity of mod ℓ representations in [49]: $\mathcal{N}_\varnothing$ is not empty (see [45] and [22]). This implies that we can suppose (and we will) that we have a section $\pi = \pi_\varnothing \colon \mathbb{T}_\varnothing \to O$. We let P_Σ denote the corresponding O-valued points of $\mathrm{Spec}(\mathbb{T}_{O,\Sigma})$ and $\mathrm{Spec}(R_{O,\Sigma})$, for each Σ. Wiles introduced the following O-modules associated to each Σ: on the one hand the cotangent spaces at the P_Σ, i.e., $P_\Sigma^* \Omega^1_{R_{O,\Sigma}/O}$ and $P_\Sigma^* \Omega^1_{\mathbb{T}_{O,\Sigma}/O}$, and on the other hand the "module of congruences" $P_\Sigma^* \mathcal{O}_{Z_\Sigma}$, defined as follows. Since $\mathbb{T}_{O,\Sigma}$ is finite free as O-module, $\mathrm{Spec}(\mathbb{Q} \otimes \mathbb{T}_{O,\Sigma})$ is the disjoint union of two open and closed subschemes $P_{\Sigma,K}$ and $Z_{\Sigma,K}$, with $P_{\Sigma,K}$ consisting of the point $P_\Sigma(\mathrm{Spec}(K))$. We let Z_Σ be the scheme theoretic closure of $Z_{\Sigma,K}$ in $\mathrm{Spec}(\mathbb{T}_{O,\Sigma})$ (note that the $\mathbb{T}_{O,\Sigma}$ are reduced by construction). These modules, that will intervene only via their lengths, are usually introduced as $\ker(P_\Sigma^*)/\ker(P_\Sigma^*)^2$ and $P_\Sigma^*\mathrm{Ann}_{\mathbb{T}_{O,\Sigma}}(\ker(P_\Sigma^*))$. (This last module has finite length if and only if $\mathrm{Spec}(\mathbb{Q}\otimes\mathbb{T}_{O,\Sigma})$ is reduced at $P_{\Sigma,K}$.)

The fact that $R_{O,\Sigma}$ represents the functor of isomorphism classes of deformations of type Σ implies the following Galois cohomological description of $P^*\Omega^1_{R_{O,\Sigma}/O}$:

$$\mathrm{Hom}_O(P^*\Omega^1_{R_{O,\Sigma}/O}, K/O) = \mathrm{H}^1_\Sigma(G_\mathbb{Q}, \mathrm{ad}^0(\rho)\otimes K/O),$$

where ρ is the representation corresponding to $P = P_\Sigma$, where $\mathrm{ad}^0(\rho)$ is the representation of $G_\mathbb{Q}$ on the sub-O-module of trace zero elements of $\mathrm{End}_O(M_\rho)$, and where $\mathrm{H}^1_\Sigma(G_\mathbb{Q}, \mathrm{ad}^0(\rho)\otimes K/O)$ denotes the subgroup of $\mathrm{H}^1(G_\mathbb{Q}, \mathrm{ad}^0(\rho)\otimes K/O)$ of classes that map, at all p, to the subgroups $L_{\Sigma,p}$ of the $\mathrm{H}^1(G_p, \mathrm{ad}^0(\rho)\otimes K/O)$ that reflect the conditions for deformations to be of type Σ. To be explicit, these $L_{\Sigma,p}$ are:

– $L_{\Sigma,p} = \mathrm{H}^1(G_p/I_p, (\mathrm{ad}^0(\rho)\otimes K/O)^{I_p})$ if $p\notin\Sigma\cup\{\ell\}$;

– $L_{\Sigma,p} = \mathrm{H}^1(G_p, \mathrm{ad}^0(\rho)\otimes K/O)$ if $p\in\Sigma$ and $p\neq\ell$;

– $L_{\Sigma,\ell}$ is the subspace of $\mathrm{H}^1(G_\ell, \mathrm{ad}^0(\rho)\otimes K/O)$ that corresponds to deformations that are Barsotti-Tate, if $\ell\notin\Sigma$;

– $L_{\Sigma,\ell}$ is the subspace of $\mathrm{H}^1(G_\ell, \mathrm{ad}^0(\rho)\otimes K/O)$ that corresponds to deformations that are semistable at ℓ, if $\ell\in\Sigma$.

The results of Poitou-Tate on local duality and global Euler characteristic show that, for M a finite discrete $G_\mathbb{Q}$-module, with a Selmer datum $L_v\subset\mathrm{H}^1(G_v, M)$ at all places v of $\mathbb{Q}$, one has:

$$\frac{\#\mathrm{H}^1_L(G_\mathbb{Q}, M)}{\#\mathrm{H}^1_{L^\perp}(G_\mathbb{Q}, M^*)} = \frac{\#\mathrm{H}^0(G_\mathbb{Q}, M)}{\#\mathrm{H}^0(G_\mathbb{Q}, M^*)}\cdot\prod_v\frac{\#L_v}{\#\mathrm{H}^0(G_v, M)},$$

where M^* is the Cartier dual $\mathrm{Hom}(M, \overline{\mathbb{Q}}^*)$ of M, and where, for each v, $L_v^\perp$ is the orthogonal of L_v. Moreover, if $L\subset L'$ are two Selmer data for M, then one has an exact sequence:

$$\begin{aligned}0\to\mathrm{H}^1_L(G_\mathbb{Q}, M)\to\mathrm{H}^1_{L'}(G_\mathbb{Q}, M)\to\prod_v L'_v/L_v&\\ \to\mathrm{H}^1_{L^\perp}(G_\mathbb{Q}, M^*)^\vee\to\mathrm{H}^1_{L'^\perp}(G_\mathbb{Q}, M^*)^\vee\to 0.&\end{aligned}$$

Having established this, Wiles first proves that $R_{O,\varnothing}\to\mathbb{T}_{O,\varnothing}$ is an isomorphism, and then that this remains so if one enlarges Σ. The argument for the first step is really amazing, he somehow manages to "patch", for a suitable sequence of Σ_n, the R_{O,Σ_n} and the $\mathbb{T}_{O,\Sigma_n}$ into power series rings, with the same number of generators, and deduce from that that $R_{O,\varnothing}\to\mathbb{T}_{O,\varnothing}$ is an isomorphism. (This patching argument was introduced in [60], and used to show that $\mathbb{T}_{O,\varnothing}$ is a complete intersection, but Faltings pointed out that one could also use the argument directly in proving $R_{O,\varnothing}\to\mathbb{T}_{O,\varnothing}$ to be an isomorphism.) We will now take a closer look at this argument, in order to see which conditions have to be satisfied by the type of deformations that one considers for it to work.

So suppose that one wants to do this argument for $R_{O,\Sigma}$. The primes p that one wants to add to Σ are congruent to 1 modulo a high power of ℓ, and such that $\overline{\rho}$ is unramified at p with distinct Frobenius eigenvalues in k^*. For such a p, and for Σ' containing p, $\rho^{\mathrm{univ}}_{O,\Sigma'}|_{G_p}$ is a direct sum of two characters, whose restrictions to I_p factor through the ℓ-part Δ_p of $\mathrm{Gal}(\mathbb{Q}_p^{\mathrm{unr}}(\mu_p)/\mathbb{Q}_p^{\mathrm{unr}}) = (\mathbb{Z}/p\mathbb{Z})^*$. Choosing one of the two Frobenius eigenvalues gives $R_{O,\Sigma'}$ the structure of an $O[\Delta_p]$-algebra. The Σ' that one wants to consider are of the form $\Sigma' = \Sigma \cup Q$, with Q a set of r elements, for some fixed r, and such that $R_{O,\Sigma'}$ can also be topologically generated by r elements. (Note that $R_{O,\Sigma'}$ is an algebra over $O[\Delta_Q]$ with $\Delta_Q = \prod_{p\in Q}\Delta_p$, and that $O[\Delta_Q]$ looks more and more as a power series ring in r variables, as the primes p are closer to 1, ℓ-adically.) Let L and L' be the Selmer data corresponding to Σ and Σ'. Then $\dim_k \prod_v L'_v/L_v = r$, so one finds, by the exact sequence above, that $\dim_k \mathrm{H}^1_{L^\perp}(G_{\mathbb{Q}}, \mathrm{ad}^0(\overline{\rho})^*) \geq \dim_k(\mathrm{H}^1_L(G_{\mathbb{Q}}, \mathrm{ad}^0(\overline{\rho})))$. But in the displayed formula above, one has $\dim_k L_p \geq \dim_k \mathrm{H}^1(G_p/I_p, \mathrm{ad}^0(\overline{\rho})^{I_p}) = \dim_k(\mathrm{H}^0(G_p, \mathrm{ad}^0(\overline{\rho}))$, for all $p \neq \ell$, whereas $\dim_k L_\infty = 0$ and $\dim \mathrm{H}^0(G_\infty, \mathrm{ad}^0(\overline{\rho})) = 1$. Moreover, in that formula one has $\#\mathrm{H}^0(G_{\mathbb{Q}}, M) = 1$ (since $\overline{\rho}$ is absolutely irreducible), and $\#\mathrm{H}^0(G_{\mathbb{Q}}, M^*) = 1$ (since the restriction of $\overline{\rho}$ to the quadratic subfield of $\mathbb{Q}(\mu_\ell)$ is absolutely irreducible). It follows that this setup can only work if $L_p = \mathrm{H}^1(G_p/I_p, \mathrm{ad}^0(\overline{\rho})^{I_p})$ for all $p \neq \ell$, and $\dim_k(L_\ell) \leq 1 + \dim_k \mathrm{ad}^0(\overline{\rho})^{G_\ell}$. This means that Σ must be $\varnothing$, and that L_ℓ must be of dimension 1, unless $\overline{\rho}|_{I_\ell} \cong \overline{\varepsilon} \oplus 1$. This last condition puts a very strong restriction on the type of local deformations at ℓ that one can use.

In order to prove that $R_{O,\varnothing} \to \mathbb{T}_{O,\varnothing}$ is an isomorphism, Taylor and Wiles use that, in their situation, the localization at $\overline{\rho}$ of $\mathrm{H}^1(X_0(N_\varnothing)(\mathbb{C}), O)$ is a free $\mathbb{T}_{O,\varnothing}$-module, and similarly for the Σ's that they choose. Such results are quite delicate to prove. In the next section we will discuss how Diamond and Fujiwara have gotten around this, and actually obtain such freeness results as a consequence of the method. In [17], the freeness assumption is not used, but the given proof still relies on q-expansions (see [17, Remark 4.15]).

Let us now briefly discuss how Wiles proved that the $R_{O,\Sigma} \to \mathbb{T}_{O,\Sigma}$ are isomorphisms. This is done by induction on the number of elements of Σ, but, in order to carry out this induction, one actually proves more, namely, that these O-algebras are complete intersections. Indeed, Wiles found a criterion for doing the induction, in terms of the changes of the lengths of $P^*\Omega^1_{R_{O,\Sigma}/O}$ and $P^*\mathcal{O}_{Z_\Sigma}$ when comparing between Σ and $\Sigma' := \Sigma \cup \{p\}$. On the Galois side, the exact sequence above gives an upper bound for the length of $P^*\Omega^1_{R_{O,\Sigma'}/O}$. On the Hecke side, [17, §4.4] gives a new proof of the lower bound for the length of $P^*\mathcal{O}_{Z'_\Sigma}$ that was proved by Wiles. This proof does not use freeness, and it nicely relates this change of length to the residue at 2 of the L-function of the symmetric square of the system of representations associated to P, giving a relation to the Bloch-Kato conjectures. Wiles' argument, which is to compute the composite $J_0(N_\Sigma) \to J_0(N_{\Sigma'}) \to J_0(N_\Sigma)$, is also sketched in [17, §4.4].

4. IMPROVEMENTS OF THE COMMUTATIVE ALGEBRA PART

The results in commutative algebra that are used in [14] and [4] are improvements of those in [64] and [60]. These improvements were found independently by Diamond [25] and Fujiwara [33], motivated by Fujiwara's work on modularity over totally real number fields. We also note that Lenstra, de Smit, Rubin, and Schoof have established isomorphism and complete intersection criteria as in Wiles, without the Gorenstein hypothesis, and without the limiting process, see [21]. Let us now state the criteria as in [25, Theorems 2.1 and 2.4].

THEOREM 4.1. — *Let k be a finite field, and $r \geq 0$ an integer.*

Let $A := k[[S_1, \ldots, S_r]]$, $B := k[[X_1, \ldots, X_r]]$, let R be a k-algebra, and let H be a non-zero R-module that has finite k-dimension. Suppose that for every $n \geq 1$ one has a commutative diagram:

$$\begin{array}{ccc} A & \xrightarrow{\phi_n} & B \\ \big\downarrow & & \big\downarrow \psi_n \\ k & \longrightarrow & R \end{array}$$

and a B-module H_n with a morphism $\pi_n \colon H_n \to H$ such that as an A-module, H_n is free over A/m_A^n, and such that the morphism $k \otimes_A H_n \to H$ induced by π_n is an isomorphism. Then H is free over R, and R is a (zero dimensional) complete intersection.

In the application of this result, k is as above, A is a projective limit of k-algebras of the form $k[\Delta_Q]$, with Q a set of r distinct primes $p \equiv 1 \bmod \ell^n$ and Δ_Q the product of the $(\mathbb{Z}/p\mathbb{Z})^*_\ell$, B is a projective limit of $k \otimes_O R_{O,Q}$'s, $R = k \otimes_O R_{O,\varnothing}$, and H and H_n come from (co)homology groups of modular curves. The freeness of H_n over A/m_A^n basically comes from standard facts about cohomology of locally constant sheaves and unramified covers of affine Riemann surfaces. The Hecke algebra $k \otimes_O \mathbb{T}_{O,\varnothing}$ is the image T of R in $\mathrm{End}_k(M)$, so the conclusion that H is free over R implies that $R = T$. The freeness version of Wiles' numerical criterion is as follows.

THEOREM 4.2. — *Let O be a complete discrete valuation ring with finite residue field k, and let R be a complete noetherian local O-algebra. Let H be an R-module, finite free over O, let $\phi \colon R \to T$ be the quotient by $\mathrm{Ann}_R(H)$, and suppose that T has a section $\pi_T \colon T \to O$. Put $\pi_R := \phi\pi_T$. Define $\Omega := H/(H[\ker(\pi_T)] + H[\mathrm{Ann}_T(\ker(\pi_T))])$. Let d be the O-rank of $H[\ker(\pi_R)]$. If Ω has finite length over O, then the following are equivalent:*

(1) $\mathrm{rank}_O H \leq d{\cdot}\mathrm{rank}_O T$ *and* $\mathrm{length}_O \Omega \geq d{\cdot}\mathrm{length}_O(\ker(\pi_R)/\ker(\pi_R)^2)$*;*
(2) $\mathrm{rank}_O H = d{\cdot}\mathrm{rank}_O T$ *and* $\Omega \cong (O/\mathrm{Fitt}_O(\ker(\pi_R)/\ker(\pi_R)^2))^d$*;*
(3) *R is a complete intersection and H is free (of rank d) over R.*

5. THE WORK OF BREUIL, CONRAD, DIAMOND, AND TAYLOR

We are now ready to discuss the work of the four authors mentioned above in [24], [14], and [4]. Before getting into any details, let us see what problems were solved in each of these three articles. In [24, Theorem 5.3], Diamond gets rid of condition (5) in Theorem 3.2. To be precise, let us state his result.

THEOREM 5.1 (Diamond). — *Let $\ell > 2$ be prime, K a finite extension of $\mathbb{Q}_\ell$, O its ring of integers, k its residue field, and $\rho\colon G_\mathbb{Q} \to \mathrm{GL}_2(O)$ an odd continuous representation such that:*

(1) *its reduction $\overline{\rho}\colon G_\mathbb{Q} \to \mathrm{GL}_2(k)$ is modular, and its restriction to the quadratic subfield of $\mathbb{Q}(\mu_\ell)$ is absolutely irreducible;*

(2) *$\rho|_{G_\ell}$ is Barsotti-Tate and $\det(\rho)|_{I_\ell} = \varepsilon|_{I_\ell}$, or, with respect to a suitable basis, $\rho|_{G_\ell}$ is of the form $\left(\begin{smallmatrix}\phi & * \\ 0 & \psi\end{smallmatrix}\right)$, with ψ unramified, $\overline{\psi} \neq \overline{\phi}$, and $\phi|_{I_\ell} = \chi\varepsilon^{k-1}|_{I_\ell}$ for some integer $k \geq 2$ and χ of finite order;*

(3) *ρ is ramified at only finitely many primes.*

Then ρ is modular.

We should note here that Theorem 3.2 is weaker than the result that was proved by Wiles. What is proved in [64] is the theorem above, with the extra condition (5) of Theorem 3.2. Let us now explain what the problem is in a case that does not satisfy this condition (5).

So suppose that $\overline{\rho}$ satisfies the conditions of Theorem 5.1, that $p \neq \ell$, that $\overline{\rho}_p = \overline{\rho}|_{G_p}$ is irreducible, but $\overline{\rho}|_{I_p}$ is reducible. Then $\overline{\rho}_p$ is of the form $\mathrm{Ind}_{\mathbb{Q}_{p^2}}^{\mathbb{Q}_p}\psi$, with $\mathbb{Q}_{p^2}$ the unramified extension of degree two of $\mathbb{Q}_p$, and $\psi\colon G_{\mathbb{Q}_{p^2}} \to k^*$ a continuous character such that $\psi^\sigma \neq \psi$. (To prove this, note that $\overline{\rho}(I_p)$ must have exactly two fixed points on $\mathbb{P}^1(\mathbb{F}_p)$, that are interchanged by Frob_p, since otherwise $\overline{\rho}_p$ would be reducible.) But then, if ℓ divides $p+1$, there are nontrivial deformations of $\overline{\rho}_p$ of the form $\rho_p = \mathrm{Ind}_{\mathbb{Q}_{p^2}}^{\mathbb{Q}_p}(\psi\mu)$ from G_p to $\mathrm{GL}_2(O')$, with O' a finite extension of O, $\psi\colon G_p \to O'^*$ the Teichmüller lift of $\psi\colon G_p \to k^*$, and with $\mu\colon G_{\mathbb{Q}_{p^2}} \to O'^*$ of order ℓ. One checks that $\det(\rho)|_{I_p}$ is the Teichmüller lift of $\det(\overline{\rho})|_{I_p}$ (use that Frob_p acts on the tame inertia group $I_p^{\mathrm{tame}} = \hat{\mathbb{Z}}^{(p)}(1)$ by multiplication by p). The whole problem arises from the fact that, on the one hand, $\overline{\rho}_p$ and ρ_p have the same Artin conductor, namely, the square of the conductor of ψ, but that, on the other hand, $\overline{\rho}_p$ admits different lifts with this conductor. This means that if we consider lifts of $\overline{\rho}$ to be minimally ramified at p if they have Artin conductor $\mathrm{cond}(\psi)^2$ at p, then we get an $L_p \subset \mathrm{H}^1(G_p, \mathrm{ad}^0(\overline{\rho}))$, in the notation of Section 3, that is nonzero, making already Wiles' method at the minimal level impossible.

The conclusion is that, on the automorphic side, levels of newforms are not fine enough invariants to work with; one should impose finer conditions on the restrictions to the G_p (at the Galois side), and corresponding conditions on the irreducible admissible representations on the other side. Wiles already notes this in the second remark following Conjecture 2.16 in [64].

The finer conditions that will be imposed are in terms of what are called "types" and "extended types" in the articles that we discuss here. An extended type, at a prime p ($p = \ell$ is allowed) is simply an isomorphism class of two-dimensional representations over $\overline{\mathbb{Q}}_\ell$ of the Weil-Deligne group W'_p of $\mathbb{Q}_p$ (see Appendix A), and then types are isomorphism classes of restrictions to I_p of extended types. The local Langlands correspondence makes extended types correspond to isomorphism classes of infinite dimensional irreducible admissible representations of $\mathrm{GL}_2(\mathbb{Q}_p)$, over $\overline{\mathbb{Q}}_\ell$. We will not discuss the proof of Theorem 5.1 here, as it is repeated in [14] and [4], with some changes, however. Diamond used, in order to simplify the representation theory at the automorphic side, the Jacquet-Langlands correspondence to work with a quaternion algebra over $\mathbb{Q}_p$ instead of the matrix algebra. In the two subsequent articles, one works directly with modular curves. We recommend [23] for an overview of [24], that does not become too technical. But, as the reader can already guess, the rest of this section will get more technical, especially on the automorphic side, because of these finer restrictions.

With Theorem 5.1 above, it is not hard to prove that all elliptic curves over $\mathbb{Q}$ that are semistable at 3 and 5 are modular; we refer to [24, §5] for details. So the only remaining problem to get modularity for all elliptic curves over $\mathbb{Q}$ is to get rid of the semistability conditions at 3 and 5. Since modularity is invariant under twisting, Diamond's result actually implies that the only elliptic curves E that remain to be dealt with are those that have potentially good reduction at 3 and 5, but that do not have a twist with good reduction at 3 and 5. Since one knows that the two j-invariants that correspond to E with more than two automorphisms are modular, the only twists to consider are quadratic twists.

The first step in the direction of relaxing the conditions at 3 and 5 was made in [14], where it is proved that E is modular if it acquires good reduction over a tame extension of $\mathbb{Q}_3$. The main new ingredient of this paper, compared to [24], is a new type of deformation problem, for a mod ℓ representation of G_ℓ. Roughly speaking, one considers deformations ρ over R of $\overline{\rho}$ over k such that the restriction of ρ to G_F, or a twist of it by a fixed quadratic character, is Barsotti-Tate, for F a fixed finite extension of $\mathbb{Q}_\ell$ with ramification index $e \leq \ell - 1$. We have seen, in Section 3, that it is crucial that the tangent space over k of the universal deformation ring of the type of deformations of $\overline{\rho}|_{G_\ell}$ that one considers be of dimension at most one. This crucial result for [14] was proved by Conrad in [13], using his description of finite free group schemes over the rings of integers of such F obtained in [12], generalizing earlier work of Fontaine in [30]. Using Conrad's result, it was then proved ([14, Theorem 7.1.2]) that any elliptic curve E over $\mathbb{Q}$ that acquires good reduction over a tame extension of $\mathbb{Q}_3$ is modular.

The final step in relaxing the conditions at 3 is done in [4]. It is the work of Breuil [3], summarized in [2], that gives a workable enough description of certain finite free group schemes over rings of integers of arbitrary finite extensions F of $\mathbb{Q}_\ell$, that makes this possible. With this tool available, it is then proved that the remaining E, i.e., those that acquire good reduction only after a wild extension of $\mathbb{Q}_3$, are modular. The article [4] consists for about 70% (of 77 pages) of the proof that, in the cases that are needed, the local deformation space at ℓ has dimension at most one.

In order to keep this section of reasonable length, we postpone the discussion of these results at ℓ to the next one, and in this one we focus more on the global aspects of the proof, and on Conjecture 1.3.1 of [4], which says for which $\rho\colon G_{\mathbb{Q}} \to \mathrm{GL}_2(O)$ one hopes to be able to prove modularity.

We will now follow [4, §1], in order to introduce the necessary terminology, and to state the conjecture just mentioned. We suppose $\ell > 2$. An extended ℓ-type is defined as an isomorphism class of two-dimensional representations of W'_ℓ over $\overline{\mathbb{Q}}_\ell$ (with open kernel), and ℓ-types are isomorphism classes of restrictions to I_ℓ of extended ℓ-types.

Suppose that τ' is an extended ℓ-type, with restriction τ to I_ℓ.

Definition 5.2. — *A continuous representation $\rho\colon G_\ell \to \mathrm{GL}_2(O)$ (with O the ring of integers in a finite extension K of $\mathbb{Q}_\ell$) is said to be of extended type τ' (resp. of type τ) if:*

(1) *ρ is potentially Barsotti-Tate;*

(2) *$\mathrm{WD}(\rho)$ (as in Appendix B) is in τ' (resp. $\mathrm{WD}(\rho)|_{I_\ell}$ is in τ);*

(3) *the character $\varepsilon^{-1}\det(\rho)$ has finite order prime to ℓ.*

Now fix a finite extension K of $\mathbb{Q}_\ell$, let O be its ring of integers and k be its residue field. Let $\overline{\rho}$ be a two-dimensional continuous representation of G_ℓ over k, say on a vector space V, such that $\mathrm{End}_{G_\ell}(V) = k$ (i.e., either $\overline{\rho}$ is absolutely irreducible, or it is a non-split extension of a character by another character). Under this last hypothesis, we have a universal deformation ring R^l_O representing the functor of deformations of $\overline{\rho}$ to complete noetherian local O-algebras with residue field k. (The superscript ℓ is there to indicate that we are considering representations of G_ℓ.) We remark that extended types will only be considered if their restriction to I_ℓ is irreducible. Now consider $\mathrm{Spec}(R^l_O)$. In it, we have a minimal closed subset that contains all deformations of $\overline{\rho}$ to finite extensions O' of O that are of type τ, and similarly for extended type τ'. These minimal closed subsets correspond to (reduced) quotients $R^l_{O,\tau}$ and $R^l_{O,\tau'}$. A deformation ρ over R of $\overline{\rho}$ is said to be *weakly of type* τ (resp.

weakly of extended type τ'), if the corresponding morphism $R^l_O \to R$ factors through $R^l_{O,\tau}$ (resp. through $R^l_{O,\tau'}$).

DEFINITION 5.3. — *A type τ (resp. an extended type τ') is* weakly acceptable *for $\overline{\rho}$ if there exists a surjection of O-algebras $O[[X]] \to R^l_{O,\tau}$ (resp. $O[[X]] \to R^l_{O,\tau'}$). A type τ (resp. an extended type τ') is* acceptable *for $\overline{\rho}$ if moreover $R^l_{O,\tau} \neq 0$ (resp. $R^l_{O,\tau'} \neq 0$), i.e., if there exists at least one ℓ-adic deformation of type τ (resp. of extended type τ'). We will also speak of $\overline{\rho}$ accepting τ (resp. τ').*

Of course, with these definitions, it is very hard to check whether a given $\overline{\rho}$ accepts a given τ or τ'. It is precisely this kind of verifications that occupy the most of [4], and it is there that crucial use is made of Conrad's and Breuil's results on finite group schemes. We note that [4] conjectures that an ℓ-adic lift of $\overline{\rho}$ is of type τ (resp. extended type τ') if and only if it is weakly of that kind (Conjecture 1.1.1 of [4]), but this has no importance for what follows. What is much more important, is that [4, Conjecture 1.3.1] tries to predict acceptability purely in computable, representation theoretic terms. In order to state this conjecture, [4] needs about 4 pages of preparation, consisting mostly of definitions. Instead of trying to state all these definitions, we will try to see where they come from.

The question one should ask oneself is: if f is a newform over $\overline{\mathbb{Q}}_\ell$, then what can one say about $\overline{\rho}_{f,\ell}\colon G_\ell \to \mathrm{GL}_2(\overline{\mathbb{F}}_\ell)$ in terms of $\pi_{f,\ell}$, assuming $\overline{\rho}_f$ irreducible? In particular, for a given $\overline{\rho}$, what are the irreducible admissible representations that occur as $\pi_{f,\ell}$, for f with $\overline{\rho}_{f,\ell} \cong \overline{\rho}$?

An answer to this question will then say for which τ and τ' there does exist an ℓ-adic lift of $\overline{\rho}$ of that type. Moreover, from the mechanism that is used to find this, one may guess under what conditions one expects $R^l_{O,\tau}$ or $R^l_{O,\tau'}$ to be topologically generated by one element.

To find the answer to the question (and for other reasons as well), [4] constructs certain ℓ-adic sheaves on certain modular curves, that pick out a non-zero part of exactly those π_f such that $\pi_{f,\ell}$ has a prescribed type (or extended type). For each τ and for each τ',with $\tau'|_{I_\ell}$ irreducible, one defines open subgroups U_τ of $\mathrm{GL}_2(\mathbb{Z}_\ell)$ and $U_{\tau'}$ of $\mathrm{GL}_2(\mathbb{Q}_\ell)$, and irreducible representations σ_τ and $\sigma_{\tau'}$ on finite dimensional $\overline{\mathbb{Q}}_\ell$-vector spaces, with open kernel. The choice of these subgroups and representations is justified by [4, Lemma 1.2.1]: for every irreducible admissible representation π of $\mathrm{GL}_2(\mathbb{Q}_\ell)$ over $\overline{\mathbb{Q}}_\ell$ one has

– $\mathrm{Hom}_{U_\tau}(\sigma_\tau, \pi) = \overline{\mathbb{Q}}_\ell$ if $\mathrm{WD}(\pi)|_{I_\ell} \cong \tau$, and $\mathrm{Hom}_{U_\tau}(\sigma_\tau, \pi) = 0$ if $\mathrm{WD}(\pi)|_{I_\ell} \ncong \tau$;

– $\mathrm{Hom}_{U_{\tau'}}(\sigma_{\tau'}, \pi) = \overline{\mathbb{Q}}_\ell$ if $\mathrm{WD}(\pi) \cong \tau'$, and $\mathrm{Hom}_{U_{\tau'}}(\sigma_{\tau'}, \pi) = 0$ if $\mathrm{WD}(\pi) \ncong \tau'$.

The fact that such subgroups and representations exist is not particular to our situation. There is a general theory, called (no surprise) type theory, whose goal it is to describe smooth representations of p-adic groups in terms of their restrictions to compact open subgroups; see [6]. Before we go on, let us mention that Khare has also asked and answered the question above, at least in the case of types, in [34].

With these (U_τ, σ_τ) and $(U_{\tau'}, \sigma_{\tau'})$ one constructs sheaves on modular curves as follows. One defines U_ℓ to be U_τ if one considers a type, and $U_{\tau'} \cap \mathrm{GL}_2(\mathbb{Z}_\ell)$ if one considers an extended type. In each case, one has a representation σ_ℓ of U_ℓ, namely, σ_τ and the restriction of $\sigma_{\tau'}$. Let $U^{(\ell)}$ be a sufficiently small open subgroup of $\mathrm{GL}_2(\hat{\mathbb{Z}}^{(\ell)})$, and let σ be the representation of $U := U_\ell U^{(\ell)}$ given by σ_ℓ. Then the modular curve and the sheaf are:

$$Y_U := \mathrm{GL}_2(\mathbb{Q})\backslash\mathrm{GL}_2(\mathbb{A})/U\mathbb{C}^*, \quad \mathcal{F}_\sigma := \mathrm{GL}_2(\mathbb{Q})\backslash\left(\mathrm{GL}_2(\mathbb{A}) \times M_\sigma^\vee\right)/U\mathbb{C}^*,$$

with M_σ the $\overline{\mathbb{Q}}_\ell[U]$-module given by σ. In the case where one considers an extended type, one also gets an automorphism w_ℓ of the pair $(Y_U, \mathcal{F}_\sigma)$. By construction, the duals of the two-dimensional Galois representations ρ that occur in $\mathrm{H}^1_!(Y_U, \mathcal{F}_\sigma)$ if one considers a type (resp. in $\mathrm{H}^1_!(Y_U, \mathcal{F}_\sigma)^{w_\ell=1}$ if one considers an extended type) for some $U^{(\ell)}$ are precisely those that correspond to newforms f such that $\pi_{f,\ell}$ are of the prescribed kind that is described by τ or τ'. Let $U' := \mathrm{GL}_2(\mathbb{Z}_\ell)U^{(\ell)}$, and consider the morphism $\pi\colon Y_U \to Y_{U'}$. Then one has $\mathrm{H}^1_!(Y_U, \mathcal{F}_\sigma) = \mathrm{H}^1_!(Y_{U'}, \pi_*\mathcal{F}_\sigma)$. In order to get information on the corresponding $\overline{\rho}_\ell$'s, one now reduces the sheaf modulo ℓ, i.e, one chooses a $\overline{\mathbb{Z}}_\ell$-lattice for σ, and one reduces modulo the maximal ideal. By construction, the Jordan-Hölder constituents of this reduction are of the form $\mathcal{F}_{n,m} := \mathrm{Sym}^n(\overline{\mathcal{F}}) \otimes \det(\overline{\mathcal{F}})^{\otimes m}$, with $0 \leq n < \ell$ and m in $\mathbb{Z}/(\ell-1)\mathbb{Z}$, and with $\overline{\mathcal{F}}$ the locally constant sheaf of two-dimensional $\overline{\mathbb{F}}_\ell$-vector spaces given by the standard representation of $\mathrm{GL}_2(\mathbb{F}_\ell)$, or, if one wishes, by the dual of the ℓ-torsion of the universal elliptic curve. Moreover, the (n,m) that occur, and their multiplicities, can be computed by representation theory (Brauer characters). But now one can use the results of Deligne and Fontaine stated in [29, Theorems 2.5 and 2.6] that describe the $\overline{\rho}_{f,\ell}$ for newforms of prime to ℓ level, and weight between 2 and $\ell+1$, in order to see which $\overline{\rho}$ occur in the $\mathrm{H}^1_!(Y_{U'}, \mathcal{F}_{n,0})$. Since $\det(\overline{\mathcal{F}})$ is simply $\overline{\mathbb{F}}_\ell(\varepsilon)$, $\mathrm{H}^1_!(Y_{U'}, \mathcal{F}_{n,m})$ is just $\mathrm{H}^1_!(Y_{U'}, \mathcal{F}_{n,0})(\varepsilon^{-m})$. Let us note that in the case of an extended type, one also has to deal with the automorphism w_ℓ; this is done in [4, §1.4]. The role played by the $\mathcal{F}_{n,m}$ explains that the dependence upon $\overline{\rho}$ of "τ (resp. τ') admits $\overline{\rho}$" is via its properties that determine the weight that Serre has attached to $\overline{\rho}|_{I_\ell}$ (see [29, Sections 2–4]).

Having seen this, we can understand what goes on behind the definition of the notion of "τ (resp. τ') admits $\overline{\rho}$" in [4, §1.3]: this means that there exist newforms f of that type such that ρ_f is a lift of $\overline{\rho}$ of the required type. What is harder to understand, is what is behind the corresponding two notions "simply admits", because [4] defines this by simply listing all elements of this relation. A reasonable guess seems that this condition means that $\overline{\pi_*\mathcal{F}_\sigma}$ has exactly one Jordan-Hölder constituent that can give rise to $\overline{\rho}^\vee$. (In fact, the freeness results in Theorems 4.1 and 4.2 and the definition of the Hecke modules that are used imply that in the situations where these theorems can be applied, $\overline{\pi_*\mathcal{F}_\sigma}$ has exactly one Jordan-Hölder constituent that can give rise to $\overline{\rho}^\vee$.) It would be interesting to know how much of the relation between $\pi_{f,\ell}$ and $\overline{\rho}_{f,\ell}$

can be computed using Fontaine's functors. Let us now state this Conjecture 1.3.1, and the two main theorems (1.4.1 and 1.4.2) of [4].

CONJECTURE 5.4. — *Let k be a finite subfield of $\overline{\mathbb{F}}_\ell$, $\overline{\rho}\colon G_\ell \to \mathrm{GL}_2(k)$ a continuous representation, τ an ℓ-type and τ' an extended ℓ-type with irreducible restriction to I_ℓ. Suppose that the centraliser of the image of $\overline{\rho}$ is k and that the image of τ is not contained in the center of $\mathrm{GL}_2(\overline{\mathbb{Q}}_\ell)$.*

(1) *τ (resp. τ') admits $\overline{\rho}$ if and only if $R^l_{O,\tau} \neq \{0\}$ (resp. $R^l_{O,\tau'} \neq \{0\}$), i.e., if and only if there is a finite extension K' of $\mathbb{Q}_\ell$ in $\overline{\mathbb{Q}}_\ell$ and a continuous representation $\rho\colon G_\ell \to \mathrm{GL}_2(O_{K'})$ which reduces to $\overline{\rho}$ and has type τ (resp. has extended type τ').*

(2) *τ (resp. τ') simply admits $\overline{\rho}$ if and only if τ (resp. τ') is acceptable for $\overline{\rho}$.*

THEOREM 5.5. — *Let $\ell > 2$ be prime, K a finite extension of $\mathbb{Q}_\ell$ in $\overline{\mathbb{Q}}_\ell$ and k its residue field. Let $\rho\colon G_\mathbb{Q} \to \mathrm{GL}_2(K)$ be an odd continuous representation, ramified at only finitely many primes. Assume that its reduction $\overline{\rho}\colon G_\mathbb{Q} \to \mathrm{GL}_2(k)$ is absolutely irreducible after restriction to the quadratic subfield of $\mathbb{Q}(\mu_\ell)$, and is modular. Further, suppose that:*

– *$\overline{\rho}|_{G_\ell}$ has centraliser k;*
– *$\rho|_{G_\ell}$ is potentially Barsotti-Tate with ℓ-type τ (resp. with extended ℓ-type τ');*
– *τ (resp. τ') admits $\overline{\rho}$;*
– *τ (resp. τ') is weakly acceptable for $\overline{\rho}$.*

Then ρ is modular.

The proof of this theorem is very parallel to the proof of [14, Theorem 7.1.1], and is just written in terms of the changes to make. The strategy is of course the same as Wiles', especially in the way that we have described it, but now one imposes, at all primes where this is required (i.e., ℓ and the so-called vexing primes of [23]), these finer restrictions to define the right notion of minimally ramified deformations. The commutative algebra that is used consists of the results that we have described in Section 4. The required Hecke modules are constructed as cohomology groups of sheaves on modular curves just as the $\mathcal{F}_\sigma$ above, with the difference that one will also have types τ_p at some primes $p \neq \ell$.

Of course, in order to apply this theorem, one has to prove that the last condition holds, i.e., that there exists a surjection $O[[X]] \to R^l_{O,\overline{\rho}}$. This condition has indeed been proved in sufficiently many cases in order to prove Theorem 1.4, by Conrad in [13] for tamely ramified types with small image, and in [4] for some more types, using Breuil's work. We will come back to this question of proving weak acceptability in the next section.

Let us now see what is still required in order to prove the theorem that all elliptic curves E over $\mathbb{Q}$ are modular. The proof of this is divided into three cases:

(1) $\overline{\rho}_{E,5}|_{\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}(\sqrt{5}))}$ is irreducible;
(2) $\overline{\rho}_{E,5}|_{\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}(\sqrt{5}))}$ is reducible, but $\overline{\rho}_{E,3}|_{\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}(\sqrt{-3}))}$ is absolutely irreducible;

(3) the remaining cases.

First of all, the last case corresponds to rational points on a few modular curves of small level. It is proved in [14, Lemma 7.2.3], with help of Elkies, that the set of all such elliptic curves has, up to isogeny and twist, just three elements, which are known to be modular by calculations. Let us consider the second case. Then E acquires semistable reduction over a tame extension of $\mathbb{Q}_3$ because $\overline{\rho}_{E,5}(I_3)$ has order dividing $(5-1)^2 5$. If a quadratic twist E' of E is semistable at 3, one switches to E', and one is in the situation considered in [24]. If not, then any ramified quadratic twist of E_K with K a ramified quadratic extension of $\mathbb{Q}_3$ has good reduction, so that one can use [13, Theorem 4.2.1]. Let us finally consider the first case. In this case, Theorem 1.4 says that $\overline{\rho}_{E,5}$ is modular. Moreover, since $5 > 3$, E acquires semi-stable reduction over a tame extension of $\mathbb{Q}_5$ of degree dividing 4 or dividing 6; in the first case, where E is potentially ordinary at 5, Theorem 5.3 of [24] applies, in the second case, there is a ramified extension K of $\mathbb{Q}_5$ of degree 3, such that a ramified quadratic twist of E_K has good, supersingular reduction, and [13, Theorem 4.2.1] applies.

So it remains to explain how Theorem 1.4 is proved. Let $\overline{\rho}$ be as in that theorem. One first twists $\overline{\rho}$ by a suitable quadratic character, such that $\overline{\rho}|_{G_3}$ falls into one of the 6 cases of [4], page 3, whose Artin conductors at 3 are 3^i, $0 \leq i \leq 5$. Then one considers elliptic curves E over $\mathbb{Q}$ such that $\overline{\rho}_{E,5}$ is isomorphic to $\overline{\rho}$. The moduli space of these is the union of two non-empty open subschemes of $\mathbb{P}^1_{\mathbb{Q}}$, hence there are plenty of such E. Using Hilbert irreducibility, and some computations by Manoharmayum [37], one can show that there exists such an E such that $\overline{\rho}_{E,3}$ is surjective on $\mathrm{GL}_2(\mathbb{F}_3)$, and such that, in the cases of conductor 3^i with $i \geq 3$, $\rho_{E,3}$ is such that Theorem 5.5 above can be applied to it, i.e., such that the type, or extended type, can be proved to be weakly acceptable for $\overline{\rho}_{E,3}$. These results are proved in [4, §2.1]. The use of an extended type is required only in the case of conductor 3^5.

6. DEFORMATION PROBLEMS AT ℓ

Let us now discuss the results concerning weak acceptability, obtained in [13], [14], and in Sections 4–9 of [4]. We recall what this means. Let $\ell > 2$ be prime, K a finite extension of $\mathbb{Q}_\ell$, with ring of integers O and residue field k. Let $\overline{\rho}$ be a two-dimensional representation of G_ℓ over k, with centraliser k. Let τ be an ℓ-type, and τ' an extended ℓ-type with irreducible restriction to I_ℓ. Then the quotients $R^l_{O,\tau}$ and $R^l_{O,\tau'}$ of the universal deformation ring R^l_O of $\overline{\rho}$ have been defined in the previous section, by taking the Zariski closures in $\mathrm{Spec}(R^l_O)$ of the sets of ℓ-adic lifts of type τ (resp. extended type τ'). Weak admissibility of τ (resp. τ') just means that there exists a surjection $O[[X]] \to R^l_{O,\tau}$ (resp. $O[[X]] \to R^l_{O,\tau'}$). So what one wants to compute is the dimension over k of the space of deformations over $k[t]/(t^2)$ of $\overline{\rho}$ that are weakly of type τ (or extended type τ'). But the way that these kind of deformations have been defined makes this impossible. So, in order to deal with this problem, one defines other deformation problems, whose universal deformation rings surject to the ones above, and for which one then proves that they admit a surjection from $O[[X]]$. The aim of this section is just to describe these new deformation problems, and to sketch the tools that are used in their study. Let us mention that there seems to be some hope that one can deal directly with the rings $R^l_{O,\tau}$ and $R^l_{O,\tau'}$ (see the beginning of [4, §4]). In what follows, we drop the superscript ℓ from the notation, as we are only considering representations of G_ℓ.

We follow [4, §4]. We first discuss some generalities, and then what happpens in the worst case, i.e., the conductor 3^5 case. Let F be a finite Galois extension of $\mathbb{Q}_\ell$, with group Γ. We let R be the ring of integers in F, and k its residue field (if we need to refer to the residue field of O, we will call it k_O). For $\mathcal{G}$ a finite free group scheme over R, of ℓ-power order, we let $\mathbb{D}(\mathcal{G})$ denote the contravariant Dieudonné module of $\mathcal{G}_k$; it is a $W(k)$-module, equipped with operators $\mathbf{F}$ and $\mathbf{V}$, such that $\mathbf{FV} = \mathbf{VF} = \ell$ and, for all x in $W(k)$: $\mathbf{F}x = \mathrm{Frob}_\ell(x)\mathbf{F}$. A *descent datum* for a finite free group scheme $\mathcal{G}$ over R is a right action of Γ on $\mathcal{G}$, compatible with its action on $\mathrm{Spec}(R)$, i.e., for each γ in Γ, one has a commutative diagram:

$$\begin{tikzcd} \mathcal{G} \arrow[r, "{[\gamma]}", "\sim"'] \arrow[d] & \mathcal{G} \arrow[d] \\ \mathrm{Spec}(R) \arrow[r, "\sim", "\mathrm{Spec}(\gamma)"'] & \mathrm{Spec}(R) \end{tikzcd}$$

such that $[\gamma_1\gamma_2] = [\gamma_2][\gamma_1]$ for all γ_1 and γ_2 in Γ. Note that, since $\mathbb{Z}_\ell \to R$ may be ramified, this is not what one should call a descent datum; however, it is a descent datum after restriction to F. In particular, we can associate in this way, to a pair $(\mathcal{G}, [\cdot])$, a group scheme over $\mathbb{Q}_\ell$. A descent datum as above gives an action of Γ on $\mathbb{D}(\mathcal{G})$, compatible with the action of Γ on $W(k)$, and commuting with $\mathbf{F}$ and $\mathbf{V}$, i.e., it becomes a module over the ring $W(k)[\mathbf{F}, \mathbf{V}][\Gamma]$ (with suitable commutation relations). The idea is now that to τ and τ', one can associate ideals I and I' of this ring, that will impose the right conditions on ℓ-adic lifts of $\overline{\rho}$ to be of type τ or of extended type τ'.

Let us now describe the kind of deformation problems that are considered. The extension F of $\mathbb{Q}_\ell$ should be taken such that the type τ (or the extended type τ') becomes unramified over it. Then one fixes a *model* over R for $\overline{\rho}$, i.e., a pair $(\mathcal{G}_0, [\cdot])$ as above, giving $\overline{\rho}$ as the module of its $\overline{\mathbb{Q}}_\ell$-points, and such that I (or I') annihilates $\mathbb{D}(\mathcal{G}_0)$. Once a model is chosen, one can consider all deformations ρ of $\overline{\rho}$, say to artinian rings, that admit a model $(\mathcal{G}, [\cdot])$ with $\mathbb{D}(\mathcal{G})$ killed by I (or by I'), with a filtration in which each successive quotient is isomorphic to $(\mathcal{G}_0, [\cdot])$. A nice condition to impose is then that such models are unique, and indeed, this can be realized in the situations that are needed (this is what [4, §4.2] is about). Let us denote by $R_{O,I}$ and $R_{O,I'}$ the universal deformation rings thus obtained.

This is where the generalities end, and where one considers each of the 3 cases (conductor 3^3, 3^4 and 3^5) separately. Let us just describe what happens in the worst case: conductor 3^5. In that case, Conjecture 1.3.1 of [4] suggests 3 extended types τ_i', $i \in \mathbb{Z}/3\mathbb{Z}$. One can take $O = \mathbb{Z}_3$. The restrictions of the τ_i' to $W_{\mathbb{Q}_3(\sqrt{-3})}$ are given by a morphism

$$\mathbb{Q}_3(\sqrt{-3})^* \longrightarrow \mathbb{Q}_3(\zeta)^*, \qquad \begin{cases} \sqrt{-3} & \mapsto & \zeta - \zeta^{-1} \\ -1 & \mapsto & -1 \\ 4 & \mapsto & 1 \\ 1+3\sqrt{-3} & \mapsto & \zeta \\ 1+\sqrt{-3} & \mapsto & \zeta^i \end{cases}$$

with ζ of order 3. This defines abelian extensions F_i of $\mathbb{Q}_3(\sqrt{-3})$ of degree 12, that are Galois over $\mathbb{Q}_3$. The ideals I_i that one uses are all three generated by: $\mathbf{F}+\mathbf{V}$, $[\gamma_4^2]+1$, and $([\gamma_3]-[\gamma_3^{-1}])[\gamma_2]-\mathbf{F}$, with γ_2, γ_3 and γ_4 certain elements of $\mathrm{Gal}(F_i/\mathbb{Q}_3)$. In particular, γ_2 is not in the inertia subgroup of $\mathrm{Gal}(F_i/\mathbb{Q}_3)$, hence the last generator of the I_i reflects that one works with an extended type. Without this condition, or even with a similar condition coming from other possible extended types, one wouldn't expect the tangent space of the deformation problem to be of dimension one.

Theorem 4.6.1 of [4] says that, for each i, there exist four models of $\overline{\rho}$ with the property that models for deformations as above are unique, and, moreover, that each deformation of $\overline{\rho}$ as above has a filtration with all successive quotients isomorphic to one of these four. Theorem 4.6.2 of [4] then says that all four have the property that the universal deformation ring is topologically generated by one element, but three of them have universal deformation ring $\mathbb{F}_3[[X]]$ ([4, Theorem 4.6.3]). Using this, it is finally proved that every 3-adic lift of $\overline{\rho}$ that is of extended type τ_i', say over a finite extension O of $\mathbb{Z}_3$, comes from a morphism $R_O \to R_{O,\tau_i'}$ that factors through the universal deformation ring $R_{O,I_i'}$ associated to this last model. Hence, finally, this proves that each of the τ_i' weakly accepts $\overline{\rho}$.

To finish this section, let us point out that proving the theorems in the preceding paragraph takes about 45 pages in [4], with about 30 of them filled with computations with Breuil's ϕ_1-modules. It is my hope that these notes will encourage readers to take a look at those pages, and understand what is going on. In order to say at least something about these modules, let us give the definition of the ℓ-torsion ϕ_1-modules over R, the category of which is anti-equivalent to that of finite free group schemes over R that are killed by ℓ.

Let R etc. be as in the second paragraph of this section. Let π be a uniformizer of R, and $E_\pi(u) = u^e - lG_\pi(u)$ its minimal polynomial over $W(k)$. Let ϕ denote the ℓ-th power map on $k[u]/(u^{el})$. An ℓ-torsion ϕ_1-module is then a triple (M, M_1, ϕ_1) with M

a finite free $k[u]/(u^{el})$-module, with M_1 a $k[u]/(u^{el})$-submodule containing u^eM, and with $\phi_1\colon M_1 \to M$ ϕ-semilinear, such that $\phi_1(M_1)$ generates M as $k[u]/(u^{el})$-module.

We note that the category just described does not depend on the choice of the uniformizer π, but that the functors giving the anti-equivalence mentioned above do depend on that choice. This fact causes a lot of trouble in [4], as one has to study the action of $\mathrm{Gal}(F/\mathbb{Q})$ on models $\mathcal{G}$ over $\mathrm{Spec}(R)$, and, of course, $\mathrm{Gal}(F/\mathbb{Q})$ will not fix the choice of π.

7. TWO RELATED RESULTS

The aim of this section is to briefly state two modularity results on two-dimensional Galois representations that are obtained by others than those mentioned in the title. They can be found in [51] and [7].

Theorem 7.1 (Skinner, Wiles). — *Let $\ell > 2$ be prime, and let $\rho\colon G_{\mathbb{Q}} \to \mathrm{GL}_2(\overline{\mathbb{Q}}_\ell)$ be an odd continuous representation, ramified at only finitely many primes, with $\det(\rho) = \psi\varepsilon^{k-1}$ with ψ of finite order, $\varepsilon\colon G_{\mathbb{Q}} \to \mathbb{Z}_\ell^*$ the cyclotomic character and $k \geq 2$ an integer. Suppose that $\rho|_{G_\ell}$ is the extension of a character ψ_2 by a character ψ_1 with $\psi_2|_{I_\ell}$ of finite order and such that $\overline{\psi_1}|_{G_\ell} \neq \overline{\psi_2}|_{G_\ell}$. If $\overline{\rho}\colon G_{\mathbb{Q}} \to \mathrm{GL}_2(\overline{\mathbb{F}}_\ell)$ is irreducible, then suppose that $\overline{\rho}$ is modular. Then ρ is modular.*

In fact, they even prove a stronger result, with $\mathbb{Q}$ replaced by an arbitrary totally real field. This generality is actually even necessary for the theorem above, since the proof involves changing the field $\mathbb{Q}$. Let us note that the representations ρ considered by Skinner and Wiles are very different from those considered by Breuil, Conrad, Diamond, and Taylor, which mostly have irreducible restrictions to any open subgroup of G_ℓ.

Theorem 7.2 (Buzzard, Dickinson, Shepherd-Barron, Taylor)

Let ρ be a continuous, irreducible, odd representation from $G_{\mathbb{Q}}$ to $\mathrm{GL}_2(\mathbb{C})$ with unsolvable image. Suppose that ρ is unramified at 2 and at 5, and that the image of $\rho(\mathrm{Frob}_2)$ in $\mathrm{PGL}_2(\mathbb{C})$ has order 3. Then ρ is modular.

Let us note that all continuous representations $\rho\colon G_F \to \mathrm{GL}_2(\mathbb{C})$ with solvable image and with F a number field are known to be associated to automorphic representations, by Hecke (in the dihedral case) and Langlands and Tunnell ([36], [61]). The strategy of the proof is explained in [57], and carried out in [8], [54], and [28]. The paper [7] mainly pulls everything together, and provides slight technical but needed improvements of previously obtained results.

In a nutshell, the strategy consists in realizing a suitable twist of ρ over a number field in $\mathbb{C}$, such that it has a reduction $\overline{\rho}$ mod 2 with values in $\mathrm{GL}_2(\mathbb{F}_4)$. Then one shows that $\overline{\rho}$ is modular, that ρ arises from an overconvergent 2-adic modular form, and finally that ρ arises from a weight one form.

8. LATEST NEWS

This section has been added at the time the final version of this text was written (June 2000). Its aim is just to direct the reader to some developments that took place after the lecture (March).

Taylor has released two preprints [58] and [59]. In the first one, he proves, using work of Skinner and Wiles ([51], [52], [53]), and of many other people, the following result.

THEOREM 8.1 (Taylor). — *Let ℓ be an odd prime, and $\rho\colon G_{\mathbb{Q}} \to \mathrm{GL}_2(\overline{\mathbb{Q}}_\ell)$ a continuous irreducible representation such that:*

(1) *ρ is unramified at all but finitely many primes;*

(2) *ρ is odd (i.e., $\det\rho(c) = -1$);*

(3) *$\rho|_{G_\ell}$ is an extension of χ_2 by $\varepsilon^n\chi_1$, with $\chi_1|_{I_\ell}$ and $\chi_2|_{I_\ell}$ of finite order and n a non-zero positive integer, such that $\varepsilon^n\chi_1\chi_2^{-1}(I_\ell)$ is not pro-ℓ.*

Then there is a totally real number field E, a regular algebraic cuspidal automorphic representation π of $\mathrm{GL}_2(\mathbb{A}_E)$ and a place λ of the field of coefficients of π above ℓ such that $\rho_{\pi,\lambda}$ (the λ-adic representation associated to π) is equivalent to $\rho|_{G_E}$.

As a consequence, such a ρ has an L-function, and this L-function is meromorphic and satisfies the expected functional equation. Also, under some mild hypothesis, it follows that ρ occurs in some $\mathrm{H}^i(X_{\overline{\mathbb{Q}}}, \mathbb{Q}_\ell(r))$, as in Conjecture 2.1. The idea that allows one to use the results of Skinner and Wiles, and others, concerning ρ such that $\overline{\rho}$ has a soluble image, is to use abelian varieties with real multiplications such that the $\overline{\rho}$ of the Theorem above is related to the ℓ-torsion, and such that the p-torsion (for some other prime p) gives a suitable soluble image. The existence of such abelian varieties is proved by applying Skolem type results to Hilbert-Blumenthal modular varieties (see for example [38]).

Ramakrishna proved in [43] that, under mild hypotheses, a continuous mod ℓ representation $\overline{\rho}\colon G_{\mathbb{Q}} \to \mathrm{GL}_2(k)$ (not supposed to be modular) can be lifted to a representation $\rho\colon G_{\mathbb{Q}} \to \mathrm{GL}_2(W(k))$ over the Witt vectors of k, with ρ unramified at almost all primes. More recently, in [44], he has proved that one can even obtain that ρ is semi-stable or crystalline at ℓ. His main innovation is to consider deformation problems that lead to a universal deformation ring $W(k)$. He requires $\rho|_{G_p}$ to be of the form $\left(\begin{smallmatrix}\varepsilon & * \\ 0 & 1\end{smallmatrix}\right)$ for suitable p. Note that this is stronger than a condition on $\rho|_{I_p}$ (until now, only conditions on $\rho|_{I_p}$ were imposed), which makes it reasonable that the deformation ring will be small. One finds a slight generalization of Ramakrishna's results in [59], where they are used to prove some more cases of the Artin conjecture.

Combined with Theorem 8.1, one obtains that $\overline{\rho}$ as above becomes modular after restriction to G_E with E a suitable totally real extension of $\mathbb{Q}$ (see [58]). This can be seen as a "potential" version of Serre's conjecture.

Khare has used Ramakrishna's work ([35]) to give another proof of certain "$R = T$" theorems. Assuming $\overline{\rho}$ to be modular, one gets an "$R = T$" theorem for Ramakrishna's deformation problem (the main result of [27] implies that Ramakrishna's lift ρ is modular). Starting with this result, Khare proves that for large enough Σ, $R_{O,\Sigma} \to \mathbb{T}_{O,\Sigma}$ is an isomorphism; his proof avoids the special arguments of Wiles and Taylor-Wiles in the minimal case. Of course, this last result suffices for proving modularity results. (It seems that from this one also easily obtains the result for all Σ.)

Breuil and Mézard have released a preprint ([5]) in which they give a conjectural description of the Samuel multiplicity of local deformation rings $R^l_{O,\tau}$ in automorphic terms. They also prove this conjecture in many cases.

APPENDIX A

GALOIS REPRESENTATIONS ASSOCIATED TO MODULAR FORMS

The aim of this section is to recall what we need about the Galois representations associated to modular forms. For simplicity, we only discuss the case of forms of weight two, so that we only need to deal with the cohomology of the constant sheaf on modular curves. We use the now standard point of view that was initiated by Deligne in [18]. As a general reference for this section, we recommend [26].

The object from which everything originates here is the Shimura datum $(\mathrm{GL}_2, \mathbb{H}^{\pm})$, with $\mathrm{GL}_2(\mathbb{R})$ acting on $\mathbb{H}^{\pm} = \mathbb{P}^1(\mathbb{C}) - \mathbb{P}^1(\mathbb{R})$ in the usual way. Let $\mathbb{A}$ denote the ring of adèles of $\mathbb{Q}$, and $\mathbb{A}_{\mathrm{f}}$ its subring of the finite adèles. For every compact open subgroup U of $\mathrm{GL}_2(\mathbb{A}_{\mathrm{f}})$, let $X^0_U(\mathbb{C})$ denote the complex analytic variety $\mathrm{GL}_2(\mathbb{Q})\backslash(\mathbb{H}^{\pm} \times \mathrm{GL}_2(\mathbb{A}_f)/U)$; it can be compactified, by adding a finite number of points, to a smooth compact Riemann surface (usually not connected) $X_U(\mathbb{C})$. We denote the associated complex algebraic curve by $X_{U,\mathbb{C}}$. The interpretation of these curves as moduli spaces of elliptic curves with level structures give models $X_{U,\mathbb{Q}}$ over $\mathbb{Q}$. The inverse limit $X_{\mathbb{Q}}$ of the $X_{U,\mathbb{Q}}$ has an action, from the right, by $\mathrm{GL}_2(\mathbb{A}_f)$. For ℓ prime, the $\overline{\mathbb{Q}}_\ell$-vector space:

$$H_\ell := \lim_{\substack{\longrightarrow \\ U}} \mathrm{H}^1(X_{U,\overline{\mathbb{Q}},\mathrm{et}}, \overline{\mathbb{Q}}_\ell)$$

has an action by $G_{\mathbb{Q}} \times \mathrm{GL}_2(\mathbb{A}_{\mathrm{f}})$. In order to understand the decomposition of H_ℓ as a representation of $\mathrm{GL}_2(\mathbb{A}_{\mathrm{f}})$, one uses the Hodge decomposition:

$$\mathrm{H}^1(X_U(\mathbb{C}), \mathbb{C}) = \Omega^1(X_U(\mathbb{C})) \oplus \overline{\Omega^1(X_U(\mathbb{C}))}.$$

On $\Omega^1(X_U(\mathbb{C}))$ one no longer has an action of $\mathrm{GL}_2(\mathbb{A}_\mathrm{f})$, but it still is a module over the Hecke algebra associated to U: the convolution algebra of compactly supported bi-U-invariant functions on $\mathrm{GL}_2(\mathbb{A}_\mathrm{f})$ (say that $\mathrm{GL}_2(\hat{\mathbb{Z}})$ has measure one). The q-expansion principle and some theory of smooth irreducible representations of the $\mathrm{GL}_2(\mathbb{Q}_p)$ show that $\Omega^1(X(\mathbb{C}))$ decomposes into a direct sum of irreducible ones, each one occurring only once:

$$\Omega^1(X(\mathbb{C})) = \lim_{\substack{\longrightarrow \\ U}} \Omega^1(X_U(\mathbb{C})) = \bigoplus_f \pi_f,$$

where f runs through the set of newforms of weight two with coefficients in $\mathbb{C}$. We recall that for all p we have a chosen embedding $\overline{\mathbb{Q}} \to \overline{\mathbb{Q}}_p$. It follows that H_ℓ decomposes as a direct sum:

$$H_\ell \cong \bigoplus_f \rho_f^\vee \otimes \pi_f,$$

with f running through the set of weight two newforms with coefficients in $\overline{\mathbb{Q}}_\ell$, and with $\rho_f\colon G_\mathbb{Q} \to \mathrm{GL}_2(\overline{\mathbb{Q}}_\ell)$ a continuous representation. We note that ρ_f is realized over any finite extension over which π_f is defined. The representation π_f is a restricted tensor product $\otimes'_p \pi_{f,p}$ over all primes, with each π_p an infinite dimensional irreducible admissible representation of $\mathrm{GL}_2(\mathbb{Q}_p)$. On the Galois side, we define, for each prime p, $\rho_{f,p} := \rho_f|_{G_p}$. With these definitions, one knows that, for $p \neq \ell$, $\pi_{f,p}$ and $\rho_{f,p}$ determine each other via a suitably normalized local Langlands correspondence. (This was first proved at the unramified places by Eichler and Shimura, then for $\pi_{f,p}$ principal series or special by Langlands, then for $p \neq 2$ by Deligne, and finally for all p by Carayol, and simplified by Nyssen.) The representation $\rho_{f,\ell}$ is usually not determined by $\pi_{f,\ell}$ (just think of the case where f corresponds to an elliptic curve with split multiplicative reduction at ℓ, where $\rho_{f,\ell}$ almost determines the elliptic curve), but Saito has shown in [47] that the (ϕ, N, G_ℓ)-module obtained by forgetting the filtration of the filtered (ϕ, N, G_ℓ)-module corresponding to $\rho_{f,\ell}$ via Fontaine's functor ([32, §10]) corresponds to $\pi_{f,\ell}$. (Actually, in the weight two case that we consider this is in fact easily deduced from the results for $p \neq \ell$; see [14, Appendix B].)

In order to fix notation, let us give a precise description of this local correspondence, so that there is no ambiguity about the normalization. To do this, we first recall that the best way to formulate the local Langlands correspondence is in terms of the Weil-Deligne group (see [56, §4] and [19]). For p prime, the Weil group W_p of $\mathbb{Q}_p$ is the subgroup of G_p consisting of elements whose image in $G_{\mathbb{F}_p}$ is in $\mathrm{Frob}_p^\mathbb{Z}$ ($\mathrm{Frob}_p\colon x \mapsto x^p$ is the arithmetic Frobenius). The Weil-Deligne group is an object W'_p that is defined so that a representation of W'_p on a finite dimensional E-vector space ($E \supset \mathbb{Q}$) is a pair (V, N) with V a continuous representation of W_p (with the discrete topology on V), and a nilpotent endomorphism N of V such that $wNw^{-1}v = pNv$ for all v in V and w in W_p mapping to Frob_p. Such a pair is called F-semisimple if V is semisimple as a representation of W_p. With these definitions, there are canonical bijections between the set of (isomorphism classes of) infinite dimensional irreducible admissible representations of $\mathrm{GL}_2(\mathbb{Q}_p)$ over $\overline{\mathbb{Q}}$, and the set of 2-dimensional F-semisimple representations of W'_p over $\overline{\mathbb{Q}}$. These bijections are such, that local L and ε-factors on both sides (suitably normalized) correspond, and they are compatible with the action of $G_{\mathbb{Q}}$ on both sides. For $p \neq 2$, this is easy, since one can easily write down the elements on both sides (on the Galois side, one uses that the wild inertia acts reducibly). For $p = 2$, this is harder; the general case was worked out by Kutzko. If V is a finite dimensional K-vector space, with K a finite extension of $\mathbb{Q}_\ell$, and $p \neq \ell$, there is an equivalence between representations of W'_p on V as above, and continuous representations of W_p on V.

Following [14], we normalize the local Langlands correspondence WD in such a way that $\rho_{f,p}|_{W_p}$, viewed as a representation of W'_p over $\overline{\mathbb{Q}}_\ell$, is isomorphic to $\overline{\mathbb{Q}}_\ell \otimes_{\overline{\mathbb{Q}}} \mathrm{WD}(\pi_{f,p})$, for each newform f with coefficients in $\overline{\mathbb{Q}}$. If $\sigma(\pi_{f,p})$ is as in [9], then we have $\mathrm{WD}(\pi_{f,p}) = \sigma(\pi_{f,p}) \otimes \chi$, where χ is the unramified character that sends Frob_p to p. If $\pi_{f,p}$ is unramified, i.e., if p does not divide the level of f, then $\rho_{f,p}$ is unramified and $\rho_{f,p}(\mathrm{Frob}_p)$ is semisimple (remember that we are in weight two) and has characteristic polynomial $X^2 - t_pX + ps_p$, where t_p and s_p are the eigenvalues of f for the Hecke and diamond operators T_p and S_p that are defined by the double cosets $U\left(\begin{smallmatrix} p & 0 \\ 0 & 1 \end{smallmatrix}\right)U$ and $U\left(\begin{smallmatrix} p & 0 \\ 0 & p \end{smallmatrix}\right)U$, with $U = \mathrm{GL}_2(\mathbb{Z}_p)$. The determinant of $\rho_{f,p}$ is $\varepsilon\chi_{\pi_{f,p}}$, where $\chi_{\pi_{f,p}}$ is the central character of $\pi_{f,p}$, viewed as a character of W_p^{ab} via the isomorphism of class field theory under which the image of p in $\mathbb{Q}_p^*/\mathbb{Z}_p^*$ corresponds to Frob_p. If χ is a continuous character $\mathbb{Q}_p^* \to \overline{\mathbb{Q}}^*$, then $\mathrm{WD}(\pi_{f,p} \otimes (\chi \det)) \cong \mathrm{WD}(\pi_{f,p}) \otimes \chi$.

To finish this section, let us recall some facts about the classification of the two-dimensional semisimple representations of W_p, over $\mathbb{C}$, say. Let ρ be such a representation, on a $\mathbb{C}$-vector space V, say. If ρ is reducible, it is a sum of two characters. Suppose ρ irreducible. Since the wild inertia subgroup I_p^{wild} of I_p acts on V via a finite p-group, it acts via two characters (possibly equal), unless $p = 2$ (recall that the dimension of an irreducible complex representation of a finite group divides the order of the group). Suppose that $\rho|I_p^{\mathrm{wild}}$ splits as a sum of two distinct characters. Then W_p acts on the set of the two corresponding lines in V, and non-trivially because ρ is irreducible. It follows that ρ becomes reducible over a quadratic extension of $\mathbb{Q}_p$. Suppose now that I_p^{wild} acts via scalars on V. Then, considering the action of W_p/I_p (which is the semi-direct product of $\mathbb{Z}$ by I_p^{tame}) on $\mathbb{P}(V)$, one sees that, again, ρ becomes reducible over a quadratic extension of $\mathbb{Q}_p$. So the conclusion is this: the two-dimensional complex semisimple representations of W_p are sums of two characters, or induced from a character of W_K with K quadratic over $\mathbb{Q}_p$, or such that $p = 2$ and I_p^{wild} acts irreducibly (these latter ones were first classified by Weil [63], in his "exercices dyadiques"; clearly, the title of [4] is inspired by this).

APPENDIX B

TYPES ASSOCIATED TO ℓ-ADIC REPRESENTATIONS AND TO ELLIPTIC CURVES

Let ℓ be any prime. We recall that an extended ℓ-type (over $\overline{\mathbb{Q}}_\ell$) is an isomorphism class of two-dimensional representations over $\overline{\mathbb{Q}}_\ell$ of the Weil-Deligne group W'_ℓ of $\mathbb{Q}_\ell$ (see Appendix A), and that types are isomorphism classes of restrictions to I_ℓ of extended ℓ-types. We want to describe how one attaches an extended type to a continuous representation $\rho\colon G_\ell \to \mathrm{GL}_2(O)$, with O the ring of integers of a finite extension K of $\mathbb{Q}_\ell$ contained in $\overline{\mathbb{Q}}_\ell$, under the assumption that ρ is potentially Barsotti-Tate. So let ρ be such a representation, and let F be a finite extension of $\mathbb{Q}_\ell$ over which ρ becomes Barsotti-Tate, i.e., such that $\rho|_{G_F}$ is isomorphic to $\mathcal{G}(\overline{\mathbb{Q}}_\ell)$ for some ℓ-divisible group with O-action over the ring of integers O_F of F. Of course, one solution to this is simply to apply Fontaine's $D_{\mathrm{st},F}$ functor as in [32, §10(b)], but in this simple case of p-divisible group schemes one can be more explicit. Another reason for doing this more explicitly is that one wants to do computations in the case of elliptic curves. For more details we refer to [14, Appendix B].

The representation ρ we have corresponds to an ℓ-divisible group $\mathcal{G}_{\mathbb{Q}_\ell}$ over $\mathbb{Q}_\ell$, with O-action. Let F be a finite Galois extension of $\mathbb{Q}_\ell$ such that $\mathcal{G}_F$ extends (uniquely, by [55, Theorem 4]) to an ℓ-divisible group $\mathcal{G}_{O_F}$ over O_F, with O-action. Let Γ denote $\mathrm{Gal}(F/\mathbb{Q}_\ell)$. For every σ in Γ, we have commutative diagrams:

$$\begin{array}{ccc} \mathcal{G}_{O_F} & \xrightarrow[\sim]{[\sigma]} & \mathcal{G}_{O_F} \\ \downarrow & & \downarrow \\ \mathrm{Spec}(O_F) & \xrightarrow[\mathrm{Spec}(\sigma)]{\sim} & \mathrm{Spec}(O_F) \end{array} \qquad \begin{array}{ccc} \mathcal{G}_{k_F} & \xrightarrow[\sim]{[\sigma]} & \mathcal{G}_{k_F} \\ \downarrow & & \downarrow \\ \mathrm{Spec}(k_F) & \xrightarrow[\mathrm{Spec}(\sigma)]{\sim} & \mathrm{Spec}(k_F) \end{array}$$

with k_F the residue field of O_F. The last diagram gives a right action of G_ℓ on $\mathcal{G}_{k_F}/k_F$. Let $d\colon W_\ell \to \mathbb{Z}$ be the morphism such that σ in W_ℓ induces the $d(\sigma)$th power of the absolute Frobenius on $\overline{\mathbb{F}}_\ell$. Then we get a morphism from W_ℓ to $(\mathbb{Q} \otimes \mathrm{End}_{k_F}(\mathcal{G}_{k_F}))^*$ by sending σ to $[\sigma]^{-1}\mathrm{Frob}_{\mathrm{abs}}^{d(\sigma)}$. Now let $\mathbb{D}$ denote the contravariant Dieudonné module functor. Then $\mathbb{Q} \otimes \mathbb{D}(\mathcal{G}_{k_F})^\vee$, with $^\vee$ denoting $W(k_F)$-dual, is a free $K \otimes_{\mathbb{Z}_\ell} W(k_F)$-module of rank two, with a left action by W_ℓ. The extended ℓ-type $\mathrm{WD}(\rho)$ associated to ρ is then the two-dimensional $\overline{\mathbb{Q}}_\ell$ vector space obtained by base change via $K \otimes_{\mathbb{Z}_\ell} W(k_F) \to \overline{\mathbb{Q}}_\ell$ (note that both K and F are subfields of $\overline{\mathbb{Q}}_\ell$); the monodromy operator is defined to be zero, as we have good reduction.

We can repeat the construction above, with $\mathcal{G}_{\mathbb{Q}_\ell}$ replaced by an elliptic curve E over $\mathbb{Q}_\ell$, with good reduction E_{O_F} over O_F. Then one gets morphisms:

$$W_\ell \longrightarrow (\mathbb{Z}[1/\ell] \otimes \mathrm{End}_{k_F}(E_{k_F}))^* \longrightarrow (\overline{\mathbb{Q}} \otimes \mathrm{End}_{k_F}(E_{k_F}))^* \cong \mathrm{GL}_2(\overline{\mathbb{Q}}).$$

More generally, one can start with a newform f with coefficients in $\overline{\mathbb{Q}}$, and then one has:

$$\mathrm{WD}(\rho_{f,\ell}) = \overline{\mathbb{Q}}_\ell \otimes_{\overline{\mathbb{Q}}} \mathrm{WD}(\pi_{f,\ell}),$$

by the results of Appendix A.

In [62] one can find a complete description of all ρ_ℓ that arise from elliptic curves over $\mathbb{Q}_\ell$, in terms of their associated filtered (ϕ, N, G_ℓ)-modules.

Acknowledgements

I would like to thank Christophe Breuil, Fred Diamond, Reinie Erné, Eyal Goren, Chandrashekhar Khare, Rutger Noot and Richard Taylor for their comments on an earlier version of this text, and Michel Gros for the reference to [6]. Attending the conference "Modularity of Elliptic Curves and Beyond" at the MSRI in Berkeley, December 6–10, 1999, was very useful, as well as the series of talks given by Brian Conrad on this subject in Rome, July 1999.

REFERENCES

[1] P. Berthelot – *Altérations de variétés algébriques.* Séminaire Bourbaki, 1995–96, exposé 815, Astérisque **241**, S.M.F. (1997), 273–311.

[2] C. Breuil – *Schémas en groupes et modules filtrés.* C. R. Acad. Sci. Paris **328** (1999), 93–97.

[3] C. Breuil – *Groupes p-divisibles, groupes finis et modules filtrés.* Ann. of Math. (2) **152** (2000), no. 2, 489–549. Mathematics.

[4] C. Breuil, B. Conrad, F. Diamond, R. Taylor – *On the modularity of elliptic curves over $\mathbb{Q}$: wild 3-adic exercises.* J. Amer. Math. Soc. **14** (2001), no. 4, 843–939.

[5] C. Breuil, A. Mezard – *Multiplicités modulaires et représentations de $\mathrm{GL}_2(\mathbb{Z}_p)$ et de $\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)$ en $\ell = p$.* Preprint, June 2000, Orsay.

[6] C.J. Bushnell – *Smooth representations of p-adic groups: the role of compact open subgroups.* Proceedings of the international congress of mathematicians, Vol. 1, 2 (Zürich, 1994), 770–779, Birkhaüser, Basel, 1995.

[7] K. Buzzard, M. Dickinson, N. Shepherd-Barron, R. Taylor – *On icosahedral Artin representations.* Duke Math. J. **109** (2001), no. 2, 283–318.

[8] K. Buzzard, R. Taylor – *Companion forms and weight one forms.* Ann. of Math. (2) **149** (1999), no. 3, 905–91

[9] H. Carayol – *Sur les représentations ℓ-adiques associées aux formes modulaires de Hilbert.* Ann. Sci. Éc. Norm. Sup. **19** (1986), 409–468.

[10] L. Clozel – *Motifs et formes automorphes.* Automorphic forms, Shimura varieties, and L-functions, Vol. I, edited by L. Clozel and J.S. Milne, Perspectives in Mathematics 10, Academic Press (1990), 77–159.

[11] P. Colmez, J.-M. Fontaine – *Construction des représentations p-adiques semi-stables.* Invent. Math. **140** (2000), no. 1, 1–43.

[12] B. Conrad – *Finite group schemes over bases with low ramification.* Compositio Math. **119** (1999), 239–320.

[13] B. Conrad – *Ramified deformation problems.* Duke Math. J. **97** (1999), 439–513.

[14] B. Conrad, F. Diamond, R. Taylor – *Modularity of certain potentially Barsotti-Tate Galois representations.* J.A.M.S. **12** (1999), 521–567.

[15] G. Cornell, J. Silverman, G. Stevens, Editors – *Modular Forms and Fermat's Last Theorem* Springer-Verlag, 1997.

[16] H. Darmon – *A proof of the full Shimura-Taniyama-Weil conjecture is announced.* Notices of the A.M.S., December 1999.

[17] H. Darmon, F. Diamond, R. Taylor – *Fermat's Last Theorem.* Elliptic curves, modular forms and Fermat's last theorem (Hong Kong, 1993). Second edition. International Press, Cambridge, MA, 1997, 2–140.

[18] P. Deligne – *Formes modulaires et représentations ℓ-adiques.* Séminaire Bourbaki, 1968–69, exposé 355, Lecture Notes in Mathematics **179**, Springer (1969), 139–172.

[19] P. Deligne – *Formes modulaires et représentations de* GL(2). Antwerp II, Lecture Notes in Math. **349**, Springer (1973), 55–106.

[20] B. De Smit, H. Lenstra – *Explicit construction of universal deformation rings.* In [15], 313–326.

[21] B. De Smit, K. Rubin, R. Schoof – *Criteria for complete intersections.* In [15], 343–355.

[22] F. Diamond – *The refined conjecture of Serre.* In Elliptic curves, modular forms and Fermat's last theorem (Hong Kong, 1993). International Press, 1995, 22–37.

[23] F. Diamond – *An extension of Wiles' results.* In [15], 475–498.

[24] F. Diamond – *On deformation rings and Hecke rings.* Ann. Math. **144** (1996), 137–166.

[25] F. Diamond – *The Taylor–Wiles construction and multiplicity one.* Invent. math. **128** (1997), 379–391.

[26] F. Diamond, J. Im – *Modular forms and modular curves.* In "Seminar on Fermat's last theorem", Canadian Mathematical Society Conference Proceedings 17, 1995 (V. Kumar Murty, editor).

[27] F. Diamond, R. Taylor – *Lifting modular mod ℓ representations.* Duke Math. J. **74** (1994), no. 2, 253–269.

[28] M. Dickinson – *On the modularity of certain* 2*-adic Galois representations.* Duke Math. J. **109** (2001), no. 2, 319–382

[29] S.J. Edixhoven – *The weight in Serre's conjectures on modular forms.* Invent. math. **109** (1992), 563–594.

[30] J.-M. Fontaine – *Groupes p-divisibles sur les corps locaux.* Astérique **47–48**, S.M.F., 1977.

[31] J.-M. Fontaine – *Représentations p-adiques semi-stables.* In "Périodes *p*-adiques," Astérisque **223**, S.M.F., 1994, p. 113–184.

[32] J.-M. FONTAINE, B. MAZUR – *Geometric Galois representations.* In "Elliptic curves, modular forms and Fermat's Last Theorem" (Hong Kong, 1993), International Press, 1995, pp. 41–78.

[33] K. FUJIWARA – *Deformation rings and Hecke rings in the totally real case.* Preprint.

[34] C. KHARE – *A local analysis of congruences in the* (p,p)*-case. II.* Invent. Math. **143** (2001), no. 1, 129–155.

[35] C. KHARE – *On isomorphisms between deformation and Hecke rings.* Work in progress.

[36] R.P. LANGLANDS – *Base change for* $GL(2)$. Ann. of Math. Studies **96**, Princeton University Press, Princeton, 1980.

[37] J. MANOHARMAYUM – *Pairs of mod* 3 *and mod* 5 *representations arising from elliptic curves.* Thesis, Cambridge University, 1998.

[38] L. MORET-BAILLY – *Groupes de Picard et problèmes de Skolem II.* Ann. Sci. ENS **22** (1989), 181–194.

[39] R. NOOT – *Abelian varieties with* ℓ*-adic Galois representation of Mumford's type* J. Reine Angew. Math. **519** (2000), 155–169

[40] J. OESTERLÉ – *Nouvelles approches du "théorème" de Fermat.* Séminaire Bourbaki, 1987–88, exposé 694, Astérisque **161–162**, S.M.F. (1988), 175–186.

[41] J. OESTERLÉ – *Travaux de Wiles (et Taylor,...), partie II.* Séminaire Bourbaki, 1994–95, exposé 804, Astérisque **237**, S.M.F. (1996), 333–355.

[42] R. RAMAKRISHNA – *On a variation of Mazur's deformation functor.* Compositio Math. **87** (1993), 269–286.

[43] R. RAMAKRISHNA – *Lifting Galois representations.* Invent. math. **138** (1999), 537–562.

[44] R. RAMAKRISHNA – *Deforming Galois representations and the conjectures of Serre and Fontaine-Mazur.* Preprint.

[45] K.A. RIBET, – *On modular representations of* $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ *arising from modular forms.* Invent. math. **100** (1990), 431–476.

[46] K. RUBIN – *Modularity of mod 5 representations.* In [15], 463–474.

[47] T. SAITO – *Modular forms and p-adic Hodge theory.* Invent. math. **129** (1997), 607–620.

[48] J-P. SERRE – *Représentations* ℓ*-adiques.* Kyoto Int. Symposium on Algebraic Number Theory, Japan Soc. for the Promotion of Science (1977), pp. 177–193, (Oeuvres, t. III, pp. 384–400).

[49] J-P. SERRE – *Sur les représentations modulaires de degré* 2 *de* $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$. Duke Math. J. **54** (1987), 179–230.

[50] J-P. SERRE – *Travaux de Wiles (et Taylor,...), partie I.* Séminaire Bourbaki, 1994–95, exposé 803, Astérisque **237**, S.M.F. (1996), 319–332.

[51] C. SKINNER, A. WILES – *Residually reducible representations and modular forms.* Pub. Math. IHES. **89** (1999), 5–126.

[52] C. SKINNER, A. WILES – *Base change and a problem of Serre.* Duke Math. J. **107** (2001), no. 1, 15–25.

[53] C. Skinner, A. Wiles – *Nearly ordinary deformations of irreducible residual representations.* Preprint.

[54] N. Shepherd-Barron, R. Taylor – *Mod* 2 *and mod* 5 *icosahedral representations.* J. Amer. Math. Soc. **10** (1997), 281–332.

[55] J. Tate – *p-divisible groups.* Proceedings of a conference on local fields (Driebergen, 1966), Springer, 1967, 158–183.

[56] J. Tate – *Number theoretic background.* Proceedings of Symposia in Pure Mathematics **33** (1979), part 2, 3–26.

[57] R. Taylor – *Icosahedral Galois representations.* Pacific J. Math.: Special issue: Olga Taussky-Todd: in memoriam (1997), 337–347.

[58] R. Taylor – *Remarks on a conjecture of Fontaine and Mazur.* Preprint, `http://www.math.harvard.edu/`

[59] R. Taylor – *On icosahedral Artin representations II.* Preprint, `http://www.math.harvard.edu/`

[60] R. Taylor, A. Wiles – *Ring-theoretic properties of certain Hecke algebras.* Ann. Math. **141** (1995), 553–572.

[61] J. Tunnell – *Artin's conjecture for representations of octahedral type.* Bull. A.M.S. **5** (1981), 173–175.

[62] M. Volkov – *Les représentations ℓ-adiques associées aux courbes elliptiques sur $\mathbb{Q}_p$.* J. Reine Angew. Math. **535** (2001), 65–101.

[63] A. Weil – *Exercices dyadiques.* Invent. math. **27** (1974), 1–22.

[64] A. Wiles – *Modular elliptic curves and Fermat's Last Theorem.* Ann. Math. **141** (1995), 443–551.

Bas EDIXHOVEN

I.R.M.A.R.
U.M.R. 6625 du CNRS
Université de Rennes I
Campus de Beaulieu
F-35042 RENNES Cedex
E-mail : `edix@maths.univ-rennes1.fr`

Séminaire BOURBAKI Mars 2000
52^{e} année, 1999-2000, n^{o} 872, p. 189 à 206

VARIÉTÉS DE FANO RÉELLES
[d'après C. Viterbo]

par **Viatcheslav KHARLAMOV**

1. INTRODUCTION

1.1. Structures algébriques sur une variété différentiable

Toute variété compacte différentiable est difféomorphe à la partie réelle d'une variété algébrique réelle sans singularités. Ce théorème est l'aboutissement des travaux de H. Seifert, J. Nash, R. Thom, A. Wallace, A. Tognoli, H. King et d'autres. Ici, une variété algébrique réelle non singulière est un ensemble défini dans un espace réel, affine ou projectif au choix, par un système d'équations polynomiales réelles sans aucun point singulier réel (voir [20] et [5] pour les détails et les problèmes encore ouverts ; notons aussi que les démonstrations se font par approximation).

Pour nous, une variété (algébrique) réelle est une variété (algébrique) complexe compacte lisse $M(\mathbb{C})$ munie d'une involution antiholomorphe, appelée *structure réelle* ou *conjugaison complexe*. L'ensemble des points fixes de cette involution est une sous-variété analytique réelle, appelée *partie réelle* de $M(\mathbb{C})$ et notée $M(\mathbb{R})$. Nous souhaitons étudier les relations entre les propriétés topologiques et géométriques de $M(\mathbb{C})$ et $M(\mathbb{R})$. Sur ce sujet, de nombreux résultats sont connus grâce à l'étude du 16-ième problème de Hilbert qui pose la question des restrictions sur la topologie de $M(\mathbb{R})$ imposées par le degré des équations le définissant ; citons en particulier les travaux d'Arnol'd, Rokhlin et leurs successeurs, voir les surveys [41], [37] et [14].

Après résolution des singularités, le théorème de Nash-Tognoli-King montre que, pour toute variété différentiable L, il existe une variété algébrique projective non singulière $M(\mathbb{C})$ stable par la conjugaison complexe et dont L est l'ensemble des points réels.

Quelques remarques concernant ce résultat s'imposent. Ici, la variété L considérée peut avoir un nombre fini quelconque de composantes connexes, alors que $M(\mathbb{C})$ peut toujours être choisie connexe (voir par exemple [36]). Notons aussi que si $M(\mathbb{C})$ est une variété projective munie d'une structure réelle, elle peut être dotée d'un plongement projectif $M(\mathbb{C}) \to P_n(\mathbb{C})$ qui conserve la structure réelle : partant d'un plongement quelconque dans $P_n(\mathbb{C})$, on restreint le plongement de Segre $P_n(\mathbb{C}) \times P_n(\overline{\mathbb{C}}) \to P_{n^2+2n}(\mathbb{C})$ au graphe de la structure réelle de M vu comme sous-variété de $M(\mathbb{C}) \times M(\overline{\mathbb{C}})$. (Notez que la structure réelle sur $P_{n^2+2n}(\mathbb{C})$ qui commute avec le plongement de Segre est équivalente à la structure réelle usuelle).

Fixons dans la suite une variété algébrique réelle dont on note M l'ensemble des points complexes, et L *une composante connexe de la partie réelle*. Il est nécessaire d'avoir choisi dans $\mathbb{C}$ une racine carrée de -1, notée i, ce qui équipe M de l'orientation complexe.

Un voisinage tubulaire de L dans M ne dépend que de L, à difféomorphisme près. En effet, la multiplication par i induit un isomorphisme entre le fibré tangent $T(L)$ de L et son fibré normal $N(L)$ dans M. De plus, les nombres d'Euler de ces fibrés satisfont la formule d'adjonction

$$e(N(L)) = (-1)^{\frac{1}{2}n(n-1)} e(T(L)), \tag{1}$$

ce qui provient du choix de l'orientation complexe sur M et de la relation $e_1 \wedge \cdots \wedge e_n \wedge ie_1 \wedge \cdots \wedge ie_n = (-1)^b e_1 \wedge ie_1 \wedge \cdots \wedge e_n \wedge ie_n$, où $b = \frac{1}{2}n(n-1)$. Notons que dès que l'on a choisi une métrique sur L, le fibré cotangent $T^*(L)$ s'identifie à $T(L)$ puis au voisinage tubulaire de L. L'orientation complexe de M restreinte à $T^*(L)$ donne l'orientation canonique (symplectique) de $T^*(L)$.

Du point de vue analytique complexe, le germe de M le long de L est déterminé par la structure analytique réelle de L. On peut noter que les voisinages tubulaires de L contiennent des tubes réels de Grauert [18]. Ce sont des variétés de Stein difféomorphes à $T^*(L)$; voir l'article de D. Burns [7] qui traite des structures réelles sur les tubes de Grauert.

1.2. Une curieuse conjecture de Nash

Il est naturel de chercher à réaliser les variétés différentiables comme variétés réelles en imposant des restrictions sur le complexifié. C'est suivant cette logique que J. Nash [29] (1952) demanda, tout en émettant des réserves, s'il est possible de réaliser toute variété différentiable compacte comme un ensemble algébrique réel sans singularités ayant un complexifié rationnel (ici, rationnel signifie birationnellement isomorphe, sur $\mathbb{R}$, à l'espace projectif ; notez qu'une variété algébrique réelle peut être rationnelle sur $\mathbb{C}$ sans l'être sur $\mathbb{R}$). En fait cette conjecture avait été réfutée, avant même d'avoir été formulée, par le théorème de Comessatti [9] (1914) de classification des modèles minimaux des surfaces réelles et rationnelles sur $\mathbb{C}$. Il en découle en effet que toute composante orientable de la partie réelle d'une surface réelle rationnelle sur $\mathbb{C}$ est une sphère ou un tore.

On peut en donner une démonstration rapide (c'est la même démarche qui mène aux inégalités d'Arnold [21]). Elle est basée sur (1). Soient M une surface réelle et rationnelle sur $\mathbb{C}$, $f \in H_2(M(\mathbb{C}); \mathbb{R})$ la classe d'homologie réalisée par une composante réelle L orientable et orientée, et h la classe d'homologie de la section hyperplane. La forme d'intersection est négative sur l'orthogonal de h, parce que la surface est rationnelle ($H^2 = H^{1,1}$). De plus $fh = 0$ puisque f et h appartiennent à deux espaces propres différents de l'action induite par la conjugaison complexe, et (1) montre que $f^2 = -\chi(L)$. Donc, $\chi(L) \geq 0$ et $g(L) \leq 1$.

En dimension trois, la conjecture est encore fausse. C'est une conséquence de la classification des modèles minimaux des variétés réelles rationnelles de dimension 3, donnée par J. Kollár [23]. Plus précisement, J. Kollár montre que parmi des composantes connexes des parties réelles de toutes les telles variétés, il n'y a au plus qu'un nombre fini de variétés à courbure sectionnelle strictement négative (voir [23] pour des résultats encore plus précis ; signalons seulement que les composantes orientables sont des sommes connexes $X \# aP_3(\mathbb{R}) \# b(S^1 \times S^2)$ où X est soit une somme connexe des espaces lenticulaires, soit une fibration de Seifert, soit un $S^1 \times S^1$-fibré au-dessus de S^1, soit son $\mathbb{Z}/2$-quotient non ramifié). En dimension supérieure, la conjecture est toujours ouverte.

Notons deux résultats positifs : toute variété compacte de dimension trois est homéomorphe à la partie réelle d'une variété algébrique projective réelle et rationnelle sur $\mathbb{R}$, mais singulière ; et toute telle variété est aussi difféomorphe à la partie réelle d'une variété algébrique sans singularité réelle, rationnelle sur $\mathbb{R}$, mais non projective (le premier est dû à R. Benedetti et A. Marin [4] ; le second à J. Kollár [24]). Signalons aussi la solution d'une version purement topologique de la conjecture de Nash en dimension trois par S. Akbulut et H. King [1], ainsi que par R. Benedetti et A. Marin [4], puis en toutes dimensions par G. Mikhalkin [28].

1.3. Variétés fortement de Fano

Les techniques employées par C. Viterbo sont celles de la géométrie symplectique. Par conséquent, dans ce qui suit, nous préférons le langage de la géométrie symplectique à celui de la géométrie algébrique, qui n'interviendra à nouveau qu'à la fin de la démonstration.

Ainsi, *une variété de Fano* est une variété compacte complexe lisse kählérienne M dont la classe de Kähler ω est égale à celle de Chern : $\omega = c_1(M)$ (variété symplectique monotone ou, en d'autres termes, à fibré anticanonique ample ; pour démontrer l'équivalence on peut faire appel au critère de Nakai-Moishezon). Toute variété de Fano M est *uniréglée*, c'est-à-dire qu'il existe deux variétés compactes complexes F et N avec une application régulière $N \to F$ à fibre générique $P_1(\mathbb{C})$, ainsi qu'une application régulière de degré fini $N \to M$. D'autre part, toute variété de Fano est simplement connexe et satisfait $\mathrm{Pic}(M) = H^2(M)$ (voir le rapport de O. Debarre [11] ; ceux qui préfèrent une démonsration moins algébrique de $\pi_1 = 1$ peuvent utiliser la métrique de Calabi-Yau et le théorème de Myers [27]).

Comme exemples de variétés de Fano, on peut citer les hypersurfaces lisses de degré $\leq n$ dans P_n, et plus généralement les intersections complètes lisses de multidegré

$(d_1, \ldots, d_k)$ dans P_n avec $n \geq d_1+\cdots+d_k$. Dans ces exemples, $-K = (n+1-\sum d_i)h$ où K est le diviseur canonique et h le diviseur de la section hyperplane. Par conséquent, si la partie réelle n'est pas vide, la structure réelle se relève dans le fibré associé à h, et se prolonge donc en une structure réelle sur P_n ayant des points réels. Or toute structure réelle sur P_n ayant des points réels est conjuguée à la structure usuelle, donc finalement les seules structures réelles sur les variétés ci-dessus sont les structures induites par celle de P_n (rappelons que lorsque n est impair, il y a des structures réelles sur P_n sans points réels : pour $n = 1$, un exemple est fourni par la conique $x^2 + y^2 + 1 = 0$).

Notons aussi que, si un tube de Grauert est une partie affine d'une variété projective lisse, cette variété projective est de Fano.

Le théorème de Viterbo ne s'applique qu'aux variétés de Fano satisfaisant deux hypothèses supplémentaires. Une variété est dite *minimalement uniréglée*, si elle est uniréglée par des courbes rationnelles dont l'aire est minimale parmi celles des courbes holomorphes de M. Une variété est *fortement de Fano*, si elle est minimalement uniréglée et s'il existe un diviseur ample H de M dont l'intersection avec une courbe rationnelle unirégulatrice est égale à 1.

Ces deux hypothèses sont vérifiées pour les intersections complètes dans P_n de multidegré $(d_1, \ldots, d_k)$ si $d_1 + \cdots + d_k < n$ (qui sont uniréglées par des droites), mais elles excluent le cas $d_1 + \cdots + d_k = n$. En dimension trois, le premier cas non couvert est celui des quartiques dans P_4.

Si M est une variété munie d'une structure réelle conj et d'une métrique de Kähler g, quitte à remplacer g par $\frac{1}{2}(g + g^{\mathrm{conj}})$, on peut se ramener au cas où la forme de Kähler Ω est anti-invariante par l'action de conj^*. Pour cette structure, $M(\mathbb{R})$ est lagrangienne.

Dans tout ce qui suit, les métriques de Kähler seront ainsi choisies.

1.4. Théorème de Viterbo et conjecture de Kollár

THÉORÈME 1.1. — *Aucune composante connexe du lieu réel d'une variété réelle fortemement de Fano n'admet de métrique à courbure sectionnelle strictement négative.*

Ce résultat de C. Viterbo [39] confirme la conjecture, toujours ouverte, de J. Kollár d'après laquelle aucune variété réelle uniréglée *de dimension* ≥ 3 n'a de composantes réelles à courbure sectionnelle strictement négative. (En dimension deux, déjà les éclatés de $P_2(\mathbb{R})$ en au moins deux points portent une métrique hyperbolique.)

D'après J. Kollár [23], la conjecture est vraie en dimension trois sauf pour un nombre fini de classes de variétés à déformation sur $\mathbb{C}$ près. Il pose aussi la question : l'existence d'une métrique hyperbolique implique-t-elle que la variété soit de type général ?

Notons deux résultats intermédiaires, d'intérêt indépendant, obtenus au cours de la démonstration du théorème 1.1 :

PROPOSITION 1.2. — *Si une composante connexe L de la partie réelle d'une variété fortement de Fano réelle admet une métrique sans géodésique contractile, il existe parmi les courbes d'aire minimale et d'intersection* 1 *avec H* (*voir 1.3*) *une courbe réelle (lisse) dont la partie réelle est non vide et contenue dans L.*

PROPOSITION 1.3. — *Soit M une variété fortement de Fano dont les courbes unirégulatrices sont d'aire α. Toute sous-variété lagrangienne L admet une membrane holomorphe à bord dans L et d'aire inférieure à α. Si de plus, L n'a pas de métrique à courbure sectionnelle strictement négative, la membrane peut être choisie d'aire inférieure à $\frac{1}{2}\alpha$.*

La courbe réelle de la proposition 1.2 est une courbe rationnelle régulatrice si H est très ample, comme il l'est, par exemple, dans le cas d'intersections complètes. Par ailleurs, si une telle courbe réelle régulatrice existe, L contient un fermé semi-algébrique d'intérieur non vide couvert par les parties réelles des courbes régulatrices réelles.

Notons aussi que parmi les courbes algébriquement équivalentes à deux fois une courbe régulatrice, il y a toujours, sans aucune hypothèse sur L et pour toute variété de Fano, suffisamment de courbes réelles lisses pour couvrir un ouvert dense de la partie réelle de M (voir [25]).

Si on avait pu étendre 1.1 et 1.2 à toutes les variétés de Fano et toutes les familles régulatrices, ces résultats auraient entraîné les deux corollaires suivants : *Le tube de Grauert* (*voir 1.1*) *d'une variété possédant une métrique à courbure sectionnelle strictement négative n'est jamais la partie affine d'une variété projective lisse. Une variété fortement de Fano réelle a au plus une composante connexe réelle admettant une métrique sans géodésique contractile.*

Dans cet exposé nous expliquerons les grandes lignes de la démonstration sans prétendre entrer dans les détails, en particulier lorsqu'il s'agit des arguments de mise en position générale.

1.5. Classification des surfaces de Fano réelles

Le cas des surfaces est extrêmement simple (et très probablement assez différent du cas des variétés de dimensions supérieures). Comme toutes les variétés de Fano de dimension donnée, les surfaces de Fano réelles (plus traditionnellement appelées sufaces de Del Pezzo), forment un nombre fini de classes à déformation (réelle) près. Leur classification à déformation près découle de celle de Comessatti des surfaces rationnelles, à l'aide des modèles anti(-bi)canoniques des surfaces. Pour les détails de la démonstration, voir [13] (et pour des informations supplémentaires, voir [34]).

Suivant la terminologie traditionnelle, nous appelons *degré* d'une surface de Del Pezzo le nombre $d = c_1^2$. Une *simplification de Morse* est une opération topologique qui consiste à enlever soit une composante sphérique, soit une anse d'une composante. Notons V_q la somme connexe de q copies de $P_2(\mathbb{R})$.

Proposition 1.4. — *À une exception près, chaque surface de Del Pezzo réelle* (M, conj) *de degré* $d \geq 1$ *est déterminée à déformation près par la topologie de* $M(\mathbb{R})$. *Dans le cas exceptionnel,* $d = 8$ *et* $M(\mathbb{R})$ *est vide, le nombre de classes de déformation est deux, et dans l'une* M/conj *est* Spin, *alors que dans l'autre non.*

En chaque degré, les types topologiques de $M(\mathbb{R})$ *sont les suivants et ceux obtenus par des suites de simplifications de Morse à partir d'eux :* $M(\mathbb{R}) = V_1$ *en degré* 9 *;* $M(\mathbb{R}) = V_2$ *et* $S^1 \times S^1$ *en degré* 8 *;* $M(\mathbb{R}) = V_3$ *en degré* 7 *;* $M(\mathbb{R}) = V_4$ *et* $S^1 \times S^1$ *en degré* 6 *;* $M(\mathbb{R}) = V_5$ *en degré* 5 *;* $M(\mathbb{R}) = V_6, S^1 \times S^1$ *et* $2S^2$ *en degré* 4 *;* $M(\mathbb{R}) = V_7$ *et* $V_1 \sqcup S^2$ *en degré* 3 *;* $M(\mathbb{R}) = V_8, 2V_1, V_2 \sqcup S^2, S^1 \times S^1$ *et* $4S^2$ *en degré* 2 *;* $M(\mathbb{R}) = V_9, V_2 \sqcup V_1, V_3 \sqcup S^2$ *et* $V_1 \sqcup 4S^2$ *en degré* 1. (*Ici,* $\sqcup$ *est la réunion disjointe et* nS^2 *est la réunion disjointe de* n *sphères de dimension* 2.)

Parmi les surfaces de Del Pezzo, les seules variétés fortement de Fano sont le plan projectif ($d = 9$) et les surfaces géométriquement réglées ($d = 8$). Les deux structures exceptionnelles du théorème précédent sont les deux structures réelles sur $P_1(\mathbb{C}) \times P_1(\mathbb{C})$ sans point réel.

1.6. Remerciements

Je voudrais remercier V. Blanlœil, O. Debarre, T. Delzant, M. Paun et J. Y. Welschinger qui m'ont aidé à décortiquer la démarche de C. Viterbo et m'ont apporté une aide précieuse au cours de la rédaction de ces notes. Je remercie aussi C. Viterbo qui a patiemment répondu à mes multiples questions, et Y. Eliashberg qui m'a fait découvrir une autre approche aux problèmes traités dans ce rapport.

2. SYSTÈMES HAMILTONIENS ASSOCIÉS

Soit M une variété complexe compacte lisse kählérienne (c'est-à-dire, une variété munie d'une forme symplectique Ω et d'une structure complexe J telle que $\Omega(\,\cdot\,, J(\,\cdot\,))$ est une métrique riemannienne). Notons n la dimension complexe de M. Considérons une sous-variété lagrangienne L de M (une sous-variété différentiable réelle de dimension réelle n dont les espaces tangents sont isotropes pour Ω ; l'exemple essentiel pour nous est donné par la partie réelle d'une variété kählérienne munie d'une structure réelle, voir 1.3). D'après le théorème de Darboux-Weinstein, il existe un voisinage U de L dans M symplectomorphe à un voisinage de la section nulle dans $T^*(L)$ (avec sa structure symplectique canonique $dp \wedge dq$, comparer 1.1). À partir de maintenant, nous les identifierons et considérerons U comme situé simultanément dans M et dans $T^*(L)$.

Fixons une métrique g sur L (sans aucun rapport avec la métrique kählérienne de M). Il est bien connu que les géodésiques γ de (L, g) sont les points critiques de l'énergie

$$E(\gamma) = \int_0^\pi |\gamma'(t)|^2 dt$$

et que les graphes des géodésiques (trajectoires géodésiques) dans $T^*(L)$ sont les points critiques de l'action hamiltonnienne

$$A(\gamma) = \int (pdq - Hdt), \quad \text{où } H = \frac{1}{2}\sum g^{ij}p_i p_j \text{ et } pdq = \sum p_i dq^i.$$

En associant à H le champ X_H défini par $\Omega(\xi, X_H) = dH(\xi)$ (ce qui est équivalent à $X_H = J\nabla H$), on retrouve le flot géodésique sur le bord ∂T_r^* des tubes $T_r^* = \{(q,p) \in T^*(L) | \sum g^{ij}p_i p_j \leq r^2\}$ de rayon r.

L'objectif est d'étendre ce tableau à M en mélangeant le flot géodésique sur le bord d'un voisinage tubulaire de L avec la géométrie de l'opérateur de Cauchy-Riemann loin de ce bord. Plus précisément, on considère l'espace

$$\widetilde{\Lambda}_0 = \{(\gamma, u) | u \in C^\infty(D^2, M), \gamma = \partial u\}$$

et on introduit sur cet espace une famille d'actions dépendant d'un paramètre $\lambda \geq 0$, $\lambda \in \mathbb{R}$,

$$A_\lambda(\gamma, u) = \int_{D^2} u^*\Omega - \lambda \int_{\partial D^2} H_{\epsilon,\rho} \circ \gamma \, dt, \quad \text{où} \quad dt = \frac{dz}{2\pi i}.$$

La fonction $H_{\epsilon,\rho} : M \to \mathbb{R}$ ne dépend que de la distance à L, elle est choisie égale à 1 au-dehors de $U = T_r^*$ et égale à $h_{\epsilon,\rho}(|p|)$, $|p|^2 = \sum g^{ij}p_i p_j$, dans U. La fonction $h_{\epsilon,\rho}$ est de classe C^∞, identiquement nulle sur $[0, \epsilon]$, croissante et strictement convexe sur $[\epsilon, \frac{1}{2}(\epsilon + \rho)]$, croissante et strictement concave sur $[\frac{1}{2}(\epsilon + \rho), \rho]$ et égale à 1 sur $[\rho, \infty[$ (bien entendu, $0 < \epsilon < \rho < r$). On suppose que $\frac{1}{2}(\epsilon + \rho)$ n'est pas parmi les longueurs des géodésiques fermées de L.

Pour un couple (γ, u) tel que γ soit un lacet réduit à un point y, on a

$$A_\lambda(\gamma, u) = \int_\alpha \Omega - \lambda H_{\epsilon,\rho}(y)$$

où α est la classe d'homologie réalisée par $u : S^2 = D^2/S^1 \to M$.

Le calcul direct de la première variation de A_λ donne

$$dA_\lambda(\delta\gamma) = -\int_{\partial D^2} \langle \delta\gamma, J\gamma' + \lambda \nabla H_{\epsilon,\rho} \rangle dt$$

où $\delta\gamma$ est un champ de vecteurs tangents à M le long de γ. Comme conséquence, les points critiques de A_0 sont les couples (γ, u) tels que γ soit un lacet constant, et $u : S^2 = D^2/S^1 \to M$ une sphère holomorphe ; et la valeur critique de A_0 correspondante à (γ, u) est $\int_\alpha \Omega$ où α est la classe d'homologie de u. Pour $\lambda > 0$, il y a trois types de points critiques suivant les trois types de γ : les lacets constants dans $M \setminus T^*_{<\rho}$, les lacets constants dans T^*_ϵ, et les trajectoires géodésiques périodiques dans $T^*_{<\rho} \setminus T^*_\epsilon$. Les valeurs critiques correspondantes aux deux premiers types de points critiques sont respectivement :

$\int_\alpha \Omega - \lambda$ si $\lambda > 0$ et le lacet est au-dehors de T^*_ρ, et $\int_\alpha \Omega$ si $\lambda > 0$ et le lacet est dans T^*_ϵ.

Toutes les trajectoires critiques γ sont des trajectoires périodiques de période 1 du flot X_H. Ce flot sur ∂T^*_τ, $\epsilon < \tau < \rho$, est le flot (co)géodésique reparamétré. Les projections des trajectoires périodiques de période 1 contenues dans ∂T^*_τ sont les géodésiques de longueur $h'(\tau)$.

3. TRAJECTOIRES DE FLOER ET COURBES HOLOMORPHES

Les trajectoires de Floer sont les analogues des séparatrices (trajectoires du flot du gradient reliant les points critiques) dans la théorie des fonctions de Morse sur les variétés de dimension finie. Étant guidé par cette analogie et le calcul de la première variation de A_λ, on définit une trajectoire de Floer reliant deux points critiques, l'un correspondant à un lacet constant sur L et l'autre à un lacet constant au-dehors de U, comme une solution $u : \mathbb{R} \times S^1 \to M$ de l'équation

$$\frac{\partial u}{\partial s} + J\frac{\partial u}{\partial t} + \lambda \nabla H_{\epsilon,\rho}(u) = 0 \tag{2}$$

telle que

$$\lim_{s\to+\infty} u(s,t) = y \in M \setminus U \quad \text{et} \quad \lim_{s\to-\infty} u(s,t) = x \in L.$$

Une écriture équivalente de (2) est $\bar\partial u = -\lambda \nabla H_{\epsilon,\rho}(u)$. En particulier, pour $\lambda = 0$ ces trajectoires de Floer sont les courbes rationnelles holomorphes $S^2 \to M$ qui passent par un point x de L et un point y de $M \setminus U$. De plus, les courbes rationnelles minimales représentent les trajectoires de Floer qui relient deux valeurs critiques consécutives de A_0.

Le groupe $\mathbb{C}^*$ agit de façon usuelle sur les solutions de l'équation (2). Dans le cas des variétés minimalement réglées, cette action est libre sur l'ensemble des solutions avec $[u] = \alpha$, α étant la classe des courbes régulatrices.

L'action est souvent réduite à S^1 en normalisant u comme suit : $\int_{s\leq 0} u^*\Omega = \frac{1}{2}\int_\alpha \Omega$. À partir de maintenant, nous identifions la sphère S^2 et le compactifié de $\mathbb{R} \times S^1$ par deux points $\pm\infty$. Ainsi, sur S^2 il y aura toujours deux points marqués.

3.1. Sélection des courbes holomorphes

Considérons *une variété fortement de Fano M minimalement réglée par des courbes rationnelles $u : S^2 \to M$ de classe $\alpha = [u] \in H_2(M;\mathbb{Z}) = \pi_2(M)$*. Posons

$$\mathrm{Mor}_\alpha = \{u : S^2 \to M | \bar\partial u = 0, [u] = \alpha\}$$

et notons $\mathrm{ev}_{\pm\infty}$ le morphismes d'évaluation $\mathrm{Mor}_\alpha \to M$, $u \mapsto u(\pm\infty)$.

L'affirmation suivante est une conséquence du lemme de cassage de Mori et du théorème de compacité de Bishop.

PROPOSITION 3.1. — *$\mathrm{Mor}_\alpha/\mathbb{C}^*$ est une variété projective, et pour tout couple de points $x \neq y$ de M, il n'existe à reparamétrisation près qu'un nombre fini (éventuellement nul) de courbes rationnelles $u : S^2 \to M$ avec $u(-\infty) = x$ et $u(+\infty) = y$.*

Notons $M(x)$ la sous-variété de M des points qui peuvent être joints à x par $u \in \mathrm{Mor}_\alpha$ (c'est-à-dire, $y \in M(x)$ si et seulement s'il existe $u \in \mathrm{Mor}_\alpha$ tel que $u(-\infty) = x$ et $u(+\infty) = y$). Par des arguments standard de position générale, on déduit de 3.1 le résultat suivant (rappelons que H est le diviseur ample tel que $H \circ \alpha = 1$, voir 1.3).

COROLLAIRE 3.2. — *Il existe un entier $k > 0$ tel que pour tout point générique x de M et toute intersection complète générique A d'hypersurfaces $H_1, \dots, H_k$ avec $H_i \in |d_i H|$ l'intersection de $\mathrm{ev}_{-\infty}^{-1}(x)$ et $\mathrm{ev}_{+\infty}^{-1}(A)$ est non vide, finie et transverse.*

(On peut compléter 3.1 et 3.2 et démontrer que $\dim_{\mathbb{C}} \mathrm{Mor}_\alpha/\mathbb{C}^* = c_1(M) \cap \alpha + n - 1$ et $k = \dim_{\mathbb{C}} M(x) = c_1(M) \cap \alpha - 1 > 0$.)

Notons le nombre de points d'intersection de $\mathrm{ev}_{-\infty}^{-1}(x)$ et $\mathrm{ev}_{+\infty}^{-1}(A)$ par $\nu(x, A)$. En fait, A peut être remplacée par une variété différentiable dans la même classe d'homologie, et le nombre de points d'intersection est alors remplacé par le nombre algébrique.

Si L est une sous-variété lagrangienne de M, le point générique x peut être choisi dans L, et la variété A dans le complémentaire d'un voisinage tubulaire U de L (aucune sous-variété complexe de M ne contient L ; et si d est pair et suffisamment grand, il existe une hypersurface $H' \in |dH|$ qui ne rencontre pas L).

Dans ce qui suit on fixe $x \in L$ et $A \subset (M \setminus U)$ ainsi choisis.

4. CASSAGE DES COURBES HOLOMORPHES

Le théorème suivant produit une géodésique fermée munie d'une membrane partiellement holomorphe et dont on contrôle l'indice i_{CZ} dit de Conley-Zehnder-Duistermaat. C'est ce résultat qui permettra à la fin soit de construire une courbe réelle, soit de réfuter l'existence d'une métrique hyperbolique.

THÉORÈME 4.1. — *Si L admet une métrique sans géodésique contractile, il existe $\lambda > 0$, une trajectoire périodique γ de période λ du flot Hamiltonien $X_{\epsilon,\rho}$, et un disque $u : D^2 \to M$ tels que $\gamma = \partial u$, $u(0) = x$, $0 < A_\lambda(\gamma, u) < \int_\alpha \Omega$ et enfin, si la métrique est de plus à courbure sectionnelle strictement négative, tels que $i_{CZ}(\gamma, u) \geq n$.*

Idée de la démonstration. Supposons que le point $x \in L$ et la variété A sont fixés comme dans 3.1. Considérons l'espace $\mathcal{M}_\lambda(\alpha, x, A)$ des trajectoires de Floer de classe α reliant x à A et introduisons le diagramme de bifurcation :

$$\Delta = \{(\lambda, m) | \lambda \geq 0, m \in \mathcal{M}_\lambda(\alpha, x, A)/\mathbb{C}^*\}.$$

Notons que $\lambda \leq \int_\alpha \Omega$ pour tout $(\lambda, m) \in \Delta$ parce que $\int_\alpha \Omega - \lambda$ est la valeur critique de A_λ au-dessus de 0 sur la trajectoire de Floer m.

LEMME 4.2. — *Si Δ n'est pas compact, il existe une suite finie de trajectoires périodiques γ_i, $1 \leq i \leq k$, et une suite de trajectoires de Floer $(\gamma_{i-1}, u_i, \gamma_i)$, $1 \leq i \leq k+1$, reliant γ_{i-1} et γ_i, telles que γ_0 soit la trajectoire constante $x \in L$, γ_{k+1} soit une trajectoire constante $y \in A$, les autres trajectoires soient dans $T^*_\rho \setminus T^*_\epsilon$, et $[u_1 \# \ldots \# u_{k+1}] = \alpha$.*

Pour la démonstration de 4.2, qui nécessite de fines renormalisations et perturbations, on renvoie à [39].

Dès qu'une telle suite de trajectoires de Floer est obtenue, on prend $u = u_1$. L'inégalité $A_\lambda(\gamma, u) > 0$ est due à la croissance de A_λ le long des trajectoires de Floer, et $A_\lambda < \int_\alpha \Omega$ découle de $k \geq 1$ et $\sum_{i=1}^{k+1} A_\lambda(\gamma_{i-1}, u_i, \gamma_i) = \int_\alpha \Omega - \lambda$ (rappelons que $H_{\epsilon,\rho}(x) = 0$ et $H_{\epsilon,\rho}(M \setminus T^*_r) = 1$).

Si la métrique est à courbure sectionnelle strictement négative, la trajectoire γ est non dégénérée. On peut alors perturber $H_{\epsilon,\rho}$ en $\tilde{H}_{\epsilon,\rho}$ générique et de Morse sur L, et obtenir, de la même façon que dans la démonstration de 4.2, une trajectoire de Floer $(\tilde{\gamma}, \tilde{u})$ avec $i_{CZ}(\tilde{\gamma}, \tilde{u}) = i_{CZ}(\gamma, u)$ et $\tilde{u}$ qui relie $\tilde{\gamma}$ à un point qui réalise le minimum de $\tilde{H}_{\epsilon,\rho}$ sur L. On obtient donc l'inégalité $i_{CZ}(\tilde{\gamma}) \geq n - \operatorname{ind} \operatorname{Hess} \tilde{H}_{\epsilon,\rho}|_L = n$ (voir [39] pour les détails techniques de cette perturbation).

Pour montrer que Δ n'est pas compact, notons que pour x et A suffisamment génériques, la structure complexe de M est régulière en chaque trajectoire de Floer $(0, m) \in \Delta$ (le fibré normal des courbes unirégulatrices génériques est ample). Donc la projection $\Delta \to \mathbb{R}_+$, $(\lambda, m) \mapsto \lambda$ est propre au-dessus d'un voisinage de $\lambda = 0$, et la partie de Δ pour laquelle λ est voisin de 0 est une variété lisse de dimension 1 dont le bord est constitué des $\nu(x, A)$ points satisfaisant $\lambda = 0$. Cette variété est orientée près de ces points et ils sont tous positifs. Si Δ était compact, la perturbation décrite dans [19] aurait donné une variété compacte orientée (l'orientation provenant de [17]) à bord orienté positif. $\square$

Si les géodésiques de L ne sont pas contractiles (cas des métriques à courbure négative), le disque D construit dans la démonstration du théorème 4.1 ainsi que les autres trajectoires de Floer $(\gamma_{i-1}, u_i, \gamma_i)$ (un disque supplémentaire plus, éventuellement, des cylindres) sortent tous de T^*_ρ. Donc, chacun contient une partie $C^i_{\epsilon,\rho}$ non vide à bord sur laquelle $u : C^i_{\epsilon,\rho} \to M$ est une courbe holomorphe propre de $M \setminus T^*_\rho$. Il conviendra d'attribuer l'indice $i = 1$ au disque du théorème 4.1 et de poser $C(u) = C_{\epsilon,\rho} = \bigcup C^i_{\epsilon,\rho}$.

PROPOSITION 4.3. — *L'aire des courbes $C^k_{\epsilon,\rho}$ est uniformément séparée de 0 et ∞ : il existe $\delta > 0$ tel que $\delta < \operatorname{aire} C^k_{\epsilon,\rho} < \operatorname{aire} C(u) < \int_\alpha \Omega$ pour tout ϵ, ρ.*

Démonstration. — On a $A_\lambda(\gamma,u) < \int_\alpha \Omega$ et

$$A_\lambda(\gamma,u) = \frac{1}{2}\sum_i \int_{(\gamma_{i-1},u,\gamma_i)} |\partial_s y|^2 + |J\partial_t u + X_H(u_s)|^2$$
$$\geq \int_{C(u)} \frac{1}{2}(|\partial_s u|^2 + |J\partial_t u|^2) = \int_{C(u)} \Omega = \operatorname{aire} C(u),$$

d'où la borne supérieure. Comme les géodésiques fermées différentes ne sont jamais homologues, chaque courbe $C^k_{\epsilon,\rho}$ sort de U. En particulier, chacune coupe $\partial T^*_{\frac{2}{3}\rho}$ et on obtient une minoration uniforme en appliquant l'inégalité de Bishop-Lelong-Wirtinger. □

5. COMPARAISON DES INDICES

Il s'agit de l'indice $i_M \in \mathbb{Z}_+$ de Morse des géodésiques, de la classe de Maslov $\mu(L) \in H^2(M,L) = H^2(M,T_\rho)$ des sous-variétés lagrangiennes L, et de l'indice $i_{CZ} \in \mathbb{Z}_+$ de Conley-Zehnder-Duistermaat des orbites périodiques non dégénérées d'un flot Hamiltonien. Pour que ce dernier indice soit bien défini, il suffit que l'orbite périodique γ soit équipée d'une membrane, c'est-à-dire d'une surface orientée $u : S \to M$ (pas forcément un disque) qui a l'orbite pour bord et dont l'orientation du bord coïncide avec la direction du flot. Une telle membrane définit alors, à homotopie près, une unique trivialisation symplectique du fibré u^*T_*M et par définition i_{CZ} vaut « l'indice spectral » du flot hamiltonien linéarisé le long γ, voir [10] et [15]. Finalement, i_{CZ} ne dépend que de la classe de u dans $H_2(M,\gamma)$. À l'aide de la même trivialisation de u^*T_*M, on peut définir la classe de Maslov comme l'obstruction à une extension sur u de la trivialisation symplectique sur $\partial u \subset L$. Rappelons que l'image de $\mu(L)$ dans $H^2(M)$ coïncide avec $2c_1(M)$.

Signalons quelques spécificités de ce choix de normalisation de l'indice de Conley-Zehnder-Duistermaat (la normalisation change d'un auteur à l'autre). Pour le graphe $\gamma \subset T^*(L)$ d'une géodésique fermée non dégénérée contractile et u dans $T^*(L)$, on a $i_{CZ}(\gamma,u) = i_M$; pour γ et u constants et placés en un point critique non dégénéré de H, on a $i_{CZ} = n - \operatorname{ind}(\operatorname{Hess} H)$ si le hessien de H est suffisamment petit.

THÉORÈME 5.1. — *Soit γ une trajectoire périodique du flot de $X_{\epsilon,\rho}$ associée à une géodésique fermée γ de L. Si u est une surface dans M dont le bord est γ,*

$$i_{CZ}(\gamma,u) = i_M(\gamma) + \langle \mu(L), u\rangle \quad \textit{ou} \quad i_{CZ}(\gamma,u) = i_M(\gamma) + \langle \mu(L), u\rangle - 1.$$

Les trajectoires γ traitées dans ce théorème sont contenues dans ∂T^*_x avec $\epsilon < x < \rho$, $x \neq \frac{1}{2}(\epsilon+\rho)$. La première formule, sans -1, a lieu si $x < \frac{1}{2}(\epsilon+\rho)$, et la deuxième pour $x > \frac{1}{2}(\epsilon+\rho)$.

Dans le cas de $M = \mathbb{C}^n$, les formules du théorème 5.1 sont établies dans [38]. La démonstration dans le cas général s'effectue de la même façon.

6. MEMBRANES HOLOMORPHES

Supposons que $M = M(\mathbb{C})$ est une variété fortement de Fano réelle et L est une composante de $M(\mathbb{R})$ qui admet une métrique à courbure sectionnelle strictement négative. (Les résultats de cette section se généralisent aux sous-variétés lagrangiennes L réelles analytiques dans une variété M fortement de Fano, sous la même hypothèse sur la métrique.)

PROPOSITION 6.1. — *Il existe une courbe holomorphe compacte C' dont le bord s'appuie sur L et qui vérifie*

$$\langle \mu(L), C' \rangle \geq n \quad et \quad \mathrm{aire}(C') = \frac{1}{2} \int_\alpha \Omega. \tag{3}$$

La démonstration s'appuie sur le phénomène remarquable de prolongement des courbes holomorphes à travers les variétés analytiques totalement réelles maximales (par exemple les variétés lagrangiennes analytiques réelles, ce qui inclut les parties réelles des variétés algébriques). Ce résultat est spécifique à cette classe de sous-variétés et il est purement local (étrangement, la preuve s'appuie à nouveau sur l'étude, cette fois-ci locale, des familles de courbes rationnelles).

THÉORÈME 6.2. — *Soient W un ouvert de $\mathbb{C}^n$ et A un sous-ensemble analytique complexe de $W \setminus \mathbb{R}^n$ de dimension 1 en tout point (une courbe holomorphe propre dans $W \setminus \mathbb{R}^n$). Si A est invariant par la conjugaison complexe, son adhérence $\bar{A} \cap W$ est un sous-ensemble analytique complexe de W.*

Pour les courbes d'aire bornée ce résultat est dû à B. Shiffman [33], et sans cette hypothèse, à H. Alexander [2].

Démonstration de 6.1. D'après le théorème 6.2, si L est la partie réelle $L = M(\mathbb{R})$ d'une variété compacte complexe $M = M(\mathbb{C})$ munie d'une structure réelle, il nous suffit de construire dans $M \setminus L$ une courbe complexe analytique C propre vérifiant (3). (Si L n'est qu'une variété lagrangienne analytique réelle, la conjugaison complexe peut être introduite près de L et le théorème s'applique de nouveau.) Une telle courbe C s'obtient à partir des courbes $C^i_{\epsilon,\rho'} \subset M \setminus T^*_\rho$, avec $\rho' \leq \rho$ et $\rho', \rho \to 0$, par le procédé diagonal, en utilisant l'inégalité de Bishop-Lelong-Wirtinger et le théorème de Bishop de compacité appliqués aux courbes $C^i_{\epsilon,\rho'} \cap (M \setminus T^*_\rho)$. Ce procédé peut être appliqué au moins deux fois, à $i = 1$ et à $i = k + 1$ (voir 4.2). Ainsi nous obtenons dans $M \setminus L$ deux courbes holomorphes propres C' et C'' telles que, d'après 4.3,

$$n, \delta < \mathrm{aire}(C'), \quad \delta < \mathrm{aire}(C''), \quad \mathrm{aire}(C') + \mathrm{aire}(C'') \leq \int_\alpha \Omega. \tag{4}$$

(En particulier, C' et C'' sont non vides.) Dans le cas de variétés réelles, on double C' à l'aide de la conjugaison complexe. L'adhérence C_1 de la courbe doublée $C' \cup \mathrm{conj}\, C'$ est une courbe holomorphe compacte (éventuellement singulière). Évidemment, aire$(C_1) = 2\,\mathrm{aire}(C')$. Grâce à (4), on a l'inégalité aire$(C_1) < 2\int_\alpha \Omega - 2\delta$, et par l'hypothèse de minimalité de α

$$\mathrm{aire}(C_1) = \int_\alpha \Omega.$$

De plus, le même raisonnement s'applique à $C_2 = C'' \cup \mathrm{conj}\, C''$ et $\mathrm{aire}(C_1) + \mathrm{aire}(C_2) \leq 2\int_\alpha \Omega$. Ceci montre que la trajectoire de Floer ne s'est cassée qu'en deux et que $C' \cup C''$ en est la limite au sens des courants (pas de multiplicités dans les limites). Par suite, $[C', \partial C'] = u_*[D, \partial D] \in H_2(M, U)$, et par conséquent $\langle \mu(L), C' \rangle = \langle \mu(L), u \rangle$. D'après 5.1 et 4.1, et comme l'indice de Morse des géodésiques est nul dans le cas des métriques hyperboliques,

$$\langle \mu(L), u \rangle = i_{CZ}(\gamma) - i_M(\gamma) = i_{CZ}(\gamma) \geq n.$$

$\square$

Fin de la démonstration de 1.1. Par la naturalité de la classe de Maslov-Chern, $\mathrm{conj}_*\, \mu = -\mu$, et on déduit de (3) que

$$c_1(M) \cap C_i \geq n, \quad \text{pour} \quad i = 1, 2. \tag{5}$$

Maintenant, il ne reste qu'à combiner quelques résultats connus de la théorie des variétés de Fano.

Supposons que M est de dimension ≥ 3 (en dimension 2 le résultat est trivial modulo la classification des surfaces de Del Pezzo). D'abord, on déduit de (5) l'inégalité $c_1(M) \cap \alpha \geq n$. D'après la minimalité de α, on obtient $c_1(M) \cap C \geq n$ pour toute courbe rationnelle C (de même que pour toute courbe holomomorphe). Le théorème de Wiśnevski [40] nous dit que dans ce cas $\mathrm{Pic}(M) = \mathbb{Z}$. Puisque $H \cap \alpha = 1$, on en déduit que $c_1(M) = mH$ avec $m \geq n$. Finalement, on est dans le cadre du théorème de Kobayashi-Ochiai [22] qui dit que M est soit $P_n(\mathbb{C})$, soit une quadrique dans $P_{n+1}(\mathbb{C})$. Sur ces variétés, les seules structures réelles ayant des points réels sont les structures usuelles (voir 1.3), donc L est ou bien $P_n(\mathbb{R})$, ou bien un produit croisé de deux sphères. Dans tous les cas le revêtement universel de L n'est pas contractile et, d'après le théorème de Cartan-Hadamard, L n'admet pas de métrique à courbure sectionelle négative.

7. REMARQUES

7.1. Théorèmes de finitude

On peut déduire des théorèmes de finitude sur $\mathbb{C}$ (voir le rapport [11]), que les variétés de Fano réelles de dimension donnée ne forment qu'un nombre fini de classes à déformation réelle près.

Si on considère une classe quelconque de variétés à déformation sur $\mathbb{C}$ près, les variétés réelles de cette classe, considérées à déformation réelle près, sont-elles en nombre fini ? Cette question est à notre connaissance ouverte. La réponse est positive pour les courbes, les variétés de type général, les variétés abéliennes et toutes les surfaces sauf peut-être les surfaces rationnelles.

Une question étroitement liée est la question de finitude des structures réelles sur une variété donnée à conjugaison près. Comme ci-dessus, la réponse est positive pour les courbes, les variétés de type général, les variétés abéliennes et toutes les surfaces sauf peut-être les surfaces rationnelles.

Pour les courbes, ces résultats sont très classiques. Pour les variétés abéliennes, ils sont dus à A. Borel et J.-P. Serre [6]. Pour les surfaces, nous les avons démontrés avec I. Itenberg, après des discussions avec C. Viterbo, en nous appuyant sur la classification des surfaces et les connaissances de leurs groupes d'automorphismes.

7.2. Espaces linéaires contenus dans des intersections complètes réelles

Si la proposition 1.2 garantit l'existence de droites réelles sur des intersections complètes sous une hypothèse géométrique, il existe un tout autre critère qui utilise une forte hypothèse d'imparité de degré, mais en revanche qui se généralise aux espaces linéaires de dimensions quelconques.

La variété des r-plans contenus dans une sous-variété V de $P^n(\mathbb{C})$ définie par des équations de multidegré $d = (d_1, \ldots, d_k)$ est le lieu des zéros d'une section d'un fibré vectoriel sur la grassmannienne $G(r, P^n)$. La décomposition en classes de Schubert de la classe de Chern de degré maximal de ce fibré est donnée dans [12].

Grâce au théorème de Bézout (et le fait que les cycles de Schubert sont réels), pour montrer l'existence d'espaces linéaires réels contenus dans des intersections complètes réelles, il suffit de montrer qu'un des coefficients dans la décomposition de la classe de Chern est impair. C'est ainsi que O. Debarre et L. Manivel démontrent (sans aucune hypothèse d'hyperbolicité) que si chaque composante de d est impaire, et toute intersection complète de multidegré d contient des r-espaces linéaires complexes (condition qu'on peut exprimer en terme d'une simple inégalité sur n, r et d), alors les intersections réelles contiennent des r-espaces réels.

Comme conséquence, la partie réelle des intersections complètes fortement de Fano de degré impair (sans aucune hypothèse d'hyperbolicité) contient un fermé semi-algébrique d'intérieur non vide couvert par des droites réelles. Le cas le plus simple est celui d'hypersurfaces cubiques dans P^4.

7.3. Classe de Maslov, orientabilité et les formes de Rokhlin-Guillou-Marin et de Viro

La classe de Maslov-Chern apparaissait déjà en géométrie algébrique réelle, partiellement grâce à ses rapports avec l'orientabilité du lieu réel. Rappelons quelques-unes de ces apparitions.

Soient $M = M(\mathbb{C})$ une variété compacte complexe munie d'une structure réelle, et $L = M(\mathbb{R})$ sa partie réelle. Le critère d'orientabilité de L le plus simple, et le plus typique, est le suivant (découvert et redécouvert par différents auteurs, voir en particulier [35], [16], [26]) : L est orientable si $c_1(M)$ est divisible par 2 et $H^1(M;\mathbb{Z}/2)=0$.

Cette observation est étroitement liée à la construction de l'indice de Maslov. En effet, la classe de Maslov est une classe de cohomologie relative $\mu \in H^2(M,L;\mathbb{Z})$, et si $H^1(M;\mathbb{Z}/2)=0$, elle se relève en une unique classe appartenant à $H^1(L;\mathbb{Z}/2)$ laquelle n'est autre que la classe de Stiefel-Whitney. Ainsi, L est orientable si $c_1 = 0 \in H^2(M;\mathbb{Z}/2)$ et $H^1(M;\mathbb{Z}/2)=0$. Même mieux, sous ces hypothèses la classe de Maslov $\mu \in H^2(M,L)$ et sa moitié $\tilde{\mu} \in H^2(M,L)$ (la dernière étant définie à l'aide des lagrangiens orientés) se relèvent (après réduction $\mod 4$) en respectivement $\mu_4 \in H^1(L;\mathbb{Z}/4)$ et $\mu_2 \in H^1(L;\mathbb{Z}/2)$, avec $\mu_4 = 2\mu_2$.

En dimension deux, comme c'était remarqué par N. Netsvetaev [30], μ_2 est la différence entre les formes quadratiques de Rokhlin [32] et de Viro [37]. Supposons pour simplifier que $H^1(M;\mathbb{Z}/2)=0$. Si $[L]=0\in H^2(M;\mathbb{Z}/2)$, *la forme de Viro* $q_V : H_1(L;\mathbb{Z}/2) \to \mathbb{Z}/2$ est définie comme suit : prenons une courbe lisse simple γ de L, multiplions son champ de vecteurs tangents par $\sqrt{-1}$ et déplaçons γ le long de ce champ dans $M \setminus L$, le nombre $q_V[\gamma]$ est alors égal à 1 plus le nombre d'enlacement entre γ déplacé et L. *La forme de Rokhlin* $q_R : H_1(L;\mathbb{Z}/2)\to\mathbb{Z}/2$ est définie si L est orientable et $[L]=w_2(M)$. Si de plus $w_2(M)=0$, elle diffère de q_V par $\chi(F)+\nu(F)$, où F est une surface plongée dans M dont le bord est réduit à γ et qui est normale à L le long de γ. D'où $q_V = q_R + \mu_2$ si $[L]=w_2(M)=0$.

Si $[L]=w_2(M)\neq 0$ (et L n'est pas forcément orientable), la forme de Rokhlin peut être remplacée par celle de Guillou-Marin $q_{RGM} : H_1(L;\mathbb{Z}/2)\to\mathbb{Z}/4$. En présence d'une courbe réelle C avec $[C]=[L]=w_2(M)$, la forme de Viro se généralise en $q_V : H_1(L;\mathbb{Z}/2)\to\mathbb{Z}/4$ et on obtient aisement l'identité $q_V = q_{RGM}+\mu_4$.

Si $H^1(M)=0$ et $c_1(M)=0$, on définit de la même façon la classe de Maslov résiduelle $\mu_0 \in H^1(L)$. Dans le cas des variétés de Calabi-Yau, à cause de l'existence de n-formes holomorphes réelles, cette classe résiduelle est nulle.

Dès que le lieu réel est orientable on s'intéresse à des orientations privilégiées. Dans le cas des courbes, lorsque $C(\mathbb{R})$ décompose $C(\mathbb{C})$, deux orientations opposées canoniques, dites orientations complexes, apparaissent sur $C(\mathbb{R})$. Ces orientations, introduites par V. Rokhlin dans les années 70, jouent depuis un rôle extrêmement important dans l'étude des courbes algébriques sur les surfaces, en particulier, dans l'étude des courbes planes. L'outil principal est la formule d'orientations complexes de Rokhlin (voir [14] pour des réferences, un survey bref du cas des courbes sur les surfaces, une formule complémentaire découverte récemment par S. Orevkov [31] et sa généralisation due à J. Y. Welschinger [42]).

7.4. Approche d'Eliashberg

Alors que nous étions en train d'achever la préparation de ce rapport, Y. Eliashberg a proposé une autre approche au théorème de Viterbo, utilisant des techniques récemment développées avec A. Givental et H. Hofer (voir l'exposé de Y. Eliashberg au congrès de Berlin). Il nous est difficile, voire impossible, d'entrer dans les détails de ce projet.

Y. Eliashberg identifie un voisinage tubulaire de L avec un tube T^*_r du cotangent T^*L, puis $T^*L \setminus L$ avec la symplectification $(\mathbb{R} \times V, d(e^\tau \alpha))$ de $V = \partial T^*_r$, où α est la structure de contact de ∂T^*_r. Ensuite il déforme la structure symplectique Ω et la structure presque complexe J. Les variétés déformées (M^t, J^t, Ω^t), où Ω^t n'est définie qu'à un facteur constant près, peuvent être vues soit comme le résultat du recollement de (T^*_r, Ω), $([0,t] \times V, d(e^\tau \alpha))$ et $(M \setminus T^*_{<r}, e^t\Omega)$, soit comme celui de $(T^*_r, e^{-t}\Omega)$, $([-t,0] \times V, d(e^\tau \alpha))$ et $(M \setminus T^*_{<r}, \Omega)$. La structure presque complexe sur T^*_r et $M \setminus T^*_{<r}$ reste la même et est choisie invariante par translation sur $[0,t] \times V$ (respectivement $[-t,0] \times V$). Lorsque $t \to +\infty$, à la limite, la variété M se déchire en T^*L et $M \setminus L$, les bouts de ces variétés étant équipés de la forme symplectique homogène $d(e^\tau \alpha)$ ($\tau \to +\infty$ dans T^*L, tandis que dans M, $\tau \to -\infty$). Il s'agit alors de comprendre ce que deviennent les courbes holomorphes dans la variété déchirée, c'est-à-dire d'obtenir une version appropriée du théorème de compacité de Gromov. N'utilisant pas de structure réelle globale, l'avantage de cet argument serait de donner des résultats sur toutes les sous-variétés lagrangiennes. L'inconvénient est que cette méthode ne semble pas permettre de construire des courbes réelles en présence d'une structure réelle.

RÉFÉRENCES

[1] S. Akbulut, H. King – *Rational structures on 3-manifolds.* Pacific J. Math., vol. 151, 1991, p. 201–204.

[2] H. Alexander – *Continuing 1-dimensional analytic sets.* Math. Ann., vol. 191, 1971, p. 143–144.

[3] V. I. Arnol'd – *On the arrangements of the ovals of real plane curves, involutions of 4-dimensional manifolds, and the arithmetic of integral quadratic forms.* Functional Anal. Appl., vol. 5, 1971, p. 169–176.

[4] R. Benedetti, A. Marin – *Déchirures de variétés de dimension* 3. Comm. Math. Helv., vol. 67, 1992, p. 514–545.

[5] J. Bochnak, M. Coste, M.-F. Roy – *Géométrie algébrique réelle.* Ergebnisse der Mathematik und ihrer Grenzgebiete, 12 ; Springer - Verlag, 1987.

[6] A. Borel, J.-P. Serre – *Théorèmes de finitude en cohomologie galoisienne.* Comm. Math. Helv., vol. 39, 1964, p. 111–164.

[7] D. Burns – *On the uniqueness and characterization of Grauert tubes.* In *Complex analysis and geometry (Trento, 1993)* Lecture Notes in Pure and Appl. Math., vol. 173, p. 119–133.

[8] F. Campana – *Remarques sur le revêtement universel des variétés kählériennes compactes.* Bull. Soc. Math. France, vol. 122, 1994, n. 2, p. 255–284.

[9] A. Comessatti – *Sulla connessione delle superfizie razionali reali.* Annali di Math., vol. 23, 1914, p. 215–283.

[10] C. Conley, E. Zehnder – *Morse type index theory for flows and periodic solutions for Hamiltonian equations.* Comm. Pure Appl. Math., vol. 37, 1984, p. 207–253.

[11] O. Debarre – *Variétés de Fano.* Sém. Bourbaki, exp. n° 827, Astérisque 245 (1997), p. 197–221.

[12] O. Debarre, L. Manivel – *Sur la variété des espaces linéaires contenus dans une intersection complète.* Math. Ann., vol. 312, 1998, n. 3, p. 549–574.

[13] A. Degtyarev, I. Itenberg, V. Kharlamov – *Real Enriques surfaces.* Submitted to Lecture Notes Math.

[14] A. Degtyarev, V. Kharlamov – *Topological properties of real algebraic varieties : du côté de chez Rokhlin.* À paraître dans Uspekhi Mat. Nauk.

[15] J. J. Duistermaat – *On the Morse index in variational calculus.* Adv. Math., vol. 21, 1976, 173–195.

[16] A. L. Edmonds – *Orientability of fixed point sets.* Proc. Amer. Math. Soc., vol. 82, 1981, n. 1, p. 120–124.

[17] A. Floer, H. Hofer, – *Coherent orientations for periodic orbit problems in symplectic geometry.* Math. Z., vol 212, 1993, n. 1, p. 13–38.

[18] H. Grauert – *On Levi's problem and the imbedding of real-analytic manifolds.* Annals of Math., vol. 68, 1958, n. 3, p. 460–472.

[19] H. Hofer, C. Viterbo – *The Weinstein Conjecture in the Presence of Holomorphic Spheres.* Commun. Pure Appl. Math., vol 45, 1992, p. 582–622.

[20] N. V. Ivanov – *Approximation of smooth manifolds by real algebraic sets.* Uspekhi Mat. Nauk, vol. 37, 1982, n. 1, p. 3–52.

[21] V. M. Kharlamov – *Topological types of nonsingular surfaces of degree* 4 *in* $\mathbb{R}P^3$. Funkcional. Anal. i Priložen., vol. 10, 1976, n. 4, p. 55–68.

[22] S. Kobayashi, T. Ochiai – *On complex manifolds with positive tangent bundles.* J. Math. Soc. Japan, vol. 22, 1970, p. 499–525.

[23] J. Kollár – *Real algebraic threefolds. II. Minimal Model Program.* J. Amer. Math. Soc., vol. 12, 1999, p. 33–83.

[24] J. Kollár – *Non-projective Nash conjecture.* exposé à Paris XI, juin 1999.

[25] J. Kollár – *Rationally connected varieties over local fields.* Ann. of Math., vol. 150, 1999, p. 357–367.

[26] V. A. Krasnov – *Characteristic classes of vector bundles on a real algebraic variety.* Izv. Akad. Nauk SSSR Ser. Mat., vol. 55, 1991, n. 4, p. 716–746.

[27] S. B. Myers – *Riemannian manifolds with positive mean curvature.* Duke Math. J., vol. 8, 1941, p. 401–404.

[28] G. Mikhalkin – *Blowup equivalence of smooth closed manifolds.* Topology, vol. 36, 1997, p. 287–299.

[29] J. Nash – *Real algebraic manifolds.* Annals of Math., vol. 56, 1952, n. 3, p. 405–421.

[30] N. Yu. Netsvetaev – *An analogue of the Maslov index.* Journal of Math. Sci., vol. 81, 1996, n. 2, p. 2535–2537.

[31] S. Yu. Orevkov – *Link theory and oval arrangements of real algebraic curves.* Topology, vol. 38, 1999, n. 4, p. 779–810.

[32] V. A. Rokhlin – *Proof of the Gudkov conjecture.* Funkts. Anal. i Prilozhen., vol. 6, 1972, n. 2, p. 62–64.

[33] B. Shiffman – *On the Continuation of Analytic Sets.* Math. Ann., vol. 185, 1970, p. 1–12.

[34] R. Silhol – *Real algebraic surfaces.* Lect. Notes Math., vol. 1392, 1989.

[35] A. J. Sommese – *Real algebraic spaces.* Ann. Scuola Norm. Sup. Pisa, vol. 4, 1977, n. 4, p. 599–612.

[36] A. Tognoli – *Quelques exemples en géométrie algébrique réelle.* in *Séminaire sur la géométrie algébrique réelle*, Publ. Math. Univ. Paris VII, vol. 24, 1986, p. 29–34.

[37] O. Ya. Viro – *Achievements in the topology of real algebraic varieties in the last 6 years.* Uspekhi Mat. Nauk, vol. 41, 1986, n. 3, p. 45–67.

[38] C. Viterbo – *A new obstruction to embedding Lagrangian tori.* Invent. Math., vol. 100, 1990, n. 2, p. 301–320.

[39] C. Viterbo – *Symplectic real algrebraic geometry.* À paraître.

[40] J. A. Wiśniewski – *On a conjecture of Mukai.* Manuscripta Math., vol. 68, 1990, p. 135–141.

[41] G. Wilson – *Hilbert's sixteenth problem.* Topology, vol. 17, 1978, n. 1, p. 53–73.

[42] J. Y. Welschinger – *J-courbes réelles à nids profonds sur les surfaces réglées.* Prépublication d'IRMA-ULP, à paraître dans les œuvres du colloque de Rokhlin, 1999, St. Petersbourg.

Viatcheslav KHARLAMOV

Université Louis Pasteur et CNRS
I.R.M.A.
7, rue René Descartes
F-67084 STRASBOURG Cedex
E-mail : `kharlam@irma.u-strasbg.fr`

Séminaire BOURBAKI Mars 2000
52e année, 1999-2000, n^o 873, p. 207 à 265

LA CORRESPONDANCE DE LANGLANDS SUR LES CORPS DE FONCTIONS [d'après Laurent Lafforgue]

par **Gérard LAUMON**

Lafforgue a récemment établi la correspondance de Langlands pour GL_r sur un corps de fonctions. Sa preuve suit la stratégie introduite, il y a plus de 25 ans, par Drinfeld pour traiter le cas $r = 2$.

Après avoir énoncé le théorème principal et ses conséquences, nous présenterons les grandes lignes de la démonstration, en renvoyant pour les détails aux publications de Lafforgue ([La 1] à [La 9]). Les sections 4 (Compactification de l'isogénie de Lang pour PGL_r) et 7 (Une variante d'un théorème de Pink) sont de nature générale et peuvent être lues indépendamment du reste du texte.

Je remercie chaleureusement L. Lafforgue pour son aide dans la préparation de cet exposé.

1. ÉNONCÉ DU THÉORÈME PRINCIPAL

On fixe dans tout cet exposé une courbe X projective, lisse et géométriquement connexe sur un corps fini $\mathbb{F}_q$ à q éléments. On note $|X|$ l'ensemble des points fermés de X.

Soit F le corps des fonctions de X. On identifie les places de F aux éléments de $|X|$. Pour chaque $x \in |X|$ on peut former le complété F_x de F en la place x. C'est un corps de valuation discrète complet. On notera encore $x : F_x^\times \to \mathbb{Z}$ sa valuation. L'anneau des entiers de F_x est l'anneau de valuation discrète $\mathcal{O}_x = \{a_x \in F_x^\times \mid x(a_x) \geq 0\} \cup \{0\}$, $\mathfrak{p}_x = \{a_x \in F_x^\times \mid x(a_x) > 0\} \cup \{0\}$ est l'unique idéal maximal de $\mathcal{O}_x$ et le corps résiduel $\kappa(x) = \mathcal{O}_x/\mathfrak{p}_x$ est une extension finie de $\mathbb{F}_q$ dont on notera $\deg(x)$ le degré.

L'anneau (topologique) des adèles de F est le produit restreint

$$\mathbb{A} = \{a = (a_x)_{x\in|X|} \mid a_x \in \mathcal{O}_x \text{ pour presque tout } x\} \subset \prod_{x\in|X|} F_x$$

où l'expression « pour presque tout x » signifie « pour tous les x sauf un nombre fini ». Le corps F se plonge diagonalement dans $\mathbb{A}$ et l'anneau topologique compact $\mathcal{O} := \prod_{x\in|X|} \mathcal{O}_x$ est un sous-anneau de $\mathbb{A}$.

On dispose d'un homomorphisme de groupes $\deg : \mathbb{A}^\times \to \mathbb{Z}$ défini par

$$\deg(a) = \sum_{x\in|X|} \deg(x)x(a_x).$$

Cet homomorphisme est surjectif et est identiquement nul sur $F^\times$ et aussi sur $\mathcal{O}^\times$. Son noyau est compact modulo $F^\times$: en d'autres termes, pour tout $a \in \mathbb{A}^\times$ de degré non nul, le groupe quotient $F^\times\backslash\mathbb{A}^\times/\mathcal{O}^\times a^{\mathbb{Z}}$ est fini.

1.1. Représentations automorphes cuspidales

Soit r un entier ≥ 1. On considère le groupe adélique $\mathrm{GL}_r(\mathbb{A})$ des matrices inversibles de taille $r \times r$ et à coefficients dans l'anneau $\mathbb{A}$, et ses sous-groupes $\mathrm{GL}_r(F)$ et

$$K = \prod_{x\in|X|} K_x = \mathrm{GL}_r(\mathcal{O}) = \prod_{x\in|X|} \mathrm{GL}_r(\mathcal{O}_x).$$

On fixe une fois pour toutes une mesure de Haar dg sur $\mathrm{GL}_r(\mathbb{A})$, et une décomposition $dg = \prod_{x\in|X|} dg_x$ de cette mesure en produit de mesures de Haar locales. Il est commode de normaliser dg et les dg_x par les conditions $\mathrm{vol}(K, dg) = 1$ et $\mathrm{vol}(K_x, dg_x) = 1$.

Une *forme automorphe cuspidale* (pour GL_r sur F) est une fonction $\varphi : \mathrm{GL}_r(\mathbb{A}) \to \mathbb{C}$ ayant les propriétés suivantes :

1) $\varphi(\gamma g) = \varphi(g)$, $\forall\gamma \in \mathrm{GL}_r(F)$, $\forall g \in \mathrm{GL}_r(\mathbb{A})$,

2) il existe un sous-groupe $K_\varphi \subset K$ d'indice fini tel que $\varphi(gk) = \varphi(g)$, $\forall g \in \mathrm{GL}_r(\mathbb{A})$, $\forall k \in K_\varphi$,

3) il existe $a \in \mathbb{A}^\times$ tel que $\deg(a) \neq 0$ et $\varphi(ga) = \varphi(g)$, $\forall g \in \mathrm{GL}_r(\mathbb{A})$,

4) pour toute décomposition non triviale $r = r_1+\cdots+r_s$ de r en entiers strictement positifs, qui définit un sous-groupe parabolique standard $P = MU \subsetneq \mathrm{GL}_r$ de radical unipotent U et de composante de Levi $M \cong \mathrm{GL}_{r_1} \times\cdots\times \mathrm{GL}_{r_s}$, le *terme constant*

$$\mathrm{GL}_r(\mathbb{A}) \to \mathbb{C},\ g \mapsto \int_{U(F)\backslash U(\mathbb{A})} \varphi(ug)du,$$

est identiquement nul (ici du est n'importe quelle mesure de Haar sur le quotient compact $U(F)\backslash U(\mathbb{A})$).

Si l'on fixe un élément a_0 de degré non nul dans $\mathbb{A}^\times$, on a

$$L_{\mathrm{cusp}} = \bigcup_{n\geq 1} L_{\mathrm{cusp}}(a_0^n)$$

où, pour tout $a \in \mathbb{A}^\times$ de degré non nul, $L_{\mathrm{cusp}}(a) \subset L_{\mathrm{cusp}}$ est le sous-espace défini en imposant ce a particulier dans la propriété 3).

Toujours pour $a \in \mathbb{A}^\times$ de degré non nul, on montre que toute fonction $\varphi \in L_{\mathrm{cusp}}(a)$ est à support compact sur $\mathrm{GL}_r(F)\backslash \mathrm{GL}_r(\mathbb{A})/a^{\mathbb{Z}}$, et on peut donc munir $L_{\mathrm{cusp}}(a)$ du produit scalaire hermitien défini positif

$$(\varphi_1, \varphi_2) := \int_{\mathrm{GL}_r(F)\backslash \mathrm{GL}_r(\mathbb{A})/a^{\mathbb{Z}}} \overline{\varphi_1(g)}\varphi_2(g)dg.$$

Le groupe $\mathrm{GL}_r(\mathbb{A})$ agit par translation à droite sur l'espace vectoriel complexe

$$L_{\mathrm{cusp}} = L_{\mathrm{cusp}}(\mathrm{GL}_r(F)\backslash \mathrm{GL}_r(\mathbb{A}))$$

des formes automorphes cuspidales. Cette représentation est *lisse* (le fixateur de tout vecteur dans L_{cusp} est un sous-groupe ouvert de $\mathrm{GL}_r(\mathbb{A})$). Comme on a fixé une mesure de Haar dg sur $\mathrm{GL}_r(\mathbb{A})$, la donnée de cette représentation équivaut à celle d'une structure de $\mathcal{H}$-module sur L_{cusp}, où l'*algèbre de Hecke*

$$\mathcal{H} = \mathcal{C}_{\mathrm{c}}^{\infty}(\mathrm{GL}_r(\mathbb{A}))$$

est l'algèbre de convolution des fonctions complexes, localement constantes et à support compact, sur $\mathrm{GL}_r(\mathbb{A})$.

Pour chaque $a \in \mathbb{A}^\times$ de degré non nul, l'action de $\mathrm{GL}_r(\mathbb{A})$ respecte le sous-espace $L_{\mathrm{cusp}}(a)$. La représentation induite sur ce sous-espace est *admissible* (pour tout sous-groupe $K' \subset K$ d'indice fini, l'espace vectoriel des invariants sous K' dans $L_{\mathrm{cusp}}(a)$ est de dimension finie). Elle est de plus *unitaire* pour le produit scalaire défini ci-dessus. La représentation de $\mathrm{GL}_r(\mathbb{A})$ sur L_{cusp} est donc semi-simple : elle admet une décomposition isotypique

$$L_{\mathrm{cusp}} \cong \bigoplus_{\pi} V_\pi^{\oplus m(\pi)}$$

où π parcourt un système de représentants des classes d'isomorphie de représentations complexes irréductibles admissibles de $\mathrm{GL}_r(\mathbb{A})$, où V_π est l'espace de π (en général de dimension infinie) et où les multiplicités $m(\pi)$ sont des entiers ≥ 0.

Définition. — *Une* représentation automorphe cuspidale irréductible (*pour* GL_r *sur* F) *est une représentation complexe admissible irréductible de* $\mathrm{GL}_r(\mathbb{A})$ *qui est isomorphe à un facteur direct de* L_{cusp}.

On notera $\mathcal{A}_r$ un système de représentants des classes d'isomorphie de ces représentations automorphes cuspidales irréductibles. Chaque $\pi \in \mathcal{A}_r$ admet un caractère central $\omega_\pi : F^\times\backslash\mathbb{A}^\times \to \mathbb{C}^\times$ qui est d'ordre fini puisque $\omega_\pi(a) = 1$ pour au moins un $a \in \mathbb{A}^\times$ de degré non nul.

Théorème (Théorème de multiplicité un, [PS 2], [Sh]). — *On a*

$$L_{\mathrm{cusp}} \cong \bigoplus_{\pi\in\mathcal{A}_r} V_\pi.$$

En d'autres termes, les multiplicités dans la décomposition isotypique de la représentation L_{cusp} *de* $\mathrm{GL}_r(\mathbb{A})$ *sont données par* $m(\pi) = 1$ *si* π *est automorphe cuspidale et* $m(\pi) = 0$ *sinon.*

Pour tout $x \in |X|$, on définit aussi l'*algèbre de Hecke locale*

$$\mathcal{H}_x = \mathcal{C}_c^\infty(\mathrm{GL}_r(F_x)).$$

C'est l'algèbre de convolution pour la mesure de Haar dg_x sur $\mathrm{GL}_r(F_x)$. L'espace de toute représentation lisse de $\mathrm{GL}_r(F_x)$ est naturellement muni d'une structure de $\mathcal{H}_x$-module.

On notera $e_{K_x} \in \mathcal{H}_x$ la fonction caractéristique de $K_x = \mathrm{GL}_r(\mathcal{O}_x)$ dans $\mathrm{GL}_r(F_x)$ et on appellera algèbre de Hecke locale *non ramifiée* la sous-algèbre

$$\mathcal{H}_x(K_x) = e_{K_x} * \mathcal{H}_x * e_{K_x} \subset \mathcal{H}_x.$$

C'est une algèbre commutative unitaire munie d'un isomorphisme, construit par Satake,

$$\mathcal{H}_x(K_x) \cong \mathbb{C}[z_1, z_1^{-1}, \ldots, z_r, z_r^{-1}]^{\mathfrak{S}_r},$$

où le groupe symétrique $\mathfrak{S}_r$ agit par permutation des indéterminées $z_1, \ldots, z_r$. En particulier, si (π_x, V_{π_x}) est une représentation admissible irréductible de $\mathrm{GL}_r(F_x)$ *non ramifiée*, c'est-à-dire telle que $V_{\pi_x}^{K_x} \neq (0)$, on a en fait

$$\dim(V_{\pi_x}^{K_x}) = 1$$

et le $\mathcal{H}_x(K_x)$-module $V_{\pi_x}^{K_x}$ de rang 1 et aussi la représentation (π_x, V_{π_x}) sont (à isomorphisme près) uniquement déterminés par la donnée d'un r-uplet non ordonné

$$(z_1(\pi_x), \ldots, z_r(\pi_x))$$

de nombres complexes non nuls, appelés les *valeurs propres de Hecke* de π_x.

L'algèbre de Hecke globale $\mathcal{H}$ est le produit tensoriel restreint

$$\mathcal{H} = \varinjlim_S \Big(\bigotimes_{x\in S} \mathcal{H}_x\Big) \otimes \Big(\bigotimes_{x\notin S} e_{K_x}\Big)$$

des algèbres de Hecke locales, où S parcourt l'ensemble des parties finies de $|X|$. Une représentation admissible irréductible (π, V_π) de $\mathrm{GL}_r(\mathbb{A})$ admet donc, pour chaque $x \in |X|$, une composante locale (π_x, V_{π_x}) qui est une représentation irréductible admissible de $\mathrm{GL}_r(F_x)$ bien définie à isomorphisme près. Pour presque tout x, π est non ramifiée en x, c'est-à-dire admet une composante locale non ramifiée en x, et π est le produit tensoriel restreint

$$\pi \cong \bigotimes_{x\in|X|}{}' \pi_x$$

de ses composantes locales. (Une fois que l'on a fixé une base v_x de $V_{\pi_x}^{K_x}$ pour chaque place x non ramifiée pour π, l'espace de la représentation du membre de droite est par définition la limite inductive

$$\varinjlim_{S} \left(\bigotimes_{x \in S} V_{\pi_x} \right) \otimes \left(\bigotimes_{x \notin S} v_x \right)$$

pour S parcourant l'ensemble des parties finies de $|X|$ qui contiennent toutes les places ramifiées pour π.)

On définit la *fonction L partielle* de π comme le produit eulérien formel

$$L^{S_\pi}(\pi, T) = \prod_{x \notin S_\pi} L(\pi_x, T) = \prod_{x \notin S_\pi} \frac{1}{\prod_{i=1}^{r} (1 - z_i(\pi_x) T^{\deg(x)})} \in \mathbb{C}[[T]]$$

où $S_\pi \subset |X|$ est l'ensemble fini des places ramifiées de π. Godement et Jacquet ([G-J]) ont démontré que le produit eulérien $L^{S_\pi}(\pi, q^{-s})$ converge absolument dans un demi-plan $\mathrm{Re}(s) > \sigma$ pour un nombre réel σ assez grand, et que la série formelle en T définie par $L^{S_\pi}(\pi, T)$ est en fait le développement d'une fraction rationnelle dans $\mathbb{C}(T)$, et même d'un polynôme dans $\mathbb{C}[T]$ si $r \geq 2$.

1.2. Représentations du groupe de Galois

On fixe une clôture séparable $\overline{F}$ de F et on note Γ_F le groupe de Galois de $\overline{F}$ sur F. Pour chaque $x \in |X|$ on choisit arbitrairement un sous-groupe de décomposition $D_x \subset \Gamma_F$ en x. Ce groupe de décomposition est le groupe de Galois Γ_{F_x} d'une certaine clôture séparable $\overline{F}_x$ de F_x. On note $I_x \subset D_x$ le sous-groupe d'inertie de ce groupe de décomposition. Le groupe quotient D_x/I_x est isomorphe au groupe de Galois $\Gamma_{\kappa(x)}$ de $\overline{\kappa(x)}$ sur $\kappa(x)$ pour une certaine clôture algébrique $\overline{\kappa(x)}$ de $\kappa(x)$ et est donc isomorphe à $\mathrm{Frob}_x^{\widehat{\mathbb{Z}}}$ où $\mathrm{Frob}_x \in \Gamma_{\kappa(x)}$ est l'élément de Frobenius géométrique (l'inverse de l'élévation à la puissance $|\kappa(x)| = q^{\deg(x)}$).

Muni de la topologie de Krull, Γ_F est un groupe topologique pro-fini, dont les représentations complexes continues et de dimension finie se factorisent nécessairement par un quotient fini. Suivant Serre et Grothendieck, on obtient une catégorie plus vaste de représentations de Γ_F en remplaçant le corps des coefficients $\mathbb{C}$ par un corps ℓ-adique. Fixons donc un nombre premier auxiliaire ℓ, distinct de la caractéristique de $\mathbb{F}_q$, et fixons une clôture algébrique $\overline{\mathbb{Q}}_\ell$ de $\mathbb{Q}_\ell$.

Une *représentation ℓ-adique* σ de Γ_F est un $\overline{\mathbb{Q}}_\ell$-espace vectoriel V_σ de dimension finie, muni d'un homomorphisme de groupes $\sigma : \Gamma_F \to \mathrm{Aut}_{\overline{\mathbb{Q}}_\ell}(V_\sigma)$, ayant les propriétés suivantes :

1) il existe une base de V_σ identifiant $\mathrm{Aut}_{\overline{\mathbb{Q}}_\ell}(V_\sigma)$ à $\mathrm{GL}_r(\overline{\mathbb{Q}}_\ell)$, où r est la dimension de V_σ, et une extension finie E_λ de $\mathbb{Q}_\ell$ contenue dans $\overline{\mathbb{Q}}_\ell$ telles que $\sigma(\Gamma_F) \subset \mathrm{GL}_r(E_\lambda) \subset \mathrm{GL}_r(\overline{\mathbb{Q}}_\ell)$,

2) $\sigma : \Gamma_F \to \mathrm{GL}_r(E_\lambda)$ est continu pour la topologie de Krull sur Γ_F et la topologie ℓ-adique sur $\mathrm{GL}_r(E_\lambda)$,

3) pour presque tout $x \in |X|$, σ est *non ramifiée* en x, c'est-à-dire que la restriction σ_x de σ au sous-groupe de décomposition $D_x \subset \Gamma_F$ est triviale sur le sous-groupe d'inertie $I_x \subset D_x$.

Ces représentations forment une catégorie abélienne où tout objet est de longueur finie.

Soit $\mathcal{G}_r$ un système de représentants des classes d'isomorphie de représentations ℓ-adiques irréductibles de rang r de Γ_F dont le déterminant (la puissance extérieure maximale) est d'ordre fini. Pour chaque $\sigma \in \mathcal{G}_r$, on note S_σ l'ensemble fini des places de F où σ est ramifiée. Pour chaque $x \notin S_\sigma$, on dispose de l'automorphisme $\sigma_x(\mathrm{Frob}_x)$ de V_σ, et on note

$$(z_1(\sigma_x), \ldots, z_r(\sigma_x))$$

les *valeurs propres de Frobenius de σ en x*, c'est-à-dire le r-uplet non ordonné d'éléments non nuls de $\overline{\mathbb{Q}}_\ell$ formé des valeurs propres de $\sigma_x(\mathrm{Frob}_x)$. La *fonction L partielle* de σ est par définition le produit eulérien formel

$$L^{S_\sigma}(\sigma, T) = \prod_{x \in |X| - S_\sigma} L(\sigma_x, T) = \prod_{x \in |X| - S_\sigma} \frac{1}{\prod_{i=1}^r (1 - z_i(\sigma_x) T^{\deg(x)})} \in \overline{\mathbb{Q}}_\ell[[T]].$$

Il résulte de la formule des points fixes de Grothendieck-Lefschetz ([Gr]) que cette série formelle est le développement d'une fraction rationnelle dans $\overline{\mathbb{Q}}_\ell(T)$, et même d'un polynôme dans $\overline{\mathbb{Q}}_\ell[T]$ si $r \geq 2$.

1.3. La correspondance de Langlands et ses conséquences

On fixe un isomorphisme de corps $\iota : \overline{\mathbb{Q}}_\ell \xrightarrow{\sim} \mathbb{C}$ (il en existe d'après l'axiome du choix). On notera encore ι les isomorphismes de $\overline{\mathbb{Q}}_\ell(T)$ (resp. $\overline{\mathbb{Q}}_\ell[[T]]$) sur $\mathbb{C}(T)$ (resp. $\mathbb{C}[[T]]$) induits par ι.

Théorème Principal. — (i) (Correspondance de Langlands) *Il existe une unique bijection*

$$\mathcal{A}_r \xrightarrow{\sim} \mathcal{G}_r,\ \pi \mapsto \sigma(\pi),$$

telle que, pour tout $\pi \in \mathcal{A}_r$ on ait l'égalité de facteurs L locaux

$$\iota(L(\sigma(\pi)_x, T)) = L(\pi_x, T),$$

pour presque tout $x \notin S_{\sigma(\pi)} \cup S_\pi$.

(ii) (Conjecture de Ramanujan-Petersson) *Pour tout $\pi \in \mathcal{A}_r$ et toute place $x \notin S_\pi$ on a*

$$|z_i(\pi_x)| = 1,\ \forall i = 1, \ldots, r.$$

Le cas $r = 1$ de ce théorème est une reformulation de la théorie du corps de classes abélien pour les corps de fonctions. Le cas $r = 2$ a été démontré par Drinfeld ([Dr 6], [Dr 7]). Le cas général est dû à Lafforgue ([La 9]).

Les conséquences de ce théorème qui sont formulées ci-dessous étaient attendues, et leur déduction du théorème principal est bien connue.

Théorème (Compatibilité avec la correspondance de Langlands locale)

Pour tout $\pi \in \mathcal{A}_r$ *et tout* $x \in |X|$, *la représentation de Galois locale* $\sigma(\pi)_x$ *de* $D_x = \Gamma_{F_x}$ *est l'image* $\sigma_x(\pi_x)$ *de la représentation locale* π_x *de* $\mathrm{GL}_r(F_x)$ *par la correspondance de Langlands locale* (*cf.* [L-R-S]). *En particulier,* π *et* $\sigma(\pi)$ *ont les mêmes facteurs* L *locaux en toutes les places de* F.

Théorème (Conjecturé par Deligne, [De] (1.2.10)). — *Soit* $\sigma \in \mathcal{G}_r$.

(i) *Le sous-corps* $E = E(\sigma)$ *de* $\overline{\mathbb{Q}}_\ell$ *engendré sur* $\mathbb{Q}$ *par les coefficients des polynômes* $\prod_{i=1}^r (1 - z_i(\sigma_x)T)$ *pour tous les* $x \notin S_\sigma$ *est un corps de nombres* (*une extension finie de* $\mathbb{Q}$).

(ii) *Pour tout nombre premier* ℓ' *distinct de la caractéristique de* $\mathbb{F}_q$, *fixons une clôture algébrique* $\overline{\mathbb{Q}}_{\ell'}$ *de* $\mathbb{Q}_{\ell'}$ *et notons* $\mathcal{G}_{r,\ell'}$ *l'ensemble* $\mathcal{G}_r$ *correspondant.*

Il existe alors une unique famille de représentations $\sigma_{\ell',\lambda'} \in \mathcal{G}_{r,\ell'}$, *indexée par les couples* (ℓ', λ') *formés d'un nombre premier* ℓ' *distinct de la caractéristique de* $\mathbb{F}_q$ *et d'un plongement* $\lambda' : E \hookrightarrow \overline{\mathbb{Q}}_{\ell'}$, *et ayant les propriétés suivantes :*

– $\sigma = \sigma_{\ell,\lambda}$ *où* λ *est l'inclusion de* E *dans* $\overline{\mathbb{Q}}_\ell$,

– *chaque* $\sigma_{\ell',\lambda'}$ *a le même ensemble de places ramifiées que* σ *et pour tout* $x \notin S_\sigma$ *on a l'égalité*

$$\prod_{i=1}^r (1 - z_i(\sigma_{\ell',\lambda',x})T) = \lambda'\left(\prod_{i=1}^r (1 - z_i(\sigma_x)T)\right) \in \overline{\mathbb{Q}}_{\ell'}[T].$$

Théorème (Conjecturé par Deligne, [De] (1.2.10)). — *Soient* Y *un schéma normal de type fini sur le corps fini* $\mathbb{F}_q$ *et* $\mathcal{L}$ *un système local* ℓ*-adique de rang* r *sur* Y. *On suppose que* $\mathcal{L}$ *est irréductible et que sa puissance extérieure maximale* $\det(\mathcal{L})$ *est d'ordre fini* ($\det(\mathcal{L})^{\otimes N}$ *est isomorphe au faisceau* ℓ*-adique constant de valeur* $\overline{\mathbb{Q}}_\ell$ *sur* Y *pour tout entier* $N \geq 1$ *assez divisible*). *Alors* $\mathcal{L}$ *est pur de poids* 0.

Une conséquence plus indirecte du théorème principal est l'énoncé dit « de descente » dans la première construction de Drinfeld (cf. [Dr 12], [F-G-K-V], [Lau 5], [Lau 6]) des faisceaux automorphes partout non ramifiés pour GL_r sur une courbe lisse, projective et géométriquement connexe sur un corps arbitraire (éventuellement de caractéristique nulle).

1.4. Historique

L'outil principal, inventé par Drinfeld et utilisé par Drinfeld et Lafforgue pour démontrer le théorème principal, est le champ modulaire des chtoucas. Avant de découvrir ce champ, Drinfeld avait introduit des variétés modulaires sur les corps de fonctions très semblables aux variétés de Shimura sur les corps de nombres : les variétés de *modules elliptiques* (ou de *faisceaux elliptiques*). L'utilisation de ces variétés avait permis de construire $\sigma(\pi)$ sous certaines conditions sur la représentation locale π_∞ en une place donnée $\infty \in |X|$. Ainsi, Drinfeld ([Dr 9] et [Dr 10]) avait construit $\sigma(\pi)$ quand $r = 2$ et π_∞ est dans la série discrète, et Flicker et Kazhdan, puis moi-même, avions généralisé cette construction de Drinfeld, pour r arbitraire, dans les cas où π_∞ est soit supercuspidale ([F-K]), soit la représentation de Steinberg ([Lau 1], [Lau 2]).

L'assertion d'unicité du théorème principal était connue depuis longtemps : elle est en effet une conséquence immédiate du théorème de densité de Čebotarev ([Se]). Il en était de même de l'injectivité de l'application $\pi \to \sigma(\pi)$ qui est automatique d'après le théorème de multiplicité un fort de Piatetski-Shapiro ([PS 2]).

Deligne avait remarqué un principe de récurrence permettant de déduire la surjectivité de l'application $\pi \to \sigma(\pi)$ de son existence. Plus précisément, si pour tout $r' = 1, \dots, r-1$, on a construit une application $\pi' \to \sigma'(\pi')$ de $\mathcal{A}_{r'} \to \mathcal{G}_{r'}$ telle que, quel que soit $\pi' \in \mathcal{A}_{r'}$, on ait l'égalité de facteurs L locaux

$$\iota(L(\sigma'(\pi')_x, T)) = L(\pi'_x, T)$$

pour presque tout $x \notin S_{\sigma'(\pi')} \cup S_{\pi'}$, alors l'équation fonctionnelle de Grothendieck ([Gr]), la formule du produit pour la constante de cette équation fonctionnelle ([Lau 7]) et le théorème inverse de Hecke, Weil et Piatetski-Shapiro ([PS 1], [C-PS]) permettent de définir une application

$$\mathcal{G}_r \to \mathcal{A}_r, \ \sigma \mapsto \pi(\sigma),$$

telle que, quel que soit $\sigma \in \mathcal{G}_r$, on ait l'égalité de facteurs L locaux

$$\iota(L(\pi(\sigma)_x, T)) = L(\sigma_x, T)$$

pour presque tout $x \notin S_\sigma \cup S_{\pi(\sigma)}$.

Lafforgue avait obtenu auparavant un énoncé très proche de la conjecture de Ramanujan-Petersson ([La 1]) comme une conséquence de la description adélique de Drinfeld des chtoucas sur les corps finis ([Dr 6]), de la formule des traces d'Arthur-Selberg ([Ar]), de la formule des points fixes de Grothendieck-Lefschetz ([Gr]), du théorème de pureté de Deligne ([De]) et des estimées de Jacquet-Shalika pour les valeurs propres de Hecke des représentations automorphes cuspidales irréductibles ([J-S 1]).

1.5. La stratégie

La stratégie utilisée par Drinfeld et Lafforgue pour définir l'application $\pi \to \sigma(\pi)$ de $\mathcal{A}_r$ dans $\mathcal{G}_r$ est inspirée des travaux de Shimura, Ihara, Deligne, Langlands, ... Sous sa forme la plus naïve elle peut se décrire comme suit :

On construit un schéma V sur F muni d'une action de l'algèbre de Hecke $\mathcal{H}$ de telle sorte que la cohomologie ℓ-adique à supports compacts

$$H^*_c(\overline{F} \otimes_F V, \overline{\mathbb{Q}}_\ell)$$

soit une représentation du produit de $\mathrm{GL}_r(\mathbb{A})$ et du groupe de Galois Γ_F.

Puis on calcule la trace de cette représentation par la formule des points fixes de Grothendieck-Lefschetz.

Enfin on compare cette formule des points fixes avec la formule des traces d'Arthur-Selberg pour prouver que la représentation

$$\bigoplus_{\pi \in \mathcal{A}_r} \pi \otimes \sigma(\pi)$$

de $\mathrm{GL}_r(\mathbb{A}) \times \Gamma_F$ que l'on cherche est exactement la partie « cuspidale » de $H_c^*(\overline{F} \otimes_F V, \overline{\mathbb{Q}}_\ell)$.

En fait, comme la cohomologie ℓ-adique ci-dessus est automatiquement définie sur $\mathbb{Q}_\ell$, il y a une obstruction (de rationalité) à mener à bien un tel programme (cf. [Ka]) et la stratégie doit être légèrement modifiée. Comme l'a proposé Drinfeld, c'est plutôt la représentation

$$\bigoplus_{\pi \in \mathcal{A}_r} \pi \otimes \sigma(\pi)^\vee \otimes \sigma(\pi)$$

du produit $\mathrm{GL}_r(\mathbb{A}) \times \Gamma_F \times \Gamma_F$, où $\sigma(\pi)^\vee$ est la représentation contragrédiente de $\sigma(\pi)$, que l'on peut espérer obtenir comme la partie cuspidale de la cohomologie ℓ-adique à supports compacts d'un schéma sur $F \otimes F$.

2. CHTOUCAS DE DRINFELD

2.1. Le champ des chtoucas

Tous les schémas (ou champs) considérés seront sur $\mathbb{F}_q$; on dira donc schéma (ou champ) au lieu de $\mathbb{F}_q$-schéma (ou $\mathbb{F}_q$-champ) et on notera simplement $U \times T$ le produit $U \times_{\mathbb{F}_q} T$ de deux tels schémas (ou champs). Pour tout schéma (ou champ) U, on notera $\mathrm{Frob}_U : U \to U$ son endomorphisme de Frobenius relativement à $\mathbb{F}_q$: si U est un schéma, Frob_U est donc l'identité sur l'espace topologique sous-jacent à U et l'élévation à la puissance q-ième sur le faisceau structural $\mathcal{O}_U$. Si $\mathcal{M}$ est un $\mathcal{O}_{U\times X}$-Module, on notera ${}^\tau\mathcal{M}$ le $\mathcal{O}_{U\times X}$-Module $(\mathrm{Frob}_U \times \mathrm{Id}_X)^*\mathcal{M}$.

Définition (Drinfeld). — *Un* chtouca à droite (*resp.* à gauche) $\widetilde{\mathcal{E}}$ *de rang* $r \geq 1$ *sur un schéma* U *est un diagramme dans la catégorie abélienne des* $\mathcal{O}_{U\times X}$*-Modules*

$$\mathcal{E} \overset{j}{\hookrightarrow} \mathcal{E}' \overset{t}{\hookleftarrow} {}^\tau\mathcal{E} \ (\textit{resp. } \mathcal{E} \overset{t}{\hookleftarrow} \mathcal{E}' \overset{j}{\hookrightarrow} {}^\tau\mathcal{E}\,)$$

où :

– $\mathcal{E}$ et $\mathcal{E}'$ sont localement libres de rang r,

– j et t sont injectifs,

– les conoyaux de j et t sont supportés par les graphes $\Gamma_\infty \subset U \times X$ et $\Gamma_o \subset U \times X$ de deux morphismes $\infty : U \to X$ et $o : U \to X$, et sont localement libres de rang 1 *sur leurs supports.*

Les morphismes ∞ et o sont appelés respectivement le pôle *et le* zéro *du chtouca.*

En d'autres termes, un chtouca à droite de rang r est la donnée d'un fibré vectoriel $\mathcal{E}$ de rang r, d'une modification élémentaire supérieure $j : \mathcal{E} \hookrightarrow \mathcal{E}'$ de $\mathcal{E}$, d'une modification élémentaire inférieure $j' : \mathcal{E}'' \hookrightarrow \mathcal{E}'$ de $\mathcal{E}'$ et d'un isomorphisme $u : {}^\tau\mathcal{E} \xrightarrow{\sim} \mathcal{E}''$. On a bien sûr une description similaire pour les chtoucas à gauche.

En faisant varier U, on définit de manière évidente le champ Cht^r des chtoucas à droite de rang r, un morphisme $(\infty, o) : \mathrm{Cht}^r \to X \times X$ et un chtouca à droite de rang r universel sur $\mathrm{Cht}^r \times X$ de pôle et de zéro les deux composantes de ce morphisme. De même, on a le champ ${}^r\mathrm{Cht}$ des chtoucas à gauche de rang r, un morphisme $(\infty, o) : {}^r\mathrm{Cht} \to X \times X$ et un chtouca à gauche de rang r universel.

On a des morphismes de Frobenius partiels

$$\mathrm{Frob}_\infty : {}^r\mathrm{Cht} \to \mathrm{Cht}^r \ \text{ et } \ \mathrm{Frob}_o : \mathrm{Cht}^r \to {}^r\mathrm{Cht}$$

au-dessus de $(\infty, o) \mapsto (\infty, \mathrm{Frob}_X(o))$ et $(\infty, o) \mapsto (\mathrm{Frob}_X(\infty), o)$, qui envoient $\widetilde{\mathcal{E}}$ sur $(\mathcal{E}' \xhookrightarrow{j} {}^\tau\mathcal{E} \xhookleftarrow{{}^\tau t} {}^\tau\mathcal{E}')$ et $(\mathcal{E}' \xhookleftarrow{t} {}^\tau\mathcal{E} \xhookrightarrow{{}^\tau j} {}^\tau\mathcal{E}')$ respectivement.

Dans la suite et sauf mention explicite du contraire, les chtoucas seront tous à droite, et on dira simplement « chtouca » pour « chtouca à droite ». Les résultats démontrés pour les chtoucas à droite ont bien entendu des analogues pour les chtoucas à gauche.

Proposition (Drinfeld). — *Le champ* Cht^r *est algébrique au sens de Deligne-Mumford. Le morphisme* $(\infty, o) : \mathrm{Cht}^r \to X \times X$ *est lisse, purement de dimension relative* $2r - 2$.

Le champ Cht^r s'écrit comme réunion disjointe

$$\mathrm{Cht}^r = \coprod_{d \in \mathbb{Z}} \mathrm{Cht}^{r,d}$$

où $\mathrm{Cht}^{r,d}$ classifie les chtoucas de degré

$$d = \deg \mathcal{E} = \deg \mathcal{E}' - 1.$$

Exemple. — Pour tout entier d, $\mathrm{Cht}^{1,d}$ peut être défini comme le produit fibré

$$\begin{array}{ccc} \mathrm{Cht}^{1,d} & \longrightarrow & \mathrm{Fib}^{1,d} \\ \big\downarrow & \square & \big\downarrow {\scriptstyle L} \\ X \times X & \xrightarrow[J]{} & \mathrm{Fib}^{1,0} \end{array}$$

où $\mathrm{Fib}^{1,d}$ est le champ algébrique des fibrés en droites de degré d sur X, J est l'application d'Abel-Jacobi qui envoie $(\infty, o) \in (X \times X)(U)$ sur le fibré en droites $\mathcal{O}_{U \times X}(\infty - o)$, et L est l'isogénie de Lang qui envoie un fibré en droites $\mathcal{L}$ sur $U \times X$ sur le fibré en droites $\mathcal{L}^{-1} \otimes_{\mathcal{O}_{U \times X}} {}^\tau\mathcal{L}$.

En particulier, $\mathrm{Cht}^{1,d}$ est de type fini et admet un espace grossier qui est un revêtement fini étale galoisien de $X \times X$.

Mais, mis à part le cas $r = 1$, *aucun des champs* $\mathrm{Cht}^{r,d}$ *n'est de type fini.*

2.2. Troncatures

Soient k un corps algébriquement clos contenant $\mathbb{F}_q$ et $\widetilde{\mathcal{E}}$ un chtouca sur (le spectre de) k. On appelle *sous-objet* de $\widetilde{\mathcal{E}}$, et on note simplement

$$\widetilde{\mathcal{F}} \subset \widetilde{\mathcal{E}},$$

la donnée de deux sous-$\mathcal{O}_{k\otimes X}$-Modules $\mathcal{F} \subset \mathcal{E}$ et $\mathcal{F}' \subset \mathcal{E}'$ tels que $\mathcal{E}/\mathcal{F}$ et $\mathcal{E}'/\mathcal{F}'$ soient localement libres de même rang et que $j(\mathcal{F}) \subset \mathcal{F}'$ et $t({}^{\tau}\mathcal{F}) \subset \mathcal{F}'$; à un tel sous-objet on peut associer son *rang*

$$\operatorname{rg} \widetilde{\mathcal{F}} = \operatorname{rg} \mathcal{F} = \operatorname{rg} \mathcal{F}'$$

et, pour chaque nombre réel α, son degré (d'indice α)

$$\deg_{\alpha} \widetilde{\mathcal{F}} = (1-\alpha) \deg \mathcal{F} + \alpha \deg \mathcal{F}'.$$

Si $0 = \widetilde{\mathcal{F}}_0 \subsetneq \widetilde{\mathcal{F}}_1 \subsetneq \cdots \subsetneq \widetilde{\mathcal{F}}_s = \widetilde{\mathcal{E}}$ est une filtration de $\widetilde{\mathcal{E}}$ par des sous-objets comme ci-dessus, on peut lui associer son polygone (d'indice α) qui est la fonction affine par morceaux

$$p : [0, r] \to \mathbb{R}$$

avec $p(0) = p(r) = 0$, dont les seules ruptures de pentes interviennent en les entiers $\operatorname{rg} \widetilde{\mathcal{F}}_{\sigma}$, $\sigma = 1, \ldots, s-1$, et qui prend en ces entiers-là la valeur

$$p(\operatorname{rg} \widetilde{\mathcal{F}}_{\sigma}) = \deg_{\alpha} \widetilde{\mathcal{F}}_{\sigma} - \frac{\operatorname{rg} \widetilde{\mathcal{F}}_{\sigma}}{r} \deg_{\alpha} \widetilde{\mathcal{E}}.$$

Proposition. — *Fixons un nombre réel* $\alpha \in [0, 1]$. *Alors, parmi tous les polygones attachés aux filtrations de* $\widetilde{\mathcal{E}}$ *comme ci-dessus, il en existe un plus grand que tous les autres, et parmi toutes les filtrations qui définissent ce polygone maximal il en existe une moins fine que tous les autres.*

Le polygone et la filtration dont la proposition ci-dessus assurent l'existence sont appelés respectivement le *polygone de Harder-Narasimhan* et la *filtration de Harder-Narasimhan* d'indice α du chtouca $\widetilde{\mathcal{E}}$.

Nous appellerons *paramètre de troncature* toute fonction continue, convexe, affine par morceaux $p : [0, r] \to \mathbb{R}_{\geq 0}$ avec $p(0) = p(r) = 0$ dont les points de ruptures de pente ont des abscisses entières.

Proposition. — *Étant donnés* $\alpha \in [0, 1]$ *et un paramètre de troncature* p, *il existe un unique ouvert* $\mathrm{Cht}^{r;\, \leq_{\alpha} p}$ *du champ* Cht^{r} *tel qu'un chtouca sur un corps algébriquement clos est dans cet ouvert si et seulement si son polygone de Harder-Narasimhan d'indice* α *est majoré par* p.

De plus, pour chaque entier d, *l'ouvert*

$$\mathrm{Cht}^{r,d;\, \leq_{\alpha} p} := \mathrm{Cht}^{r;\, \leq_{\alpha} p} \cap \mathrm{Cht}^{r,d}$$

du champ $\mathrm{Cht}^{r,d}$ *est de type fini.*

On voit donc que, pour chaque entier d, le champ algébrique localement de type fini $\mathrm{Cht}^{r,d}$ est réunion filtrante des ouverts de type fini $\mathrm{Cht}_{\alpha}^{r,d;\,\leq p}$.

2.3. Structures de niveau sur les chtoucas

Comme les courbes elliptiques, les chtoucas admettent des structures de niveau. Un *niveau* est ici un sous-schéma fermé fini

$$N = \mathrm{Spec}(\mathcal{O}_N) \subset X.$$

Si $\mathcal{E}$ est un fibré vectoriel sur $U \times X$ pour un schéma (ou un champ) U, on notera simplement $\mathcal{E}_N$ la restriction de $\mathcal{E}$ au fermé $U \times N \subset U \times X$. Si $\mathcal{M}$ est un $\mathcal{O}_{U\times N}$-Module on notera encore ${}^{\tau}\mathcal{M}$ le $\mathcal{O}_{U\times N}$-Module $(\mathrm{Frob}_U \times \mathrm{Id}_N)^*\mathcal{M}$. Pour tout fibré vectoriel $\mathcal{E}$ sur $U \times X$ on a évidemment ${}^{\tau}(\mathcal{E}_N) \cong ({}^{\tau}\mathcal{E})_N$.

On considère le champ algébrique

$$\mathrm{Tr}_N^r$$

qui associe au schéma U la catégorie des couples $(\mathcal{F}, u)$ formés d'un fibré vectoriel de rang r sur $U \times N$ et d'un isomorphisme de fibrés vectoriels $u : {}^{\tau}\mathcal{F} \xrightarrow{\sim} \mathcal{F}$. Si l'on note $\mathrm{GL}_r^N = \mathrm{Res}_{N/\mathrm{Spec}(\mathbb{F}_q)} \mathrm{GL}_r$ le schéma en groupes sur $\mathbb{F}_q$ restriction à la Weil du schéma en groupes linéaire GL_r sur N et par $h \mapsto \tau(h) = \mathrm{Frob}_{\mathrm{GL}_r}(h)$ son endomorphisme de Frobenius (relatif à $\mathbb{F}_q$), ce champ n'est autre que le quotient

$$\mathrm{Tr}_N^r = [\mathrm{GL}_r^N / {}^{\tau}\mathrm{Ad}(\mathrm{GL}_r^N)]$$

de GL_r^N par l'action de GL_r^N sur lui-même par la conjugaison tordue

$${}^{\tau}\mathrm{Ad}(h)(g) = \tau(h)^{-1} g h.$$

Comme l'*isogénie de Lang*

$$L : \mathrm{GL}_r^N \to \mathrm{GL}_r^N,\ h \mapsto L(h) = \tau(h)^{-1} h,$$

est un torseur sous le groupe fini $\mathrm{GL}_r(\mathcal{O}_N) = \mathrm{GL}_r^N(\mathbb{F}_q)$ (pour son action par translation à droite sur GL_r^N), Tr_N^r est canoniquement isomorphe au champ classifiant $[\mathrm{Spec}(\mathbb{F}_q)/\mathrm{GL}_r(\mathcal{O}_N)]$.

On ne munira ici de structures de niveau que les chtoucas dont le pôle et le zéro ne rencontrent pas N. Si

$$\widetilde{\mathcal{E}} = (\mathcal{E} \xhookrightarrow{j} \mathcal{E}' \xhookleftarrow{t} {}^{\tau}\mathcal{E})$$

est un tel chtouca sur U on remarque que les restrictions

$$j_N : \mathcal{E}_N \to \mathcal{E}'_N \text{ et } t_N : {}^{\tau}(\mathcal{E}_N) \cong ({}^{\tau}\mathcal{E})_N \to \mathcal{E}'_N$$

des flèches j et t sont des isomorphismes de $\mathcal{O}_{U\times N}$-Modules. On a donc un morphisme de champs algébriques

$$\mathrm{Cht}^r \times_{X^2} (X - N)^2 \longrightarrow \mathrm{Tr}_N^r,\ \widetilde{\mathcal{E}} \mapsto (\mathcal{E}_N, j_N^{-1} \circ t_N).$$

DÉFINITION (Drinfeld). — *Soit* $\widetilde{\mathcal{E}}$ *un chtouca de rang* r *sur un schéma* U *dont le pôle et le zéro se factorisent par l'immersion ouverte* $X - N \hookrightarrow X$. *Une* structure (principale) de niveau N *sur ce chtouca est la donnée d'un isomorphisme*

$$\iota : \mathcal{O}_{U\times N}^r \xrightarrow{\sim} \mathcal{E}_N$$

tel que

$$\iota = (j_N^{-1} \circ t_N) \circ {}^\tau\iota : \mathcal{O}_{U\times N}^r \xrightarrow{\sim} \mathcal{E}_N.$$

On définit de manière évidente le champ Cht_N^r des chtoucas de rang r, dont le pôle et le zéro ne rencontrent pas N, et avec structures de niveau N. On a un carré cartésien

$$\begin{array}{ccc} \mathrm{Cht}_N^r & \longrightarrow & \mathrm{Tr}_N^{r,\tau} \\ \big\downarrow & \square & \big\downarrow \\ \mathrm{Cht}^r \times_{X^2} (X-N)^2 & \longrightarrow & \mathrm{Tr}_N^r \end{array}$$

où la flèche horizontale du bas est la flèche de restriction à N définie ci-dessus, où $\mathrm{Tr}_N^{r,\tau}$ est le champ qui associe au schéma U la catégorie des triplets $(\mathcal{F}, u, \iota)$ formés d'un objet $(\mathcal{F}, u)$ de $\mathrm{Tr}_N^r(U)$ et d'un isomorphisme de $\mathcal{O}_{U\times N}$-Modules $\iota : \mathcal{O}_{U\times N}^r \xrightarrow{\sim} \mathcal{F}$ tel que $\iota = u \circ {}^\tau\iota$, et où la flèche verticale de droite est l'oubli de la composante ι. On remarquera que la composante ι de $(\mathcal{F}, u, \iota)$ détermine uniquement la composante u et donc que $\mathrm{Tr}_N^{r,\tau}$ est canoniquement isomorphe au champ $[\mathrm{GL}_r^N / \mathrm{GL}_r^N]$ quotient de GL_r^N par l'action de GL_r^N sur lui-même par translation à droite, c'est-à-dire au schéma réduit à un point $\mathrm{Spec}(\mathbb{F}_q)$. Par conséquent, la flèche verticale de droite du carré ci-dessus se décrit encore, soit comme la flèche $[\mathrm{GL}_r^N / \mathrm{GL}_r^N] \to [\mathrm{GL}_r^N / {}^\tau\mathrm{Ad}(\mathrm{GL}_r^N)]$ induite par l'isogénie de Lang L, soit comme le $\mathrm{GL}_r(\mathcal{O}_N)$-torseur canonique $\mathrm{Spec}(\mathbb{F}_q) \to [\mathrm{Spec}(\mathbb{F}_q)/ \mathrm{GL}_r(\mathcal{O}_N)]$.

La flèche verticale de gauche est donc un $\mathrm{GL}_r(\mathcal{O}_N)$-torseur et Cht_N^r est un champ algébrique, lisse purement de dimension $2r-2$ sur $(X-N)^2$.

3. HOMOMORPHISMES COMPLETS ET CHTOUCAS ITÉRÉS

3.1. Homomorphismes complets

Soient k un corps et V et W deux k-espaces vectoriels de dimension r. On notera $H(V, W)$ le schéma des uplets

$$(u_1, \ldots, u_r; \lambda_1, \ldots, \lambda_{r-1})$$

où, pour $\rho = 1, \ldots, r$, u_ρ est une application linéaire non nulle de $\wedge^\rho V$ dans $\wedge^\rho W$ et où $\lambda_1, \ldots, \lambda_{r-1}$ sont des scalaires. Les conditions

– u_1 est un isomorphisme,
– $\lambda_1, \ldots, \lambda_{r-1}$ sont inversibles,
– $\wedge^\rho u_1 = \lambda_1^{\rho-1}\lambda_2^{\rho-2}\cdots\lambda_{\rho-1}u_\rho$ pour $\rho = 2, \ldots, r$,

définissent un sous-schéma localement fermé $\Omega^\circ(V,W)$ de ce schéma affine et la projection $(u_1;\lambda_1,\ldots,\lambda_{r-1})$ identifie $\Omega^\circ(V,W)$ au schéma $\mathrm{Isom}(V,W)\times_k \mathbb{G}_{\mathrm{m},k}^{r-1}$. Nous noterons

$$\Omega(V,W)\subset H(V,W)$$

l'adhérence schématique de $\Omega^\circ(V,W)$ dans $H(V,W)$.

Le schéma $\mathbb{A}_k^{r-1}$ est naturellement muni d'un diviseur à croisements normaux, réunion de $r-1$ diviseurs lisses, à savoir les diviseurs $\{\lambda_\rho=0\}$ pour $\rho=1,\ldots,r-1$. Ce diviseur à croisement normaux définit de la manière habituelle une stratification localement fermée de $\mathbb{A}_k^{r-1}$, indexée par les parties R de $[r-1]=\{1,\ldots,r-1\}$ et de R-ième strate

$$\mathbb{A}_{k,R}^{r-1}=\{(\lambda_1,\ldots,\lambda_{r-1})\mid \lambda_\rho=0\Leftrightarrow \rho\in R\}\cong \mathbb{G}_{\mathrm{m},k}^{[r-1]-R}.$$

Par construction, on dispose d'un morphisme

$$(\lambda_1,\ldots,\lambda_{r-1}):\Omega(V,W)\to \mathbb{A}_k^{r-1}.$$

et on peut relever à $\Omega(V,W)$ le diviseur et la stratification de $\mathbb{A}_k^{r-1}$ que l'on vient d'introduire. En particulier, pour tout sous-ensemble $R\subset[r-1]$, on notera

$$\Omega_R(V,W)\subset\Omega(V,W)$$

l'image réciproque par ce morphisme de la strate localement fermée $\mathbb{A}_{k,R}^{r-1}\subset\mathbb{A}_k^{r-1}$. On vérifie que $\Omega_\varnothing(V,W)=\Omega^\circ(V,W)$.

Proposition. — *Pour chaque $R=\{r_1,r_2,\ldots,r_{s-1}\}\subset[r-1]$ où $0=r_0<r_1<r_2<\cdots<r_{s-1}<r_s=r$, $\Omega_R(V,W)$ est isomorphe au schéma des uplets*

$$(V^\bullet,W_\bullet,(v_\sigma)_{\sigma=1,\ldots,s-1};(\lambda_\rho)_{\rho\in[r-1]-R})$$

où :

– les λ_ρ sont des scalaires inversibles ;

– $V^\bullet=(V=V^0\supsetneq V^1\supsetneq\cdots\supsetneq V^s=(0))$ est une filtration décroissante par des sous-espaces vectoriels de codimensions $0=r_0$, $r_1,\ldots$, $r_s=r$;

– $W_\bullet=((0)=W_0\subsetneq W_1\subsetneq\cdots\subsetneq W_s=W)$ est une filtration croissante par des sous-espaces vectoriels de dimensions $0=r_0$, $r_1,\ldots$, $r_s=r$;

– $v_\sigma:V^{\sigma-1}/V^\sigma\xrightarrow{\sim}W_\sigma/W_{\sigma-1}$ est un isomorphisme d'espaces vectoriels.

A fortiori, le morphisme $(\lambda_\rho)_{\rho\in[r-1]-R}:\Omega_R(V,W)\to\mathbb{G}_{\mathrm{m},k}^{[r-1]-R}$ est lisse de dimension relative r^2.

Remarque. — L'écriture $R=\{r_1,r_2,\ldots,r_{s-1}\}$ où $0=r_0<r_1<r_2<\cdots<r_{s-1}<r_s=r$ identifie les parties de $[r-1]$ aux décompositions $r=(r_1-r_0)+(r_2-r_1)+\cdots+(r_s-r_{s-1})$ de r en entiers strictement positifs.

Corollaire. — *Le morphisme $(\lambda_1,\ldots,\lambda_{r-1})$ est lisse purement de dimension relative r^2 et $\Omega^\circ(V,W)\subset\Omega(V,W)$ est l'ouvert complémentaire d'un diviseur à croisements normaux, réunion des $r-1$ diviseurs lisses $\{\lambda_\rho=0\}$ pour $\rho=1,\ldots,r-1$.*

Le tore $\mathbb{G}_{\mathrm{m},k}^{r-1} = \{(\mu_1, \ldots, \mu_{r-1})\}$ agit librement sur $\Omega(V, W)$ par

$$u_1 \mapsto u_1,\ u_2 \mapsto \mu_1^{-1} u_2,\ u_3 \mapsto \mu_1^{-2}\mu_2 u_3,\ \ldots,\ u_r \mapsto \mu_1^{1-r}\mu_2^{2-r}\cdots\mu_{r-1}u_r,$$

et

$$\lambda_1 \mapsto \mu_1\lambda_1,\ \lambda_2 \mapsto \mu_2\lambda_2,\ \ldots,\ \lambda_{r-1} \mapsto \mu_{r-1}\lambda_{r-1},$$

et le quotient

$$\widetilde{\mathrm{Hom}}(V, W) := \Omega(V, W)/\mathbb{G}_{\mathrm{m},k}^{r-1}$$

est un schéma quasi-projectif et lisse, qui contient comme ouvert dense $\mathrm{Isom}(V, W)$ avec pour fermé complémentaire une réunion de $r - 1$ diviseurs lisses à croisements normaux. C'est par définition le *schéma des homomorphismes complets de* V dans W. Par construction, on a un morphisme représentable, lisse et purement de dimension relative r^2, de champs algébriques

$$\widetilde{\mathrm{Hom}}(V, W) \to [\mathbb{A}_k^{r-1}/\mathbb{G}_{\mathrm{m},k}^{r-1}],$$

et le diviseur à croisements normaux ci-dessus est l'image réciproque par ce morphisme du diviseur à croisements normaux évident sur le champ algébrique $[\mathbb{A}_k^{r-1}/\mathbb{G}_{\mathrm{m},k}^{r-1}]$.

Si $V = W = k^r$ on notera encore $\widetilde{\mathrm{gl}}_{r,k}$ le k-schéma $\widetilde{\mathrm{Hom}}(k^r, k^r)$ des *endomorphismes complets* de k^r. La strate ouverte de $\widetilde{\mathrm{gl}}_{r,k}$ est bien entendu $\mathrm{GL}_{r,k}$.

Remarque. — (i) Le morphisme

$$u_1 : \widetilde{\mathrm{Hom}}(V, W) \to \mathrm{Hom}(V, W) - \{0\}$$

peut aussi être obtenu comme le composé

$$\widetilde{\mathrm{Hom}}(V, W) = H_{r-1} \to \cdots \to H_1 \to H_0 = \mathrm{Hom}(V, W) - \{0\}$$

où $H_1 \to H_0$ est l'éclatement de H_0 le long du fermé des homomorphismes de rang 1 et où, pour $\rho = 2, \ldots, r-1$, $H_\rho \to H_{\rho-1}$ est l'éclatement de $H_{\rho-1}$ le long du transformé strict du fermé de H_0 formé des homomorphismes de rang $\leq \rho$.

(ii) Le groupe multiplicatif agit par homothétie sur les homomorphismes complets. Le quotient $\widetilde{\mathrm{gl}}_{r,k}/\mathbb{G}_{\mathrm{m},k}$ est un schéma projectif, isomorphe à la compactification de De Concini et Procesi de $\mathrm{PGL}_{r,k}$.

Pour tout k-schéma U, la catégorie $[\mathbb{A}_k^{r-1}/\mathbb{G}_{\mathrm{m},k}^{r-1}](U)$ a pour objets les uplets

$$((\mathcal{L}_1, \lambda_1), \ldots, (\mathcal{L}_{r-1}, \lambda_{r-1}))$$

où $\mathcal{L}_1, \ldots, \mathcal{L}_{r-1}$ sont des $\mathcal{O}_U$-Modules inversibles et où $\lambda_1, \ldots, \lambda_{r-1}$ sont des sections globales de ces fibrés en droites. On peut donc voir un U-homomorphisme complet de V dans W, c'est-à-dire un U-point de $\widetilde{\mathrm{Hom}}(V, W)$, comme un uplet

$$u = (u_1, \ldots, u_r; (\mathcal{L}_1, \lambda_1), \ldots, (\mathcal{L}_{r-1}, \lambda_{r-1}))$$

où $((\mathcal{L}_1, \lambda_1), \ldots, (\mathcal{L}_{r-1}, \lambda_{r-1})) \in \mathrm{ob}\,[\mathbb{A}_k^{r-1}/\mathbb{G}_{\mathrm{m},k}^{r-1}](U)$ et où

$$u_\rho : \bigwedge^\rho V \otimes_k \mathcal{L}_1^{\otimes(\rho-1)} \otimes \mathcal{L}_2^{\otimes(\rho-2)} \otimes \cdots \otimes \mathcal{L}_{\rho-1} \to \bigwedge^\rho W$$

est un homomorphisme partout non nul de $\mathcal{O}_U$-Modules pour $\rho = 1, \ldots, r$. Un homomorphisme complet u est donc bien un homomorphisme u_1 *complété* par des données supplémentaires et peut être noté commodément $u : V \Rightarrow W$.

Soient maintenant S un k-schéma et $\mathcal{V}$, $\mathcal{W}$ deux $\mathcal{O}_S$-modules localement libres de rang constant r. Si U est un S-schéma, on peut considérer plus généralement les uplets

$$u = (u_1, \ldots, u_r; (\mathcal{L}_1, \lambda_1), \ldots, (\mathcal{L}_{r-1}, \lambda_{r-1}))$$

où $((\mathcal{L}_1, \lambda_1), \ldots, (\mathcal{L}_{r-1}, \lambda_{r-1})) \in \mathrm{ob}\,[\mathbb{A}_k^{r-1}/\mathbb{G}_{\mathrm{m},k}^{r-1}](U)$ et où

$$u_\rho : \bigwedge^{\rho} \mathcal{V}_U \otimes \mathcal{L}_1^{\otimes(\rho-1)} \otimes \mathcal{L}_2^{\otimes(\rho-2)} \otimes \cdots \otimes \mathcal{L}_{\rho-1} \to \bigwedge^{\rho} \mathcal{W}_U,$$

est un homomorphisme partout non nul de $\mathcal{O}_U$-Modules pour $\rho = 1, \ldots, r$. (On a noté $\mathcal{V}_U$ et $\mathcal{W}_U$ les restrictions de $\mathcal{V}$ et $\mathcal{W}$ à U.) On dira alors qu'un tel uplet est un *U-homomorphisme complet de $\mathcal{V}$ dans $\mathcal{W}$*, et on utilisera la notation $u : \mathcal{V} \Rightarrow \mathcal{W}$, si, pour toutes trivialisations locales $\mathcal{V} \cong V \otimes_k \mathcal{O}_S$ et $\mathcal{W} \cong W \otimes_k \mathcal{O}_S$, ce uplet est un U-homomorphisme complet de V dans W.

On définit ainsi un S-schéma $\widetilde{\mathrm{Hom}}(\mathcal{V}, \mathcal{W})$, muni d'un morphisme représentable, lisse et purement de dimension relative r^2, de champs algébriques

$$\widetilde{\mathrm{Hom}}(\mathcal{V}, \mathcal{W}) \to S \times_k [\mathbb{A}_k^{r-1}/\mathbb{G}_{\mathrm{m},k}^{r-1}]$$

Le S-schéma quasi-projectif $\widetilde{\mathrm{Hom}}(\mathcal{V}, \mathcal{W})$ contient le S-schéma $\mathrm{Isom}(\mathcal{V}, \mathcal{W})$ comme un ouvert dense, et le fermé complémentaire est un diviseur à croisement normaux relatifs sur S, réunion de $r-1$ diviseurs lisses.

3.2. Chtoucas itérés

DÉFINITION. — *Un* pré-chtouca itéré *de rang r sur un schéma U est la donnée :*

– *d'un diagramme*

$$\mathcal{E} \overset{j}{\hookrightarrow} \mathcal{E}' \overset{j'}{\hookleftarrow} \mathcal{E}''$$

et de deux morphismes $\infty : U \to X$ et $o : U \to X$, où $\mathcal{E}$ est un fibré vectoriel de rang r sur $U \times X$, j est une modification élémentaire supérieure de $\mathcal{E}$ le long du graphe de ∞ et j' est une modification élémentaire inférieure de $\mathcal{E}'$ le long du graphe de o,

– *de $\mathcal{O}_U$-Modules inversibles $\mathcal{L}_1, \ldots, \mathcal{L}_{r-1}$ munis de sections globales $\lambda_1, \ldots, \lambda_{r-1}$,*

– *pour chaque entier $\rho = 1, \ldots, r$, d'un homomorphisme de $\mathcal{O}_{U\times X}$-Modules*

$$u_\rho : \wedge^\rho({}^\tau\mathcal{E}) \otimes \mathcal{L}_1^{\otimes(q-1)(\rho-1)} \otimes \mathcal{L}_2^{\otimes(q-1)(\rho-2)} \otimes \cdots \otimes \mathcal{L}_{\rho-1}^{\otimes(q-1)} \to \wedge^\rho \mathcal{E}'',$$

de telle sorte que le uplet

$$u = (u_1, \ldots, u_r; (\mathcal{L}_1^{\otimes(q-1)}, \lambda_1^{q-1}), \ldots, (\mathcal{L}_{r-1}^{\otimes(q-1)}, \lambda_{r-1}^{q-1})) : {}^\tau\mathcal{E} \Rightarrow \mathcal{E}''$$

soit un homomorphisme complet.

En faisant varier U, on définit de manière évidente le champ $\mathcal{C}^r$ des pré-chtoucas itérés de rang r. Il n'est pas difficile de vérifier que c'est un champ algébrique (au sens d'Artin), localement de type fini, muni d'un morphisme de champs

$$((\infty, o), ((\mathcal{L}_1, \lambda_1), \dots, (\mathcal{L}_{r-1}, \lambda_{r-1}))) : \mathcal{C}^r \to X \times X \times [\mathbb{A}^{r-1}/\mathbb{G}_{\mathrm{m}}^{r-1}].$$

En particulier, pour tout $R \subset [r-1]$, on dispose de la strate localement fermée $\mathcal{C}^r_R \subset \mathcal{C}^r$ image réciproque de la strate $[\mathbb{A}^{r-1}_R/\mathbb{G}_{\mathrm{m}}^{r-1}] \subset [\mathbb{A}^{r-1}/\mathbb{G}_{\mathrm{m}}^{r-1}]$.

D'après la proposition du paragraphe 3.1, si $R = \{r_1, \dots, r_{s-1}\}$ où $0 = r_0 < r_1 < \cdots < r_{s-1} < r_s = r$, la donnée d'un U-point de $\mathcal{C}^r_R$ au-dessus d'un U-point $((\mathcal{L}_1, \lambda_1), \dots, (\mathcal{L}_{r-1}, \lambda_{r-1}))$ de $[\mathbb{A}^{r-1}_R/\mathbb{G}_{\mathrm{m}}^{r-1}]$ équivaut aux données suivantes :

– un diagramme

$$\mathcal{E} \xhookrightarrow{j} \mathcal{E}' \xhookleftarrow{j'} \mathcal{E}''$$

et deux morphismes $\infty : U \to X$ et $o : U \to X$, où $\mathcal{E}$ est un fibré vectoriel de rang r sur $U \times X$, j est une modification élémentaire supérieure de $\mathcal{E}$ le long du graphe de ∞ et j' est une modification élémentaire inférieure de $\mathcal{E}'$ le long du graphe de o ;

– une filtration décroissante

$${}^\tau\mathcal{E} = \mathcal{F}^0 \supsetneq \mathcal{F}^1 \supsetneq \cdots \supsetneq \mathcal{F}^{s-1} \supsetneq \mathcal{F}^s = (0)$$

de ${}^\tau\mathcal{E}$ par des sous-$\mathcal{O}_{U\times X}$-Modules localement facteurs directs, de corangs $0 = r_0$, $r_1, \dots, r_{s-1}$, $r_s = r$;

– une filtration croissante

$$(0) = \mathcal{E}''_0 \subsetneq \mathcal{E}''_1 \subsetneq \cdots \subsetneq \mathcal{E}''_{s-1} \subsetneq \mathcal{E}''_s = \mathcal{E}''$$

de $\mathcal{E}''$ par des sous-$\mathcal{O}_{U\times X}$-Modules localement facteurs directs, de rangs $0 = r_0, r_1, \dots$, r_{s-1}, $r_s = r$;

– une famille d'isomorphismes de $\mathcal{O}_{U\times X}$-Modules

$$v_\sigma : (\mathcal{F}^{\sigma-1}/\mathcal{F}^\sigma) \otimes \left(\bigotimes_{\sigma'=1}^{\sigma-1} {}^\tau\mathcal{L}_{r_{\sigma'}}\right) \xrightarrow{\sim} (\mathcal{E}''_\sigma/\mathcal{E}''_{\sigma-1}) \otimes \left(\bigotimes_{\sigma'=0}^{\sigma-1} \mathcal{L}_{r_{\sigma'}}\right), \ \sigma = 1, \dots, s.$$

Pour un tel U-point notons

$$\mathcal{E}'_\sigma = j'(\mathcal{E}''_\sigma) \subset \mathcal{E}', \ \mathcal{E}_\sigma = j^{-1}(\mathcal{E}'_\sigma) \subset \mathcal{E}, \ \forall \sigma = 0, 1, \dots, s-1, \text{ et } \mathcal{E}'_s = \mathcal{E}', \ \mathcal{E}_s = \mathcal{E},$$

$$\mathcal{G}_1 = \mathcal{E}_1, \ \mathcal{G}'_1 = \mathcal{E}'_1 \text{ et } \mathcal{G}_\sigma = \mathcal{E}_\sigma/\mathcal{E}_{\sigma-1}, \ \mathcal{G}'_\sigma = \mathcal{F}^{\sigma-1} \cap {}^\tau\mathcal{E}_\sigma, \ \forall \sigma = 2, \dots, s.$$

On dispose de l'homomorphisme $\mathcal{G}_1 \hookrightarrow \mathcal{G}'_1$ induit par j et du composé

$${}^\tau\mathcal{G}_1 \hookrightarrow {}^\tau\mathcal{E} \twoheadrightarrow {}^\tau\mathcal{E}/\mathcal{F}^1 \xrightarrow{\sim} \mathcal{E}''_1 \xrightarrow{j'} \mathcal{G}'_1,$$

et, pour chaque $\sigma = 2, \dots, r$, on dispose du composé

$$\mathcal{G}'_\sigma \otimes \left(\bigotimes_{\sigma'=0}^{\sigma-1} {}^\tau\mathcal{L}_{r_{\sigma'}}\right) \hookrightarrow \mathcal{F}^{\sigma-1} \otimes \left(\bigotimes_{\sigma'=0}^{\sigma-1} {}^\tau\mathcal{L}_{r_{\sigma'}}\right) \twoheadrightarrow (\mathcal{F}^{\sigma-1}/\mathcal{F}^\sigma) \otimes \left(\bigotimes_{\sigma'=0}^{\sigma-1} {}^\tau\mathcal{L}_{r_{\sigma'}}\right)$$

$$\xrightarrow{\sim} (\mathcal{E}''_\sigma/\mathcal{E}''_{\sigma-1}) \otimes \left(\bigotimes_{\sigma'=0}^{\sigma-1} \mathcal{L}_{r_{\sigma'}}\right) \xrightarrow{\overline{j}'} (\mathcal{E}'_\sigma/\mathcal{E}'_{\sigma-1}) \otimes \left(\bigotimes_{\sigma'=0}^{\sigma-1} \mathcal{L}_{r_{\sigma'}}\right)$$

où $\overline{j}'$ est induite par j', de l'homomorphisme

$$\overline{j} : \mathcal{G}_\sigma \otimes \left(\bigotimes_{\sigma'=0}^{\sigma-1} \mathcal{L}_{r_{\sigma'}}\right) \to (\mathcal{E}'_\sigma/\mathcal{E}'_{\sigma-1}) \otimes \left(\bigotimes_{\sigma'=0}^{\sigma-1} \mathcal{L}_{r_{\sigma'}}\right)$$

induit par j, et du composé

$$\mathcal{G}'_\sigma \hookrightarrow {}^\tau\mathcal{E}_\sigma \twoheadrightarrow {}^\tau\mathcal{E}_\sigma/{}^\tau\mathcal{E}_{\sigma-1} = {}^\tau\mathcal{G}_\sigma.$$

On a donc des diagrammes

$$\widetilde{\mathcal{G}}_1 = (\mathcal{G}_1 \hookrightarrow \mathcal{G}'_1 \hookleftarrow {}^\tau\mathcal{G}_1)$$

et

$$\widetilde{\mathcal{G}}_\sigma = \left(\begin{array}{l} \qquad \mathcal{G}_\sigma \otimes \left(\bigotimes_{\sigma'=0}^{\sigma-1} \mathcal{L}_{r_{\sigma'}}\right) \\ \qquad\qquad\qquad \downarrow \\ (\mathcal{E}'_\sigma/\mathcal{E}'_{\sigma-1}) \otimes \left(\bigotimes_{\sigma'=0}^{\sigma-1} \mathcal{L}_{r_{\sigma'}}\right) \hookleftarrow \mathcal{G}'_\sigma \otimes \left(\bigotimes_{\sigma'=0}^{\sigma-1} {}^\tau\mathcal{L}_{r_{\sigma'}}\right) \hookrightarrow {}^\tau\mathcal{G}_\sigma \otimes \left(\bigotimes_{\sigma'=0}^{\sigma-1} {}^\tau\mathcal{L}_{r_{\sigma'}}\right) \end{array}\right)$$

pour $\sigma = 2, \ldots, s$.

LEMME. — *Pour tout $\rho = 1, \ldots, r$, notons u°_ρ la restriction de u_ρ à l'ouvert de $U \times X$ complémentaire des graphes des morphismes ∞ et o, et identifions les restrictions de $\mathcal{E}$ et $\mathcal{E}''$ à cet ouvert à l'aide de ${j'}^{-1} \circ j$. Avec ces notations les deux conditions ci-dessous sont équivalentes :*

(a) *il n'existe aucun point géométrique de U tel que la fibre en ce point d'un des u°_ρ soit nilpotente.*

(a′) *pour tout $\sigma = 1, \ldots, s-1$, on a $({}^\tau j)(\mathcal{F}^\sigma) \cap {}^\tau\mathcal{E}'_\sigma = (0)$ dans ${}^\tau\mathcal{E}'$.*

De plus, si ces conditions équivalentes sont vérifiées, le plus grand ouvert de $U \times X$ où toutes les flèches des diagrammes $\widetilde{\mathcal{G}}_\sigma$ ci-dessus sont des isomorphismes rencontre chaque fibre de la projection canonique $U \times X \to U$ suivant un ouvert dense.

La condition (a) a été introduite par Drinfeld en rang $r = 2$.

Suivant Lafforgue considérons les conditions supplémentaires suivantes :

(b) $\mathcal{E}'/\mathcal{E}'_\sigma$ est un $\mathcal{O}_{U\times S}$-Module localement libre pour $\sigma = 0, 1, \ldots, s-1$,

(c) pour tout $\sigma = 1, \ldots, s$, l'homomorphisme composé $\mathcal{E}''_\sigma \xhookrightarrow{j'} \mathcal{E}' \twoheadrightarrow \mathcal{E}'/j(\mathcal{E})$ est surjectif,

(d) pour tout $\sigma = 1, \ldots, s-1$, on a $\mathcal{F}^{\sigma-1} + {}^\tau\mathcal{E}_\sigma = {}^\tau\mathcal{E}$, La condition (c) assure en particulier que

$$\overline{j} : \mathcal{G}_\sigma \otimes \left(\bigotimes_{\sigma'=0}^{\sigma-1} \mathcal{L}_{r_{\sigma'}} \right) \to (\mathcal{E}'_\sigma / \mathcal{E}'_{\sigma-1}) \otimes \left(\bigotimes_{\sigma'=0}^{\sigma-1} \mathcal{L}_{r_{\sigma'}} \right)$$

est un isomorphisme. On peut donc récrire les diagrammes $\widetilde{\mathcal{G}}_\sigma$, $\sigma = 2, \ldots, s$, sous la forme

$$\widetilde{\mathcal{G}}_\sigma = \left(\mathcal{G}_\sigma \otimes \left(\bigotimes_{\sigma'=0}^{\sigma-1} \mathcal{L}_{r_{\sigma'}} \right) \hookleftarrow \mathcal{G}'_\sigma \otimes \left(\bigotimes_{\sigma'=0}^{\sigma-1} {}^\tau\mathcal{L}_{r_{\sigma'}} \right) \hookrightarrow {}^\tau\mathcal{G}_\sigma \otimes \left(\bigotimes_{\sigma'=0}^{\sigma-1} {}^\tau\mathcal{L}_{r_{\sigma'}} \right) \right).$$

LEMME. — (i) *Sous les conditions* (a), (b), (c) *et* (d) *le diagramme* $\widetilde{\mathcal{G}}_1$ *est un chtouca à droite de rang* r_1 *et de pôle* ∞, *les diagrammes* $\widetilde{\mathcal{G}}_2, \ldots, \widetilde{\mathcal{G}}_s$ *sont des chtoucas à gauche, le zéro de* $\widetilde{\mathcal{G}}_\sigma$ *est le pôle de* $\widetilde{\mathcal{G}}_{\sigma+1}$ *pour* $\sigma = 1, \ldots, s-1$ *et le zéro de* $\widetilde{\mathcal{G}}_s$ *est* o.

(ii) *Il existe un unique sous-champ ouvert* $\overline{\mathrm{Cht}}^r \subset \mathcal{C}^r$ *tel que pour chaque* $R = \{r_1, \ldots, r_{s-1}\}$ *comme ci-dessus la trace de cet ouvert sur* $\mathcal{C}^r_R$ *soit définie par les conditions* (a), (b), (c) *et* (d).

Le champ $\overline{\mathrm{Cht}}^r$ est par définition le champ des *chtoucas itérés* de rang r. Il contient comme ouvert dense le champ des chtoucas Cht^r. Tout comme un chtouca ordinaire, un chtouca itéré admet un degré, à savoir le degré du fibré vectoriel sous-jacent $\mathcal{E}$, et on a un découpage de $\overline{\mathrm{Cht}}^r$ en composantes

$$\overline{\mathrm{Cht}}^r = \coprod_{d\in\mathbb{Z}} \overline{\mathrm{Cht}}^{r,d}$$

indexées par ce degré. Bien sûr, pour chaque entier d, $\mathrm{Cht}^{r,d}$ est un ouvert dense de $\overline{\mathrm{Cht}}^{r,d}$.

Pour chaque $R = \{r_1, \ldots, r_{s-1}\} \subset [r-1]$ on notera $\overline{\mathrm{Cht}}^r_R$ l'intersection de $\overline{\mathrm{Cht}}^r$ avec la strate $\mathcal{C}^r_R$. C'est un sous-champ localement fermé de $\overline{\mathrm{Cht}}^r$, qui n'est autre que l'ouvert Cht^r pour $R = \varnothing$, et $\overline{\mathrm{Cht}}^r$ est la réunion disjointes des $\overline{\mathrm{Cht}}^r_R$.

Il résulte du dernier lemme que l'on a un morphisme fini, surjectif et radiciel de champs

$$\overline{\mathrm{Cht}}^r_R \to \mathrm{Cht}^{r_1} \times_X {}^{r_2-r_1}\mathrm{Cht} \cdots \times_X {}^{r_s-r_{s-1}}\mathrm{Cht}\,.$$

De plus on a un morphisme de Frobenius partiel (au-dessus de l'endomorphisme $\mathrm{Id}_X \times \mathrm{Frob}_X$ de $X \times X$)

$$\begin{aligned}&\mathrm{Cht}^{r_1} \times_X {}^{r_2-r_1}\mathrm{Cht} \cdots \times_X {}^{r_s-r_{s-1}}\mathrm{Cht}\\ &\qquad \to \mathrm{Cht}^R := \mathrm{Cht}^{r_1} \times_X \mathrm{Cht}^{r_2-r_1} \times_{X,\mathrm{Frob}_X} \cdots \times_{X,\mathrm{Frob}_X} \mathrm{Cht}^{r_s-r_{s-1}}\end{aligned}$$

qui envoie $(\widetilde{\mathcal{G}}_1, \widetilde{\mathcal{G}}_2, \ldots, \widetilde{\mathcal{G}}_s)$ sur

$$(\widetilde{\mathcal{G}}_1, \mathrm{Frob}_o(\widetilde{\mathcal{G}}_2), \ldots, \mathrm{Frob}_o(\widetilde{\mathcal{G}}_s))$$

qui est lui aussi fini, surjectif et radiciel. Par composition on a donc un morphisme fini, surjectif et radiciel de champs

$$\overline{\mathrm{Cht}}_R^r \to \mathrm{Cht}^R .$$

Le but de ce dernier morphisme est un champ de Deligne-Mumford séparé et lisse purement de dimension relative

$$(2r_1 - 2) + (2(r_2 - r_1) - 2) + \cdots + (2(r_s - r_{s-1}) - 2) = 2r - 2s$$

au-dessus de $X \times X^{s-1} \times X$ par le morphisme $(\infty = \infty_1, \infty_2, \ldots, \infty_s, o_s = o)$. Il se décompose en

$$\mathrm{Cht}^R = \coprod_{d_\bullet \in \mathbb{Z}^s} \mathrm{Cht}^{R,d_\bullet}$$

où

$$\mathrm{Cht}^{R,d_\bullet} := \mathrm{Cht}^{r_1,d_1} \times_X \mathrm{Cht}^{r_2-r_1,d_2} \times_{X,\mathrm{Frob}_X} \cdots \times_{X,\mathrm{Frob}_X} \mathrm{Cht}^{r_s-r_{s-1},d_s} .$$

On notera

$$\overline{\mathrm{Cht}}_R^r = \coprod_{d_\bullet \in \mathbb{Z}^s} \overline{\mathrm{Cht}}_R^{r,d_\bullet}$$

la décomposition correspondante de la source. On vérifie que

$$\overline{\mathrm{Cht}}_R^{r,d} := \overline{\mathrm{Cht}}_R^r \cap \overline{\mathrm{Cht}}^{r,d} = \coprod_{\substack{d_\bullet \in \mathbb{Z}^s \\ d_1 + \cdots d_s = d-s+1}} \overline{\mathrm{Cht}}_R^{r,d_\bullet}$$

(le décalage $-s+1$ provient des Frobenius partiels).

3.3. Chtoucas itérés et troncature

Soient d un entier et $p : [0,r] \to \mathbb{R}$ un paramètre de troncature. Pour chaque $\rho = 0, 1, \ldots, r$, notons $\widetilde{p}(\rho)$ l'unique entier appartenant à l'intervalle de longueur 1

$$]p(\rho) + \frac{\rho}{r}d - 1, p(\rho) + \frac{\rho}{r}d].$$

On définit alors des entiers $d_1, \ldots, d_s$ par

$$d_1 = \widetilde{p}(r_1) \text{ et } d_\sigma = \widetilde{p}(r_\sigma) - \widetilde{p}(r_{\sigma-1}) - 1,\ \forall \sigma = 2, \ldots, s,$$

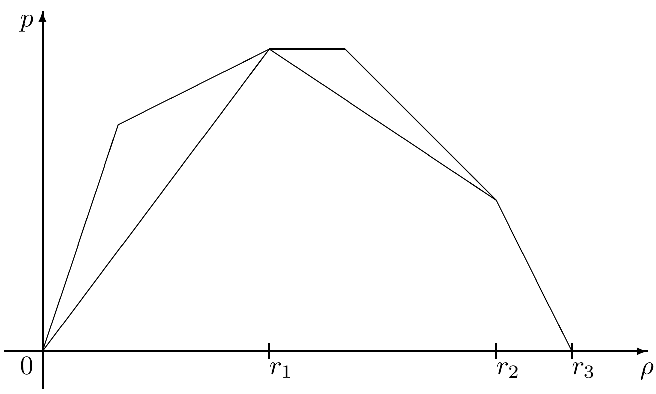
et des paramètres de troncature $p_1 : [0, r_1] \to \mathbb{R}, \ldots, p_s : [0, r_s - r_{s-1}] \to \mathbb{R}$ en imposant que

$$p_1(\rho_1) = \widetilde{p}(\rho_1) - \frac{\rho_1}{r_1}d_1,\ \forall \rho_1 = 1, \ldots, r_1 - 1,$$

et

$$p_\sigma(\rho_\sigma) = \widetilde{p}(r_{\sigma-1} + \rho_\sigma) - \widetilde{p}(r_{\sigma-1}) - 1 - \frac{\rho_\sigma}{r_\sigma - r_{\sigma-1}}d_\sigma,\ \forall \rho_\sigma = 1, \ldots, r_\sigma - r_{\sigma-1}.$$

À décalage et normalisation près, les p_σ sont essentiellement les restrictions de p aux intervalles $[r_{\sigma-1}, r_\sigma]$ comme dans la figure ci-dessous :



On note

$$\mathrm{Cht}^{R,d;\,\leq p} = \mathrm{Cht}^{r_1,d_1;\,\leq p_1} \times_X \mathrm{Cht}^{r_2-r_1,d_2;\,\leq p_2} \times_{X,\mathrm{Frob}_X} \cdots$$
$$\cdots \times_{X,\mathrm{Frob}_X} \mathrm{Cht}^{r_s-r_{s-1},d_s;\,\leq p_2} \subset \mathrm{Cht}^R .$$

Définition. — *Un paramètre de troncature* $p : [0,r] \to \mathbb{R}$ *est dit* μ*-convexe pour un nombre réel* $\mu \geq 0$ *si l'on a*

$$(p(\rho) - p(\rho-1)) - (p(\rho+1) - p(\rho)) \geq \mu, \ \forall 1 \leq \rho \leq r-1.$$

Une propriété est dite vraie « pour tout paramètre de troncature assez convexe » s'il existe un nombre réel $\mu \geq 0$ *tel que la propriété soit vraie pour tout paramètre de troncature qui est* μ*-convexe.*

On remarquera que les paramètres de troncature p_σ définis ci-dessus sont automatiquement $(\mu-2)$-convexes dès que p est μ-convexe.

Théorème. — *Pour tout entier* d *et tout paramètre de troncature* $p : [0,r] \to \mathbb{R}$ *qui est* 2*-convexe, il existe un ouvert*

$$\overline{\mathrm{Cht}}^{r,d;\,\leq p} \subset \overline{\mathrm{Cht}}^{r,d}$$

de la composante d*-ième du champ des chtoucas itérés ayant les propriétés suivantes :*

– *pour tout* $R \subset [r-1]$, $\overline{\mathrm{Cht}}_R^{r,d;\,\leq p} := \overline{\mathrm{Cht}}_R^{r,d} \cap \overline{\mathrm{Cht}}^{r,d;\,\leq p}$ *est l'image réciproque par le morphisme fini, surjectif et radiciel* $\overline{\mathrm{Cht}}_R^{r} \to \mathrm{Cht}^R$ *de l'ouvert* $\mathrm{Cht}^{R,d;\,\leq p}$ *;*

– *le morphisme de champs* $(\infty, o) : \overline{\mathrm{Cht}}^{r,d;\,\leq p} = \overline{\mathrm{Cht}}^{r;\,\leq p} \cap \overline{\mathrm{Cht}}^{r,d} \to X \times X$ *est propre (et donc séparé et de type fini) ;*

– *si l'on suppose de plus que* p *est assez convexe relativement au genre de la courbe* X*, le morphisme de champs* (∞, o) *ci-dessus est lisse purement de dimension relative* $2r-2$ *et le fermé complémentaire dans* $\overline{\mathrm{Cht}}^{r,d;\,\leq p}$ *de la strate ouverte* $\mathrm{Cht}^{r,d;\,\leq p} = \overline{\mathrm{Cht}}_{\varnothing}^{r,d;\,\leq p}$ *est un diviseur à croisements normaux relatif sur* $X \times X$*, qui est réunion de* $r-1$ *diviseurs lisses et dont la stratification canonique est celle par les* $\overline{\mathrm{Cht}}_R^{r,d;\,\leq p}$ *pour* R *parcourant les parties non vides de* $[r-1]$*.*

On posera

$$\overline{\mathrm{Cht}}^{r;\leq p} = \coprod_{d\in\mathbb{Z}} \overline{\mathrm{Cht}}^{r,d;\leq p} \subset \overline{\mathrm{Cht}}^{r}$$

et

$$\overline{\mathrm{Cht}}_R^{r;\leq p} = \coprod_{d\in\mathbb{Z}} \overline{\mathrm{Cht}}_R^{r,d;\leq p} \subset \overline{\mathrm{Cht}}^{r;\leq p}.$$

3.4. Structures de niveau « sans facteur carré » sur les chtoucas itérés

Soit

$$\mathcal{E} \xhookrightarrow{j} \mathcal{E}' \xhookleftarrow{j'} \mathcal{E}'' \xLeftarrow{u} {}^\tau\mathcal{E},$$

un chtouca itéré sur un schéma U ;

$$u = (u_1, \ldots, u_r; (\mathcal{L}_1^{\otimes(q-1)}, \lambda_1^{q-1}), \ldots, (\mathcal{L}_{r-1}^{\otimes(q-1)}, \lambda_{r-1}^{q-1}))$$

est donc un homomorphisme complet de ${}^\tau\mathcal{E}$ dans $\mathcal{E}''$. Si le pôle et le zéro de ce chtouca itéré ne rencontrent pas N, les restrictions

$$j_N : \mathcal{E}_N \to \mathcal{E}'_N \text{ et } j'_N : \mathcal{E}''_N \to \mathcal{E}'_N$$

des flèches j et j' sont des isomorphismes de $\mathcal{O}_{U\times N}$-Modules localement libres de rang r et la restriction

$$u_N = (u_{1,N}, \ldots, u_{r,N}; (\mathcal{L}_1^{\otimes(q-1)}, \lambda_1^{q-1}), \ldots, (\mathcal{L}_{r-1}^{\otimes(q-1)}, \lambda_{r-1}^{q-1})) : {}^\tau\mathcal{E}_N \Longrightarrow \mathcal{E}_N$$

est un homomorphisme complet. Le morphisme de restriction à N

$$\mathrm{Cht}^r \times_{X^2} (X-N)^2 \to \mathrm{Tr}_N^r$$

se prolonge donc en un morphisme

$$\overline{\mathrm{Cht}}^r \times_{X^2} (X-N)^2 \to \overline{\mathrm{Tr}}_N^r$$

où $\overline{\mathrm{Tr}}_N^r$ est le champ qui associe au schéma U la catégorie des uplets

$$(\mathcal{F}, u_1, \ldots, u_r; (\mathcal{L}_1, \lambda_1), \ldots, (\mathcal{L}_{r-1}, \lambda_{r-1}))$$

formés d'un fibré $\mathcal{F}$ de rang r sur $U \times N$, de fibrés en droites $\mathcal{L}_1, \ldots, \mathcal{L}_{r-1}$ sur U munis de sections globales $\lambda_1, \ldots, \lambda_{r-1}$ et d'un homomorphisme complet

$$u = (u_1, \ldots, u_r; (\mathcal{L}_1^{\otimes(q-1)}, \lambda_1^{q-1}), \ldots, (\mathcal{L}_{r-1}^{\otimes(q-1)}, \lambda_{r-1}^{q-1})) : {}^\tau\mathcal{F} \Longrightarrow \mathcal{F}.$$

L'ouvert Tr_N^r de $\overline{\mathrm{Tr}}_N^r$ est l'image réciproque par le morphisme

$$((\mathcal{L}_1, \lambda_1), \ldots, (\mathcal{L}_{r-1}, \lambda_{r-1})) : \overline{\mathrm{Tr}}_N^r \to [\mathbb{A}^{r-1}/\mathbb{G}_{\mathrm{m}}^{r-1}]$$

de la strate ouverte $\mathrm{Spec}(\mathbb{F}_q) = [\mathbb{G}_{\mathrm{m}}^{r-1}/\mathbb{G}_{\mathrm{m}}^{r-1}] \subset [\mathbb{A}^{r-1}/\mathbb{G}_{\mathrm{m}}^{r-1}]$.

THÉORÈME. — *Considérons le morphisme de champs algébriques*

$$\overline{\mathrm{Cht}}^{\,r} \times_{X^2} (X-N)^2 \to \overline{\mathrm{Tr}}_N^{\,r} \times (X-N)^2$$

de composante le morphisme précédant et la seconde projection canonique.

Pour tout entier d et tout paramètre de troncature $p : [0,r] \to \mathbb{R}$ assez convexe par rapport à X et N, la restriction

$$\overline{\mathrm{Cht}}^{\,r,d;\,\leq p} \times_{X^2} (X-N)^2 \to \overline{\mathrm{Tr}}_N^{\,r} \times (X-N)^2$$

de ce morphisme à l'ouvert $\overline{\mathrm{Cht}}^{\,r,d;\,\leq p} \subset \overline{\mathrm{Cht}}^{\,r}$ est lisse purement de dimension relative $2r-2$.

Supposons maintenant que N soit *« sans facteur carré »*, c'est-à-dire *réduit.* Le cas essentiel est alors celui d'un niveau N réduit à un seul point fermé de X que l'on supposera ici rationnel sur le corps fini de base $\mathbb{F}_q$ pour alléger les notations. Dans ce cas, on a simplement $\mathrm{GL}_r^N = \mathrm{GL}_r$ et le champ $\overline{\mathrm{Tr}}_N^{\,r}$ est le produit fibré

$$\overline{\mathrm{Tr}}_N^{\,r} = [\widetilde{\mathrm{gl}}_r \,/ {}^{\tau}\mathrm{Ad}(\mathrm{GL}_r)] \times_{[\mathbb{A}^{r-1}/\mathbb{G}_{\mathrm{m}}^{r-1}],\langle q-1\rangle} [\mathbb{A}^{r-1}/\mathbb{G}_{\mathrm{m}}^{r-1}]$$

où $\widetilde{\mathrm{gl}}_r$ est le schéma des endomorphismes complets du $\mathbb{F}_q$-espace vectoriel $\mathbb{F}_q^r$, où le morphisme $[\widetilde{\mathrm{gl}}_r \,/ {}^{\tau}\mathrm{Ad}(\mathrm{GL}_r)] \to [\mathbb{A}^{r-1}/\mathbb{G}_{\mathrm{m}}^{r-1}]$ est induit par le morphisme canonique $\widetilde{\mathrm{gl}}_r \to [\mathbb{A}^{r-1}/\mathbb{G}_{\mathrm{m}}^{r-1}]$ et où $\langle q-1\rangle$ est l'endomorphisme de $[\mathbb{A}^{r-1}/\mathbb{G}_{\mathrm{m}}^{r-1}]$ induit par l'élévation à la puissance $q-1$ des coordonnées.

Lafforgue définit alors le champ $\overline{\mathrm{Cht}}_N^{\,r,d;\,\leq p}$ de *chtoucas itérés avec structure de niveau N* (sans facteur carré) par le carré cartésien

$$\begin{array}{ccc}
\overline{\mathrm{Cht}}_N^{\,r,d;\,\leq p} & \longrightarrow & \overline{\mathrm{Tr}}_N^{\,r,\tau} \\
\big\downarrow & \square & \big\downarrow \\
\overline{\mathrm{Cht}}^{\,r,d;\,\leq p} \times_{X^2} (X-N)^2 & \longrightarrow & \overline{\mathrm{Tr}}_N^{\,r}
\end{array}$$

où la flèche verticale de droite

$$\begin{aligned}
\overline{\mathrm{Tr}}_N^{\,r,\tau} &:= [\widetilde{\mathrm{gl}}_r^{\,\tau} \,/\, \mathrm{GL}_r] \times_{[\mathbb{A}^{r-1}/\mathbb{G}_{\mathrm{m}}^{r-1}],\langle q-1\rangle} [\mathbb{A}^{r-1}/\mathbb{G}_{\mathrm{m}}^{r-1}] \\
&\qquad \to [\widetilde{\mathrm{gl}}_r \,/ {}^{\tau}\mathrm{Ad}(\mathrm{GL}_r)] \times_{[\mathbb{A}^{r-1}/\mathbb{G}_{\mathrm{m}}^{r-1}],\langle q-1\rangle} [\mathbb{A}^{r-1}/\mathbb{G}_{\mathrm{m}}^{r-1}] = \overline{\mathrm{Tr}}_N^{\,r}
\end{aligned}$$

est induite par un morphisme projectif équivariant $\overline{L} : \widetilde{\mathrm{gl}}_r^{\,\tau} \to \widetilde{\mathrm{gl}}_r$ qui prolonge l'isogénie de Lang $L : \mathrm{GL}_r \to \mathrm{GL}_r$.

Par construction, le schéma $\widetilde{\mathrm{gl}}_r^{\,\tau}$ est muni d'un morphisme lisse vers un « champ torique » (voir la sous-section suivante) $\mathcal{T}^{r,\tau}$, qui relève le morphisme $\widetilde{\mathrm{gl}}_r \to [\mathbb{A}^{r-1}/\mathbb{G}_{\mathrm{m}}^{r-1}]$. On a donc un morphisme lisse de champs

$$\overline{\mathrm{Tr}}_N^{\,r,\tau} \to \mathcal{T}^{r,\tau} \times_{[\mathbb{A}^{r-1}/\mathbb{G}_{\mathrm{m}}^{r-1}],\langle q-1\rangle} [\mathbb{A}^{r-1}/\mathbb{G}_{\mathrm{m}}^{r-1}]$$

et il résulte du théorème ci-dessus que $\overline{\mathrm{Cht}}_N^{\,r,d;\,\leq p}$ est lisse sur le champ

$$\Big(\mathcal{T}^{r,\tau} \times_{[\mathbb{A}^{r-1}/\mathbb{G}_{\mathrm{m}}^{r-1}],\langle q-1\rangle} [\mathbb{A}^{r-1}/\mathbb{G}_{\mathrm{m}}^{r-1}]\Big) \times (X-N)^2.$$

En particulier, comme on sait résoudre les singularités d'un champ torique en raffinant l'éventail qui le définit, on sait aussi résoudre les singularités du champ $\overline{\mathrm{Cht}}_N^{r,d;\,\leq p}$.

La section suivante est consacrée à la construction de $\widetilde{\mathrm{gl}}_r^{\tau}$.

4. COMPACTIFICATION DE L'ISOGÉNIE DE LANG POUR PGL_r

4.1. Champ des pavages

Pour prolonger l'isogénie de Lang pour GL_r au-dessus du schéma des endomorphismes complets, Lafforgue est amené à construire une famille de champs toriques $(\mathcal{T}^{r,n})_{n\geq 0}$ qu'il appelle les *champs des pavages*.

Une *variété torique* est un triplet formé d'un tore T, d'une immersion ouverte $T \hookrightarrow \overline{T}$ d'image dense de T dans un schéma de type fini normal $\overline{T}$, et d'une action de T sur $\overline{T}$ qui prolonge l'action de T sur lui-même par translation. Une telle variété torique est associée à un éventail. Rappelons tout d'abord rapidement cette construction.

Soient Y un $\mathbb{Z}$-module libre de rang fini et $Y_{\mathbb{R}} = \mathbb{R} \otimes_{\mathbb{Z}} Y$. Un *cône convexe polyédral rationnel* est une partie localement fermée σ de Y de la forme

$$\sigma = \mathbb{R}_{>0}\, y_1 + \cdots + \mathbb{R}_{>0}\, y_n \subset Y_{\mathbb{R}}$$

pour des vecteurs $y_1, \ldots, y_n \in Y \subset Y_{\mathbb{R}}$. L'adhérence $\overline{\sigma}$ de σ est le cône convexe fermé

$$\sigma = \mathbb{R}_{\geq 0}\, y_1 + \cdots + \mathbb{R}_{\geq 0}\, y_n \subset Y_{\mathbb{R}}.$$

On définit les faces de σ comme les cônes convexes polyédraux rationnels τ de la forme $\overline{\tau} = H \cap \overline{\sigma}$ pour un hyperplan H de $Y_{\mathbb{R}}$ ne rencontrant pas σ. Un *éventail* est une famille finie E de cônes convexes polyédraux rationnels deux à deux disjoints, appelés les *cellules* de E et ayant les propriétés suivantes :

– toute face d'une cellule de E est encore une cellule de E,

– tout couple (σ_1, σ_2) de cellules de E admet une face commune $\tau \in E$ telle que $\overline{\sigma}_1 \cap \overline{\sigma}_2 = \overline{\tau}$.

– le cône convexe polyédral rationnel $\{0\}$ est une cellule de E.

Si $\sigma = \mathbb{R}_{>0}\, y_1 + \cdots + \mathbb{R}_{>0}\, y_n \subset Y_{\mathbb{R}}$ est un cône convexe polyédral rationnel ne contenant aucune droite vectorielle de $Y_{\mathbb{R}}$, la collection formée de σ et de toutes ses faces est un éventail. On associe à cet éventail particulier la variété torique affine

$$T = Y \otimes_{\mathbb{Z}} \mathbb{G}_{\mathrm{m}} = \mathrm{Spec}(\mathbb{F}_q[Y^{\vee}]) \hookrightarrow \overline{T}(\sigma) = \mathrm{Spec}(\mathbb{F}_q[Y^{\vee} \cap \sigma^{\vee}])$$

où $Y^{\vee} := \mathrm{Hom}_{\mathbb{Z}}(Y, \mathbb{Z})$ et

$$\sigma^{\vee} = \{y^{\vee} : Y_{\mathbb{R}} \to \mathbb{R} \mid y^{\vee}(y_1), \ldots, y^{\vee}(y_n) \geq 0\} \subset Y_{\mathbb{R}}^{\vee}$$

est le cône fermé dual de σ. On vérifie que $Y^{\vee} \cap \sigma^{\vee}$ est un monoïde à engendrement fini, saturé et contenant un système de générateurs du groupe $Y^{\vee}$. L'exemple type est bien entendu le plongement torique affine standard $\mathbb{G}_{\mathrm{m}}^{r-1} \hookrightarrow \mathbb{A}^{r-1}$ qui est obtenu par la construction ci-dessus pour $Y = \mathbb{Z}^{r-1}$ et $\sigma = (\mathbb{R}_{>0})^{r-1} \subset \mathbb{R}^{r-1} = Y_{\mathbb{R}}$.

Si τ est une face de σ, on a $\sigma^\vee \subset \tau^\vee$ et l'immersion ouverte $T \hookrightarrow \overline{T}(\sigma)$ se factorise par l'immersion ouverte $T \hookrightarrow \overline{T}(\tau)$ en

$$T \hookrightarrow \overline{T}(\tau) \hookrightarrow \overline{T}(\sigma).$$

En particulier, si σ_1 et σ_2 sont deux cellules de E et si $\overline{\tau} = \overline{\sigma}_1 \cap \overline{\sigma}_2$, on a une donnée de recollement

$$\overline{T}(\sigma_1) \hookleftarrow \overline{T}(\tau) \hookrightarrow \overline{T}(\sigma_2)$$

pour les variétés toriques affines $T \hookrightarrow \overline{T}(\sigma_1)$ et $T \hookrightarrow \overline{T}(\sigma_2)$.

La variété torique $T \hookrightarrow \overline{T}$ associée à l'éventail E est obtenue en recollant les variétés toriques affines $T \hookrightarrow \overline{T}(\sigma)$ pour $\sigma \in E$ à l'aide des données de recollements ci-dessus.

Les orbites pour l'action de T sur $\overline{T}$ sont en nombre fini et forment une stratification en parties localement fermées de $\overline{T}$. On peut indexer ces orbites par les cellules de E : l'orbite $\overline{T}_\sigma$ est contenue dans la carte affine $\overline{T}(\sigma)$ et est en fait l'unique orbite fermée pour l'action de T sur cette variété torique affine. L'orbite $\overline{T}_\sigma$ est dans l'adhérence de l'orbite $\overline{T}_\tau$ si et seulement si τ est une face de σ. En particulier, $\overline{T}_{\{0\}} = T$ est l'unique orbite ouverte dans $\overline{T}$. Chaque orbite $\overline{T}_\sigma$ a par construction un point marqué $\overline{t}_\sigma$.

Le *champ torique* associé à l'éventail E est le champ algébrique quotient

$$\overline{\mathcal{T}} = [\overline{T}/T].$$

Ce champ algébrique, qui est normal, de type fini et de dimension 0, n'a qu'un nombre fini de points, à savoir les points définis par les $\overline{t}_\sigma \in \overline{T}$. La gerbe résiduelle d'un tel point est le classifiant du fixateur de $\overline{t}_\sigma$ dans T.

Soient maintenant k un corps et n un entier ≥ 0. Considérons l'espace affine $\mathbb{R}^{r,n} = \{x = (x_0, \dots, x_n) \in \mathbb{R}^{n+1} \mid x_0 + x_1 + \cdots + x_n = r\}$ et son réseau $\mathbb{Z}^{r,n} = \{x = (x_0, \dots, x_n) \in \mathbb{Z}^{n+1} \mid x_0 + x_1 + \cdots + x_n = r\}$. On a le simplexe

$$S_{\mathbb{R}}^{r,n} = \{x = (x_0, \dots, x_n) \in (\mathbb{R}_{\geq 0})^{n+1} \mid x_0 + x_1 + \cdots + x_n = r\} \subset \mathbb{R}^{r,n}$$

et l'ensemble de ses points entiers

$$S^{r,n} = \mathbb{Z}^{r,n} \cap S_{\mathbb{R}}^{r,n}.$$

On appelle *pavé entier* toute partie P de $S_{\mathbb{R}}^{r,n}$ d'intérieur P° non vide et de la forme

$$P = \{x \in \mathbb{R}^{r,n} \mid \sum_{j \in J} x_j \geq d_J,\ \forall J\}$$

pour (d_J) une famille d'entiers indexés par les parties J de $\{0, 1, \dots, n\}$ qui est convexe au sens où

$$d_\varnothing = 0,\ d_{\{0,1,\dots,n\}} = r \text{ et } d_{J'} + d_{J''} \leq d_{J' \cup J''} + d_{J' \cap J''},\ \forall J', J''.$$

On appelle *pavage entier* de $S_{\mathbb{R}}^{r,n}$ toute famille $\mathbf{P}$ de pavés entiers telle que

$$S_{\mathbb{R}}^{r,n} = \bigcup_{P\in\mathbf{P}} P$$

et que $P_1^\circ \cap P_2^\circ = \varnothing$ pour tous $P_1 \neq P_2$ dans $\mathbf{P}$.

Une application $S^{r,n} \to \mathbb{R}$ est dite affine si elle est la restriction d'une application affine $\mathbb{R}^{r,n} \to \mathbb{R}$. Soient $Y_{\mathbb{R}}^{r,n}$ le $\mathbb{R}$-espace vectoriel quotient de $\mathbb{R}^{S^{r,n}} = \{S^{r,n} \to \mathbb{R}\}$ par le sous-espace des applications affines, et $Y^{r,n}$ l'image dans ce quotient du réseau $\mathbb{Z}^{S^{r,n}} \subset \mathbb{R}^{S^{r,n}}$.

Pour tout pavage entier $\mathbf{P}$ de $S_{\mathbb{R}}^{r,n}$, notons $\sigma_{\mathbf{P}} \subset Y_{\mathbb{R}}$ l'ensemble des classes y de fonctions $S^{r,n} \to \mathbb{R}$ qui, pour chaque pavé $P \in \mathbf{P}$, admettent un représentant y_P : $S^{r,n} \to \mathbb{R}$ à valeurs positives tel que

$$P = \{x \in S_{\mathbb{R}}^{r,n} \mid y_P(x) = 0\}.$$

Le pavage entier $\mathbf{P}$ est dit *admissible* si σ_P est non vide.

LEMME (Lafforgue). — *Pour chaque pavage entier admissible* $\mathbf{P}$ *de* $S_{\mathbb{R}}^{r,n}$, $\sigma_{\mathbf{P}}$ *est un cône convexe polyédral rationnel dans* $Y_{\mathbb{R}}$. *De plus,* $\sigma_{\mathbf{Q}}$ *est une face de* $\sigma_{\mathbf{P}}$ *si et seulement si le pavage entier admissible* $\mathbf{Q}$ *est plus grossier que* $\mathbf{P}$.

La famille des $\sigma_{\mathbf{P}}$, *pour* $\mathbf{P}$ *parcourant l'ensemble des pavages entiers admissibles de* $S_{\mathbb{R}}^{r,n}$, *est un éventail.*

On appellera *champ des pavages* le k-champ torique

$$\mathcal{T}^{r,n} = [\overline{T}^{r,n}/T^{r,n}]$$

associé à cet éventail. On vérifie que l'on a une suite exacte

$$1 \to \mathbb{G}_{\mathrm{m},k} \to \mathbb{G}_{\mathrm{m},k}^{n+1} \times \mathbb{G}_{\mathrm{m},k} \to \mathbb{G}_{\mathrm{m},k}^{S^{r,n}} \to T^{r,n} \to 1$$

où la deuxième flèche envoie z sur $(z,\ldots,z;z^r)$ et la troisième est donnée par

$$(u_0,u_1,\ldots,u_n;z) \mapsto (u_0^{i_0}u_1^{i_1}\cdots u_n^{i_n}z^{-1})_{i\in S^{r,n}}.$$

Par construction, le k-schéma $\overline{T}^{r,n}$ est normal de type fini et l'action de $T^{r,n}$ sur $\overline{T}^{r,n}$ n'a qu'un nombre fini d'orbites. Ces orbites sont indexées par les pavages entiers admissibles de $S^{r,n}$ et chacune a un point marqué $\overline{t}_{\mathbf{P}}$. Le champ des pavages $\mathcal{T}^{r,n}$ est donc algébrique normal de type fini et n'a qu'un nombre fini de points définis par les points $\overline{t}_{\mathbf{P}}$. Le point défini par $\overline{t}_{\mathbf{P}}$ est dans l'adhérence du point défini par $\overline{t}_{\mathbf{Q}}$ si le pavage entier admissible $\mathbf{P}$ raffine le pavage entier admissible $\mathbf{Q}$. Le point ouvert correspond au pavage le plus grossier, c'est-à-dire au pavage trivial (un seul pavé).

Exemple. — (i) Pour $n=0$, $S_{\mathbb{R}}^{r,n} = S^{r,n} = \{0\}$ et le champ des pavages est simplement $\mathrm{Spec}(k)$. Pour $n=1$ on peut identifier $S_{\mathbb{R}}^{r,n}$ à l'intervalle $[0,r] \subset \mathbb{R}$ et les pavés entiers sont précisément les intervalles $[i, j]$ où $i < j$ sont des entiers compris entre 0 et r. Un pavage entier est donc une décomposition

$$[0, r] = [0, r_1] \cup [r_1, r_1 + r_2] \cup \cdots \cup [r_1 + \cdots + r_{s-1}, r_1 + \cdots + r_{s-1} + r_s]$$

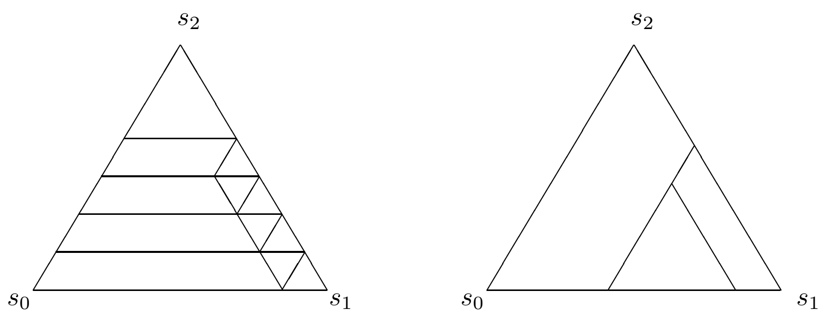
de l'intervalle $[0, r]$ où $r_1 + \cdots + r_s = r$ est une décomposition de r en entiers strictement positifs. Le champ des pavages n'est autre dans ce cas que le champ torique $[\mathbb{A}_k^{r-1}/\mathbb{G}_{\mathrm{m},k}^{r-1}]$.

(ii) Pour $n = 2$ on peut identifier $S_{\mathbb{R}}^{r,n}$ à un triangle équilatéral dans $\mathbb{R}^2$ dont les trois côtés sont de longueur r. On numérote les sommets $s_0 = s_3, s_1, s_2$ de ce triangle et, pour $i = 0, 1, 2$, on désigne par x_i la distance au sommet s_i d'un point du côté $s_i s_{i+1}$. Par tout point de $S_{\mathbb{R}}^{r,n}$ passent trois droites D_{12}, D_{20} et D_{01} parallèles aux côtés du triangle $s_1 s_2$, $s_2 s_0$ et $s_0 s_1$, et on peut repérer un tel point par les coordonnées x_0, x_1 et x_2 des intersections $D_{12} \cap s_0 s_1$, $D_{20} \cap s_1 s_2$ et $D_{01} \cap s_2 s_0$. Par définition, $S^{r,n} \subset S_{\mathbb{R}}^{r,n}$ est formé des points à coordonnées (x_0, x_1, x_2) entières.

Appelons droite entière toute droite qui est parallèle à l'un des trois côtés du triangle $S_{\mathbb{R}}^{r,n}$ et qui passe par un point de $S^{r,n}$; appelons segment entier tout segment d'une droite entière dont les deux extrémités sont des points de $S^{r,n}$.

Si l'on trace toutes les droites entières, on obtient un pavage constitué de petits triangles équilatéraux de côtés de longueur 1 qui est entier admissible. Il est plus fin que tout autre pavage entier : tout pavé entier est une réunion d'une famille de ces petits triangles équilatéraux et est un hexagone, éventuellement dégénéré, dont les côtés sont des segments entiers. La variété torique $T^{r,n} \hookrightarrow \overline{T}^{r,n}$ est donc affine (Cette dernière propriété n'est plus satisfaite dès que $n \geq 3$).

Des deux pavages entiers ci-dessous, seul celui de gauche (qui joue un rôle dans la compactification de l'isogénie de Lang) est admissible.



(iii) Si $r = 2$, le champ torique $\mathcal{T}^{r,n}$ est lisse, mais ce n'est plus le cas dès que $n \geq 2$ et $r \geq 3$.

Pour chaque entier $0 \leq m < n$ et chaque application injective $\{0,1,\ldots,m\} \hookrightarrow \{0,1,\ldots,n\}$, on a une identification induite de $S_{\mathbb{R}}^{r,m}$ à une face de $S_{\mathbb{R}}^{r,n}$ et tout pavage entier admissible de $S_{\mathbb{R}}^{r,n}$ induit par restriction un pavage entier admissible de $S_{\mathbb{R}}^{r,m}$. Dualement, on en déduit des projections $T^{r,n} \to T^{r,m}$, $\overline{T}^{r,n} \to \overline{T}^{r,m}$ et $\mathcal{T}^{r,n} \to \mathcal{T}^{r,m}$.

4.2. Compactifications toroïdales des $\mathrm{PGL}_r^{n+1} / \mathrm{PGL}_r$ pour $n \leq 2$

Considérons un espace vectoriel V de dimension r sur un corps k, que l'on verra aussi comme un k-schéma affine isomorphe à $\mathbb{A}_k^r$, et notons $G = \mathrm{GL}(V)$ le k-schéma des automorphismes linéaires de V. Pour tout entier $n \geq 0$, le k-schéma en groupes G^{n+1} agit de manière évidente sur V^{n+1} et donc sur $\bigwedge^r V^{n+1}$.

On peut décomposer la représentation $\bigwedge^r V^{n+1}$ de G^{n+1} en la somme directe de représentations irréductibles

$$\bigwedge^r V^{n+1} = \bigoplus_{i \in S^{r,n}} \bigwedge^{i_0} V \otimes \bigwedge^{i_1} V \otimes \cdots \otimes \bigwedge^{i_n} V$$

où $S^{r,n}$ est l'ensemble des points entiers du simplexe introduit dans la sous-section précédente. Le tore $\mathbb{G}_{\mathrm{m},k}^{S^{r,n}}$ agit sur $\bigwedge^r V^{n+1}$ par

$$(t = (t_i)_{i \in S^{r,n}}, x = \oplus_{i \in S^{r,n}} x_i) \mapsto \oplus_{i \in S^{r,n}} t_i x_i,$$

et cette action commute à celle de G^{n+1}. On en déduit une action du k-schéma en groupes produit $G^{n+1} \times \mathbb{G}_{\mathrm{m},k}^{S^{r,n}}$ sur l'espace projectif $\mathbb{P}(\bigwedge^r V^{n+1})$ des droites de $\bigwedge^r V^{n+1}$.

On envoie le produit $G \times \mathbb{G}_{\mathrm{m},k}^{n+1} \times \mathbb{G}_{\mathrm{m},k}$ dans $G^{n+1} \times \mathbb{G}_{\mathrm{m},k}^{S^{r,n}}$ par

$$(g; u_0, u_1, \ldots, u_n; z) \mapsto (u_0^{-1} g, u_1^{-1} g, \ldots, u_n^{-1} g; (u_0^{i_0} u_1^{i_1} \cdots u_n^{i_n} z^{-1})_{i \in S^{r,n}}).$$

L'homomorphisme ainsi défini n'est pas injectif (il admet pour noyau $\mathbb{G}_{\mathrm{m},k}$ plongé par $z \mapsto (z\, \mathrm{Id}_V; z, \ldots, z, z^r)$), mais on notera quand même $(G^{n+1} \times \mathbb{G}_{\mathrm{m},k}^{S^{r,n}})/(G \times \mathbb{G}_{\mathrm{m},k}^{n+1} \times \mathbb{G}_{\mathrm{m},k})$ le quotient de $G^{n+1} \times \mathbb{G}_{\mathrm{m},k}^{S^{r,n}}$ par l'image de cet homomorphisme.

Le plongement diagonal $V \hookrightarrow V^{n+1}$ induit un plongement

$$\bigwedge^r V \hookrightarrow \bigwedge^r V^{n+1}$$

et définit un point $\omega \in \mathbb{P}(\bigwedge^r V^{n+1})$ dont le fixateur est précisément l'image de $G \times \mathbb{G}_{\mathrm{m},k}^{n+1} \times \mathbb{G}_{\mathrm{m},k}$ dans $G^{n+1} \times \mathbb{G}_{\mathrm{m},k}^{S^{r,n}}$. On a donc un plongement localement fermé de k-schémas

$$(G^{n+1} \times \mathbb{G}_{\mathrm{m},k}^{S^{r,n}})/(G \times \mathbb{G}_{\mathrm{m},k}^{n+1} \times \mathbb{G}_{\mathrm{m},k}) \hookrightarrow \mathbb{P}\left(\bigwedge^r V^{n+1}\right), \ (g,t) \mapsto (g,t) \cdot \omega.$$

Par construction l'image de ce plongement est contenue dans l'ouvert

$$\left[\prod_{i \in S^{r,n}} \left(\left(\bigwedge^{i_0} V \otimes \bigwedge^{i_1} V \otimes \cdots \otimes \bigwedge^{i_n} V\right) - \{0\}\right)\right] \Big/ \mathbb{G}_{\mathrm{m},k} \subset \mathbb{P}\left(\bigwedge^r V^{n+1}\right).$$

des droites de $\bigwedge^r V^{n+1}$ dont la projection sur chaque facteur direct $\bigwedge^{i_0} V \otimes \bigwedge^{i_1} V \otimes \cdots \otimes \bigwedge^{i_n} V$ n'est pas nulle.

Considérons alors le morphisme

$$(G^{n+1} \times \mathbb{G}_{\mathrm{m},k}^{S^{r,n}})/(G \times \mathbb{G}_{\mathrm{m},k}^{n+1} \times \mathbb{G}_{\mathrm{m},k}) \hookrightarrow \overline{T}^{r,n} \times_k \left[\prod_{i\in S^{r,n}} \left(\left(\bigwedge^{i_0} V \otimes \bigwedge^{i_1} V \otimes \cdots \otimes \bigwedge^{i_n} V\right) - \{0\}\right)\right] \Big/ \mathbb{G}_{\mathrm{m},k}$$

dont les composantes sont le composé de la projection

$$(G^{n+1} \times_k \mathbb{G}_{\mathrm{m},k}^{S^{r,n}})/(G \times \mathbb{G}_{\mathrm{m},k}^{n+1} \times \mathbb{G}_{\mathrm{m},k}) \twoheadrightarrow \mathbb{G}_{\mathrm{m},k}^{S^{r,n}}/(\mathbb{G}_{\mathrm{m},k}^{n+1} \times \mathbb{G}_{\mathrm{m},k}) = T^{r,n}$$

et de l'immersion ouverte

$$T^{r,n} \hookrightarrow \overline{T}^{r,n},$$

et l'immersion localement fermée ci-dessus. Ce morphisme est une immersion localement fermée. Notons

$$\Omega^n(V) \subset \left(\left[\prod_{i\in S^{r,n}} \left(\left(\bigwedge^{i_0} V \otimes \bigwedge^{i_1} V \otimes \cdots \otimes \bigwedge^{i_n} V\right) - \{0\}\right)\right] \Big/ \mathbb{G}_{\mathrm{m},k}\right) \times_k \overline{T}^{r,n}$$

l'adhérence schématique de son image. et

$$\overline{\Omega}^n(V) \to \prod_{i\in S^{r,n}} \mathbb{P}\left(\bigwedge^{i_0} V \otimes \bigwedge^{i_1} V \otimes \cdots \otimes \bigwedge^{i_n} V\right)$$

le quotient de $\Omega^n(V)$ par l'action libre du tore $\mathbb{G}_{\mathrm{m},k}^{S^{r,n}}/\mathbb{G}_{\mathrm{m},k}$ (où $\mathbb{G}_{\mathrm{m},k}$ est plongé diagonalement dans $\mathbb{G}_{\mathrm{m},k}^{S^{r,n}}$).

Théorème. — *La première projection*

$$\Omega^n(V) \to \left[\prod_{i\in S^{r,n}} \left(\left(\bigwedge^{i_0} V \otimes \bigwedge^{i_1} V \otimes \cdots \otimes \bigwedge^{i_n} V\right) - \{0\}\right)\right] \Big/ \mathbb{G}_{\mathrm{m},k}$$

et son quotient par les actions libres du tore $\mathbb{G}_{\mathrm{m},k}^{S^{r,n}}/\mathbb{G}_{\mathrm{m},k}$

$$\overline{\Omega}^n(V) \to \prod_{i\in S^{r,n}} \mathbb{P}\left(\bigwedge^{i_0} V \otimes \bigwedge^{i_1} V \otimes \cdots \otimes \bigwedge^{i_n} V\right)$$

sont projectives. Le k-schéma $\overline{\Omega}^n(V)$ *est donc une compactification projective de* $\overline{G}^{n+1}/\overline{G}$*, où on a noté* $\overline{G} = \mathrm{PGL}(V)$*.*

La seconde projection

$$\Omega^n(V) \to \overline{T}^{r,n}$$

est équivariante relativement au morphisme de tores $\mathbb{G}_{\mathrm{m},k}^{S^{r,n}}/\mathbb{G}_{\mathrm{m},k} \twoheadrightarrow T^{r,n}$ et induit par passage au quotient un morphisme de k-champs algébriques

$$\overline{\Omega}^n(V) = \Omega^n(V)/(\mathbb{G}_{\mathrm{m},k}^{S^{r,n}}/\mathbb{G}_{\mathrm{m},k}) \to [\overline{T}^{r,n}/T^{r,n}] = \mathcal{T}^{r,n}.$$

Plus généralement, pour tout entier $0 \leq m < n$ et toute application injective $\{0,1,\ldots,m\} \hookrightarrow \{0,1,\ldots,n\}$, l'inclusion $S^{r,m} \hookrightarrow S^{r,n}$ et la projection $G^{n+1} \twoheadrightarrow G^{m+1}$ correspondantes induisent un épimorphisme de tores $T^{r,n} \twoheadrightarrow T^{r,m}$, un prolongement torique $\overline{T}^{r,n} \to \overline{T}^{r,m}$ de cet épimorphisme, un morphisme de champs toriques

$$\mathcal{T}^{r,n} = [\overline{T}^{r,n}/T^{r,n}] \to [\overline{T}^{r,m}/T^{r,m}] = \mathcal{T}^{r,m}$$

et des morphismes de k-schémas $\Omega^n(V) \to \Omega^m(V)$ et $\overline{\Omega}^n(V) \to \overline{\Omega}^m(V)$, dits *de face*, au-dessus de $\overline{T}^{r,n} \to \overline{T}^{r,m}$ et $\mathcal{T}^{r,n} \to \mathcal{T}^{r,m}$ respectivement.

En particulier, les $\overline{\Omega}^n(V)$ et les morphismes de face $\overline{\Omega}^n(V) \to \overline{\Omega}^m(V)$ sont sous-jacents à un k-schéma simplicial $\overline{\Omega}^\bullet(V)$ qui contient comme ouvert dense le k-schéma simplicial classifiant $\overline{G}^{\bullet+1}/\overline{G}$.

Théorème. — *Pour $n = 0,1,2$, la seconde projection*

$$\Omega^n(V) \to \overline{T}^{r,n}$$

est lisse, purement de dimension relative nr^2, et le morphisme de champs qu'on en déduit par passage au quotient

$$\overline{\Omega}^n(V) \to \mathcal{T}^{r,n}$$

est lisse, purement de dimension relative $n(r^2-1)$.

Plus généralement, pour tous entiers $0 \leq m < n \leq 2$ et toute application injective $\{0,1,\ldots,m\} \hookrightarrow \{0,1,\ldots,n\}$, les morphismes « de face »

$$\Omega^n(V) \to \Omega^m(V) \times_{\overline{T}^{r,m}} \overline{T}^{r,n} \quad \textit{et} \quad \overline{\Omega}^n(V) \to \overline{\Omega}^m(V) \times_{\mathcal{T}^{r,m}} \mathcal{T}^{r,n}$$

sont lisses, purement de dimensions relatives $(n-m)r^2$ et $(n-m)(r^2-1)$ respectivement.

Remarques. — (o) Dans [La 8], Lafforgue a cru avoir démontré ce dernier théorème quel que soit l'entier $n \geq 0$, mais la preuve de l'assertion (i) du lemme 3 de loc. cit. contient un erreur. En fait la projection $\Omega^n(V) \to \overline{T}^{r,n}$ n'est pas lisse en général (le plus « petit » contre-exemple intervient pour $n = 3$ et $r = 4$).

(i) Pour $n = 0$, $\Omega^n(V)$ n'est autre que $\mathrm{Spec}(k)$. Pour $n = 1$, le carré commutatif

$$\begin{tikzcd}
(\mathrm{GL}(V)^2 \times_k \mathbb{G}_{\mathrm{m},k}^{r+1})/(\mathrm{GL}(V) \times_k \mathbb{G}_{\mathrm{m},k}^2 \times_k \mathbb{G}_{\mathrm{m},k}) \arrow[r, "\sim"] \arrow[d] \arrow[dr, phantom, "+"] & \mathrm{GL}(V) \times_k \mathbb{G}_{\mathrm{m},k}^{r-1} \arrow[d] \\
\mathbb{G}_{\mathrm{m},k}^{r+1}/(\mathbb{G}_{\mathrm{m},k}^2 \times_k \mathbb{G}_{\mathrm{m},k}) \arrow[r, "\sim"] & \mathbb{G}_{\mathrm{m},k}^{r-1}
\end{tikzcd}$$

où les deux flèches horizontales

$$(g_0, g_1, (\mu_\rho)_{\rho=0,\ldots,r}) \mapsto \left(\frac{\mu_1}{\mu_0} g_0 g_1^{-1}, \frac{\mu_0\mu_2}{\mu_1^2}, \ldots, \frac{\mu_{r-2}\mu_r}{\mu_{r-1}^2}\right)$$

et

$$(\mu_\rho)_{\rho=0,1,\ldots,r} \mapsto \left(\frac{\mu_0\mu_2}{\mu_1^2}, \ldots, \frac{\mu_{r-2}\mu_r}{\mu_{r-1}^2}\right)$$

sont des isomorphismes, se prolonge en un carré commutatif

$$\begin{array}{ccc} \Omega^1(V) & \xrightarrow{\sim} & \Omega(V,V) \\ \downarrow & + & \downarrow \\ \overline{T}^{r,1} & \xrightarrow{\sim} & \mathbb{A}_k^{r-1} \end{array}$$

où, de nouveau, les deux flèches horizontales sont des isomorphismes. (La première flèche verticale est celle du théorème ci-dessus et la seconde $\Omega(V,V) \to \mathbb{A}_k^{r-1}$ a été construite dans la section 3.) En particulier, on peut identifier le k-schéma $\widetilde{\mathrm{End}}(V)$ des endomorphismes complets de V à un quotient de $\Omega^1(V)$ par un tore $\mathbb{G}_{\mathrm{m},k}^r$ et, en divisant en plus par les homothéties, on en déduit un isomorphisme

$$\overline{\Omega}^1(V) \xrightarrow{\sim} \widetilde{\mathrm{End}}(V)/\mathbb{G}_{\mathrm{m},k}$$

qui échange les projections

$$\overline{\Omega}^1(V) \to \mathcal{T}^{r,n} \text{ et } \widetilde{\mathrm{End}}(V)/\mathbb{G}_{\mathrm{m},k} \to [\mathbb{A}_k^{r-1}/\mathbb{G}_{\mathrm{m},k}^{r-1}].$$

Le k-schéma projectif $\overline{\Omega}^1(V)$ est donc la compactification de De Concini et Procesi de $\mathrm{PGL}(V)$.

(ii) Le morphisme $\overline{\Omega}^n(V) \to (\overline{\Omega}^1(V))^{n+1}$ produit des morphismes induits par les applications injectives $\{0,1\} \hookrightarrow \{0,1,\ldots,n\}$ données par $\{0,1\} \mapsto \{0,1\}, \{0,1\} \mapsto \{1,2\}, \ldots, \{0,1\} \mapsto \{n,0\}$, envoie $\overline{G}^{n+1}/\overline{G}$ sur le fermé d'équation $g_0g_1\cdots g_n = 1$ dans $(\overline{G}^2/\overline{G})^{n+1} \cong \overline{G}^{n+1}$. Ainsi, $\overline{\Omega}^n(V)$ réalise une « compactification » du morphisme de multiplication de n éléments dans $\overline{G}$.

Rappelons que le k-schéma $\overline{T}^{r,n}$ est stratifié par les orbites de $T^{r,n}$ et que ces orbites sont indexées par les pavages entiers admissibles $\mathbf{P}$ de $S_{\mathbb{R}}^{r,n}$ et possèdent chacune un point marqué $\overline{t}_{\mathbf{P}} \in \overline{T}^{r,n}$. L'image réciproque par la projection $\Omega^n(V) \to \overline{T}^{r,n}$ de cette stratification est évidemment une stratification de $\Omega^n(V)$ par des parties localement fermées $\Omega_{\mathbf{P}}^n(V)$, indexées elles aussi par les pavages entiers admissibles $\mathbf{P}$ de $S_{\mathbb{R}}^{r,n}$. L'inclusion dans la strate $\Omega_{\mathbf{P}}^n(V)$ de la fibre $\Omega^n(V)_{\overline{t}_{\mathbf{P}}}$ de $\Omega^n(V) \to \overline{T}^{r,n}$ en $\overline{t}_{\mathbf{P}}$ induit un isomorphisme de schémas

$$(\Omega^n(V)_{\overline{t}_{\mathbf{P}}} \times_k \mathbb{G}_{\mathrm{m},k}^{S^{r,n}})/(\mathbb{G}_{\mathrm{m},k}^{S^{r,n}})_{\overline{t}_{\mathbf{P}}} \xrightarrow{\sim} \Omega_{\mathbf{P}}^n(V)$$

où $(\mathbb{G}_{\mathrm{m},k}^{S^{r,n}})_{\overline{t}_{\mathbf{P}}}$ est le stabilisateur de $\overline{t}_{\mathbf{P}}$ dans $\mathbb{G}_{\mathrm{m},k}^{S^{r,n}}$.

En particulier, si $\mathbf{P}$ est le pavage trivial, $\Omega^n(V)_{\overline{t}_{\mathbf{P}}} = G^{n+1}/G$ et $\Omega_{\mathbf{P}}^n$ est l'ouvert $(G^{n+1} \times \mathbb{G}_{\mathrm{m},k}^{S^{r,n}})/(G \times \mathbb{G}_{\mathrm{m},k}^{n+1} \times \mathbb{G}_{\mathrm{m},k})$ de $\Omega^n(V)$.

Lafforgue appelle *k-schéma des graphes recollés dans V^{n+1}* associé à un pavage entier admissible $\mathbf{P}$ de $S_{\mathbb{R}}^{r,n}$, et note $\mathrm{Gr}_{\mathbf{P}}^n(V)$, le schéma de modules des familles $(W_P)_{P\in\mathbf{P}}$ de sous-espaces de dimension r dans V^{n+1} indexés par les pavés P de $\mathbf{P}$ telles que :

– pour tout pavé $P \in \mathbf{P}$ et tout $J \subset \{0, 1, \ldots, n\}$, on a

$$\dim(W_P \cap V^J) = \min_{i \in P \cap S^{r,n}} \left\{ \sum_{j \in J} i_j \right\},$$

– pour tous pavés $P', P'' \in \mathbf{P}$ ayant en commun un bord d'équation $\sum_{j \in J} x_j = d_J$ pour un certain $J \subset \{0, 1, \ldots, n\}$ et

$$d_J = \min_{i' \in P' \cap S^{r,n}} \left\{ \sum_{j \in J} i'_j \right\} = \max_{i'' \in P'' \cap S^{r,n}} \left\{ \sum_{j \in J} i''_j \right\},$$

on a

$$i_J^{-1}(W_{P'}) = \mathrm{pr}_J(W_{P''}) \text{ et } \mathrm{pr}_{J^c}(W_{P'}) = i_{J^c}^{-1}(W_{P''})$$

– où J^c est le complémentaire de J dans $\{0, 1, \ldots, n\}$ et $i_J : V^J \hookrightarrow V^{n+1}$, $i_{J^c} : V^{J^c} \hookrightarrow V^{n+1}$ et $\mathrm{pr}_J : V^{n+1} \twoheadrightarrow V^J$, $\mathrm{pr}_{J^c} : V^{n+1} \twoheadrightarrow V^{J^c}$ sont les inclusions et projections canoniques.

Théorème. — *Pour $n = 0, 1, 2$ et pour chaque pavage entier admissible $\mathbf{P}$ de $S_{\mathbb{R}}^{r,n}$, les coordonnées de Plücker induisent un isomorphisme du k-schéma des graphes recollés $\mathrm{Gr}_{\mathbf{P}}^n(V)$ sur la fibre $\Omega^n(V)_{\bar{t}_{\mathbf{P}}}$.*

4.3. Compactification de l'isogénie de Lang

Le corps de base est de nouveau $\mathbb{F}_q$. On notera encore simplement τ l'endomorphisme de Frobenius d'un schéma.

Rappelons que l'on a défini un éventail $(\sigma_{\mathbf{P}})_{\mathbf{P}}$ dans l'espace vectoriel réel $Y_{\mathbb{R}}^{r,2}$ quotient de $\mathbb{R}^{S^{r,2}} = \{S^{r,2} \to \mathbb{R}\}$ par le sous-espace des applications affines et que l'on a noté $T^{r,2} \hookrightarrow \overline{T}^{r,2}$ la variété torique correspondante.

Considérons maintenant le sous-espace de $\mathbb{R}^{S^{r,2}}$ formé des applications

$$y : S^{r,2} = \{(i_0, i_1, i_2) \in \mathbb{N}^3 \mid i_0 + i_1 + i_2 = r\} \to \mathbb{R}$$

telles que

$$y(0, i_1, i_2) = qy(i_1, 0, i_2), \ \forall (0, i_1, i_2) \in S^{r,2},$$

et son image $Y_{\mathbb{R}}^{r,\tau}$ dans $Y_{\mathbb{R}}^{r,2}$. On remarquera que la donnée d'une telle application y équivaut à la donnée de sa restriction à

$$S^{r,\tau} = \{(i_0, i_1, i_2) \in \mathbb{N}^3 \mid i_0 + i_1 + i_2 = r \text{ et } i_0 \neq 0\} \subset S^{r,2}$$

Appelons *pavage entier q-admissible* tout pavage $\mathbf{P}$ de $S_{\mathbb{R}}^{r,2}$ tel que le cône convexe polyédral rationnel $\sigma_{\mathbf{P}} \subset Y_{\mathbb{R}}^{r,2}$ rencontre le sous-espace $Y_{\mathbb{R}}^{r,\tau}$. La famille des intersections $\sigma_{\mathbf{P}}^{\tau} := \sigma_{\mathbf{P}} \cap Y_{\mathbb{R}}^{r,\tau}$ pour $\mathbf{P}$ parcourant l'ensemble des pavages entiers q-admissibles est un éventail dans $Y_{\mathbb{R}}^{r,\tau}$ et définit donc une variété torique normale

$$T^{r,\tau} \hookrightarrow \overline{T}^{r,\tau}.$$

De la même façon que l'on a une suite exacte

$$1 \to (\mathbb{G}_{\mathrm{m}}^3 \times \mathbb{G}_{\mathrm{m}})/\mathbb{G}_{\mathrm{m}} \to \mathbb{G}_{\mathrm{m}}^{S^{r,2}} \to T^{r,2} \to 1$$

on a une suite exacte

$$1 \to \mathbb{G}_{\mathrm{m}} \to \mathbb{G}_{\mathrm{m}}^{S^{r,\tau}} \to T^{r,\tau} \to 1$$

où la deuxième flèche envoie u sur $(u^{i_0+qi_1})_{i\in S^{r,\tau}}$. On vérifie que le plongement

$$\mathbb{G}_{\mathrm{m}}^{S^{r,\tau}} \hookrightarrow \mathbb{G}_{\mathrm{m}}^{S^{r,2}}$$

qui envoie $(t_i)_{i\in S^{r,\tau}}$ sur $(t_i)_{i\in S^{r,2}}$, où $t_{(0,i_1,i_2)} = t^q_{(i_1,0,i_2)}$ si $i_1 \neq 0$ et $t_{(0,0,r)} = 1$, induit un plongement $T^{r,\tau} \hookrightarrow T^{r,2}$ qui se prolonge en une immersion fermée

$$\overline{T}^{r,\tau} \hookrightarrow \overline{T}^{r,2}.$$

On identifiera dans la suite $\overline{T}^{r,\tau}$ à l'image de cette immersion fermée.

On a vu que les trois applications strictement croissantes $\{0,1\} \hookrightarrow \{0,1,2\}$, $(0,1) \mapsto (1,2),(0,2),(0,1)$, induisent trois morphismes toriques $p_0,p_1,p_2 : \overline{T}^{r,2} \to \overline{T}^{r,1}$ et trois morphismes de face $\pi_0,\pi_1,\pi_2 : \Omega^{r,2} \to \Omega^{r,1}$ au-dessus de ceux-ci. Par construction, la relation

$$p_0 = \tau \circ p_1$$

est vérifiée sur le fermé $\overline{T}^{r,\tau}$ de $\overline{T}^{r,2}$.

Soit alors

$$\Omega^{r,\tau} \subset \Omega^{r,2} \times_{\overline{T}^{r,2}} \overline{T}^{r,\tau}$$

le sous-schéma fermé défini par l'équation $\pi_0 = \tau \circ \pi_1$ dans $\Omega^{r,1}$. L'action évidente du sous-tore $\mathbb{G}_{\mathrm{m}}^{S^{r,\tau}} \subset \mathbb{G}^{S^{r,2}}$ sur $\Omega^{r,2} \times_{\overline{T}^{r,2}} \overline{T}^{r,\tau}$ stabilise ce fermé. On notera π_1^τ, π_2^τ les restrictions des morphismes π_1,π_2 à $\Omega^{r,\tau}$. Elles relèvent les restrictions $p_1^\tau, p_2^\tau : \overline{T}^{r,\tau} \to \overline{T}^{r,1}$ de p_1,p_2 et sont $\mathbb{G}_{\mathrm{m}}^{S^{r,\tau}}$-équivariantes.

Théorème. — *La seconde projection $\Omega^{r,\tau} \to \overline{T}^{r,\tau}$ est $\mathbb{G}_{\mathrm{m}}^{S^{r,\tau}}$-équivariante et lisse purement de dimension relative r^2.*

La stratification par les $T^{r,\tau}$-orbites de $\overline{T}^{r,\tau}$ induit par image réciproque une stratification de $\Omega^{r,\tau}$ dont les strates $\Omega^{r,\tau}_{\mathbf{P}}$ sont indexées par les pavages entiers q-admissibles. Un tel pavage $\mathbf{P}$ induit des pavages entiers convexes admissibles $\mathbf{Q}_0 = \mathbf{Q}_1$ et $\mathbf{Q}_2$ sur les côtés $\{x_0 = 0\}$, $\{x_1 = 0\}$ et $\{x_2 = 0\}$ de $S^{r,2}_{\mathbb{R}}$ identifiés à $S^{r,1}_{\mathbb{R}}$ et on a des morphismes de face $p_i : \Omega^{r,\tau}_{\mathbf{P}} \to \Omega^{r,1}_{\mathbf{Q}_i}$ pour $i = 0,1,2$. Ces strates et ces morphismes de face admettent des descriptions en termes des schémas de graphes recollés $\mathrm{Gr}^{r,2}_{\mathbf{P}}$ et $\mathrm{Gr}^{r,1}_{\mathbf{Q}}$.

En particulier, pour le pavage trivial, on vérifie que la strate ouverte $\Omega^{r,\tau}_{\mathbf{P}}$ est isomorphe au fermé

$$(\{(g_0,g_1,g_2) \mid g_1g_2^{-1} = \tau(g_0g_2^{-1})\} \times \mathbb{G}_{\mathrm{m}}^{S^{r,\tau}})/(\mathrm{GL}_r \times \mathbb{G}_{\mathrm{m}}) \subset (\mathrm{GL}_r^3 \times \mathbb{G}_{\mathrm{m}}^{S^{r,\tau}})/(\mathrm{GL}_r \times \mathbb{G}_{\mathrm{m}})$$

où le quotient est pris pour le plongement $(g,u) \mapsto ((u^{-1}g, u^{-q}g, g), (u^{i_0+qi_1})_{i\in S^{r,\tau}})$. On peut identifier ce fermé au schéma

$$\mathrm{GL}_r \times (\mathbb{G}_{\mathrm{m}}^{S^{r,\tau}}/\mathbb{G}_{\mathrm{m}})$$

par le morphisme $((g_0,g_1,g_2),t) \mapsto (g_2g_0^{-1}, t)$. Si l'on identifie $\Omega^{r,1}$ à $\mathrm{GL}_r \times (\mathbb{G}_{\mathrm{m}}^{S^{r,1}}/\mathbb{G}_{\mathrm{m}})$ par le morphisme $((g_0,g_1),t) \mapsto (g_1g_0^{-1},t)$, les projections $p_1,p_2 : \Omega_{\mathbf{P}}^{r,\tau} \to \Omega_{\mathbf{Q}}^{r,1}$, où $\mathbf{Q} = \mathbf{Q}_0 = \mathbf{Q}_1 = \mathbf{Q}_2$ est le pavage trivial de $S_{\mathbb{R}}^{r,1}$, sont induites par les morphismes $\mathrm{GL}_r \to \mathrm{GL}_r$ qui envoient g sur g et sur $\tau(g)^{-1}g$ respectivement et les morphismes $\mathbb{G}_{\mathrm{m}}^{S^{r,\tau}} \to \mathbb{G}_{\mathrm{m}}^{S^{r,1}}$ qui envoient t sur $(t_{(i_0,0,i_1)})_{i\in S^{r,1}}$ (avec $t_{(0,0,r)}=1$) et sur $(t_{(i_0,i_1,0)})_{i\in S^{r,1}}$ (avec $t_{(0,r,0)} = t_{r,0,0}^q$) respectivement.

Soit $S^{r,\tau,\circ} = S^{r,\tau} - \{(1,0,r-1)\}$. L'inclusion $\mathbb{G}_{\mathrm{m}}^{S^{r,\tau,\circ}} \subset \mathbb{G}_{\mathrm{m}}^{S^{r,\tau}}$ est une section de l'épimorphisme $\mathbb{G}_{\mathrm{m}}^{S^{r,\tau}} \twoheadrightarrow T^{r,\tau}$.

Théorème. — *Le quotient* $\widetilde{\mathrm{gl}}_r^{\tau}$ *de* $\Omega^{r,\tau}$ *par l'action libre du tore* $\mathbb{G}_{\mathrm{m}}^{S^{r,\tau,\circ}}$ *contient* GL_r *comme un ouvert dense. Il est lisse sur le champ torique* $\mathcal{T}^{r,\tau} = [\overline{T}^{r,\tau}/T^{r,\tau}]$ *et il est muni de deux morphismes projectifs sur le schéma des homomorphismes complets* $\widetilde{\mathrm{gl}}_r$ *qui s'insèrent dans les deux carrés commutatifs*

$$\begin{array}{ccc} \mathrm{GL}_r & \hookrightarrow & \widetilde{\mathrm{gl}}_r^{\tau} \\ {\scriptstyle \mathrm{Id}}\downarrow & + & \downarrow \\ \mathrm{GL}_r & \hookrightarrow & \widetilde{\mathrm{gl}}_r \end{array} \quad \mathit{et} \quad \begin{array}{ccc} \mathrm{GL}_r & \hookrightarrow & \widetilde{\mathrm{gl}}_r^{\tau} \\ {\scriptstyle L}\downarrow & + & \downarrow \\ \mathrm{GL}_r & \hookrightarrow & \widetilde{\mathrm{gl}}_r \end{array}$$

où $L : g \to \tau(g)^{-1}g$ *est l'isogénie de Lang.*

Remarque. — En passant au quotient par l'action libre du tore $\mathbb{G}_{\mathrm{m}}^{r,\tau}$ tout entier, on obtient un schéma *projectif* $\overline{\Omega}^{r,\tau}$, muni d'un morphisme vers la compactification de De Concini et Procesi de PGL_r qui prolonge l'isogénie de Lang $g \mapsto \tau(g)^{-1}g$ de ce schéma en groupes.

5. ALLURE DES NOMBRES DE LEFSCHETZ

On fixe dans toute cette section un niveau arbitraire $N = \mathrm{Spec}(\mathcal{O}_N) \hookrightarrow X$.

5.1. Correspondances de Hecke et endomorphismes de Frobenius partiels

Tout élément $g \in \mathrm{GL}_r(\mathbb{A})$ définit une correspondance dite *de Hecke*

$$c = (c_1,c_2) : \mathrm{Cht}_N^r(g) \to \mathrm{Cht}_N^r \times_{(X-N)^2} \mathrm{Cht}_N^r$$

où $\mathrm{Cht}_N^r(g)$ est un champ de Deligne-Mumford et c_1, c_2 sont des morphismes représentables étales. Cette correspondance ne dépend que de la double classe K_NgK_N où K_N est le sous-groupe d'indice fini $\mathrm{Ker}(\mathrm{GL}_r(\mathcal{O}) \twoheadrightarrow \mathrm{GL}_r(\mathcal{O}_N))$ de $K = \mathrm{GL}_r(\mathcal{O})$. Si S_g est la réunion du support $|N| \subset |X|$ de N et de l'ensemble fini des places x de X telles que $g_x \notin F_x^\times K_{N,x} \subset \mathrm{GL}_r(F_x)$, c_1 et c_2 sont finies au-dessus de l'ouvert $(X-S_g)^2$ de $(X-N)^2$.

Si $a \in \mathbb{A}^\times$, la double classe $F^\times a\,\mathrm{Ker}(\mathcal{O}^\times \twoheadrightarrow \mathcal{O}_N^\times)$ dans $F^\times \backslash \mathbb{A}^\times / \mathrm{Ker}(\mathcal{O}^\times \twoheadrightarrow \mathcal{O}_N^\times)$ est la classe d'isomorphie d'un $\mathcal{O}_X$-Module inversible $\mathcal{L}$ muni d'une trivialisation de sa restriction à N. Si $g \in aK_N \subset \mathrm{GL}_r(\mathbb{A})$, on a $\mathrm{Cht}_N^r(g) = \mathrm{Cht}_N^r$, c_1 est l'identité et c_2 est l'automorphisme de Cht_N^r qui envoie un chtouca $\widetilde{\mathcal{E}}$ muni d'une structure de niveau N sur le produit tensoriel $\mathcal{L} \otimes \widetilde{\mathcal{E}}$ muni de la structure de niveau N évidente. On notera simplement a cet automorphisme.

Soit Λ_X le schéma intersection de tous les ouverts de $X \times X$ complémentaires des images réciproques de la diagonale par les endomorphismes $\mathrm{Frob}_X^n \times \mathrm{Id}_X$ et $\mathrm{Id}_X \times \mathrm{Frob}_X^n$ de $X \times X$ pour tous les entiers $n \geq 0$. Au-dessus de Λ_X, il n'y a plus de différence entre chtoucas à gauche et chtoucas à droite : un chtouca à droite $\mathcal{E} \xhookrightarrow{j} \mathcal{E}' \xhookleftarrow{t} {}^\tau\mathcal{E}$ dont le couple pôle-zéro (∞, o) se factorise par $\Lambda_X \subset X \times X$ définit un chtouca à gauche $\mathcal{E} \xhookleftarrow{t'} \mathcal{E}'' \xhookrightarrow{j'} {}^\tau\mathcal{E}$ de pôle $\infty' = \infty$ et de zéro $o' = o$, où $\mathcal{E}'' = \mathcal{E} \times_{j,\mathcal{E}',t} {}^\tau\mathcal{E}$ et t', j' sont les deux projections, et vice versa. Au-dessus de Λ_X, les morphismes de Frobenius partiels introduits en 2.1 peuvent donc être vus comme des endomorphismes, dits encore de Frobenius partiels,

$$\mathrm{Frob}_\infty \ \text{ et } \ \mathrm{Frob}_o : \Lambda_X \times_{X^2} \mathrm{Cht}_N^r \to \Lambda_X \times_{X^2} \mathrm{Cht}_N^r$$

au-dessus de $\mathrm{Frob}_X \times \mathrm{Id}_X$ et $\mathrm{Id}_X \times \mathrm{Frob}_X$ respectivement.

On fixe dans la suite un élément $a \in \mathbb{A}^\times$ de degré $\deg(a) > 0$ et on considère le quotient

$$\mathrm{Cht}_N^r / a^{\mathbb{Z}} \cong \coprod_{d=1}^{r\deg(a)} \mathrm{Cht}_N^{r,d}\,.$$

Les correspondances de Hecke et les endomorphismes de Frobenius partiels que nous venons d'introduire passent au quotient par $a^{\mathbb{Z}}$.

Pour tout paramètre de troncature $p : [0,r] \to \mathbb{R}$ et tout $\alpha \in [0,1]$, $a^{\mathbb{Z}}$ stabilise l'ouvert $\mathrm{Cht}_N^{r;\leq\alpha p}$ de Cht_N^r et l'ouvert quotient

$$\mathrm{Cht}_N^{r;\leq\alpha p} / a^{\mathbb{Z}} \subset \mathrm{Cht}_N^r / a^{\mathbb{Z}}$$

est de type fini. *Par contre, cet ouvert n'est stabilisé ni par les endomorphismes de Frobenius partiels* Frob_∞, Frob_o, *ni par les correspondances de Hecke* (*sauf celles qui sont associées à des* $g \in \mathbb{A}^\times K$). C'est la difficulté majeure qu'a dû surmonter Lafforgue.

5.2. Nombres de Lefschetz

Si ∞ et o sont deux points fermés de X le sous-schéma fini $\infty \times o \subset X \times X$ a exactement $\delta(\infty, o)$ points fermés, où $\delta(\infty, o)$ est le plus grand commun diviseur de $\deg(\infty)$ et $\deg(o)$. Le corps résiduel $\kappa(\xi)$ de chaque $\xi \in \infty \times o$ est une extension composée de $\kappa(\infty)$ et $\kappa(o)$, et son degré $\deg(\xi)$ sur $\mathbb{F}_q$ est le plus petit commun multiple

$$\mu(\infty, o) = \frac{\deg(\infty)\deg(o)}{\delta(\infty, o)}$$

de $\deg(\infty)$ et $\deg(o)$. Pour chaque $\xi \in \infty \times o$, les suites de points fermés de $X \times X$,

$$n \mapsto (\mathrm{Frob}_X^n \times \mathrm{Id}_X)(\xi) \text{ et } n \mapsto (\mathrm{Id}_X \times \mathrm{Frob}_X^n)(\xi)$$

sont en fait à valeurs dans $\infty \times o$ et sont périodiques de périodes $\deg(\infty)$ et $\deg(o)$ respectivement.

Fixons $g \in \mathrm{GL}_r(\mathbb{A})$, deux points fermés distincts ∞ et o de $X - S_g$, un entier t, deux entiers $n_\infty, n_o \geq 1$ tels que $\deg(o)n_o - \deg(\infty)n_\infty = t$, un point fermé ξ de $\infty \times o$, un paramètre de troncature $p : [0, r] \to \mathbb{R}$ et un nombre réel $\alpha \in [0, 1]$.

On note $\mathrm{Cht}_{N,\xi}^r/a^{\mathbb{Z}}$ la fibre en ξ de la projection canonique $\mathrm{Cht}_N^r/a^{\mathbb{Z}} \to (X-N)^2$. C'est un champ algébrique de Deligne-Mumford lisse purement de dimension relative $2r-2$ sur le corps fini $\kappa(\xi)$. On note

$$c_\xi = (c_{1,\xi}, c_{2,\xi}) : \mathrm{Cht}_{N,\xi}^r(g)/a^{\mathbb{Z}} \to \mathrm{Cht}_{N,\xi}^r/a^{\mathbb{Z}} \times_{\kappa(\xi)} \mathrm{Cht}_{N,\xi}^r/a^{\mathbb{Z}}$$

la fibre en ξ de la correspondance de Hecke définie par g. Les deux composantes $c_{1,\xi}$ et $c_{2,\xi}$ sont représentables finies étales.

Puisque ∞ et o sont distincts, le point fermé ξ est dans $\Lambda_X \subset X^2$. Les puissances $\mathrm{Frob}_\infty^{\deg(\infty)}$ et $\mathrm{Frob}_o^{\deg(o)}$ des endomorphismes de Frobenius partiels sont donc définies sur la fibre $\mathrm{Cht}_{N,\xi}^r/a^{\mathbb{Z}}$ et elles la stabilisent.

DÉFINITION. — *Le nombre de Lefschetz*

$$\mathrm{Lef}_\xi(g \times \mathrm{Frob}_\infty^{\deg(\infty)n_\infty} \times \mathrm{Frob}_o^{\deg(o)n_o}, \mathrm{Cht}_N^{r;\leq\alpha p}/a^{\mathbb{Z}})$$

est la somme

$$\sum_y \frac{1}{|\mathrm{Aut}(y)|}$$

où y parcourt l'ensemble des classes d'isomorphie de points dans $\mathrm{Cht}_{N,\xi}^r(g)/a^{\mathbb{Z}}$ munis d'un isomorphisme

$$\begin{aligned} c_{1,\xi}(y) &\cong (\mathrm{Frob}_\infty^{\deg(\infty)n_\infty} \circ \mathrm{Frob}_o^{\deg(o)n_o})(c_{2,\xi}(y)) \\ &\cong (\mathrm{Frob}_o^{\deg(o)n_o} \circ \mathrm{Frob}_\infty^{\deg(\infty)n_\infty})(c_{2,\xi}(y)) \end{aligned} \tag{1}$$

et tels que

$$c_{1,\xi}(y) \in \mathrm{Cht}_{N,\xi}^{r;\leq\alpha p}/a^{\mathbb{Z}} \subset \mathrm{Cht}_{N,\xi}^r/a^{\mathbb{Z}},$$

et où $\mathrm{Aut}(y)$ *est le groupe fini des automorphismes du point fixe y.*

En utilisant

– la description adélique de Drinfeld des points de $\mathrm{Cht}_{N,\xi}^r$ ([Dr 5], [La 1]),

– le cas particulier du « Lemme Fondamental » prouvé par Drinfeld ([Lau 1]),

– la décomposition spectrale de Langlands ([Lan 2], [Mo-Wa]) et la formule des traces d'Arthur-Selberg ([Ar], [La 1]),

– la formule de Cauchy pour calculer les intégrales qui apparaissent dans la formule des traces,

Lafforgue démontre alors :

THÉORÈME. — *Si* $\deg(\infty)$ *et* $\deg(o)$ *sont assez grands en fonction de la double classe* $K_N g K_N$ *et de l'entier* t *et si* p *est assez convexe en fonction du niveau* N *et de la double classe* $K_N g K_N$*, la moyenne de chacune des deux suites périodiques*

$$k \mapsto \mathrm{Lef}_{(\mathrm{Frob}_X^k \times \mathrm{Id}_X)(\xi)}(g \times \mathrm{Frob}_\infty^{\deg(\infty)n_\infty} \times \mathrm{Frob}_o^{\deg(o)n_o}, \mathrm{Cht}_N^{r;\leq\alpha p}/a^{\mathbb{Z}})$$

et

$$k \mapsto \mathrm{Lef}_{(\mathrm{Id}_X \times \mathrm{Frob}_X^k)(\xi)}(g \times \mathrm{Frob}_\infty^{\deg(\infty)n_\infty} \times \mathrm{Frob}_o^{\deg(o)n_o}, \mathrm{Cht}_N^{r;\leq\alpha p}/a^{\mathbb{Z}}),$$

de période le p.g.c.d. de $\deg(o)$, $\deg(\infty)$ *et* $r!$*, est égale à l'expression spectrale suivante*

$$\begin{aligned} &\sum_{\substack{\pi\in\mathcal{A}_r(K_N)\\ \omega_\pi(a)=1}} \mathrm{Tr}_\pi(1_{K_N g K_N}) q^{(\deg(\infty)n_\infty+\deg(o)n_o)\frac{r-1}{2}} S_\infty^{(-n_\infty)}(\pi) S_o^{(n_o)}(\pi) \\ &+ \sum_{\substack{1\leq r'<r\\ 1\leq r''<r}} \sum_{\substack{\pi'\in\mathcal{A}_{r'}(K_N)\\ \pi''\in\mathcal{A}_{r''}(K_N)}} \mathrm{Tr}_{\pi',\pi''}^{\leq\alpha p}(1_{K_N g K_N}, \deg(\infty)n_\infty) S_\infty^{(-n_\infty)}(\pi') S_o^{(n_o)}(\pi'') \end{aligned} \tag{2}$$

où

– $\mathcal{A}_r(K_N) \subset \mathcal{A}_r$ *est un système de représentants des classes d'isomorphie de représentations automorphes cuspidales irréductibles de* $\mathrm{GL}_r(\mathbb{A})$ *qui admettent au moins un vecteur non nul fixe sous* K_N,

– $1_{K_N g K_N}$ *est la fonction caractéristique de* $K_N g K_N \subset \mathrm{GL}_r(\mathbb{A})$ *et* $\mathrm{Tr}_\pi(1_{K_N g K_N})$ *est la trace de l'opérateur* $\pi(1_{K_N g K_N})$,

– *pour* $x = \infty, o$ *et* ν *un entier positif,* $S_x^{(\nu)}(\pi) = z_1(\pi_x)^\nu + \cdots + z_r(\pi_x)^\nu$ *est la somme des* ν*-ièmes puissances des valeurs propres de Hecke de la composante locale non ramifiée* π_x *de* π,

– $\nu \mapsto \mathrm{Tr}_{\pi',\pi''}^{\leq\alpha p}(1_{K_N g K_N}, \nu)$ *est une fonction complexe de l'entier* ν*, qui ne dépend pas des places* $o, \infty \in |X| - S_g$*, qui est de la forme*

$$\sum_{z\in\mathbb{C}^\times} P_z(\nu) z^\nu$$

pour une famille à support fini de polynômes $P_z(u) \in \mathbb{C}[u]$, $z \in \mathbb{C}^\times$*, et qui est identiquement nulle sauf pour un nombre fini de couples* (π', π'').

6. REPRÉSENTATIONS GALOISIENNES

6.1. Représentations de Galois r-négligeables

Rappelons que F est le corps des fonctions rationnelles sur la courbe X et que l'on a fixé une clôture séparable $\overline{F}$ de F. On désignera par $\overline{\mathbb{F}}_q$ la clôture algébrique de $\mathbb{F}_q$ dans $\overline{F}$. On identifiera le groupe de Galois $\Gamma_{\mathbb{F}_q} := \mathrm{Gal}(\overline{\mathbb{F}}_q/\mathbb{F}_q)$ à $\widehat{\mathbb{Z}}$ en choisissant pour générateur topologique de $\Gamma_{\mathbb{F}_q}$ l'élément de Frobenius géométrique (l'inverse de l'élévation à la puissance q-ième dans $\overline{\mathbb{F}}_q$).

Notons E le corps des fractions de $F\otimes F$, c'est-à-dire le corps des fonctions rationnelles sur la surface $X\times X$, et fixons une clôture algébrique $\overline{E}$ de E et un plongement $\overline{F}\otimes_{\overline{\mathbb{F}}_q}\overline{F}\hookrightarrow\overline{E}$ au-dessus du plongement $F\otimes F\hookrightarrow E$. On a donc défini un point géométrique $\overline{\delta}:\mathrm{Spec}(\overline{E})\to X\times X$ localisé en le point générique $\delta=\mathrm{Spec}(E)$ de $X\times X$ et tel que ses images par les deux projections canoniques de $X\times X$ se factorisent par le point géométrique $\overline{\eta}:\mathrm{Spec}(\overline{F})\to X$ localisé au point générique η de X.

Pour tout ouvert non vide U de X, le groupe fondamental de Grothendieck $\pi_1(U,\overline{\eta})$ est le quotient de $\Gamma_F=\mathrm{Gal}(\overline{F}/F)$ qui classifie les extensions finies de F dans $\overline{F}$ qui sont non ramifiées en toutes les places $x\in|U|$. Il admet comme quotient le groupe de Galois $\Gamma_{\mathbb{F}_q}=\widetilde{\mathbb{Z}}$. De même, pour tout ouvert non vide V de $X\times X$, le groupe fondamental de Grothendieck $\pi_1(V,\overline{\delta})$ est un quotient du groupe de Galois $\mathrm{Gal}(\overline{E}/E)$ et admet $\widetilde{\mathbb{Z}}$ comme quotient.

Soit V un ouvert non vide de $X\times X$ de la forme $U_1\times U_2$ pour des ouverts U_1 et U_2 de X. Cet ouvert est stable par les endomorphismes $\mathrm{Frob}_X\times\mathrm{Id}_X$ et $\mathrm{Id}_X\times\mathrm{Frob}_X$ de $X\times X$ et les foncteurs « images réciproques » par ces endomorphismes sont des équivalences, quasi-inverses l'une de l'autre, de la catégorie des revêtements finis étales de V sur elle-même (le foncteur image réciproque par $\mathrm{Frob}_V=\mathrm{Frob}_{U_1}\times\mathrm{Frob}_{U_2}$ est canoniquement isomorphe à l'identité). On peut donc considérer la catégorie des revêtements finis étales V' de V munis d'un isomorphisme $(\mathrm{Frob}_{U_1}\times\mathrm{Id}_{U_2})^*V'\xrightarrow{\sim}V'$ ou, ce qui revient au même, d'un isomorphisme $(\mathrm{Id}_{U_1}\times\mathrm{Frob}_{U_2})^*V'\xrightarrow{\sim}V'$. Le foncteur fibre en $\overline{\delta}$ identifie cette catégorie à la catégorie des ensembles finis munis de l'action d'un groupe pro-fini $\widetilde{\pi}_1(U_1\times U_2,\overline{\delta})$ qui s'insère dans un diagramme commutatif

$$\begin{array}{ccccccccc}
1 & \longrightarrow & \pi_1(U_1\times U_2,\overline{\delta}) & \longrightarrow & \widetilde{\pi}_1(U_1\times U_2,\overline{\delta}) & \longrightarrow & \widehat{\mathbb{Z}} & \longrightarrow & 0\\
 & & \downarrow & & \downarrow & & \| & & \\
0 & \longrightarrow & \widehat{\mathbb{Z}} & \longrightarrow & \widehat{\mathbb{Z}}^2 & \longrightarrow & \widehat{\mathbb{Z}} & \to & 0
\end{array}$$

à lignes exactes et à flèches verticales surjectives. Tout revêtement fini étale V' de $V=U_1\times U_2$ de la forme $V'=U'_1\times U'_2$ où chaque U'_i est un revêtement fini étale de U_i, est canoniquement muni d'un isomorphisme $(\mathrm{Frob}_{U_1}\times\mathrm{Id}_{U_2})^*V'\xrightarrow{\sim}V'$. On a donc un homomorphisme de $\widetilde{\pi}_1(U_1\times U_2,\overline{\delta})$ dans $\pi_1(U_1,\overline{\eta})\times\pi_1(U_2,\overline{\eta})$ qui prolonge l'homomorphisme $\pi_1(U_1\times U_2,\overline{\delta})\to\pi_1(U_1,\overline{\eta})\times\pi_1(U_2,\overline{\eta})$ défini par les deux projections canoniques de $U_1 \times U_2$, et qui induit l'identité entre les quotients $\widehat{\mathbb{Z}}^2$ de sa source et de son but.

Proposition (Drinfeld). — *L'homomorphisme défini ci-dessus*

$$\widetilde{\pi}_1(U_1 \times U_2, \overline{\delta}) \to \pi_1(U_1, \overline{\eta}) \times \pi_1(U_2, \overline{\eta}),$$

est un isomorphisme.

On a fixé dans la section 1 un nombre premier ℓ distinct de la caractéristique de $\mathbb{F}_q$ et une clôture algébrique $\overline{\mathbb{Q}}_\ell$ de $\mathbb{Q}_\ell$.

Rappelons que, pour tout ouvert non vide U de X (resp. V de $X \times X$), le foncteur fibre en $\overline{\eta}$ (resp. $\overline{\delta}$) est une équivalence de la catégorie des systèmes locaux ℓ-adiques (c'est-à-dire des $\overline{\mathbb{Q}}_\ell$-faisceaux lisses) sur U (resp. V) sur la catégorie des représentations ℓ-adiques de $\pi_1(U, \overline{\eta})$ (resp. $\pi_1(V, \overline{\delta})$). Pour tout système local ℓ-adique L sur U (resp. M sur V) et tout point fermé x de U (resp. y de V), la classe de conjugaison dans $\pi_1(U, \overline{\eta})$ (resp. $\pi_1(V, \overline{\delta})$) des éléments de Frobenius géométriques en x (resp. y) définit donc une classe de conjugaison d'endomorphismes de $L_{\overline{\eta}}$ (resp. $M_{\overline{\delta}}$). On note $\mathrm{Tr}(\mathrm{Frob}_x, L)$ (resp. $\mathrm{Tr}(\mathrm{Frob}_y, M)$) sa trace et $\det(1 - T\,\mathrm{Frob}_x, L)$ (resp. $\det(1 - T\,\mathrm{Frob}_y, M)$) son polynôme caractéristique.

Soient U un ouvert non vide de X et M un système local ℓ-adique *irréductible* sur $U \times U$. On considère les propriétés suivantes :

– M est de la forme $L_1 \boxtimes L_2$ pour des systèmes locaux ℓ-adiques (irréductibles) L_1 sur L_2 sur U et, en particulier, il existe des fonctions $\Delta_1, \Delta_2 : |U \times U| \to 1 + T\overline{\mathbb{Q}}_\ell[T]$ telles que

$$\det(1 - T\,\mathrm{Frob}_\xi, M) = \Delta_1(\infty) * \Delta_2(o)$$

pour tout $\xi \in |U \times U|$ de projections $\infty, o \in |U|$, où l'opération $*$ est définie par

$$\prod_{i_1=1}^{r_1} (1 - \alpha_{1,i_1} T) * \prod_{i_2=1}^{r_2} (1 - \alpha_{2,i_2} T) = \prod_{i_1,i_2} (1 - \alpha_{1,i_1} \alpha_{2,i_2} T).$$

– Il existe un isomorphisme $(\mathrm{Frob}_U \times \mathrm{Id}_U)^* M \xrightarrow{\sim} M$ et, en particulier, la fonction $\xi \mapsto \det(1 - T\,\mathrm{Frob}_\xi, M)$ est constante sur $|\infty \times o|$ pour chaque couple (∞, o) de points fermés de U.

La propriété (i) implique évidemment la propriété (ii) et il devrait résulter de la proposition de Drinfeld que les propriétés (i) et (ii) sont en fait équivalentes.

Lafforgue démontre un résultat partiel dans cette direction.

Proposition. — *On suppose de plus que M est ponctuellement pur au sens de* [De] *et qu'il existe des fonctions $\Delta_1, \Delta_2 : |U \times U| \to 1 + T\overline{\mathbb{Q}}_\ell[T]$ telles que*

$$\det(1 - T\,\mathrm{Frob}_\xi, M) = \Delta_1(\infty) * \Delta_2(o)$$

pour tout $\xi \in |U \times U|$ dont les projections $\infty, o \in |U|$ sont distinctes.

Alors, la propriété (ii) *implique la propriété* (i).

Définition. — *Si U est un ouvert de X, un système local ℓ-adique sur $U\times U$ est dit* r-négligeable *si tous ses sous-quotients irréductibles sont des facteurs directs de systèmes locaux de la forme $L_1\boxtimes L_2$ où L_1 et L_2 sont des systèmes locaux ℓ-adiques irréductibles* de rangs $\leq r-1$ *sur U.*

Si λ est une unité ℓ-adique, on notera par $\overline{\mathbb{Q}}_\ell^{(\lambda)}$ le $\overline{\mathbb{Q}}_\ell$-faisceau sur $\mathrm{Spec}(\mathbb{F}_q)$ défini par le caractère $\widehat{\mathbb{Z}}\to\overline{\mathbb{Q}}_\ell^\times$, $1\mapsto\lambda$, du groupe de Galois $\Gamma_{\mathbb{F}_q}$. Si F est un faisceau ℓ-adique sur un ($\mathbb{F}_q$-)schéma S, on notera suivant Deligne $F^{(\lambda)}$ le produit tensoriel de F par l'image réciproque sur S de $\overline{\mathbb{Q}}_\ell^{(\lambda)}$. En particulier, $F^{(q^{-1})}$ n'est autre que la torsion à la Tate $F(1)$ de F. Rappelons que tout système local ℓ-adique L sur un schéma S normal de type fini peut s'écrire sous la forme $L=M^{(\lambda)}$ où λ est une unité ℓ-adique et M est un système local ℓ-adique dont le déterminant (c'est-à-dire la puissance extérieure maximale) est *d'ordre fini.*

Remarque. — Comme $\pi_1(X\times X,\overline{\delta})$ diffère du produit $\pi_1(X,\overline{\eta})\times\pi_1(X,\overline{\eta})$ par une copie de $\widehat{\mathbb{Z}}$, il arrive que, pour deux systèmes locaux ℓ-adiques irréductibles L_1 et L_2 sur X, le système local $L_1\boxtimes L_2$ sur $X\times X$ ne soit pas irréductible. C'est le cas si on a $L_1\cong L_1^{(\lambda)}$ et $L_2\cong L_2^{(\lambda^{-1})}$ pour une racine de l'unité $\lambda\neq 1$ d'ordre divisant à la fois le rang de L_1 et celui de L_2. Cependant un tel système local $L_1\boxtimes L_2$ est toujours semi-simple et ses facteurs irréductibles sont images les uns des autres par les puissances de $\mathrm{Frob}_X\times\mathrm{Id}_X$.

6.2. La récurrence

Soient $N\subset X$ un niveau « sans facteur carré », $p:[0,r]\to\mathbb{R}$ un paramètre de troncature assez convexe par rapport au genre de la courbe X et à N, et d un entier. On a vu dans la section 3 comment Lafforgue construit un morphisme de champs algébriques
$$\overline{\mathrm{Cht}}_N^{r,d;\leq p}\to\overline{\mathrm{Tr}}_N^{r,\tau}\times(X-N)^2$$
lisse purement de dimension relative $2r-2$, dont le composé avec la projection sur $(X-N)^2$ est propre, et dont la restriction au-dessus de l'ouvert $\mathrm{Spec}(\mathbb{F}_q)=\mathrm{Tr}_N^{r,\tau}\subset\overline{\mathrm{Tr}}_N^{r,\tau}$ est le morphisme pôle-zéro $\mathrm{Cht}_N^{r,d;\leq p}\to(X-N)^2$.

Par construction (cf. 3.4), le champ $\widetilde{\mathrm{Tr}}_N^{r,\tau}$ est lisse sur $\mathcal{T}^{r,\tau}\times_{[\mathbb{A}^{r-1}/\mathbb{G}_{\mathrm{m}}^{r-1}],\langle q-1\rangle}[\mathbb{A}^{r-1}/\mathbb{G}_{\mathrm{m}}^{r-1}]$ où $\mathcal{T}^{r,\tau}$ est un champ torique. On sait donc construire des résolutions des singularités $\widetilde{\mathrm{Tr}}_N^{r,\tau}\to\overline{\mathrm{Tr}}_N^{r,\tau}$. Fixons-en une et notons
$$\widetilde{\mathrm{Cht}}_N^{r,d;\leq p}=\overline{\mathrm{Cht}}_N^{r,d;\leq p}\times_{\overline{\mathrm{Tr}}_N^{r,\tau}}\widetilde{\mathrm{Tr}}_N^{r,\tau}.$$

Le morphisme de champs
$$\widetilde{\mathrm{Cht}}_N^{r,d;\leq p}\to(X-N)^2$$
prolonge le morphisme pôle-zéro $\mathrm{Cht}_N^{r,d;\leq p}\to(X-N)^2$ et est encore propre. Il est de plus lisse purement de dimension relative $2r-2$, et le fermé complémentaire de l'ouvert

$\mathrm{Cht}_N^{r,d;\leq p} \subset \widetilde{\mathrm{Cht}}_N^{r,d;\leq p}$ est un diviseur à croisements normaux relatif sur $(X-N)^2$ qui est réunion de diviseurs lisses sur $(X-N)^2$.

En prenant la réunion disjointe sur les entiers d des champs $\widetilde{\mathrm{Cht}}_N^{r,d;\leq p}$ et en divisant par l'action de $a^{\mathbb{Z}}$, on obtient un morphisme de champs

$$\widetilde{\mathrm{Cht}}_N^{r;\leq p}/a^{\mathbb{Z}} \to (X-N)^2$$

qui prolonge le morphisme pôle-zéro $\mathrm{Cht}_N^{r;\leq p}/a^{\mathbb{Z}} \to (X-N)^2$, qui est propre et lisse purement de dimension relative $2r-2$, et pour lequel le fermé complémentaire de l'ouvert $\mathrm{Cht}_N^{r;\leq p}/a^{\mathbb{Z}} \subset \widetilde{\mathrm{Cht}}_N^{r;\leq p}/a^{\mathbb{Z}}$ est un diviseur à croisements normaux relatif sur $(X-N)^2$ qui est réunion de diviseurs lisses sur $(X-N)^2$.

Le théorème suivant est une description partielle de la cohomologie ℓ-adique des variétés de chtoucas compactifiées.

THÉORÈME. — *Pour tout niveau « sans facteur carré » N et tout paramètre de troncature p assez convexe par rapport au genre de la courbe X et à N, les images directes supérieures du faisceau constant de valeur $\overline{\mathbb{Q}}_\ell$ par le morphisme canonique $\widetilde{\mathrm{Cht}}_N^{r;\leq p}/a^{\mathbb{Z}} \to (X-N)^2$ sont toutes des systèmes locaux ℓ-adiques $(r+1)$-négligeables.*

Lafforgue prouve le théorème principal de la section 1 et le théorème ci-dessus (convenablement généralisé à des niveaux N arbitraires) par une récurrence simultanée sur r. Plus précisément, pour tout entier $r' \geq 1$, il considère les propriétés suivantes :

– $\mathrm{P}_1(r')$ Quel que soit $\pi' \in \mathcal{A}_{r'}$, il existe $\sigma(\pi') \in \mathcal{G}_{r'}$ tel que l'on ait l'égalité de facteurs L locaux

$$\iota(L(\sigma(\pi')_x, T)) = L(\pi'_x, T)$$

pour presque toute place $x \notin S_{\sigma(\pi')} \cup S_{\pi'}$.

– $\mathrm{P}_2(r')$ Quel que soit $\sigma' \in \mathcal{G}_{r'}$, il existe $\pi(\sigma') \in \mathcal{A}_{r'}$ tel que l'on ait l'égalité de facteurs L locaux

$$L(\pi(\sigma')_x, T) = \iota(L(\sigma'_x, T))$$

pour presque toute place $x \notin S_{\pi(\sigma')} \cup S_{\sigma'}$.

– $\mathrm{P}_3(r')$ Quels que soient $\pi' \in \mathcal{A}_{r'}$ et $x \in |X| - S_{\pi'}$, les valeurs propres de Hecke $z_1(\pi'), \ldots, z_{r'}(\pi')$ sont toutes de valeur absolue 1.

– $\mathrm{P}_4(r')$ Le théorème ci-dessus (convenablement généralisé à des niveaux N arbitraires) est démontré en rang r'.

Et il prouve :

THÉORÈME. — *Fixons un entier $r \geq 1$. Alors les propriétés $\mathrm{P}_1(r')$, $\mathrm{P}_3(r')$ et $\mathrm{P}_4(r')$ pour $r' = 1, \ldots, r-1$ impliquent ces mêmes propriétés pour $r' = r$.*

Comme on l'a déjà signalé en 1.4, les propriétés $\mathrm{P}_1(1), \ldots, \mathrm{P}_1(r-1)$ impliquent les propriétés $\mathrm{P}_2(1), \ldots, \mathrm{P}_2(r-1), \mathrm{P}_2(r)$, et en particulier la correspondance de Langlands en tout rang $r' < r$. Le théorème ci-dessus implique donc bien le théorème principal de 1.3. Ces mêmes propriétés impliquent aussi des propriétés renforcées où l'on impose que $S_{\sigma(\pi')} = S_{\pi'}$ et les égalités de facteurs L locaux pour toutes les places $x \notin S_{\pi'}$.

Nous terminerons cet exposé en esquissant la preuve du théorème ci-dessus. *Nous nous restreindrons aux niveaux N « sans facteur carré » et aux représentations automorphes $\pi \in \mathcal{A}_r(K_N) \subset \mathcal{A}_r$*. Dans la suite de cette section nous nous concentrerons sur les aspects galoisiens.

6.3. Les strates du bord sont r-négligeables

On fixe pour toute la suite de la section 6 un niveau N « sans facteur carré ».

Pour tout paramètre de troncature $p : [0, r] \to \mathbb{R}$ assez convexe en fonction du genre de la courbe X et de N et pour tout entier ν, on note

$$H_N^{r;\,\leq p,\,\nu},\ \widetilde{H}_N^{r;\,\leq p,\,\nu} \text{ et } \widetilde{H}_{N,\partial}^{r;\,\leq p,\,\nu}$$

les ν-ièmes images directes supérieures à supports compacts du faisceau constant $\overline{\mathbb{Q}}_\ell$ par les morphismes de champs

$$\mathrm{Cht}_N^{r;\,\leq p}/a^{\mathbb{Z}} \to (X-N)^2,\ \widetilde{\mathrm{Cht}}_N^{r;\,\leq p}/a^{\mathbb{Z}} \to (X-N)^2$$

et

$$\left(\widetilde{\mathrm{Cht}}_N^{r;\,\leq p} - \mathrm{Cht}_N^{r;\,\leq p}\right)/a^{\mathbb{Z}} \to (X-N)^2.$$

Ce sont des systèmes locaux ℓ-adiques sur $(X-N)^2$ nuls pour $\nu \notin \{0, 1, \ldots, 4r-4\}$ (et même pour $\nu \notin \{0, 1, \ldots, 4r-6\}$ dans le cas de $\widetilde{H}_{N,\partial}^{r;\,\leq p,\,\nu}$), et on a une suite exacte longue

$$\cdots \to H_N^{r;\,\leq p,\nu} \to \widetilde{H}_N^{r;\,\leq p,\,\nu} \to \widetilde{H}_{N,\partial}^{r;\,\leq p,\,\nu} \to H_N^{r;\,\leq p,\,\nu+1} \to \cdots.$$

Le diviseur à croisements normaux relatif $\left(\widetilde{\mathrm{Cht}}_N^{r;\,\leq p} - \mathrm{Cht}_N^{r;\,\leq p}\right)/a^{\mathbb{Z}}$ sur $(X-N)^2$ est réunion d'une famille finie de diviseurs lisses. Il est donc muni d'une stratification canonique. Une *strate fermée du bord* est une intersection d'une sous-famille non vide de cette famille. La relation d'inclusion définit une relation d'ordre sur les strates fermées du bord. Une *strate ouverte du bord* est le complémentaire dans une strate fermée du bord de la réunion des strates fermées du bord strictement plus petite. Par exemple, pour $N = \varnothing$, les strates ouvertes du bord sont les $\overline{\mathrm{Cht}}_R^{r;\,\leq p}/a^{\mathbb{Z}}$ pour $\varnothing \neq R = \{r_1, \ldots, r_{s-1}\} \subset [r-1]$ définies en 3.3, et les strates fermées du bord sont les adhérences des strates ouvertes du bord.

Toute strate du bord (fermée ou ouverte) est lisse sur $(X-N)^2$, et les images directes supérieures à supports compacts de $\overline{\mathbb{Q}}_\ell$ par la restriction à une telle strate du morphisme pôle-zéro sont toutes des systèmes locaux ℓ-adiques sur $(X-N)^2$, qu'il est commode d'appeler *images directes supérieures de la strate*.

Proposition. — *Supposons que les propriétés* $\mathrm{P}_1(1)\ldots,\mathrm{P}_1(r-1)$ *soient vérifiées. Alors, les systèmes locaux* ℓ*-adiques*

– *images directes supérieures des strates fermées du bord,*
– *images directes supérieures des strates ouvertes du bord*
– $\widetilde{H}_{N,\partial}^{r;\,\leq p,\,\nu}$, $0\leq\nu\leq 4r-6$,

sont tous r*-négligeables.*

Esquissons la preuve de cette proposition pour $N=\varnothing$. On procède par récurrence sur r. Compte tenu de la suite exacte longue ci-dessus, les hypothèses de la proposition et de récurrence assurent que les images directes supérieures par les morphismes

$$\mathrm{Cht}^{r',d';\,\leq p'}\to X\times X$$

sont toutes $(r'+1)$-négligeables quels que soient les entiers $1\leq r'<r$ et d' et le paramètre de troncature p' assez convexe par rapport au genre de X.

Chaque strate $\overline{\mathrm{Cht}}_R^{r,d;\,\leq p}$, $\varnothing\neq R=\{r_1,\ldots,r_{s-1}\}\subset[r-1]$, est finie et radicielle sur un champ $\mathrm{Cht}^{R,d;\,\leq p}$ et lui est donc équivalente du point de vue de la cohomologie ℓ-adique. Par hypothèse de récurrence, tous les sous-quotients irréductibles des images directes supérieures par le morphisme

$$(\infty=\infty_1,o_1=\infty_2,\ldots,o_{s-1}=\infty_s,o_s=o):\mathrm{Cht}^{R,d;\,\leq p}\to X\times X^{s-1}\times X$$

sont des facteurs directs de faisceaux de la forme

$$L_1'\boxtimes\left(\mathop{\boxtimes}_{j=1}^{s-1}(L_j''\otimes L_{j+1}')\right)\boxtimes L_s''$$

pour des systèmes locaux ℓ-adiques irréductibles L_1',L_1'' de rangs $\leq r_1<r$, L_2',L_2'' de rangs $\leq r_2-r_1<r$, ... et L_s',L_s'' de rangs $\leq r-r_{s-1}<r$ sur la courbe X. Par suite, les images directes supérieures par le morphisme $(\infty,o):\overline{\mathrm{Cht}}_R^{r,d;\,\leq p}\to X\times X$ se décomposent en systèmes locaux ℓ-adiques qui sont des facteurs directs de faisceaux de la forme

$$(L_1'\boxtimes L_s'')\otimes H^\nu\left(\overline{\mathbb{F}}_q\otimes_{\mathbb{F}_q}X^{s-1},\mathop{\boxtimes}_{j=1}^{s-1}(L_j''\otimes L_{j+1}')\right),$$

et tous leurs sous-quotients irréductibles sont des facteurs directs de systèmes locaux de la forme $(L_1'\boxtimes L_s'')^{(\lambda)}$ où λ est une valeur propre de Frob_q agissant sur $H^\nu\left(\overline{\mathbb{F}}_q\otimes_{\mathbb{F}_q}X^{s-1},\boxtimes_{j=1}^{s-1}(L_j''\otimes L_{j+1}')\right)$; la proposition s'en suit.

Le cas général est un peu plus difficile mais similaire en utilisant le fait essentiel que les $\overline{\mathrm{Cht}}_N^{r;\,\leq p}/a^{\mathbb{Z}}$ et $\widetilde{\overline{\mathrm{Cht}}}_N^{r;\,\leq p}/a^{\mathbb{Z}}$ sont construits à partir des $\overline{\mathrm{Cht}}^{r;\,\leq p}/a^{\mathbb{Z}}$ en formant un produit fibré au-dessus du champ $\overline{\mathrm{Tr}}_N^r$. On a besoin d'une formule de comptage supplémentaire où interviennent les fonctions d'Euler-Poincaré introduites par Kottwitz (cf. [Lau 1]) et un argument inspiré de [Ha-Ta].

6.4. Séparation de la partie cuspidale de la cohomologie

Soit S un schéma de type fini sur $\mathbb{F}_q$. Un système local ℓ-adique *virtuel* M sur $U \times U$ est une combinaison linéaire formelle $M = \sum_L m_M(L)[L]$ où L parcourt un ensemble de représentants des classes d'isomorphie de systèmes locaux ℓ-adiques irréductibles sur S (fixé une fois pour toute) et où $L \mapsto m_M(L)$ est une fonction sur cet ensemble à valeurs *rationnelles* et à support fini. Le coefficient $m_M(L) \in \mathbb{Q}$ est appelé la *multiplicité* de L dans M. On dit que L *intervient dans* M ou encore que L *est un constituant de* M si $m_M(L) \neq 0$. Un système local ℓ-adique M définit un système local ℓ-adique virtuel $[M]$: $m_M(L)$ est la multiplicité de L dans n'importe quelle filtration de Jordan-Hölder de M.

On dira qu'un système local ℓ-adique virtuel $M = \sum_L m_M(L)[L]$ sur S est effectif si $m_M(L) \geq 0$ quel que soit L. On dira qu'il est *pur de poids* un entier w si tous les L qui interviennent dans M sont ponctuellement purs de poids w. On dira qu'il est *mixte de poids contenus dans* un ensemble W d'entiers si chaque L qui intervient dans M est ponctuellement pur de poids appartenant à W.

Soit maintenant U un ouvert non vide de X. Un système local ℓ-adique *virtuel* M sur $U \times U$ est dit r*-négligeable* si tous les L *qui interviennent dans* M sont r-négligeables. Si $M = \sum_L m_M(L)[L]$ est un système local ℓ-adique virtuel sur $U \times U$, on appelera *morceau non* r*-négligeable* de M et on notera $M_{\mathrm{cusp}} = \sum_L m_{M_{\mathrm{cusp}}}(L)[L]$ le système local ℓ-adique virtuel sur $U \times U$ défini par $m_{M_{\mathrm{cusp}}}(L) = 0$ si L est r-négligeable et $m_{M_{\mathrm{cusp}}}(L) = m_M(L)$ sinon.

En confrontant la formule des points fixes de Grothendieck-Lefschetz ([Gr])

$$\mathrm{Lef}_\xi\left(\mathrm{Frob}^{\deg(\xi)n}_{\mathrm{Cht}_N^{r;\leq p}/a^{\mathbb{Z}}}\right) = \sum_\nu (-1)^\nu \,\mathrm{Tr}(\mathrm{Frob}^n_\xi, H_N^{r;\leq p,\,\nu})$$

et le cas particulier $g = 1$ et $\deg(\infty)n_\infty = \deg(o)n_o = \deg(\xi)n$ de la formule pour la moyenne de nombres de Lefschetz énoncée en 5.2, on voit que sans changer la valeur de l'expression

$$\sum_{\substack{1\leq r'<r\\ 1\leq r''<r}} \sum_{\substack{\pi'\in\mathcal{A}_{r'}(K_N)\\ \pi''\in\mathcal{A}_{r''}(K_N)}} \mathrm{Tr}^{\leq_\alpha p}_{\pi',\pi''}(1_{K_N}, \deg(\xi)n) S_\infty^{\left(-\frac{\deg(\xi)}{\deg(\infty)}n\right)}(\pi') S_o^{\left(\frac{\deg(\xi)}{\deg(o)}n\right)}(\pi'')$$

on peut y remplacer les fonctions $\nu \mapsto \mathrm{Tr}^{\leq p}_{\pi',\pi''}(1_{K_N}, \nu)$ par des fonctions de la forme $\sum_{z\in\mathbb{C}^\times} c_z z^\nu$ pour une famille à support fini de scalaires $c_z \in \mathbb{Q}$, $z \in \mathbb{C}^\times$.

L'hypothèse de récurrence permet de remplacer les π' et π'' qui interviennent dans cette même expression par des systèmes locaux ℓ-adiques irréductibles et purs de poids 0 sur $X - N$. Compte tenu du théorème de pureté de Deligne ([De]) on en déduit :

Proposition. — *Si* $p : [0, r] \to \mathbb{R}$ *est un paramètre de troncature assez convexe en fonction du genre de la courbe* X *et de* N*, il existe un système local* ℓ*-adique virtuel* *mixte de poids entiers* $\widetilde{H}^{r;\,\leq p}_{N,\mathrm{cusp}}$ *sur* $(X-N)^2$ *tel que la différence*

$$\widetilde{H}^{r;\,\leq p}_{N,\mathrm{cusp}} - \frac{1}{r!}\sum_{k=1}^{r!}\sum_{\nu}(-1)^{\nu}[(\mathrm{Frob}_X^k\times \mathrm{Id}_X)^* H_N^{r;\,\leq p,\,\nu}]$$

soit r*-négligeable et que, pour tout couple* (∞, o) *de points fermés* distincts *de* $X-N$*, tout point fermé* $\xi\in\infty\times o$ *et tout entier* $n\geq 1$*, on ait*

$$\mathrm{Tr}(\mathrm{Frob}_{\xi}^n, \widetilde{H}^{r;\,\leq p}_{N,\mathrm{cusp}}) = \sum_{\substack{\pi\in\mathcal{A}_r(K_N)\\ \omega_\pi(a)=1}} q^{(r-1)\deg(\xi)n}\,\mathrm{Tr}_\pi(1_{K_N}) S_\infty^{\left(-\frac{\deg(\xi)}{\deg(\infty)}n\right)}(\pi) S_o^{\left(\frac{\deg(\xi)}{\deg(o)}n\right)}(\pi).$$

L'encadrement de Jacquet et Shalika ([J-S 1])

$$q^{-\frac{1}{2}} < |z_i(\pi_x)|^{\frac{1}{\deg(x)}} < q^{\frac{1}{2}},\ \forall i=1,\dots,r,\ \forall x\in|X-N|,$$

pour les valeurs propres de Hecke de tout $\pi\in\mathcal{A}_r(K_N)$ assure que $\widetilde{H}^{r;\,\leq p}_{N,\mathrm{cusp}}$ est mixte de poids contenus dans $\{2r-3, 2r-2, 2r-1\}$ et qu'il a des constituants de poids $2r-3$ si et seulement si il en a de poids $2r-1$.

Par des arguments de fonctions L pour les système locaux ℓ-adiques sur la surface $(X-N)^2$, Lafforgue déduit de la proposition ci-dessus, du théorème de pureté de Deligne et des résultats de 6.3 :

Corollaire. — (i) *Le système local* ℓ*-adique virtuel* $\widetilde{H}^{r;\,\leq p}_{N,\mathrm{cusp}}$ *sur* $(X-N)^2$ *est effectif, pur de poids* $2r-2$ *et n'admet aucun constituant* r*-négligeable.*

(ii) *Les systèmes locaux* ℓ*-adiques* $\widetilde{H}_N^{r;\,\leq p,\,\nu}$*,* $\nu\neq 2r-2$*, sont tous* r*-négligeables et* $\widetilde{H}^{r;\,\leq p}_{N,\mathrm{cusp}}$ *est exactement le morceau non* r*-négligeable du système local* ℓ*-adique virtuel*

$$\frac{1}{r!}\sum_{k=1}^{r!}[(\mathrm{Frob}_X^k\times\mathrm{Id}_X)^*\widetilde{H}_N^{r;\,\leq p,\,2r-2}].$$

Pour tous paramètres de troncatures $p\leq q$ qui sont assez convexes par rapport au genre de X et au niveau N, on a un carré commutatif

$$\begin{array}{ccc} H_N^{r;\,\leq q,\,*} & \longrightarrow & \widetilde{H}_N^{r;\,\leq q,\,*} \\ \uparrow & + & \downarrow \\ H_N^{r;\,\leq p,\,*} & \longrightarrow & \widetilde{H}_N^{r;\,\leq p,\,*} \end{array}$$

où les deux flèches horizontales et la flèche verticale montante de gauche sont les prolongements par zéro pour les inclusions de $\mathrm{Cht}_N^{r;\,\leq p}\subset\widetilde{\mathrm{Cht}}_N^{r;\,\leq p}$, $\mathrm{Cht}_N^{r;\,\leq q}\subset\widetilde{\mathrm{Cht}}_N^{r;\,\leq q}$ et $\mathrm{Cht}_N^{r;\,\leq p}\subset\mathrm{Cht}_N^{r;\,\leq q}$ et la flèche verticale descendante de droite est induite par la correspondance dans $\widetilde{\mathrm{Cht}}_N^{r;\,\leq p}\times\widetilde{\mathrm{Cht}}_N^{r;\,\leq q}$ qui est l'adhérence du graphe de l'inclusion $\mathrm{Cht}_N^{r;\,\leq p}\subset\mathrm{Cht}_N^{r;\,\leq q}$. La proposition de 6.3, le corollaire ci-dessus et l'existence de ce carré commutatif impliquent facilement le lemme suivant qui est crucial pour la suite.

Lemme. — *Pour tous paramètres de troncatures $p \leq q$ qui sont assez convexes par rapport au genre de X et au niveau N, les noyau et conoyau de l'homomorphisme de prolongement par zéro*

$$H_N^{r;\, \leq p,\, 2r-2} \to H_N^{r;\, \leq q,\, 2r-2}$$

sont r-négligeables.

En utilisant ce lemme, Lafforgue prouve que le ind-système local ℓ-adique « de rang infini » sur $(X-N)^2$

$$H_N^{r;\, 2r-2} = \varinjlim_p H_N^{r;\, \leq p,\, 2r-2},$$

où p parcourt l'ensemble des paramètres de troncature qui sont assez convexes par rapport au genre de la courbe X et à N, a la propriété suivante :

Lemme. — *Il existe une unique filtration finie*

$$F^\bullet = ((0) = F^0 \subseteq F^1 \subsetneq F^2 \subsetneq \cdots \subsetneq F^{2u+1} \subsetneq F^{2u} \subsetneq \cdots \subsetneq F^T = H_N^{r;\, 2r-2}$$

telle que :

– pour tout entier $u \geq 0$ tel que $2u+1 \leq T$, F^{2u+1}/F^{2u} est la somme de tous les sous-systèmes locaux ℓ-adiques (de rang fini) r-négligeables de $H_N^{r;\, 2r-2}/F^{2u}$,

– pour tout entier $u \geq 0$ tel que $2u+2 \leq T$, F^{2u+2}/F^{2u+1} est la somme de tous les sous-systèmes locaux ℓ-adiques (de rang fini) $H_N^{r;\, 2r-2}/F^{2u+1}$ dont aucun sous-quotient n'est r-négligeable,

– si p est un paramètre de troncature assez convexe par rapport au genre de X et au niveau N et si on note $F^{\leq p,\, \bullet}$ la filtration sur $H_N^{r;\, \leq p,\, 2r-2}$ induite par $F^\bullet$, alors, pour tout entier $u \geq 0$, le plongement

$$F^{\leq p,\, 2u+2}/F^{\leq p,\, 2u+1} \hookrightarrow F^{2u+2}/F^{2u+1}$$

est un isomorphisme.

Considérons alors le système local ℓ-adique (de rang fini) sur $(X-N)^2$

$$H^r_{N,\mathrm{cusp}} = \bigoplus_{u \geq 0} F^{2u+2}/F^{2u+1}.$$

Rassemblant les résultats de cette sous-section, on obtient finalement :

Proposition. — (i) *Pour tout paramètre de troncature p assez convexe par rapport au genre de X et au niveau N, on a l'égalité de systèmes locaux ℓ-adiques virtuels*

$$\widetilde{H}^{r;\, \leq p}_{N,\mathrm{cusp}} = \frac{1}{r!} \sum_{k=1}^{r!} [(\mathrm{Frob}^k_{X-N} \times \mathrm{Id}_{X-N})^* H^r_{N,\mathrm{cusp}}].$$

(ii) *Le système local ℓ-adique $H^r_{N,\mathrm{cusp}}$ est ponctuellement pur de poids $2r-2$.*

(iii) *Pour tout couple (∞, o) de points fermés distincts de $X - N$, tout point fermé ξ de $\infty \times o$ et tout entier n, on a*

$$\frac{1}{r!}\sum_{k=1}^{r!} \mathrm{Tr}(\mathrm{Frob}_\xi^n, (\mathrm{Frob}_{X-N}^k \times \mathrm{Id}_{X-N})^* H_{N,\mathrm{cusp}}^r)$$
$$= q^{(r-1)\deg(\xi)n} \sum_{\substack{\pi\in\mathcal{A}_r(K_N)\\ \omega_\pi(a)=1}} \mathrm{Tr}_\pi(1_{K_N}) S_\infty^{\left(-\frac{\deg(\xi)}{\deg(\infty)}n\right)}(\pi) S_o^{\left(\frac{\deg(\xi)}{\deg(o)}n\right)}(\pi).$$

7. UNE VARIANTE D'UN THÉORÈME DE PINK

7.1. Le cas des schémas

Soient S un schéma de type fini et lisse (sur $\mathbb{F}_q$) et $p : X \to S$ un morphisme de schémas propre et lisse purement de dimension relative d. On se donne un diviseur $Y \subset X$ à croisements normaux relatif sur S qui est réunion $Y = Y_1 \cup \cdots \cup Y_m$ de m diviseurs lisses sur S. Pour toute partie I de $\{1, \ldots, m\}$ on note $Y_I = \bigcap_{i\in I} Y_i$ et $X_I = Y_I - \bigcup_{J \supsetneq I} Y_J$. On a donc $X_\varnothing = X - Y \subset Y_\varnothing = X$ et X est la réunion disjointe de ses parties localement fermées X_I. De plus, pour chaque $I \subset \{1, \ldots, m\}$, la restriction p_I de p à Y_I est propre et lisse purement de dimension relative $d - |I|$, et X_I est un ouvert dense de Y_I.

On considère un endomorphisme $f : S \to S$ et un morphisme fini de schémas

$$\Gamma_\varnothing \to Z_\varnothing := X_\varnothing \times_{f,S} X_\varnothing \subset X_\varnothing \times X_\varnothing$$

où « $\times_{f,S}$ » est une notation abrégée pour « $\times_{p_\varnothing \circ f, S, p_\varnothing}$ ». On suppose que la première projection $\mathrm{pr}_{\varnothing,1} : \Gamma_\varnothing \to X_\varnothing$ est étale. Le S-schéma est donc lisse purement de dimension relative d sur S. En particulier il est normal.

Alors, si s est un point fermé de S et si $n \geq 1$ est un entier tel que

$$\mathrm{Frob}_S^n(f(s)) = f(\mathrm{Frob}_S^n(s)) = s,$$

la fibre $\Gamma_{\varnothing,s} \to Z_{\varnothing,s} \subset X_{\varnothing,f(s)} \times X_{\varnothing,s}$ en s de $\Gamma_\varnothing$ coupe transversalement la fibre $X_{\varnothing,f(s)} \to X_{\varnothing,s} \times X_{\varnothing,f(s)}$ en $(s, f(s))$ du graphe $\delta_\varnothing^n = (\mathrm{Frob}_{X_\varnothing}^n, \mathrm{Id}_{X_\varnothing}) : X_\varnothing \to X_\varnothing \times X_\varnothing$. On note

$$\mathrm{Lef}_s(\Gamma_\varnothing \times \mathrm{Frob}_{X_\varnothing}^n, X_\varnothing)$$

le nombre des points d'intersection.

On note $\Gamma \to Z := X \times_{f,S} X \subset X \times X$ le morphisme fini de schémas obtenu par normalisation de Z dans $\Gamma_\varnothing$. On se propose de donner une interprétation cohomologique de $\mathrm{Lef}_s(\Gamma_\varnothing \times \mathrm{Frob}_X^n, X_\varnothing)$ sous une hypothèse de « stabilité de l'ouvert $X_\varnothing \subset X$ au voisinage des points fixes de Γ ».

7.2. Stabilité au voisinage des points fixes

On note $\mathrm{pr}_1, \mathrm{pr}_2 : \Gamma \to X$ les deux projections canoniques de Γ et, pour chaque entier $n \geq 0$, δ^n le graphe $(\mathrm{Frob}_X^n, \mathrm{Id}_X) : X \to X \times X$.

Définition. — *On dira que Γ stabilise $X_\varnothing$ au voisinage de ses points fixes s'il existe un ouvert $U \subset X \times X$ contenant toutes les intersections de l'image de Γ dans $Z \subset X \times X$ avec les images des graphes de Frobenius $\delta^n(X) \subset X \times X$, $n \geq 0$, et tel que*

$$\mathrm{pr}_1(\mathrm{pr}_2^{-1}(X_\varnothing) \cap U_\Gamma) \subset X_\varnothing$$

où $U_\Gamma \subset \Gamma$ est la trace de U sur Γ.

Lemme. — *Pour que Γ stabilise $X_\varnothing$ au voisinage de ses points fixes, il faut et il suffit que Γ satisfasse au critère valuatif suivant :*

$(*)$ *Pour tout anneau de valuation discrète A de corps des fractions K et de corps résiduel k et pour tout point $\gamma : \mathrm{Spec}(A) \to \Gamma$ tel que $\mathrm{pr}_1(\gamma(\mathrm{Spec}(K))) \in Y(K)$ et $\mathrm{pr}_2(\gamma(\mathrm{Spec}(K))) \in X_\varnothing(K)$, l'image de $\gamma(\mathrm{Spec}(k)) \in \Gamma(k)$ dans $Z \subset X \times X$ n'est dans l'image d'aucun des graphes de Frobenius δ^n, $n \in \mathbb{N}$.*

Comme l'a fait Pink dans [Pi], Lafforgue introduit l'éclatement $\pi : \widetilde{Z} \to Z$ de $Z = X \times_{f,S} X$ le long du fermé réunion des $Y_i \times_{f,S} Y_i$, $1 \leq i \leq m$. Cet éclatement est aussi le produit fibré au-dessus de Z des éclatements $\widetilde{Z}_i \to Z$, $1 \leq i \leq m$, de Z le long des $Y_i \times_{f,S} Y_i$. Il est donc muni de diviseurs E_i, $1 \leq i \leq m$, qui sont les images réciproques dans $\widetilde{Z}$ des diviseurs exceptionnels dans les $\widetilde{Z}_i$. Pour toute partie non vide $I \subset \{1, \ldots, m\}$, on vérifie que

$$E_I := \bigcap_{i \in I} E_i \subset \bigcap_{i \in I} \pi^{-1}(Y_i \times_{f,S} Y_i) = \pi^{-1}(Y_I \times_{f,S} Y_I),$$

que E_I est le produit fibré sur Z des diviseurs exceptionnels dans les $\widetilde{Z}_i$ pour $i \in I$ et des $\widetilde{Z}_i$ pour $i \notin I$, que E_I est lisse sur S purement de dimension relative $2d - |I|$ et que la restriction de $\pi_I : E_I \to Y_I \times_{f,S} Y_I$ au-dessus de l'ouvert $X_I \times_{f,S} X_I$ est un fibré projectif de rang $|I|$. Les E_i sont donc des diviseurs lisses sur S et leur réunion $E = E_1 \cup \cdots \cup E_m$ est un diviseur à croisements normaux relatif sur S.

Par construction $\pi : \widetilde{Z} \to Z$ est un isomorphisme au-dessus de $Z_\varnothing$. On notera $\widetilde{\Gamma} \to \widetilde{Z}$ la normalisation de $\Gamma_\varnothing$ dans $\widetilde{Z}$. Bien entendu π « envoie » $\widetilde{\Gamma}$ dans Γ.

Pour tout point fermé $s \in S$ la fibre $\pi_s : \widetilde{Z}_s \to Z_s$ en s de π est l'éclatement de Z_s le long de la réunion des fermés $Y_{i,s} \times_{f,f(s)} Y_{i,f(s)}$ de $Z_s = X_s \times_{f,f(s)} X_{f(s)}$. Pour tout point fermé $s \in S$ et tout entier $n \geq 0$ tel que $\mathrm{Frob}_S^n(f(s)) = f(\mathrm{Frob}_S^n(s)) = s$, la fibre $\delta^n_{(s,f(s))} : X_{f(s)} \to Z_s \subset X_s \times X_{f(s)}$ en $(s, f(s))$ du graphe de Frobenius se relève canoniquement en un morphisme

$$\widetilde{\delta}^n_{(s,f(s))} : X_{f(s)} \to \widetilde{Z}_s$$

puisque, pour chaque $i = 1, \dots, m$, l'image réciproque par $\delta^n_{(s,f(s))}$ du fermé $Y_{i,s} \times_{f,f(s)} Y_{i,f(s)} \subset Z_s$ est le diviseur $Y_{i,f(s)} \subset X_{f(s)}$.

La proposition suivante prolonge un résultat de Pink ([Pi]).

Proposition. — *Supposons que Γ stabilise $X_\varnothing$ au voisinage de ses points fixes. Alors, quitte à remplacer f par*

$$f^{(n_0)} := \mathrm{Frob}_S^{n_0} \circ f = f \circ \mathrm{Frob}_S^{n_0}$$

et $\Gamma_\varnothing$ par

$$\Gamma_\varnothing^{(n_0)} := (X_\varnothing \times X_\varnothing) \times_{\mathrm{Frob}^{n_0}_{X_\varnothing} \times \mathrm{Id}_{X_\varnothing}, X_\varnothing \times X_\varnothing} \Gamma_\varnothing.$$

pour un entier $n_0 \geq 0$ assez grand, la normalisation $\widetilde{\Gamma}$ vérifie la propriété suivante : pour tout point fermé $s \in S$ et tout entier $n > 0$ tel que $\mathrm{Frob}_S^n(f(s)) = f(\mathrm{Frob}_S^n(s)) = s$, l'image de $\widetilde{\delta}^n_{f,f(s)}$ dans $\widetilde{Z}_s$ ne rencontre pas l'image de $\widetilde{\Gamma}_s$ en dehors de $Z_{\varnothing,s} \subset \widetilde{Z}_s$.

7.3. Formule des points fixes

Rappelons que l'on a fixé un nombre premier ℓ distinct de la caractéristique de $\mathbb{F}_q$. Considérons la cohomologie ℓ-adique de X et Z au-dessus de S

$$Rp_*\mathbb{Q}_\ell \text{ et } Rq_*\mathbb{Q}_\ell \in D^{\mathrm{b}}_{\mathrm{c}}(S, \mathbb{Q}_\ell)$$

où $q : X \times_{f,S} X \to S$ est la projection canonique, et plus généralement les cohomologies ℓ-adique des Y_I au-dessus de S

$$Rp_{I,*}\mathbb{Q}_\ell \text{ et } Rq_{I,*}\mathbb{Q}_\ell \in D^{\mathrm{b}}_{\mathrm{c}}(S, \mathbb{Q}_\ell)$$

où $q_I : Y_I \times_{f,S} Y_I \to S$ est la projection canonique. La formation de ces cohomologies commute aux changements de base $S' \to S$. En particulier, la fibre en un point géométrique $\overline{s}$ de S de $Rp_{I,*}\mathbb{Q}_\ell$ est la cohomologie ℓ-adique $R\Gamma(Y_{I,\overline{s}}, \mathbb{Q}_\ell)$ de la fibre de p_I en ce point.

Nos hypothèses assurent que les faisceaux de cohomologie $R^j p_{I,*}\mathbb{Q}_\ell$ (resp. $R^j q_{I,*}\mathbb{Q}_\ell$) sont tous lisses sur S et qu'ils sont nuls pour les j qui ne sont pas dans l'intervalle $[0, 2(d-|I|)]$ (resp. $[0, 4(d-|I|])$.

Soient $I \subset \{0, 1, \dots, m\}$, n un entier ≥ 1 et $u_I \in H^0(S, R^{2(d-|I|)} q_{I,*}\mathbb{Q}_\ell)(d-|I|)$ une classe de cohomologie. Alors :

– l'endomorphisme de Frobenius $\mathrm{Frob}^n_{Y_I}$ qui relève Frob^n_S, induit un homomorphisme

$$\mathrm{Frob}^n_{Y_I} : (\mathrm{Frob}^n_S)^* Rp_{I,*}\mathbb{Q}_\ell \to Rp_{I,*}\mathbb{Q}_\ell,$$

– d'après la formule de Künneth et la dualité de Poincaré, u_I peut être vu comme un homomorphisme

$$u_I : f^* Rp_{I,*}\mathbb{Q}_\ell \to Rp_{I,*}\mathbb{Q}_\ell,$$

– les homomorphismes composés

$$\mathrm{Frob}_{Y_I}^n \circ (\mathrm{Frob}_S^n)^*(u_I) : (\mathrm{Frob}_S^n)^* f^* Rp_{I,*}\mathbb{Q}_\ell \to Rp_{I,*}\mathbb{Q}_\ell$$

et

$$u_I \circ f^*(\mathrm{Frob}_{Y_I}^n) : f^*(\mathrm{Frob}_S^n)^* Rp_{I,*}\mathbb{Q}_\ell \to Rp_{I,*}\mathbb{Q}_\ell$$

coïncident.

On notera

$$u_I \times \mathrm{Frob}_{Y_I}^n : (\mathrm{Frob}_S^n)^* f^* Rp_{I,*}\mathbb{Q}_\ell = f^*(\mathrm{Frob}_S^n)^* Rp_{I,*}\mathbb{Q}_\ell \to Rp_{I,*}\mathbb{Q}_\ell$$

ces homomorphismes composés. Pour tout point géométrique $\overline{s}$ de S qui vérifie $f(\mathrm{Frob}_S^n(\overline{s})) = \mathrm{Frob}_S^n(f(\overline{s})) = \overline{s}$, $u_I \times \mathrm{Frob}_{Y_I}^n$ induit un endomorphisme de la cohomologie $R\Gamma(Y_{I,\overline{s}}, \mathbb{Q}_\ell)$ de la fibre en $\overline{s}$ de $p_I : Y_I \to S$ dont on notera

$$\mathrm{tr}(u_I \times \mathrm{Frob}_{Y_I}^n, R\Gamma(Y_{I,\overline{s}}, \mathbb{Q}_\ell)) \in \mathbb{Q}_\ell$$

la trace.

Le « cycle de codimension d » Γ sur $X \times_{f,S} X$ admet une classe de cohomologie

$$\mathrm{cl}(\Gamma) \in H^0(S, R^{2d}q_*\mathbb{Q}_\ell)(d).$$

De même, si on note $\widetilde{q} = q \circ \pi : \widetilde{Z} \to S$, $\widetilde{\Gamma}$ admet une classe de cohomologie

$$\mathrm{cl}(\widetilde{\Gamma}) \in H^0(S, R^{2d}\widetilde{q}_*\mathbb{Q}_\ell)(d).$$

Pour chaque $I \subset \{1, \ldots, m\}$ non vide, on a un homomorphisme de faisceaux

$$R^{2d}\widetilde{q}_*\mathbb{Q}_\ell(d) \to R^{2d}\widetilde{q}_{I,*}\mathbb{Q}_\ell(d) \to R^{2(d-|I|)}q_{I,*}\mathbb{Q}_\ell(d-|I|)$$

où $\widetilde{q}_I : E_I \to S$ est la projection canonique, où la première flèche est l'homomorphisme de restriction au fermé $E_I \subset \widetilde{Z}$ et où la seconde flèche est duale de l'homomorphisme

$$\pi_I^* : R^{2(d-|I|)}q_{I,*}\mathbb{Q}_\ell(d-|I|) \to R^{2(d-|I|)}\widetilde{q}_{I,*}\mathbb{Q}_\ell(d-|I|).$$

On note $\mathrm{cl}(\Gamma)_I$ l'image de $\mathrm{cl}(\widetilde{\Gamma})$ par cet homomorphisme composé.

Plus généralement on peut refaire cette construction après avoir remplacé f par $f^{(n_0)}$ et $\Gamma_\varnothing$ par $\Gamma_\varnothing^{(n_0)}$ pour un entier $n_0 \geq 0$ (cf. la proposition de 7.2). Pour chaque $I \subset \{1, \ldots, m\}$ non vide, on obtient une section globale $\mathrm{cl}(\Gamma^{(n_0)})_I$ de $R^{2(d-|I|)}q_{I,*}^{(n_0)}\mathbb{Q}_\ell(d-|I|)$, où $q_I^{(n_0)} : Y_I \times_{f^{(n_0)},S} Y_I \to S$ est la projection canonique. L'image directe de cette section globale par le morphisme radiciel $\mathrm{Frob}_{Y_I}^{n_0} \times \mathrm{Id}_{Y_I} : Y_I \times_{f^{(n_0)},S} Y_I \to Y_I \times_{f,S} Y_I$ est une section globale notée $\mathrm{cl}(\Gamma)_I^{(n_0)}$ de $R^{2(d-|I|)}q_{I,*}\mathbb{Q}_\ell(d-|I|)$.

Avec ces notations, Lafforgue prouve la généralisation suivante de la formule des points fixes de Grothendieck-Lefschetz ([Gr]).

THÉORÈME. — *Supposons que* Γ *stabilise l'ouvert* $X_\varnothing$ *au voisinage de ses points fixes et soit* $n_0 \geq 0$ *un entier vérifiant la conclusion de la proposition de* 7.3.

Alors, pour tout point fermé s *de* S *et tout entier* $n > n_0$ *tels que* $\mathrm{Frob}_S^n(f(s)) = f(\mathrm{Frob}_S^n(s)) = s$, *on a*

$$\begin{aligned}\mathrm{Lef}_s(\Gamma_\varnothing \times \mathrm{Frob}_{X_\varnothing}^n, X_\varnothing) &= \mathrm{tr}(\mathrm{cl}(\Gamma) \times \mathrm{Frob}_X^n, R\Gamma(X_{\overline{s}}, \mathbb{Q}_\ell)) \\ &\quad + \sum_{\substack{I \subset \{1,\dots,m\} \\ I \neq \varnothing}} (-1)^{|I|}\, \mathrm{tr}(\mathrm{cl}(\Gamma)_I^{(n_0)} \times \mathrm{Frob}_{Y_I}^n, R\Gamma(Y_{I,\overline{s}}, \mathbb{Q}_\ell)).\end{aligned} \tag{3}$$

7.4. Extension à certains champs algébriques

Comme le montre Lafforgue, les résultats de cette section valent encore après avoir remplacé le S-schéma $p : X \to S$, le diviseur $Y = Y_1 \cup \cdots \cup Y_m$ et le morphisme de schémas $\Gamma_\varnothing \to Z_\varnothing$ par un S-champ algébrique $p : \mathcal{X} \to S$, un diviseur $\mathcal{Y}$ dans $\mathcal{X}$ et un morphisme de champs algébriques $\Gamma_\varnothing \to \mathcal{Z}_\varnothing$ qui vérifient les propriétés suivantes :

– le morphisme $p : \mathcal{X} \to S$ est de type fini,

– au voisinage de chacun de ses points le champ $\mathcal{X}$ admet un recouvrement fini et plat par un espace algébrique muni de l'action d'un groupe fini qui agit transitivement sur les fibres de ce recouvrement,

– le morphisme p satisfait au critère valuatif de propreté,

– le morphisme p est lisse purement de dimension relative d,

– $\mathcal{Y}$ est une réunion de diviseurs $\mathcal{Y}_1, \dots, \mathcal{Y}_m$ qui sont lisses sur S et est à croisement normaux relatif sur S,

– $\mathcal{X}_\varnothing$ est un S-champ algébrique de Deligne-Mumford, le morphisme $\Gamma_\varnothing \to \mathcal{Z}_\varnothing = \mathcal{X}_\varnothing \times_{f,S} \mathcal{X}_\varnothing$ est représentable fini et la première projection de $\Gamma_\varnothing$ sur $\mathcal{X}_\varnothing$ est étale.

8. FIN DE LA RÉCURRENCE

8.1. Ce qu'il reste à démontrer

Soit $N \subset X$ un niveau « sans facteur carré ». Pour tout $g \in \mathrm{GL}_r(\mathbb{A})$ la correspondance de Hecke

$$c = (c_1, c_2) : \mathrm{Cht}_N^r(g)/a^{\mathbb{Z}} \to (\mathrm{Cht}_N^r/a^{\mathbb{Z}}) \times_{(X-N)^2} (\mathrm{Cht}_N^r/a^{\mathbb{Z}})$$

envoie $\mathrm{Cht}_N^{r;\leq p}/a^{\mathbb{Z}}$ dans $\mathrm{Cht}_N^{r;\leq q}/a^{\mathbb{Z}}$ dès que $q - p$ est assez convexe en fonction de $K_N g K_N$. Par conséquent, le ind-système local ℓ-adique $H_N^{r;\,2r-2}$ défini en 6.4 est muni d'une action de la $\overline{\mathbb{Q}}_\ell$-algèbre de Hecke $\mathcal{H}(K_N) = \mathcal{C}_c(\mathrm{GL}_r(\mathbb{A})//K_N)$ des fonctions à valeurs dans $\overline{\mathbb{Q}}_\ell$, K_N-invariantes à gauche et à droite et à supports compacts sur $\mathrm{GL}_r(\mathbb{A})$.

De même, les morphismes de Frobenius partiels induisent un isomorphisme

$$(\mathrm{Frob}_{X-N}\times \mathrm{Id}_{X-N})^* H_N^{r;\,2r-2} \xrightarrow{\sim} H_N^{r;\,2r-2}$$

qui commute à l'action de $\mathcal{H}(K_N)$.

Comme la construction faite en 6.4 du système local ℓ-adique $H^r_{N,\mathrm{cusp}}$ à partir de $H_N^{r;\,2r-2}$ est canonique, $H^r_{N,\mathrm{cusp}}$ est lui aussi muni d'une action de l'algèbre de Hecke $\mathcal{H}(K_N)$ et d'un isomorphisme $(\mathrm{Frob}_{X-N}\times \mathrm{Id}_{X-N})^* H^r_{N,\mathrm{cusp}} \xrightarrow{\sim} H^r_{N,\mathrm{cusp}}$. Dans l'énoncé de la dernière proposition de la section 6.4, on peut donc remplacer la moyenne $\frac{1}{r!}\sum_{k=1}^{r!}[(\mathrm{Frob}^k_{X-N}\times \mathrm{Id}_{X-N})^* H^r_{N,\mathrm{cusp}}]$ par $[H^r_{N,\mathrm{cusp}}]$.

Dans cette dernière section nous allons montrer :

Proposition. — *Pour tout $g\in \mathrm{GL}_r(\mathbb{A})$, tout couple (∞, o) de points fermés distincts de $X - S_g$ (où $S_g \supset |N|$ est l'ensemble fini de points fermés de X défini en* 5.1*), tout point fermé ξ de $\infty\times o$ et tout entier n, on a*

$$\begin{aligned}\mathrm{Tr}(&1_{K_N g K_N}\times \mathrm{Frob}^n_\xi, H^r_{N,\mathrm{cusp}})\\ &= q^{(r-1)\deg(\xi)n}\sum_{\substack{\pi\in\mathcal{A}_r(K_N)\\ \omega_\pi(a)=1}} \mathrm{Tr}_\pi(1_{K_N g K_N}) S_\infty^{\left(-\frac{\deg(\xi)}{\deg(\infty)}n\right)}(\pi) S_o^{\left(\frac{\deg(\xi)}{\deg(o)}n\right)}(\pi).\end{aligned}$$

Cette proposition généralise la partie (iii) de la dernière proposition de la sous-section 6.4 et des arguments standard permettent d'en déduire les propriétés $\mathrm{P}_1(r)$, $\mathrm{P}_2(r)$ (pour les $\pi\in\mathcal{A}_r(K_N)$) et $\mathrm{P}_3(r)$ (pour les N « sans facteur carré »).

8.2. Trace des opérateurs de Hecke sur $\widetilde{H}^{r;\,\leq p}_{N,\mathrm{cusp}}$

Lemme. — *Soient U un ouvert non vide de X et M un système local ℓ-adique sur $U\times U$ muni d'un endomorphisme $h : M\to M$. Alors, une fois fixé un ensemble de représentants des classes d'isomorphie de systèmes locaux ℓ-adiques irréductibles sur $U\times U$, il existe une unique application à support fini $L\mapsto \lambda_L$ de cet ensemble dans $\overline{\mathbb{Q}}_\ell$ telle que*

$$\mathrm{Tr}(h\circ\gamma, M_{\overline{\delta}}) = \sum_L \lambda_L \,\mathrm{Tr}(\gamma, L_{\overline{\delta}})$$

pour tout $\gamma\in\pi_1(U\times U, \overline{\delta})$.

Si $[M]=\sum_L m_M(L)[L]$ est le système local ℓ-adique virtuel sur $U\times U$ associé à M et si $M_{\mathrm{cusp}}=\sum_L m_{M_{\mathrm{cusp}}}(L)[L]$ est son morceau non r-négligeable (cf. 6.4), on notera

$$\mathrm{Tr}_{M_{\mathrm{cusp}}}(h\circ\gamma) = \sum_{\substack{L\\ m_M(L)\neq 0}} \frac{m_{M_{\mathrm{cusp}}}(L)}{m_M(L)}\lambda_L \,\mathrm{Tr}(\gamma, L_{\overline{\delta}})$$

pour tout $\gamma\in\pi_1(U\times U, \overline{\delta})$.

Bien entendu, on va appliquer ceci au morceau non r-négligeable $\widetilde{H}_{N,\mathrm{cusp}}^{r;\,\leq p}$ de la cohomologie ℓ-adique $\widetilde{H}_N^{r;\,\leq p,\,*}$ et à l'endomorphisme de $\widetilde{H}_N^{r;\,\leq p,\,2r-2}$ induit par une correspondance de Hecke.

Plus précisément, soit $p : [0,r] \to \mathbb{R}$ un paramètre de troncature assez convexe par rapport au genre de la courbe X et à N et soit $g \in \mathrm{GL}_r(\mathbb{A})$. D'une part, on a déjà vu que la correspondance de Hecke $c : \mathrm{Cht}_N^r(g) \to \mathrm{Cht}_N^r \times_{(X-N)^2} \mathrm{Cht}_N^r$ envoie $\mathrm{Cht}_N^{r;\,\leq p}$ dans $\mathrm{Cht}_N^{r;\,\leq q}$ pour un paramètre de troncature $q \geq p$ convenable et qu'elle induit un morphisme entre les cohomologies à supports compacts $h : H_N^{r;\,\leq p,\,*} \to H_N^{r;\,\leq q,\,*}$. D'autre part, cette correspondance induit par restriction un cycle

$$c^{\leq p} : \mathrm{Cht}_N^{r;\,\leq p}(g) \to \mathrm{Cht}_N^{r;\,\leq p} \times_{(X-N)^2} \mathrm{Cht}_N^{r;\,\leq p},$$

puis par normalisation une correspondance

$$\widetilde{c}^{\leq p} : \widetilde{\mathrm{Cht}}_N^{r;\,\leq p}(g) \to \widetilde{\mathrm{Cht}}_N^{r;\,\leq p} \times_{(X-N)^2} \widetilde{\mathrm{Cht}}_N^{r;\,\leq p},$$

et donc un endomorphisme noté $\widetilde{h}$ de la cohomologie ℓ-adique $\widetilde{H}_N^{r;\,\leq p,\,*}$, d'où une trace

$$\mathrm{Tr}_{\widetilde{H}_{N,\mathrm{cusp}}^{r;\,\leq p}}(\widetilde{h} \circ \gamma)$$

pour tout $\gamma \in \pi_1((X-N)^2, \overline{\delta})$.

Quitte à agrandir q, on peut supposer que q est assez convexe par rapport au genre de la courbe X et à N. On alors un diagramme commutatif

$$\begin{array}{ccccc}
H_N^{r;\,\leq q,\,*} & & \xrightarrow{\qquad\qquad\qquad} & & \widetilde{H}_N^{r;\,\leq q,\,*} \\
{\scriptstyle h}\Big\uparrow & & + & & \Big\downarrow \\
H_N^{r;\,\leq p,\,*} & \longrightarrow & \widetilde{H}_N^{r;\,\leq p,\,*} & \underset{\widetilde{h}}{\longrightarrow} & \widetilde{H}_N^{r;\,\leq p,\,*}
\end{array}$$

où les flèches h et $\widetilde{h}$ sont celles définies ci-dessus et les autres flèches sont les mêmes que dans le carré commutatif de 6.4. Par un argument facile utilisant les résultats de 6.4, Lafforgue en déduit :

Proposition. — *Pour tout paramètre de troncature p assez convexe en fonction de N, tout point fermé ξ de $(X-N)^2$ et tout entier n, on a*

$$\mathrm{Tr}(1_{K_N g K_N} \times \mathrm{Frob}_\xi^n, H_{N,\mathrm{cusp}}^r) = \mathrm{Tr}_{\widetilde{H}_{N,\mathrm{cusp}}^{r;\,\leq p}}(\widetilde{h} \circ \mathrm{Frob}_\xi^n).$$

8.3. Application de la variante du théorème de Pink

Fixons un paramètre de troncature $p : [0,r] \to \mathbb{R}$ assez convexe par rapport au genre de la courbe X et à N, et un élément $g \in \mathrm{GL}_r(\mathbb{A})$. On a défini un ensemble fini $S_g \supset |N|$ de points fermés de X.

On cherche à appliquer les résultats généraux du chapitre 7 à

$$S = (X - S_g)^2,\ f = \mathrm{Id}_S : S \to S,$$

$$\mathcal{X}_\varnothing = (X-S_g)^2 \times_{(X-N)^2} (\mathrm{Cht}_N^{r;\,\leq p}/a^{\mathbb{Z}}) \subset (X-S_g)^2 \times_{(X-N)^2} (\widetilde{\mathrm{Cht}}_N^{r;\,\leq p}/a^{\mathbb{Z}}) = \mathcal{X},$$

$$(c_1, c_2) : (X - S_g)^2 \times_{(X-N)^2} (\mathrm{Cht}_N^r(g)/a^{\mathbb{Z}}) = \Gamma_\varnothing \to \mathcal{X}_\varnothing \times_S \mathcal{X}_\varnothing,$$

et

$$\Gamma \to \mathcal{X} \times_S \mathcal{X}$$

la normalisation de $\mathcal{X}_\varnothing \times_S \mathcal{X}_\varnothing$ dans $\Gamma_\varnothing$. Le fermé complémentaire $\mathcal{Y} = \mathcal{X} - \mathcal{X}_\varnothing$ est un diviseur à croisements normaux relatif sur $(X - S_g)^2$, qui est réunion $\mathcal{Y} = \mathcal{Y}_1 \cup \cdots \cup \mathcal{Y}_m$ de diviseurs lisses sur $(X - S_g)^2$.

Pour chaque entier $n_0 \geq 0$, on dispose des classes de cohomologie suivantes :

– $\mathrm{cl}(\Gamma)$ dans la cohomologie de $\mathcal{X} \times_S \mathcal{X}$ au-dessus de $(X - S_g)^2$,

– $\mathrm{cl}(\Gamma)_I^{(n_0)}$ dans la cohomologie de $\mathcal{Y}_I \times_S \mathcal{Y}_I$ au-dessus de $(X - S_g)^2$ pour tout $I \subset \{1, \ldots, m\}$ non vide.

Pour tout ouvert U de X notons Λ_U le schéma intersection de tous les ouverts de $U \times U$ complémentaires des images réciproques de la diagonale par les endomorphismes $\mathrm{Frob}_U^n \times \mathrm{Id}_U$ et $\mathrm{Id}_U \times \mathrm{Frob}_U^n$ de $U \times U$ pour tous les entiers $n \geq 0$, c'est-à-dire l'intersection de $\Lambda_X \subset X \times X$ avec $U \times U$.

Nous avons besoin du théorème suivant de Lafforgue, qui prolonge en rang arbitraire un résultat établi par Drinfeld en rang 2 ([Dr 7]).

Théorème. — *Supposons que p est assez convexe et $U \subset X - S_g$ est assez petit relativement à N et à $K_N g K_N$. Il existe alors un ouvert de Zariski $V \subset U \times U$, contenant Λ_U, tel que, après avoir remplacé $S = (X - S_g)^2$ par V et $\mathcal{X}_\varnothing \subset \mathcal{X}$ et $\Gamma_\varnothing \subset \Gamma$ par leurs restrictions à cet ouvert de S, la correspondance Γ stabilise l'ouvert $\mathcal{X}_\varnothing$ au voisinage de ses points fixes.*

En d'autres termes, l'hypothèse du théorème de la sous-section 7.3 est vérifiée au-dessus de $V \supset \Lambda_U$. Par conséquent, si $n_0 \geq 0$ est l'entier de loc. cit., pour tout point géométrique $\overline{\xi}$ localisé en un point fermé ξ de Λ_U et tout entier n tel que $\deg(\xi)n > n_0$, on a la formule de points fixes

$$\begin{aligned} \mathrm{Lef}_\xi(\Gamma_\varnothing \times \mathrm{Frob}_{\mathcal{X}_\varnothing}^{\deg(\xi)n}, \mathcal{X}_\varnothing) &= \mathrm{tr}(\mathrm{cl}(\Gamma) \times \mathrm{Frob}_{\mathcal{X}}^{\deg(\xi)n}, R\Gamma(\mathcal{X}_{\overline{\xi}}, \mathbb{Q}_\ell)) \\ &\quad + \sum_{I \neq \varnothing} (-1)^{|I|} \mathrm{tr}(\mathrm{cl}(\Gamma)_I^{(n_0)} \times \mathrm{Frob}_{\mathcal{Y}_I}^{\deg(\xi)n}, R\Gamma(\mathcal{Y}_{I,\overline{\xi}}, \mathbb{Q}_\ell)). \end{aligned} \tag{4}$$

On remarquera que, pour chaque $I \subset \{1, \ldots, m\}$ non vide, il existe une famille $(m_I(L))_L$ de scalaires dans $\overline{\mathbb{Q}}_\ell$, indexée par un ensemble de représentants des systèmes locaux ℓ-adiques irréductibles sur $(X - S_g)^2$, à support fini, telle que

$$\mathrm{tr}(\mathrm{cl}(\Gamma)_I^{(n_0)} \times \mathrm{Frob}_{\mathcal{Y}_I}^{\deg(\xi)n}, R\Gamma(\mathcal{Y}_{I,\overline{\xi}}, \mathbb{Q}_\ell)) = \sum_L m_I(L)\,\mathrm{Tr}(\mathrm{Frob}_\xi^n, L),$$

et que $m_I(L) = 0$ pour tout L qui n'est pas r-négligeable d'après 6.3.

Dans la suite de cette sous-section, on identifie $\overline{\mathbb{Q}}_\ell$ à $\mathbb{C}$ par l'isomorphisme fixé en 1.3.

Dans la formule de points fixes ci-dessus, on peut remplacer $\overline{\xi}$ par $(\mathrm{Frob}_X^k \times \mathrm{Id}_X)(\overline{\xi})$ pour $k = 1, \ldots, r!$ et faire la moyenne des $r!$ expressions obtenues. Le premier membre de cette moyenne est égal à

$$\sum_{\substack{\pi \in \mathcal{A}_r(K_N) \\ \omega_\pi(a)=1}} \mathrm{Tr}_\pi(1_{K_N g K_N}) q^{(r-1)\deg(\xi) n} S_\infty^{\left(-\frac{\deg(\xi)}{\deg(\infty)} n\right)}(\pi) S_o^{\left(\frac{\deg(\xi)}{\deg(o)} n\right)}(\pi)$$

$$(5) \qquad + \sum_{\substack{1 \leq r' < r \\ 1 \leq r'' < r}} \sum_{\substack{\pi' \in \mathcal{A}_{r'}(K_N) \\ \pi'' \in \mathcal{A}_{r''}(K_N)}} \mathrm{Tr}_{\pi',\pi''}^{\leq_\alpha p}(1_{K_N g K_N}, \deg(\xi) n) S_\infty^{\left(-\frac{\deg(\xi)}{\deg(\infty)} n\right)}(\pi') S_o^{\left(\frac{\deg(\xi)}{\deg(o)} n\right)}(\pi'')$$

d'après 5.2. Compte tenu de la proposition de la sous-section précédente et de l'hypothèse de récurrence $\mathrm{P}_1(r')$ pour tout $r' < r$ (cf. 6.2), on a donc une expression pour la différence

$$\mathrm{Tr}(1_{K_N g K_N} \times \mathrm{Frob}_\xi^n, H_{N,\mathrm{cusp}}^r) \\ - q^{(r-1)\deg(\xi) n} \sum_{\substack{\pi \in \mathcal{A}_r(K_N) \\ \omega_\pi(a)=1}} \mathrm{Tr}_\pi(1_{K_N g K_N}) S_\infty^{\left(-\frac{\deg(\xi)}{\deg(\infty)} n\right)}(\pi) S_o^{\left(\frac{\deg(\xi)}{\deg(o)} n\right)}(\pi)$$

de la forme

$$\sum_L P_L(n \deg(\xi)) \,\mathrm{Tr}(\mathrm{Frob}_\xi^n, L)$$

où $(\nu \mapsto P_L(\nu))_L$ est une famille à support fini de fonctions complexes de l'entier ν de la forme

$$P_L(\nu) = \sum_z P_{L,z}(\nu) z^\nu$$

pour une famille à support fini de polynômes $P_{L,z}(u) \in \mathbb{C}[u]$, $z \in \mathbb{C}^\times$, qui sont identiquement nuls quand L n'est pas r-négligeable.

En développant les polynômes $P_{L,z}(u)$ et en utilisant l'indépendance linéaire des fonctions $\nu \mapsto \nu^k z^\nu$ de l'entiers ν pour $k \in \mathbb{Z}$ et $z \in \mathbb{C}^\times$, on voit tout d'abord que l'on ne perd rien en replaçant chaque $P_{L,z}(u)$ par son terme constant $P_{L,z}(0)$.

En utilisant la correspondance de Langlands et la conjecture de Ramanujan-Petersson déjà connues en les rangs $< r$ et un dernier argument de fonctions L, Lafforgue conclut alors la preuve de la proposition de 8.1 et la récurrence.

RÉFÉRENCES

Travaux de L. Lafforgue

[La 1] L. Lafforgue – *Chtoucas de Drinfeld et conjecture de Ramanujan-Petersson*, Astérisque, 243 (1997).

[La 2] L. Lafforgue – Sur la conjecture de Ramanujan-Petersson pour les corps de fonctions. I : Étude géométrique, *C. R. Acad. Sci. Paris* 322 (1996), 605-608.

[La 3] L. Lafforgue – Sur la conjecture de Ramanujan-Petersson pour les corps de fonctions. II : Étude spectrale, *C. R. Acad. Sci. Paris* 322 (1996), 707-710.

[La 4] L. Lafforgue – Sur la dégénérescence des chtoucas de Drinfeld, *C. R. Acad. Sci. Paris* 323 (1996), 491-494.

[La 5] L. Lafforgue – Compactification de l'isogénie de Lang et dégénérescence des structures de niveau simple des chtoucas de Drinfeld, *C. R. Acad. Sci. Paris* 325 (1997), 1309-1312.

[La 6] L. Lafforgue – Chtoucas de Drinfeld et applications, dans *Proceedings of the International Congress of Mathematicians, Vol. II* (*Berlin, 1998*), Doc. Math., (1998), 563-570.

[La 7] L. Lafforgue – Une compactification des champs classifiant les chtoucas de Drinfeld, *J. Amer. Math. Soc.* 11 (1998), 1001-1036.

[La 8] L. Lafforgue – Pavages des simplexes, schémas de graphes recollés et compactification des $\mathrm{PGL}_r^{n+1} / \mathrm{PGL}_r$, *Invent. Math.* 136 (1999), 233-271.

[La 9] L. Lafforgue – La correspondance de Langlands sur les corps de fonctions, manuscrit final, (2000).

Articles de V.G. Drinfeld sur les chtoucas

[Dr 1] V. G. Drinfeld – Commutative subrings of certain noncommutative rings, *Funct. Anal. and its Appl.* 11 (1977), 9-12.

[Dr 2] V. G. Drinfeld – Proof of the global Langlands conjecture for GL(2) over a function field, *Funct. Anal. and its Appl.* 11 (1977), 223-225.

[Dr 3] V. G. Drinfeld – A proof of Petersson's conjecture for function fields, *Uspehi Mat. Nauk* 32 (1977), 209-210.

[Dr 4] V. G. Drinfeld – Langlands' conjecture for GL(2) over functional fields, dans *Proceedings of the International Congress of Mathematicians* (*Helsinki, 1978*), Acad. Sci. Fennica, (1980), 565-574.

[Dr 5] V. G. Drinfeld – Moduli varieties of F-sheaves, *Funct. Anal. and its Appl.* 21 (1987), 107-122.

[Dr 6] V. G. Drinfeld – Proof of the Petersson conjecture for GL(2) over a global field of characteristic p, *Funct. Anal. and its Appl.* 22 (1988), 28-43.

[Dr 7] V. G. Drinfeld – Cohomology of compactified moduli varieties of F-sheaves of rank 2, *J. of Soviet Math.* 46 (1989), 1789-1821.

[Dr 8] V. G. Drinfeld – On my paper : "Cohomology of compactified moduli varieties of F-sheaves of rank 2", *Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov.* (*LOMI*)*, Anal. Teor. Chisel i Teor. Funktsii. 9* 168 (1988), 45–47.

Autres publications sur la correspondance de Langlands

[Ca] H. Carayol – Variétés de Drinfeld compactes, d'après Laumon, Rapoport et Stuhler, dans *Séminaire Bourbaki* 1991/92, Astérisque 206, (1992), 369-409.

[D-H] P. Deligne, D. Husemoller – Survey of Drinfeld modules, dans *Current trends in arithmetical algebraic geometry* (*Arcata, 1985*), Contemp. Math. 67, (1987), 25-91.

[Dr 9] V.G. Drinfeld – Elliptic modules, *Math. USSR Sbornik* 23 (1974), 561-592.

[Dr 10] V.G. Drinfeld – Elliptic modules II, *Math. USSR Sbornik* 31 (1977), 159-170.

[Dr 11] V. G. Drinfeld – Number of two-dimensional irreducible representations of the fundamental group of a curve over a finite field, *Functional Anal. and its Appl.* 15 (1982), 294-295.

[F-K] Y. Flicker, D. Kazhdan – Geometric Ramanujan conjecture and Drinfeld reciprocity law, dans *Number theory, trace formulas and discrete groups*, Academic Press, (1989), 201-218.

[H-K] G. Harder, D. Kazhdan – Automorphic forms on GL_2 over function fields (after V. G. Drinfeld), *Proc. Sym. Pure Math.* 33, Part 2 (1979), 357-379.

[Ka] D. Kazhdan – An introduction to Drinfeld's "shtuka", *Proc. Sym. Pure Math.* 33, Part 2 (1979), 347-356.

[Lau 1] G. Laumon – *Cohomology of Drinfeld modular varieties, Part I* (*Geometry, counting of points and local harmonic analysis*), Cambridge University Press (1995).

[Lau 2] G. Laumon – *Cohomology of Drinfeld modular varieties, Part II* (*Automorphic forms, trace formulas and Langlands correspondence*), Cambridge University Press (1996).

[Lau 3] G. Laumon – Drinfeld Shtukas, dans *CIME Session "Vector bundles on Curves. New Directions"* (*Cetraro, juin 1995*), Lecture Notes in Mathematics 1649, (1996), 50-109.

[Lau 4] G. Laumon – The Langlands Correspondence for Function Fields following Laurent Lafforgue, dans *Current Developments in Mathematics Conference* (*Harvard University, novembre 1999*).

[L-R-S] G. Laumon, M. Rapoport et U. Stuhler – D-elliptic sheaves and the Langlands correspondence, *Inv. Math.* 113 (1993), 217-338.

Sur la correspondance de Langlands géométrique

[Dr 12] V. G. Drinfeld – Two-dimensional ℓ-adic representations of the fundamental group of a curve over a finite field and automorphic forms on GL(2), *Amer. J. Math.* 105 (1983), 85-114.

[Dr 13] V. G. Drinfeld – Two-dimensional ℓ-adic representations of the Galois group of a global field of characteristic p and automorphic forms on GL(2), *J. of Soviet Math.* 36 (1987), 93-105.

[F-G-K-V] E. Frenkel, D. Gaitsgory, D. Kazhdan, K. Vilonen – Geometric realization of Whittaker functions and the Langlands conjecture, *J. Amer. Math. Soc.* 11 (1998), 451-484.

[Lau 5] G. Laumon – Correspondance de Langlands géométrique pour les corps de fonctions, *Duke Math. J.* 54 (1987), 309-359.

[Lau 6] G. Laumon – Faisceaux automorphes pour GL_n : la première construction de Drinfeld, prépublication électronique, réf. alg-geom/9511004.

Autres références de géométrie algébrique

[De] P. Deligne – La conjecture de Weil. II, *Publ. Math. IHES* 52 (1980), 137-252.

[DC-P] C. De Concini, C. Procesi – Complete symmetric varieties, dans *CIME Session "Invariant Theory"* (*Montecatini*), Lecture Notes in Math. 996, Springer-Verlag, (1982), 1-44.

[Fa] G. Faltings – Explicit resolution of local singularities of moduli-spaces, *Journal für die reine und angewandte Mathematik* 483 (1997), 183-196.

[Gr] A. Grothendieck – Formule de Lefschetz et rationalité des fonctions L, Sém. Bourbaki 1964/65, exp. n° 279, collection hors série Astérisque **9** (1995), 41-55, et dans *Dix exposés sur la cohomologie des schémas*, North Holland, (1968), 31-45.

[H-N] G. Harder, M.S. Narasimhan – On the cohomology groups of moduli spaces of vector bundles on curves, *Math. Ann.* 212 (1975), 215-248.

[H-T] M. Harris, R. Taylor – On the geometry and cohomology of some simple Shimura varieties, Institut de Mathématiques de Jussieu, Prépublication 227, (1999).

[Lak] D. Laksov – Completed quadrics and linear maps, dans *Algebraic geometry, Bowdoin, 1985* (*Brunswick, Maine, 1985*), Proc. Sympos. Pure Math., 46, Part 2, Amer. Math. Soc., (1987), 371-387.

[Lau 7] G. Laumon – Transformation de Fourier, constantes d'équations fonctionnelles et conjecture de Weil, *Publ. Math. I.H.E.S.* 65 (1987), 131-210.

[L-M] G. Laumon, L. Moret-Bailly – *Champs algébriques*, Springer-Verlag (1999).

[Pi] R. Pink – On the calculation of local terms in the Lefschetz-Verdier trace formula and its application to a conjecture of Deligne, *Ann. Math.* 135 (1992), 483-525.

[SD] B. Saint-Donat – Affine Embeddings, dans *Toroidal Embeddings I*, Lecture Notes in Math. 339, Springer-Verlag, (1973), 1-40.

[Se] J.-P. Serre – *Abelian ℓ-Adic Representations and Elliptic Curves*, Addison-Wesley (1988).

Autres références automorphes

[Ar] J. Arthur – A trace formula for reductive groups I : terms associated to classes in $G(\mathbb{Q})$, *Duke Math. J.* 45 (1978), 911-953.

[C-PS] J. W. Cogdell, I. I. Piatetski-Shapiro – Converse theorems for GL_n, *Publ. Math. IHES* 79 (1994), 157-214.

[G-J] R. Godement, H. Jacquet – Zeta functions of simple algebras, Lecture Notes in Math. 260, Springer-Verlag, (1972).

[J-PS-S] H. Jacquet, I. I Piatetski-Shapiro, J.A. Shalika – Rankin-Selberg convolutions, *Amer. J. Math.* 105 (1983), 367-464.

[J-S 1] H. Jacquet, J.A. Shalika – On Euler products and the classification of automorphic representations I, *Amer. J. Math.* 103 (1981), 499-558.

[J-S 2] H. Jacquet, J.A. Shalika – On Euler products and the classification of automorphic representations II, *Amer. J. Math.* 103 (1981), 777-815.

[Lan 1] R.P. Langlands – Problems in the theory of automorphic forms, dans *Lectures in modern analysis and applications III*, Lecture Notes in Math. 170, Springer-Verlag, (1970), 18-61.

[Lan 2] R.P. Langlands – *On the functional equations satisfied by Eisenstein series*, Lecture Notes in Math. 544, Springer-Verlag (1976).

[M-W] C. Moeglin, J.-L. Waldspurger – *Décomposition spectrale et séries d'Eisenstein, Une paraphrase de l'Écriture*, Birkhäuser (1994).

[PS 1] I.I. Piatetski-Shapiro – Zeta functions of $\mathrm{GL}(n)$, prépublication de l'université du Maryland, (1976).

[PS 2] I.I. Piatetski-Shapiro – Multiplicity one theorem, *Proc. Sym. Pure Math.* 33, Part 1 (1979), 209-212.

[Sh] J.A. Shalika – The multiplicity one theorem for GL_n, *Ann. of Math.* 100 (1974), 171-193.

Gérard LAUMON

Université de Paris-Sud
et CNRS, UMR 8628
Mathématiques, Bât. 425
F-91405 Orsay Cedex (France)
E-mail : `Gerard.Laumon@math.u-psud.fr`

Séminaire BOURBAKI Mars 2000
52^{e} année, 1999-2000, n^{o} 874, p. 267 à 297

MOTIVIC MEASURES

by **Eduard LOOIJENGA**

1. INTRODUCTION

An n-jet of an arc in an algebraic variety is a one parameter Taylor series of length n in that variety. To be precise, if the variety X is defined over the algebraically closed field k, then it is a $k[[t]]/(t^{n+1})$-valued point of X. The set of such n-jets are the closed points of a variety $\mathcal{L}_n(X)$ also defined over k and the arc space of X, $\mathcal{L}(X)$, is the projective limit of these. Probably Nash [24] was the first to study arc spaces in a systematic fashion (the paper in question was written in 1968). He concentrated on arcs based at a given point of X and observed that to each irreducible component of this 'provariety' there corresponds in an injective manner an irreducible component of the preimage of this point in any resolution of X. He asked the (still unanswered) question how to identify these components on a given resolution. The renewed interest in arc spaces has a different origin, however. Batyrev [4] proved that two connected projective complex manifolds with trivial canonical bundle which are birationally equivalent must have the same Betti numbers. This he showed by first lifting the data to a situation over a discrete valution ring with finite residue field and then exploiting a p-adic integration technique. (Such a p-adic integration approach to problems in complex algebraic geometry had also been used by Denef and Loeser [12] in their work on topological zeta functions attached to singular points of complex varieties.) When Kontsevich learned of Batyrev's result he saw how this proof could be made to work in a complex setting using arc spaces. The new proof also gave more: equality of Hodge numbers, and even an isomorphism of Hodge structures with rational coefficients. The underlying technique, now going under the name of *motivic integration*, has led to an avalanche of applications. These include new (so-called *stringy*) invariants of singularities, a complex analogue of the Igusa zeta function, a motivic version of the Thom-Sebastiani property and the motivic

McKay correspondence. Some of these were covered in a recent talk by Reid [26] in this seminar.

The idea is simple if we keep in mind an analogous, more classical situation. Consider the case of a complete discrete valuation ring (R, m) with finite residue field F. There is a Haar measure on the Boolean algebra consisting of the cosets of powers of m that takes the value 1 on R (so it is also a probability measure). This induces one on a suitable Boolean algebra of subsets of the set of R-valued points of any scheme that is flat of pure dimension and of finite type over $\operatorname{Spec}(R)$. Associated to this measure is a function that essentially counts the number of 'points' in each reduction modulo m^k: the *Igusa zeta function*, introduced by Weil, and intensively studied by Igusa, Denef and Loeser (and reported on by Denef in this seminar [11]). A missing case was that of equal characteristic zero: $\mathcal{O} = k[[t]]$, $k \supset \mathbb{Q}$. The proposal of Kontsevich is to give $\mathcal{O}$ a measure that takes values in a Grothendieck ring of k-varieties in which the class of the affine line, $\mathbb{L}$, is invertible: the value on the ideal (t^n) is then simply $\mathbb{L}^{-n}$ (or $\mathbb{L}^{1-n}$, which is sometimes more convenient). If $\mathcal{X}$ is a suitable $\mathcal{O}$-scheme, then we obtain a measure on the set of sections as before, but now with values in this Grothendieck ring. The corresponding zeta function is a very fine bookkeeping device, for it does its counting in a ring that is huge. There is no a priori reason to restrict to the case of equal characteristic, for Kontsevich's idea makes sense for any complete discrete valuation ring. Indeed, with little extra effort the material in Sections 2, 3 and 9 can be generalized to that context.

This report concerns mainly work of Denef and Loeser. Some of their results are presented here somewhat differently, and this is why more proofs are provided than one perhaps expects of the write up of a seminar talk. References to the sources are in general given after the section titles, rather than in the statements of theorems.

I thank Jan Denef for inviting me for a short visit to Leuven to discuss the material exposed here. I am also indebted to Maxim Kontsevich and especially to Jan Denef for comments on previous versions, from which this text has greatly benefitted (though remaining errors are my responsability only). This applies in particular to the motivic Thom-Sebastiani theorem and a word of explanation is in order here. In the original version I had introduced (albeit somewhat implicitly) a binary operator on a certain Grothendieck ring of motives, called here quasi-convolution. Quasi-convolution is almost associative, but not quite, and since I thought this to be a serious defect, I passed to the universal associative quotient. But in a recent overview, Denef and Loeser [19] noted that there is no need for this: the property one wants (which is another than associativity) holds already without passing that quotient. As this no longer justifies its introduction, I thought it best to take advantage of their observation and rewrite things accordingly.

2. THE ARC SPACE AND ITS MEASURE [14], [23]

Throughout the talk we fix a complete discrete valuation ring $\mathcal{O}$ whose residue field k is assumed to be algebraically closed and of characteristic zero. The spectrum of $\mathcal{O}$ is denoted $\mathbb{D}$ with generic point $\mathbb{D}^{\times}$ and closed point o. A uniformizing parameter is often denoted by t so that $\mathcal{O} = k[[t]]$. The assumption that k be algebraically closed is for convenience only: in most situations this restriction is unnecessary or can be avoided.

The symbol $\mathbb{N}$ stands for the set of nonnegative integers.

The Grothendieck ring of varieties

Consider the *Grothendieck ring* $K_0(\mathcal{V}_k)$ of reduced k-varieties: this is the abelian group generated by the isomorphism classes of such varieties, subject to the relations $[X - Y] = [X] - [Y]$, where Y is closed in X. The product over k turns it into a ring. Note that if we restrict ourselves to smooth varieties we get the same ring: the reason is that every k-variety X admits a stratification (i.e., a filtration by closed subschemes $X = X^0 \supset X^1 \supset \cdots \supset X^{d+1} = \varnothing$ such that $X^k - X^{k+1}$ is smooth) and that any two such admit a common refinement. The latter property implies that $[X] := \sum_k [X^k - X^{k+1}]$ is unambiguously defined. In fact, $K_0(\mathcal{V}_k)$ is generated by the classes of complete nonsingular varieties, for any smooth variety U admits a completion $\overline{U}$ by adding a normal crossing divisor and then $[U] = \sum (-1)^i [\overline{U}^i]$, where $\overline{U}^i$ stands for the normalization of the codimension i skeleton of the resulting stratification. Włodarczyk's weak factorization theorem (in the form of the main theorem of [1]) can be used to show that relations of the following simple type suffice: if X is smooth projective and $\tilde{X} \to X$ is obtained by blowing up a smooth closed subvariety $Y \subset X$ with exceptional divisor $\tilde{Y}$, then $[\tilde{X}] - [\tilde{Y}] = [X] - [Y]$.

We denote the class of the affine line $\mathbb{A}^1$ by $\mathbb{L}$ and we write M_k for the localization $K_0(\mathcal{V}_k)[\mathbb{L}^{-1}]$. Recall that a subset of a variety X is called *constructible* if it is a finite union of (locally closed) subvarieties. Any constructible subset C of X defines an element $[C] \in M_k$. The constructible subsets of X form a Boolean algebra and so we obtain in a tautological manner a M_k-valued measure μ_X defined on this Boolean algebra. More generally, a morphism $f : Y \to X$ defines on that same algebra an M_k-valued measure $f_*\mu_Y$: assign to a constructible subset of X its preimage in Y.

The ring M_k is interesting, big, and hard to grasp. Fortunately, there are several characteristics of M_k (i.e., ring homomorphisms from M_k to a ring) that are well understood. We describe some of these in decreasing order of complexity under the assumption that k is a subfield of $\mathbb{C}$. The first example is the Grothendieck ring $K_0(\mathrm{HS})$ of the category of Hodge structures. A *Hodge structure* consists of a finite dimensional $\mathbb{Q}$-vector space H, a finite bigrading $H \otimes \mathbb{C} = \oplus_{p,q \in \mathbb{Z}} H^{p,q}$ such that $H^{p,q}$ is the complex conjugate of $H^{q,p}$ and each *weight summand*, $\oplus_{p+q=m} H^{p,q}$, is defined over $\mathbb{Q}$. There are evident notions of tensor product and morphism of Hodge structures so that we get an abelian category HS with tensor product. The Grothendieck construction produces a group $K_0(\mathrm{HS})$, elements of which are representable as a formal difference of Hodge structures $[H]-[H']$ and $[H]=[H']$ if and only if H and H' are isomorphic. The tensor product makes it a ring.

For every complex variety X, the cohomology with compact supports, $H^r_c(X;\mathbb{Q})$, comes with a natural finite increasing filtration $W_\bullet H^r_c(X;\mathbb{Q})$, *the weight filtration*, such that the associated graded $\mathrm{Gr}^W_\bullet H^r_c(X;\mathbb{Q})$ underlies a Hodge structure having $\mathrm{Gr}^W_m H^r_c(X;\mathbb{Q})$ as weight m summand. We assign to X the *Hodge characteristic*(1)

$$\chi_{\mathrm{h}}(X) := \sum_r (-1)^r [H^r_c(X;\mathbb{Q})] \in K_0(\mathrm{HS})$$

If $Y\subset X$ is closed subvariety, then the exact sequence

$$\cdots \to H^r_c(X-Y)\to H^r_c(X)\to H^r_c(Y)\to H^{r+1}_c(X-Y)\to \dots$$

is compatible in a strong sense with the Hodge data. This implies the additivity property $\chi_{\mathrm{h}}(X)=\chi_{\mathrm{h}}(X-Y)+\chi_{\mathrm{h}}(Y)$. For the affine line $\mathbb{A}^1$, $H^r_c(\mathbb{A}^1;\mathbb{Q})$ is nonzero only for $r=2$; the cohomology group $H^2_c(\mathbb{A}^1;\mathbb{Q})$ is one-dimensional and of type $(1,1)$. So $\chi_{\mathrm{h}}(\mathbb{A}^1)$ (usually denoted as $\mathbb{Q}(-1)$) is invertible. It follows that χ_{h} factorizes over M_k. If we only care for dimensions, then we compose with the ring homomorphism $K_0(\mathrm{HS})\to\mathbb{Z}[u,u^{-1},v,v^{-1}]$, $[H]\mapsto\sum_{p,q}\dim(H^{p,q})u^pv^q$, to get the *Hodge number characteristic* $\chi_{\mathrm{hn}}: M_k\to\mathbb{Z}[u,u^{-1},v,v^{-1}]$. It takes $\mathbb{L}$ to uv. The *weight characteristic* $\chi_{\mathrm{wt}}: M_k\to\mathbb{Z}[w,w^{-1}]$ is obtained if we go further down along the map $\mathbb{Z}[u,u^{-1},v,v^{-1}]\to\mathbb{Z}[w,w^{-1}]$ that sends both u and v to w. Evaluating the latter at $w=1$ gives the ordinary(2) Euler characteristic $\chi_{\mathrm{top}}: M_k\to\mathbb{Z}$.

In the spirit of this discussion is the following question raised by Kapranov [22]:

Question 2.1. — Let X be a variety over k. If $\sigma_n(X)\in M_k$ denotes the class of its nth symmetric power, is then

$$Z_X(T) := 1+\sum_{n=1}^{\infty}\sigma_n(X)T^n \in M_k[[T]]$$

a rational function in the sense that it determines an element in a suitable localization of $M_k[T]$? (Since the logarithmic derivative Z'/Z defines an additive map $M_k\to M_k[[T]]$, we may restrict ourselves here to the case of a smooth variety.) Does it

(1) As all our characteristics are compactly supported we omit the otherwise desirable subscript c from the notation.

(2) A complex algebraic variety can be compactified within its homotopy type by giving it a topological boundary that is stratifyable into strata of odd dimension. This boundary has zero Euler characteristic, hence the compactly supported Euler characteristic of the variety is its ordinary Euler characteristic.

satisfy a functional equation when X is smooth and complete? Kapranov shows that the answer to both questions is yes in case $\dim(X) \leq 1$.

A measure on the space of sections

Let us call a $\mathbb{D}$-*variety* a separated reduced scheme that is flat and of finite type over $\mathbb{D}$ and whose closed fiber is reduced. Given a $\mathbb{D}$-variety $\mathcal{X}/\mathbb{D}$ with closed fiber X, then the set of its sections up to order n, $\mathcal{X}_n$, is the set of closed points of a k-variety (also denoted $\mathcal{X}_n$) naturally associated to $\mathcal{X}$. It is obtained from $\mathcal{X}$ modulo $\mathfrak{m}^{n+1}$ essentially by Weil restriction of scalars [20]. So $\mathcal{X}_0 = X$. The set $\mathcal{X}_\infty$ of sections of $\mathcal{X} \to \mathbb{D}$ is the projective limit of these and is therefore the set of closed points of a provariety. If $\mathcal{X}/\mathbb{D}$ is of the form $X \times \mathbb{D} \to \mathbb{D}$, with X a k-variety, then we are dealing with the space of *n-jets* (of curves) on X and the *arc space* of X, here denoted by $\mathcal{L}_n(X)$ resp. $\mathcal{L}(X)$.

For $m \geq n$ we have a forgetful morphism $\pi_n^m : \mathcal{X}_m \to \mathcal{X}_n$. (When $n = 0$, we shall often write π_X^m, π_X instead of π_0^m, π_0.) A fiber of π_n^{n+1} lies in an affine space over the Zariski tangent space of the base point. In case X is smooth, it is in fact an affine space over the tangent space of the base point: π_n^{n+1} has then the structure of a torsor over the tangent bundle. A theorem of Greenberg [21] asserts that there exists a constant c such that the image of π_n equals the image of π_n^{cn}. So $\pi_n(\mathcal{X}_\infty)$ is constructible.

The goal is to define a measure on an interesting algebra of subsets of $\mathcal{X}_\infty$ in such a way that its direct image under π_X is the tautological measure μ_X when X is smooth. (This will lead us to deviate from the definition of Denef-Loeser and Batyrev by a factor $\mathbb{L}^d$ and to adopt the one used in [26] instead.) For this we assume that $\mathcal{X}$ is of pure relative dimension d and we say that a subset A of $\mathcal{X}_\infty$ is *stable* if for some $n \in \mathbb{N}$ we have

– $\pi_n(A)$ is constructible in $\mathcal{X}_n$ and $A = \pi_n^{-1}\pi_n(A)$,

– for all $m \geq n$ the projection $\pi_{m+1}(A) \to \pi_m(A)$ is a piecewise trivial fibration (that is, trivial relative to a decomposition into subvarieties) with fiber an affine space of dimension d.

The second condition is of course superfluous in case $\mathcal{X}/\mathbb{D}$ is smooth. It is clear that $\dim \pi_m(A) - md$ is independent of the choice of $m \geq n$; we call this the *(virtual) dimension* $\dim A$ of A. The same is true for the class $[\pi_m(A)]\mathbb{L}^{-md} \in M_k$; we denote that class by $\tilde{\mu}_{\mathcal{X}}(A)$. The collection of stable subsets of $\mathcal{X}$ is a Boolean ring (i.e., is closed under finite union and difference) on which $\tilde{\mu}_{\mathcal{X}}$ defines a finite additive measure. A theorem of Denef-Loeser (see Theorem 9.1) ensures that there are plenty of stable sets.

In order to extend the measure to a bigger collection of interesting subsets of $\mathcal{X}_\infty$ we need to complete M_k. Given $m \in \mathbb{Z}$, let $F_m M_k$ be the subgroup of M_k spanned by the $[Z]\mathbb{L}^{-r}$ with $\dim Z \leq m + r$. This is a filtration of M_k as a ring:

$F_mM_k.F_nM_k \subset F_{m+n}M_k$. So the separated completion of M_k with respect to this filtration,

$$\hat{M}_k := \varprojlim M_k/F_mM_k \quad (m \to -\infty \text{ in this limit}),$$

to which we will refer as the *dimensional completion*, is also a ring. The kernel of the natural map $M_k \to \hat{M}_k$ is $\cap_m F_mM_k$, of course. It is not known whether this is zero[(3)]. In case $k \subset \mathbb{C}$, the Hodge characteristic extends to this completion:

$$\chi_{\mathrm{h}} : \hat{M}_k \to \hat{K}_0(\mathrm{HS}).$$

Here $\hat{K}_0(\mathrm{HS})$ is defined in a similar way as $\hat{M}_k$ with 'dimension' replaced by 'weight'. The assertion follows from the fact that the weights in the compactly supported cohomology of a variety of dimension d are $\leq 2d$. Likewise we can extend the characteristics counting Hodge numbers or weight numbers (with values Laurent power series in the reciprocals of their variables). This does not apply to the Euler characteristic, but in many cases of interest the weight characteristic gives a rational function in w that has no pole at $w = 1$. Its value there is then a good substitute.

We will be mostly concerned with the composite of $\tilde{\mu}_{\mathcal{X}}$ and the completion map, for it is this measure that we shall extend. We call this the *motivic measure* on $\mathcal{X}$ and denote it by $\mu_{\mathcal{X}}$. Let us say that a subset $A \subset \mathcal{X}_\infty$ is *measurable* if for every (negative) integer m there exist a stable subset $A_m \subset \mathcal{X}_\infty$ and a sequence $(C_i \subset \mathcal{X}_\infty)_{i=0}^\infty$ of stable subsets such that the symmetric difference $A\Delta A_m$ is contained in $\cup_{i\in\mathbb{N}} C_i$ with $\dim C_i < m$ for all i and $\dim C_i \to -\infty$, for $i \to \infty$.

PROPOSITION 2.2. — *The measurable subsets of $\mathcal{X}_\infty$ make up a Boolean subring and $\mu_{\mathcal{X}}$ extends as a measure to this ring by*

$$\mu_{\mathcal{X}}(A) := \lim_{m\to-\infty} \mu_{\mathcal{X}}(A_m).$$

In particular, the above limit exists in $\hat{M}_k$ and its value only depends on A.

The proof is based on

LEMMA 2.3. — *Let $\mathcal{X}/\mathbb{D}$ be of pure dimension and $A \subset \mathcal{X}_\infty$ a stable subset. If $\mathcal{C} = \{C_i\}_{i=1}^\infty$ is a countable covering of A by stable subsets with $\dim C_i \to -\infty$ as $i \to \infty$, then A is covered by a finite subcollection of $\mathcal{C}$.*

Proof. — Let $n \in \mathbb{N}$ be such that $A = \pi_n^{-1}\pi_n(A)$. Suppose that A is not covered by a finite subcollection of $\mathcal{C}$. Choose $k \in \mathbb{N}$ such that $\dim C_i < -(n+2)d$ for $i > k$ and let $u_{n+1} \in \pi_{n+1}(A \setminus \cup_{i\leq k} C_i)$. We have $\pi_{n+1}^{-1}u_{n+1} \subset A$. This set is not covered by a finite subcollection of $\mathcal{C}$, for clearly $\pi_{n+1}^{-1}(u_{n+1})$ is not covered by $\{C_i\}_{i\leq k}$ and for $i > k$, $C_i \cap \pi_{n+1}^{-1}(u)$ is of positive codimension in $\pi_{n+1}^{-1}(u)$.

(3) This issue is avoided if we work with the adic completion $\mathbb{Z}((L^{-1})) \otimes_{\mathbb{Z}[L]} K_0(\mathcal{V}_k)$ instead, but in practice this is too small. Nevertheless, it seems that in all applications we are dealing with elements lying in the localization $\mathbb{Q}(L) \otimes_{\mathbb{Z}[L]} K_0(\mathcal{V}_k)$.

With induction we find a sequence $\{u_m \in \mathcal{L}_m(X)\}_{m>n}$ so that for all $m > n$ u_{m+1} lies over u_m and $\pi^{-1}(u_m)$ is not covered by a finite subcollection of $\mathcal{C}$. The sequence defines an element $u \in \mathcal{X}$. Since $\pi_n(u) \in \pi_n(A)$, we have $u \in A$ and so $u \in C_i$ for some i. But if C_i is stable at level $m > n$, then $\pi_m^{-1}(u_m) \subset C_i$, which contradicts a defining property of u_m. □

For $k = \mathbb{C}$, the condition $\lim_{i\to\infty} \dim C_i = -\infty$ is unnecessary, for we may then use the Baire property of $\mathbb{C}$ instead [5].

Proof of 2.2. — Suppose we have another solution $A\Delta A'_m \subset \cup_{i\in\mathbb{N}} C'_i$ with A'_m and C'_i stable, $\dim(C'_i) < m$ for all i and $\dim C'_i \to -\infty$ as $i \to \infty$. It is enough to prove that the dimension of the stable set $A_m \Delta A'_m$ is $< m$. Since $A_m\Delta A'_m \subset \cup_{i\in\mathbb{N}}(C_i \cup C'_i)$, Lemma 2.3 applies and we find that $A_m\Delta A'_m \subset \cup_{i\le N}(C_i \cup C'_i)$ for some N. Since every term has dimension $< m$, this is also true for $A_m\Delta A'_m$. □

So a countable union of stable sets $A = \cup_{n\in\mathbb{N}} A_n$ with $\lim_{n\to\infty} \dim A_n = -\infty$ is measurable and $\mu_{\mathcal{X}}(A) = \lim_{n\to\infty} \mu_{\mathcal{X}}(\cup_{k\le n} A_k)$.

Remark 2.4. — Given a $\mathbb{D}$-variety $\mathcal{X}$, then for any $d \in \mathbb{N}$ there is a d-measure $\mu^d_{\mathcal{X}}$ that induces $\mu_{\mathcal{Y}}$ on $\mathcal{Y}_\infty$ for any $\mathbb{D}$-subvariety $\mathcal{Y}$ of pure dimension d. We expect this measure to extend to a much bigger collection of subsets of $\mathcal{X}$ so that if $f : \mathcal{X} \to \mathcal{S}$ is a dominant $\mathbb{D}$-morphism of pure relative dimension d, then every fiber of $f_* : \mathcal{X}_\infty \to \mathcal{S}_\infty$ is $\mu^d_{\mathcal{X}}$-measurable.

Here is a sample of the results of Denef and Loeser on the rationality of Poincaré series [14].

Theorem 2.5. — *Let X be a k-variety. Then $\sum_{n=0}^{\infty} \mu_{\mathcal{X}}(\pi_n\mathcal{L}(X))T^n \in M_k[[T]]$ is a rational expression in T with each factor in the denominator of the form $1 - \mathbb{L}^a T^b$ where $a \in \mathbb{Z}$ and b is a positive integer.*

We will not discuss its proof, since this theorem is not used in what follows. Denef and Loeser derive this by means of Kontsevich's transformation rule discussed below, which is applied to a suitable projective resolution $\mathcal{X}$, and a theorem about semialgebraic sets, due to Pas [25]. It is likely that this theorem still holds for the space of sections of any $\mathbb{D}$-variety.

3. THE TRANSFORMATION RULE [23], [14], [16]

We describe two results that are at the basis of the theory. The proofs are relegated to Section 9.

Proposition 3.1. — *For a $\mathbb{D}$-variety $\mathcal{X}/\mathbb{D}$ of pure dimension, the preimage of any constructible subset under $\pi_n : \mathcal{X}_\infty \to \mathcal{X}_n$ is measurable. In particular, $\mathcal{X}_\infty$ is measurable. If $\mathcal{Y} \subset \mathcal{X}$ is nowhere dense, then $\mathcal{Y}_\infty$ is of measure zero.*

For $\mathcal{X}/\mathbb{D}$ of pure relative dimension we have the notion of an integrable function $\Phi : \mathcal{X}_\infty \to \hat{M}_k$: this requires the fibers of Φ to be measurable and the sum $\sum_a \mu_{\mathcal{X}}(\Phi^{-1}(a))a$ to converge, i.e., there are at most countably many nonzero terms $(\mu_{\mathcal{X}}(\Phi^{-1}(a_i))a_i)_{i\in\mathbb{N}}$ and we have $\mu_{\mathcal{X}}(\Phi^{-1}(a_i))a_i \in F_{m_i}\hat{M}_k$ with $\lim_{i\to\infty} m_i = -\infty$. The motivic integral of Φ is then by definition the value of this series:

$$\int \Phi \, d\mu_{\mathcal{X}} = \sum_i \mu_{\mathcal{X}}(\Phi^{-1}(a_i))a_i.$$

We have a similar notion for maps with values in topological $\hat{M}_k$-modules. An important example arises from an ideal $\mathcal{I} \subset \mathcal{O}_X$: such an ideal defines a function $\mathrm{ord}_{\mathcal{I}} : \mathcal{X}_\infty \to \mathbb{N} \cup \{\infty\}$ by assigning to $\gamma \in \mathcal{X}_\infty$ the multiplicity of $\gamma^*\mathcal{I}$. The condition $\mathrm{ord}_{\mathcal{I}}\,\gamma = n$ only depends on the n-jet of γ and this defines a constructible subset $C_n \subset \mathcal{X}_n$. Hence the fibers of $\mathrm{ord}_{\mathcal{I}}$ are measurable. We shall see that the function

$$\mathbb{L}^{-\mathrm{ord}_{\mathcal{I}}} : \mathcal{X}_\infty \to \hat{M}_k$$

is integrable.

There is a beautiful transformation rule for motivic integrals under modifications. Let $H : \mathcal{Y} \to \mathcal{X}$ be a morphism of $\mathbb{D}$-varieties of pure dimension d. We define the *Jacobian ideal* $\mathcal{J}_H \subset \mathcal{O}_{\mathcal{Y}}$ of H as 0th Fitting ideal of $\Omega_{\mathcal{Y}/\mathcal{X}}$. This has the nice property that its formation commutes with base change. The following theorem generalizes an unpublished theorem of Kontsevich.

Theorem 3.2. — *Let $H : \mathcal{Y} \to \mathcal{X}$ be a $\mathbb{D}$-morphism of pure dimensional $\mathbb{D}$-varieties with $\mathcal{Y}/\mathbb{D}$ smooth. If A is a measurable subset of $\mathcal{Y}_\infty$ with $H\big|_A$ injective, then HA is measurable and $\mu_{\mathcal{X}}(HA) = \int_A \mathbb{L}^{-\mathrm{ord}_{\mathcal{J}_H}} \, d\mu_{\mathcal{Y}}$.*

4. THE BASIC FORMULA [14]

A relative Grothendieck ring

It is convenient to be able to work in a relative setting. Given a variety S, denote by $K_0(\mathcal{V}_S)$ the Grothendieck ring of S-varieties and by M_S its localization with respect $\mathbb{L}$. The ring M_S can be dimensionally completed as usual. Notice that an element of M_S defines a M_k-valued measure on the Boolean algebra of constructible subsets of S. Often measures are naturally represented this way. For instance, the preceding shows that for all $n \in \mathbb{N}$, the direct image of $\mu_{\mathcal{X}}$ on $\mathcal{X}_n$ is given by an element $\mu_{\mathcal{X},n} \in \hat{M}_{\mathcal{X}_n}$. (Notice that $\mu_{\mathcal{X},n}$ is then the direct image of $\mu_{\mathcal{X},n+1}$.)

A morphism $f : S' \to S$ induces a ring homomorphism $f^* : M_S \to M_{S'}$. This makes $M_{S'}$ a M_S-module. We also have a direct image $f_* : M_{S'} \to M_S$ that is a homomorphism of M_S-modules. Notice that f itself defines an element $[f] \in M_S$; this is also the image of $1 \in M_{S'}$ under f_*.

There are corresponding characteristics. For instance, the ordinary Euler characteristic χ_{top} becomes a ring homomorphism from M_S to the Grothendieck ring of constructible $\mathbb{Q}$-vector spaces on S. This ring is generated by direct images of irreducible local systems of $\mathbb{Q}$-vector spaces over smooth irreducible subvarieties Z of S. (A better choice is to take the intersection cohomology sheaf in S of this local system along Z; this has the advantage that it only depends on the generic point of Z.)

Similarly, the Hodge characteristic χ_{h} takes values in a ring $K_0(\mathrm{HS}_S)$ that is generated by variations of Hodge structures over a smooth subvariety of S. The homomorphisms f^* and f_* persist on this level: $f : S' \to S$ induces homomorphisms $f^* : K_0(\mathrm{HS}_S) \to K_0(\mathrm{HS}_{S'})$ and $f_* : K_0(\mathrm{HS}_{S'}) \to K_0(\mathrm{HS}_S)$.

The basic computation

A case of interest is when the base variety is $(\mathbb{N} \times \mathbb{G}_m)^r$. This fails to be finite type, but that is of no consequence and we identify $\hat{M}_{(\mathbb{N}\times\mathbb{G}_m)^r}$ with $\hat{M}_{\mathbb{G}_m^r}[[T_1, \ldots, T_r]]$ in the obvious way.

We use a uniformizing parameter of $\mathcal{O}$ to define

$$\mathrm{ac} : \mathcal{L}(\mathbb{A}^1) - \{0\} \to \mathbb{N} \times \mathbb{G}_m,$$

by assigning to γ its order $\mathrm{ord}(\gamma)$ resp. the first nonzero coefficient of γ (ac stands for *angular component*). Integration along ac sends a $\hat{M}_k$-valued measure on $\mathcal{L}(\mathbb{A}^1)$ to an element of $\hat{M}_{\mathbb{G}_m}[[T]]$. The prime example is when this measure is given by a regular function $f : \mathcal{X} \to \mathbb{A}^1$ on a $\mathbb{D}$-variety $\mathcal{X}$ of pure relative dimension: this induces a map $f_* : \mathcal{X}_\infty \to \mathcal{L}(\mathbb{A}^1)$ and we then define

$$\mathrm{ac}_f : \mathcal{X}_\infty \xrightarrow{\ f_*\ } \mathcal{L}(\mathbb{A}^1) \xrightarrow{\ \mathrm{ac}\ } \mathbb{N} \times \mathbb{G}_m,$$

so that $[\mathrm{ac}_f] \in \hat{M}_{\mathbb{G}_m}[[T]]$. More generally, given a morphism $f = (f_1, \ldots, f_r) : \mathcal{X} \to \mathbb{A}^r$, we abbreviate

$$\mathrm{ac}_{X,f} := (\pi_X, \mathrm{ac}_{f_1}, \ldots \mathrm{ac}_{f_r}) : \mathcal{X}_\infty \to X \times (\mathbb{N} \times \mathbb{G}_m)^r.$$

So $[\mathrm{ac}_{X,f}] \in \hat{M}_{X\times\mathbb{G}_m^r}[[T_1, \ldots, T_r]]$.

Conventions 4.1. — If E is a simple normal crossing hypersurface on a smooth k-variety Y, then we adhere to the following notation throughout the talk: $(E_i)_{i\in\mathrm{irr}(E)}$ denotes the collection of irreducible components of E (so these are all smooth by assumption) and for any subset $I \subset \mathrm{irr}(E)$, E_I° stands for the locus of $p \in \tilde{X}$ with $p \in E_i$ if and only if $i \in I$. (With this convention, $E_\varnothing^\circ = Y - E$.) We denote the complement of the zero section of the normal bundle of E_i by U_{E_i} (so this is a

$\mathbb{G}_m$-bundle over E_i) and U_I designates the fiber product of the bundles $U_{E_i}|E_I^\circ$, $i \in I$ (a $\mathbb{G}_m^I$-bundle whose total space has the same dimension as Y).

If $\mathcal{E}$ is a simple normal crossing hypersurface on a $\mathbb{D}$-variety $\mathcal{Y}/\mathbb{D}$ with $\mathcal{Y}$ smooth, then we shall always assume that its union with the closed fiber Y has also normal crossings. The notational conventions are as above to the extent that restriction or intersection with Y is indicated by switching from calligraphic to roman font (e.g., $E_i = \mathcal{E}_i \cap Y$). If Y is smooth, then we may identify $\mathrm{irr}(E)$ with a subset of $\mathrm{irr}(\mathcal{E})$. (An equality if $\mathcal{E}$ has no component in Y.)

The following proposition accounts for many of the rationality assertions in [14].

Proposition 4.2. — *Let $\mathcal{X}/\mathbb{D}$ be a $\mathbb{D}$-variety of pure relative dimension and $H : \mathcal{Y} \to \mathcal{X}$ a resolution of singularities. Let $\mathcal{E}$ be a simple normal crossing hypersurface on $\mathcal{Y}$ that has no irreducible component in the closed fiber Y. Assume that the Jacobian ideal $\mathcal{J}_H$ of H is principal and has divisor $\sum_i(\nu_i - 1)\mathcal{E}_i$ (so $\nu_i \geq 1$). Let for $\rho = 1, \dots, r$, $f_\rho : \mathcal{X} \to \mathbb{A}^1$ be a regular function such that $f_\rho H$ has zero divisor $\sum_i N_{i,\rho}\mathcal{E}_i$ and put $N_i := (N_{i,1}, \cdots N_{i,r}) \in \mathbb{N}^r$, $i \in \mathrm{irr}(E)$. Then*

$$[\mathrm{ac}_{X,f}] = \sum_{I \subset \mathrm{irr}(E)} [U_I/X \times \mathbb{G}_m^r] \prod_{i \in I} (\mathbb{L}^{\nu_i} T^{-N_i} - 1)^{-1} \text{ in } \hat{M}_{X \times \mathbb{G}_m^r}[[T_1, \dots, T_r]],$$

where $U_I \to X \times \mathbb{G}_m^r$ has first component projection onto $E_I^\circ \subset X$ followed by the restriction of H and second component induced by fH.

Proof. — Given $m \in \mathbb{N}^{\mathrm{irr}(E)}$, consider the set $\mathcal{Y}(m)$ of $\gamma \in \mathcal{Y}_\infty$ with order m_i along $\mathcal{E}_i$. So for $\gamma \in \mathcal{Y}(m)$ we have $\mathrm{ord}_{\mathcal{J}_H}(\gamma) = \sum_i m_i(\nu_i - 1)$ and $\mathrm{ord}_{f_\rho H}(\gamma) = \sum_i m_i N_{\rho,i}$. If $\mathrm{supp}(m) \subset \mathrm{irr}(E)$ is the support of m, then we have a natural projection $e_m : \mathcal{Y}(m) \to U_{\mathrm{supp}(m)}$. Its composite with the morphism $U_{\mathrm{supp}(m)} \to X \times \mathbb{G}_m^r$ is a restriction of $\mathrm{ac}_{X,fH} := (\pi_X H, \mathrm{ac}_{f_1 H}, \dots, \mathrm{ac}_{f_r H}) : \mathcal{Y}_\infty \to X \times (\mathbb{N} \times \mathbb{G}_m)^r$ with $\mathbb{N}^r$-component $\sum_i m_i N_i$. In other words,

$$[\mathrm{ac}_{X,fH}\,|_{\mathcal{Y}(m)}] = [U_{\mathrm{supp}(m)}/X \times \mathbb{G}_m^r]\mathbb{L}^{-\sum_i m_i} T^{\sum_i m_i N_i}.$$

So the transformation formula 3.2 yields

$$\begin{aligned}
[\mathrm{ac}_{X,f}] &= \sum_{m \in \mathbb{N}^{\mathrm{irr}(E)}} [U_{\mathrm{supp}(m)}/X \times \mathbb{G}_m^r] \prod_{i \in \mathrm{supp}(m)} \left(\mathbb{L}^{-m_i - m_i(\nu_i - 1)} T^{m_i N_i}\right) \\
&= \sum_{I \subset \mathrm{irr}(E)} [U_I/X \times \mathbb{G}_m^r] \prod_{i \in I} \frac{\mathbb{L}^{-\nu_i} T^{N_i}}{1 - \mathbb{L}^{-\nu_i} T^{N_i}}.
\end{aligned}$$

□

If we drop the assumption that $\mathcal{E}$ has no irreducible component in Y, then the above formula must be somewhat modified: now each irreducible component of Y contributes with an expression of the above form times a monomial in $\mathbb{L}^{-1}$ and $T_1, \dots, T_r$.

COROLLARY 4.3. — *In the situation of 4.2, the class of* $(\pi_X, \mathrm{ord}_f) : \mathcal{X}_\infty \to X \times \mathbb{N}^r$ *in* $\hat{M}_X[[T_1, \ldots, T_r]]$ *equals*

$$\sum_{I \subset \mathrm{irr}(E)} [E_I^\circ/X] \prod_{i \in I} \frac{\mathbb{L} - 1}{\mathbb{L}^{\nu_i} T^{-N_i} - 1}.$$

In particular, the direct image of $\mu_{\mathcal{X}}$ *on* X *is represented by*

$$\sum_{I \subset \mathrm{irr}(E)} [E_I^\circ/X] \prod_{i \in I} [\mathbb{P}^{\nu_i - 1}]^{-1}.$$

Proof. — Since U_I is a $\mathbb{G}_m^I$-bundle over E_I°, the class of the projection $U_I \to X$ is $(\mathbb{L}-1)^{|I|}$ times the class of $E_I^\circ \to X$. □

This corollary shows that $\mathcal{X}_\infty$ is measurable so that the measurable subsets of $\mathcal{X}_\infty$ form in fact a Boolean algebra. It also implies that the Hodge number characteristic of $\mathcal{X}_\infty$ is an element of $\mathbb{Q}[u, v][(uv)^N - 1)^{-1} \mid N = 1, 2, \ldots]$ on which the Euler characteristic takes the value $\sum_{I \subset \mathrm{irr}(E)} \chi_{\mathrm{top}}(E_I^\circ) \prod_{i \in I} \nu_i^{-1}$.

Remark 4.4. — We can also express the direct image of $\mu_{\mathcal{X}}$ on X in terms of the closed subvarieties E_I: if $\mathrm{irr}'(E)$ denotes the set of $i \in \mathrm{irr}(E)$ with $\nu_i \geq 2$, then

$$\sum_{I \subset \mathrm{irr}'(E)} (-\mathbb{L})^{|I|} [E_I/X] \prod_{i \in I} \frac{[\mathbb{P}^{\nu_i - 2}]}{[\mathbb{P}^{\nu_i - 1}]}.$$

All varieties appearing in this expression are proper over X and nonsingular. So it gives rise to an element of a complex cobordism ring of X localized away from the classes of the complex projective varieties. This class, and the values that various genera take on it, might deserve closer study.

5. THE MOTIVIC NEARBY FIBER [13], [18]

An equivariant Grothendieck ring

Let G be an affine algebraic group. We consider varieties X with *good* G-action, where 'good' means that every orbit is contained in an affine open subset. For instance, a representation of G on a k-vector space V is good. For a fixed variety S with G-action, we define the Grothendieck group $K_0^G(\mathcal{V}_S)$ as generated by isomorphism types of S-varieties with good G-action modulo the usual equivalence relation (defined by pairs) and the relation that declares that every finite dimensional representation ρ of G has the same class as the trivial representation of the same degree (i.e., $\mathbb{L}^{\deg(\rho)}$).

In case the action on S is trivial, the product makes $K_0^G(\mathcal{V}_S)$ a $K_0(\mathcal{V}_S)$-algebra. If moreover G is finite abelian, then assigning to a variety X with good G-action its G-orbit space $\overline{X} := G\backslash X$ augments this as a $K_0(\mathcal{V}_S)$-module:

$$K_0^G(\mathcal{V}_S) \to K_0(\mathcal{V}_S), \quad a \mapsto \bar{a}.$$

(Not as an algebra, for the orbit space of a product is in general not the product of orbit spaces.) That this is well-defined follows from the lemma below. (We do not know whether this holds for arbitrary finite G.)

LEMMA 5.1. — *Let be given a representation of a finite abelian group G on a k-vector space V of finite dimension n. Then the class of $\overline{V}$ in $K_0(\mathcal{V}_k)$ is $\mathbb{L}^n$.*

Proof. — Let $V = \oplus_{\chi\in\hat{G}} V_\chi$ be the eigenspace decomposition of the G-action. Given a subset $I \subset \hat{G}$, denote by V_I the set of vectors in V whose V_χ-component is nonzero if and only if $\chi \in I$. We have a natural projection $\overline{V_I} \to \prod_{\chi\in I} \mathbb{P}(V_\chi)$. This has the structure of a torus bundle, the torus in question being a quotient of $\mathbb{G}_m^I$ by a finite subgroup. So the class of $\overline{V_I}$ in M_k is $(\mathbb{L}-1)^{|I|}$ times the class of $\prod_{\chi\in I}\mathbb{P}(V_\chi)$. Since V_I has also that structure, the classes of $\overline{V_I}$ and V_I in M_k coincide. Hence the same is true for $\overline{V}$ and V. □

Similarly we can form $M_S^G := K_0^G(\mathcal{V}_S)[\mathbb{L}^{-1}]$ and its dimensional completion. The class of an S-variety Z/S with G-action in M_S^G or $\hat{M}_S^G$ is denoted by $[Z/S; G]$. If G is abelian and acts trivially on S, then we have corresponding augmentations taking values in M_S and its completion.

There are corresponding characteristics in case $k \subset \mathbb{C}$. For instance, the ordinary Euler characteristic defines a ring homomorphism from M_k^G to the Grothendieck ring $K_0^G(\mathbb{Q})$ of finite dimensional representations of G over $\mathbb{Q}$ and more generally, we have a ring homomorphism χ_{top}^G from M_S^G to the Grothendieck ring of constructible sheaves with G-action on S, $K_0^G(\mathbb{Q}_S)$. Similarly, there is a Hodge character $\chi_{\text{h}}^G : M_S^G \to K_0^G(\text{HS}_S)$.

The case $G = \hat{\mu}$

We will mostly (but not exclusively) be concerned with the case when G is a group of roots of unity. We have the Grothendieck ring $M_S^{\hat{\mu}}$ of varieties with a topological action of the procyclic group $\hat{\mu} = \lim_{\leftarrow} \mu_n$ (such an action factorizes through a finite quotient μ_n). The inverse automorphism of $\hat{\mu}$, $\zeta \mapsto \zeta^{-1}$, defines an involution * in $M_S^{\hat{\mu}}$.

The group of continuous characters of $\hat{\mu}$ is naturally isomorphic with $\mathbb{Q}/\mathbb{Z}$, with the involution * acting as multiplication by -1; the projection $\hat{\mu} \to \mu_n$ followed by the inclusion $\mu_n \subset \mathbb{G}_m$ corresponds to $\frac{1}{n}$ (mod $\mathbb{Z}$). In other words, $K_0^{\hat{\mu}}(\mathbb{C}) \cong \mathbb{Z}[e^\alpha \mid \alpha \in \mathbb{Q}/\mathbb{Z}]$. For every positive integer n there is a rational irreducible representation χ_n of μ_n, namely the field $\mathbb{Q}(\mu_n)$, regarded as $\mathbb{Q}$-vector space. These make up an additive basis of $K_0^{\hat{\mu}}(\mathbb{Q})$. The image of χ_n in $\mathbb{Z}[\mathbb{Q}/\mathbb{Z}]$ is $\sum_{(k,n)=1} e^{k/n}$, which allows us to regard $K_0^{\hat{\mu}}(\mathbb{Q})$ as a subring of $\mathbb{Z}[e^\alpha \mid \alpha \in \mathbb{Q}/\mathbb{Z}]$.

The so-called mapping torus construction gives rise to an M_k-linear map

$$M_k^{\hat{\mu}} \to M_{\mathbb{G}_m}$$

with the property that composition with the direct image homomorphism $M_{\mathbb{G}_m} \to M_k$ is $(\mathbb{L}-1)$ times the augmentation $M_k^{\hat{\mu}} \to M_k$. It is defined as follows. If X is a variety with good μ_n, then its *mapping torus* is the étale locally trivial fibration $\mathbb{G}_m \times^{\mu_n} X \to \mathbb{G}_m$ whose total space is the orbit space of the μ_n-action on $\mathbb{G}_m \times X$ defined by $\zeta(\lambda, x) = (\lambda\zeta^{-1}, \zeta x)$ and for which the projection is induced by $(\lambda, x) \mapsto \lambda^n$. Notice that the fiber over $1 \in \mathbb{G}_m$ can be identified with X and that the monodromy is given by the action of μ_n on X. The projection on the second factor induces a morphism $\mathbb{G}_m \times^{\mu_n} X \to \mathbb{G}_m \to \overline{X}$ that has the structure of a piecewise $\mathbb{G}_m$-bundle. So the image of $\mathbb{G}_m \times^{\mu_n} X$ in M_k is $(\mathbb{L}-1)[\overline{X}]$. If $m = kn$ is a positive multiple of n and we let μ_m act on X via $\mu_m \to \mu_n$, then $(\lambda, x) \mapsto (\lambda^k, x)$ identifies the two fibrations over $\mathbb{G}_m$ and so we have a map as asserted. This generalizes at once to the case where we have a base variety with trivial $\hat{\mu}$-action.

$\mathrm{Aut}(\mathbb{D})$-equivariance

The automorphism group $\mathrm{Aut}(\mathbb{D})$ can be identified with the group of formal power series $k[[t]]$ with nonzero constant term where the group law is given by substitution. It acts on the arc space of any k-variety by composition: $h(\gamma) := \gamma h^{-1}$. If the variety is of pure dimension, then this action is free outside negligible subset. Clearly, a morphism of k-varieties induces an $\mathrm{Aut}(\mathbb{D})$-equivariant map between their arc spaces. Since we end up with more than just a $\mathrm{Aut}(\mathbb{D})$-invariant measure on an arc space, it is worthwhile to explicate this structure by means of a definition. If $\mathbb{D}_n$ denotes the subscheme of $\mathbb{D}$ defined by the ideal (t^{n+1}), then $\mathrm{Aut}(\mathbb{D}_n)$ (which has the same underlying variety as the group of units of $k[[t]]/(t^{n+1})$) acts naturally on $\mathcal{L}_n(X)$. For $n \geq 1$, the kernel of $\mathrm{Aut}(\mathbb{D}_{n+1}) \to \mathrm{Aut}(\mathbb{D}_n)$ can be identified with $\mathbb{G}_a$. Its action is trivial on $(\pi_1^{n+1})^{-1}(0)$ and free on the complement $(\pi_1^{n+1})^{-1}(TX - \{0\})$. By choosing a constructible section of the latter we lift the direct image homomorphism $(\pi_n^{n+1})_*$ to a map

$$M_{\mathcal{L}_{n+1}(X)}^{\mathrm{Aut}(\mathbb{D}_{n+1})} \to M_{\mathcal{L}_n(X)}^{\mathrm{Aut}(\mathbb{D}_n)}.$$

The result is easily seen to be independent of this choice.

Definition 5.2. — *An equivariant motivic measure on $\mathcal{L}(X)$ is a collection $\lambda = (\Lambda_n \in \hat{M}_{\mathcal{L}_n(X)}^{\mathrm{Aut}(\mathbb{D}_n)})_{n=1}^{\infty}$, so that Λ_n is the direct image of Λ_{n+1} for all n.*

It is clear that such a collection determines an $\hat{M}_k$-valued measure on the stable subsets. The definition is so devised that the measure $\mu_{\mathcal{L}(X)}$ constructed earlier comes from an equivariant motivic measure.

This notion is of particular interest when the variety in question is a smooth curve C and we are given a closed point $o \in C$. An $\mathrm{Aut}(\mathbb{D})$-orbit in $\mathcal{L}(C, o)$ is given by a positive integer n that may also take the value ∞. If n is finite, then this orbit projects onto the set of nonzero elements of $(\pi_{n-1}^n)^{-1}(0) \cong T_{C,o}^{\otimes n}$. The group $\mathrm{Aut}(\mathbb{D}_n)$ acts on the latter orbit through $\mathrm{Aut}(\mathbb{D}_1) \cong \mathbb{G}_m$ with $\mu_n \subset \mathbb{G}_m$ as isotropy group. So the value of λ on a fiber over $T_{C,o}^{\otimes n} - \{0\}$ is naturally an element λ_n of $M_k^{\mu_n}$. We call the generating series

$$\lambda(T) := \sum_{n=1}^{\infty} \lambda_n T^n$$

the *zeta function* of λ. It is not hard to verify that this series determines λ completely. This is particularly so if we view λ as a $\hat{M}_k$-valued measure on $\mathcal{L}(C, o)$. For instance, its value on the preimage in $\mathcal{L}(C, o)$ of a constructible subset A of $\mathcal{L}_m(C, o)$ consisting of order n-arcs (with $n \leq m$) is $\mathbb{L}^{n-m}[A]\overline{\lambda}_n$. Notice that the series $\sum_{n=1}^{\infty}(\mathbb{L}-1)\overline{\lambda}_n$ converges to the full integral of λ.

A motivic zeta function

Given a pure dimensional variety X and a flat morphism $X \to \mathbb{A}^1$, let $X_0 := f^{-1}(0)$ and denote by f the restriction $(X, X_0) \to (\mathbb{A}^1, 0)$. Then the direct image of $\mu_{\mathcal{L}(X,X_0)}$ (regarded as an equivariant measure) on $X_0 \times \mathcal{L}(\mathbb{A}^1, 0)$ is then also equivariant. We will (perhaps somewhat ambiguously) refer to this measure as the direct image of the motivic measure of $\mathcal{L}(X, X_0)$ on $X_0 \times \mathcal{L}(\mathbb{A}^1, 0)$. Its zeta function is denoted by

$$S(f) = \sum_{n=1}^{\infty} S_n(f) T^n \in \hat{M}_{X_0}^{\hat{\mu}}[[T]].$$

We now assume that X is smooth and connected. The smoothness of X ensures that the preimage in $\mathcal{L}(X, X_0)$ of a stable subset of $\mathcal{L}(\mathbb{A}^1, 0)$ of level n is stable of level n, so that $S_n(f)$ already is defined as an element of M_{X_0} (but we shall not bring out the distinction in our notation). The series $S(f)$ can be computed from an embedded resolution of the zero set of f, $H : Y \to X$ of X, as in 4.1. We assume here that the preimage E of X_0 is a simple normal crossing hypersurface that contains the exceptional set. Let m be a positive integer that is divided by all the coefficients N_i of the divisor (f) on the irreducible components of E. If we make a base change of $\tilde{f} := fH$ over the mth power map $\mathbb{A}^1 \to \mathbb{A}^1$ and normalize, then we get a μ_m-covering $\tilde{Y} \to Y$. Let $\tilde{E}_I^\circ$ be a connected component of the preimage of E_I° in $\tilde{Y}$. The restriction $\tilde{E}_I^\circ \to E_I^\circ$ is unramified, and has μ_m-stabilizer of $\tilde{E}_I^\circ$ as its Galois group. The latter is easily seen to be the subgroup μ_{N_I}, where $N(I) := \gcd\{N_i \mid i \in I\}$. This defines

$$[\tilde{E}_I^\circ / Y; \mu_{N(I)}] \in M_Y^{\mu_{N_I}}.$$

This element lies over X_0 if I is nonempty, an assumption we make from now on. We wish to compare it with $U_I(1) \subset U_I$, the fiber over 1 of the projection $U_I \to \mathbb{G}_m$ induced by $\tilde{f}$. This projection has weights $(N_i)_{i\in I}$ relative to the $\mathbb{G}_m^I$-action and so $\prod_{i\in I} \mu_{N_i} \subset \mathbb{G}_m^I$ preserves $U_I(1)$. This finite group contains a monodromy action by $\mu_{N(I)}$: write $N(I) = \sum_{i\in I} \alpha_i N_i$ and embed $\mathbb{G}_m$ in $\mathbb{G}_m^I$ by $t \mapsto (t^{\alpha_i})_{i\in I}$ (since the $(\alpha_i)_{i\in I}$ are relatively prime, this is an embedding indeed). Notice that the projection $U_I \to \mathbb{G}_m$ is homogeneous of degree N_I relative to the action of this one parameter subgroup. This implies that $\mu_{N(I)} \subset \mathbb{G}_m$ may serve as monodromy group. (There are a priori several choices for this action, but they are all E_I°-isomorphic.)

LEMMA 5.3. — *In* $M_{Y_0}^{\hat{\mu}}$ *we have* $[U_I(1)/Y_0; \mu_{N(I)}] = (\mathbb{L}-1)^{|I|-1}[\tilde{E}_I^\circ/Y_0; \mu_{N(I)}]$.

Proof. — One verifies that the Stein factorization of the projection $U_I(1) \to E_I^\circ$ has $\tilde{E}_I^\circ \to E_I^\circ$ as finite factor with $U_I(1) \to \tilde{E}_I^\circ$ being an algebraic torus bundle of rank $|I|-1$. In view of Lemma 5.1 the equivariant class of the latter is $(\mathbb{L}-1)^{|I|-1}$ times the equivariant class of the base. The lemma follows. □

Much of the work of Denef-Loeser on motivic integration centers around the following

THEOREM 5.4. — *The following identity holds in* $M_{X_0}^{\hat{\mu}}[[T]]$*:*

$$S(f) = \sum_{\varnothing \neq I \subset \operatorname{irr}(E)} (\mathbb{L}-1)^{|I|-1}[\tilde{E}_I^\circ/X_0; \mu_{N(I)}] \prod_{i\in I} (\mathbb{L}^{\nu_i}T^{-N_i}-1)^{-1}.$$

Proof. — Start with the identity of Proposition 4.2 (with $r=1$). Omit at both sides the constant terms (on the right this amounts to summing over nonempty I only), and restrict the resulting identity to the fiber over $1 \in \mathbb{G}_m$. If we take into account the monodromies and use Lemma 5.3, we get the asserted identity, at least if we take our coefficients in $\hat{M}_{X_0}^{\hat{\mu}}$. Inspection of the proof shows that this actually holds in $M_{X_0}^{\hat{\mu}}[[T]]$. □

So the expression at the righthand side is independent of the resolution, something that is not at all evident a priori. Since it lies in the $M_{X_0}^{\hat{\mu}}$-subalgebra of $M_{X_0}^{\hat{\mu}}[[T]]$ generated by the fractions $(L^\nu T^{-N}-1)^{-1}$ with $\nu, N > 0$, $S(f)$ has a value at $T=\infty$:

$$S(f)\big|_{T=\infty} = -\sum_{\varnothing \neq I \subset \operatorname{irr}(E)} (1-\mathbb{L})^{|I|-1}[\tilde{E}_I^\circ/X_0; \mu_{N(I)}]$$

Comparison with ordinary monodromy

The element $-S(f)\big|_{T=\infty}$ has an interpretation in terms of the nearby cycle sheaf of f as we shall now explain.

Suppose first that $k=\mathbb{C}$. Let $\widetilde{X - X_0} \to X - X_0 \subset X$ be the pull-back along f of the universal covering $\exp : \mathbb{C} \to \mathbb{C}^\times \subset \mathbb{C}$. Take the full direct image of the constant sheaf $\mathbb{Q}_{\widetilde{X-X_0}}$ on X and restrict to X_0: this defines ψ_f as an element of the derived category of constructible sheaves on X_0. Let $\sigma : \widetilde{X-X_0} \to \widetilde{X-X_0}$ be a generator of the covering transformation that induces in $\mathbb{C}$ translation over $-2\pi\sqrt{-1}$. This generator has the property that its action in ψ_f is the monodromy.

Let $H : Y \to X$ be a resolution as in 4.1. In the same way, $\psi_{\tilde{f}}$ is defined as an element of the derived category of constructible sheaves on the zero set Y_0 of $\tilde{f}$. The full direct image of $\psi_{\tilde{f}}$ on X_0 is equal to ψ_f.

An elementary calculation shows that the stalk of $\psi_{\tilde f}$ at a point of E_I° is the cohomology of N_I copies of a real torus of dimension N_I-1. More precisely, the restriction of $\psi_{\tilde f}$ to E_I° is naturally representable as the full direct image of the constant sheaf on $U_I(1)$ (an algebraic torus bundle of dimension N_I-1 over $\tilde E_I^\circ$) under the projection $U_I(1)\to E_I^\circ$. We have a canonical isomorphism $H^k(\mathbb{G}_m^r;\mathbb{Q})\cong H_c^{k+r}(\mathbb{G}_m^r;\mathbb{Q})$ and hence the Euler characteristic $\sum_k(-1)^k[H^k(\mathbb{G}_m^r;\mathbb{Q})]$ in $K_0(HS)$ is $(-1)^r$ times the Euler characteristic $\sum_k(-1)^k[H_c^k(\mathbb{G}_m^r;\mathbb{Q})]$. In other words, it is the value of χ_h on $(1-\mathbb{L})^r$. Hence, if Z is a subvariety of E_I° with preimage $\tilde Z$ in $\tilde E_I^\circ$, then $\sum_k(-1)^k[H_c^k(Z;\psi_{\tilde f})]$ is the value of χ_h on $(1-\mathbb{L})^{|I|-1}[\tilde Z;\mu_{N(I)}]$. This shows that ψ_f and $-S(f)\big|_{T=\infty}$ have the same Hodge characteristic. We therefore put

$$[\psi_f]:=-S(f)\big|_{T=\infty}=\sum_{\varnothing\neq I\subset\operatorname{irr}(E)}(1-\mathbb{L})^{|I|-1}[\tilde E_I^\circ/X_0;\mu_{N(I)}].$$

We refer to $[\psi_f]$ as the *nearby cycle class* of f along X_0. Its component in the augmentation submodule,

$$[\phi_f]:=[\psi_f]-\overline{[\psi_f]}\in M_{X_0}^{\hat\mu},$$

is by definition the *vanishing cycle class* of f.

Let S be a variety with trivial $\hat\mu$-action. Given a S-variety Z with a good topological $\hat\mu$-action, then for any positive integer n the fixed point locus of $\ker(\hat\mu\to\mu_n)$ in Z is a S-variety which inherits a good μ_n-action. This defines a homomorphism of M_S-algebras

$$\operatorname{Tr}_n: M_S^{\hat\mu}\to M_S^{\mu_n}.$$

If $\sigma\in\hat\mu$ generates a dense subgroup of $\hat\mu$, then the fixed point locus of $\ker(\hat\mu\to\mu_n)$ is also the fixed point locus of σ^n. In case $k\subset\mathbb{C}$, a Lefschetz fixed point formula (applied to a partition of Z by orbit type) implies that $\chi_h[Z^{\sigma^n}]$ equals the trace of σ^n in $\chi_h[Z]$. So we may then think of $Tr_n[Z]$ as the motivic trace of σ^n. This is why the following proposition is a motivic version of a result of A'Campo [2].

PROPOSITION 5.5 (see Denef-Loeser [18]). — *The series $S(f)$ and $\sum_{n=1}^\infty\operatorname{Tr}_n[\psi_f]T^n$ in $M_{X_0}^{\mu_n}[[T]]$ are congruent modulo $\mathbb{L}-1$.*

Proof. — The monodromy σ acts on $\tilde E_I^\circ$ as a covering transformation of order N_I. So σ^n has no fixed point if N_I does not divide n and is equal to all of $\tilde E_I^\circ$ otherwise. It follows from formula for the nearby cycle class that

$$\operatorname{Tr}_n[\psi_f]=\sum_{I\subset\operatorname{irr}(E),N_I|n}(1-\mathbb{L})^{|I|-1}[\tilde E_i^\circ/X_0].$$

So

$$\sum_{n=1}^{\infty} \mathrm{Tr}_n[\psi_f]T^n = \sum_{n=1}^{\infty} \sum_{I\subset \mathrm{irr}(E), N_I|n} (1-\mathbb{L})^{|I|-1}[\tilde{E}_I^\circ/X_0]T^n$$

$$= \sum_{\varnothing\neq I\subset \mathrm{irr}(E)} (1-\mathbb{L})^{|I|-1}[\tilde{E}_I^\circ/X_0]\sum_{k\geq 1} T^{kN_I} = \sum_{\varnothing\neq I\subset \mathrm{irr}(E)} (1-\mathbb{L})^{|I|-1}[\tilde{E}_I^\circ/X_0]\frac{T^{N_I}}{1-T^{N_I}}.$$

If we reduce modulo $(\mathbb{L}-1)$ only the terms with I a singleton remain. Theorem 5.4 shows that this has the same reduction modulo $(\mathbb{L}-1)$ as $S(f)$. □

6. THE MOTIVIC ZETA FUNCTION OF DENEF-LOESER [13]

This function is a motivic analogue of Igusa's local zeta function. It captures slightly less than the function $S(f)$, but has the virtue that it is defined in greater generality. First we introduce two homomorphisms of Grothendieck rings.

An arrow $M_S^{\mu_{rn}} \to M_S^{\mu_n}$ is defined by assigning to a variety with good μ_{rn}-action its orbit space with respect to the subgroup $\mu_r \subset \mu_{rn}$ (with a residual action of μ_n). The totality of these arrows forms a projective system whose limit we denote by $M_S(\hat{\mu})$. This is not the same as $M_S^{\hat{\mu}}$, but there is certainly a natural ring homomorphism

$$\rho : M_S^{\hat{\mu}} \to M_S(\hat{\mu}).$$

It is given by assigning to a variety X with good $\hat{\mu}$-action, the system $(X_n)_n$, where X_n is the orbit space of X by the kernel of $\hat{\mu} \to \mu_n$ endowed with the residual action of μ_n.

We next define the *Kummer map*

$$M_{S\times\mathbb{G}_m} \to M_S(\hat{\mu}), \quad [f] \mapsto [f]^{1/\infty}.$$

Given a S-variety Y and a morphism $f : Y \to \mathbb{G}_m$, then for every positive integer n, let $f^{1/n} : Y(f^{1/n}) \to \mathbb{G}_m$ be the pull-back of f over the nth power map $[n] : \mathbb{G}_m \to \mathbb{G}_m$. So $Y(f^{1/n})$ is the hypersurface in $\mathbb{G}_m \times Y$ defined by $f(z) = u^n$. The projection of $Y(f^{1/n}) \to Y$ is a μ_n-covering and thus defines an element $[f]^{1/n}$ of $M_S^{\mu_n}$. Notice that $Y(f^{1/n})$ is the orbit space of $Y(f^{1/nr})$ relative to the subgroup $\mu_r \subset \mu_{rn}$. Hence the $[f]^{1/n}$'s define an element $[f]^{1/\infty} \in M_S(\hat{\mu})$.

The following lemma is a straightforward exercise.

LEMMA 6.1. — *The composition of the mapping torus construction and the Kummer map is equal to* $(\mathbb{L}-1)\rho$.

For $\mathcal{X}$ a smooth $\mathbb{D}$-variety of pure relative dimension d, define the *Denef-Loeser zeta function* by

$$I(f) := \mathbb{L}^{-d} \sum_{n=0}^{\infty} [\mathrm{ac}_{f,n}]^{1/\infty}\, \mathbb{L}^{-sn} \in M_X(\hat{\mu})[[\mathbb{L}^{-s}]],$$

where $\mathbb{L}^{-s}$ is just a variable with a suggestive notation. Then Corollary 4.3 and Lemma 6.1 yield

THEOREM 6.2. — *The following identity holds in $M_X(\hat{\mu})[[\mathbb{L}^{-s}]]$:*

$$I(f) = \mathbb{L}^{-d} \sum_{I \subset \mathrm{irr}(E)} \rho[\tilde{E}_I^\circ / X; \mu_{N(I)}] \prod_{i \in I} \frac{\mathbb{L} - 1}{\mathbb{L}^{\nu_i + sN_i} - 1}.$$

Putting $\mathbb{L} = 1$

Consider the $\mathbb{Z}[L, L^{-1}]$-subalgebra S of $\mathbb{Q}(L, L^{-s})$ generated by the rational functions $(L-1)(L^{n+sN}-1)^{-1}$, $n, N \geq 1$. The spectrum of S contains the generic point of the exceptional divisor of the blow up of $(1,1)$ in $\mathbb{G}_m \times \mathbb{A}^1$. The corresponding specialization is the evalation homomorphism $S \to \mathbb{Q}(s)$ which sends $(L-1)(L^{n+sN}-1)^{-1}$ to $(n+sN)^{-1}$. According to Theorem 6.2, $I(f)$ lies in $S \otimes_{\mathbb{Z}[L,L^{-1}]} M_X(\hat{\mu})$. Evaluation at $\mathbb{L} = 1$ yields

$$I(f)\big|_{\mathbb{L}=1} = \sum_{I \subset \mathrm{irr}(D)} \rho[\tilde{E}_I^\circ / X; \mu_{N(I)}] \prod_{i \in I} \frac{1}{\nu_i + sN_i} \in M_X(\hat{\mu})/(\mathbb{L} - 1) \otimes_{\mathbb{Z}} \mathbb{Q}(s).$$

This is the motivic incarnation of the topological zeta function considered earlier by Denef and Loeser in [12]. At the time the resolution independence of this function was established using Theorem 6.3 below.

Comparison with Igusa's p-adic zeta function

Suppose we are given a complete discrete valuation ring (R, m) of characteristic zero whose residue field $F = R/m$ has finite cardinality q. Then R contains all the $(q-1)$st roots of unity μ_{q-1} and this group projects isomorphically onto $F^\times$. Let K be the quotient field of R. If we choose a uniformizing parameter $\pi \in m - m^2$, then the collection $(\zeta\pi^k)_{\zeta \in \mu_{q-1}, k \in \mathbb{Z}}$ is a system of representatives of $K^\times/(1+m)$. Define

$$\mathrm{ac}^s : K \to \mathbb{Z}[\mu_{q-1}][q^{-s}]$$

by assigning to $u \in \zeta\pi^k + m^{k+1}$ the value ζq^{-ks} and 0 to 0. (Here q^{-s} is just the name of a variable; the righthand side can be more canonically understood as the group algebra of $K^\times/(1+m)$.) There is a natural (additive) Haar measure μ on the the Boolean ring of subsets of K generated by the cosets of powers of m that takes the value 1 on R. It takes values in $\mathbb{Z}[q^{-1}]$. Given an $f \in R[x_1, \ldots, x_d]$ whose reduction mod m is nonzero, then its *Igusa local zeta function* is defined by

$$Z(f) := \int_{R^m} \mathrm{ac}^s f(x) d\mu(x),$$

where R^m is endowed with the product measure. We regard this as an element of $\mathbb{Q}[\mu_{q-1}][[q^{-s}]]$: the coefficient of ζq^{-ns} is the volume of $f^{-1}(\zeta\pi^n + m^{n+1})$. (It is customary to let s be a complex number—the series then converges in a right half plane—and to compose with a complex character $\mu_{q-1} \to \mathbb{C}^\times$.)

Let us write $\mathcal{X}$ for $\mathrm{Spec}(R[x_1, \dots, x_d])$ and regard f as a morphism $\mathcal{X} \to \mathbb{A}^1_R$ over $\mathrm{Spec}(R)$. Suppose we have an embedded resolution $H : \mathcal{Y} \to \mathcal{X}$ of the zero locus of f over $\mathrm{Spec}(R)$ with a simple normal crossing hypersurface $\mathcal{E}$ relative to $\mathrm{Spec}(R)$ (so no irreducible component in the closed fiber). Then we get an embedded resolution of the closed fiber $Y \to X$ with simple normal crossing divisor E. Make a base change of $fH : \mathcal{Y} \to \mathbb{A}^1_R$ over the $(q-1)$st power map $[q-1] : \mathbb{A}^1_R \to \mathbb{A}^1_R$ and normalize; this gives a μ_{q-1}-covering $\tilde{\mathcal{Y}} \to \mathcal{Y}$. We now get a covering $\hat{E}_I^\circ \to E_I^\circ$ defined over F with Galois group $\mu_{N_q(I)}$, where $N_q(I) := \gcd(q-1, (N_i)_{i\in I})$ over F in much the same way as before. The $\mu_{N_q(I)}$-set $\hat{E}_I^\circ(F)$ determines an element

$$\#[\hat{E}_I^\circ; \mu_{q-1}] \in \mathbb{Q}[\mu_{N_q(I)}] \subset \mathbb{Q}[\mu_{q-1}].$$

Denef proved earlier [10] the following analogue of 6.2:

Theorem 6.3 (Denef). — *In this situation we have*

$$Z(f) = q^{-d} \sum_{I \subset \mathrm{irr}(E)} \#[\hat{E}_I^\circ; \mu_{q-1}] \prod_{i\in I} \frac{q-1}{q^{\nu_i + sN_i} - 1},$$

where ν_i and N_i have the usual meaning.

As appears from 6.2, $Z(f)$ is what we get from the value of $I(f_{\bar{K}})$ on $X_0(\bar{K})$ (with $\bar{K}$ an algebraic closure of K) if we replace classes in $M_{\bar{K}}$ by the number of F-rational points in their F-counterparts (so that we substitute q for $\mathbb{L}$) and pass from $\hat{\mu}$ to μ_{q-1}. This should be understood on a more conceptual level that involves a Grothendieck ring $M^{\mu_{q-1}}_{\mathrm{Spec}(R)}$ which specializes to both $M^{\mu_{q-1}}_{\mathrm{Spec}(\bar{K})}$ and $\mathbb{Q}[\mu_{q-1}][[q^{-s}]]$, and avoids resolution.

7. MOTIVIC CONVOLUTION [15]

Join and quasi-convolution

Consider the Fermat curve J_n in $\mathbb{G}_m^2$ defined by $u^n + v^n = 1$. Notice that it is invariant under the subgroup $\mu_n^2 \subset \mathbb{G}_m^2$. If d is a positive divisor of n, then the μ_d^2-orbit space of J_n is $J_{n/d}$. In particular, the μ_n^2-orbit space of J_n is J_1, an affine line less two points. Given varieties X and Y with good μ_n-action, then we have the variety with $\mu_n \times \mu_n$-action

$$J_n(X,Y) := J_n \times^{(\mu_n \times \mu_n)} (X \times Y).$$

(If a group G acts well on varieties A and B, then $A \times^G B$ stands for quotient of $A \times B$ by the equivalence relation $(ga, b) \sim (a, gb)$ with G acting well on it by $g[a, b] := [ga, b] = [a, gb]$.) Let μ_n act on $J_n(X, Y)$ diagonally: $\zeta[(u, v), (x, y)] := [(\zeta u, \zeta v), (x, y)]$. The natural map $J_n(X, Y) \to J_1$ is étale locally trivial. If Y has trivial μ_n-action, then $J_n(X, Y) = J_n(X, pt) \times Y$ and the variety $J_n(X, pt)$ can be identified with $(\mathbb{G}_m - \{\mu_n\}) \times^{\mu_n} X$. The latter has the structure of a piecewise $\mathbb{G}_m$-bundle over $\overline{X}$ from which a copy of X has been removed. Similarly, the natural projection of $\overline{J_n(X, Y)} \to \overline{X} \times \overline{Y}$ is a piecewise $\mathbb{G}_m$-bundle from which a copy of $\overline{X \times Y}$ has been removed.

The construction is perhaps better understood in terms of the fibrations over $\mathbb{G}_m$ defined by the mapping torus construction. Recall that for a variety X with μ_n-action, its mapping torus $\mathbb{G}_m \times^{\mu_n} X$ fibers over $\mathbb{G}_m$ by $[\lambda, x] \mapsto \lambda^n$ with $\{1\} \times X$ mapping to the fiber over 1. The monodromy is the given μ_n-action on X. If Y is another variety with μ_n-action, then the composite

$$(\mathbb{G}_m \times^{\mu_n} X) \times (\mathbb{G}_m \times^{\mu_n} Y) \longrightarrow \mathbb{G}_m \times \mathbb{G}_m \subset \mathbb{G}_a \times \mathbb{G}_a \xrightarrow{+} \mathbb{G}_a$$

is a fibration over $\mathbb{G}_m$. The fiber over $1 \in \mathbb{G}_a$ is identified as $J_n(X, Y)$ and the monodromy is the given μ_n-action on $J_n(X, Y)$ defined above.

Clearly, $J_n(X, Y) \cong J_n(Y, X)$. If m is a divisor of n and the action of μ_n on X and Y is through μ_m, then $J_m(X, Y) = J_n(X, Y)$. So this induces a binary operation, the *join*

$$J : M_k^{\hat{\mu}} \times M_k^{\hat{\mu}} \to M_k^{\hat{\mu}}.$$

The preceding discussion shows that the join is commutative and bilinear over M_k and that (i) $J(a, 1) = (\mathbb{L} - 1)\overline{a} - a$ and (ii) $\overline{J(a, b)} = (\mathbb{L} - 1)\overline{a}\overline{b} - \overline{ab}$, where we recall that $a \in M_k^{\hat{\mu}} \mapsto \overline{a} \in M_k$ is the augmentation defined by 'passing to the orbit space'. This suggests to define another binary operation $*$, the *quasi-convolution*, on $M_k^{\hat{\mu}}$ by:

$$a * B := -J(a, b) + (\mathbb{L} - 1)\overline{ab}.$$

The quasi-convolution is commutative and bilinear over M_k, whereas the properties (i) and (ii) come down to

(i) 1 is a unit for $*$: $a * 1 = a$ (and hence $a * \overline{b} = a\overline{b}$) and

(ii) $\overline{a * b} = \overline{ab}$.

Neither the join nor the quasi-convolution is associative, but we do have:

(iii) $a * (b * c) - (\mathbb{L} - 1)\overline{a(b * c)} + (\mathbb{L} - 1)^2\overline{a}\overline{bc}$ is symmetric in a, b and c,

which shows that the quasi-convolution is associative modulo elements of M_k. This property is seen as follows. Let J_n^2 denotes the Fermat surface in $\mathbb{G}_m^3$ defined by $u^n + v^n + w^n = 1$ and consider the morphism

$$J_n \times J_n \to J_n^{(2)}, \quad ((u_1, v_1), (u_2, v_2)) \mapsto (u, v, w) = (u_1, v_1 u_1, v_1 u_2).$$

This morphism is equivariant with respect to the action of μ_n on $J_n \times J_n$ that is diagonal on the first factor and trivial on the second and the diagonal action μ_n on

$J_n^{(2)}$. It also factorizes over the orbit space of $J_n \times J_n$ with respect to the μ_n action defined by $\zeta((u_1, v_1), (u_2, v_2)) = ((u_1, \zeta^{-1}v_1), (\zeta u_2, \zeta v_2))$. One easily verifies that this identifies the orbit space for this action with in $J_n^{(2)} - K_n$, where $K_n \subset J_n^{(2)}$ is defined by $u^n = 1$. A choice of an nth root α of -1, identifies K_n with $\mu_n \times \mathbb{G}_m \times \mu_n$ via $(u, v, w) \mapsto (u, v, \alpha w/v)$. The μ_n^3-action on K_n carries in an obvious manner to $\mu_n \times \mathbb{G}_m \times \mu_n$.

It follows from these observations that if X, Y, Z are varieties with good μ_n-action, then $J_n^{(2)} \times^{\mu_n^3} X \times Y \times Z$ decomposes as a μ_n-variety into two pieces that can be identified with $J_n(X, J_n(Y, Z))$ and $X \times (\mathbb{G}_m \times^{\mu_n} (Y \times Z))$ respectively. The factor $\mathbb{G}_m \times^{\mu_n} (Y \times Z)$ has the structure of a $\mathbb{G}_m$-bundle over $\overline{Y \times Z}$. Passing now to $M_k^{\hat{\mu}}$ we find that

$$J(a, J(b, c)) + (\mathbb{L} - 1)a\overline{bc}$$

is symmetric in a, b, c and this is equivalent to property (iii) above.

Join and quasi-convolution extend to $\hat{M}_k^{\hat{\mu}}$ and admit relative variants.

Formation of the spectrum

Join and quasi-convolution also descend to the Grothendieck ring $K_0^{\hat{\mu}}(HS)$ of Hodge structures with $\hat{\mu}$-action. We need:

Lemma 7.1 (Shioda-Katsura, [27]). — *Given $(\alpha, \beta) \in (\mathbb{Q}/\mathbb{Z})^2$, then for every common denominator n of α and β, the Hodge type of the eigenspace $I_{\alpha,\beta}$ of $\mu_n \times \mu_n$ in $H_c^1(J_n)$ with character $(\alpha, \beta) \in (n^{-1}\mathbb{Z}/\mathbb{Z})^2$ is independent of n and we have $\dim I_{\alpha,\beta} = 1$ for $(\alpha, \beta) \neq (0, 0)$ and $\dim I_{0,0} = 2$. If $\alpha \in \mathbb{Q}/\mathbb{Z} \mapsto \tilde{\alpha} \in [0, 1[$ is the obvious section, then $I_{\alpha,\beta}$ is of Hodge type*

$$\begin{cases} (0,1) \text{ if } \alpha \neq 0 \neq \beta \text{ and } 0 < \tilde{\alpha} + \tilde{\beta} < 1, \\ (1,0) \text{ if } 1 < \tilde{\alpha} + \tilde{\beta} < 2 \text{ and} \\ (0,0) \text{ otherwise: } \alpha = 0 \text{ or } \beta = 0 \text{ or } \alpha + \beta = 0. \end{cases}$$

The only other nonzero group is $H_c^2(J_n)$, which is isomorphic to $\mathbb{Q}(-1)$ and has trivial character $(0, 0)$.

Corollary 7.2. — *If $H, H' \in K_0^{\hat{\mu}}(HS)$, then*

$$H * H' = H_0 \otimes H'_0 + \sum_{\alpha \neq 0} H_\alpha \otimes H'_{-\alpha}(-1) + \sum_{\alpha+\beta \neq 0} H_\alpha \otimes H'_\beta \otimes I_{\alpha,\beta}.$$

Anderson [3] investigated Hodge structures with $\hat{\mu}$-action using a notion of a *fractional Hodge structure*. For us such a structure will consist of a complex vector space V defined over $\mathbb{Q}$ with a complex decomposition $V = \oplus_{p,q \in \mathbb{Q}; p+q \in \mathbb{Z}} V^{p,q}$ such that $V^{q,p}$ is the complex conjugate of $V^{p,q}$ and $\oplus_{p+q=n} V^{p,q}$ is defined over $\mathbb{Q}$ for every $n \in \mathbb{Z}$. They form an abelian category $HS(\mathbb{Q})$ with tensor product. Anderson associates to a Hodge structure H with $\hat{\mu}$-action a fractional Hodge structure $\sigma(H)$ whose underlying vector space is H, leaves the bidegrees on H_0 unaltered and increases the bidegrees of H_α by $(\tilde{\alpha}, 1-\tilde{\alpha})$ if $\alpha \neq 0$. We shall refer to this operation as the *formation of the spectrum*. It defines an additive functor and hence a homomorphism of groups $\mathrm{sp} : K_0^{\hat{\mu}}(HS) \to K_0(HS(\mathbb{Q}))$. This is not a ring homomorphism, but Corollary 7.2 shows that sp takes quasi-convolution to the tensor product:

$$\mathrm{sp}(H * H') = \mathrm{sp}(H) \otimes \mathrm{sp}(H').$$

Convolution

In what follows we need the (additive) group structure on the affine line, so we write $\mathbb{G}_a$ instead of $\mathbb{A}^1$. We have a bijection $\mathcal{L}(\mathbb{G}_a, 0) \cong \mathfrak{m}$, defined by assigning to $\gamma \in \mathcal{L}(\mathbb{G}_a, 0)$ the pull-back of the standard coordinate on $\mathbb{G}_a$.

Let $\lambda = (\Lambda_n)_n$ and $\lambda' = (\Lambda'_n)_n$ be equivariant measures on $\mathcal{L}(\mathbb{G}_a, 0)$. Then $\lambda \times \lambda' := (\Lambda_n \times \Lambda'_n)_{n=1}^\infty$ defines a measure on the algebra of stable subsets of $\mathcal{L}(\mathbb{G}_a, 0)^2$ (that is, preimages of constructible subset of some truncation $\mathcal{L}_n(\mathbb{G}_a, 0)^2)$. For instance, if $C \subset \mathcal{L}_n(\mathbb{G}_a, 0)^2$ is constructible and consists of pairs of truncated arcs of fixed order (k, l) (with $k, l \leq n$), then the value of $\lambda \times \lambda'$ on the preimage of C in $\mathcal{L}(\mathbb{G}_a, 0)^2)$ is $\lambda_k \lambda'_l [C] \mathbb{L}^{-2n}$.

The direct image of $\lambda \times \lambda'$ under the addition morphism $\mathrm{add} : \mathbb{G}_a \times \mathbb{G}_a \to \mathbb{G}_a$, $\lambda * \lambda' := (\mathcal{L}_n(\mathrm{add})(\Lambda_n \times \Lambda'_n))_{n=1}^\infty$, is an equivariant measure on $\mathcal{L}(\mathbb{G}_a, 0)$, called the *convolution* of λ and λ'.

Lemma 7.3. — *The zeta function of $\lambda * \lambda'$ is determined by those of λ and λ':*

$$(\lambda * \lambda')_n = -(\lambda_n * \lambda'_n) + (\mathbb{L} - 1) \sum_{i \leq n} \mathbb{L}^{i-n} \overline{\lambda_i \lambda'_i)} + (\mathbb{L} - 1) \sum_{i > n} (\lambda_n \overline{\lambda'_i} + \overline{\lambda_i} \lambda'_n).$$

Proof. — The preimage of $t^n + \mathfrak{m}^{n+1}$ in $\mathfrak{m} \times \mathfrak{m}$ under $\mathcal{L}(\mathrm{add})$ decomposes into the following pieces: $(t^n + \mathfrak{m}^{n+1}) \times \mathfrak{m}^{n+1}$, $\mathfrak{m}^{n+1} \times (t^n + \mathfrak{m}^{n+1})$ and for $i = 1, \ldots, n$ the preimage $\tilde{C}_{n,i}$ of the subset $C_{n,i} \subset ((\mathfrak{m}^i - \mathfrak{m}^{i+1})/\mathfrak{m}^{n+1})^2$ of pairs $(\alpha_i t^i + \cdots + \alpha_n t^n, \beta_i t^i + \cdots + \beta_n t^n)$ with $\alpha_k + \beta_k = 0$ for $k = i, \ldots, n-1$ and $\alpha_n + \beta_n = 1$. We must evaluate $\lambda \times \lambda'$ on each of these (relative to the diagonal μ_n-action). The first piece gives $\lambda_n \sum_{i > n} (\mathbb{L} - 1) \overline{\lambda'_i}$ and the second the same expression with λ and λ' interchanged. Since $[C_{n,i}] = [(\mathfrak{m}^i - \mathfrak{m}^{i+1})/\mathfrak{m}^{n+1}] = (\mathbb{L} - 1)\mathbb{L}^{n-i}$, we find that for $i < n$, the value of $\lambda \times \lambda'$ on $\tilde{C}_{n,i}$ equals $(\mathbb{L} - 1)\mathbb{L}^{i-n} \overline{\lambda_i \lambda'_i}$ (the action of μ_n is trivial here). Notice that $C_{n,n}$ is embedded in $(\mathfrak{m}^n - \mathfrak{m}^{n+1}/\mathfrak{m}^{n+1})^2$ as J_1 in $\mathbb{G}_m^2$. From the above discussion one sees that $\lambda \times \lambda'$ takes on this set the value $J(\lambda_n, \lambda'_n)$. If we substitute the defining equation for $*$, the Lemma follows. $\square$

This lemma suggests a notion of a convolution operator for series $\lambda(T) = \sum_{n=1}^\infty \lambda_n T^n \in \hat{M}_k^{\hat{\mu}}[[T]]$ with the property that the *mass* $(\mathbb{L} - 1) \sum_{n=1}^\infty \overline{\lambda}_n$ converges.

For a $\mathbb{Z}[L, L^{-1}]$-module M we set

$$M\langle T \rangle := M[T][\frac{1}{T^N - L^\nu} \mid \nu \in \mathbb{Z}, N = 1, 2, 3, \ldots].$$

Expanding the denominators $(1 - T^N L^{-\nu})^{-1}$ in T embeds $M\langle T\rangle$ in $M[[T]]$ and expanding $(1 - T^{-N} L^{\nu})^{-1}$ in T^{-1} embeds $M\langle T\rangle$ in $M[[T^{-1}]][T]$.

According to Theorem 5.4, $S(f) \in \hat{M}^{\hat{\mu}}\langle T\rangle$.

THEOREM 7.4 (Abstract Thom-Sebastiani property). — *Let λ and λ' be equivariant measures on $\mathcal{L}(\mathbb{G}_a, 0)$ whose zeta functions lie in $\hat{M}^{\hat{\mu}}\langle T\rangle$. Then $\lambda * \lambda'$ has this property, too. If moreover λ and λ' have zero mass and zeta functions converging at $T = \infty$, then $\lambda * \lambda'(T)$ has these properties as well and $(\lambda * \lambda')(\infty) = \lambda(\infty) * \lambda'(\infty)$.*

COROLLARY 7.5. — *Let X and Y be smooth connected varieties and $f : X \to \mathbb{G}_a$, $g : Y \to \mathbb{G}_a$ nonconstant morphisms with zero fibers X_0 and Y_0. Let $f*g : X\times Y \to \mathbb{G}_a$ be defined by $(f * g)(x, y) := f(x) + g(y)$. Then the restriction of $[\phi_{f*g}]$ to $X_0 \times Y_0$ and the exterior $*$-product $[\phi_f] * [\phi_g] \in M_{X_0\times Y_0}$ coincide.*

If we apply the Hodge number characteristic followed by formation of the spectrum, then we recover the Thom-Sebastiani property for the spectrum, proved earlier by Varchenko in case f and g have isolated singularities and by M. Saito [28] in general.

For the proof of Theorem 7.4 we need the following

LEMMA 7.6. — *Let M and N be $\mathbb{Z}[L, L^{-1}]$-modules and let $a \in M\langle T\rangle$ and $b \in N\langle T\rangle$ both be zero at $T = 0$ and regular at $T = \infty$. If $\sum_{k>0} a_k T^k$ resp. $\sum_{k>0} b_k T^k$ are their expansions at 0, then $\sum_{k>0}(a_k \otimes b_k)T^k$ is the expansion at zero of a $c \in (M \otimes_{\mathbb{Z}[L,L^{-1}]} N)\langle T\rangle$ whose value at $T = \infty$ equals $-a(\infty) \otimes b(\infty)$.*

Proof. — It is easy to see that it suffices to prove this for $M = N = \mathbb{Z}[L, L^{-1}]$. The idea of the proof in this case is inspired by a paper of Deligne [9]. Fix for the moment $L \in \mathbb{C} - \{0\}$. Let $r_0 > 0$ be a radius of convergence for the two expansions. Let $T \in \mathbb{C}$ be such that $|T| < r_0^2$ and choose $|T|/r_0 < r < r_0$. Consider the integral

$$c(T) := \frac{1}{2\pi\sqrt{-1}} \int_{|\tau|=r} a(T/\tau)b(\tau)\frac{d\tau}{\tau}.$$

On the circle of integration the expansions converge uniformly and absolutely and so

$$c(T) = \frac{1}{2\pi\sqrt{-1}} \int_{|\tau|=r} \sum_{k,l\in\mathbb{N}} a_k T^k b^l \tau^{k-l} \frac{d\,\tau}{\tau}.$$

Since summation and integration may be interchanged, only the terms with $k = l$ remain and hence $c(T) = \sum_{k\in\mathbb{N}} a_k b_k T^k$. If P_a resp. P_b denotes the set of poles of a resp. b, then the integrand has polar set $TP_a^{-1} \cup P_b$ (there is no pole in 0 or ∞) and the poles enclosed by the circle of integration are those in TP_a^{-1}. By the theory of residues, $-c(T)$ must then be equal to the sum of the residues of the integrand at P_b. This description no longer requires $|T| < r_0^2$ and defines an analytic extension of c to the complement of P_aP_b. This extension is easily seen to be meromorphic at P_aP_b. To compute its behavior at ∞, we note that $a(T/\tau)$ converges for $T \to \infty$ on a neighborhood of P_b absolutely (with all its derivatives) to the constant function $a(\infty)$.

So as $T \to \infty$, $-c(T)$ tends to the sum of the residues of $a(\infty)b(\tau)\tau^{-1}\, d\tau$ at P_b. This sum is opposite to the residue at the remaining pole ∞, hence equal to $a(\infty)b(\infty)$. In particular, c is a rational function with polar set contained in $P_a P_b$.

Assume now that $a, b \in \tilde{R}$. A pole of an element of R in $\mathbb{C}^\times \times \mathbb{C}$ satisfies an equation $T^N = L^\nu$ for certain integers $N > 0$, $\nu \geq 0$. A product of such poles satisfies a similar equation, and this implies that a product of c and a finite set of polynomials of the form $T^N - L^\nu$ is in $\mathbb{C}[L, L^{-1}, T]$. Since the expansion of c at $T = 0$ has integral coefficients, this product lies in $\mathbb{Z}[L, L^{-1}, T]$. $\square$

Proof of Theorem 7.4. — We start with the convolution formula 7.3. It says that

$$(\lambda * \lambda')(T) = -\sum_{n>0} \lambda_n * \lambda_n T^n + (\mathbb{L}-1) \sum_{0<i\leq n} \overline{\lambda_i \lambda'_i} \mathbb{L}^{i-n} T^n + (\mathbb{L}-1) \sum_{i>n>0} (\lambda_n \overline{\lambda}'_i + \overline{\lambda}_i \lambda'_n) T^n.$$

We now assume that λ and λ' are massless so that $(\mathbb{L}-1)\sum_{i>n} \overline{\lambda}_i = -(\mathbb{L}-1)\sum_{i=1}^n \overline{\lambda}_i$ and similarly for λ'. We then have

$$\begin{aligned}(\lambda * \lambda)(T) = -\sum_{n>0} \lambda_n * \lambda'_n T^n + (\mathbb{L}-1) \sum_{0<i\leq n} \overline{\lambda_i \lambda'_i} \mathbb{L}^{i-n} T^n + \\ + (\mathbb{L}-1) \sum_{n>0} \overline{\lambda_n \lambda'_n} T^n - (\mathbb{L}-1) \sum_{0<i\leq n} (\lambda_n \overline{\lambda}'_i + \overline{\lambda}_i \lambda'_n) T^n.\end{aligned}$$

We consider each series on the right separately. By Lemma 7.6, $-\sum_{n>0} \lambda_n * \lambda'_n T^n$ is in the in $\hat{M}_k^{\hat{\mu}}\langle T\rangle$ with value at ∞ equal to $\lambda(\infty) * \lambda'(\infty)$. We also have

$$(\mathbb{L}-1) \sum_{0<i\leq n} \overline{\lambda_i \lambda'_i} \mathbb{L}^{i-n} T^n = (\mathbb{L}-1) \sum_{i>0} \sum_{k\geq 0} \overline{\lambda_i \lambda'_i} \mathbb{L}^{-k} T^{k+i} = \frac{\mathbb{L}-1}{1-\mathbb{L}^{-1}T} \sum_{i>0} \overline{\lambda_i \lambda'_i} T^i.$$

By the same 7.6 the righthand side is in $\hat{M}_k^{\hat{\mu}}\langle T\rangle$ and takes the value zero at ∞. Since

$$\sum_{0<i\leq n} \overline{\lambda}_i T^n = -(T-1)^{-1} \sum_{i>0} \overline{\lambda}_i T^i$$

is in $\hat{M}_k^{\hat{\mu}}\langle T\rangle$ with value zero at ∞ it follows from 7.6 that the same is true for $(\mathbb{L}-1)\sum_{0<i\leq n} (\overline{\lambda}_i \lambda'_n) T^n$. Likewise for $(\mathbb{L}-1)\sum_{0<i\leq n} (\lambda_n \overline{\lambda}'_i) T^n$. So $(\lambda * \lambda')(T)$ is in $\hat{M}_k^{\hat{\mu}}\langle T\rangle$ and has value $\lambda(\infty) * \lambda'(\infty)$ at ∞. $\square$

8. THE MCKAY CORRESPONDENCE [6], [16], [26]

Suppose a group G of finite order m acts well and effectively on a smooth connected variety U of dimension d. This defines an orbifold $p : U \to U_G$ with underlying variety $G\backslash U$. Let us write X for the orbifold U_G. We also fix a primitive mth root of unity ζ_m.

Let $g \in G$ and let U^g be its fixed point set in U. The action of g in the normal bundle of U^g decomposes that bundle into a direct sum of eigensubbundles

$$\nu_{U/U^g} = \oplus_{k=1}^{m-1} \nu_g^k,$$

where ν_g^k has eigenvalue ζ_m^k. We like to think of ν_g^k as the pull-back of a fractional bundle on a subvariety of X whose virtual rank is k/m times that of ν_g^k. A more formal discussion involves the extension $M_X[\mathbb{L}^{1/m}]$ of M_X obtained by adjoining an mth root of $\mathbb{L}$. To be precise, let $w(g) := \sum_k \frac{k}{m}\,\mathrm{rk}(\nu_g^k)$, considered as locally constant function $U^g \to m^{-1}\mathbb{Z}$, and let $\mathbb{L}_{U^g}^{w(g)}$ be the element of $M_{U^g}[\mathbb{L}^{1/m}] \subset M_U[\mathbb{L}^{1/m}]$ that this defines. Then $\sum_{g\in G} \mathbb{L}_{U^g}^{w(g)}$ is the image under p^* of

$$W(X) = \sum_{[g]\in\mathrm{conj}(G)} \sum_{i\in\pi_0(U^g)} [(G_i\backslash U_i^g)/X]\mathbb{L}^{w_i} \in M_X[\mathbb{L}^{1/m}].$$

Here U_i^g is the connected component of U^g labeled by i, G_i is the G-stabilizer of this component, and w_i the value of $w(g)$ on U_i^g. The sum is over a system of representatives of the conjugacy classes of G and can be rewritten as one over the orbifold strata of X (see Reid [26]): the decomposition of U into connected strata by orbit type (a stratum is a connected component of the locus of points with given G-stabilizer) induces a partition of X into orbifolds and $W(X)$ has the form $\sum_S [S]W_S$, where the sum is over the orbifold strata, and W_S is a polynomial in $\mathbb{L}^{1/m}$. We will see that $W(X)$ can be understood as the class of an obstruction bundle for lifting arcs in X to arcs in U.

The McKay correspondence identifies $W(X)$ in terms of a resolution of X:

Theorem 8.1 (Batyrev [6], Denef-Loeser [16]). — *Let $H : Y \to X$ be a resolution of the orbifold X whose exceptional divisor E has simple normal crossings. With the usual meaning of E_I° and with ν_i^* as defined below we have the following identity in $\hat{M}_X[\mathbb{L}^{1/m}]$:*

$$W(X) = \sum_{I\subset \mathrm{irr}(D)} [E_I^\circ/X] \prod_{i\in I} \frac{\mathbb{L}-1}{\mathbb{L}^{\nu_i^*}-1}.$$

The statement does not involve arc spaces, but the proof does. It could well be that the identity is already valid in $M_X[\mathbb{L}^{1/m}]$. The relative simplicity of the lefthand side has implications for the righthand side, one of which is that all the 'non-Tate' material in a fiber of H must cancel out in the sum. For that same reason the lefthand side is hardly affected if we apply the weight character relative to X to it, that is, if we take the image of $W(X)$ in the Grothendieck ring of constructible $\mathbb{Z}((w^{-1/m}))$-modules on X: just substitute w^2 for $\mathbb{L}$.

We first seek an orbifold measure on $\mu_{\mathcal{L}(X)}^{\mathrm{orb}}$ on $\mathcal{L}(X)$ with the property that for every G-invariant measurable $A \subset \mathcal{L}(U)$ we have

$$\mu_{\mathcal{L}(X)}^{\mathrm{orb}}(p_*A) := \overline{\mu_{\mathcal{L}(U)}(A)},$$

where the righthand side should be interpreted as follows: think of $\mu_{\mathcal{L}(U)}(A)$ as an element of $\hat{M}_k^G$, and then let $\overline{\mu_{\mathcal{L}(U)}(A)}$ be the image of $\mu_{\mathcal{L}(U)}(A)$ under the augmentation $\hat{M}_k^G \to \hat{M}_k$. Since $p_* : \mathcal{L}(U) \to \mathcal{L}(X)$ need not be surjective, this will not characterize the orbifold measure a priori. But it suggests how to define it: suppose that the Jacobian ideal $\mathcal{J}_p$ has constant order e along A. Then the usual measure of $\mathcal{L}(X)$ pulled back to A is $\mathbb{L}^{-e}\mu_{\mathcal{L}(U)}|A$. We therefore want the orbifold measure restricted to $p_*(A)$ to be the restriction of $\mathbb{L}^e\mu_{\mathcal{L}(X)}$. This can be done as follows. Let r be a positive integer such that $(\Omega_U^d)^{\otimes r}$ descends to an invertible sheaf $\omega_X^{(r)}$ on X. (So for every $u \in U$, G_u acts on on the tangent space T_uU with determinant an rth root of unity.) There is a natural homomorphism $(\Omega_X^d)^{\otimes r} \to \omega_X^{(r)}$ whose kernel is the torsion of $(\Omega_X^d)^{\otimes r}$. The image of this homomorphism has the form $\mathcal{I}^{(r)}\omega_X^{(r)}$ for an ideal $\mathcal{I}^{(r)}$. We set

$$\mu_{\mathcal{L}(X)}^{\mathrm{orb}} := \mathbb{L}^{\mathrm{ord}_{\mathcal{I}^{(r)}}/r}\mu_{\mathcal{L}(X)}.$$

It is a measure that takes values in $\hat{M}_k[\mathbb{L}^{1/r}]$.

Lemma 8.2. — *The pull-back of $\mu_{\mathcal{L}(X)}^{\mathrm{orb}}$ under p^* is a measure that assigns to any G-invariant measurable subset A of $\mathcal{L}(U)$ the image of A under the augmentation map $\hat{M}_k^G \to \hat{M}_k$.*

Proof. — If we apply p^* to the identity $(\Omega_X^d)^{\otimes r}/tors = \mathcal{I}^{(r)}\omega_X^{(r)}$ we get $\mathcal{J}_p^r(\Omega_U^d)^{\otimes r} = p^*(\mathcal{I}^{(r)})\omega_U^{\otimes r}$. Since $\Omega_U^d = \omega_U$, it follows that $p^*(\mathcal{I}^{(r)}) = \mathcal{J}_p^r$. So $\mu_{\mathcal{L}(X)}^{\mathrm{orb}}$ pulls back under p^* to $\mathbb{L}^{-\mathrm{ord}_{\mathcal{J}_p}+p^*(\mathcal{I}^{(r)})/r}\mu_{\mathcal{L}(U)} = \mu_{\mathcal{L}(U)}$. The rest is left to the reader. □

The following lemma describes the direct image of $\mu_{\mathcal{L}(X)}^{\mathrm{orb}}$ on X in terms of a resolution of X: let $Y \to X$ be a resolution of singularities with simple normal crossing divisor E. We have $H^*\omega_X^{(r)} = \tilde{\mathcal{I}}^{(r)}\omega_Y^{\otimes r}$ for some fractional ideal $\tilde{\mathcal{I}}^{(r)}$ on Y. It is known that the multiplicity m_i of E_i in this ideal is $> -r$. So $\nu_i^* := 1 + m_i/r$ is positive. Entirely analogous to the proof of Theorem 4.2 one derives:

Lemma 8.3. — *The direct image of $\mu_{\mathcal{L}(X)}^{\mathrm{orb}}$ on X is represented by the class*

$$\sum_{I\subset \mathrm{irr}(E)} [E_I^\circ/X] \prod_{i\in I} \frac{\mathbb{L}-1}{\mathbb{L}^{\nu_i^*}-1} \in \hat{M}_X[\mathbb{L}^{1/r}].$$

Let $\mathcal{L}'(X)$ be the set of arcs in X not contained in the discriminant of $p : U \to X$. This is a subset of full measure. We decompose $\mathcal{L}'(X)$ according to the ramification behavior of $p : U \to X$. Let $[m] : \mathbb{D} \to \mathbb{D}$ be the mth power map and denote the parameter of the domain by $t^{1/m}$. We regard ζ_m (through its action on the domain) as generator of the Galois group of $[m]$. For $\gamma \in \mathcal{L}'(X)$, $\gamma[m]$ lifts to a morphism $\tilde{\gamma} : \mathbb{D} \to \tilde{X}$ and this lift is unique up to conjugation with G. Given the lift, there is a $g \in G$ such that $g\tilde{\gamma} = \tilde{\gamma}\zeta_m$. Its conjugacy class $[g]$ in G only depends on γ. This conjugacy class determines the isomorphism type of the G-covering over γ: if m' is the order of g, then $\gamma^*(p)$ is isomorphic to $G \times^{\langle g \rangle} \mathbb{D} \to \mu_{m'} \backslash \mathbb{D}$, with g acting on $\mathbb{D}$ as multiplication by $\zeta^{m/m'}$. Notice that $\tilde{\gamma}(o)$ is in the fixed point set U^g. The 'fractional lifts' $\tilde{\gamma}$ that so arise are like arcs in the total space of the normal bundle $\oplus_k \nu_g^k$ of U^g (based at the zero section) which in the ν_g^k-direction develop as $t^{k/m}$ times a power series in t.

Denote the set of arcs in $\mathcal{L}'(X)$ belonging to the conjugacy class of $[g]$ of g by $\mathcal{L}(X, [g])$. The McKay correspondence now results from:

LEMMA 8.4. — *The subset $\mathcal{L}(X, [g])$ is measurable for $\mu^{\mathrm{orb}}_{\mathcal{L}(X)}$ and the restriction of $\mu^{\mathrm{orb}}_{\mathcal{L}(X)})$ to this subset is represented by the class $[(G_g \backslash U^g)/X]\mathbb{L}^{w(g^{-1})} \in M_X[\mathbb{L}^{1/m}]$, where G_g is the G-stabilizer of U^g.*

The proof is a calculation which we only discuss in a heuristic fashion. The elements of $\mathcal{L}(X, [g])$ correspond to G_g-orbits of fractional lifts as described above. In view of our definition of orbifold measure, we need to argue that these fractional lifts are represented by the element $\mathbb{L}^{w(g^{-1})}_{U^g}$. If $r_1, \ldots, r_{m-1}$ are positive integers, then the arcs in $\oplus_k \nu_g^k$ of U^g based at the zero section and which in the ν_g^k-direction have order r_k make up a constructible subset of $\mathcal{L}(\oplus_k \nu_g^k)$ whose class is easily seen to be equal to $\mathbb{L}^w_{U^g}$, with $w = \sum_k (1 - r_k) \operatorname{rk}(\nu_g^k)$. The fact is that this also holds for the fractional values $r_k = k/m$. So in that case we have $w = \sum_k (1 - k/m) \operatorname{rk}(\nu_g^k) = w(g^{-1})$.

9. PROOF OF THE TRANSFORMATION RULE [14]

Let $\mathcal{X}/\mathbb{D}$ be a $\mathbb{D}$-variety of pure relative dimension d. The dth Fitting ideal of $\Omega_{\mathcal{X}/\mathbb{D}}$ defines the locus where $\mathcal{X}$ fails to be smooth over $\mathbb{D}$; we denote that ideal by $\mathcal{J}(\mathcal{X}/\mathbb{D})$. Locally this ideal is obtained as follows: if $\mathcal{X}$ is given as a closed subset of $(\mathbb{A}^{d+l})_{\mathbb{D}}$, then $\mathcal{J}_{\mathcal{X}/\mathbb{D}}$ is the restriction to $\mathcal{X}$ of the ideal generated by the determinants $\det((\partial f_j / \partial x_{i_k})^l_{j,k=1})$, where $f_1, \ldots, f_l$ are taken from the ideal $I_{\mathcal{X}} \subset \mathcal{O}[x_1, \ldots, x_{d+l}]$ defining $\mathcal{X}$ and $1 \leq i_1 < \cdots < i_l \leq d + l$.

Let $\gamma \in \mathcal{X}_\infty$ be such that $\gamma^* \mathcal{J}(\mathcal{X}/\mathbb{D})$ has finite order e. This implies that γ maps $\mathbb{D}^\times$ to the part $(\mathcal{X}/\mathbb{D})_{\mathrm{reg}}$ where $\mathcal{X}$ is smooth over $\mathbb{D}$. In particular, $\gamma^* \Omega_{\mathcal{X}/\mathbb{D}}$ is a $\mathcal{O}$-module of rank d. Since the formation of a Fitting ideal commutes with base change, the dth Fitting ideal of $\gamma^* \Omega_{\mathcal{X}/\mathbb{D}}$ will be $\mathfrak{m}^e$. This means that the torsion of $\gamma^* \Omega_{\mathcal{X}/\mathbb{D}}$ has length e.

It is clear that $\mathrm{Der}_{\mathcal{O}}(\mathcal{O}_{\mathcal{X},\gamma(o)}, \mathcal{O}) \cong \mathrm{Hom}_{\mathcal{O}}(\gamma^* \Omega_{\mathcal{X}/\mathbb{D}}, \mathcal{O})$ is a free $\mathcal{O}$-module of rank d (where $\mathcal{O}$ is a $\mathcal{O}_{\mathcal{X},\gamma(o)}$-module via $\gamma*$). The fiber over o, $\mathrm{Hom}_{\mathcal{O}}(\gamma^* \Omega_{\mathcal{X}/\mathbb{D}}, \mathcal{O}) \otimes_{\mathcal{O}} k$, is d-dimensional subspace of the Zariski tangent space $T_{X,\gamma(o)}$, which we shall denote by $\hat{T}_{X,\gamma}$. Any $\mathcal{O}$-homomorphism $\gamma^* \Omega_{\mathcal{X}/\mathbb{D}} \to \mathcal{O}/\mathfrak{m}^{n+1}$ that kills the torsion lifts to a $\mathcal{O}$-homomorphism $\gamma^* \Omega_{\mathcal{X}/\mathbb{D}} \to \mathcal{O}$. This is automatic when $n \geq e$ and so $\hat{T}_{X,\gamma}$ only depends on the e-jet of γ. The space $\hat{T}_{X,\gamma}$ has a simple geometric interpretation: it is the 'limiting position' of the tangent space along the fibers of $\mathcal{X}/\mathbb{D}$ at the generic point of $\gamma(\mathbb{D})$ in the closed point $\gamma(o)$.

If $\gamma' : \mathbb{D} \to \mathcal{X}$ has the same n-jet as γ, then γ^* and γ'^* differ by a homomorphism $\mathcal{O}_{\mathcal{X},\gamma(o)} \to \mathfrak{m}^{n+1}$. The reduction modulo $\mathfrak{m}^{2(n+1)}$ of this homomorphism is a $\mathcal{O}$-derivation, i.e., defines an element of $\mathrm{Hom}_{\mathcal{O}}(\gamma^*\Omega_{\mathcal{X}/\mathbb{D}}, \mathfrak{m}^{n+1}/\mathfrak{m}^{2(n+1)})$. Its reduction modulo $\mathfrak{m}^{n+2}$ will lie in $\hat{T}_{X,\gamma} \otimes \mathfrak{m}^{n+1}/\mathfrak{m}^{n+2}$, provided that $n \geq e$. The next lemma shows that every element of this k-vector space so arises.

Lemma 9.1. — *Assume that $n \geq e$. The fiber of $\pi_{n+1}\mathcal{X}_\infty \to \pi_n\mathcal{X}_\infty$ over $\pi_n(\gamma)$ is an affine space with translation space $\hat{T}_{X,\gamma} \otimes_k \mathfrak{m}^{n+1}/\mathfrak{m}^{n+2}$. This defines an affine space bundle of rank d over the locus of $\pi_n\mathcal{X}_\infty$ defined by $\mathrm{ord}_{\mathcal{J}(\mathcal{X}/\mathbb{D})} \leq n$.*

Proof. — Assume that $\mathcal{X}$ is given as a closed subset of $(\mathbb{A}^{d+l})_{\mathbb{D}}$ as above. There exist $f_1, \dots, f_l \in I_{\mathcal{X}}$ and $1 \leq i_1 < \dots < i_l \leq d+l$ such that the Jacobian matrix $\det((\partial f_j/\partial x_{i_k})_{j,k=1}^l)$ has order e along γ, whereas for any other matrix thus formed the order is $\geq e$. By means of a coordinate change we may arrange that

$$\gamma^* df_j \equiv t^{e_j} dx_j \pmod{t^{e_j+1}(dx_{j+1}, \dots, dx_{d+l})}, \quad j = 1, \dots l,$$

so that $e = \sum_j e_j$. The subspace of $\mathbb{A}_k^{d+l}$ spanned by the last d basis vectors is then just $\hat{T}_{X,\gamma}$.

We investigate which $u_0 \in k^{d+l}$ appear as the constant coefficient of an $u \in \mathcal{O}^{d+l}$ with the property that $\gamma + t^{n+1}u \in \mathcal{X}_\infty$. We first do this for the complete intersection defined by $f_1, \dots, f_l$. This complete intersection contains $\mathcal{X}$ and the irreducible component that contains the image of γ lies in $\mathcal{X}$. So we want $f_j(\gamma + t^{n+1}u) = 0$ for $j = 1, \dots, l$. By expanding at γ this amounts to identities of the form

$$t^{n+1} D_\gamma f_j(u) + t^{2(n+1)} F_j(u) = 0, \quad j = 1, \dots, l,$$

with $D_\gamma f_j$ the derivative of f_j at γ and $F_j \in \mathcal{O}[x_1, \dots, x_{d+l}]$. Equivalently:

$$t^{-e_j} D_\gamma f_j(u) + t^{n+1-e_j} F_j(u) = 0, \quad j = 1, \dots, l.$$

All the terms are regular and the reduction modulo t yields the jth unit vector in k^{d+l}. Hensel's lemma says that a solution u exists if and only if u_0 solves this set of equations modulo t. This just means that $u_0 \in \hat{T}_{X,\gamma}$. In particular, we see that for all $k \in \mathbb{N}$, $\pi_{n+k}\pi_n^{-1}\pi_n(\gamma)$ is isomorphic to an affine space and hence is irreducible. This implies that all elements of $\pi_n^{-1}\pi_n(\gamma)$ map to the same irreducible component of the common zero locus of $f_1, \dots, f_l$. It follows that $\pi_n^{-1}\pi_n(\gamma) \subset \mathcal{X}_\infty$. The last assertion is easy. □

Proof of Proposition 3.1. — Suppose that $\mathcal{X}$ is of pure relative dimension d. Let C_e denote the subset of $\mathcal{X}_\infty$ defined by $\mathrm{ord}_{\mathcal{J}(\mathcal{X}/\mathbb{D})} = e$. It is clear that $C_e = \pi_e^{-1}\pi_e(C_e)$. It follows from Greenberg's theorem [21] that $\pi_e(C_e)$ is constructible. Hence C_e is stable by Lemma 9.1. We have $\cup_e C_e = \mathcal{X}_\infty - (\mathcal{X}_{\mathrm{sing}})_\infty$. In view of Lemma 2.3 it now suffices to see that $\dim \pi_e(C_e) - de \to -\infty$ as $e \to \infty$. This is not difficult. □

Let $H : \mathcal{Y} \to \mathcal{X}$ be a $\mathbb{D}$-morphism of $\mathbb{D}$-varieties of pure relative dimension d. Recall that the Jacobian ideal $\mathcal{J}_H$ of H is the 0th Fitting ideal of $\Omega_{\mathcal{Y}/\mathcal{X}}$. Suppose $\gamma \in \mathcal{Y}_\infty$ is such that $\mathcal{J}_H$ has finite order e along γ. Then γ resp. $H\gamma$ maps the generic point $\mathbb{D}^\times$ to $(\mathcal{Y}/\mathbb{D})_{\mathrm{reg}}$ resp. $(\mathcal{X}/\mathbb{D})_{\mathrm{reg}}$. We have an exact sequence of $\mathcal{O}$-modules

$$(H\gamma)^*\Omega_{\mathcal{X}/\mathbb{D}} \to \gamma^*\Omega_{\mathcal{Y}/\mathbb{D}} \to \gamma^*\Omega_{\mathcal{Y}/\mathcal{X}} \to 0.$$

The base change property of Fitting ideals implies that the length of $\gamma^*\Omega_{\mathcal{Y}/\mathcal{X}}$ must be e. So if $\gamma^*\Omega_{\mathcal{Y}/\mathbb{D}}$ is torsion free and $n \geq e$, then the kernel of the map

$$D_\gamma^{(n)} : \mathrm{Hom}_{\mathcal{O}}(\gamma^*\Omega_{\mathcal{Y}/\mathbb{D}}, \mathcal{O}/\mathfrak{m}^{n+1}) \to \mathrm{Hom}_{\mathcal{O}}((\gamma H)^*\Omega_{\mathcal{X}/\mathbb{D}}, \mathcal{O}/\mathfrak{m}^{n+1})$$

induced by the derivative of H is contained in $\mathrm{Hom}_{\mathcal{O}}(\gamma^*\Omega_{\mathcal{Y}/\mathbb{D}}, \mathfrak{m}^{n+1-e}/\mathfrak{m}^{n+1})$, can be identified with $\mathrm{Hom}_{\mathcal{O}}(\gamma^*\Omega_{\mathcal{Y}/\mathcal{X}}, \mathcal{O}/\mathfrak{m}^{n+1})$, and is of length e. The proof of Theorem 3.2 now rests on the

Key lemma 9.2. — *Suppose $\mathcal{Y}/\mathbb{D}$ smooth and let $A \subset \mathcal{Y}_\infty$ be a stable subset of level l: $A = \pi_l^{-1}\pi_l(A)$. Assume that $H\big|_A$ is injective and that $\mathrm{ord}_{\mathcal{J}_H}\big|_A$ is constant equal to $e < \infty$. If $n \geq \sup\{2e, l+e, \mathrm{ord}_{\mathcal{J}(\mathcal{X}/\mathbb{D})}\big|_{HA}\}$, then $H_n : \pi_n A \to H_n\pi_n A$ has the structure of affine-linear bundle of dimension e.*

Proof. — Let $\gamma \in A$ and put $x := \gamma(o)$, $y := H(x)$. Suppose $\gamma' \in A$ is such that $H\gamma'$ and $H\gamma$ have the same n-jet. We first show that γ and γ' have the same $(n-e)$-jet. We do this by constructing a $\gamma_1 \in \mathcal{Y}_\infty$ (by successive approximation) with the same $(n-e)$-jet as γ and with $H\gamma_1 = H\gamma'$. Since $n - e \geq l$, we will have $\gamma_1 \in A$ and our injectivity assumption then implies $\gamma_1 = \gamma'$.

The difference $(H\gamma)^* - (H\gamma')^*$ defines a $\mathcal{O}$-derivation $\mathcal{O}_{\mathcal{X},x} \to \mathfrak{m}^{n+1}/\mathfrak{m}^{2(n+1)}$ over γ^* and hence a $\tilde{v} \in \mathrm{Hom}_{\mathcal{O}}((\gamma H)^*\Omega_{\mathcal{X}/\mathbb{D}}, \mathfrak{m}^{n+1}/\mathfrak{m}^{2n+2})$. Since $n \geq \mathrm{ord}_{H\gamma}\mathcal{J}(\mathcal{X}/\mathbb{D})$, this element annihilates the torsion of $(\gamma H)^*\Omega_{\mathcal{X}/\mathbb{D}}$. This is then also true for its reduction modulo $\mathfrak{m}^{n+2}$ and it follows from the fact that $n \geq e$ that this reduction is of the form $D_\gamma^{(n+1)}(u)$ for some $u \in \mathrm{Hom}_{\mathcal{O}}(\gamma^*\Omega_{\mathcal{Y}/\mathbb{D}}, \mathfrak{m}^{n-e+1}/\mathfrak{m}^{n+2})$. Regard u as a $\mathcal{O}$-derivation $\mathcal{O}_{\mathcal{Y},y} \to \mathfrak{m}^{n-e+1}/\mathfrak{m}^{n+2}$ and let $\gamma_1 \in \mathcal{Y}_\infty$ be such that $\gamma_1^* - \gamma^*$ represents u. Then $\pi_{n-e}(\gamma_1) = \pi_{n-e}(\gamma)$ and $\pi_{n+1}(H\gamma_1) = \pi_{n+1}(H\gamma)$. Replace γ by γ_1 and continue with induction on n.

So $(\gamma')^* - \gamma^*$ defines a $\mathcal{O}$-derivation $\mathcal{O}_{\mathcal{X},x} \to \mathfrak{m}^{n-e+1}/\mathfrak{m}^{2(n-e+1)}$ and hence a $\mathcal{O}$-derivation $\mathcal{O}_{\mathcal{X},x} \to \mathfrak{m}^{n-e+1}/\mathfrak{m}^{n+1}$ (because $n \geq 2e$). The latter is zero if and only if $\pi_n(\gamma') = \pi_n(\gamma)$. This proves that the fiber of $H_n\big|_{\pi_n A}$ through $\pi_n(\gamma)$ is an affine space over the kernel of $D_\gamma^{(n)}$, $\mathrm{Hom}_{\mathcal{O}}(\gamma^*\Omega_{\mathcal{Y}/\mathcal{X}}, \mathcal{O}/\mathfrak{m}^{n+1})$, which has length e.

The last assertion is easy. □

Proof of 3.2. — It is enough to prove this for A stable. In that case the theorem follows in a straightforward manner from Lemma 9.2. □

REFERENCES

[1] D. Abramovich, K. Karu, K. Matsuki, J. Włodarczyk, *Torification and Factorization of Birational Maps*, 30 p., `arXiv.org/abs/math/9904135`.

[2] N. A'Campo, *La fonction zêta d'une monodromie*, Comment. Math. Helv. **50** (1975), 233–248.

[3] G. Anderson, *Cyclotomy and an extension of the Taniyama group*, Comp. Math. **57** (1986), 153–217.

[4] V. Batyrev, *Birational Calabi–Yau n-folds have equal Betti numbers*, in New trends in algebraic geometry, Klaus Hulek *et al.*, eds., CUP, 1999, 1–11.

[5] V. Batyrev, *Stringy Hodge numbers of varieties with Gorenstein canonical singularities*, in Integrable systems and algebraic geometry (Kobe/Kyoto, 1997), 1–32, World Sci. Publishing, River Edge, NJ, 1998.

[6] V. Batyrev, *Non-Archimedean integrals and stringy Euler numbers of log-terminal pairs*, J. Eur. Math. Soc. **1** (1999), 5–33.

[7] V. Batyrev and D. Dais, *Strong McKay correspondence, string-theoretic Hodge numbers and mirror symmetry*, Topology **35** (1996), 901–929.

[8] A. Craw, *An introduction to motivic integration*, 23 pp., `arXiv.org/abs/math/9911179`.

[9] P. Deligne, *Intégration sur un cycle évanescent.* Invent. Math. **76** (1984), no. 1, 129–143.

[10] J. Denef, *Local zeta functions and Euler characteristics.* Duke Math. J. **63** (1991), no. 3, 713–721.

[11] J. Denef, *Report on Igusa's local zeta function*, in *Séminaire Bourbaki, volume 1990/91*, exposé 741. Astérisque **201-202-203** (1991), 359–386.

[12] J. Denef, F. Loeser, *Caractéristiques d'Euler-Poincaré, fonctions zêta locales et modifications analytiques*, J. Amer. Math. Soc. **5** (1992), 705–720.

[13] J. Denef, F. Loeser, *Motivic Igusa zeta functions*, J. Algebraic Geom., **7** (1998), 505–537.

[14] J. Denef, F. Loeser, *Germs of arcs on singular algebraic varieties and motivic integration*, Invent. Math., **135** (1999), 201–232.

[15] J. Denef, F. Loeser, *Motivic exponential integrals and a motivic Thom-Sebastiani Theorem*, Duke Math. J., **99** (1999), 285–309.

[16] J. Denef, F. Loeser, *Motivic integration, quotient singularities and the McKay correspondence*, 20 p., `arXiv.org/abs/math/9903187`.

[17] J. Denef, F. Loeser, *Definable sets, motives and P-adic integrals*, 45 p., `arXiv.org/abs/math/9910107`.

[18] J. Denef, F. Loeser, *Lefschetz numbers of the monodromy and truncated arcs*, 10 p., `arXiv.org/abs/math/0001105`.

[19] J. Denef, F. Loeser, *Geometry on arc spaces of algebraic varieties*, 22 p., arXiv.org/abs/math/0006050.

[20] M. Greenberg, *Schemata over local rings*, Ann. Math. **73** (1961), 624–648.

[21] M. Greenberg, *Rational points in discrete valuation rings*, Publ. Math. I.H.E.S. **31** (1966), 59–64.

[22] M. Kapranov, *The elliptic curve in the S-duality theory and Eisenstein series for Kac-Moody groups*, 41 pp., arXiv.org/abs/math/0001005.

[23] M. Kontsevich, Lecture at Orsay (December 7, 1995).

[24] J. Nash Jr., *Arc structure of singularities*, Duke Math. J., **81** (1995), 31–38.

[25] J. Pas, *Uniform p-adic cell decomposition and local zeta functions*, J. f. d. reine u. angew. Math. **399** (1989), 137–172.

[26] M. Reid, *La correspondance de McKay* (en anglais), in *Séminaire Bourbaki*, exposé 867, Novembre 1999, 20 p., in this volume and at arXiv.org/abs/math/9911195.

[27] T. Shioda, T. Katsura, *On Fermat varieties*, Tohuku math. J. **31** (1979), 97–115.

[28] J. Steenbrink, *The spectrum of hypersurface singularities*, in *Théorie de Hodge, Luminy 1987*, Astérisque **179-180** (1989), 163–184.

[29] A. Varchenko, *Asymptotic Hodge structure in the vanishing cohomology*, Math. USSR Izvestija **18** (1982), 469-512.

[30] W. Veys, *The topological zeta function associated to a function on a normal surface germ*, Topology **38** (1999), no. 2, 439–456.

[31] W. Veys, *Zeta functions and 'Kontsevich invariants' on singular varieties*, 27 pp., arXiv.org/abs/math.AG/0003025.

Eduard LOOIJENGA

Mathematisch Instituut
Universiteit Utrecht
Postbus 80.010, NL-3508 TA Utrecht
Nederland
E-mail : looijeng@math.uu.nl

Séminaire BOURBAKI Juin 2000
52^{e} année, 1999-2000, n^{o} 875, p. 299 à 339

VERTEX ALGEBRAS AND ALGEBRAIC CURVES

by **Edward FRENKEL**

1. INTRODUCTION

Vertex operators appeared in the early days of string theory as local operators describing propagation of string states. Mathematical analogues of these operators were discovered in representation theory of affine Kac-Moody algebras in the works of Lepowsky–Wilson [LW] and I. Frenkel–Kac [FK]. In order to formalize the emerging structure and motivated in particular by the I. Frenkel–Lepowsky–Meurman construction of the Moonshine Module of the Monster group, Borcherds gave the definition of vertex algebra in [B1]. The foundations of the theory were subsequently laid down in [FLM, FHL]; in particular, it was shown in [FLM] that the Moonshine Module indeed possessed a vertex algebra structure. In the meantime, Belavin, Polyakov and Zamolodchikov [BPZ] initiated the study of two-dimensional conformal field theory (CFT). Vertex algebras can be seen in retrospect as the mathematical equivalent of the central objects of CFT called the chiral symmetry algebras. Moreover, the key property of associativity of vertex algebras is equivalent to the property of operator product expansion in CFT, which goes back to the pioneering works of Polyakov and Wilson. Thus, vertex algebras may be thought of as the mathematical language of two-dimensional conformal field theory.

Vertex algebras have turned out to be extremely useful in many areas of mathematics. They are by now ubiquitous in representation theory of infinite-dimensional Lie algebras. They have also found applications in such fields as algebraic geometry, theory of finite groups, modular functions and topology. Recently Beilinson and Drinfeld have introduced a remarkable geometric version of vertex algebras which they called chiral algebras [BD3]. Chiral algebras give rise to some novel concepts and techniques which are likely to have a profound impact on algebraic geometry.

In this talk we review the theory of vertex algebras with a particular emphasis on their algebro-geometric interpretation and applications. We start in §2 with the axiomatic definition of vertex algebra, which is somewhat different from, but equivalent to Borcherds' original definition (see [FKRW, K2]). We then discuss some of their most important properties and give the first examples, which come from infinite-dimensional Lie algebras, such as Heisenberg, affine Kac-Moody and the Virasoro algebra. More unconventional examples of vertex algebras which are not generated by a Lie algebra, such as the $\mathcal{W}$–algebras, are reviewed in §3.

In §4 we explain how to make vertex algebras coordinate independent, thus effectively getting rid of the formal variable omnipresent in the standard algebraic approach. This is achieved by attaching to each conformal vertex algebra a vector bundle with a flat connection on a formal disc, equipped with an intrinsic operation. The formal variable is restored when we choose a coordinate on the disc; the independence of the operation from the choice of coordinate follows from the fact that the group of changes of coordinates is an "internal symmetry" of the vertex algebra. Once we obtain a coordinate independent object, we can give a rigorous definition of the space of conformal blocks associated to a conformal vertex algebra and an algebraic curve X (see §5). From the physics point of view, conformal blocks give rise to "chiral correlation functions" on powers of X with singularities along the diagonals.

As we vary the curve X and other data on X reflecting the internal symmetry of a vertex algebra (such as G–bundles), the corresponding spaces of coinvariants, which are the duals of the spaces of conformal blocks, combine into a sheaf on the relevant moduli space. This sheaf carries the structure of a (twisted) $\mathcal{D}$–module, as explained in §6. One can gain new insights into the structure of moduli spaces from the study of these $\mathcal{D}$–modules. For instance, one obtains a description of the formal deformation spaces of the complex structure or a G–bundle on a curve in terms of certain sheaves on the symmetric powers of the curve [BG, BD2, Gi]. Thus, vertex algebras appear as the local objects controlling deformations of curves with various extra structures. This raises the possibility that more exotic vertex algebras, such as $\mathcal{W}$–algebras, also correspond to some still unknown moduli spaces.

Finally, in §7 we discuss the relation between vertex algebras and the Beilinson–Drinfeld chiral algebras. We review briefly the description of chiral algebras as factorization algebras, i.e., sheaves on the Ran space of finite subsets of a curve, satisfying certain compatibilities. Using this description, Beilinson and Drinfeld have introduced the concept of chiral homology, which can be thought of as a derived functor of the functor of coinvariants.

The formalism of vertex and chiral algebras appears to be particularly suitable for the construction of the conjectural geometric Langlands correspondence between automorphic $\mathcal{D}$–modules on the moduli space of G–bundles on a smooth projective curve X over $\mathbb{C}$ and flat ${}^LG$–bundles on X, where G is a simple algebraic group and ${}^LG$ is the Langlands dual group (see [BD2]). The idea is to construct the automorphic $\mathcal{D}$–modules corresponding to flat ${}^LG$–bundles as the sheaves of coinvariants (or, more generally, chiral homology) of a suitable vertex algebra. We mention below two such constructions (both due to Beilinson and Drinfeld): in one of them the relevant vertex algebra is associated to an affine Kac-Moody algebra of critical level (see §6.5), and in the other it is the chiral Hecke algebra (see §7).

Another application of vertex algebras to algebraic geometry is the recent construction by Malikov, Schechtman and Vaintrob [MSV] of a sheaf of vertex superalgebras on an arbitrary smooth algebraic variety, called the chiral deRham complex, which is reviewed in §6.6.

Vertex algebras form a vast and rapidly growing subject, and it is impossible to cover all major results (or even give a comprehensive bibliography) in one survey. For example, because of lack of space, I have not discussed such important topics as the theory of conformal algebras [K2, K3] and their chiral counterpart, Lie* algebras [BD2]; quantum deformations of vertex algebras [B3, EK, FR]; and the connection between vertex algebras and integrable systems.

Most of the material presented below (note that §§4–6 contain previously unpublished results) is developed in the forthcoming book [BF].

I thank A. Beilinson for answering my questions about chiral algebras and D. Ben-Zvi for helpful comments on the draft of this paper. The support from the Packard Foundation and the NSF is gratefully acknowledged.

2. DEFINITION AND FIRST PROPERTIES OF VERTEX ALGEBRAS

2.1. Let R be a $\mathbb{C}$–algebra. An R–valued *formal power series* (or formal distribution) in variables $z_1, z_2, \ldots, z_n$ is an arbitrary infinite series

$$A(z_1, \ldots, z_n) = \sum_{i_1 \in \mathbb{Z}} \cdots \sum_{i_n \in \mathbb{Z}} a_{i_1,\ldots,i_n} z_1^{i_1} \cdots z_n^{i_n},$$

where each $a_{i_1,\ldots,i_n} \in R$. These series form a vector space denoted by $R[[z_1^{\pm 1}, \ldots, z_n^{\pm 1}]]$. If $P(z_1, \ldots, z_n) \in R[[z_1^{\pm 1}, \ldots, z_n^{\pm 1}]]$ and $Q(w_1, \ldots, w_m) \in R[[w_1^{\pm 1}, \ldots, w_m^{\pm 1}]]$, then their product is a well-defined element of $R[[z_1^{\pm 1}, \ldots, z_n^{\pm 1}, w_1^{\pm 1}, \ldots, w_m^{\pm 1}]]$. In general, a product of two formal power series in the same variables does not make sense, but the product of a formal power series by a polynomial (i.e., a series, such that $a_{i_1,\ldots,i_n} = 0$ for all but finitely many indices) is always well-defined.

Let V be a $\mathbb{Z}_+$–graded vector space $V = \oplus_{n=0}^{\infty} V_n$ with finite-dimensional homogeneous components. An endomorphism T of V is called homogeneous of degree n if $T(V_m) \subset V_{n+m}$. Denote by $\operatorname{End} V$ the vector space of linear endomorphisms of V, which are finite linear combinations of homogeneous endomorphisms. This is a $\mathbb{Z}$–graded algebra.

A *field* of conformal dimension $\Delta \in \mathbb{Z}_+$ is an $\operatorname{End} V$–valued formal power series in z,

$$\phi(z) = \sum_{j \in \mathbb{Z}} \phi_j z^{-j-\Delta}$$

where each ϕ_j is a homogeneous linear endomorphism of V of degree $-j$. Two fields $\phi(z)$ and $\psi(z)$ are called mutually *local* if there exists $N \in \mathbb{Z}_+$, such that

$$(z-w)^N [\phi(z), \psi(w)] = 0 \tag{1}$$

(as an element of $\operatorname{End} V[[z^{\pm 1}, w^{\pm 1}]]$).

Now we can formulate the axioms of vertex algebra.

Definition 2.1. — *A* vertex algebra *is a collection of data:*

– space of states, *a $\mathbb{Z}_+$–graded vector space $V = \oplus_{n=0}^{\infty} V_n$, with* $\dim(V_n) < \infty$*;*
– vacuum vector $|0\rangle \in V_0$*;*
– shift operator $T : V \to V$ *of degree one;*
– vertex operation, *a linear map $Y(\cdot, z) : V \to \operatorname{End} V[[z, z^{-1}]]$ taking each $A \in V_n$ to a field of conformal dimension n.*

These data are subject to the following axioms:

– (vacuum axiom) $Y(|0\rangle, z) = \mathrm{Id}_V$. *Furthermore, for any $A \in V$ we have $Y(A,z)|0\rangle \in A + zV[[z]]$ (i.e., $Y(A,z)|0\rangle$ has a well–defined value at $z = 0$, which is equal to A).*
– (translation axiom) *For any $A \in V$, $[T, Y(A,z)] = \partial_z Y(A,z)$ and $T|0\rangle = 0$.*
– (locality axiom) *All fields $Y(A,z)$ are mutually local with each other.*

Remark 2.2. — It is easy to adopt the above definition to the supercase (see, e.g., [K2]). Then V is a superspace, and the above structures and axioms should be modified appropriately; in particular, we need to replace the commutator by the supercommutator in the definition of locality. Then we obtain the definition of vertex superalgebra. We mostly consider below purely even vertex superalgebras, but general vertex superalgebras are very important; for instance, $N = 2$ superconformal vertex superalgebras appear in physical models relevant to mirror symmetry.

The above conditions on V can be relaxed: it suffices to require that for any $A \in V$, $v \in V$, we have: $A_n \cdot v = 0$ for n large enough. It is not necessary to require that $\dim V_n < \infty$ and even that V is graded (in that case however one needs to be careful when dealing with dual spaces). We impose the above stronger conditions in order to simplify the exposition.

It is straightforward to define homomorphisms between vertex algebras, vertex subalgebras, ideals and quotients. If V_1 and V_2 are two vertex algebras, then $V_1 \otimes V_2$ carries a natural vertex algebra structure.

2.2. Example: commutative vertex algebras

Let V be a $\mathbb{Z}$–graded commutative algebra (with finite-dimensional homogeneous components) with a unit and a derivation T of degree 1. Then V carries a canonical structure of vertex algebra. Namely, we take the unit of V as the vacuum vector $|0\rangle$, and define the operation Y by the formula

$$Y(A,z) = \sum_{n\geq 0} \frac{1}{n!} m(T^n A) z^n,$$

where for $B \in V$, $m(B)$ denotes the operator of multiplication by B on V. It is straightforward to check that all axioms of vertex algebra are satisfied; in fact, the locality axiom is satisfied in a strong sense: for any $A, B \in V$, we have: $[Y(A,z), Y(B,w)] = 0$ (so we have $N = 0$ in formula (1)).

Conversely, let V be a vertex algebra, in which locality holds in the strong sense (we call such a vertex algebra *commutative*). Then locality and vacuum axioms imply that $Y(A,z) \in \operatorname{End} V[[z]]$ for all $A \in V$ (i.e., there are no terms with negative powers of z). Denote by Y_A the endomorphism of V, which is the constant term of $Y(A,z)$, and define a bilinear operation $\circ$ on V by setting $A \circ B = Y_A \cdot B$. By construction, $Y_A Y_B = Y_B Y_A$ for all $A, B \in V$. This implies commutativity and associativity of $\circ$ (see, e.g., [Li]). We obtain a commutative and associative product on V, which respects the $\mathbb{Z}$–gradation. Furthermore, the vacuum vector $|0\rangle$ is a unit, and the operator T is a derivation with respect to this product. Thus, we see that the notion of commutative vertex algebra is equivalent to that of $\mathbb{Z}$–graded commutative associative algebra with a unit and a derivation of degree 1.

Remark 2.3. — The operator T may be viewed as the generator of infinitesimal translations on the formal additive group with coordinate z. Therefore a commutative vertex algebra is the same as a commutative algebra equipped with an action of the formal additive group. Thus one may think of general vertex algebras as meromorphic generalizations of commutative algebras with an action of the formal additive group. This point of view has been developed by Borcherds [B3], who showed that vertex algebras are “singular commutative rings” in a certain category. He has also considered generalizations of vertex algebras, replacing the formal additive group by other (formal) groups or Hopf algebras.

2.3. Non-commutative example: Heisenberg vertex algebra

Consider the space $\mathbb{C}((t))$ of Laurent series in one variable as a commutative Lie algebra. We define the Heisenberg Lie algebra $\mathcal{H}$ as follows. As a vector space, it is the direct sum of the space of formal Laurent power series $\mathbb{C}((t))$ and a one-dimensional space $\mathbb{C}\mathbf{1}$, with the commutation relations

$$[f(t), g(t)] = -\operatorname{Res} f dg \mathbf{1}, \qquad [\mathbf{1}, f(t)] = 0. \tag{2}$$

Here Res denotes the (-1)st Fourier coefficient of a Laurent series. Thus, $\mathcal{H}$ is a one-dimensional central extension of the commutative Lie algebra $\mathbb{C}((t))$. Note that the relations (2) are independent of the choice of local coordinate t. Thus we may define a Heisenberg Lie algebra canonically as a central extension of the space of functions on a punctured disc without a specific choice of formal coordinate t.

The Heisenberg Lie algebra $\mathcal{H}$ is topologically generated by the generators $b_n = t^n$, $n \in \mathbb{Z}$, and the central element $\mathbf{1}$, and the relations between them read

$$[b_n, b_m] = n\delta_{n,-m}\mathbf{1}, \qquad [\mathbf{1}, b_n] = 0. \tag{3}$$

The subspace $\mathbb{C}[[t]] \oplus \mathbf{1}$ is a commutative Lie subalgebra of $\mathcal{H}$. Let π be the $\mathcal{H}$–module induced from the one-dimensional representation of $\mathbb{C}[[t]] \oplus \mathbf{1}$, on which $\mathbb{C}[[t]]$ acts by 0 and $\mathbf{1}$ acts by 1. Equivalently, we may describe π as the polynomial algebra $\mathbb{C}[b_{-1}, b_{-2}, \dots]$ with $b_n, n < 0$, acting by multiplication, b_0 acting by 0, and $b_n, n > 0$, acting as $n\dfrac{\partial}{\partial b_{-n}}$. The operators b_n with $n < 0$ are known in this context as creation operators, since they "create the state b_n from the vacuum 1", while the operators b_n with $n \geq 0$ are the annihilation operators, repeated application of which will kill any vector in π. The module π is called the Fock representation of $\mathcal{H}$.

We wish to endow π with a structure of vertex algebra. This involves the following data (Definition 2.1):

– $\mathbb{Z}_+$ grading: $\deg b_{j_1} \dots b_{j_k} = -\sum_i j_i$.

– Vacuum vector: $|0\rangle = 1$.

– The shift operator T defined by the rules: $T \cdot 1 = 0$ and $[T, b_i] = -ib_{i-1}$.

We now need to define the fields $Y(A, z)$. To the vacuum vector $|0\rangle = 1$, we are required to assign $Y(|0\rangle, z) = \mathrm{Id}$. The key definition is that of the field $b(z) = Y(b_{-1}, z)$, which generates π in an appropriate sense. Since $\deg(b_{-1}) = 1$, $b(z)$ needs to have conformal dimension one. We set

$$b(z) = \sum_{n\in\mathbb{Z}} b_n z^{-n-1}.$$

The Reconstruction Theorem stated below implies that once we have defined $Y(b_{-1}, z)$, there is (at most) a unique way to extend this definition to other vectors of π. This is not very surprising, since as a commutative algebra with derivation T, π is freely generated by b_{-1}. Explicitly, the fields corresponding to other elements of π are constructed by the formula

$$Y(b_{-n_1} b_{-n_2} \dots b_{-n_k}, z) = \frac{1}{(n_1 - 1)! \dots (n_k - 1)!} {:}\partial_z^{n_1 - 1} b(z) \dots \partial_z^{n_k - 1} b(z){:}. \tag{4}$$

The columns in the right hand side of the formula stand for the *normally ordered product*, which is defined as follows. First, let ${:}b_{n_1} \dots b_{n_k}{:}$ be the monomial obtained from $b_{n_1} \dots b_{n_k}$ by moving all b_{n_i} with $n_i < 0$ to the left of all b_{n_j} with $n_j \geq 0$ (in other words, moving all "creation operators" $b_n, n < 0$, to the left and all "annihilation operators" $b_n, n \leq 0$, to the right). The important fact that makes this definition correct is that the operators b_n with $n < 0$ (resp., $n \geq 0$) commute with each other, hence it does not matter how we order the creation (resp., annihilation) operators among themselves (this property does not hold for more general vertex algebras, and in those cases one needs to use a different definition of normally ordered product, which is given below). Now define $:\partial_z^{m_1} b(z) \dots \partial_z^{m_k} b(z):$ as the power series in z obtained from the ordinary product $\partial_z^{m_1} b(z) \dots \partial_z^{m_k} b(z)$ by replacing each term $b_{n_1} \dots b_{n_k}$ with $:b_{n_1} \dots b_{n_k}:$.

PROPOSITION 2.4. — *The Fock representation π with the structure given above satisfies the axioms of vertex algebra.*

The most difficult axiom to check is locality. The Reconstruction Theorem stated below allows us to reduce it to verifying locality property for the "generating" field $b(z)$. This is straightforward from the commutation relations (3):

$$[b(z), b(w)] = \sum_{n,m\in\mathbb{Z}} [b_n, b_m] z^{-n-1} w^{-m-1} = \sum_{n\in\mathbb{Z}} [b_n, b_{-n}] z^{-n-1} w^{n-1}$$
$$= \sum_{n\in\mathbb{Z}} n z^{-n-1} w^{n-1} = \partial_w \delta(z-w),$$

where we use the notation $\delta(z-w) = \sum_{m\in\mathbb{Z}} w^m z^{-m-1}$.

Now $\delta(z-w)$ has the property that $(z-w)\delta(z-w) = 0$. In fact, the annihilator of the operator of multiplication by $(z-w)$ in $R[[z^{\pm 1}, w^{\pm 1}]]$ equals $R[[w^{\pm 1}]] \cdot \delta(z-w)$ (note that the product of $\delta(z-w)$ with any formal power series in z or in w is well-defined). More generally, the annihilator of the operator of multiplication by $(z-w)^N$ in $R[[z^{\pm 1}, w^{\pm 1}]]$ equals $\oplus_{n=0}^{N-1} R[[w^{\pm 1}]] \cdot \partial_w^n \delta(z-w)$ (see [K1]). This implies in particular that $(z-w)^2 [b(z), b(w)] = 0$, and hence the field $b(z)$ is local with itself. Proposition 2.4 now follows from the Reconstruction Theorem below.

2.4. Reconstruction Theorem

We state a general result, which provides a "generators–and–relations" approach to the construction of vertex algebras. Let V be a $\mathbb{Z}_+$–graded vector space, $|0\rangle \in V_0$ a non-zero vector, and T a degree 1 endomorphism of V. Let S be a countable ordered set and $\{a^\alpha\}_{\alpha\in S}$ be a collection of homogeneous vectors in V, with a^α of degree Δ_α. Suppose we are also given fields

$$a^\alpha(z) = \sum_{n\in\mathbb{Z}} a_n^\alpha z^{-n-\Delta_\alpha}$$

on V, such that the following hold:

– For all α, $a^\alpha(z)|0\rangle \in a^\alpha + zV[[z]]$;
– $T|0\rangle = 0$ and $[T, a^\alpha(z)] = \partial_z a^\alpha(z)$ for all α;
– All fields $a^\alpha(z)$ are mutually local;

– V is spanned by lexicographically ordered monomials

$$a^{\alpha_1}_{-\Delta_{\alpha_1}-j_1}\dots a^{\alpha_m}_{-\Delta_{\alpha_m}-j_m}|0\rangle$$

(including $|0\rangle$), where $j_1 \geq j_2 \geq \dots \geq j_m \geq 0$, and if $j_i = j_{i+1}$, then $\alpha_i \geq \alpha_{i+1}$ with respect to the order on the set S.

THEOREM 2.5 ([FKRW]). — *Under the above assumptions, the above data together with the assignment*

$$Y(a^{\alpha_1}_{-\Delta_{\alpha_1}-j_1}\dots a^{\alpha_m}_{-\Delta_{\alpha_m}-j_m}|0\rangle, z) = \frac{1}{j_1!\dots j_m!}{:}\partial_z^{j_1} a^{\alpha_1}(z)\dots\partial_z^{j_m} a^{\alpha_m}(z){:}, \tag{5}$$

define a vertex algebra structure on V.

Here we use the following general definition of the normally ordered product of fields. Let $\phi(z), \psi(w)$ be two fields of respective conformal dimensions Δ_ϕ, Δ_ψ and Fourier coefficients ϕ_n, ψ_n. The normally ordered product of $\phi(z)$ and $\psi(z)$ is by definition the formal power series

$${:}\phi(z)\psi(z){:} = \sum_{n\in\mathbb{Z}}\left(\sum_{m\leq-\Delta_\phi}\phi_m\psi_{n-m} + \sum_{m>-\Delta_\phi}\psi_{n-m}\phi_m\right)z^{-n-\Delta_\phi-\Delta_\psi}.$$

This is a field of conformal dimension $\Delta_\phi + \Delta_\psi$. The normal ordering of more than two fields is defined recursively from right to left, so that by definition ${:}A(z)B(z)C(z){:} = {:}A(z)({:}B(z)C(z){:}){:}$. It is easy to see that in the case of the Heisenberg vertex algebra π this definition of normal ordering coincides with the one given in §2.3.

2.5. The meaning of locality

The product $\phi(z)\psi(w)$ of two fields is a well–defined $\operatorname{End} V$–valued formal power series in $z^{\pm1}$ and $w^{\pm1}$. Given $v \in V$ and $v^\vee \in V^\vee = \oplus_{n\geq0} V_n^*$, consider the matrix coefficient

$$\langle v^\vee|\phi(z)\psi(w)|v\rangle \in \mathbb{C}[[z^{\pm1}, w^{\pm1}]].$$

Since $V_n = 0$ for $n < 0$, we find by degree consideration that it belongs to $\mathbb{C}((z))((w))$, the space of formal Laurent series in w, whose coefficients are formal Laurent series in z. Likewise, $\langle v^\vee|\psi(w)\phi(z)|v\rangle$ belongs to $\mathbb{C}((w))((z))$.

As we have seen in §2.2, the condition that the fields $\phi(z)$ and $\psi(w)$ literally commute is too strong, and it essentially keeps us in the realm of commutative algebra. However, there is a natural way to relax this condition, which leads to the more general notion of locality. Let $\mathbb{C}((z, w))$ be the field of fractions of the ring $\mathbb{C}[[z, w]]$; its elements may be viewed as meromorphic functions in two formal variables. We have natural embeddings

$$\mathbb{C}((z))((w)) \longleftarrow \mathbb{C}((z, w)) \longrightarrow \mathbb{C}((w))((z)), \tag{6}$$

which are simply the inclusions of $\mathbb{C}((z,w))$ into its completions in two different topologies, corresponding to the z and w axes. For example, the images of $1/(z-w) \in \mathbb{C}((z,w))$ in $\mathbb{C}((z))((w))$ and in $\mathbb{C}((w))((z))$ are equal to

$$\delta(z-w)_- = \sum_{n\geq 0} w^n z^{-n-1}, \qquad -\delta(z-w)_+ = -\frac{1}{z}\sum_{n<0} w^n z^{-n-1},$$

respectively.

Now we can relax the condition of commutativity of two fields by requiring that for any $v \in V$ and $v^\vee \in V^\vee$, the matrix elements $\langle v^\vee|\phi(z)\psi(w)|v\rangle$ and $\langle v^\vee|\psi(w)\phi(z)|v\rangle$ are the images of the same element $f_{v^\vee,v}$ of $\mathbb{C}((z,w))$ in $\mathbb{C}((z))((w))$ and $\mathbb{C}((w))((z))$, respectively. In the case of vertex algebras, we additionally require that

$$f_{v^\vee,v} \in \mathbb{C}[[z,w]][z^{-1},w^{-1},(z-w)^{-1}]$$

for all $v \in V, v^\vee \in V^\vee$, and that the order of pole of $f_{v^\vee,v}$ is universally bounded, i.e., there exists $N \in \mathbb{Z}_+$, such that $(z-w)^N f_{v^\vee,v} \in \mathbb{C}[z^{\pm 1},w^{\pm 1}]$ for all $v, v^\vee$. The last condition is equivalent to the condition of locality given by formula (1).

From the analytic point of view, locality of the fields $\phi(z)$ and $\phi(w)$ means that the matrix element $\langle v^\vee|\phi(z)\psi(w)|v\rangle$ is well-defined in the domain $|z| > |w|$ whereas $\langle v^\vee|\psi(w)\phi(z)|v\rangle$ is well-defined in the domain $|w| > |z|$, but both can be analytically continued to the same function $f_{v^\vee,v}(z,w)$. Then their commutator (considered as a distribution) is the difference between boundary values of holomorphic functions, and hence is a delta-like distribution supported on the diagonal. In the simplest case, this amounts to the formula $\delta(z-w) = \delta(z-w)_- + \delta(z-w)_+$, which is a version of the Sokhotsky–Plemelj formula well-known in complex analysis.

2.6. Associativity

Now we state the "associativity" property of vertex algebras. Consider the $\operatorname{End} V$–valued formal power series

$$Y(Y(A,z-w)B,w) = \sum_{n\in\mathbb{Z}} Y(A_n\cdot B,w)(z-w)^{-n-\Delta_A}$$

in w and $z-w$. By degree reasons, this is an element of $\operatorname{End} V((w))((z-w))$ (i.e., the order of pole at $z=w$ is bounded). By mapping $(z-w)^{-j}$ to $(\delta(z-w)_-)^j$ (its expansion in w/z), we obtain an embedding of $\operatorname{End} V((w))((z-w))$ into $\operatorname{End} V[[z^{\pm 1},w^{\pm 1}]]$.

Proposition 2.6 ([FHL, K2]). — *Any vertex algebra V satisfies the following associativity property: for any $A,B \in V$ we have the equality of formal power series*

$$Y(A,z)Y(B,w) = Y(Y(A,z-w)B,w) = \sum_{n\in\mathbb{Z}} \frac{Y(A_n\cdot B,w)}{(z-w)^{n+\Delta_A}}, \tag{7}$$

where by the right hand side we understand its image in $\operatorname{End} V[[z^{\pm 1},w^{\pm 1}]]$.

Formula (7) is called the operator product expansion (OPE). From the physics point of view, it manifests the important property in quantum field theory that the product of two fields at nearby points can be expanded in terms of other fields and the small parameter $z-w$. From the analytic point of view, this formula expresses the fact that the matrix elements of the left and right hand sides of the formula, well-defined in the appropriate domains, can be analytically continued to the same rational functions in z, w, with poles only at $z=0, w=0$, and $z=w$.

For example, in the case of Heisenberg algebra, we obtain:

$$b(z)b(w) = \frac{1}{(z-w)^2} + \sum_{n\geq 0} \frac{1}{n!} :\partial_w^n b(w) b(w): (z-w)^n. \tag{8}$$

Using the Cauchy formula, one can easily extract the commutation relations between the Fourier coefficients of the fields $Y(A,z)$ and $Y(B,w)$ from the singular (at $z=w$) terms of their OPE. The result is

$$[A_m, B_k] = \sum_{n>-\Delta_A} \begin{pmatrix} m+\Delta_A-1 \\ n+\Delta_A-1 \end{pmatrix} (A_n \cdot B)_{m+k}. \tag{9}$$

2.7. Correlation functions

Formula (7) and the locality property have the following "multi–point" generalization. Let V^* be the space of all linear functionals on V.

Proposition 2.7 ([FHL]). — *Let $A_1, \ldots, A_n \in V$. For any $v \in V, \varphi \in V^*$, and any permutation σ on n elements, the formal power series in $z_1, \ldots, z_n$,*

$$\varphi\left(Y(A_{\sigma(1)}, z_{\sigma(1)}) \ldots Y(A_{\sigma(n)}, z_{\sigma(n)})|0\rangle\right) \tag{10}$$

is the expansion in $\mathbb{C}((z_{\sigma(1)}))\ldots((z_{\sigma(n)}))$ of an element $f^\sigma_{A_1,\ldots,A_n}(z_1,\ldots,z_n)$ of $\mathbb{C}[[z_1,\ldots,z_n]][(z_i-z_j)^{-1}]_{i\neq j}$ which satisfies the following properties: it does not depend on σ (so we suppress it in the notation);

$$f_{A_1,\ldots,A_n}(z_1,\ldots,z_n) = f_{Y(A_i,z_i-z_j)A_j,A_1,\ldots,\widehat{A_i},\ldots,\widehat{A_j},\ldots,A_n}(z_j,z_1,\ldots,\widehat{z_i},\ldots,\widehat{z_j},\ldots,z_n)$$

for all $i \neq j$; and $\partial_{z_i} f_{A_1,\ldots,A_n}(z_1,\ldots,z_n) = f_{A_1,\ldots,TA_i,\ldots,A_n}(z_1,\ldots,A_n)$.

Thus, to each $\varphi \in V^*$ we can attach a collection of matrix elements (10), the "n–point functions" on $\operatorname{Spec}\mathbb{C}[[z_1,\ldots,z_n]][(z_i-z_j)^{-1}]_{i\neq j}$. They satisfy a symmetry condition, a "bootstrap" condition (which describes the behavior of the n–point function near the diagonals in terms of $(n-1)$–point functions) and a "horizontality" condition. We obtain a linear map from V^* to the vector space $\mathcal{F}_n$ of all collections $\{f_{A_1,\ldots,A_m}(z_1,\ldots,z_m), A_i \in V\}_{m=1}^n$ satisfying the above conditions. This map is actually an isomorphism for each $n \geq 1$. The inverse map $\mathcal{F}_n \to V^*$ takes $\{f_{A_1,\ldots,A_m}(z_1,\ldots,z_m), A_i \in V\}_{m=1}^n \in \mathcal{F}_n$ to the functional φ on V, defined by the formula $\varphi(A) = f_A(0)$. Thus, we obtain a "functional realization" of V^*. In the case when V is generated by fields such that the singular terms in their OPEs are linear combinations of the same fields and their derivatives, we can simplify this functional realization by considering only the n–point functions of the generating fields.

For example, in the case of the Heisenberg vertex algebra we consider for each $\varphi \in \pi^*$ the n–point functions

$$\omega_n(z_1, \ldots, z_n) = \varphi(b(z_1) \ldots b(z_n)|0\rangle). \tag{11}$$

By Proposition 2.7 and the OPE (8), these functions are symmetric and satisfy the bootstrap condition

$$\omega_n(z_1, \ldots, z_n) = \left(\frac{\omega_{r-2}(z_1, \ldots, \widehat{z_i}, \ldots, \widehat{z_j}, \ldots, z_n)}{(z_i - z_j)^2} + \text{regular} \right). \tag{12}$$

Let Ω_∞ be the vector space of infinite collections $(\omega_n)_{n \geq 0}$, where

$$\omega_n \in \mathbb{C}[[z_1, \ldots, z_n]][(z_i - z_j)^{-1}]_{i \neq j}$$

satisfy the above conditions. Using the functions (11), we obtain a map $\pi^* \to \Omega_\infty$. One can show that this map is an isomorphism.

3. MORE EXAMPLES

3.1. Affine Kac-Moody algebras

Let $\mathfrak{g}$ be a simple Lie algebra over $\mathbb{C}$. Consider the formal loop algebra $L\mathfrak{g} = \mathfrak{g}((t))$ with the obvious commutator. The affine algebra $\widehat{\mathfrak{g}}$ is defined as a central extension of $L\mathfrak{g}$. As a vector space, $\widehat{\mathfrak{g}} = L\mathfrak{g} \oplus \mathbb{C}K$, and the commutation relations read: $[K, \cdot] = 0$ and

$$[A \otimes f(t), B \otimes g(t)] = [A, B] \otimes f(t)g(t) + (\operatorname{Res}_{t=0} f dg (A, B))K. \tag{13}$$

Here $(\cdot, \cdot)$ is an invariant bilinear form on $\mathfrak{g}$ (such a form is unique up to a scalar multiple; we normalize it as in [K1] by the requirement that $(\alpha_{\max}, \alpha_{\max}) = 2$).

Consider the Lie subalgebra $\mathfrak{g}[[t]]$ of $L\mathfrak{g}$. If $f, g \in \mathbb{C}[[t]]$, then $\operatorname{Res}_{t=0} f dg = 0$. Hence $\mathfrak{g}[[t]]$ is a Lie subalgebra of $\widehat{\mathfrak{g}}$. Let $\mathbb{C}_k$ be the one–dimensional representation of $\mathfrak{g}[[t]] \oplus \mathbb{C}K$, on which $\mathfrak{g}[[t]]$ acts by 0 and K acts by the scalar $k \in \mathbb{C}$. We define the vacuum representation of $\widehat{\mathfrak{g}}$ of level k as the representation induced from $\mathbb{C}_k$: $V_k(\mathfrak{g}) = \operatorname{Ind}_{\mathfrak{g}[[t]] \oplus \mathbb{C}K}^{\widehat{\mathfrak{g}}} \mathbb{C}_k$.

Let $\{J^a\}_{a=1, \ldots \dim \mathfrak{g}}$ be a basis of $\mathfrak{g}$. Denote $J^a_n = J^a \otimes t^n \in L\mathfrak{g}$. Then $J^a_n, n \in \mathbb{Z}$, and K form a (topological) basis for $\widehat{\mathfrak{g}}$. By the Poincaré–Birkhoff–Witt theorem, $V_k(\mathfrak{g})$ has a basis of lexicographically ordered monomials of the form $J^{a_1}_{n_1} \ldots J^{a_m}_{n_m} v_k$, where v_k is the image of $1 \in \mathbb{C}_k$ in $V_k(\mathfrak{g})$, and all $n_i < 0$. We are now in the situation of the Reconstruction Theorem, and hence we obtain a vertex algebra structure on $V_k(\mathfrak{g})$, such that

$$Y(J^a_{-1} v_k, z) = J^a(z) := \sum_{n \in \mathbb{Z}} J^a_n z^{-n-1}.$$

The fields corresponding to other monomials are obtained by formula (5). Explicit computation shows that the fields $J^a(z)$ (and hence all other fields) are mutually local.

3.2. Virasoro algebra

The Virasoro algebra Vir is a central extension of the Lie algebra $\operatorname{Der}\mathbb{C}((t)) = \mathbb{C}((t))\frac{\partial}{\partial t}$. It has (topological) basis elements $L_n = -t^{n+1}, n \in \mathbb{Z}$, and C satisfying the relations that C is central and

$$[L_n, L_m] = (n-m)L_{n+m} + \frac{n^3-n}{12}\delta_{n,-m}C$$

(a coordinate independent version is given in §5.2). These relations imply that the Lie algebra $\operatorname{Der}\mathbb{C}[[t]] = \mathbb{C}[[t]]\frac{\partial}{\partial t}$ generated by $L_n, n \geq 0$, is a Lie subalgebra of Vir. Consider the induced representation $\mathrm{Vir}_c = \operatorname{Ind}^{Vir}_{\operatorname{Der}\mathbb{C}[[t]]\oplus\mathbb{C}C}\mathbb{C}_c$, where $\mathbb{C}_c$ is a one–dimensional representation, on which C acts as c and $\operatorname{Der}\mathbb{C}[[t]]$ acts by zero. By the Poincaré–Birkhoff–Witt theorem, Vir_c has a basis consisting of monomials of the form $L_{j_1}\dots L_{j_m}v_c, j_1 \leq j_2 \leq j_m \leq -2$, where v_c is the image of $1 \in \mathbb{C}_c$ in the induced representation. By Reconstruction Theorem, we obtain a vertex algebra structure on Vir_c, such that

$$Y(L_{-2}v_c, z) = T(z) := \sum_{n\in\mathbb{Z}} L_n z^{-n-2}.$$

Note that the translation operator is equal to L_{-1}.

The Lie algebra $\operatorname{Der}\mathcal{O}$ generates infinitesimal changes of coordinates. As we will see in §4, it is important to have this Lie algebra (and even better, the whole Virasoro algebra) act on a given vertex algebra by "internal symmetries". This property is formalized by the following definition.

Definition 3.1. — *A vertex algebra V is called* conformal, *of central charge $c \in \mathbb{C}$, if V contains a non-zero* conformal vector $\omega \in V_2$, *such that the corresponding vertex operator $Y(\omega, z) = \sum_{n\in\mathbb{Z}} L_n z^{-n-2}$ satisfies: $L_{-1} = T$, $L_0|_{V_n} = n\,\mathrm{Id}$, and $L_2\omega = \frac{1}{2}c|0\rangle$.*

These conditions imply that the operators $L_n, n \in \mathbb{Z}$, satisfy the commutation relations of the Virasoro algebra with central charge c. We also obtain a non-trivial homomorphism $\mathrm{Vir}_c \to V$, sending $L_{-2}v_c$ to ω.

For example, the vector $\frac{1}{2}b_{-1}^2 + \lambda b_{-2}$ is a conformal vector in π for any $\lambda \in \mathbb{C}$. The corresponding central charge equals $1 - 12\lambda^2$. The Kac-Moody vertex algebra $V_k(\mathfrak{g})$ is conformal if $k \neq -h^\vee$, where $h^\vee$ is the dual Coxeter number of $\mathfrak{g}$. The conformal vector is given by the Sugawara formula $\frac{1}{2(k+h^\vee)}\sum_a (J^a_{-1})^2 v_k$, where $\{J^a\}$ is an orthonormal basis of $\mathfrak{g}$. Thus, each $\widehat{\mathfrak{g}}$–module from the category $\mathcal{O}$ of level $k \neq -h^\vee$ is automatically a module over the Virasoro algebra.

3.3. Boson–fermion correspondence

Let $\mathcal{C}$ be the Clifford algebra associated to the vector space $\mathbb{C}((t)) \oplus \mathbb{C}((t))dt$ equipped with the inner product induced by the residue pairing. It has topological generators $\psi_n = t^n, \psi^*_n = t^{n-1}dt, n \in \mathbb{Z}$, satisfying the anti-commutation relations

$$[\psi_n, \psi_m]_+ = [\psi^*_n, \psi^*_m]_+ = 0, \qquad [\psi_n, \psi^*_m]_+ = \delta_{n,-m}. \tag{14}$$

Denote by $\bigwedge$ the fermionic Fock representation of $\mathcal{C}$, generated by vector $|0\rangle$, such that $\psi_n|0\rangle = 0$, $n \geq 0$, $\psi^*_n|0\rangle = 0$, $n > 0$. This is a vertex superalgebra with

$$Y(\psi_{-1}|0\rangle, z) = \psi(z) := \sum_{n\in\mathbb{Z}} \psi_n z^{-n-1}, \qquad Y(\psi^*_0|0\rangle, z) = \psi^*(z) := \sum_{n\in\mathbb{Z}} \psi^*_n z^{-n}.$$

Boson–fermion correspondence establishes an isomorphism between $\bigwedge$ and a vertex superalgebra built out of the Heisenberg algebra $\mathcal{H}$ from §2.3. For $\lambda \in \mathbb{C}$, let π_λ be the $\mathcal{H}$–module generated by a vector $\mathbf{1}_\lambda$, such that $b_n \cdot \mathbf{1}_\lambda = \lambda\delta_{n,0}\mathbf{1}_\lambda, n \geq 0$. For $N \in \mathbb{Z}_{>0}$, set $V_{\sqrt{N}\mathbb{Z}} = \oplus_{m\in\mathbb{Z}}\pi_{m\sqrt{N}}$. This is a vertex algebra (resp., superalgebra) for any even N (resp., odd N), which contains $\pi_0 = \pi$ as a vertex subalgebra. The gradation is given by the formulas $\deg b_n = -n, \deg \mathbf{1}_\lambda = \lambda^2/2$ (note that it can be half-integral). The translation operator is $T = \frac{1}{2}\sum_{n\in\mathbb{Z}} b_n b_{-1-n}$, so $T \cdot \mathbf{1}_\lambda = \lambda b_{-1}\mathbf{1}_\lambda$. In order to define the vertex operation on $V_{\sqrt{N}\mathbb{Z}}$, it suffices to define the fields $Y(\mathbf{1}_{m\sqrt{N}}, z)$. They are determined by the identity

$$\partial_z Y(\mathbf{1}_\lambda, z) = Y(T \cdot \mathbf{1}_\lambda, z) = \lambda : b(z)Y(\mathbf{1}_\lambda, z) : \quad .$$

Explicitly,

$$Y(\mathbf{1}_\lambda, z) = S_\lambda z^{\lambda b_0} \exp\left(-\lambda \sum_{n<0} \frac{b_n}{n} z^{-n}\right) \exp\left(-\lambda \sum_{n>0} \frac{b_n}{n} z^{-n}\right),$$

where $S_\lambda : \pi_\mu \to \pi_{\mu+\lambda}$ is the shift operator, $S_\lambda \cdot \mathbf{1}_\mu = \mathbf{1}_{\mu+\lambda}, [S_\lambda, b_n] = 0, n \neq 0$.

The *boson–fermion correspondence* is an isomorphism of vertex superalgebras $\bigwedge \simeq V_\mathbb{Z}$, which maps $\psi(z)$ to $Y(\mathbf{1}_{-1}, z)$ and $\psi^*(z)$ to $Y(\mathbf{1}_1, z)$. For more details, see, e.g., [K2].

3.4. Rational vertex algebras

Rational vertex algebras constitute an important class of vertex algebras, which are particularly relevant to conformal field theory [dFMS]. In order to define them, we first need to give the definition of a module over a vertex algebra.

Let V be a vertex algebra. A vector space M is called a *V–module* if it is equipped with the following data:

– gradation: $M = \oplus_{n\in\mathbb{Z}_+} M_n$;

– operation $Y_M : V \to \operatorname{End} M[[z, z^{-1}]]$, which assigns to each $A \in V_n$ a field $Y_M(A, z)$ of conformal dimension n on M;

subject to the following conditions:

– $Y_M(|0\rangle, z) = \mathrm{Id}_M$;

– $Y_M(A,z)Y_M(B,w) = Y_M(Y(A,z-w)B,w)$, in the sense of Proposition 2.6.

A conformal vertex algebra V (see Definition 3.1) is called *rational* if every V–module is completely reducible. It is shown in [DLM] that this condition implies that V has *finitely many* inequivalent simple modules, and the graded components of each simple V–module M are finite-dimensional. Furthermore, the gradation operator on M coincides with the Virasoro operator L_0^M up to a shift. Hence we can attach to a simple V–module M its *character* $\operatorname{ch} M = \operatorname{Tr}_M q^{L_0^M - c/24}$, where c is the central charge of V. Set $q = e^{2\pi i\tau}$. Zhu [Z1] has shown that if V satisfies a certain finiteness condition, then the characters give rise to holomorphic functions in τ on the upper-half plane. Moreover, he proved the following remarkable

THEOREM 3.2. — *Let V be a rational vertex algebra satisfying Zhu's finiteness condition, and $\{M_1,\ldots,M_n\}$ be the set of all inequivalent simple V–modules. Then the vector space spanned by* $\operatorname{ch} M_i, i=1,\ldots,n$, *is invariant under the action of* $SL_2(\mathbb{Z})$.

This result has the following heuristic explanation. To each vertex algebra we can attach the sheaves of coinvariants on the moduli spaces of curves (see §6). It is expected that the sheaves associated to a rational vertex algebra satisfying Zhu's condition are vector bundles with projectively flat connection. The characters of simple V–modules should form the basis of the space of horizontal sections of the corresponding bundle on the moduli of elliptic curves near $\tau = \infty$. The $SL_2(\mathbb{Z})$ action is just the monodromy action on these sections.

Here are some examples of rational vertex algebras.

(1) Let L be an even positive definite lattice in a real vector space W. One can attach to it a vertex algebra V_L in the same way as in §3.3 for $L = \sqrt{N}\mathbb{Z}$. Namely, we define the Heisenberg Lie algebra associated to $W_{\mathbb{C}}$ as the Kac-Moody type central extension of the commutative Lie algebra $W_{\mathbb{C}}((t))$. Its Fock representation π_W is a vertex algebra. The vertex structure on π_W can be extended to $V_L = \pi_W \otimes \mathbb{C}[L]$, where $\mathbb{C}[L]$ is the group algebra of L [B1, FLM]. Then V_L is rational, and its inequivalent simple modules are parameterized by L'/L where L' is the dual lattice [D1]. The corresponding characters are theta-functions. The vertex algebra V_L is the chiral symmetry algebra of the free bosonic conformal field theory compactified on the torus W/L.

Note that if we take as L an arbitrary integral lattice, then V_L is a vertex superalgebra.

(2) Let $\mathfrak{g}$ be a simple Lie algebra and $k \in \mathbb{Z}_+$. Let $L_k(\mathfrak{g})$ be the irreducible quotient of the $\widehat{\mathfrak{g}}$–module $V_k(\mathfrak{g})$. This is a rational vertex algebra, whose modules are integrable representations of $\widehat{\mathfrak{g}}$ of level k [FZ]. The corresponding conformal field theory is the Wess-Zumino-Witten model (see [dFMS]).

(3) Let $L_{c(p,q)}$ be the irreducible quotient of $\mathrm{Vir}_{c(p,q)}$ (as a module over the Virasoro algebra), where $c(p,q) = 1 - 6(p-q)^2/pq$, $p,q > 1$, $(p,q) = 1$. This is a rational vertex

algebra [Wa], whose simple modules form the "minimal model" of conformal field theory defined by Belavin, Polyakov, and Zamolodchikov [BPZ] (see also [dFMS]).

(4) The Moonshine Module vertex algebra $V^\natural$ (see below).

Zhu [Z1] has attached to each vertex algebra V an associative algebra $A(V)$, such that simple V–modules are in one-to-one correspondence with simple $A(V)$–modules.

3.5. Orbifolds and the Monster

Let G be a group of automorphisms of a vertex algebra V. Then the space V^G of G–invariants in V is a vertex subalgebra of V. When G is a finite group, *orbifold theory* allows one to construct V^G–modules from twisted V–modules. For a finite order automorphism g of V, one defines a g–twisted V–module by modifying the above axioms of V–module in such a way that $Y_M(A,z)$ has monodromy λ^{-1} around the origin, if $g \cdot A = \lambda v$ (see [DLM]). Then conjecturally all V^G–modules can be obtained as V^G–submodules of g–twisted V–modules for $g \in G$.

Now suppose that V is a *holomorphic* vertex algebra, i.e., rational with a unique simple module, namely itself. Let G be a cyclic group of automorphisms of V of order n generated by g. Then the following pattern is expected to hold (see [D2]): (1) for each $h \in G$ there is a unique simple h–twisted V–module V_h; (2) V_h has a natural G–action, so we can write $V_h = \oplus_{i=0}^{n-1} V_h(i)$, where $V_h(i) = \{v \in V_h | g \cdot v = e^{2\pi i k/n} v\}$; then each $V_h(i)$ is a simple V^G–module, and these are all simple V^G–modules; (3) The vector space $\widetilde{V} = \oplus_{h \in G} V_h(0)$ carries a canonical vertex algebra structure, which is holomorphic.

A spectacular example is the construction of the *Moonshine Module* vertex algebra $V^\natural$ by I. Frenkel, Lepowsky and Meurman [FLM]. In this case V is the vertex algebra V_Λ associated to the Leech lattice Λ (it is self-dual, hence V_Λ is holomorphic), and g is constructed from the involution -1 on Λ. Then $V^\natural = V_\Lambda(0) \oplus V_{\Lambda,g}(0)$ is a vertex algebra, whose group of automorphisms is the Monster group [FLM]. Moreover, $V^\natural$ is holomorphic [D2]. Conjecturally, $V^\natural$ is the unique holomorphic vertex algebra V with central charge 24, such that $V_1 = 0$. Its character is the modular function $j(\tau) - 744$. More generally, for each element x of the Monster group, consider the Thompson series $\mathrm{Tr}_{V^\natural}\, x q^{L_0 - 1}$. Conway and Norton conjectured, and Borcherds proved [B2] that these are Hauptmoduls for genus zero subgroups of $SL_2(\mathbb{R})$.

3.6. Coset construction

Let V be a vertex algebra, and W its vertex subalgebra. Let $C(V,W)$ be the vector subspace of V spanned by vectors v, such that $Y(A,z) \cdot v \in V[[z]]$ for all $A \in W$. Then $C(V,W)$ is a vertex subalgebra of V, which is called the *coset vertex algebra* of the pair (V,W). In the case $W = V$, the vertex algebra $C(V,V)$ is commutative, and is called the *center* of V. An example of a coset vertex algebra is provided by the famous Goddard–Kent–Olive construction [GKO], which identifies $C(L_1(\mathfrak{sl}_2) \otimes L_k(\mathfrak{sl}_2), L_{k+1}(\mathfrak{sl}_2))$ with $L_{c(k+2,k+3)}$ (see §3.4). For other examples, see [dFMS].

3.7. BRST construction and $\mathcal{W}$–algebras

For $A \in V$, let $y(A) = \operatorname{Res} Y(A,z)$. Then associativity implies: $[y(A), Y(B,z)] = Y(y(A)\cdot B, z)$, and so $y(A)$ is an infinitesimal automorphism of V. Now suppose that $V^\bullet$ is a vertex (super)algebra with an additional $\mathbb{Z}$–gradation and $A \in V^1$ is such that $y(A)^2 = 0$. Then $(V^\bullet, y(A))$ is a complex, and its cohomology is a graded vertex (super)algebra. Important examples of such complexes are provided by *topological vertex superalgebras* introduced by Lian and Zuckerman [LZ]. In this case the cohomology is a graded commutative vertex algebra, but it has an additional structure of Batalin-Vilkovisky algebra.

Another example is the BRST complex of quantum hamiltonian reduction. We illustrate it in the case of quantum Drinfeld-Sokolov reduction [FF3], which leads to the definition of $\mathcal{W}$–algebras. Let $\mathfrak{g}$ be a simple Lie algebra of rank ℓ, and $\mathfrak{n}$ its upper nilpotent subalgebra.

Let $\mathcal{C}$ be the Clifford algebra associated to the vector space $\mathfrak{n}((t)) \oplus \mathfrak{n}^*((t))dt$ equipped with the inner product induced by the residue pairing. It has topological generators $\psi_{\alpha,n} = e_\alpha \otimes t^n, \psi^*_{\alpha,n} = e^*_\alpha \otimes t^{n-1}dt, \alpha \in \Delta_+, n \in \mathbb{Z}$, satisfying the anti-commutation relations (14). Let $\bigwedge_{\mathfrak{n}}$ be its Fock representation, generated by vector $|0\rangle$, such that $\psi_{\alpha,n}|0\rangle = 0, n \geq 0, \psi^*_{\alpha,n}|0\rangle = 0, n > 0$. This is a vertex superalgebra which is the tensor product of several copies of the vertex superalgebra $\bigwedge$ from §3.3. Introduce an additional $\mathbb{Z}$–gradation on $\mathcal{C}$ and $\bigwedge_{\mathfrak{n}}$ by setting $\deg \psi^*_{\alpha,n} = -\deg \psi_{\alpha,n} = 1, \deg |0\rangle = 0$.

Now consider the vertex superalgebra $C^\bullet_k(\mathfrak{g}) = V_k(\mathfrak{g}) \otimes \bigwedge^\bullet_{\mathfrak{n}}$. It carries a canonical differential d_{st} of degree 1, the standard differential of semi-infinite cohomology of $\mathfrak{n}((t))$ with coefficients in V_k [Fe1, FGZ]; it is equal to the residue of a field from $C^\bullet_k(\mathfrak{g})$ (see [FF3]). Let $\chi = \sum_{i=1}^{\ell} \operatorname{Res} \psi^*_{\alpha_i}(z)$ be the Drinfeld-Sokolov character [DS] of $\mathfrak{n}((t))$. Then $d = d_{\mathrm{st}} + \chi$ is a differential on $C^\bullet_k(\mathfrak{g})$, and the cohomology $H^\bullet_k(\mathfrak{g})$ of this differential is a vertex superalgebra.

THEOREM 3.3. — *$H^0_k(\mathfrak{g})$ is a vertex algebra generated by elements W_i of degrees $d_i + 1$, $i = 1,\ldots,\ell$, where d_i is the ith exponent of $\mathfrak{g}$ (in the sense of the Reconstruction Theorem), and $H^i_k(\mathfrak{g}) = 0, i \neq 0$.*

This theorem is proved in [FF3, FF4] for generic k and in [dBT] for all k. The vertex algebra $H^0_k(\mathfrak{g})$ is called the $\mathcal{W}$–*algebra* associated to $\mathfrak{g}$ and denoted by $\mathcal{W}_k(\mathfrak{g})$. We have: $\mathcal{W}_k(\mathfrak{sl}_2) = \mathrm{Vir}_{c(k)}$, where $c(k) = 1 - 6(k+1)^2/(k+2)$. For $\mathfrak{g} = \mathfrak{sl}_3$, the $\mathcal{W}$–algebra was first constructed by Zamolodchikov, and for $\mathfrak{g} = \mathfrak{sl}_N$ by Fateev and Lukyanov [FL]. Since $V_k(\mathfrak{g})$ is conformal for $k \neq -h^\vee$ (see §3.2), $\mathcal{W}_k(\mathfrak{g})$ is also conformal, with W_1 playing the role of conformal vector. On the other hand, $\mathcal{W}_{-h^\vee}(\mathfrak{g})$ is a commutative vertex algebra, which is isomorphic to the center of $V_{-h^\vee}(\mathfrak{g})$ [FF3] (see §6.5). The simple quotient of $\mathcal{W}_k(\mathfrak{g})$ for $k = -h^\vee + p/q$, where p, q are relatively prime integers greater than or equal to $h^\vee$, is believed to be a rational vertex algebra.

Moreover, if $\mathfrak{g}$ is simply-laced and $q = p + 1$, this vertex algebra is conjecturally isomorphic to the coset algebra of $(L_1(\mathfrak{g}) \otimes L_{p-2}(\mathfrak{g}), L_{p-1}(\mathfrak{g}))$, see [FKW].

For any $V_k(\mathfrak{g})$–module M, the cohomology of the complex $(M \otimes \bigwedge^\bullet_{\mathfrak{n}}, d)$ is a $\mathcal{W}_k(\mathfrak{g})$–module. This defines a functor, which was studied in [FKW].

The $\mathcal{W}$–algebras exhibit a remarkable duality: $\mathcal{W}_k(\mathfrak{g}) \simeq \mathcal{W}_{k'}({}^L\mathfrak{g})$, where ${}^L\mathfrak{g}$ is the Langlands dual Lie algebra to $\mathfrak{g}$ and $(k + h^\vee)r^\vee = (k' + {}^L h^\vee)^{-1}$ (here $r^\vee$ denotes the maximal number of edges connecting two vertices of the Dynkin diagram of $\mathfrak{g}$), see [FF3]. In the limit $k \to -h^\vee$ it becomes the isomorphism of Theorem 6.3.

4. COORDINATE INDEPENDENT DESCRIPTION OF VERTEX ALGEBRA STRUCTURE

Up to this point, we have discussed vertex algebras in the language of formal power series. The vertex operation is a map $Y : V \to \operatorname{End} V[[z, z^{-1}]]$, or, equivalently, an element of the completed tensor product $V^*((z)) \widehat{\otimes} \operatorname{End} V$. Thus, we can view Y as an $\operatorname{End} V$–valued section of a vector bundle over the punctured disc $D^\times = \operatorname{Spec} \mathbb{C}((z))$ with fiber V^*. The question is whether one can define this bundle in such a way that this section is canonical, i.e., independent of the choice of z. In order to do that, we need a precise description of the transformation properties of Y. Once we understand what type of geometric object Y is, we will be able to give a global geometric meaning to vertex operators on arbitrary curves.

4.1. The group $\operatorname{Aut} \mathcal{O}$

Denote by $\mathcal{O}$ the complete topological $\mathbb{C}$–algebra $\mathbb{C}[[z]]$, and let $\operatorname{Aut} \mathcal{O}$ be the group of continuous automorphisms of $\mathcal{O}$. Such an automorphism is determined by its action on the generator $z \in \mathbb{C}[[z]]$. Thus as a set this group consists of elements $\rho(z) \in \mathbb{C}[[z]]$ of the form $\rho_1 z + \rho_2 z^2 + \ldots$, with $\rho_1 \in \mathbb{C}^\times$, endowed with the composition law $(\rho * \mu)(z) = \mu(\rho(z))$. It is easy to see that $\operatorname{Aut} \mathcal{O}$ is a proalgebraic group, $\varprojlim \operatorname{Aut} \mathbb{C}[[z]]/(z^n)$. It is the semi-direct product $\mathbb{G}_m \ltimes \operatorname{Aut}_+ \mathcal{O}$, where $\operatorname{Aut}_+ \mathcal{O}$ is the subgroup that consists of the transformations of the form $z \to z + a_2 z^2 + \cdots$, and $\mathbb{G}_m$ is the group of rescalings $z \to az, a \neq 0$. Furthermore, $\operatorname{Aut}_+ \mathcal{O}$ is a prounipotent group. The Lie algebra of $\operatorname{Aut} \mathcal{O}$ is $\operatorname{Der}_0 \mathcal{O} = z\mathbb{C}[[z]]\partial_z$, which is a semi–direct product of $\mathbb{C}\partial_z$ and $\operatorname{Der}_+ \mathcal{O} = z^2\mathbb{C}[[z]]\partial_z$. Any representation of $\operatorname{Der}_0 \mathcal{O}$, on which $z\partial_z$ is diagonalizable with integral eigenvalues and $\operatorname{Der}_+ \mathcal{O}$ acts locally nilpotently, can be exponentiated to a representation of $\operatorname{Aut} \mathcal{O}$.

Given a vertex algebra and a field $Y(A, z)$, it makes sense to consider a new field $Y(A, \rho(z))$. We now seek an action of $\operatorname{Aut} \mathcal{O}$ on V, $A \mapsto R(\rho) \cdot A$, such that $Y(A, \rho(z))$ is related to $Y(R(\rho)A, z)$ in some reasonable way. In the theory of vertex algebras, actions of $\operatorname{Aut} \mathcal{O}$ usually arise from the action of the Virasoro algebra. Namely, let V be a conformal vertex algebra (see Definition 3.1). Then the Fourier coefficients

L_n of $Y(\omega, z)$ satisfy the commutation relations of the Virasoro algebra with central charge c. The operators $L_n, n \geq 0$, then define an action of the Lie algebra $\mathrm{Der}_+ \mathcal{O}$ (L_n corresponds to $-t^{n+1}\partial_t$). It follows from the axioms of vertex algebra that this action can be exponentiated to an action of $\mathrm{Aut}\,\mathcal{O}$. For $f(z) \in \mathrm{Aut}\,\mathcal{O}$, denote by $R(f) : V \to V$ the corresponding operator.

Given a vector field $v = \sum_{n \geq -1} v_n z^{n+1} \partial_z$, we assign to it the operator $\mathbf{v} = -\sum_{n \geq -1} v_n L_n$. From formula (9) we obtain:

$$[\mathbf{v}, Y(A, w)] = -\sum_{m \geq -1} \frac{1}{(m+1)!} (\partial_w^{m+1} v(w)) Y(L_m A, w). \tag{15}$$

By "exponentiating" this formula, we obtain the following identity, due to Y.-Z. Huang [Hu]:

$$Y(A, t) = R(\rho) Y(R(\rho_t)^{-1} A, \rho(t)) R(\rho)^{-1}, \tag{16}$$

for any $\rho \in \mathrm{Aut}\,\mathbb{C}[[z]]$ (here $\rho_t(z) = \rho(t+z) - \rho(t)$, considered as formal power series in z with coefficients in $\mathbb{C}[[t]]$).

4.2. Vertex algebra bundle

Now we can give a coordinate-free description of the operation Y. Let X be a smooth complex curve. Given a point $x \in X$, denote by $\mathcal{O}_x$ the completion of the local ring of x, by $\mathfrak{m}_x$ its maximal ideal, and by $\mathcal{K}_x$ the field of fractions of $\mathcal{O}_x$. A formal coordinate t_x at x is by definition a topological generator of $\mathfrak{m}_x$. Consider the set of pairs (x, t_x), where $x \in X$ and t_x is a formal coordinate at x. This is the set of points of a scheme $\widehat{X}$ of infinite type, which is a principal $\mathrm{Aut}\,\mathcal{O}$–bundle over X. Its fiber at $x \in X$ is the $\mathrm{Aut}\,\mathcal{O}$–torsor $\mathcal{L}_x$ of all formal coordinates at x. Given a finite-dimensional $\mathrm{Aut}\,\mathcal{O}$–module V, let $\mathcal{V} = \widehat{X} \underset{\mathrm{Aut}\,\mathcal{O}}{\times} V$ be the vector bundle associated to V and $\widehat{X}$. The fiber of $\mathcal{V}$ at $x \in X$ is the $\mathcal{L}_x$–twist of V, $\mathcal{V}_x = \mathcal{L}_x \underset{\mathrm{Aut}\,\mathcal{O}}{\times} V$.

A conformal vertex algebra is an $\mathrm{Aut}\,\mathcal{O}$–module, which has a filtration $V_{\leq i} := \oplus_{k=0}^i V_k$ by finite-dimensional $\mathrm{Aut}\,\mathcal{O}$–submodules. We obtain the directed system $(\mathcal{V}_{\leq i})$ of the corresponding vector bundles of finite rank on X and embeddings $\mathcal{V}_{\leq i} \hookrightarrow \mathcal{V}_{\leq j}, i \leq j$. We will denote this system simply by $\mathcal{V}$, thinking of it as an inductive limit of bundles of finite rank. Likewise, by $\mathcal{V}^*$ we will understand the inverse system of bundles $(\mathcal{V}_{\leq i})^*$ and surjections $(\mathcal{V}_{\leq j})^* \twoheadrightarrow (\mathcal{V}^i)^*$, thinking of it as a projective limit of bundles of finite rank.

The $\mathrm{Aut}\,\mathcal{O}$–bundle $\widehat{X}$ on X above carries an action of the Lie algebra $\mathrm{Der}\,\mathcal{O}$, which is compatible with the $\mathrm{Aut}\,\mathcal{O}$–action and simply transitive, i.e., $\mathrm{Der}\,\mathcal{O} \widehat{\otimes} \mathcal{O}_{\widehat{X}} \simeq \Theta_{\widehat{X}}$. Since V is a $\mathrm{Der}\,\mathcal{O}$–module, by the general construction of §6.1 we obtain a flat connection $\nabla : \mathcal{V} \to \mathcal{V} \otimes \Omega$ on $\mathcal{V}$. Locally, ∇ can be written as $d + L_{-1} \otimes dz$, where d is the deRham differential. The connection ∇ on $\mathcal{V}$ gives us a connection ∇^* on the dual bundle $\mathcal{V}^*$.

To be precise, $\mathcal{V}$ is not an inductive limit of flat bundles, since $\nabla(\mathcal{V}_{\leq i}) \subset \mathcal{V}_{i+1} \otimes \Omega$, and likewise, $\mathcal{V}^*$ is not a projective limit of flat bundles (it is a $\mathcal{D}$–module on X, which is not quasicoherent as an $\mathcal{O}$–module).

Now let us restrict $\mathcal{V}^*$ to the punctured disc $D_x^\times = \operatorname{Spec} \mathcal{K}_x$ around $x \in X$. We want to define an $\operatorname{End} \mathcal{V}_x$–valued meromorphic section $\mathcal{Y}_x$ of $\mathcal{V}^*$ on $D_x^\times$. Pick a formal coordinate z at x. With this choice, identify $\mathcal{V}_x$ with V and trivialize $\mathcal{V}|_{D_x}$. We define our section $\mathcal{Y}_x$ in this trivialization through its matrix coefficients: to each triple $c \in V$, $c' \in V^*$ and a section $F : D_x \to V$ we need to assign a meromorphic function on the disc, which is $\mathbb{C}$–linear in c and c' and $\mathcal{O}$–linear in F. It suffices to assign such a function to triples c, c', F, where F is the constant section equal to $A \in V$ with respect to our trivialization. We set this function to be equal to $\langle c'|Y(A,z)|c\rangle$.

THEOREM 4.1. — *The section $\mathcal{Y}_x$ defined this way is canonical, i.e., independent of the choice of formal coordinate z at x. Moreover, $\mathcal{Y}_x$ is horizontal: $\nabla^* \mathcal{Y}_x = 0$.*

Thus, we see that a conformal vertex algebra gives rise to a vector bundle on any smooth curve and a canonical horizontal section of the restriction of this bundle to the neighborhood of each point with values in the endomorphisms of the corresponding twist of V. The fact that this section is canonical is equivalent to formula (16). Flatness follows from the formula $\partial_z Y(B,z) = Y(L_{-1}B, z)$, which holds in any vertex algebra.

Remark 4.2. — Let $\mathcal{V}^{\boxtimes n}$ be the n–fold external tensor power of $\mathcal{V}$. Consider the n–point functions (10). According to Proposition 2.7, these are elements of

$$\mathbb{C}[[z_1, \dots, z_n]][(z_i - z_j)^{-1}]_{i \neq j}.$$

Trivializing the restriction of $\mathcal{V}^{\boxtimes n}$ to $D_x^n = \operatorname{Spec} \mathbb{C}[[z_1, \dots, z_n]]$ using the coordinate z, we obtain a $\mathcal{V}_x^*$–valued section of $(\mathcal{V}^{\boxtimes n})^*|_{D_x^n}$ with poles on the diagonals. Then this section is canonical, i.e., independent of the choice of z, and horizontal.

4.3. Examples

If W is an $\operatorname{Aut} \mathcal{O}$–stable subspace of V, it gives rise to a subbundle $\mathcal{W}$ of $\mathcal{V}$. The universal section $\mathcal{Y}_x$ then gives rise to an $\operatorname{End} \mathcal{V}_x$–valued section of $\mathcal{W}^*|_{D_x^\times}$. This way one obtains many familiar geometric objects.

Suppose $A \in V_\Delta$ satisfies $L_n \cdot A = 0, L_0 A = \Delta A$. Then subspace $\mathbb{C}A \subset V$ is a one-dimensional $\operatorname{Aut} \mathcal{O}$ submodule of V. The line bundle associated to this module is nothing but the line bundle $\Omega^{-\Delta}$ of Δ–differentials on X. Thus we obtain a line subbundle $\mathcal{L}_A$ of $\mathcal{V}$. Our section $\mathcal{Y}_x$ therefore gives us an $\operatorname{End} \mathcal{V}_x$–valued section of $\mathcal{L}_A^{-1} = \Omega^\Delta$. In other words, the $\operatorname{End} \mathcal{V}_x$–valued Δ–differential $Y(A,z)(dz)^\Delta$ does not depend on the choice of formal coordinate z at x.

Next, consider the Heisenberg vertex algebra $V = \pi$, with conformal structure given by the vector $\frac{1}{2} b_{-1}^2 + b_{-2}$. For each smooth curve X, π gives rise to a vector bundle that we denote by Π. The first piece π_0 of our filtration on $\pi = \mathbb{C}[b_n]_{n<0}$

is the one-dimensional subspace spanned by the vacuum vector 1. This is a trivial representation of the group $\operatorname{Aut} \mathcal{O}$. Hence it gives rise to a trivial subbundle $\mathcal{O}_X \subset \Pi$. The next piece in the filtration, $\pi_{\leq 1} = \mathbb{C}1 \oplus \mathbb{C}b_{-1}$, gives rise to a rank two subbundle of Π, which we denote by $\mathcal{B}$. Its dual bundle is an extension

$$0 \to \Omega \to \mathcal{B}^* \to \mathcal{O}_X \to 0. \tag{17}$$

Our universal section $\mathcal{Y}_x$ gives rise to an $\operatorname{End} \Pi_x$–valued section of $\mathcal{B}^*|_{D_x^\times}$, which projects onto the section $\operatorname{End} \Pi_x$–valued section of $\mathcal{O}_X$ equal to Id (this follows from the vacuum axiom). Explicit computation shows that the space of sections of $\mathcal{B}^*$ which project onto the section 1 of $\mathcal{O}_X$ (equivalently, splittings of (17)) is canonically isomorphic to the space of *affine connections* (they may also be described as affine structures, see [Gu]). Thus, we conclude that the field $b(z)$ (or, more precisely, the expression $\partial_z + b(z)$) transforms as an $\operatorname{End} \Pi_x$–valued affine connection on $D_x^\times$. Similarly, we obtain connections on G–bundles from the subspace $V_k(\mathfrak{g})_{\leq 1}$ of the Kac-Moody vertex algebra $V_k(\mathfrak{g})$.

Let V be a conformal vertex algebra with central charge c. It follows from the definition that the vector space $\mathbb{C}|0\rangle \oplus \mathbb{C}\omega$ is $\operatorname{Aut} \mathcal{O}$–stable. Hence it gives rise to a rank two subbundle $\mathcal{T}_c$ of $\mathcal{V}$. Its dual bundle is an extension

$$0 \to \Omega^2 \to \mathcal{T}_c^* \to \mathcal{O}_X \to 0. \tag{18}$$

The universal section $\mathcal{Y}_x$ gives rise to an $\operatorname{End} \mathcal{V}_x$–valued section of $\mathcal{T}_c^*$ over the punctured disc, which projects onto $\mathrm{Id} \in \operatorname{End} \mathcal{V}_x \otimes \mathcal{O}_x$. Explicit calculation shows that the space of sections of $\mathcal{T}_c^*$ projecting onto $1 \in \Gamma(\mathcal{O}_X)$ (equivalently, splittings of (18)) is isomorphic to the space of self–adjoint second order differential operators $\rho : \Omega^{-\frac{1}{2}} \to \Omega^{\frac{3}{2}}$ with constant symbol $\frac{c}{6}$. Locally, such an operator can be written as $\frac{c}{6}\partial_z^2 + q(z)$. Operators of this form with symbol 1 are known as *projective connections* (they may also be described as projective structures, see [Gu]). Thus, we see that for $c \neq 0$ the field $T(z)$ (called the stress tensor in conformal field theory), or more precisely, the expression $\partial_z^2 + \frac{6}{c}T(z)$ transforms as an $\operatorname{End} \mathcal{V}_x$–valued projective connection on the punctured disc.

4.4. General twisting property

We have seen in §4.2 that the vertex operation Y is in some sense invariant under the $\operatorname{Aut} \mathcal{O}$–action (see formula (16)). Therefore Y gives rise to a well-defined operation on the twist of V by any $\operatorname{Aut} \mathcal{O}$–torsor. This "twisting property" follows from the fact that $\operatorname{Aut} \mathcal{O}$ acts on V by "internal symmetries", that is by exponentiation of Fourier coefficients of the vertex operator $Y(\omega, z)$. It turns out that vertex algebras exhibit a similar twisting property with respect to any group $\mathcal{G}$ of internal symmetry, i.e., a group (or more generally, an ind-group) obtained by exponentiation of Fourier coefficients of vertex operators. Using associativity, one can obtain an analogue of formula (16) for the transformations of Y under the action of $\mathcal{G}$. This formula means that we get a well-defined operation on the twist of V by any $\mathcal{G}$–torsor.

For example, let V be a vertex algebra with a $\widehat{\mathfrak{g}}$*–structure* of level $k \neq -h^\vee$, i.e., one equipped with a homomorphism $V_k(\mathfrak{g}) \to V$, whose image is not contained in $\mathbb{C}|0\rangle$. Recall that $V_k(\mathfrak{g})$ is a conformal vertex algebra if $k \neq -h^\vee$ (see §3.2). Given such a homomorphism $V_k(\mathfrak{g}) \to V$, the Fourier coefficients of the corresponding fields $Y(\omega, z)$ and $Y(J^a, z)$ generate an action of the Lie algebra $Vir \ltimes \widehat{\mathfrak{g}}$ on V. Suppose that the action of its Lie subalgebra $\mathrm{Der}_0\, \mathcal{O} \ltimes \mathfrak{g}(\mathcal{O})$ can be exponentiated to an action of the group $\mathrm{Aut}\, \mathcal{O} \ltimes G(\mathcal{O})$ on V (then we say that this group acts on V by internal symmetries). Let $\mathcal{P}$ be a G–bundle on a smooth curve X. Denote by $\widehat{\mathcal{P}}$ the principal $\mathrm{Aut}\, \mathcal{O} \ltimes G(\mathcal{O})$–bundle over X, whose fiber at x consists of pairs (z, s), where z is a formal coordinate at x and s is a trivialization of $\mathcal{P}|_{D_x}$. Let $\mathcal{V}^{\mathcal{P}}$ be the $\widehat{\mathcal{P}}$–twist of V. Then the vertex operation Y on V gives rise to a canonical section $\mathcal{Y}_x^{\mathcal{P}}$ of $(\mathcal{V}^{\mathcal{P}})^*|_{D_x^\times}$ with values in $\mathrm{End}\, \mathcal{V}_x^{\mathcal{P}}$.

5. CONFORMAL BLOCKS

5.1. The results of the previous section allow us to assign to a conformal vertex algebra a vector bundle with connection on any smooth curve X, and a family of local structures on it. We can now associate to any compact curve X an invariant of the vertex algebra structure.

Definition 5.1. — *Let V be a conformal vertex algebra, X a smooth projective curve, and x a point of X. A linear functional φ on $\mathcal{V}_x$ is called a* conformal block *if $\varphi(\mathcal{Y}_x \cdot A) \in \Gamma(D_x^\times, \mathcal{V}^*)$ can be extended to a regular section of $\mathcal{V}^*$ on $X \backslash x$ for all $A \in \mathcal{V}_x$.*

The set of conformal blocks is a vector subspace of $\mathcal{V}_x^*$, denoted by $C(X, x, V)$.

Let $\varphi \in C(X, x, V)$, and $A \in \mathcal{V}_x$. Denote by φ_A the corresponding section of $\mathcal{V}^*$ over $X \backslash x$. If W is an $\mathrm{Aut}\, \mathcal{O}$–stable subspace of V, then φ_A can be projected onto a section of $\mathcal{W}^*$ over $X \backslash x$. Note that according to the vacuum axiom, $\varphi_{|0\rangle}$ is actually a regular section of $\mathcal{V}^*$ over the whole X, and so is its projection. In particular, taking as W the two-dimensional subspace of V spanned by $|0\rangle$ and ω, we assign to each conformal block a projective connection on X (if $c \neq 0$). Denote $\varphi(\cdot|0\rangle)$ as $\langle \cdot \rangle_\varphi$. If we choose a formal coordinate z at x, then we can write this projective connection as $\partial_z^2 + \frac{6}{c}\langle T(z) \rangle_\varphi$. If $c = 0$, then we obtain a quadratic differential $\langle T(z) \rangle_\varphi dz^2$.

Remark 5.2. — Let V be a vertex algebra with $\widehat{\mathfrak{g}}$–structure of level $k \neq -h^\vee$, and $\mathcal{P}$ a G–bundle on X. A conformal block *twisted by* $\mathcal{P}$ is by definition a linear functional φ on $\mathcal{V}_x^{\mathcal{P}}$, such that $\varphi(\mathcal{Y}_x^{\mathcal{P}} \cdot A)$ (see §4.4) can be extended to a regular section of $(\mathcal{V}^{\mathcal{P}})^*$ on $X \backslash x$ for all $A \in \mathcal{V}_x^{\mathcal{P}}$. We denote the corresponding space by $C^{\mathcal{P}}(X, x, V)$.

5.2. Examples

In the case of the Kac-Moody vertex algebra $V_k(\mathfrak{g})$ the Definition 5.1 can be simplified. In this case the subspace $(\mathfrak{g}\otimes t^{-1})v_k$ is $\operatorname{Aut}\mathcal{O}$–invariant, and therefore we have a surjection $(\mathcal{V}_k(\mathfrak{g}))^* \twoheadrightarrow \mathfrak{g}^* \otimes \Omega$. Hence for each $\varphi \in C(X, x, \pi)$, $A \in \mathcal{V}_x$, we obtain a regular $\mathfrak{g}^*$–valued one-form on $X\backslash x$, whose restriction to $D_x^\times$ is $\varphi(J(z)\cdot A)dz$, where $J(z) = \sum_a J_a \otimes J^a(z)$.

Now observe that any $f = \sum f_n z^n \in \mathfrak{g}\otimes\mathcal{K}_x$ gives rise to an operator on $\mathcal{V}_k(\mathfrak{g})_x$, $\widetilde{f} = \operatorname{Res}_x(f, J(z))dz = \sum f_n b_n$, which does not depend on the choice of the coordinate z. Since by definition $\varphi(J(z)\cdot A)dz$ extends to a regular one-form on $X\backslash x$, we obtain from the residue theorem that $\varphi(\widetilde{f}\cdot A) = 0$ for all $f \in \mathfrak{g}_{x,\mathrm{out}} = \mathfrak{g}\otimes\mathbb{C}[X\backslash x]$. Hence we obtain a map from $C(X, x, V_k(\mathfrak{g}))$ to the space of $\mathfrak{g}_{x,\mathrm{out}}$–invariant functionals on $\mathcal{V}_k(\mathfrak{g})_x$.

Lemma 5.3. — *The space of conformal blocks $C(X, x, V_k(\mathfrak{g}))$ is isomorphic to the space of $\mathfrak{g}_{x,\mathrm{out}}$–invariant functionals on $\mathcal{V}_k(\mathfrak{g})_x$.*

Thus we recover the common definition of conformal blocks as $\mathfrak{g}_{x,\mathrm{out}}$–invariants. They have been extensively studied recently because of the relation to the moduli spaces of G–bundles on curves (see §6.4 below). More generally, we have: $C^{\mathcal{P}}(X, x, V_k(\mathfrak{g})) = \operatorname{Hom}_{\mathfrak{g}^{\mathcal{P}}_{x,\mathrm{out}}}(\mathcal{V}_k(\mathfrak{g})^{\mathcal{P}}_x, \mathbb{C})$, where $\mathfrak{g}^{\mathcal{P}}_{x,\mathrm{out}}$ is the Lie algebra of sections of $\mathcal{P}\underset{G}{\times}\mathfrak{g}$ over $X\backslash x$.

The situation becomes more subtle in the case of the Virasoro vertex algebra Vir_c. The analogue of $\mathfrak{g}_{x,\mathrm{out}}$ is then the Lie algebra $\operatorname{Vect}(X\backslash x)$ of vector fields on $X\backslash x$. By analogy with the Kac-Moody case, we would like to define a homomorphism $\operatorname{Vect}(X\backslash x) \to \operatorname{End}\mathcal{V}ir_{c,x}$ sending $\xi(z)\partial_z \in \operatorname{Vect}(X\backslash x)$ to $\operatorname{Res}_x \xi(z)T(z)dz$. But it is not clear whether this map is independent of the choice of formal coordinate z, unless $c = 0$ (when the field $T(z)$ transforms as a quadratic differential). According to §4.3, if $c \neq 0$, then $\partial_z^2 + \frac{6}{c}T(z)$ is naturally a section of a rank two bundle $\mathcal{T}_c^*$. Hence it can be paired with sections of the rank two bundle $\mathcal{T}_c\otimes\Omega$ (we could split this bundle by choosing a projective structure on X, but we prefer not to do that). One can show that the sheaf $\widehat{\mathcal{T}}_c = (\mathcal{T}_c\otimes\Omega)/d\mathcal{O}$, which is an extension of Θ by $\Omega/d\mathcal{O}$, has a canonical Lie algebra structure (see [Wi, BS]), such that $Vir_x = \Gamma(D_x^\times, \widehat{\mathcal{T}}_c)$ is isomorphic to the Virasoro algebra and acts on $\mathcal{V}ir_{c,x}$. The map $\Gamma(X\backslash x, \widehat{\mathcal{T}}_c) \to Vir_x$ factors through $\operatorname{Vect}(X\backslash x)$, and hence the above map $\operatorname{Vect}(X\backslash x) \to \operatorname{End}\mathcal{V}ir_{c,x}$ is well-defined. One can then show that the space of conformal blocks $C(X, x, \mathrm{Vir}_c)$ is isomorphic to the space of $\operatorname{Vect}(X\backslash x)$–invariant functionals on $\mathcal{V}ir_{c,x}$.

5.3. General invariance condition

For general vertex algebras the analogue of the above invariance condition is defined as follows. Let $\mathcal{V} \xrightarrow{\nabla} \mathcal{V}\otimes\Omega$, where $\mathcal{V}\otimes\Omega$ is placed in degree 0, be the de Rham complex of the flat vector bundle $\mathcal{V}$, considered as a complex of sheaves on the curve X. Here

∇ is the connection operator defined in §4.2 (recall that $\nabla = d + L_{-1} \otimes dz$). For $\Sigma \subset X$, denote by $H^0_{\mathrm{dR}}(\Sigma, \mathcal{V} \otimes \Omega)$ the degree 0 cohomology of the restriction of this complex to Σ. If Σ is affine or $D^\times_x$, then $H^0_{\mathrm{dR}}(\Sigma, \mathcal{V} \otimes \Omega)$ is simply the quotient of $\Gamma(\Sigma, \mathcal{V} \otimes \Omega)$ by the image of ∇.

There is a linear map $\gamma : \Gamma(D^\times_x, \mathcal{V} \otimes \Omega) \to \operatorname{End} \mathcal{V}_x$ sending $\mu \in \Gamma(D^\times_x, \mathcal{V} \otimes \Omega)$ to a linear operator $O_\mu = \operatorname{Res}_x \langle \mathcal{Y}_x, \mu \rangle$ on $\mathcal{V}_x$. Since the residue of a total derivative vanishes, γ factors through $U(\mathcal{V}_x) = H^0_{\mathrm{dR}}(D^\times_x, \mathcal{V} \otimes \Omega)$. If we choose a formal coordinate z at x, then we can identify $U(\mathcal{V}_x)$ with $U(V)$, the completion of the span of all Fourier coefficients of all vertex operators $Y(A, z), A \in V$. According to formula (9), $U(\mathcal{V}_x)$ is a Lie algebra. Note that in the case of Kac-Moody and Virasoro vertex algebras, $U(\mathcal{V}_x)$ contains $\widehat{\mathfrak{g}}_x$ and Vir_x, respectively, as Lie subalgebras.

Given an open $\Sigma \subset X$, we have a canonical map $H^0_{\mathrm{dR}}(\Sigma, \mathcal{V} \otimes \Omega) \to U(\mathcal{V}_x)$. Denote its image by $U_\Sigma(\mathcal{V}_x)$. It follows from the Beilinson-Drinfeld chiral algebra formalism (see §7) that $U_\Sigma(\mathcal{V}_x)$ is a Lie subalgebra of $U(\mathcal{V}_x)$. The image of $U_\Sigma(\mathcal{V}_x)$ in $\operatorname{End} \mathcal{V}_x$ is the span of all operators O_μ for $\mu \in \Gamma(\Sigma, \mathcal{V} \otimes \Omega)$. The residue theorem implies:

Proposition 5.4. — *The space of conformal blocks $C(X, x, V)$ coincides with the space of $U_{X \backslash x}(\mathcal{V}_x)$–invariant functionals on $\mathcal{V}_x$.*

Therefore the dual space to the space of conformal blocks is the *space of coinvariants* $H(X, x, V) = \mathcal{V}_x / U_{X \backslash x}(\mathcal{V}_x) \cdot \mathcal{V}_x$.

The definition of the spaces $C(X, x, V)$ and $H(X, x, V)$ can be generalized to the case of multiple (distinct) points $x_1, \ldots x_n$ on the curve X, at which we "insert" arbitrary conformal V–modules $M_1, \ldots, M_n$. A V–module M is called conformal if the gradation operator on M coincides, up to a shift, with the Fourier coefficient L^M_0 of the field $Y_M(\omega, z)$. Assume for simplicity that the eigenvalues of L^M_0 on all modules M_i are integers. Then each M_i is an $\operatorname{Aut} \mathcal{O}$–module, and so we can attach to it a vector bundle $\mathcal{M}_i$ on X with a flat connection (a general conformal V–module gives rise to a vector bundle with connection on the space of pairs (x, τ_x), where $x \in X$ and τ_x is a non-zero tangent vector to X at x).

The space of conformal blocks $C_V(X, (x_i), (M_i))^n_{i=1}$ is then by definition the space of linear functionals φ on $\mathcal{M}_{1,x_1} \otimes \cdots \otimes \mathcal{M}_{n,x_n}$, such that for any $A_i \in \mathcal{M}_{i,x_i}, i = 1, \ldots, n$, there exists a section of $\mathcal{V}^*$ over $X \setminus \{x_1, \ldots, x_n\}$, whose restriction to each $D^\times_{x_i}$ equals

$$\varphi(A_1 \otimes \cdots \otimes \mathcal{Y}_{M_i} \cdot A_i \otimes \cdots \otimes A_n).$$

Informally, one can say that the local sections of $\mathcal{V}^*$ over the discs around the points, obtained by acting with vertex operators at those points, can be "glued together" into a single meromorphic section of $\mathcal{V}^*$. There is a canonical isomorphism

$$C_V(X, (x_1, \ldots, x_n, y), (M_1, \ldots, M_n, V)) \simeq C_V(X, (x_1, \ldots, x_n), (M_1, \ldots, M_n)), \tag{19}$$

given by $\varphi \mapsto \varphi|_{M_1 \otimes \cdots \otimes M_N \otimes |0\rangle}$.

Remark 5.5. — Let V be a rational vertex algebra (see §3.4). Then we obtain a functor from the category of pointed algebraic curves with insertions of simple V–modules at the points to the category of vector spaces, which sends $(X,(x_i),(M_i))$ to $C_V(X,(x_i),(M_i))_{i=1}^n$. This is a version of *modular functor* corresponding to V [Se, Ga, Z2]. It is known that in the case of rational vertex algebras $L_k(\mathfrak{g})$ or $L_{c(p,q)}$ the spaces of conformal blocks are finite-dimensional. Moreover, the functor can be extended to the category of pointed stable curves with insertions. It then satisfies a factorization property, which expresses the space of conformal blocks associated to a singular curve with a double point in terms of conformal blocks associated to its normalization. The same is expected to be true for general rational vertex algebras.

5.4. Functional realization

The spaces $C_V(X,(x_i),(M_i))_{i=1}^n$ with varying points $x_1,\ldots,x_n$ can be organized into a vector bundle on $\overset{\circ}{X}{}^n := X^n \backslash \Delta$ (where Δ is the union of all diagonals) with a flat connection, which is a subbundle of $(\mathcal{M}_1 \boxtimes \cdots \boxtimes \mathcal{M}_n)^*$. Consider for example the spaces $C_V(X,(x_i),(V))_{i=1}^n$. By (19), all of them are canonically isomorphic to each other and to $C_V(X,x,V)$. Hence the corresponding bundles of conformal blocks are canonically trivialized. For $\varphi \in C_V(X,x,V)$, let φ_n denote the corresponding element of $C_V(X,(x_i),(V))_{i=1}^n$. Let $A_i(x_i)$ be local sections of $\mathcal{V}$ near x_i. Evaluating our conformal blocks on them, we obtain what physicists call the chiral correlation functions

$$\varphi_n(A_1(x_1)\otimes\cdots\otimes A_n(x_n)) \sim \langle A_1(x_1)\ldots A_n(x_n)\rangle_\varphi. \tag{20}$$

Moreover, we can explicitly describe these functions when x_i's are very close to each other, in the neighborhood of a point $x \in X$: if we choose a formal coordinate z at x and denote the coordinate of the point x_i by z_i, then

$$\varphi_n(A_1(z_1)\otimes\cdots\otimes A_n(z_n)) = \varphi(Y(A_1,z_1)\ldots Y(A_n,z_n)|0\rangle). \tag{21}$$

We obtain that the restriction of the n–point chiral correlation function to $\overset{\circ}{D}{}^n_x = D^n_x\backslash\Delta := \operatorname{Spec}\mathbb{C}[[z_1,\ldots,z_n]][(z_i-z_j)^{-1}]_{i\neq j}$ coincides with the n–point function corresponding to the functional φ introduced in §2.7.

According to Remark 4.2, the matrix elements given by the right hand side of formula (21) give rise to a horizontal section of $(\mathcal{V}^{\boxtimes n})^*$ on $\overset{\circ}{D}{}^n_x$. Therefore by Proposition 2.7 we obtain an embedding $\mathcal{V}_x^* \hookrightarrow \Gamma_\nabla(\overset{\circ}{D}{}^n_x,(\mathcal{V}^{\boxtimes n})^*)$, where Γ_∇ stands for the space of horizontal sections. If $\varphi \in \mathcal{V}_x^*$ is a conformal block, then Definition 5.1 implies that the image of $\varphi \in C_V(X,x,V)$ in $\Gamma_\nabla(\overset{\circ}{D}{}^n_x,(\mathcal{V}^{\boxtimes n})^*)$ extends to a horizontal section φ_n of $(\mathcal{V}^{\boxtimes n})^*$ over $\overset{\circ}{X}{}^n$. Thus, we obtain an embedding $C_V(X,x,V) \hookrightarrow \Gamma(\overset{\circ}{X}{}^n,(\mathcal{V}^{\boxtimes n})^*) = \Gamma_\nabla(X^n, j_*j^*(\mathcal{V}^{\boxtimes n})^*)$, where $j : \overset{\circ}{X}{}^n \hookrightarrow X^n$.

Imposing the bootstrap conditions on the diagonals from Proposition 2.7, we obtain a quotient $\mathcal{V}_n^*$ of $j_* j^* (\mathcal{V}^{\boxtimes n})^*$ (the precise definition of $\mathcal{V}_n^*$ uses the operation $\mathcal{Y}^{(2)}$ introduced in Theorem 7.1). Since by construction our sections satisfy the bootstrap conditions, we have embeddings $\mathcal{V}_x^* \hookrightarrow \Gamma_\nabla(D_x^n, \mathcal{V}_n^*)$ and $C_V(X, x, V) \hookrightarrow \Gamma_\nabla(X^n, \mathcal{V}_n^*)$. A remarkable fact is that these maps are isomorphisms for all $n > 1$. Thus, using correlation functions we obtain *functional realizations* of the (dual space of) a vertex algebra and its spaces of conformal blocks.

Remark 5.6. — Note that neither $(\mathcal{V}^{\boxtimes n})^*$ nor $\mathcal{V}_n^*$ is quasicoherent as an $\mathcal{O}_{X^n}$–module. It is more convenient to work with the dual sheaves $\mathcal{V}^{\boxtimes n}$ and $\mathcal{V}_n$, which are quasicoherent. The sheaf $\mathcal{V}_n$ on X^n is defined by Beilinson and Drinfeld in their construction of factorization algebra corresponding to $\mathcal{V}$ (see §7). Note that the space $\Gamma_\nabla(X, \mathcal{V}_n^*)$ is dual to the top deRham cohomology $H^n_{\mathrm{dR}}(X, \mathcal{V}_n \otimes \Omega_{X^n})$ (which is therefore isomorphic to $H(X, x, V)$).

In the case of Heisenberg algebra, the functional realization can be simplified, because we can use the sections of the exterior products of the canonical bundle Ω instead of the bundle Π^* (here we choose $\frac{1}{2} b_{-1}^2$ as the conformal vector, so that $b(z)dz$ transforms as a one-form). Namely, as in §2.7 we assign to $\varphi \in \Pi_x^*$ the collection of polydifferentials

$$\varphi(b(z_1) \ldots b(z_n)|0\rangle) dz_1 \ldots dz_n.$$

These are sections of $\Omega^{\boxtimes n}(2\Delta)$ over D_x^n, which are symmetric and satisfy the bootstrap condition (12). Let Ω_∞ be the sheaf on X, whose sections over $U \subset X$ are collections of sections of $\Omega^{\boxtimes n}(2\Delta)$ over U^n, for $n \geq 0$, satisfying these conditions (the bootstrap condition makes sense globally because of the identification $\Omega^{\boxtimes 2}(2\Delta)/\Omega^{\boxtimes 2}(\Delta) \simeq \mathcal{O}$). Then we have canonical isomorphisms $\Pi_x^* \simeq \Gamma(D_x, \Omega_\infty), C(X, x, \pi) \simeq \Gamma(X, \Omega_\infty)$.

The same construction can be applied to the Kac-Moody vertex algebras, see [BD1].

Remark 5.7. — If we consider conformal blocks with general V–module insertions, then the horizontal sections of the corresponding flat bundle of conformal blocks over $\overset{\circ}{X}{}^n$ will have non-trivial monodromy around the diagonals. For example, in the case of the Kac-Moody vertex algebra $V_k(\mathfrak{g})$ and $X = \mathbb{P}^1$, these sections are solutions of the Knizhnik-Zamolodchikov equations, and the monodromy matrices are given by the R–matrices of the quantum group $U_q(\mathfrak{g})$, see [TK, SV2].

6. SHEAVES OF CONFORMAL BLOCKS ON MODULI SPACES

In the previous section we associated the space of conformal blocks to a vertex algebra V, a smooth curve X and a point x of X. Now we want to understand the behavior of these spaces as we move x along X and vary the complex structure on X. More precisely, we wish to organize the spaces of conformal blocks into a sheaf on the moduli space $\mathfrak{M}_{g,1}$ of smooth pointed curves of genus g (here and below by moduli space we mean the corresponding moduli stack in the smooth topology; $\mathcal{D}$–modules on algebraic stacks are defined in [BB, BD2]). Actually, for technical reasons we prefer to work with the spaces of coinvariants (also called covacua), which are dual to the spaces of conformal blocks. One can construct in a straightforward way a quasicoherent sheaf on $\mathfrak{M}_{g,1}$, whose fiber at (X, x) is the space of coinvariants attached to (X, x). But in fact this sheaf also carries a structure of (twisted) $\mathcal{D}$–module on $\mathfrak{M}_{g,1}$, i.e., we can canonically identify the (projectivizations of) the spaces of coinvariants attached to infinitesimally nearby points of $\mathfrak{M}_{g,1}$ (actually, this $\mathcal{D}$–module always descends to $\mathfrak{M}_g$). The key fact used in the proof of this statement is the "Virasoro uniformization" of $\mathfrak{M}_{g,1}$. Namely, the Lie algebra $\operatorname{Der} \mathcal{K}$ acts transitively on the moduli space of triples (X, x, z), where (X, x) are as above and z is a formal coordinate at x; this action is obtained by "gluing" X from D_x and $X \backslash x$ (see Theorem 6.1).

In the case when the vertex algebra V is rational, the corresponding twisted $\mathcal{D}$–module on $\mathfrak{M}_g$ is believed to be coherent, i.e., isomorphic to the sheaf of sections of a vector bundle of finite rank equipped with a projectively flat connection (this picture was first suggested by Friedan and Shenker [FrS]). This is known to be true in the case of the Kac-Moody vertex algebra $L_k(\mathfrak{g})$ [TUY] (see also [So]) and the minimal models of the Virasoro algebra [BFM]. Moreover, in those cases the $\mathcal{D}$–module on $\mathfrak{M}_g$ can be extended to a $\mathcal{D}$–module with regular singularities on the Deligne-Mumford compactification $\overline{\mathfrak{M}}_g$, and the dimensions of the fibers at the boundary are equal to those in $\mathfrak{M}_g$. This allows one to compute these dimensions by "Verlinde formula" from the dimensions of the spaces of coinvariants attached to $\mathbb{P}^1$ with three points. The latter numbers, called the fusion rules, can in turn be found from the matrix of the modular transformation $\tau \mapsto -1/\tau$ acting on the space of characters (see Theorem 3.2). The same pattern is believed to hold for other rational vertex algebras, but as far as I know this has not been proved in general.

The construction of the $\mathcal{D}$–module structure on the sheaf of coinvariants is a special case of the general formalism of localization of modules over Harish-Chandra pairs, due to Beilinson-Bernstein [BB], which we now briefly recall.

6.1. Generalities on localization

A *Harish–Chandra pair* is a pair $(\mathfrak{g}, K)$ where $\mathfrak{g}$ is a Lie algebra, K is a Lie group, such that $\mathfrak{k} = \operatorname{Lie} K$ is embedded into $\mathfrak{g}$, and an action Ad of K on $\mathfrak{g}$ compatible with the adjoint action of K on $\mathfrak{k} \subset \mathfrak{g}$ and the action of $\mathfrak{k}$ on $\mathfrak{g}$.

A $(\mathfrak{g}, K)$*–action* on a scheme Z is the data of an action of $\mathfrak{g}$ on Z (that is, a homomorphism $\rho : \mathfrak{g} \to \Theta_Z$), together with an action of K on Z, such that (1) the differential of the K–action is the restriction of the $\mathfrak{g}$–action to $\mathfrak{k}$, and (2) $\rho(Ad_k(a)) = k\rho(a)k^{-1}$. A $(\mathfrak{g}, K)$–action is called *transitive* if the map $\mathfrak{g} \otimes \mathcal{O}_Z \to \Theta_Z$ is surjective, and *simply transitive* if this map is an isomorphism.

A $(\mathfrak{g}, K)$–*structure* on a scheme S is a principal K–bundle $\pi : Z \to S$ together with a simply transitive $(\mathfrak{g}, K)$–action on Z which extends the fiberwise action of K.

An example of $(\mathfrak{g}, K)$–structure is the homogeneous space $S = G/K$, where G is a finite–dimensional Lie group, $\mathfrak{g} = \operatorname{Lie} G$, K is a Lie subgroup of G, and we take the obvious right action of $(\mathfrak{g}, K)$ on $\widehat{X} = G$. In the finite–dimensional setting, any space with a $(\mathfrak{g}, K)$–structure is locally of this form. However, in the infinite-dimensional setting this is no longer so. An example is the $(\operatorname{Der} \mathcal{O}, \operatorname{Aut} \mathcal{O})$–structure $Z = \widehat{X} \to X$ on a smooth curve X, which we used above. This example can be generalized to the case when S is an arbitrary smooth scheme. Define $\widehat{S}$ to be the scheme of pairs $(x, \vec{t}_x)$, where $x \in S$ and $\vec{t}_x$ is a system of formal coordinates at x. Set $\mathfrak{g} = \operatorname{Der} \mathbb{C}[[z_1, \ldots, z_n]], K = \operatorname{Aut} \mathbb{C}[[z_1, \ldots, z_n]]$. Then $\widehat{S}$ is a $(\mathfrak{g}, K)$–structure on S, first considered by Gelfand, Kazhdan and Fuchs [GKF].

Let V be a $(\mathfrak{g}, K)$–module, i.e., a vector space together with actions of $\mathfrak{g}$ and K satisfying the obvious compatibility condition. Then we define a flat vector bundle $\mathcal{V}$ on any variety S with a $(\mathfrak{g}, K)$–structure Z. Namely, as a vector bundle, $\mathcal{V} = Z \underset{K}{\times} V$. Since by assumption the $\mathfrak{g}$–action on Z is simply transitive, it gives rise to a flat connection on the trivial vector bundle $Z \times V$ over Z, which descends to $\mathcal{V}$, because the actions of K and $\mathfrak{g}$ are compatible. In the special case of the $(\operatorname{Der} \mathcal{O}, \operatorname{Aut} \mathcal{O})$–structure $\widehat{X}$ over a smooth curve X we obtain the flat bundle $\mathcal{V}$ on X from §4.2.

Now consider the case when the $\mathfrak{g}$–action is transitive but not simply transitive (there are stabilizers). Then we can still construct a vector bundle $\mathcal{V}$ on S equipped with a K–equivariant action of the Lie algebroid $\mathfrak{g}_Z = \mathfrak{g} \otimes \mathcal{O}_Z$, but we cannot obtain an action of Θ_Z on $\mathcal{V}$ because the map $a : \mathfrak{g}_Z \to \Theta_Z$ is no longer an isomorphism. Nevertheless, Θ_Z will act on any sheaf, on which $\operatorname{Ker} a$ acts by 0. In particular, the sheaf $V \otimes \mathcal{O}_Z / \operatorname{Ker} a \cdot (V \otimes \mathcal{O}_Z) \simeq \mathcal{D}_Z \otimes_{U\mathfrak{g}} V$ gives rise to a K–equivariant $\mathcal{D}$–module on Z, and hence to a $\mathcal{D}$–module on S, denoted $\Delta(V)$ and called the *localization* of V on S.

More generally, suppose that V is a module over a Lie algebra $\mathfrak{l}$, which contains $\mathfrak{g}$ as a Lie subalgebra and carries a compatible K–action. Suppose also that we are given a K–equivariant Lie algebra subsheaf $\widetilde{\mathfrak{l}}$ of the constant sheaf of Lie algebras $\mathfrak{l} \otimes \mathcal{O}_Z$, which is preserved by the natural action of the Lie algebroid $\mathfrak{g}_Z$. Then if $\widetilde{\mathfrak{l}}$ contains $\operatorname{Ker} a$, the sheaf $V \otimes \mathcal{O}_Z / \widetilde{\mathfrak{l}} \cdot (V \otimes \mathcal{O}_Z)$ is a K–equivariant $\mathcal{D}$–module on Z, which descends to a $\mathcal{D}$–module $\widetilde{\Delta}(V)$ on S.

Let us describe the fibers of $\widetilde{\Delta}(V)$ (considered as an $\mathcal{O}_S$–module). For $s \in S$, let Z_s be the fiber of Z at s, and $\mathfrak{l}_s = Z_s \underset{K}{\times} \mathfrak{l}$, $\mathcal{V}_s = Z_s \underset{K}{\times} V$. The fibers of $\widetilde{\mathfrak{l}}$ at the points of $Z_s \subset M$ give rise to a well-defined Lie subalgebra $\widetilde{\mathfrak{l}}_s$ of $\mathfrak{l}_s$. Then the fiber of $\widetilde{\Delta}(V)$ at $s \in S$ is canonically isomorphic to the space of coinvariants $\mathcal{V}_s / \widetilde{\mathfrak{l}}_s \cdot \mathcal{V}_s$.

6.2. Localization on the moduli space of curves

We apply the above formalism in the case when S is the moduli space $\mathfrak{M}_{g,1}$ of smooth pointed curves of genus $g > 1$, and Z is the moduli space $\widehat{\mathfrak{M}}_{g,1}$ of triples (X, x, z), where $(X, x) \in \mathfrak{M}_{g,1}$ and z is a formal coordinate at x. Clearly, $\widehat{\mathfrak{M}}_{g,1}$ is an $\operatorname{Aut}\mathcal{O}$–bundle over $\mathfrak{M}_{g,1}$.

THEOREM 6.1 (cf. [ADKP, BS, Ko, TUY]). — *$\widehat{\mathfrak{M}}_{g,1}$ carries a transitive action of* $\operatorname{Der}\mathcal{K}$ *compatible with the* $\operatorname{Aut}\mathcal{O}$*–action along the fibers.*

The action of the corresponding ind–group $\operatorname{Aut}\mathcal{K}$ is defined by the "gluing" construction. If (X, x, z) is an R–point of $\widehat{\mathfrak{M}}_{g,1}$, and $\rho \in \operatorname{Aut}\mathcal{K}(R)$, we construct a new R–point (X_ρ, x_ρ, z_ρ) of $\widehat{\mathfrak{M}}_{g,1}$ by "gluing" the formal neighborhood of x in X with $X\backslash x$ with a "twist" by ρ.

Now we are in the situation of §6.1, with $\mathfrak{g} = \operatorname{Der}\mathcal{K}$ and $K = \operatorname{Aut}\mathcal{O}$. Denote by $\mathcal{A}$ the Lie algebroid $\operatorname{Der}\mathcal{K}\widehat{\otimes}\mathcal{O}_{\widehat{\mathfrak{M}}_{g,1}}$ on $\widehat{\mathfrak{M}}_{g,1}$. By construction, the kernel of the corresponding homomorphism $a : \mathcal{A} \to \Theta$ is the subsheaf $\mathcal{A}_{\mathrm{out}}$ of $\mathcal{A}$, whose fiber at $(X, x) \in \mathfrak{M}_{g,1}$ is $\operatorname{Vect}(X\backslash x)$. Applying the construction of §6.1 to $V = \mathrm{Vir}_0$ (the Virasoro vertex algebra with $c = 0$), we obtain a $\mathcal{D}$–module $\Delta(\mathrm{Vir}_0)$ on $\mathfrak{M}_{g,1}$, whose fibers are the spaces of coinvariants $\mathrm{Vir}_0 / \operatorname{Vect}(X\backslash x) \cdot \mathrm{Vir}_0$.

More generally, let V be an arbitrary conformal vertex algebra with central charge 0. Then it is a module over the Lie algebra $\mathfrak{l} = U(V)$ from §5.3. Let $\widetilde{\mathfrak{l}} = \mathcal{U}(V)_{\mathrm{out}}$ be the subsheaf of $U(V)\widehat{\otimes}\mathcal{O}_{\widehat{\mathfrak{M}}_{g,1}}$, whose fiber at $(X, x, z) \in \mathfrak{M}_{g,1}$ equals $U_{X\backslash x}(V)$ (see §5.3). Note that $U(V)$ contains $\operatorname{Der}\mathcal{K}$, and $\mathcal{U}(V)_{\mathrm{out}}$ contains $\mathcal{A}_{\mathrm{out}}$. Moreover, we have:

LEMMA 6.2. — *$\mathcal{U}(V)_{\mathrm{out}}$ is preserved by the action of the Lie algebroid $\mathcal{A}$.*

This is equivalent to the statement that $U_{X\backslash x}(V)$ and $U_{X_\rho\backslash x_\rho}(V)$ are conjugate by ρ, which follows from the fact that formula (16) is valid for any $\rho \in \operatorname{Aut}\mathcal{K}$.

So we are again in the situation of §6.1, and hence we obtain a $\mathcal{D}$–module $\widetilde{\Delta}(V)$ on $\mathfrak{M}_{g,1}$, whose fiber at (X, x) is precisely the space of coinvariants $H(X, x, V) = \mathcal{V}_x/U_{X\backslash x}(\mathcal{V}_x) \cdot \mathcal{V}_x$ (see §5.3), which is what we wanted.

In the case of vertex algebras with non-zero central charge, we need to modify the general construction of §6.1 as follows. Suppose that $\mathfrak{g}$ has a central extension $\widehat{\mathfrak{g}}$ which splits over $\mathfrak{k}$, and such that the extension

$$0 \to \mathcal{O}_Z \to \widehat{\mathfrak{g}} \otimes \mathcal{O}_Z \to \mathfrak{g}_Z \to 0 \tag{22}$$

splits over the kernel of $a : \mathfrak{g}_Z \to \Theta_Z$. Then the quotient $\mathfrak{g}'_Z$ of $\widehat{\mathfrak{g}} \otimes \mathcal{O}_Z$ by $\operatorname{Ker} a$ is an extension of Θ_Z by $\mathcal{O}_Z$, with a natural Lie algebroid structure. The enveloping algebra of this Lie algebroid is a sheaf $\mathcal{D}'_Z$ of twisted differential operators (TDO) on Z, see [BB]. Further, $\mathfrak{g}'_Z$ descends to a Lie algebroid on S, and gives rise to a sheaf of TDO $\mathcal{D}'_S$.

Now if V is a $(\widehat{\mathfrak{g}}, K)$–module, we obtain a $\mathcal{D}'_Z$–module $\mathcal{D}'_Z \otimes_{U\widehat{\mathfrak{g}}} V$. By construction, it is K–equivariant, and hence descends to a $\mathcal{D}'_S$–module on S. More generally, suppose that $\widehat{\mathfrak{g}}$ is a Lie subalgebra of $\widehat{\mathfrak{l}}$, and we are given a subsheaf $\widetilde{\mathfrak{l}}$ of $\widehat{\mathfrak{l}} \otimes \mathcal{O}_Z$, containing $\operatorname{Ker} a$ and preserved by the action of $\widehat{\mathfrak{g}} \otimes \mathcal{O}_Z$. Then the sheaf $V \otimes \mathcal{O}_Z / \widetilde{\mathfrak{l}} \cdot V \otimes \mathcal{O}_Z$ is a K–equivariant $\mathcal{D}'_Z$–module on Z, which descends to a $\mathcal{D}'_S$–module (still denoted by $\widetilde{\Delta}(V)$). Its fibers are the coinvariants $\mathcal{V}_s/\widetilde{\mathfrak{l}}_s \cdot \mathcal{V}_s$.

Let V be a conformal vertex algebra with an arbitrary central charge c. Then by the residue theorem the corresponding sequence (22) with $\widehat{\mathfrak{g}} = Vir$ splits over $\operatorname{Ker} a = \mathcal{A}_{\mathrm{out}}$. Hence there exists a TDO sheaf $\mathcal{D}_c$ on $\mathfrak{M}_{g,1}$ and a $\mathcal{D}_c$–module $\widetilde{\Delta}(V)$ on $\mathfrak{M}_{g,1}$ whose fiber at (X, x) is the space of coinvariants $H(X, x, V)$. Actually, $\widetilde{\Delta}(V)$ descends to $\mathfrak{M}_g$. Explicit computation shows that the sheaf of differential operators on the determinant line bundle over $\mathfrak{M}_g$ corresponding to the sheaf of relative λ–differentials on the universal curve is $\mathcal{D}_{c(\lambda)}$, where $c(\lambda) = -12\lambda^2 + 12\lambda - 2$ (see [BS, BFM]).

It is straightforward to generalize the above construction to the case of multiple points with arbitrary V–module insertions $M_1, \ldots, M_n$. In that case we obtain a twisted $\mathcal{D}$–module on the moduli space $\mathfrak{M}'_{g,n}$ of n–pointed curves with non-zero tangent vectors at the points.

6.3. Localization on other moduli spaces

In the previous section we constructed twisted $\mathcal{D}$–modules on the moduli spaces of pointed curves using the interior action of the Harish-Chandra pair $(Vir, \operatorname{Aut} \mathcal{O})$ on conformal vertex algebras and its modules. This construction can be generalized to the case of other moduli spaces if we consider interior actions of other Harish–Chandra pairs.

For example, consider the Kac-Moody vertex algebra $V_k(\mathfrak{g})$. It carries an action of the Harish-Chandra pair $(\widehat{\mathfrak{g}}, G(\mathcal{O}))$, where G is the simply-connected group with Lie algebra $\mathfrak{g}$. The corresponding moduli space is the moduli space $\mathfrak{M}_G(X)$ of G–bundles on a curve X. For $x \in X$, let $\widehat{\mathfrak{M}}_G(X)$ be the moduli spaces of pairs $(\mathcal{P}, s)$, where $\mathcal{P}$ is a G–bundle on X, and s is a trivialization of $\mathcal{P}|_{D_x}$. Then $\widehat{\mathfrak{M}}_G(X)$ is a principal $G(\mathcal{O}_x)$–bundle over $\mathfrak{M}_G(X)$, equipped with a compatible $\mathfrak{g}(\mathcal{K}_x)$–action. The latter is given by the gluing construction similar to that explained in the proof of Theorem 6.1.

Now we can apply the construction of §6.1 to the $(\widehat{\mathfrak{g}}_x, G(\mathcal{O}_x))$–module $\mathcal{V}_k(\mathfrak{g})_x$, and obtain a twisted $\mathcal{D}$–module on $\mathfrak{M}_G(X)$. Its fiber at $\mathcal{P} \in \mathfrak{M}_G(X)$ is the space of coinvariants $\mathcal{V}_k(\mathfrak{g})_x/\mathfrak{g}^{\mathcal{P}}_{x,\mathrm{out}} \cdot \mathcal{V}_k(\mathfrak{g})_x$, where $\mathfrak{g}^{\mathcal{P}}_{x,\mathrm{out}}$ is the Lie algebra of sections of $\mathcal{P} \underset{G}{\times} \mathfrak{g}$ over $X \backslash x$.

More generally, let V be a vertex algebra equipped with a $\widehat{\mathfrak{g}}$–structure of level $k \neq -h^\vee$ (see §4.4). Then V carries an action of the Harish-Chandra pair $(\widehat{\mathfrak{g}}_x, G(\mathcal{O}_x))$. In the same way as in the case of the Virasoro algebra, we attach to V

a twisted $\mathcal{D}$–module on $\mathfrak{M}_G(X)$, whose fiber at $\mathcal{P}$ is the twisted space of coinvariants $H^{\mathcal{P}}(X, x, V)$, dual to the space $C^{\mathcal{P}}(X, x, V)$ defined in Remark 5.2.

For the Heisenberg vertex algebra π, the corresponding moduli space is the Jacobian $J(X)$ of X. The bundle $\widehat{J}(X)$ over $J(X)$, which parameterizes line bundles together with trivializations on D_x, carries a transitive action of the abelian Lie algebra $\mathcal{K}$. Applying the above construction, we assign to any conformal vertex algebra equipped with an embedding $\pi \to V$ a twisted $\mathcal{D}$–module on $J(X)$, whose fiber at $\mathcal{L} \in J(X)$ is the twisted space of coinvariants $H^{\mathcal{L}}(X, x, V)$. The corresponding TDO sheaf is the sheaf of differential operators acting on the theta line bundle on $J(X)$.

By considering the action of the semi-direct product $Vir \ltimes \widehat{\mathfrak{g}}$ on V we obtain a twisted $\mathcal{D}$-module on the moduli space of curves and G–bundles on them. We can further generalize the construction by inserting modules at points of the curve, etc.

6.4. Local and global structure of moduli spaces

In the simplest cases the twisted $\mathcal{D}$–modules obtained by the above localization construction are the twisted sheaves $\mathcal{D}$ themselves (considered as left modules over themselves). For example, the sheaf $\Delta(V_k(\mathfrak{g}))$ on $\mathfrak{M}_G(X)$ is the sheaf $\mathcal{D}_k$ of differential operators acting on the kth power of the determinant line bundle $\mathcal{L}_G$ on $\mathfrak{M}_G(X)$ (the ample generator of the Picard group of $\mathfrak{M}_G(X)$). The dual space to the stalk of $\mathcal{D}_k$ at $\mathcal{P} \in \mathfrak{M}_G(X)$ is canonically identified with the space of sections of $\mathcal{L}_G^{\otimes k}$ on the formal neighborhood of $\mathcal{P}$ in $\mathfrak{M}_G(X)$. But according to our construction, this space is also isomorphic to the space of conformal blocks $C^{\mathcal{P}}(X, x, V_k(\mathfrak{g}))$. In particular, the coordinate ring of the formal deformation space of a G–bundle $\mathcal{P}$ on a curve X is isomorphic to $C^{\mathcal{P}}(X, x, V_0(\mathfrak{g}))$. Thus, using the description of conformal blocks in terms of correlation functions (see §5.4), we obtain a realization of the formal deformation space and a line bundle on it in terms of polydifferentials on powers of the curve X satisfying bootstrap conditions on the diagonals (see [BG, BD2, Gi]). On the other hand, if we replace X by D_x, the corresponding space of polydifferentials becomes $V_k(\mathfrak{g})^*$ (more precisely, its twist by the torsor of formal coordinates at x). Therefore the vertex algebra $V_k(\mathfrak{g})$ may be viewed as the local object responsible for deformations of G–bundles on curves.

Similarly, the Virasoro vertex algebra Vir_c and the Heisenberg vertex algebra π may be viewed as the local objects responsible for deformations of curves and line bundles on curves, respectively.

Optimistically, one may hope that any one–parameter family of “Verma module type” vertex algebras (such as Vir_c or $V_k(\mathfrak{g})$) is related to a moduli space of curves with some additional structures. It would be interesting to identify the deformation problems related to the most interesting examples of such vertex algebras (for instance, the $\mathcal{W}$–algebras of §3.7), and to construct the corresponding global moduli spaces. One can attach a “formal moduli space” to any vertex algebra V by taking the double quotient of the ind–group, whose Lie algebra is $U(V)$, by its “in” and “out” subgroups.

But this moduli space is very big, even in the familiar examples of Virasoro and Kac-Moody algebras. The question is to construct a smaller formal moduli space with a line bundle, whose space of sections is isomorphic to the space of conformal blocks $C(X, x, V)$.

While "Verma module type" vertex algebras are related to the local structure of moduli spaces, rational vertex algebras (see §3.4) can be used to describe the global structure.

For instance, the space of conformal blocks $C(X, x, L_k(\mathfrak{g})) \simeq \mathrm{Hom}_{\mathfrak{g}_{\mathrm{out}}}(L_k(\mathfrak{g}), \mathbb{C})$, is isomorphic to $\Gamma_k = \Gamma(\mathfrak{M}_G(X), \mathcal{L}_G^{\otimes k})$, see [BL, Fa, KNR]. The space $C(X, x, L_k(\mathfrak{g}))$ satisfies a factorization property: its dimension does not change under degenerations of the curve into curves with nodal singularities. This property allows one to find this dimension by computing the spaces of conformal blocks in the case of $\mathbb{P}^1$ and three points with $L_k(\mathfrak{g})$–module insertions (fusion rules). The resulting formula for $\dim C(X, x, L_k) = \dim \Gamma_k$ is called the Verlinde formula, see [TUY, So].

Since $\mathcal{L}_G$ is ample, we can recover the moduli space of semi-stable G–bundles on X as the Proj of the graded ring $\oplus_{k \geq 0} \Gamma_{Nk} = \oplus_{k \geq 0} C(X, x, L_{Nk}(\mathfrak{g}))$ for large N (the ring of "non-abelian theta functions"). Thus, if we could define the product on conformal blocks in a natural way, we would obtain a description of the moduli space. Using the correlation function description of conformal blocks Feigin and Stoyanovsky [FS] have identified the space $C(X, x, L_k(\mathfrak{g}))$ with the space of sections of a line bundle on the power of X satisfying certain conditions. For example, when $\mathfrak{g} = \mathfrak{sl}_2$ it is the space of sections of the line bundle $\Omega(2nx)^{\boxtimes nk}$ over X^{nk}, which are symmetric and vanish on the diagonals of codimension k (here $x \in X$, and $n \geq g$). In these terms the multiplication $\Gamma_k \times \Gamma_l \to \Gamma_{k+l}$ is just the composition of exterior tensor product and symmetrization of these sections. Thus we obtain an explicit description of the coordinate ring of the moduli space of semi-stable SL_2–bundles. On the other hand, replacing in the above description X by D_x, we obtain a functional realization of $L_k(\mathfrak{sl}_2)^*$.

Similarly, the space of conformal blocks corresponding to the lattice vertex super-algebra $V_{\sqrt{N}\mathbb{Z}}$ (see §3.3) may be identified with the space of theta functions of order N on $J(X)$, so we obtain the standard "functional realization" of the Jacobian $J(X)$.

6.5. Critical level

When $k = -h^\vee$ (which is called the critical level), the vertex algebra $V_k(\mathfrak{g})$ is not conformal. Nevertheless, $V_{-h^\vee}(\mathfrak{g})$ still carries an action of $\mathrm{Der}\,\mathcal{O}$ which satisfies the same properties as for non-critical levels. Therefore we can still attach to $V_{-h^\vee}$ a twisted $\mathcal{D}$–module on $\mathfrak{M}_G(X)$ (but we cannot vary the curve). This $\mathcal{D}$–module is just the sheaf $\mathcal{D}'$ of differential operators acting on the square root of the canonical bundle on $\mathfrak{M}_G(X)$ (see [BD2]).

Recall from §3.6 that the center of a vertex algebra V is the coset vertex algebra of the pair (V, V). This is a commutative vertex subalgebra of V. The center of $V_k(\mathfrak{g})$ is simply its subspace of $\mathfrak{g}[[z]]$–invariants. It is easy to show that this subspace equals $\mathbb{C}v_k$ if $k \neq -h^\vee$. In order to describe the center $\mathfrak{z}(\widehat{\mathfrak{g}})$ of $V_{-h^\vee}(\mathfrak{g})$ we need the concept of *opers*, due to Beilinson and Drinfeld [BD2].

Let G_{ad} be the adjoint group of $\mathfrak{g}$, and B_{ad} be its Borel subgroup. By definition, a $\mathfrak{g}$–oper on a smooth curve X is a G_{ad}–bundle on X with a (flat) connection ∇ and a reduction to B_{ad}, satisfying a certain transversality condition [BD2]. For example, an $\mathfrak{sl}_2$–oper is the same as a projective connection. The set of $\mathfrak{g}$–opers on X is the set of points of an affine space, denoted $\mathrm{Op}_{\mathfrak{g}}(X)$, which is a torsor over $\oplus_{i=1}^{\ell}\Gamma(X,\Omega^{d_i+1})$, where d_i's are the exponents of $\mathfrak{g}$. Denote by $A_{\mathfrak{g}}(X)$ the ring of functions on $\mathrm{Op}_{\mathfrak{g}}(X)$.

The above definition can also be applied when X is replaced by the disc $D = \mathrm{Spec}\,\mathbb{C}[[z]]$ (see [DS]). The corresponding ring $A_{\mathfrak{g}}(D)$ has a natural $(\mathrm{Der}\,\mathcal{O}, \mathrm{Aut}\,\mathcal{O})$–action, and is therefore a commutative vertex algebra (see §2.2). It is isomorphic to a limit of the $\mathcal{W}$–vertex algebra $\mathcal{W}_k(\mathfrak{g})$ as $k \to \infty$ (see §3.7), and because of that it is called the classical $\mathcal{W}$–algebra corresponding to $\mathfrak{g}$. For example, $A_{\mathfrak{sl}_2}(D)$ is a limit of Vir_c as $c \to \infty$. The following result was conjectured by Drinfeld and proved in [FF3].

Theorem 6.3. — *The center $\mathfrak{z}(\widehat{\mathfrak{g}})$ of $V_{-h^\vee}(\mathfrak{g})$ is isomorphic, as a commutative vertex algebra with* $\mathrm{Der}\,\mathcal{O}$*–action, to $A_{{}^L\mathfrak{g}}(D)$, where ${}^L\mathfrak{g}$ is the Langlands dual Lie algebra to $\mathfrak{g}$.*

In fact, more is true: both of these commutative vertex algebras carry what can be called Poisson vertex algebra (or coisson algebra, in the terminology of [BD3]) structures, which are preserved by this isomorphism.

The localization of $\mathfrak{z}(\widehat{\mathfrak{g}})$ on $\mathfrak{M}_G(X)$ is the "constant" $\mathcal{D}'$–module $H(X, x, \mathfrak{z}(\widehat{\mathfrak{g}})) \otimes \mathcal{O}_{\mathfrak{M}_G(X)}$. By functoriality, it is a commutative subalgebra of $\mathcal{D}'$. As shown in [BD3], the space of coinvariants of a commutative vertex algebra A is canonically isomorphic to the ring of functions on $\Gamma_\nabla(X, \mathrm{Spec}\,\mathcal{A})$. In our case we obtain: $\Gamma_\nabla(X, \mathrm{Spec}\,\mathcal{A}_{{}^L\mathfrak{g}}) \simeq \mathrm{Op}_{{}^L\mathfrak{g}}(X)$, and so $H(X, x, A_{{}^L\mathfrak{g}}(D)) \simeq A_{{}^L\mathfrak{g}}(X)$. Theorem 6.3 then implies:

Corollary 6.4. — *There is a homomorphism of algebras $A_{{}^L\mathfrak{g}}(X) \to \Gamma(\mathfrak{M}_G(X), \mathcal{D}')$.*

Beilinson and Drinfeld have shown in [BD2] that this map is an embedding, and if G is simply-connected, it is an isomorphism. In that case, each ${}^LG$–oper ρ gives us a point of the spectrum of the commutative algebra $\Gamma(\mathfrak{M}_G(X), \mathcal{D}')$, and hence a homomorphism $\widetilde{\rho} : \Gamma(\mathfrak{M}_G(X), \mathcal{D}') \to \mathbb{C}$. The $\mathcal{D}'$–module $\mathcal{D}'/(\mathcal{D}' \cdot m_{\widetilde{\rho}})$ on $\mathfrak{M}_G(X)$, where $m_{\widetilde{\rho}}$ is the kernel of $\widetilde{\rho}$, is holonomic. It is shown in [BD2] that the corresponding untwisted $\mathcal{D}$–module is a Hecke eigensheaf attached to ρ by the geometric Langlands correspondence.

6.6. Chiral deRham complex

Another application of the general localization pattern of §6.1 is the recent construction, due to Malikov, Schechtman and Vaintrob [MSV], of a sheaf of vertex superalgebras on any smooth scheme S called the chiral deRham complex. Under certain conditions, this sheaf has a purely even counterpart, the sheaf of chiral differential operators. Here we give a brief review of the construction of these sheaves.

Let Γ_N be the Weyl algebra associated to the symplectic vector space $\mathbb{C}((t))^N \oplus (\mathbb{C}((t))dt)^N$. It has topological generators $a_{i,n}, a^*_{i,n}, i = 1, \ldots, N; n \in \mathbb{Z}$, with relations

$$[a_{i,n}, a^*_{j,m}] = \delta_{i,j}\delta_{n,-m}, \qquad [a_{i,n}, a_{j,m}] = [a^*_{i,n}, a^*_{j,m}] = 0.$$

Denote by H_N the Fock representation of Γ_N generated by the vector $|0\rangle$, satisfying $a_{i,n}|0\rangle = 0$, $n \geq 0$; $a^*_{i,m}|0\rangle = 0, m > 0$. This is a vertex algebra, such that

$$Y(a_{i,-1}|0\rangle, z) = a(z) = \sum_{n\in\mathbb{Z}} a_{i,n} z^{-n-1}, \qquad Y(a^*_{i,0}|0\rangle, z) = a^*(z) = \sum_{n\in\mathbb{Z}} a^*_{i,n} z^{-n}.$$

Let $\bigwedge^\bullet_N$ be the fermionic analogue of H_N defined in the same way as in §3.7. This is the Fock module over the Clifford algebra with generators $\psi_{i,n}$, $\psi^*_{i,n}$, $i = 1, \ldots, N, n \in \mathbb{Z}$, equipped with an additional $\mathbb{Z}$–gradation. It is shown in [MSV] that the vertex algebra structure on H_N may be extended to $\widehat{H}_N = H_N \underset{\mathbb{C}[a^*_{i,0}]}{\otimes} \mathbb{C}[[a^*_{i,0}]]$. Let $\widehat{\Omega}^\bullet_N = \widehat{H}_N \otimes \bigwedge^\bullet_N$. It carries a differential $d = \operatorname{Res} \sum_i a_i(z)\psi^*_i(z)$, which makes it into a complex.

Denote by W_N (resp., $\mathcal{O}_N$, Ω^1_N) the topological Lie algebra of vector fields (resp., ring of functions, module of differentials) on $\operatorname{Spec} \mathbb{C}[[t_1, \ldots, t_N]]$. Define a map $W_N \to U(\widehat{\Omega}^\bullet_N)$ (the completion of the Lie algebra of Fourier coefficients of fields from $\widehat{\Omega}^\bullet_N$) sending

$$f(t_i)\partial_{t_j} \mapsto \operatorname{Res}\left({:}f(a^*_i(z))a_j(z){:} + \sum_{k=1}^N {:}{:}(\partial_{t_k} f)(a^*_i(z))\psi^*_k(z)\psi_j(z){:}{:} \right).$$

Explicit computation shows that it is a homomorphism of Lie algebras (this is an example of "supersymmetric cancellation of anomalies") [FF2, MSV]. We obtain a W_N–action on $\widehat{\Omega}^\bullet_N$, which can be exponentiated to an action of the Harish-Chandra pair $(W_N, \operatorname{Aut} \mathcal{O}_N)$. Furthermore, because the $(W_N, \operatorname{Aut} \mathcal{O}_N)$–action comes from the residues of vertex operators, it preserves the vertex algebra structure on V (so the group $\operatorname{Aut} \mathcal{O}_N$ is an example of the group of internal symmetries from §4.4). Recall from §6.1 that any smooth scheme S of dimension N carries a canonical $(W_N, \operatorname{Aut} \mathcal{O}_N)$–structure $\widehat{S}$. Applying the construction of §6.1, we attach to $\widehat{\Omega}^\bullet_N$ a $\mathcal{D}$–module on S. The sheaf of horizontal sections of this $\mathcal{D}$–module is a sheaf of vertex superalgebras (with a structure of complex) on S. This is the *chiral deRham complex* of S, introduced in [MSV]. For its applications to mirror symmetry and elliptic cohomology, see [BoL, Bo1].

Now consider a purely bosonic analogue of this construction (i.e., replace $\widehat{\Omega}_N$ with $\widehat{H}_N$). The Lie algebra $U(\widehat{H}_N)$ of Fourier coefficients of the fields from $\widehat{H}_N$ has a filtration $A_N^{\leq i}, i \geq 0$, by the order of $a_{i,n}$'s. Note that $A_N^{\leq 1}$ is a Lie subalgebra of $U(\widehat{H}_N)$ and A_N^0 is its (abelian) ideal. Denote by $\mathcal{T}_N$ the Lie algebra $A_N^{\leq 1}/A_N^0$. Intuitively, $U(\widehat{H}_N)$ is the algebra of differential operators on the topological vector space $\mathbb{C}((t))^N$ with the standard filtration, $A^{(0)}$ (resp., $\mathcal{T}_N$) is the space of functions (resp., the Lie algebra of vector fields) on it, and $\widehat{H}_N$ is the module of "delta–functions supported on $\mathbb{C}[[t]]^N$".

In contrast to the finite-dimensional case, the extension

$$0 \to A_N^0 \to A_N^{\leq 1} \to \mathcal{T}_N \to 0$$

does not split [FF2]. Because of that, the naive map $W_N \to U(\widehat{H}_N)$ sending $f(t_i)\partial_{t_j}$ to $\operatorname{Res} {:}f(a_i^*(z))a_j(z){:}$ is not a Lie algebra homomorphism. Instead, we obtain an extension

$$0 \to \Omega_N^1/d\mathcal{O}_N \to \widetilde{W}_N \to W_N \to 0 \tag{23}$$

of W_N by its module $\Omega_N^1/d\mathcal{O}_N$, which is embedded into $A_N^{\leq 1}$ by the formula $g(t_i)dt_j \mapsto \operatorname{Res} g(a_i^*(z))\partial_z a_j^*(z)$. If the sequence (23) were split, we would obtain a $(W_N, \operatorname{Aut}\mathcal{O}_N)$–action on $\widehat{H}_N$ and associate to $\widehat{H}_N$ a $\mathcal{D}$–module on S in the same way as above.

But in reality the extension (23) does not split for $N > 1$ (note that $\Omega_1^1/d\mathcal{O}_1 = 0$). Hence in general we only have an action on $\widehat{H}_N$ of the Harish-Chandra pair $(\widetilde{W}_N, \widetilde{\operatorname{Aut}\mathcal{O}_N})$, where $\widetilde{\operatorname{Aut}\mathcal{O}_N}$ is an extension of $\operatorname{Aut}\mathcal{O}_N$ by the additive group $\Omega_N^1/d\mathcal{O}_N$. This action again preserves the vertex algebra structure on V. To apply the construction of §6.1, we need a lifting of the $(W_N, \operatorname{Aut}\mathcal{O}_N)$–structure $\widehat{S}$ on S to a $(\widetilde{W}_N, \widetilde{\operatorname{Aut}\mathcal{O}_N})$–structure. Such liftings form a *gerbe* on S (in the analytic topology) with the lien $\Omega_S^1/d\mathcal{O}_S \simeq \Omega_{S,\mathrm{cl}}^2$, the sheaf of closed holomorphic two-forms on S. The equivalence class of this gerbe in $H^2(S, \Omega_{S,\mathrm{cl}}^2)$ is computed in [GMS]. If this class equals 0, then we can lift $\widehat{S}$ to an $(\widetilde{W}_N, \widetilde{\operatorname{Aut}\mathcal{O}_N})$–structure $\widetilde{S}$ on S, and the set of isomorphism classes of such liftings is an $H^1(S, \Omega_{S,\mathrm{cl}}^2)$–torsor. Applying the construction of §6.1, we can then attach to $\widehat{H}_N$ a $\mathcal{D}$–module on S. The sheaf of horizontal sections of this $\mathcal{D}$–module is a sheaf of vertex algebras on S. This is the sheaf of *chiral differential operators* on S (corresponding to $\widetilde{S}$) introduced in [MSV, GMS]. As shown in [GMS], the construction becomes more subtle in Zariski topology; in this case one naturally obtains a gerbe with the lien $\Omega_S^2 \to \Omega_{S,\mathrm{cl}}^3$.

In the case when S is the flag manifold G/B of a simple Lie algebra $\mathfrak{g}$, one can construct the sheaf of chiral differential operators in a more direct way. The action of $\mathfrak{g}$ on the flag manifold G/B gives rise to a homomorphism $\mathfrak{g}((t)) \to \mathcal{T}_N$, where $N = \dim G/B$. It is shown in [Wa, FF1, FF2] that it can be lifted to a homomorphism $\widehat{\mathfrak{g}} \to A_N^{\leq 1}$ of level $-h^\vee$ (Wakimoto realization). Moreover, it gives rise to a homomorphism of vertex algebras $V_{-h^\vee}(\mathfrak{g}) \to H_N$. Hence we obtain an embedding of the constant subalgebra $\mathfrak{g}$ of $\widehat{\mathfrak{g}}$ into $\widetilde{W}_N$, and a $(\mathfrak{g}, B)$–action on $\widehat{H}_N$. On the other hand, G/B has a natural $(\mathfrak{g}, B)$–structure, namely G. Applying the localization construction to the Harish-Chandra pair $(\mathfrak{g}, B)$ instead of $(\widetilde{W}_N, \widetilde{\mathrm{Aut}\mathcal{O}_N})$, we obtain the sheaf of chiral differential operators on G/B – note that it is uniquely defined in this case (see [MSV]).

7. CHIRAL ALGEBRAS

In §4.2, we gave a coordinate independent description of the vertex operation $Y(\cdot, z)$. The formalism of chiral algebras invented by Beilinson and Drinfeld [BD3] (see [G] for a review) is based on a coordinate–free realization of the operator product expansion (see §2.6). In the definition of $\mathcal{Y}_x$ we acted by $Y(A, z)$ on B, placed at the fixed point $x \in X$. In order to define the OPE invariantly, we need to let x move along X as well. This is one as follows. Recall that to each conformal vertex algebra V we have assigned a vector bundle $\mathcal{V}$ on any smooth curve X. Choose a formal coordinate z at $x \in X$, and use it to trivialize $\mathcal{V}|_{D_x}$. Then $\mathcal{V} \boxtimes \mathcal{V}(\infty\Delta)$ is a sheaf on $D_x^2 = \operatorname{Spec} \mathbb{C}[[z, w]]$ associated to the $\mathbb{C}[[z, w]]$–module $V \otimes V[[z, w]][(z - w)^{-1}]$. Let $\Delta_! \mathcal{V}$ be a sheaf on D_x^2 associated to the $\mathbb{C}[[z, w]]$–module $V[[z, w]][(z-w)^{-1}]/V[[z, w]]$. Recall from §4.2 that we have a flat connection on $\mathcal{V}$. Hence $\mathcal{V}$ is a vector bundle with a flat connection, i.e., a $\mathcal{D}$–module, on D_x. The sheaves $\mathcal{V} \boxtimes \mathcal{V}(\infty\Delta)$ and $\Delta_! \mathcal{V}$ are not vector bundles, but they have natural structures of $\mathcal{D}$–modules on D_x^2 (independent of the choice of z). Because of that, we switch to the language of $\mathcal{D}$–modules. The following result is proved in the same way as in Theorem 4.1 (see also [HL]).

Theorem 7.1. — *Define a map $\mathcal{Y}_x^{(2)} : \mathcal{V} \boxtimes \mathcal{V}(\infty\Delta) \to \Delta_! \mathcal{V}$ by the formula*

$$\mathcal{Y}_x^{(2)}(f(z, w)A \boxtimes B) = f(z, w)Y(A, z - w) \cdot B \mod V[[z, w]].$$

Then $\mathcal{Y}_x^{(2)}$ is a homomorphism of $\mathcal{D}$–modules, which is independent of the choice of the coordinate z.

At this point it is convenient to pass to the right $\mathcal{D}$–module on X corresponding to the left $\mathcal{D}$–module $\mathcal{V}$. Recall that if $\mathcal{F}$ is a left $\mathcal{D}$ module on a smooth variety Z, then $\mathcal{F}^r := \mathcal{F} \otimes \Omega_Z$, where Ω_Z denotes the canonical sheaf on Z, is a right $\mathcal{D}$–module on Z. Denote $\Delta : X \hookrightarrow X^2, j : X^2 \backslash \Delta(X) \hookrightarrow X^2$.

Corollary 7.2. — *The vertex algebra structure on V gives rise to a homomorphism of right $\mathcal{D}$–modules on X^2, $\mu : j^* j_* \mathcal{V}^r \boxtimes \mathcal{V}^r \to \Delta_! \mathcal{V}^r$ satisfying the following conditions:*
– (skew-symmetry) $\mu(f(x, y)A \boxtimes B) = -\sigma_{12} \circ \mu(f(y, x)B \boxtimes A)$*;*
– (Jacobi identity) $\mu(\mu(f(x, y, z)A \boxtimes B) \boxtimes C) + \sigma_{123} \circ \mu(\mu(f(y, z, x)B \boxtimes C) \boxtimes A)$

$$+\sigma_{123}^{-1} \circ \mu(\mu(f(z, x, y)C \boxtimes A) \boxtimes B) = 0.$$

– (vacuum) *we are given an embedding* $\Omega \hookrightarrow \mathcal{V}^r$ *compatible with the natural homomorphism* $j_*j^*\Omega \boxtimes \Omega \to \Delta_!\Omega$.

Beilinson and Drinfeld define a *chiral algebra* on a curve X as a right $\mathcal{D}$–module $\mathcal{A}$ on X equipped with a homomorphism $j_*j^*\mathcal{A} \boxtimes \mathcal{A} \to \Delta_!\mathcal{A}$ satisfying the conditions of Corollary 7.2 (see [BD3, G]).

The axioms of chiral algebra readily imply that for any chiral algebra $\mathcal{A}$, its deRham cohomology $H^\bullet_{\mathrm{dR}}(\Sigma, \mathcal{A})$ is a (graded) Lie algebra for any open $\Sigma \subset X$. In particular, for an affine open $\Sigma \subset X$ we obtain the result mentioned in §5.3 that $H^0_{\mathrm{dR}}(\Sigma, \mathcal{V}^r)$ is a Lie algebra.

Define the right $\mathcal{D}$–module $\mathcal{V}_2^r$ on X^2 as the kernel of the homomorphism μ, and let $\mathcal{V}_2 = \mathcal{V}_2^r \otimes \Omega_{X^2}^{-1}$ be the corresponding left $\mathcal{D}$–module. There is a canonical isomorphism $H(X, x, V) \simeq H^2_{\mathrm{dR}}(X^2, \mathcal{V}_2^r)$ (see [G], Prop. 5.1). It is dual to the isomorphism $C(X, x, V) \simeq \Gamma_\nabla(X^2, \mathcal{V}_2^*)$ mentioned in §5.4.

Beilinson and Drinfeld give a beautiful description of chiral algebras as *factorization algebras*. By definition, a factorization algebra is a collection of quasicoherent $\mathcal{O}$–modules $\mathcal{F}_n$ on $X^n, n \geq 1$, satisfying a factorization condition, which essentially means that the fiber of $\mathcal{F}_n$ at $(x_1, \ldots, x_n)$ is isomorphic to $\otimes_{s \in S}(\mathcal{F}_1)_s$, where $S = \{x_1, \ldots, x_n\}$. For example, $j^*\mathcal{F}_2 = j^*(\mathcal{F}_1 \boxtimes \mathcal{F}_1)$ and $\Delta^*\mathcal{F}_2 = \mathcal{F}_1$. Intuitively, such a collection may be viewed as an $\mathcal{O}$–module on the *Ran space* $\mathcal{R}(X)$ of all finite non-empty subsets of X. In addition, it is required that $\mathcal{F}_1$ has a global section (unit) satisfying natural properties. It is proved in [BD3] that under these conditions each $\mathcal{F}_n$ is automatically a left $\mathcal{D}$–module. Moreover, the right $\mathcal{D}_X$–module $\mathcal{F}_1^r = \mathcal{F}_1 \otimes \Omega_X$ acquires a canonical structure of chiral algebra, with μ given by the composition $j^*j_*\mathcal{F}_1^r \boxtimes \mathcal{F}_1^r = j^*j_*\mathcal{F}_2^r \to \Delta_!\Delta^!\mathcal{F}_2^r = \Delta_!\mathcal{F}^r$. In fact, there is an equivalence between the categories of factorization algebras and chiral algebras on X [BD2].

For a chiral algebra $\mathcal{V}^r$, the corresponding sheaf $\mathcal{F}_n$ on X^n is the sheaf $\mathcal{V}_n$ mentioned in §5.4 (its dual is the sheaf of chiral correlation functions). Using these sheaves, Beilinson and Drinfeld define "chiral homology" $H_i^{\mathrm{ch}}(X, \mathcal{V}^r), i \geq 0$, of a chiral algebra $\mathcal{V}^r$. Intuitively, this is (up to a change of cohomological dimension) the deRham cohomology of the factorization algebra corresponding to $\mathcal{V}$, considered as a $\mathcal{D}$–module on the Ran space $\mathcal{R}(X)$. The 0th chiral homology of $\mathcal{V}^r$ is isomorphic to $H^n_{\mathrm{dR}}(X^n, \mathcal{V}_n \otimes \Omega_{X^n})$ for all $n > 1$ and hence to the space of coinvariants $H(X, x, V)$. The full chiral homology functor may therefore be viewed as a derived coinvariants functor.

As an example, we sketch the Beilinson-Drinfeld construction of the factorization algebra corresponding to the Kac-Moody vertex algebra $V_0(\mathfrak{g})$ [BD2] (see also [G]). Let Gr_n be the ind-scheme over X^n whose fiber $Gr_{x_1,\ldots,x_n}$ over $(x_1, \ldots, x_n) \in X$ is the moduli space of pairs $(\mathcal{P}, t)$, where $\mathcal{P}$ is a G–bundle on X, and t is its trivialization on $X \backslash \{x_1, \ldots, x_n\}$ (e.g., the fiber of Gr_1 over any $x \in X$ is isomorphic to the affine Grassmannian $G(\mathcal{K})/G(\mathcal{O})$). Let e be the section of Gr_n corresponding to the trivial G–bundle, and $\mathcal{A}_{x_1\ldots,x_n}$ be the space of delta–functions on $Gr_{x_1,\ldots,x_n}$ supported at e.

These are fibers of a left $\mathcal{D}$–module $\mathcal{A}_n$ on X^n. Claim: $\{\mathcal{A}_n\}$ form a factorization algebra. The factorization property for $\{\mathcal{A}_n\}$ follows from the factorization property of the ind-schemes Gr_n: namely, $Gr_{x_1,\dots,x_n} = \prod_{s\in S} Gr_s$, where $S = \{x_1,\dots,x_n\}$. The chiral algebra corresponding to the factorization algebra $\{\mathcal{A}_n\}$ is $\mathcal{V}_0(\mathfrak{g})^r$. To obtain the factorization algebra corresponding to $\mathcal{V}_k(\mathfrak{g})^r$ with $k \neq 0$, one needs to make a twist by a line bundle on Gr_1.

A spectacular application of the above formalism is the Beilinson-Drinfeld construction of the *chiral Hecke algebra* associated to a simple Lie algebra $\mathfrak{g}$ and a negative integral level $k < -h^\vee$. The corresponding vertex algebra $S_k(\mathfrak{g})$ is a module over $\widehat{\mathfrak{g}} \times {}^L\mathfrak{g}$: $S_k(\mathfrak{g}) = \oplus_{\lambda\in {}^LP^+} M_{\lambda,k} \otimes V_\lambda^*$, where V_λ is a finite-dimensional ${}^L\mathfrak{g}$–module with highest weight λ, and $M_{\lambda,k}$ is the irreducible highest weight $\widehat{\mathfrak{g}}$–module of level k with highest weight $(k+h^\vee)\lambda$. Here ${}^LG$ is the group of adjoint type corresponding to ${}^L\mathfrak{g}$, and ${}^LP^+$ is the set of its dominant weights; we identify the dual $\mathfrak{h}^*$ of the Cartan subalgebra of $\mathfrak{g}$ with $\mathfrak{h} = {}^L\mathfrak{h}^*$ using the normalized invariant inner product on $\mathfrak{g}$. In particular, $S_k(\mathfrak{g})$ is a module over $V_k(\mathfrak{g}) = M_{0,k}$. Note that using results of [KL] one can identify the tensor category of ${}^L\mathfrak{g}$–modules with a subcategory of the category $\mathcal{O}$ of $\widehat{\mathfrak{g}}$–modules of level k, so that V_λ corresponds to $M_{\lambda,k}$. The analogues of $S_k(\mathfrak{g})$ when $\mathfrak{g}$ is replaced by an abelian Lie algebra with a non-degenerate inner product are lattice vertex algebras V_L from §3.4.

As a chiral algebra, the chiral Hecke algebra is the right $\mathcal{D}$–module $\mathcal{S}_k(\mathfrak{g})^r$ on X corresponding to the twist of $S_k(\mathfrak{g})$ by $\widehat{X}$. The corresponding factorization algebra is constructed analogously to the above construction of $V_k(\mathfrak{g})$ using certain irreducible $\mathcal{D}$–modules on Gr_1 corresponding to V_λ; see [G] for the construction in the abelian case (it is also possible to construct a vertex algebra structure on $S_k(\mathfrak{g})$ directly, using results of [KL]). Moreover, for any flat ${}^L\mathfrak{g}$–bundle $\mathcal{E}$ on X, the twist of $\mathcal{S}_k(\mathfrak{g})^r$ by $\mathcal{E}$ with respect to the ${}^L\mathfrak{g}$–action on $S_k(\mathfrak{g})$ is also a chiral algebra on X. Conjecturally, the complex of sheaves on $\mathfrak{M}_G(X)$ obtained by localization of this chiral algebra (whose fibers are its chiral homologies) is closely related to the automorphic $\mathcal{D}$–module that should be attached to $\mathcal{E}$ by the geometric Langlands correspondence.

REFERENCES

[ADKP] E. Arbarello, C. De CONCINI, V. Kac and C. Procesi – *Moduli spaces of curves and representation theory*, Comm. Math. Phys. **117** (1988) 1–36.

[BL] A. Beauville and Y. Laszlo – *Conformal blocks and generalized theta functions*, Comm. Math. Phys. **164** (1994) 385–419.

[BB] A. Beilinson and J. Bernstein – *A Proof of Jantzen Conjectures*, Advances in Soviet Mathematics **16**, Part 1, pp. 1–50, AMS 1993.

[BD1] A. Beilinson and V. Drinfeld – *Affine Kac–Moody algebras and polydifferentials*, Int. Math. Res. Notices **1** (1994) 1–11.

[BD2] A. Beilinson and V. Drinfeld – *Quantization of Hitchin's Integrable System and Hecke eigensheaves*. Preprint.

[BD3] A. Beilinson and V. Drinfeld – *Chiral Algebras*. Preprint.

[BFM] A. Beilinson, B. Feigin and B. Mazur – *Introduction to Algebraic Field Theory on Curves*. Preprint.

[BG] A. Beilinson and V. Ginzburg – *Infinitesimal structure of moduli spaces of G–bundles*, Duke Math. J. Imrn **4** (1992) 63–74.

[BS] A. Beilinson and V. Schechtman – *Determinant bundles and Virasoro algebras*, Comm. Math. Phys. **118** (1988) 651–701.

[BPZ] A. Belavin, A. Polyakov and A. Zamolodchikov – *Infinite conformal symmetries in two–dimensional quantum field theory*, Nucl. Phys. **B241** (1984) 333–380.

[BF] D. Ben-ZVI and E. Frenkel – *Vertex algebras and algebraic curves*, book in preparation.

[B1] R. Borcherds – *Vertex algebras, Kac–Moody algebras and the monster*, Proc. Nat. Acad. Sci. USA **83** (1986) 3068–3071.

[B2] R. Borcherds – *Monstrous moonshine and monstrous Lie superalgebras*, Invent. Math. **109** (1992) 405–444.

[B3] R. Borcherds – *Quantum vertex algebras*, Preprint `math.QA/9903038`.

[Bo1] L. Borisov – *Introduction to the vertex algebra approach to mirror symmetry*, Preprint `math.AG/9912195`.

[BoL] L. Borisov and A. Libgober – *Elliptic genera of toric varieties and applications to mirror symmetry*, Invent. Math. **140** (2000) 453–485.

[dBT] J. de BOER and T. Tjin – *The relation between quantum $\mathcal{W}$–algebras and Lie algebras*, Comm. Math. Phys. **160** (1994) 317-332.

[dFMS] P. di FRANCESCO, P. Mathieu and D. Senechal – *Conformal Field Theory*. Springer–Verlag 1997.

[D1] C. Dong – *Vertex algebras associated with even lattices*, J. Algebra **161** (1993) 245–265.

[D2] C. Dong – *Representations of the moonshine module vertex operator algebra*, Contemp. Math. **175** (1994) 27–36.

[DLM] C. Dong, H. Li and G. Mason – *Twisted representations of vertex operator algebras*, Math. Ann. **310** (1998) 571–600.

[DS] V. Drinfeld and V. Sokolov – *Lie algebras and KdV type equations*, J. Sov. Math. **30** (1985) 1975-2036.

[EK] P. Etingof and D. Kazhdan – *Quantization of Lie bialgebras.* V, Preprint `math.QA/9808121`.

[Fa] G. Faltings – *A proof of the Verlinde formula*, J. Alg. Geom. **3** (1994) 347–374.

[FL] V. Fateev and S. Lukyanov – *The models of two-dimensional conformal quantum field theory with $\mathbb{Z}_n$ symmetry*, Int. J. Mod. Phys. **A3** (1988), 507–520.

[Fe1] B. Feigin – *The semi-infinite cohomology of Kac–Moody and Virasoro Lie algebras*, Russ. Math. Surv. **39**, No. 2 (1984) 155–156.

[FF1] B. Feigin and E. Frenkel – *A family of representations of affine Lie algebras*, Russ. Math. Surv. **43**, No. 5 (1988) 221–222.

[FF2] B. Feigin and E. Frenkel – *Affine Kac–Moody algebras and semi–infinite flag manifolds*, Comm. Math. Phys. **128** (1990) 161–189.

[FF3] B. Feigin and E. Frenkel – *Affine Kac–Moody algebras at the critical level and Gelfand–Dikii algebras*, Int. Jour. Mod. Phys. **A7**, Suppl. 1A (1992) 197–215.

[FF4] B. Feigin and E. Frenkel – *Integrals of Motion and Quantum Groups*, in Lect. Notes in Math. **1620**, pp. 349–418, Springer–Verlag 1996.

[FS] B. Feigin and A. Stoyanovsky – *Realization of a modular functor in the space of differentials, and geometric approximation of the moduli space of G–bundles*, Funct. Anal. Appl. **28** (1994) 257–275.

[FKW] E. Frenkel, V. Kac and M. Wakimoto – *Characters and fusion rules for $\mathcal{W}$–algebras via quantized Drinfeld-Sokolov reduction*, Comm. Math. Phys. **147** (1992), 295–328.

[FKRW] E. Frenkel, V. Kac, A. Radul and W. Wang – *$\mathcal{W}_{1+\infty}$ and $\mathcal{W}_N$ with central charge N*, Comm. Math. Phys. **170** (1995) 337–357.

[FR] E. Frenkel and N. Reshetikhin – *Towards deformed chiral algebras*, Preprint `q-alg/9706023`.

[FK] I. Frenkel and V. Kac – *Basic representations of affine Lie algebras and dual resonance models*, Invent. Math. **62** (1980) 23–66.

[FGZ] I. Frenkel, H. Garland and G. Zuckerman – *Semi-infinite cohomology and string theory*, Proc. Nat. Acad. Sci. U.S. A. **83** (1986) 8442–8446.

[FLM] I. Frenkel, J. Lepowsky and A. Meurman – *Vertex Operator Algebras and the Monster.* Academic Press 1988.

[FHL] I. Frenkel, Y.-Z. Huang and J. Lepowsky – *On axiomatic approaches to vertex operator algebras and modules.* Mem. Amer. Math. Soc. **104** (1993), no. 494.

[FZ] I. Frenkel and Y. Zhu – *Vertex operator algebras associated to representations of affine and Virasoro algebras*, Duke Math. J. **60** (1992) 123–168.

[FrS] D. Friedan and S. Shenker – *The analytic geometry of two-dimensional conformal field theory*, Nucl. Phys. **B281** (1987) 509–545.

[G] D. Gaitsgory – *Notes on 2D Conformal Field Theory and String Theory*, in Quantum fields and strings: a course for mathematicians, Vol. 2, pp. 1017–1089, AMS 1999.

[Ga] K. Gawędzki – *Conformal field theory*, Sém. Bourbaki, Exp. 704, Astérisque **177-178** (1989) 95–126.

[GKF] I.M. Gelfand, D.A. Kazhdan and D.B. Fuchs – *The actions of infinite–dimensional Lie algebras*, Funct. Anal. Appl. **6** (1972) 9–13.

[Gi] V. Ginzburg – *Resolution of diagonals and moduli spaces*, in The moduli space of curves, Progress in Math. **129**, pp. 231–266, Birkhäuser 1995.

[Go] P. Goddard – *Meromorphic conformal field theory*, in Infinite-dimensional Lie algebras and groups, V. Kac (ed.), pp. 556–587, World Scientific 1989.

[GKO] P. Goddard, A. Kent and D. Olive – *Unitary representations of the Virasoro and super-Virasoro algebras*, Comm. Math. Phys. **103** (1986) 105–119.

[GMS] V. Gorbounov, F. Malikov and V. Schechtman – *Gerbes of chiral differential operators*. I, `math.AG/9906116`; II, `math.AG/0003170`.

[Gu] R. Gunning – *Lectures on Riemann Surfaces*. Princeton University Press 1966.

[Hu] Y.-Z. Huang – *Two-dimensional conformal geometry and vertex operator algebras*. Progress in Math. **148**. Birkhäuser 1997.

[HL] Y.-Z. Huang and J. Lepowsky – *On the $\mathcal{D}$–module and formal variable approaches to vertex algebras*, in Topics in geometry, pp. 175–202, Birkhäuser 1996.

[K1] V. Kac – *Infinite-dimensional Lie algebras*, Third Edition. Cambridge University Press 1990.

[K2] V. Kac – *Vertex Algebras for Beginners*, Second Edition. AMS 1998.

[K3] V. Kac – *Formal distribution algebras and conformal algebras*, in Proc. XXIIth ICMP, Brisbane, 1994, pp. 80–96, International Press 1999.

[KL] D. Kazhdan and G. Lusztig – *Tensor structures arising from affine Lie algebras* IV, J. of AMS **7** (1993) 383–453.

[Ko] M. Kontsevich – *The Virasoro algebra and Teichmüller spaces*, Funct. Anal. Appl. **21** (1987), no. 2, 156–157.

[KNR] S. Kumar, M.S. Narasimhan and A. Ramanathan – *Infinite Grassmannians and moduli spaces of G–bundles*, Math. Ann. **300** (1993) 395–423.

[LW] J. Lepowsky and R.L. Wilson – *Construction of the affine Lie algebra* $A_1^{(1)}$, Comm. Math. Phys. **62** (1978) 43–53.

[Li] H. Li – *Local systems of vertex operators, vertex superalgebras and modules*, J. Pure Appl. Alg. **109** (1996) 143–195.

[LZ] B. Lian and G. Zuckerman – *New perspectives on the BRST-algebraic structure of string theory*, Comm. Math. Phys. **154** (1993) 613–646.

[MSV] A. Malikov, V. Schechtman and A. Vaintrob – *Chiral deRham complex*, Comm. Math. Phys. **204** (1999) 439–473.

[SV2] V. Schechtman and A. Varchenko – *Quantum groups and homology of local systems*, in Algebraic Geometry and Analytic Geometry, M. Kashiwara and T. Miwa (eds.), pp. 182–191, Springer–Verlag 1991.

[Se] G. Segal – *The Definition of Conformal Field Theory*, unpublished manuscript.

[So] C. Sorger – *La formule de Verlinde*, Sém. Bourbaki, Exp. 793, Astérisque **237** (1996) 87–114.

[TK] A. Tsuchiya and Y. Kanie – *Vertex operators in conformal field theory on $\mathbb{P}^1$ and monodromy representations of the braid group*, in Adv. Stud. Pure Math **16**, pp. 297–372, Academic Press 1988.

[TUY] A. Tsuchiya, K. Ueno and Y. Yamada – *Conformal field theory on universal family of stable curves with gauge symmetries*, Adv. Stud. Pure Math. **19**, pp. 459–566, Academic Press 1989.

[Wa] M. Wakimoto – *Fock representations of affine Lie algebra $A_1^{(1)}$*, Comm. Math. Phys. **104** (1986) 605–609.

[Wa] W. Wang – *Rationality of Virasoro vertex operator algebras*, Duke Math. J. Imrn **7** (1993) 197–211.

[Wi] E. Witten – *Quantum field theory, Grassmannians and algebraic curves*, Comm. Math. Phys **113** (1988) 529–600.

[Z1] Y. Zhu – *Modular invariance of characters of vertex operator algebras*, J. Ams **9** (1996) 237–302.

[Z2] Y. Zhu – *Global vertex operators on Riemann surfaces*, Comm. Math. Phys. **165** (1994) 485–531.

Edward FRENKEL

Department of Mathematics
University of California
Berkeley, CA 94720, USA
E-mail : `frenkel@math.berkeley.edu`

Séminaire BOURBAKI Juin 2000
52^{e} année, 1999-2000, n^{o} 876, p. 341 à 368

DERIVATIVES OF EISENSTEIN SERIES AND GENERATING FUNCTIONS FOR ARITHMETIC CYCLES

by **Stephen S. KUDLA**

The classical formula of Siegel and Weil identifies the values of Siegel–Eisenstein series at certain critical points as integrals of theta functions. When the critical point is the center of symmetry for the functional equation, the Fourier coefficients of the values of the 'even' Siegel–Eisenstein series thus contain arithmetic information about the representations of quadratic forms. It is natural to ask for an arithmetic interpretation of the *derivative* of the 'odd' series at their center of symmetry.

I would like to report on my work on a family of identities relating the Fourier expansions of the derivatives of certain Siegel–Eisenstein series at their center of symmetry, on one side, and generating functions for the degrees of 0–cycles on moduli schemes for abelian varieties, on the other. On the one hand, such identities can be viewed as generalizations of the Siegel–Weil formula to the case of the derivative. On the other hand, the identities imply that the generating functions in question, which are given as power series in q with coefficients arising from arithmetical algebraic geometry, are in fact the q-expansions of modular forms. This work grows out of results obtained in collaboration with Steve Rallis [18], [19], [20] and with John Millson [15], [16], [17]. More recent progress has been made in collaboration with Michael Rapoport [21], [22], [23] and Tonghai Yang [24], [25]. At present, the identities have been fully established only in certain special cases as explained below. Nonetheless, these examples, together with partial results in higher dimensions, suggest the outline of a more extensive theory.

An additional origin of the investigation described here was the study of the triple product L-function at the center of the critical strip, in collaboration with Michael Harris [9] and with Benedict Gross [7]. In particular, a Siegel–Eisenstein series is a key ingredient in the Rankin-Selberg integral representation of this L–function. Thus the occurrence of arithmetic geometric quantities in the central derivatives of the Eisenstein series should reflect their appearance in the central derivative of the L–function, and hence should provide a relation to the Gross–Zagier formula [8].

Section 1 contains two examples, one recalling the work of Hirzebruch and Zagier on the modular generating functions for curves on a Hilbert–Blumenthal surface and the second illustrating a generating function in the simplest arithmetic case, involving the derivative of a classical Eisenstein series of weight 1. In section 2, the incoherent Siegel–Eisenstein series, which should be related to arithmetic generating series, are defined in general. Section 3 reviews the results of [15],[16],[17] on generating functions for cycles on locally symmetric spaces. These results suggest what one should hope for in the arithmetic case. In section 4, the generating function for 0–cycles on an arithmetic surface attached to a Shimura curve is defined, and Conjecture 4.7 relates it to the central derivative of an incoherent Eisenstein series of weight $\frac{3}{2}$ and genus 2. The comparison of the nonsingular Fourier coefficients of the two objects is discussed in sections 5 and 6. Section 7 contains a brief survey of results in higher dimensional cases as well as a second look at the simplest example of section 1. Some speculations about further developments are made in section 8.

1. Two examples
2. Central derivatives of Siegel–Eisenstein series
3. Generating functions in the geometric case
4. Generating functions for arithmetic 0–cycles; the case of Shimura curves
5. Non-singular Fourier coefficients
6. Green's functions and Whittaker functions
7. Further results
8. Final remarks

I would like to warmly thank J.-B. Bost, G. Henniart and M. Rapoport for detailed comments and advice on the original draft of this report.

Notation

$\mathbb{Q}$, $\mathbb{A}$, $\mathbb{A}_f$, $\mathbb{A}^\times$ denote the rational numbers, the adèles, the finite adèles, and the idèles of $\mathbb{Q}$ respectively.

$\hat{\mathbb{Z}} = \varprojlim_n \mathbb{Z}/n\mathbb{Z}$, and, for any $\mathbb{Z}$–module M, $\hat{M} = M \otimes_{\mathbb{Z}} \hat{\mathbb{Z}}$.

$\mathbb{F}_q$ denotes the finite field with q elements, and $\bar{\mathbb{F}}_q$ denotes an algebraic closure of it.

$(\ ,\ )_p$ (resp. $(\ ,\ )_{\mathbb{A}}$) denotes the quadratic Hilbert symbol for $\mathbb{Q}_p$ (resp. $\mathbb{A}$).

ψ is a fixed nontrivial character of $\mathbb{A}/\mathbb{Q}$.

$\mathfrak{H}_g = \{\tau = u + iv \in \mathrm{Sym}_g(\mathbb{C}) \mid v > 0\}$ is the Siegel space of genus g;

$\mathfrak{H} = \mathfrak{H}_1$ is the upper half–plane.

$e(x) = e^{2\pi i x}$

$\mathrm{Sym}_n(R) = \{x \in M_n(R) \mid {}^t x = x\}$, the space of $n \times n$ symmetric matrices.

$$\mathrm{diag}(a_1, \ldots, a_n) = \begin{pmatrix} a_1 & & \\ & \ddots & \\ & & a_n \end{pmatrix} \in \mathrm{Sym}_n(R).$$

1. TWO EXAMPLES

To fix ideas, it may be useful to consider two examples in classical language. The first of these illustrates the construction of generating functions for curves on a complex surface. More precisely, it gives a compact quotient version of the Hirzebruch–Zagier generating function for curves on a Hilbert–Blumenthal surface and a similar generating function for 0–cycles on such a surface. The second example illustrates the arithmetic case where cycles on moduli spaces for abelian varieties are defined by imposing extra endomorphisms. The example involves CM elliptic curves and the generating function is identified as the central derivative of an Eisenstein series of weight 1.

1.1. The case of a complex surface.

The results of this section are special cases of joint work with John Millson [15], [16], [17]. Let V, $(\ ,\ )$ be a 4–dimensional rational vector space with a symmetric bilinear form of signature $(2,2)$. Let $Q(x) = \frac{1}{2}(x,x)$ be the associated quadratic form. Fix a lattice L in V on which the form is $\mathbb{Z}$–valued and let

$$\Gamma \subset \{\gamma \in SO(V)(\mathbb{Q})^{+} \mid \gamma L = L\}$$

be a subgroup of finite index, where $SO(V)(\mathbb{R})^{+}$ is the identity component of the Lie group $SO(V)(\mathbb{R})$ and $SO(V)(\mathbb{Q})^{+} = SO(V)(\mathbb{Q}) \cap SO(V)(\mathbb{R})^{+}$. The space D of negative 2–planes in $V(\mathbb{R})$ is isomorphic to the product $\mathfrak{H}\times\mathfrak{H}$ of two copies of the upper half–plane $\mathfrak{H}$, and the quotient $S = \Gamma\backslash D$ is (the complex points of) a quasi–projective variety. Now assume that the space V is anisotropic so that S is projective. This assumption eliminates complications coming from the compactification of the cusps, which are a significant issue in the Hilbert–Blumenthal case considered by Hirzebruch and Zagier.

For a vector $x \in V(\mathbb{Q})$ with $Q(x) > 0$, let

$$D_x = \{z \in D \mid z \perp x\}$$

be the set of negative 2–planes orthogonal to x, so that $D_x \subset D$ is isomorphic to an embedded upper half–plane $\mathfrak{H} \subset \mathfrak{H} \times \mathfrak{H}$. The image $Z(x,\Gamma)$ of D_x in the quotient S is a compact curve, the image of the quotient $\Gamma_x\backslash D_x$, where Γ_x is the stabilizer of x. Note that the curve $Z(x,\Gamma)$ depends only on the Γ orbit of x. Associated to a positive integer t, there is a finite sum of such curves

$$Z(t,L) = \sum_{\substack{x\in L\\ Q(x)=t\\ \bmod \Gamma}} Z(x,\Gamma), \tag{1.1}$$

parametrized by the Γ–orbits in the set of lattice vectors of length t. This is the analogue of the Hirzebruch–Zagier curve T_N on a Hilbert–Blumenthal surface [11]. Let

$[Z(t,L)] \in H^2(S,\mathbb{Q})$ be the cohomology class of $Z(t,L)$, and let $[Z(0,L)] \in H^2(S,\mathbb{Q})$ be the cohomology class of the invariant form[(1)]

$$\omega = -\frac{1}{4\pi}\big(\operatorname{Im}(z_1)^{-2}\,dz_1 \wedge d\bar z_1 + \operatorname{Im}(z_2)^{-2}\,dz_2 \wedge d\bar z_2\big) \tag{1.2}$$

on $\mathfrak{H}\times\mathfrak{H}$. As in [11], one can form a generating function

$$\phi_1(\tau,L) = [Z(0,L)] + \sum_{t\in\mathbb{Z}_{>0}} [Z(t,L)]\,q^t, \tag{1.3}$$

where $q = e(\tau)$, for an auxiliary variable $\tau = u+iv \in \mathfrak{H}$. The analogue of the result of Hirzebruch–Zagier [11] is a special case of [15], [16], [17], see Theorem 3.1 below.

Theorem 1.1. — *The generating function $\phi_1(\tau,L)$ is an elliptic modular form of weight 2, valued in $H^2(S,\mathbb{C})$.*

Taking the intersection product with the cohomology class of an arbitrary curve C on S one obtains:

Corollary 1.2. — *For a curve C on S, the generating function for intersection numbers*

$$\phi_1(\tau,L)\cdot[C] = \operatorname{vol}(C) + \sum_{t>0}[Z(t,L)]\cdot[C]\,q^t$$

is a modular form of weight 2. Here $\operatorname{vol}(C) = \int_C \omega$.

One can also define a generating function for 0–cycles on S as follows. For a pair of vectors $x = [x_1,x_2] \in V(\mathbb{Q})^2$ with matrix of inner products $Q(x) = \frac{1}{2}((x_i,x_j))_{i,j} \in \operatorname{Sym}_2(\mathbb{Q})$, there is an associated cycle $D_x = D_{x_1}\cap D_{x_2} \subset D$. If $Q(x)$ is nonsingular, then D_x is a point when $Q(x)$ is positive definite and is empty otherwise. If $Q(x)$ has rank 1, then $D_x = D_{x_1} = D_{x_2}$ is a curve when $Q(x)\geq 0$ (the components x_1 and x_2 are colinear since V is anisotropic) and is empty otherwise. Finally, if $Q(x)=0$, then $x=0$ and $D_x = D$. Let $Z(x,\Gamma)$ be the image of D_x in the quotient S; again, this depends only on the Γ–orbit of x. For $T\in\operatorname{Sym}_2(\mathbb{Z})$, let

$$Z(T,L) = \sum_{\substack{x\in L^2\\ Q(x)=T\\ \bmod \Gamma}} Z(x,\Gamma). \tag{1.4}$$

If $T>0$ is positive definite, $Z(T,L)$ is either empty or a finite sum of points on S. If $T\geq 0$, has rank 1, then $Z(T,L)$ is either empty or a finite sum of curves on S, and if $T=0$, then $Z(0,L)=S$. Let $[Z(T,L)] \in H^{2r(T)}(S,\mathbb{Q})$ be the cohomology class of $Z(T,L)$, where $r(T) = \operatorname{rank}(T)$. The generating function in this case is

$$\phi_2(\tau,L) = \sum_{T\in\operatorname{Sym}_2(Z)_{\geq 0}} [Z(T,L)]\cup[\omega]^{2-r(T)}\,q^T, \tag{1.5}$$

(1) This is twice the form used in [11], p.104, since the $Z(t,L)$'s of (1.1) are twice the corresponding cycles in [11].

where $\tau \in \mathfrak{H}_2$, the Siegel space of genus 2, and $q^T = e(\mathrm{tr}(T\tau))$. Note that the terms for singular T's are obtained by shifting by suitable powers of $[\omega]$. The coefficients of this generating function lie in $H^4(S,\mathbb{C})$, and [15], [16], [17] yield the following result.

THEOREM 1.3. — *The generating function $\phi_2(\tau, L)$ is a Siegel modular form of weight 2 and genus 2 valued in $H^4(S,\mathbb{C})$.*

Applying the degree map $H^4(S,\mathbb{C}) \to \mathbb{C}$, one obtains a scalar valued Siegel modular form.

COROLLARY 1.4. — *The generating function*

$$\deg(\phi_2(\tau, L)) = \mathrm{vol}(S) + \sum_{\substack{T \in \mathrm{Sym}_2(\mathbb{Z})_{\geq 0} \\ r(T)=1}} \mathrm{vol}(Z(T,L))\, q^T + \sum_{T \in \mathrm{Sym}_2(\mathbb{Z})_{>0}} \deg(Z(T,L))\, q^T$$

is a Siegel modular form of weight 2 and genus 2.

In particular, the positive definite Fourier coefficients of $\deg(\phi_2(\tau, L))$ are the degrees of the 0–cycles $Z(T,L)$ on the surface S. The volumes of curves on S are taken with respect to the restriction of the invariant (1,1)–form ω of (1.2) and

$$\mathrm{vol}(S) = \int_S \omega^2.$$

Theorems 1.1 and 1.3 are proved by constructing a theta function $\theta_1(\tau, L)$ for $\tau \in \mathfrak{H}$, resp. $\theta_2(\tau, L)$ for $\tau \in \mathfrak{H}_2$, valued in *closed* (1,1)–forms, resp. closed $(2,2)$–forms on S. The generating function is the cohomology class of this theta function, i.e., $\phi_i(\tau, L) = [\theta_i(\tau, L)]$ for $i = 1, 2$, and hence is modular. In addition, the generating function of Corollary 1.2, resp. Corollary 1.4, is obtained as the integral of $\theta_1(\tau, L)$ over the curve C, resp. $\theta_2(\tau, L)$ over S. For suitable Γ's, this last integral over S is a constant multiple of the group theoretic integral of the theta function which occurs in the Siegel–Weil formula, and hence coincides with a special value of a Siegel–Eisenstein series of genus 2 at the point $s = \frac{1}{2}$, [13] and Proposition 3.2 below.

COROLLARY 1.5. — *There is a nonzero constant c such that*

$$\deg(\phi_2(\tau, L)) = c \cdot E(\tau, \frac{1}{2}, L)$$

for a suitable Siegel–Eisenstein series $E(\tau, s, L)$ of genus 2 and weight 2.

In the case in which S is a product of modular curves, such a geometric interpretation of the Fourier coefficients of a Siegel–Eisenstein series was observed by Gross and Keating [6].

1.2. Interlude.

More general results of this type, [15], [16], [17], and [13], concerning generating functions for cycles of codimension n on Shimura varieties X defined by rational quadratic forms of signature $(m-2,2)$ are discussed in section 3 below. Note that the complex dimension of X is $m-2$. The main aim of this report is to explain the first steps in establishing a similar theory in the arithmetic case. Roughly speaking, this means the following. First, one wants to consider cycles of codimension n on integral models $\mathfrak{X}$ of the Shimura varieties X and their classes in the arithmetic Chow groups $\widehat{CH}^r(\mathfrak{X})$, [4]. For $2 \leq m \leq 4$, integral models can be obtained as moduli spaces of suitable abelian varieties, and cycles can be defined as the loci where the abelian varieties in question are equipped with additional endomorphisms of a certain type. The theta functions valued in the deRham complex are not available in this context, and so it is not clear at present how to define generating functions for cycles of arbitrary codimension. If one considers 0–cycles, however, one may apply the arithmetic degree map $\widehat{\deg} : \widehat{CH}^{m-1}(\mathfrak{X}) \to \mathbb{R}$. One may then look for an analogue of Corollary 1.5 and Proposition 3.2, where the Siegel–Eisenstein series will now have genus $n = m-1$ and the critical point will be $s_0 = \frac{m}{2} - \frac{n+1}{2} = 0$, i.e., the central point on the real axis for the functional equation of the Eisenstein series. Moreover, it turns out that the 'correct' Eisenstein series of genus n and weight $\frac{n+1}{2}$ will have a zero at this point, so that one should look at its first derivative $E'(\tau, 0, L)$. The case of the arithmetic surfaces attached to Shimura curves is discussed at length below in sections 4–6, and the analogue of Corollary 1.5 is given in Conjecture 4.7. The simplest example, however, occurs for $m = 2$. This case involves only classical objects, e.g., elliptic curves with complex multiplication and Eisenstein series of weight 1 for SL_2.

1.3. Derivatives of Eisenstein series of weight 1.

This section describes the simplest case in which the derivative at $s = 0$ of an Eisenstein series can be identified with a generating function for the arithmetic degrees of 0–cycles on a moduli scheme. These results are joint work with Michael Rapoport and Tonghai Yang [24]. Fix a prime $d \equiv 3 \mod 4$ with $d > 3$, and let $\boldsymbol{k} = \mathbb{Q}(\sqrt{-d})$ be the corresponding imaginary quadratic field with ring of integers $\mathcal{O}_{\boldsymbol{k}}$ and associated Dirichlet character χ_d. Let

$$\Lambda(s, \chi_d) = \pi^{-\frac{s+1}{2}}\, \Gamma\left(\frac{s+1}{2}\right) L(s, \chi_d)$$

where $L(s, \chi_d)$ is the Dirichlet L–series of χ_d. For a nonzero integer $n \in \mathbb{Z}$, let $\rho(n)$ be the number of ideals in $\mathcal{O}_{\boldsymbol{k}}$ of norm n. For example, note that, for a prime p, $\rho(p) = 0$ if p is inert in $\boldsymbol{k}$, $\rho(p) = 2$ is p is split in $\boldsymbol{k}$, and $\rho(d) = 1$.

There are two normalized Eisenstein series of weight 1 for $\Gamma = \mathrm{SL}_2(\mathbb{Z})$ attached to $\boldsymbol{k}$. For $\tau = u + iv \in \mathfrak{H}$, and $s \in \mathbb{C}$ with $\mathrm{Re}(s) > 1$, let

$$E^*_{\pm}(\tau, s) = v^{\frac{s}{2}}\, d^{\frac{s+1}{2}}\, \Lambda(s+1, \chi_d) \sum_{\gamma \in \Gamma_\infty \backslash \Gamma} (c\tau + d)^{-1} |c\tau + d|^{-s}\, \Phi_{\pm}(\gamma),$$

where, for $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma$,

$$\Phi_{\pm}(\gamma) = \begin{cases} \chi_d(a) & \text{if } c \equiv 0 \mod (d) \\ \pm i\, d^{-\frac{1}{2}}\, \chi_d(c) & \text{if } c \text{ is prime to } d. \end{cases}$$

The entire analytic continuations of these series in s satisfy the functional equations

$$E^*_{\pm}(\tau, -s) = \pm E^*_{\pm}(\tau, s).$$

A general construction of series of this sort is described in section 2 below. A case of the Siegel–Weil formula due to Hecke describes the value at $s = 0$ of the even series:

$$E^*_{+}(\tau, 0) = 2h_{\boldsymbol{k}} + 4 \sum_{t=1}^{\infty} \rho(t)\, q^t = 2 \sum_{\mathfrak{a}} \vartheta(\tau, \mathfrak{a}),$$

where $h_{\boldsymbol{k}}$ is the class number of $\boldsymbol{k}$, the ideal $\mathfrak{a}$ runs over representatives of the ideal classes of $\boldsymbol{k}$, and $\theta(\tau, \mathfrak{a})$ is the binary theta series attached to $\mathfrak{a}$. For the odd series, $E^*_{-}(\tau, 0) = 0$, and the function of interest is the (negative of the) leading term

$$\phi(\tau, d) = -\frac{\partial}{\partial s}\big\{E^*_{-}(\tau, s)\big\}\big|_{s=0}.$$

Theorem 1.6. — *The modular form $\phi(\tau, d)$ of weight 1 has Fourier expansion*

$$\phi(\tau, d) = a_0(v) + \sum_{t<0} a_t(v)\, q^t + \sum_{t=1}^{\infty} a_t\, q^t,$$

where, for $t > 0$,

$$a_t = 2\log(d)\,(\mathrm{ord}_d(t) + 1)\,\rho(t) + 2 \sum_{p \neq d} \log(p)\,(\mathrm{ord}_p(t) + 1)\,\rho(t/p),$$

where the sum runs over primes p inert in $\boldsymbol{k}$,

$$a_0 = -h_{\boldsymbol{k}} \big(\log(d) + \log(v) + 2\frac{\Lambda'(1, \chi_d)}{\Lambda(1, \chi_d)}\big),$$

and, for $t < 0$,

$$a_t(v) = -2\mathrm{Ei}(-4\pi |t| v)\, \rho(-t).$$

Here

$$-\mathrm{Ei}(-x) = \int_1^{\infty} u^{-1}\, e^{-ux}\, du$$

is the exponential integral.

The idea now is to give an interpretation of these coefficients as the degrees in the sense of arithmetic geometry of certain 0–cycles on the coarse moduli scheme $\mathcal{M}$ for elliptic curves (E,ι) with complex multiplication $\iota : \mathcal{O}_{\boldsymbol{k}} \hookrightarrow \mathrm{End}(E)$ by $\mathcal{O}_{\boldsymbol{k}}$. This scheme over $\mathcal{O}_{\boldsymbol{k}}$ can be identified with $\mathrm{Spec}(\mathcal{O}_H)$, where H is the Hilbert class field of $\boldsymbol{k}$. For such a curve (E,ι), the space of special endomorphisms is the $\mathbb{Z}$-module

$$V(E,\iota) = \{x \in \mathrm{End}(E) \mid x\,\iota(a) = \iota(\bar{a})\,x \text{ for all } a \in \mathcal{O}_{\boldsymbol{k}}\ \}. \tag{1.6}$$

This space has a $\mathbb{Z}$-valued quadratic form Q defined by $x^2 = -Q(x)\cdot id_E$. For $t \in \mathbb{Z}$, let $Z(t)$ be the coarse moduli scheme whose points over an algebraically closed field correspond to triples (E,ι,x) where $x \in V(E,\iota)$ with $Q(x)=t$. The scheme $Z(t) \to \mathcal{M}$ is the locus of (E,ι)'s with an extra multiplication, anticommuting with the action of $\mathcal{O}_{\boldsymbol{k}}$. Such extra endomorphisms can only exist for a supersingular curve E in characteristic p for a prime p which is not split in $\boldsymbol{k}$. Then $Z(t) = \mathrm{Spec}(R(t))$ where $R(t)$ is an Artin algebra in which p is nilpotent. Let

$$\widehat{\deg}(Z(t)) = \log|R(t)|.$$

The second main result of [24] is a calculation of this degree; this calculation depends in an essential way on the results of Gross [5].

Theorem 1.7. — *For $t>0$,*

$$\widehat{\deg}(Z(t)) = a_t,$$

and hence

$$\phi(\tau,d) = \sum_{t>0} \widehat{\deg}(Z(t))\,q^t + a_0(v) + \sum_{t<0} a_t(v)\,q^t.$$

A sort of geometric interpretation of the remaining terms will be discussed in section 7 below, cf. also [24].

2. CENTRAL DERIVATIVES OF SIEGEL–EISENSTEIN SERIES

A general construction of ‘even’ and ‘odd’ Siegel–Eisenstein series is best described in representation theoretic language, and is connected with the Siegel–Weil formula at the central critical point. These series, which will be called coherent and incoherent series respectively, for reasons explained below, will have integral or half–integral weight depending on the parity of the dimension of the relevant quadratic spaces. Hence it is necessary to work on the metaplectic group.

Let $G = \mathrm{Sp}_{2n}$ be the symplectic group of rank n over $\mathbb{Q}$ and let $P = MN$ be the maximal parabolic with Levi factor $M \simeq \mathrm{GL}_n$ and unipotent radical $N \simeq \mathrm{Sym}_n$. Let

$$G_{\mathbb{A}} = \begin{cases} \mathrm{M}p_{2n}(\mathbb{A}) & \text{if } n \text{ is even} \\ \mathrm{Sp}_{2n}(\mathbb{A}) & \text{if } n \text{ is odd,} \end{cases}$$

where $\mathrm{M}p_{2n}(\mathbb{A})$ is the twofold metaplectic cover of $\mathrm{Sp}_{2n}(\mathbb{A})$, and let $P_{\mathbb{A}}$ and $M_{\mathbb{A}}$ be the subgroups of $G_{\mathbb{A}}$ corresponding to P and M. Let $G_{\mathbb{Q}}$ be $\mathrm{Sp}_{2n}(\mathbb{Q})$ for n odd resp. the image of this group in $\mathrm{M}p_{2n}(\mathbb{A})$ under the canonical splitting, if n is even. For each place $p \leq \infty$ of $\mathbb{Q}$, there are groups G_p, P_p and M_p, defined analogously.

A quadratic character χ of $\mathbb{A}^\times/\mathbb{Q}^\times$ determines a character $\chi = \chi^\psi$ of $M_{\mathbb{A}}$, trivial on $M_{\mathbb{Q}} = M_{\mathbb{A}} \cap G_{\mathbb{Q}}$, and for $s \in \mathbb{C}$, one has the degenerate principal series representation

$$I(s,\chi) = \mathrm{Ind}_{P_{\mathbb{A}}}^{G_{\mathbb{A}}}(\chi|\ |^s)$$

of $G_{\mathbb{A}}$. (The character χ^ψ depends on the fixed choice of the nontrivial additive character ψ of $\mathbb{A}/\mathbb{Q}$ in the metaplectic case.)

For $\Phi(s) \in I(s,\chi)$, the Siegel-Eisenstein series is defined for $\mathrm{Re}(s) > \frac{n+1}{2}$ by

$$E(g,s,\Phi) = \sum_{\gamma \in P_{\mathbb{Q}} \backslash G_{\mathbb{Q}}} \Phi(\gamma g, s).$$

From the standard theory of Eisenstein series one knows that this function has a meromorphic analytic continuation to the whole s–plane and satisfies a functional equation relating s and $-s$. In addition, it has no poles on the line $\mathrm{Re}(s) = 0$ (unitary axis). In particular, there is an intertwining map

$$E(0) : I(0,\chi) \longrightarrow \mathcal{A}(G), \qquad \Phi(0) \mapsto E(0,\Phi) \tag{2.1}$$

from the degenerate principal series at $s = 0$ to the space $\mathcal{A}(G)$ of automorphic forms on $G_{\mathbb{A}}$.

The image and kernel of this map can be described in terms of representations associated to quadratic forms as follows.

A rational vector space V of dimension m with a nondegenerate quadratic form Q determines a quadratic character χ_V of $\mathbb{A}^\times/\mathbb{Q}^\times$ by

$$\chi_V(x) = (x, (-1)^{\frac{m(m-1)}{2}} \det(Q))_{\mathbb{A}},$$

where $(\ ,\ )_{\mathbb{A}}$ is the global quadratic Hilbert symbol. For such a space (V,Q), there is a Weil representation ω_V of $G_{\mathbb{A}}$ on the Schwartz space $S(V(\mathbb{A})^n)$, determined by ψ. This gives rise to a $G_{\mathbb{A}}$– intertwining map

$$\lambda_V : S(V(\mathbb{A})^n) \longrightarrow I(s_0,\chi_V), \qquad \lambda_V(\varphi)(g) = \big(\omega_V(g)\varphi\big)(0),$$

where $s_0 = \frac{m}{2} - \frac{n+1}{2}$. Specializing to the case $\chi_V = \chi$ and $m = n+1$, one obtains an irreducible constituent $\Pi(V) = \lambda_V\big(S(V(\mathbb{A})^n)\big)$ of $I(0,\chi)$.

Similarly, for each place $p \leq \infty$, there is an analogous local construction which yields an irreducible constituent $\Pi_p(V_p)$ of the local induced representation $I_p(0,\chi_{V_p})$ of G_p associated to each quadratic space V_p over $\mathbb{Q}_p$ of dimension $n+1$ and character χ_{V_p}. Then, for a global space V, one has

$$\Pi(V) \simeq \otimes'_{p\leq\infty} \Pi_p(V_p) \subset I(0,\chi) = \otimes'_{p\leq\infty} I_p(0,\chi_p)$$

where $V_p = V \otimes_{\mathbb{Q}} \mathbb{Q}_p$ and the primes indicate the restricted tensor products. For a finite prime p, there are precisely two possible quadratic spaces V_p^+ and V_p^- over $\mathbb{Q}_p$, for a fixed n and character χ_p. They are distinguished by their Hasse invariants $\epsilon_p(V_p^{\pm}) = \pm 1$, and, in fact, [19],

$$I_p(0, \chi_p) = \Pi_p(V_p^+) \oplus \Pi_p(V_p^-).$$

For $p = \infty$, the quadratic spaces of dimension $n+1$ and character χ_∞ are determined by their signature, and fall into two groups according to their Hasse invariant. The local induced representation $I_\infty(0, \chi_\infty)$ is the direct sum of the corresponding $\Pi_\infty(V_\infty)$'s, [18]. For example, for $n = 4$ the quadratic spaces over $\mathbb{R}$ of signatures $(5, 0)$, and $(1, 4)$ have Hasse invariant $+1$, while that of signature $(3, 2)$ has Hasse invariant -1, and

$$I_\infty(0, \chi_\infty) = \Pi(5, 0) \oplus \Pi(3, 2) \oplus \Pi(1, 4),$$

in the obvious notation. Here $\chi_\infty = 1$.

If a collection of local quadratic spaces $\mathcal{C} = \{V_p\}$ is the set of localizations of a global space V, then the product formula for the Hasse invariants asserts that

$$\epsilon(\mathcal{C}) := \prod_{p \leq \infty} \epsilon_p(V_p) = 1.$$

Such a collection and the Eisenstein series associated to it will be called *coherent.* On the other hand, the collection $\mathcal{C}$ of local quadratic spaces obtained by choosing one prime p_0 (e.g., $p_0 = \infty$) and switching the space V_{p_0} to a space V'_{p_0} with the opposite Hasse invariant has

$$\epsilon(\mathcal{C}) := \epsilon_{p_0}(V_{p_0}) \prod_{\substack{p \leq \infty \\ p \neq p_0}} \epsilon_p(V_p) = -1,$$

so that such a collection cannot be the set of localizations of any global quadratic space. In this case, the collection $\mathcal{C}$ and the Eisenstein series associated to it will be called *incoherent.* The irreducible admissible representation

$$\Pi(\mathcal{C}) = \Pi_{p_0}(V'_{p_0}) \otimes \left(\otimes_{\substack{p \leq \infty \\ p \neq p_0}} \Pi_p(V_p) \right)$$

of $G_{\mathbb{A}}$ is also a constituent of $I(0, \chi)$. Then there is a direct sum decomposition

$$I(0, \chi) = \left(\bigoplus_V \Pi(V) \right) \oplus \left(\bigoplus_{\mathcal{C}} \Pi(\mathcal{C}) \right)$$

where V runs over all global quadratic spaces of dimension $n+1$ and character χ, and $\mathcal{C}$ runs over all incoherent collections, wh ere $\epsilon(\mathcal{C}) = -1$, as just described. One then obtains a description of the kernel and image of the map $E(0)$ in terms of the $\Pi(V)$'s and $\Pi(\mathcal{C})$'s, [14], [20], [9].

THEOREM 2.1. — (i)

$$\ker(E(0)) = \bigoplus_{\epsilon(\mathcal{C})=-1} \Pi(\mathcal{C}).$$

(ii) *Each automorphic representation* $\Pi(V)$ *in the image*

$$\mathrm{Im}(E(0)) \simeq \bigoplus_{V} \Pi(V)$$

coincides with space of (regularized) theta integrals

$$I(g,\varphi) = \int_{O(V)(\mathbb{Q})\backslash O(V)(\mathbb{A})} \theta(g,h;\varphi)\,dh, \tag{2.2}$$

attached to the global quadratic space V. *Here, for* $\varphi \in S(V(\mathbb{A})^n)$, $g \in G_{\mathbb{A}}$ *and* $h \in O(V)(\mathbb{A})$,

$$\theta(g,h;\varphi) = \sum_{x\in V(\mathbb{Q})} (\omega(g)\varphi)(h^{-1}x).$$

The integral $I(g,\varphi)$ must be defined by a regularization procedure [20] whenever V is isotropic.

Part (ii) of Theorem 1.2 is essentially the Siegel–Weil formula in the present context, [20], [33]. Note that the theta functions involve global arithmetic, e.g., the number of solutions of diophantine equations of the form $Q(x) = T$ for $T \in Sym_n(\mathbb{Z})$ and $x \in L^n$ for a lattice $L \subset V(\mathbb{Q})$, whereas the Eisenstein series is constructed from local data.

Problem. — What is the arithmetic content of the first derivative $E'(g,0,\Phi)$ when $\Phi \in \Pi(\mathcal{C})$ with $\epsilon(\mathcal{C}) = -1$?

Remark 2.2. — The results to be discussed in the remainder of the talk suggest an answer to this question, at least for the following particular case:

Let V be a rational quadratic space, as above, with signature $(n-1,2)$, and let $\mathcal{C}$ be the collection of local quadratic spaces obtained from $\{V_p\}$ by replacing V_∞ by the space V'_∞ of signature $(n+1,0)$. Let $\varphi'_\infty \in S((V'_\infty)^n)$ be the Gaussian $\varphi'_\infty(x) = \exp(-\pi Q'(x))$, and let $\Phi_\infty^{(n+1)/2}(s) \in I_\infty(s,\chi_\infty)$ be the corresponding section; it is the unique eigenvector for K_∞ of weight $\frac{n+1}{2}$. For any $\varphi \in S(V(\mathbb{A}_f)^n)$, with corresponding section $\Phi_f(s) \in I_f(s,\chi_f)$,

$$\Phi(0) = \Phi_\infty^{\frac{n+1}{2}}(0) \otimes \Phi_f(0) \in \Pi(\mathcal{C}).$$

Then the central derivative $E'(g,0,\Phi)$ should be related to a generating function for the degrees of 0–cycles on an integral model of the Shimura variety associated to the group $\mathrm{GSpin}(V)$.

Remark 2.3. — For comparison with the generating functions considered below, it is convenient to write the Eisenstein series in a more classical language. For $\tau = u+iv \in \mathfrak{H}_n$, the Siegel space of genus n, and for $\varphi \in S(V(\mathbb{A}_f)^n)$, let

$$(2.3)\qquad E(\tau, s, \varphi) := \det(v)^{-\frac{1}{2}(s+\frac{n+1}{2})} E(g_\tau, s, \Phi)$$
$$= \sum_{\gamma \in \Gamma_\infty \backslash \Gamma} \det(c\tau + d)^{-\frac{n+1}{2}} |\det(c\tau+d)|^{-s} \, (\omega_f(\gamma)\varphi)(0),$$

where $\Gamma = \mathrm{Sp}_{2n}(\mathbb{Z})$, $\Gamma_\infty = \Gamma \cap P(\mathbb{Q})$, ω_f denotes the action of $G_{\mathbb{A}_f}$ on $S(V(\mathbb{A}_f)^n)$ via the Weil representation, and $g_\tau = \begin{pmatrix} v^{\frac{1}{2}} & uv^{-\frac{1}{2}} \\ & v^{-\frac{1}{2}} \end{pmatrix}$.

3. GENERATING FUNCTIONS IN THE GEOMETRIC CASE

This section describes the results of [15], [16], [17], and [13] on generating functions for the cohomology classes of special cycles in the case of $O(n-1, 2)$, special cases of which were described in section 1.1. These results suggest what one might hope to prove in the arithmetic case.

For a rational quadratic space V of signature $(n-1, 2)$, let $H = \mathrm{GSpin}(V)$ and let D be the space of oriented negative 2–planes in $V(\mathbb{R})$. The space D is isomorphic to two copies of a bounded domain of type IV in $\mathbb{C}^{n-1}$, [10], [28], and the group $H(\mathbb{R})$ acts on it by holomorphic automorphisms. For a compact open subgroup $K \subset H(\mathbb{A}_f)$, the orbit space

$$X_K(\mathbb{C}) = H(\mathbb{Q}) \backslash D \times H(\mathbb{A}_f)/K$$

is the set of complex points of a quasi–projective variety X_K defined over $\mathbb{Q}$, the Shimura variety attached to H, D and K. The variety X_K is in fact projective if V is anisotropic and smooth if K is sufficiently small.

Fix $r \in \mathbb{Z}$ with $1 \le r \le n-1$. For $x \in V(\mathbb{R})^r$, let

$$D_x = \{ z \in D \mid x \perp z \}$$

be the set of z's which are orthogonal to all components of x. If the matrix $T = Q(x) = \frac{1}{2}((x_i, x_j))$ is positive definite, i.e., if the components of x span a positive r–plane, then D_x has complex codimension r in D. If, in addition, $x \in V(\mathbb{Q})^r$, then $x^\perp$ is a rational quadratic space of signature $(n-r-1, 2)$, the stabilizer H_x of x in H is isomorphic to $\mathrm{GSpin}(x^\perp)$, and there is a natural map of Shimura varieties

$$Z(x, K) : H_x(\mathbb{Q}) \backslash D_x \times H_x(\mathbb{A}_f)/H_x(\mathbb{A}_f) \cap K \longrightarrow X_K(\mathbb{C}),$$

giving a cycle of codimension r on $X_K(\mathbb{C})$. Given a function $\varphi \in S(V(\mathbb{A}_f)^r)^K$ and $T \in \mathrm{Sym}_r(\mathbb{Q})_{>0}$, there is a weighted linear combination $Z(T, \varphi)$ of such cycles [13], and the resulting cohomology classes $[Z(T,\varphi)] \in H^{2r}(X_K)$, where $H^*(X_K)$ is the cohomology of $X_K(\mathbb{C})$ with complex coefficients [17]. Examples are given by (1.1) and (1.4) in section 1.1, where φ is the characteristic function of $(\hat{L})^r \subset V(\mathbb{A}_f)^r$, for $\hat{L} = L \otimes_{\mathbb{Z}} \hat{\mathbb{Z}}$. If T is only positive semi–definite with $\mathrm{rank}(T) = r(T)$, the associated cycles have codimension $r(T)$ and their cohomology classes lie in $H^{2r(T)}(X_K)$.

To form a generating function, let $\tau = u + iv \in \mathfrak{H}_r$, the Siegel space of genus r, and let $q^T = e(\mathrm{tr}(\tau T))$.

Theorem 3.1 ([17]). — *For $\varphi \in S(V(\mathbb{A}_f)^r)^K$, and for a suitable choice of a Kähler form ω on $X_K(\mathbb{C})$, the generating series*

$$\phi_r(\tau, \varphi) = \sum_{T \in \mathrm{Sym}_r(\mathbb{Q})_{\geq 0}} [Z(T, \varphi)] \cdot [\omega]^{r - r(T)} \, q^T$$

is the q–expansion of a holomorphic Siegel modular form of weight $\frac{n+1}{2}$ and genus r valued in $H^{2r}(X_K)$.

The proof of this result depends on a construction of a theta function taking values in the space of closed $2r$ forms on $X_K(\mathbb{C})$. This method is quite general and applies to the locally symmetric spaces associated to $O(p,q)$, $U(p,q)$ and $Sp(p,q)$, cf. [15], [16], and [17].

Specializing to the case where V is anisotropic and $r = n-1$ and applying $\deg : H^{2(n-1)}(X_K) \to \mathbb{C}$, one obtains a holomorphic Siegel modular form of genus $n-1$ and weight $\frac{n+1}{2}$,

(3.1)
$$\phi_{\deg}(\tau, \varphi) = \deg\big(\phi_{n-1}(\tau, \varphi)\big) = \sum_{T \in \ \mathrm{Sym}_{n-1}(\mathbb{Q})_{\geq 0}} \deg\big(\, [Z(T, \varphi)] \cdot [\omega]^{n-1-r(T)} \,\big)\, q^T.$$

By the Siegel–Weil formula [20], this form is, in turn, a value of an Eisenstein series. More precisely, let

$$G'_{\mathbb{A}} = \begin{cases} \mathrm{M}p_{2(n-1)}(\mathbb{A}) & \text{if } n \text{ is even} \\ \mathrm{Sp}_{2(n-1)}(\mathbb{A}) & \text{if } n \text{ is odd.} \end{cases}$$

The machinery described in section 2 carries over, except that the map $\lambda_V : S(V(\mathbb{A})) \longrightarrow I(s_0, \chi)$ now takes values in the induced representation at the point $s_0 = \frac{1}{2}$. Let $E(g, s, \Phi)$ be the Siegel–Eisenstein series associated to the section $\Phi_\infty^{\frac{n+1}{2}}(s) \otimes \Phi_f(s)$, where $\Phi_f(\frac{1}{2}) = \lambda_V(\varphi)$ for the weight function $\varphi \in S(V(\mathbb{A}_f)^{n-1})^K$. The Siegel–Weil formula for the anisotropic space V then implies the following generalization of Corollary 1.5 above [13]:

Proposition 3.2. — *For $\tau = u + iv \in \mathfrak{H}_{n-1}$, and for a weight function $\varphi \in S(V(\mathbb{A}_f)^{n-1})^K$,*

$$\phi_{\deg}(\tau, \varphi) = \mathrm{vol}(X_K(\mathbb{C}))\ \det(v)^{-\frac{n+1}{4}} E(g_\tau, \frac{1}{2}, \Phi).$$

where
$$\mathrm{vol}(X_K(\mathbb{C})) = \int_{X_K(\mathbb{C})} \omega^{n-1}.$$
In particular, the positive definite Fourier coefficients of this Eisenstein series are the degrees of the (weighted) 0–cycles $Z(T,\varphi)$ on X_K.

4. GENERATING FUNCTIONS FOR ARITHMETIC 0–CYCLES: THE CASE OF SHIMURA CURVES

This section describes the generating function for the arithmetic degrees of 0–cycles on the arithmetic surface $\mathfrak{X}$ associated to a Shimura curve X over $\mathbb{Q}$. The series of interest will be analogous to the series $\deg(\phi_2(\tau,L))$ for a complex surface given in Corollary 1.4 and will have the form

$$\hat{\phi}_{\deg}(\tau) = \sum_{T\in \mathrm{Sym}_2(\mathbb{Z})} \widehat{\deg}(\widehat{\mathfrak{Z}}(T,v))\, q^T, \tag{4.1}$$

where $\tau = u+iv \in \mathfrak{H}_2$ and the $\widehat{\mathfrak{Z}}(T,v)$'s are certain classes in $\widehat{CH}^2(\mathfrak{X})$, the top arithmetic Chow group of the arithmetic surface $\mathfrak{X}$, [4]. As in the second example of section 1, the definition of the relevant cycles will depend on a modular interpretation. To give a more detailed explanation, it is convenient to begin with the geometric situation of section 3.

Fix an indefinite division quaternion algebra B over $\mathbb{Q}$. The space $V = \{x \in B \mid \mathrm{tr}(x) = 0\}$, equipped with the restriction of the reduced norm of B, is a three dimensional quadratic space over $\mathbb{Q}$ of signature $(1,2)$. The group $H = B^\times \simeq \mathrm{GSpin}(V)$ acts on V by conjugation. The Shimura curve X_K associated to a compact open subgroup $K \subset H(\mathbb{A}_f)$, is a moduli space for abelian surfaces with $\mathcal{O}_B$–action and level structure, where $\mathcal{O}_B$ is a maximal order in B. For example, suppose that $K = \big(\mathcal{O}_B \otimes_{\mathbb{Z}} \hat{\mathbb{Z}}\big)^\times$. Then $X = X_K$ is the (coarse) moduli scheme over $\mathbb{Q}$ for pairs (A,ι) consisting of an abelian surface A together with an action $\iota : \mathcal{O}_B \hookrightarrow \mathrm{End}(A)$ of $\mathcal{O}_B$, and $X(\mathbb{C}) \simeq \Gamma_B\backslash\mathfrak{H}$, where $\Gamma_B = (\mathcal{O}_B)^\times_+$ is the group of norm 1 units on $\mathcal{O}_B$.

The 0–cycles on X defined in section 3 can also be defined by specifying additional endomorphisms.

Definition 4.1. — *For an abelian surface (A,ι) with $\mathcal{O}_B$–action, the space of special endomorphisms is*
$$V(A,\iota) = \{x \in \mathrm{End}(A) \mid \iota(b)x = x\iota(b) \quad \textit{and} \quad \mathrm{tr}(x) = 0\}.$$
This space is equipped with a $\mathbb{Z}$–valued quadratic form defined, for a special endomorphism $x \in V(A,\iota)$, by $x^2 = -Q(x)\cdot 1_A$.

Definition 4.2. — *For $t \in \mathbb{Z}_{>0}$, the special cycle $Z(t)$ is the locus of triples (A,ι,x) where $x \in V(A,\iota)$ with $Q(x) = t$.*

In fact, one then has $Z(t) = Z(t, \varphi)$ where $Z(t, \varphi)$ is the 0–cycle on $X = X_K$ defined in section 3 for the weight function $\varphi \in S(V(\mathbb{A}_f))^K$, the characteristic function of the set $V(\mathbb{A}_f) \cap \big(\mathcal{O}_B \otimes_{\mathbb{Z}} \hat{\mathbb{Z}}\big)$.

The 0–cycle $Z(t)$ on X is rational over $\mathbb{Q}$ and is analogous to the set of CM points on the modular curve associated to the imaginary quadratic field $\mathbb{Q}(\sqrt{-t})$. For $\tau = u + iv \in \mathfrak{H}$, the upper halfplane and $q = e(\tau)$, the degree generating function (3.1) in the present case

$$\phi_{\deg}(\tau) = \operatorname{vol}(X) + \sum_{t>0} \deg(Z(t))\, q^t \tag{4.2}$$

is the value at $s = \frac{1}{2}$ of an Eisenstein series of weight $\frac{3}{2}$ for a congruence subgroup of $SL_2(\mathbb{Z})$, cf. Proposition 3.2. Here $\operatorname{vol}(X)$ is the volume of $X(\mathbb{C}) \simeq \Gamma_B\backslash\mathfrak{H}$ with respect to $-\frac{1}{2\pi}\, y^{-2}\, dx \wedge dy$.

Now consider the arithmetic case. Let $\mathfrak{X}$ be the coarse moduli scheme over $S = \operatorname{Spec}(\mathbb{Z})$ for abelian surfaces (A, ι) with $\mathcal{O}_B$–action satisfying Drinfeld's 'special' condition [2]. The arithmetic surface $\mathfrak{X}$ has generic fiber $\mathfrak{X}_{\mathbb{Q}} = X$, the canonical model of the Shimura curve; $\mathfrak{X}$ has good reduction at all primes $p \nmid D(B)$, where $D(B)$ is the product of the primes p such that $B_p = B \otimes_{\mathbb{Q}} \mathbb{Q}_p$ is a division algebra. The points of $\mathfrak{X}$ over an algebraically closed field $\boldsymbol{k}$ correspond to isomorphism classes of (A, ι)'s over $\boldsymbol{k}$.

For an abelian scheme (A, ι) over a connected base, the space of special endomorphisms $V(A, \iota)$ with its $\mathbb{Z}$–valued quadratic form Q is defined as before.

Definition 4.3. — *For $T \in \operatorname{Sym}_2(\mathbb{Z})_{>0}$, the arithmetic special cycle $\mathfrak{Z}(T)$ is the locus of triples (A, ι, x), where $x = [x_1, x_2] \in V(A, \iota)^2$ is a pair of special endomorphisms $x_i \in V(A, \iota)$ with $Q(x) = \frac{1}{2}((x_i, x_j))_{i,j} = T$.*

Proposition 4.4 ([23]). — *The cycle $\mathfrak{Z}(T)$ is either empty or is supported in the set of supersingular points of a single fiber $\mathfrak{X}_p$, where p is determined by T, as described in Lemma 4.5 below. If $p \nmid D(B)$, or if $p|D(B)$ but $p \nmid T$, then $\mathfrak{Z}(T)$ is a 0–cycle on $\mathfrak{X}_p$. If $p|D(B)$ and $p|T$, then $\mathfrak{Z}(T)$ is a union, with multiplicities, of components of the fiber $\mathfrak{X}_p$ (and some additional embedded components).*

Sketch of proof. — The proof of this result illustrates the way in which the basic structure of the cycle $\mathfrak{Z}(T)$ is determined by the space $V(A, \iota)$. Observe that, for a geometric point (A, ι) of $\mathfrak{X}$ and viewing A up to isogeny,

$$\operatorname{End}^0(A) = \operatorname{End}(A) \otimes_{\mathbb{Z}} \mathbb{Q} \simeq \begin{cases} B & \text{if } A \text{ is simple,} \\ M_2(C) & \text{if } A \simeq E \times E \text{ with } E \text{ ordinary,} \\ M_2(\mathbb{B}) & \text{if } A \simeq E \times E \text{ with } E \text{ supersingular,} \end{cases}$$

for an elliptic curve E. In the second case, $\operatorname{End}^0(E) \simeq C$ is an imaginary quadratic field which splits B, and, in the third case, which occurs only in characteristic $p > 0$,

$\mathrm{End}^0(E) \simeq \mathbb{B}$ is the quaternion algebra over $\mathbb{Q}$ ramified at ∞ and p. In this case, write $M_2(\mathbb{B}) \simeq B \otimes B^{(p)}$ where $B^{(p)}$ is the definite quaternion algebra over $\mathbb{Q}$ whose local invariants differ from those of B precisely at ∞ and p. Then, in the three cases, $\mathrm{End}^0(A,\iota) \simeq \mathbb{Q}$, C, and $B^{(p)}$ respectively, and

$$V^0(A,\iota) = V(A,\iota) \otimes_{\mathbb{Z}} \mathbb{Q} \simeq \begin{cases} 0 & \\ C^0 & = \{x \in C \mid \mathrm{tr}(x) = 0\}, \\ V^{(p)} & = \{x \in B^{(p)} \mid \mathrm{tr}(x) = 0\}. \end{cases}$$

It follows that $\mathfrak{Z}(T)_{\mathbb{Q}} = \varnothing$, i.e., the cycle $\mathfrak{Z}(T)$ has no points in characteristic 0. If $\mathfrak{Z}(T)$ meets the fiber $\mathfrak{X}_p$ at p, then $\mathfrak{Z}(T) \cap \mathfrak{X}_p$ is contained in the supersingular locus of $\mathfrak{X}_p$, and the rational quadratic space $V^{(p)}$ represents T, i.e., there exists a pair of vectors $x = [x_1, x_2] \in V^{(p)}(\mathbb{Q})$ such that $Q^{(p)}(x) = T$. This last condition implies that $\mathfrak{Z}(T)$ is supported in a single fiber, due to the following simple observation.

Lemma 4.5. — *(i) The quadratic spaces $V^{(p)}$ have the same determinant as V, i.e.,* $\det(V^{(p)}) = \det(V) \in \mathbb{Q}^\times/\mathbb{Q}^{\times,2}$.

(ii) For a given $T \in \mathrm{Sym}_2(\mathbb{Q})$ with $\det(T) \neq 0$, there is a unique three dimensional rational quadratic space V_T with $\det(V_T) = \det(V)$ which represents T; the quadratic form on V_T has matrix

$$\begin{pmatrix} T & \\ & \det(V)/\det(T) \end{pmatrix}. \tag{4.3}$$

Thus, if V_T is not isomorphic to any of the $V^{(p)}$'s then $\mathfrak{Z}(T)$ is empty, while if $V_T \simeq V^{(p)}$ for some p, then $\mathfrak{Z}(T)$ is supported in the supersingular locus of $\mathfrak{X}_p$.

Definition 4.6. — *A matrix $T \in \mathrm{Sym}_2(\mathbb{Z})_{>0}$ will be called* irregular *if $V_T \simeq V^{(p)}$ where $p|D(B)$ and $p|T$. Otherwise T will be called* regular.

The assertions about the irregular case require a more detailed argument, using the p-adic uniformization of the formal completion of $\mathfrak{X}_p$, [23]. □

The arithmetic Chow group $\widehat{CH}^2(\mathfrak{X})$ is generated by pairs $(\mathfrak{Z}, g)$ where $\mathfrak{Z}$ is a 0–cycle on $\mathfrak{X}$ and g is a smooth $(1,1)$–form on $X(\mathbb{C})$, modulo a suitable equivalence, [4], [32]. There is a degree map $\widehat{\deg} : \widehat{CH}^2(\mathfrak{X}) \to \mathbb{R}$.

For $T \in \mathrm{Sym}_2(\mathbb{Z})_{>0}$ regular, let

$$\widehat{\mathfrak{Z}}(T, v) = (\mathfrak{Z}(T), 0) \in \widehat{CH}^2(\mathfrak{X}). \tag{4.4}$$

Then $\mathfrak{Z}(T) = \mathrm{Spec}(R(T))$ for an Artin ring $R(T)$ in which p is nilpotent, and the corresponding positive definite coefficients of the generating function (4.1) are given by

$$\widehat{\deg}(\widehat{\mathfrak{Z}}(T, v)) = \log |R(T)|. \tag{4.5}$$

For nonsingular $T \in \mathrm{Sym}_2(\mathbb{Z})$ with signature $(1,1)$ or $(0,2)$, set

$$\widehat{\mathfrak{Z}}(T,v) = (0, \mathfrak{g}(T,v)) \in \widehat{CH}^2(\mathfrak{X}), \tag{4.6}$$

where $\mathfrak{g}(T,v)$ is a smooth $(1,1)$–form, depending on T and v, which is described in (6.3) of section 6.

Finally, if $T \in \mathrm{Sym}_2(\mathbb{Z})_{\geq 0}$ has rank 1, then, in effect, only one special endomorphism had been imposed, so that there is an associated divisor $\mathfrak{Z}(T)$ on $\mathfrak{X}$. A Green's function $\Xi(T,v)$ for this divisor is constructed in section 6, below. This function continues to make sense when $T \leq 0$. There is a resulting class

$$\widehat{\mathfrak{Z}}(T,v) = (\mathfrak{Z}(T), \Xi(T,v)) \in \widehat{CH}^1(\mathfrak{X}), \tag{4.7}$$

where $\mathfrak{Z}(T)$ is empty when $T \leq 0$. Let $\hat{\boldsymbol{\omega}}_{\mathfrak{X}}$ be the relative dualizing sheaf of $\mathfrak{X}$ over $\mathrm{Spec}(\mathbb{Z})$, with metric coming from the uniformization of $X(\mathbb{C})$ by D, viewed as an element of $\widehat{CH}^1(\mathfrak{X})$ via the isomorphism $\widehat{\mathrm{Pic}}(\mathfrak{X}) \simeq \widehat{CH}^1(\mathfrak{X})$, and let

$$\hat{\boldsymbol{\omega}}_{\mathfrak{X}}(v) = \hat{\boldsymbol{\omega}}_{\mathfrak{X}} + (0, \log\det(v)) \in \widehat{CH}^1(\mathfrak{X}). \tag{4.8}$$

This class plays a role analogous to that of $[\omega]$, the Kähler class, in section 3 above.

Using the arithmetic intersection pairing $\widehat{CH}^1(\mathfrak{X}) \times \widehat{CH}^1(\mathfrak{X}) \to \widehat{CH}^2(\mathfrak{X})$, the full arithmetic generating function (4.1), analogous to $\phi_{\deg}$ given in Corollary 1.4 and (3.1) in the geometric case, is then

$$\hat{\phi}_{\deg}(\tau) = \widehat{\deg}(\hat{\boldsymbol{\omega}}_{\mathfrak{X}}(v)^2) + \sum_{\mathrm{rank}(T)=1} \widehat{\deg}\big(\, \widehat{\mathfrak{Z}}(T,v) \cdot \hat{\boldsymbol{\omega}}_{\mathfrak{X}}(v)\,\big)\, q^T \\ + \sum_{\substack{T \\ \det(T)\neq 0}} \widehat{\deg}\big(\, \widehat{\mathfrak{Z}}(T,v)\,\big)\, q^T. \tag{4.9}$$

where suitable terms have been added[2] for irregular T.

In the present situation, the construction of Remark 2.3 yields an incoherent Eisenstein series $E(\tau,s,\varphi)$ of weight $\frac{3}{2}$ associated to $\varphi \in S(V(\mathbb{A}_f)^2)$, the characteristic function of $(\widehat{\mathcal{O}_B} \cap V(\mathbb{A}_f))^2$.

Conjecture 4.7. — *The generating function $\hat{\phi}_{\deg}(\tau)$ is a Siegel modular form of weight $\frac{3}{2}$, more precisely*

$$\hat{\phi}_{\deg}(\tau) = \mathrm{vol}(X(\mathbb{C})) \cdot E'(\tau, 0, \varphi). \tag{4.10}$$

The main results in this direction assert that many of the Fourier coefficients of the two series coincide. Recall that $D(B)$ is the product of the primes p at which B_p is division. Also put $V^{(\infty)} = V$, so that there is a rational quadratic space associated to each place of $\mathbb{Q}$. By Lemma 4.5, a given nonsingular $T \in \mathrm{Sym}_2(\mathbb{Q})$ is represented by at most one of these spaces. If T is positive definite, this space, if it exists, must

[2] by a still not very satisfactory procedure

be one of the $V^{(p)}$'s for a finite prime p, while if T has signature $(1,1)$ or $(0,2)$, then this space can only be $V = V^{(\infty)}$.

THEOREM 4.8 ([14]). — *Suppose that $T \in \mathrm{Sym}_2(\mathbb{Z})$ is nonsingular.*

(i) *If T is not represented by any $V^{(p)}$, then the T–th Fourier coefficient of both sides of* (4.10) *vanish.*

(ii) *If T is represented by $V^{(p)}$ with $p \nmid 2D(B)$, including $p = \infty$, then the T–th Fourier coefficients of the two sides of* (4.10) *agree, i.e.,*

$$\widehat{\deg}(\widehat{\mathfrak{Z}}(T,v))\, q^T = \mathrm{vol}(X(\mathbb{C})) \cdot E'_T(\tau, 0, \varphi). \tag{4.11}$$

In fact, work in progress [25] should extend (4.11) to all T of rank 1.

The proof of (ii) is similar to the proof of the main identity at the heart of the Gross–Zagier formula [8]; it amounts to an explicit computation of the two quantities, one from arithmetic geometry, the other from automorphic forms. A sketch of the proof is given in the next two sections.

5. NON–SINGULAR FOURIER COEFFICIENTS

The Fourier coefficients of the central derivative of an incoherent Eisenstein series $E(\tau, s, \varphi)$ associated to a rational quadratic space V of signature $(n-1, 2)$, as defined as in Remark 2.3, have an interesting structure.

For each prime $p \leq \infty$, define a quadratic space $V^{(p)}$ of dimension $n+1$ and the same determinant as that of V as follows. Let $V^{(\infty)} = V$. For a finite prime p, $V^{(p)}$ has signature $(n+1, 0)$ and local Hasse invariants at finite primes ℓ determined by

$$\epsilon_\ell(V^{(p)} \otimes_{\mathbb{Q}} \mathbb{Q}_\ell) = \begin{cases} \epsilon_\ell(V) & \text{if } \ell \neq p, \text{ and} \\ -\epsilon_p(V_p) & \text{if } \ell = p. \end{cases} \tag{5.1}$$

For each nonsingular $T \in \mathrm{Sym}_n(\mathbb{Q})$, there is a rational quadratic space V_T of dimension $n+1$ and the same determinant as V defined by equation (4.3).

It is well known that if the function $\varphi = \otimes_{p<\infty}\varphi_p \in S(V(\mathbb{A}_f)^n)$ is factorizable and $\mathrm{Re}(s) > \frac{n+1}{2}$, then each nonsingular Fourier coefficient of $E(\tau, s, \varphi)$ is given by a product

$$E_T(\tau, s, \varphi) = W^{\frac{n+1}{2}}_{T,\infty}(\tau, s) \cdot \prod_{p<\infty} W_{T,p}(s, \varphi_p)$$

of local (degenerate) Whittaker functions, [20]. In classical language, the archimedean factor is a confluent hypergeometric function of a matrix argument studied by Shimura [30], while the product over the finite primes is the Siegel series.

PROPOSITION 5.1 ([14], section 6). — (i) *If V_T is not isomorphic to any $V^{(p)}$, then*

$$E'_T(\tau, 0, \varphi) = 0.$$

(ii) *If $V_T \simeq V^{(p)}$ for a finite prime, then $W_{T,p}(0, \varphi_p) = 0$ and*

$$E'_T(\tau, 0, \varphi) = q^T \cdot W_{T,p}(s, \varphi_p)'\big|_{s=0} \cdot A_T^{(p)}(\varphi),$$

where $A_T^{(p)}(\varphi)$ is, up to a factor at p, the Fourier coefficient of a theta integral (cf. (2.2)*) attached to $V^{(p)}$.*

(iii) *If $V_T \simeq V^{(\infty)} = V$,*

$$E'_T(\tau, 0, \varphi) = W_{T,\infty}^{\frac{n+1}{2}}(\tau, s)'\big|_{s=0} \cdot A_T^{(\infty)}(\varphi),$$

where $A_T^{(\infty)}(\varphi)$ is the T–th Fourier coefficient of a theta integral attached to V.

Idea of proof of Theorem 4.8 *for $p < \infty$.* Restricting to the case $n = 2$ as in section 4, suppose that $T \in \mathrm{Sym}_2(\mathbb{Z})$ is nonsingular with $V_T \simeq V^{(p)}$ for a finite prime $p \nmid 2D(B)$. In this case, using Proposition 5.1, the identity to be proved in (ii) of Theorem 4.8 amounts to

$$\widehat{\deg}(\widehat{\mathfrak{Z}}(T, v)) = \widehat{\deg}((\mathfrak{Z}(T), 0)) = \mathrm{vol}(X(\mathbb{C})) \cdot W_{T,p}(s, \varphi_p)'\big|_{s=0} \cdot A_T^{(p)}(\varphi). \tag{5.2}$$

This identity, which is of the same nature as the identities between heights and Fourier coefficients involved in the Gross-Zagier formula, is proved by computing the two sides explicitly.

On the geometric side, since T is regular, $\mathfrak{Z}(T)$ is a collection of supersingular points on $\mathfrak{X}_p$, each counted with a certain multiplicity. This multiplicity is the length of the local Artin ring associated to the deformations of the triple (A, ι, x), where x is a pair of special endomorphisms. By the Serre–Tate Theorem, one can pass to the deformations of $(A(p), \iota, x)$ where $A(p)$ is corresponding p–divisible group with the action ι of $(\mathcal{O}_B)_p = \mathcal{O}_B \otimes_{\mathbb{Z}} \mathbb{Z}_p \simeq M_2(\mathbb{Z}_p)$ and a pair of special endomorphisms x. Since A is isogenous to $E \times E$ for a supersingular elliptic curve E, one reduces, via the idempotents in $(\mathcal{O}_B)_p$, to the problem of deforming $(E(p), x)$ for a pair $x = [x_1, x_2]$ of endomorphisms of the p–divisible group of such a curve. Note that $E(p)$ is a p–divisible formal group of dimension 1 and height 2. The length of the associated Artin ring is then obtained by specializing a beautiful result of Gross and Keating [6].

They consider the deformations of a collection $(\mathbb{X}, \mathbb{X}', y)$ where $\mathbb{X}$ and $\mathbb{X}'$ are formal groups of dimension 1 and height 2 over $\bar{\mathbb{F}}_p$ and $y = [y_1, y_2, y_3]$ is a triple of nonzero isogenies $y_i : \mathbb{X} \to \mathbb{X}'$. Note that the universal deformation ring of the pair $(\mathbb{X}, \mathbb{X}')$ is $W[[t, t']]$, where $W = W(\bar{\mathbb{F}}_p)$ is the ring of Witt vectors of $\bar{\mathbb{F}}_p$.

PROPOSITION 5.2 (Gross-Keating, [6], Proposition 5.4). — *Suppose that the matrix Q of inner products of the triple $y = [y_1, y_2, y_3]$ with respect to the degree quadratic form on* $\mathrm{Hom}(\mathbb{X}, \mathbb{X}')$ *has invariants $a_1 \leq a_2 \leq a_3$. For p odd, this means that*

$Q \in \mathrm{Sym}_3(\mathbb{Z}_p)$ is $GL_3(\mathbb{Z}_p)$ equivalent to $\mathrm{diag}(\epsilon_1 p^{a_1}, \epsilon_2 p^{a_2}, \epsilon_3 p^{a_3})$ for units $\epsilon_i \in \mathbb{Z}_p^\times$. Then the length of the deformation ring of $(\mathbb{X}, \mathbb{X}', y)$ is given by:

$$\sum_{i=0}^{a_1-1} (i+1)(a_1+a_2+a_3-3i)\, p^i + \sum_{i=a_1}^{(a_1+a_2-2)/2} (a_1+1)(2a_1+a_2+a_3-4i)\, p^i + \frac{1}{2}(a_1+1)(a_3-a_2+1)\, p^{(a_1+a_2)/2} \tag{5.3}$$

if a_1+a_2 is even, and

$$\sum_{i=0}^{a_1-1} (i+1)(a_1+a_2+a_3-3i)\, p^i + \sum_{i=a_1}^{(a_1+a_2-1)/2} (a_1+1)(2a_1+a_2+a_3-4i)\, p^i$$

if a_1+a_2 is odd.

Specialized to the case where y_1 is an isomorphism, so that $a_1 = 0$, one obtains an explicit formula for the multiplicity $\delta_p(T)$ of a point in $\mathfrak{Z}(T)$, and, in particular, observes that this multiplicity depends only on the $\mathrm{GL}_2(\mathbb{Z}_p)$–equivalence class of T, and not on the particular point. Thus, the left hand side of (5.2) has the form $(\delta_p(T) \cdot (\ \#\ \text{points in } \mathfrak{Z}(T)))$.

On the other hand, for $p \neq 2$, the quantity $W_{T,p}(s, \varphi_p)'\big|_{s=0}$ on the analytic side of (5.2) can be computed from the result of Kitaoka [12]. More precisely, the quadratic form on the lattice $L_p = (\mathcal{O}_B \otimes \mathbb{Z}_p) \cap V_p$ has matrix $S_0 = \mathrm{diag}(1, 1, -1)$ for a suitable basis. Let

$$S_r = \mathrm{diag}(1, 1, -1, \underbrace{1, \dots, 1}_{r}, \underbrace{-1, \dots, -1}_{r})$$

be the quadratic form obtained from S_0 by adding r hyperbolic planes. Then

$$W_{T,p}(r, \varphi_p) = \gamma_p \cdot \alpha_p(S_r, T),$$

for a root of unity γ_p, where

$$\alpha_p(S_r, T) = \lim_{t \to \infty} p^{-t(3+4r)}\, |\{x \in M_{3+2r,2}(\mathbb{Z}_p/p^t\mathbb{Z}_p) \mid {}^t x S_r x \equiv T \mod (p^t)\}|$$

is the classical representation density of T by S_r as defined by Siegel, [12]. This quantity has the form $\alpha_p(S_r, T) = A_p(q^{-r}, T)$ for a polynomial $A_p(X, T)$, which was first calculated in this case by Kitaoka [12][(3)]. When T is such that $W_{T,p}(0, \varphi_p) = 0$, and $T \in \mathrm{Sym}_2(\mathbb{Z}_p)$ is $GL_2(\mathbb{Z}_p)$–equivalent to $\mathrm{diag}(\epsilon_1 p^a, \epsilon_2 p^b)$, then Kitaoka's formula yields

$$W_{T,p}(s, \varphi_p)'\big|_{s=0} = -\log(p)\, \gamma_p\, \frac{\partial}{\partial X}\{A_p(X, T)\}\big|_{X=1}, \tag{5.4}$$

(3) Recently, it has been calculated in general, for $p \neq 2$, by F. Sato and Y. Hironaka, [29].

that is

$$\log(p)\,\gamma_p\,(1-p^{-2})\times\begin{cases}\displaystyle\sum_{j=0}^{(a-1)/2}(a+b-4j)\,p^j & \text{if } a \text{ is odd,}\\[2ex] \displaystyle\sum_{j=0}^{(a-2)/2}(a+b-4j)\,p^j+\tfrac{1}{2}(b-a+1)\,p^{a/2} & \\ & \text{if } a \text{ is even and } b \text{ is odd.}\end{cases}$$

Remarkably,

$$W_{T,p}(s,\varphi_p)'\big|_{s=0}=\log(p)\,\gamma_p\,(1-p^{-2})\,\delta_p(T), \tag{5.5}$$

where $\delta_p(T)$ is the multiplicity computed via the Gross–Keating formula! Finally, by a straightforward counting argument, the number of points in $\mathfrak{Z}(T)$ is given by $A_T^{(p)}(\varphi)$, up to simple factors, independent of T, which compensate for the extra p^2-1 and the factor $\mathrm{vol}(X(\mathbb{C}))$. □

6. GREEN'S FUNCTIONS AND WHITTAKER FUNCTIONS

The classes $\widehat{\mathfrak{Z}}(T,v)$ in the arithmetic Chow groups of integral models of Shimura curves involve the Green's functions defined in [14], section 11. This section describes the construction of such functions. First, suppose that V is a rational quadratic space of signature $(n-1,2)$, as in section 3, and recall that for $t\in\mathbb{Q}_{>0}$ and for $\varphi\in S(V(\mathbb{A}_f))^K$, there is a divisor $Z(t,\varphi)$ on X_K, given as a weighted sum of the images in X_K of the divisors D_x in D. A Green's function of logarithmic type for $Z(t,\varphi)$ in the sense of Gillet–Soulé [4] can be constructed by averaging rapidly decreasing Green's functions for D_x's. Next, in the case of a Shimura curve ($n=2$), the smooth $(1,1)$–form $\mathfrak{g}(T,v)$ used in the construction of the terms for indefinite T's in the generating function of arithmetic degrees is defined via the star product. The notation is that of section 3.

For an oriented negative 2–plane $z\in D$, let pr_z be the projection to z with respect to the orthogonal decomposition $V(\mathbb{R})=z+z^{\perp}$. For $x\in V(\mathbb{R})$, $x\neq 0$, the quantity

$$R(x,z)=-(\mathrm{pr}_z(x),\mathrm{pr}_z(x))$$

is nonnegative and vanishes if and only if $z\in D_x$. Let

$$\xi(x,z)=-\mathrm{Ei}(-2\pi R(x,z)),$$

where, for $r>0$,

$$-\mathrm{Ei}(-r)=\int_1^\infty\frac{e^{-rt}}{t}\,dt$$

is the exponential integral. Since $-\mathrm{Ei}(-r)$ decays exponentially as r goes to infinity and behaves like $-\log(r)+O(1)$ as r goes to 0, the function $\xi(x)$ has a logarithmic singularity along the (possibly empty) divisor D_x and decays very rapidly away from D_x. In addition, it satisfies the Green's equation:

$$dd^c\xi(x) + \delta_{D_x} = \mu(x)$$

for a smooth $(1,1)$–form $\mu(x)$ on D, and hence defines a Green's form of log type for D_x, in the sense of Gillet-Soulé.

Note that $R(hx, hz) = R(x, z)$ for $h \in \mathrm{GSpin}(V)(\mathbb{R})$, so that $\xi(hx, hz) = \xi(x, z)$ as well. For $v > 0$ and $t \in \mathbb{Q}_{>0}$, and a weight function $\varphi \in S(V(\mathbb{A}_f))^K$, the sum

$$\mathfrak{g}(t, v, \varphi)(z, h) = \sum_{\substack{x \in V(\mathbb{Q}) \\ Q(x)=t}} \varphi(h^{-1}x)\, \xi(x, z) \tag{6.1}$$

depends only on the orbit $H(\mathbb{Q})(z, h)K$ of the point $(z, h) \in D \times H(\mathbb{A}_f)$. Thus, $\mathfrak{g}(t, v, \varphi)$ defines a Green's function of logarithmic type for the divisor $Z(t, \varphi)$, defined in section 3 (for $r = 1$), on X_K.

In the case $n = 2$, for the space V considered in section 4, and a pair of vectors $x = [x_1, x_2] \in V(\mathbb{R})^2$ such that $\det(Q(x)) \neq 0$, the integral

$$\Lambda(x) = \int_D \xi(x_1) * \xi(x_2)$$

of the $*$–product of the associated Green's functions [4] is well defined and satisfies $\Lambda(hx) = \Lambda(x)$ for all $h \in \mathrm{GSpin}(V)(\mathbb{R})$. This implies that $\Lambda(x)$ actually only depends on the matrix of inner products $Q(x)$, i.e.,

$$\Lambda(x) =: \Lambda_0(Q(x)). \tag{6.2}$$

One can view $\Lambda(x)$ as the 'archimedean height pairing' of the 0–cycles D_{x_1} and D_{x_2} in D. In fact this quantity has the following rather surprising additional invariance:

THEOREM 6.1 ([14], section 13). — *For any $k \in SO(2)$, and any x,*

$$\Lambda(xk) = \Lambda(x).$$

Note that, even when $Q(x_1) > 0$ and $Q(x_2) > 0$, so that one initially has a pair of 0–cycles D_{x_1} and D_{x_2} in D, eventually, after rotation, one encounters a pair of vectors x_1' and x_2' with $Q(x_2') < 0$, so that the cycle $D_{x_2'}$ has vanished (!) and is replaced, in some sense, by the geodesic arc $\{z \in D \mid x_2' \in z\}$. Nonetheless, the 'archimedean height pairing' $\Lambda(x)$ is invariant under such a deformation.

It follows that, writing $v \in \mathrm{Sym}_2(\mathbb{R})_{>0}$ as $v = a^t a$ for $a \in \mathrm{GL}_2(\mathbb{R})^+$, the quantity $\Lambda(xa)$ depends only on v and not on the choice of a. Then, for a nonsingular $T \in \mathrm{Sym}_2(\mathbb{Z})$, one has a smooth $(1,1)$–form

$$\mathfrak{g}(T, v) = \Bigg(\sum_{\substack{x \in V(\mathbb{Q})^2 \\ Q(x)=T \mod \Gamma_B}} \varphi(x)\, \Lambda(xa) \Bigg) \mu \tag{6.3}$$

on X, where $\mu = y^{-2}\,dx\wedge dy$, φ is the characteristic function of the set $\big(\widehat{\mathcal{O}_B}^{\,2} \cap V(\mathbb{A}_f)^2\big)$, and $\Gamma_B = \mathcal{O}_B^\times$. This is the form used in the definition (4.6) of $\widehat{\mathfrak{Z}}(T,v)$ when T has signature $(1,1)$ or $(0,2)$. Note that the sum is nonempty precisely when V represents T.

Idea of proof of Theorem 4.8 *for* $p=\infty$. Again using Proposition 5.1, the identity to be proved in this case is

$$\widehat{\deg}(\widehat{\mathfrak{Z}}(T,v))\,q^T = \mathrm{vol}(X(\mathbb{C}))\cdot W_{T,\infty}^{\frac{n+1}{2}}(\tau,s)'\big|_{s=0}\cdot A_T^{(\infty)}(\varphi).$$

The left hand side is simply

$$q^T\cdot\Lambda_0({}^t aTa)\cdot \sum_{\substack{x\in V(\mathbb{Q})^2\\ Q(x)=T \mod \Gamma_B}} \varphi(x), \tag{6.4}$$

where $v = a^t a$, as before and Λ_0 is given by (6.2). Then some transformations of the integral representations of the matrix argument confluent hypergeometric function of Shimura's paper [30] together with a manipulation of the integral defining Λ shows that

$$W_{T,\infty}^{\frac{n+1}{2}}(\tau,s)'\big|_{s=0} = c_\infty\cdot\Lambda_0({}^t aTa)\cdot q^T,$$

where c_∞ is an innocuous constant. Again, the sum in (6.4) counts pairs of lattice vectors modulo Γ_B and coincides with $A_T^{(\infty)}$, the Fourier coefficient of the theta integral, up to a constant which absorbs c_∞ and provides the required $\mathrm{vol}(X(\mathbb{C}))$ factor. $\square$

7. FURTHER RESULTS

One would like to identify the central derivative $E'(\tau,0,\varphi)$ of the incoherent Eisenstein series (2.3) as a generating function for arithmetic degrees in the general case. In the series of papers [24], [22], [21] the cases $n=1$, 3 and 4 are considered. In each of these cases (and also for $n=5$), the Shimura variety X associated to $H=\mathrm{GSpin}(V)$ is of PEL type, i.e., can be interpreted as a moduli space for abelian varieties (A,ι) with a specified endomorphism ring, due to the existence of an accidental isomorphism. This allows one to give a modular definition of an integral model $\mathfrak{X}_K$ of X_K, at least over $\mathrm{Spec}(\mathbb{Z}[N^{-1}])$ for a suitable N depending on the compact open subgroup K. For each abelian scheme (A,ι), there is a space of special endomorphisms $V(A,\iota)$, equipped with a $\mathbb{Z}$–valued quadratic form. Special cycles on $\mathfrak{X}$ are then defined by imposing collections of such endomorphisms, as in section 4 and the example of section 1.3 above.

The case $n=1$. This case is considered in [24] and is described in classical language in section 1.3 above. From the point of view now developed it amounts to the following. An imaginary quadratic field $\boldsymbol{k} = \mathbb{Q}(\sqrt{-d})$ with quadratic form $Q(x) = -N_{\boldsymbol{k}/\mathbb{Q}}(x)$ gives a rational quadratic space (V,Q) of signature $(0,2)$. The group $\mathrm{GSpin}(V)$ is then the torus over $\mathbb{Q}$ with $\mathrm{GSpin}(V)(\mathbb{Q}) \simeq \boldsymbol{k}^\times$, and $\mathfrak{X} \simeq \mathrm{Spec}(\mathcal{O}_H)$ is the restriction of scalars to $\mathrm{Spec}(\mathbb{Z})$ of the coarse moduli space over $\mathcal{O}_{\boldsymbol{k}}$ of elliptic curves (A, ι) with complex multiplication by $\mathcal{O}_{\boldsymbol{k}}$. In this case the space of special endomorphisms (1.6) is

$$V(A, \iota) = \{x \in \mathrm{End}(A) \mid x\iota(b) = \iota(\bar{b})x\}.$$

This space is zero unless A is a supersingular elliptic curve in characteristic p, where p is not split in $\boldsymbol{k}$, in which case, $V(A, \iota) \otimes_{\mathbb{Z}} \mathbb{Q} \simeq V^{(p)}$, the rational quadratic space given by $V^{(p)} = \boldsymbol{k}$ with quadratic form $Q^{(p)}(x) = p\, N_{\boldsymbol{k}/\mathbb{Q}}(x)$. For $\tau = u + iv \in \mathfrak{H}$, the generating function is then given by

$$\hat{\phi}_{\deg}(\tau) = \sum_{t \in \mathbb{Z}} \widehat{\deg}(\widehat{\mathfrak{Z}}(t, v))\, q^t. \tag{7.1}$$

Here, for $t > 0$, $\widehat{\mathfrak{Z}}(t, v) = (\mathfrak{Z}(t), 0) \in \widehat{CH}^1(\mathfrak{X})$ where $\mathfrak{Z}(t)$ is the locus of (A, ι, x)'s where $x \in V(A, \iota)$ with $Q(x) = -x^2 = t$. For $t < 0$, $\widehat{\mathfrak{Z}}(t, v) = (0, \mathfrak{g}(t, v)) \in \widehat{CH}^1(\mathfrak{X})$, where $\mathfrak{g}(t, v)$ is the function on $X(\mathbb{C}) \simeq H(\mathbb{Q})\backslash D \times H(\mathbb{A}_f)/K$ given by (6.1). Here $K \simeq \widehat{\mathcal{O}}_{\boldsymbol{k}}^\times$. Note that in this case, D consists of two points (via the two possible orientations of $V(\mathbb{R})$), and $R(x, z) = 2N_{\boldsymbol{k}/\mathbb{Q}}(x)$, so that $\xi(xa, z) = -\mathrm{Ei}(-4\pi v N_{\boldsymbol{k}/\mathbb{Q}}(x))$, precisely as in [24]. This gives an improved version of Theorem 1.7:

THEOREM 7.1 ([24]). — *When $d \equiv 3 \mod (4)$ is a prime and for a suitable definition of the constant term $\widehat{\deg}(\widehat{\mathfrak{Z}}(0, v))$,*

$$\hat{\phi}_{\deg}(\tau) = \mathrm{vol}(X(\mathbb{C})) \cdot E'(\tau, 0, \varphi),$$

where φ is the characteristic function of $\widehat{\mathcal{O}_{\boldsymbol{k}}} \subset V(\mathbb{A}_f)$ and $\mathrm{vol}(X(\mathbb{C})) = h_{\boldsymbol{k}}$ is the class number of $\boldsymbol{k}$.

The result is proved by a direct calculation of both sides, using the results of Gross [5] to compute multiplicities on the moduli space. The restriction to prime discriminant is only made to streamline the calculations.

The higher dimensional cases treated so far exhibit some new phenomena.

The case $n = 3$. This case is considered in [22]. The incoherent Eisenstein series (2.2) associated to a rational quadratic space of signature $(2, 2)$ will have weight 2 and genus 3. To define the relevant moduli problem and generating function for arithmetic degrees, let $C(V) = C^+(V) \oplus C^-(V)$ be the Clifford algebra of V with its $\mathbb{Z}_2$–grading. Then the center of $C^+(V)$ is a real quadratic field $\boldsymbol{k}$ (possibly $\boldsymbol{k} = \mathbb{Q} \oplus \mathbb{Q}$), and $C^+(V)$ has the form $B_0 \otimes_{\mathbb{Q}} \boldsymbol{k}$ for an indefinite quaternion algebra B_0 over $\mathbb{Q}$. The associated Shimura variety X is a surface whose complex points parametrize polarized abelian varieties (A, ι) of dimension 8 with an action of a maximal order in $C(V) \otimes \boldsymbol{k}$. Included among the X's are products of modular curves ($\boldsymbol{k} = \mathbb{Q} \oplus \mathbb{Q}$, $B_0 = M_2(\mathbb{Q})$), products of Shimura curves ($\boldsymbol{k} = \mathbb{Q} \oplus \mathbb{Q}$, B_0 division), Hilbert-Blumenthal surfaces ($\boldsymbol{k}$ a field, $B_0 = M_2(\mathbb{Q})$) and their twisted (quaternionic) analogues ($\boldsymbol{k}$ a field, B_0 division). An integral model $\mathfrak{X}$ of X over $\mathrm{Spec}(\mathbb{Z}[N^{-1}])$, is defined as the moduli space of polarized abelian schemes with such an action, level structure, etc. The space of special endomorphisms of a given (A,ι) is then

$$V(A,\iota)=\{x\in \mathrm{End}(A)\mid x\,\iota(c\otimes a)=\iota(c\otimes \bar{a})\,x,\ \text{and } x^*=x\}$$

where $*$ denotes the Rosati involution of A. As before, this space has a $\mathbb{Z}$–valued quadratic form defined by $x^2=Q(x)\cdot 1_A$. Again, a key point is that, for (A,ι) supersingular, the space $V(A,\iota)\otimes\mathbb{Q}\simeq V^{(p)}$ is the 4 dimensional rational quadratic space defined by (5.1) above. For $T\in \mathrm{Sym}_3(\mathbb{Z})_{>0}$, the special cycle $\mathfrak{Z}(T)$ is the locus of (A,ι,x)'s where $x\in V(A,\iota)^3$ with $Q(x)=T$. This cycle is either empty or is supported in the supersingular locus $\mathfrak{X}_p^{\mathrm{s.s.}}$ of the fiber at p for the unique prime p for which $V_T\simeq V^{(p)}$. Here one assumes that $p\nmid N$, so that p is a prime of good reduction; in particular, p is not ramified in $\boldsymbol{k}$.

If a prime $p\nmid N$ splits in $\boldsymbol{k}$, then the supersingular locus $\mathfrak{X}_p^{\mathrm{s.s.}}$ consists of a finite set of points. If p is inert in $\boldsymbol{k}$, then $\mathfrak{X}_p^{\mathrm{s.s.}}$ has dimension 1 and is, in fact, a union of $\mathbb{P}^1$'s [31], [22], section 4, crossing at $\mathbb{F}_{p^2}$ rational points.

Let $\varphi\in S(V(\mathbb{A}_f)^3)$ be the characteristic function of $L\otimes\hat{\mathbb{Z}}$ for a lattice $L\subset V(\mathbb{Q})$ such that L_p is self dual for all $p\nmid N$, and let $E(\tau,s,\varphi)$ be the associated incoherent Eisenstein series (2.3) of weight 2.

THEOREM 7.2 ([22]). — *Suppose that $T\in \mathrm{Sym}_3(\mathbb{Z})_{>0}$.*

(i) *If V_T is not isomorphic to any $V^{(p)}$, then $\mathfrak{Z}(T)$ is empty and $E'_T(\tau,0,\varphi)=0$.*

(ii) *If $V_T\simeq V^{(p)}$ for a prime $p\nmid N$ which splits in $\boldsymbol{k}$, then there is a class $\widehat{\mathfrak{Z}}(T)=(\mathfrak{Z}(T),0)\in\widehat{CH}^3(\mathfrak{X})$ and*

$$\widehat{\deg}(\widehat{\mathfrak{Z}}(T))\cdot q^T=C\cdot E'_T(\tau,0,\varphi),\tag{7.2}$$

for a constant C independent of T.

(iii) *Suppose that $V_T\simeq V^{(p)}$ for a prime $p\nmid N$ which is inert in $\boldsymbol{k}$. If $p\nmid T$, then $\mathfrak{Z}(T)$ consists of a finite number of points, there is a class $\widehat{\mathfrak{Z}}(T)=(\mathfrak{Z}(T),0)\in\widehat{CH}^3(\mathfrak{X})$ and, again,*

$$\widehat{\deg}(\widehat{\mathfrak{Z}}(T))\cdot q^T=C\cdot E'_T(\tau,0,\varphi).\tag{7.3}$$

If $p\mid T$, then $\mathfrak{Z}(T)$ is a union of components of the supersingular locus $\mathfrak{X}_p^{\mathrm{s.s.}}$.

Here, $T\in \mathrm{Sym}_3(\mathbb{Z})_{>0}$ will be called irregular if $V_T\simeq V^{(p)}$ with $p\nmid N$ inert in $\boldsymbol{k}$ and $p\mid T$. The situation for $p\mid N$, e.g., p ramified in $\boldsymbol{k}$, has not yet been studied.

Here, as in (4.5), $\widehat{\deg}(\widehat{\mathfrak{Z}}(T))=\log|R(T)|$ for the Artin ring $R(T)$ defining $\mathfrak{Z}(T)$, so that the fact that $\mathfrak{X}$ is only defined over $\mathrm{Spec}(\mathbb{Z}[N^{-1}])$ and need not be proper will not be important. These issues will be essential, however, if one wants to define the full generating function $\hat{\phi}_{\deg}(\tau)$ and compare it to $E'(\tau,0,\varphi)$. The proof of (7.2) and (7.3), where $\mathfrak{Z}(T)$ is a 0–cycle, again comes down to a relation between derivatives of representation densities for quadratic forms and the result of Gross and Keating described in Proposition 5.2 above, now in its full generality. Indeed, that result was obtained in connection with the study of the derivative of a Siegel–Eisenstein series of weight 2, due to the connection of such a series with the triple product L–function [3], [27], [9], [7].

The case $n = 4$. This case is considered in [21]. Here the Shimura varieties are (twisted) Siegel 3–folds, and the pattern is similar to that for $n = 3$. The one new point is that the 'regularity' condition on $T \in \mathrm{Sym}_4(\mathbb{Z})$ required to obtain a 0–cycle in the supersingular locus (which is again a curve with $\mathbb{P}^1$ components) becomes: $V_T \simeq V^{(p)}$ and T represents 1 over $\mathbb{Z}_p$. One then obtains an analogue of Theorem 6.2, with the comparison again based on [6].

8. FINAL REMARKS

Beyond the range of the accidental isomorphisms, i.e., for $n \geq 6$, the Shimura varieties associated to $\mathrm{GSpin}(V)$ for rational quadratic spaces V of signature $(n-1, 2)$ are no longer of PEL type, so that the modular interpretation of points, special endomorphisms, and other tools used before are no longer available. Instead, it will be necessary to work with integral models defined by suitable types of Hodge classes, etc., [26]. Presumably there is a good notion of special endomorphisms or special Hodge classes in this situation which cut out the required cycles. This theory remains to be established.

Even in the range $2 \leq n \leq 5$, many difficulties lie in the way of a full treatment of the generating function for $\hat{\phi}_{\mathrm{deg}}(\tau)$. For example, there is the problem of the contribution of irregular T's, which occur even in the case of good reduction. The contribution of the singular T's is 'global' and will presumably involve models over $\mathbb{Z}$, detailed information about the fibers of bad reduction, etc. Much work remains to be done.

In addition, there is the question of defining modular arithmetic generating functions $\hat{\phi}_r(\tau)$, analogous to the series $\phi_r(\tau)$ of Theorem 3.1 above, valued in arithmetic Chow groups $\widehat{CH}^r(\mathfrak{X})$ for arbitrary codimension. Of course the image of such series under the cycle class map should be the generating series discussed in section 3. Recent work of Borcherds [1], which gives a generating function involving the classes of the divisors $Z(t, \varphi)$ on $X_K(\mathbb{C})$ in the *Chow group* $CH^1(X_K(\mathbb{C}))$, rather than in cohomology, will be relevant here.

In all cases, it remains to work out the modifications required in the case of non-compact quotient.

REFERENCES

[1] R. E. Borcherds, *The Gross-Kohnen-Zagier Theorem in higher dimensions*, Duke Math. J. **97** (1999), pp. 219–233.

[2] J.-F. Boutot and H. Carayol, *Uniformisation p–adiques des courbes de Shimura*, Astérisque **196–197** (1991), 45–158.

[3] P. Garrett, *Decomposition of Eisenstein series: Rankin triple products*, Annals of Math. **125** (1987), 209–235.

[4] H. Gillet and C. Soulé, *Arithmetic intersection theory*, Jour of IHES, **72** (1990), 94–174.

[5] B. H. Gross, *On canonical and quasi-canonical liftings*, Inventiones Math. **84** (1986), 321–326.

[6] B. H. Gross and K. Keating, *On the intersection of modular correspondences*, Invent. Math. **112** (1993), 225–245.

[7] B. H. Gross and S. Kudla, *Heights and the central critical values of triple product L-functions*, Compositio Math., **81** (1992), 143–209.

[8] B. H. Gross and D. Zagier, *Heegner points and the derivatives of L-series*, Inventiones math. **84** (1986), 225–320.

[9] M. Harris and S. Kudla, *The central critical value of a triple product L-function*, Annals of Math. **133** (1991), 605-672.

[10] S. Helgason, *Differential Geometry and Symmetric Spaces*, Academic Press, 1962.

[11] F. Hirzebruch and D. Zagier, *Intersection numbers of curves on Hilbert modular surfaces and modular forms of Nebentypus*, Inventiones Math. **36** (1976), 57–113.

[12] Y. Kitaoka, *A note on local densities of quadratic forms*, Nagoya Math. J. **92** (1983), 145–152.

[13] S. Kudla, *Algebraic cycles on Shimura varieties of orthogonal type*, Duke Math. J. **86** (1997), 39–78.

[14] S. Kudla, *Central derivatives of Eisenstein series and height pairings*, Annals of Math. **146** (1997), 545–646.

[15] S. Kudla and J. Millson, *The theta correspondence and harmonic forms I*, Math. Annalen, **274** (1986), 353–378.

[16] S. Kudla and J. Millson, *The theta correspondence and harmonic forms II*, Math. Annalen, **277** (1987), 267–314.

[17] S. Kudla and J. Millson, *Intersection numbers of cycles on locally symmetric spaces and Fourier coefficients of holomorphic modular forms in several complex variables*, IHES **71** (1990), 121–172.

[18] S. Kudla and S. Rallis, *Degenerate principal series and invariant distributions*, Israel J. Math., **69** (1990), 25–45

[19] S. Kudla and S. Rallis, *Ramified degenerate principal series representations for $Sp(n)$*, Israel J. Math., **78** (1992), 209–256.

[20] S. Kudla and S. Rallis, *A regularized Siegel-Weil formula: the first term identity*, Annals of Math. **140** (1994), 1–80.

[21] S. Kudla and M. Rapoport, *Cycles on Siegel 3–folds and derivatives of Eisenstein series*, Ann. École Norm. Sup. **33** (2000), 695–756.

[22] S. Kudla and M. Rapoport, *Arithmetic Hirzebruch–Zagier cycles*, J. reine angew. Math. **515** (1999), 155–244.

[23] S. Kudla and M. Rapoport, *Heights on Shimura curves and p–adic uniformization*, Inventiones Math. (to appear)

[24] S. Kudla, M. Rapoport and T. Yang, *On the derivative of an Eisenstein series of weight 1*, Int. Math. Res. Notices No. 7 (1999), 347–385.

[25] S. Kudla, M. Rapoport and T. Yang (in preparation).

[26] B. Moonen, *Models of Shimura varieties in mixed characteristics*, in Galois Representations in Arithmetic Algebraic Geometry, A. J. Scholl and R. L. Taylor, eds., London Math. Society Lecture Note Series **254**, Cambridge Univ. Press, 1998.

[27] I.I. Piatetski-Shapiro and S. Rallis, *Rankin triple L–functions*, Compositio Math. **64** (1987), 31–115.

[28] I. Satake, *Algebraic Structures of Symmetric Domains*, Publ. of the Math. Soc. of Japan, **14**, Iwanami Shoten and Princeton Univ. Press, 1980.

[29] F. Sato and Y. Hironaka, *Local densities of representations of quadratic forms over p–adic integers (the non-dyadic case)*, J. Number Theory **83** (2000), 106–136.

[30] G. Shimura, *Confluent hypergeometric functions on tube domains*, Math. Annalen **260** (1982), 269–302.

[31] H. Stamm, *On the reduction of the Hilbert-Blumenthal-moduli scheme with $\Gamma_0(p)$-level structure*, Forum Math. **9** (1997), 405–455.

[32] C. Soulé, D. Abramovich, J.-F. Burnol, and J. Kramer, Lectures on Arakelov Geometry, Cambridge Stud. Adv. Math., vol 33, Cambridge U. Press, 1992.

[33] W. J. Sweet, *The metaplectic case of the Weil–Siegel formula*, thesis, Univ. of Maryland, 1990.

[34] T. Yang, *An explicit formula for local densities of quadratic forms*, J. Number Theory **72** (1998), 309–356.

Stephen S. KUDLA

Department of Mathematics
University of Maryland
College Park, MD 20742
USA
E-mail : ssk@math.umd.edu

Séminaire BOURBAKI Juin 2000
52e année, 1999-2000, nO 877, p. 369 à 388

POLYNOMIAL FUNCTORS OVER FINITE FIELDS
[after Franjou, Friedlander, Henn, Lannes, Schwartz, Suslin]

by **Teimuraz PIRASHVILI**

INTRODUCTION

Let K be a field and let **Vect** be the category of vector spaces over K. Let $\mathcal{V}$ be the full subcategory of **Vect** consisting of finite dimensional vector spaces and let $\mathcal{F}$ be the category of all covariant functors $\mathcal{V} \to \mathbf{Vect}$. A functor $F : \mathcal{V} \to \mathcal{V}$ is called a *polynomial functor* if the maps $\mathrm{Hom}(V, W) \to \mathrm{Hom}(FV, FW)$ given by $f \mapsto F(f)$ are polynomial for all V and W. If these maps are homogeneous of degree d, then F is called *homogeneous of degree d*. Let us recall that a map $g : X \to Y$ between finite dimensional vector spaces is called *polynomial* if after choosing bases it is given by m polynomials with n variables, where $n = \dim X$ and $m = \dim Y$. Over finite fields there is an essential difference between polynomials and maps obtained by polynomials. This yields another version of polynomial functors – strict polynomial functors, which was recently proposed by Friedlander and Suslin [14]. There is a functor $\mathcal{P} \to \mathcal{F}$ from the category of strict polynomial functors to the category of functors, which is no longer an embedding. The category $\mathcal{F}$ is closely related to the category of unstable modules over the Steenrod algebra thanks to the remarkable result of Henn, Lannes and Schwartz [16], while the category $\mathcal{P}$ is related to the theory of polynomial representations of the general linear group.

The aim of this talk is to describe some important properties of the categories $\mathcal{F}$ and $\mathcal{P}$. We start with some general remarks on the category $\mathcal{F}$ and then we describe relationship between the category $\mathcal{F}$ and the category of unstable modules over the Steenrod algebra. In the section 2 we deal with the category $\mathcal{P}$ and its relations with the category of polynomial representation of GL_n. In the next section we discuss the theorem of Franjou-Friedlander-Scorichenko-Suslin [13] on the comparison between $\mathrm{Ext}_{\mathcal{P}}$ and $\mathrm{Ext}_{\mathcal{F}}$. In the last section we discuss the Betley-Pirashvili conjecture [2] involving Waldhausen's stable K-theory. Recently this conjecture was proved for finite fields by Betley [1] and Suslin (see Appendix of [13]) independently.

1. THE CATEGORY $\mathcal{F}$

1.1. General Properties

In what follows $\otimes_K$ and Hom_K are denoted by $\otimes$ and Hom respectively. For any vector space V one denotes the dual vector space by $V^\diamondsuit$.

The category $\mathcal{F}$ of all functors from the category $\mathcal{V}$ of finite dimensional vector spaces to the category of all vector spaces is an abelian category. For any $V \in \mathcal{V}$ one defines $P_V := K[\mathrm{Hom}(V,-)]$. Here $K[S]$ denotes the free vector space generated by a set S. For any $F \in \mathcal{F}$ one has a natural isomorphism $\mathrm{Hom}_{\mathcal{F}}(P_V, F) \cong F(V)$ thanks to the Yoneda lemma and therefore P_V is a projective object in $\mathcal{F}$.

For an object $F \in \mathcal{F}(K)$ one defines the *dual* DF of F by $V \mapsto (F(V^\diamondsuit))^\diamondsuit$. Let $J_W = DP_W$. Then $\mathrm{Hom}_{\mathcal{F}}(F, J_V) \cong (DF)(V)$ for all $F \in \mathcal{F}$. Hence J_W is injective. Moreover P_V and J_V, $V \in \mathcal{V}$, are respectively projective and injective generators of the category $\mathcal{F}$. Clearly $P_V \otimes P_W \cong P_{V\oplus W}$ and $J_V \otimes J_W \cong J_{V\oplus W}$.

Let us recall the definition of the degree of functors due to Eilenberg and Mac Lane [10]. Let $F \in \mathcal{F}$ be a functor with $F(0) = 0$. The *second cross-effect* of F is given by

$$F(X \mid Y) := ker(F(X \oplus Y) \longrightarrow F(X) \oplus F(Y)).$$

Clearly $F(X \oplus Y) \cong F(X) \oplus F(Y) \oplus F(X \mid Y)$. In order to define the third cross-effect of F, one can consider the second cross-effect $F(X \mid Y)$ as a functor on X and then take the second cross-effect of it. In this way one gets all higher cross-effects. One easily shows that the n-th cross-effect $F(X_1 \mid \cdots \mid X_n)$ of F is the same as the kernel of the natural homomorphism:

$$F(X_1 \oplus \cdots \oplus X_n) \to \bigoplus_{i=1}^{n} F(X_1 \oplus \cdots \oplus \hat{X}_i \oplus \cdots \oplus X_n).$$

A functor F is called of *degree* n if the $(n+1)$-st cross-effect of F vanishes and the n-th cross-effect is nonzero. In this case we write $deg(F) = n$. We let $\mathcal{F}_n$ be the category of all functors $F \in \mathcal{F}$ of degree $\leq n$. The examples of functors of degree n are functors T^n, S^n, Λ^n, Γ^n. Here $T^n(V)$ (resp. S^nV, Λ^nV, Γ^nV) is the n-th tensor (resp. symmetric, exterior, divided) power of V. Clearly $DS^n \cong \Gamma^n$, while $D\Lambda^n \cong \Lambda^n$ and $DT^n \cong T^n$.

A functor F is called *analytic* if it is an union of subfunctors of finite degrees. We let $\mathcal{F}_\omega$ be the category of all analytic functors. The functors $T^*(V) = \bigoplus_{d\geq 0} T^d(V)$, $\Lambda^*(V)$, $\Gamma^*(V)$ are examples of analytic functors.

If K is a finite field, then any map between finite dimensional vector spaces is polynomial. A consequence of this phenomenon is the fact that over such fields the functor J_V is analytic for all $V \in \mathcal{V}$. Indeed, since $J_V \cong J_K^{\otimes(\dim V)}$ it is sufficient to consider only J_K. Clearly $J_K(V) \cong Maps(V^\diamondsuit, K)$ and therefore it has a natural commutative algebra structure. Moreover if q is the number of elements in K, then the equality $a^q = a$ holds in $J_K(V)$. Hence the natural embedding $V \subset Maps(V^\diamondsuit, K)$

yields a homomorphism of algebras $S^*_q(V) \to J_K(V)$. It is not difficult to prove that this homomorphism is an isomorphism. Here $S^*_q(V)$ is the quotient of the symmetric algebra $S^*(V)$ by the ideal generated by $v^q - v, v \in V$. Since the quotient of an analytic functor is analytic, J_V is analytic too. Thus for finite fields the category $\mathcal{F}_\omega$ is an abelian category with sufficiently many injective objects.

1.2. The quotient category $\mathcal{F}_d/\mathcal{F}_{d-1}$

In this section we assume for simplicity that K is a prime field, although almost everything works for general rings (see [26]).

Let $F \in \mathcal{F}_n$. Since $F(X_1 \mid \cdots \mid X_n)$ is additive with respect to each variable we see that

$$F(X_1 \mid \cdots \mid X_n) \cong (cr_n F) \otimes (X_1 \otimes \cdots \otimes X_n),$$

where $cr_n F := F(K \mid \cdots \mid K) \subset F(\bigoplus_{i=1}^n K)$. The action of the symmetric group Σ_n on $\bigoplus_{i=1}^n K$ by permutation of summands yields an action of Σ_n on $cr_n F$. In this way one obtains an exact functor $cr_n : \mathcal{F}_n \to \Sigma_n$-*mod*. Clearly $cr_n F = 0$ if and only if $F \in \mathcal{F}_{n-1}$. For any $M \in \Sigma_n$-*mod*, one denotes by $M^\flat$ and $M^\sharp$ the functors given by $M^\flat(X) = (M \otimes X^{\otimes n})_{\Sigma_n}$ and $M^\sharp(X) := (M \otimes X^{\otimes n})^{\Sigma_n}$. Here for a Σ_n-module A, we let A_{Σ_n} and A^{Σ_n} be the coinvariants and invariants under the action of Σ_n. These functors are related by the *norm* homomorphism $N : M^\flat \to M^\sharp$, which is induced by the action of $\sum_{\sigma \in \Sigma_n} \sigma$.

PROPOSITION 1.1 ([26]). — *The functors $M \mapsto M^\flat$ and $M \mapsto M^\sharp$ define left and right adjoints to the functor $cr_n : \mathcal{F}_n \to \Sigma_n$-mod. Moreover $cr_n M^\flat \cong M \cong cr_n M^\sharp$. Thus the kernels and cokernels of the natural maps $(cr_n F)^\flat \to F$, $F \to (cr_n F)^\sharp$ are functors of degree $< n$. For any $F \in \mathcal{F}_n$ the composition $(cr_n F)^\flat \to F \to (cr_n F)^\sharp$ coincides with the norm homomorphism.*

Since the norm homomorphism is an isomorphism over the rationals it follows that in this case for any functor F of degree d one has an isomorphism $F \cong cr_d(F)^\flat \oplus F'$, where $deg(F') \leq d-1$. Using this fact it is not difficult to prove that any functor of finite degree has a unique decomposition $\bigoplus_d F_d$, where F_d is of the form $M^\flat$ and M is a representation of Σ_d. Moreover any natural transformation between functors F_d and F_l for $d \neq l$ vanishes. Hence over the rationals one has an equivalence of categories $\mathcal{F}_n \cong \prod_{i=1}^n (\Sigma_i\text{-}mod)$ (see also Appendix A of [24]).

For $char(K) > 0$ we only have an equivalence of categories $\mathcal{F}_n/\mathcal{F}_{n-1} \cong \Sigma_n$-*mod*.

If M is a simple representation of Σ_n then the image of the norm homomorphism $M^\flat \to M^\sharp$ is a simple object in $\mathcal{F}$ and any simple object of $\mathcal{F}$ is of this form. This is a formal consequence of Proposition 1.1 (compare with Proposition 4.7 of [21]). Recently Piriou and Schwartz [30] proved that for any simple functor $F \in \mathcal{F}$ one has $\mathrm{Ext}^1_{\mathcal{F}}(F, F) = 0$, unlike to the situation with representations of symmetric groups.

1.3. Unstable modules over the Steenrod algebra and $\mathcal{F}$

In the 90's Henn, Lannes and Schwartz [16] discovered a remarkable link between the category $\mathcal{F}(\mathbf{F}_p)$ and representations of the Steenrod algebra. For simplicity we consider only the case $K = \mathbf{F}_2$.

We let $\mathcal{A}$ denote the mod-2 Steenrod algebra. Let us recall that it is generated by the elements Sq^i of degree i, $i > 0$ and these generators satisfy the Adem relations (see [S1]). A (graded) module $M^* = \bigoplus_{i \geq 0} M^i$ over $\mathcal{A}$ is said to be *unstable* if for every $x \in M$ one has $Sq^i x = 0$ if $i > |x|$. Here $|x|$ denotes the degree of x. Note that for any space X, the mod-2 cohomology $H^*(X)$ is an example of an unstable $\mathcal{A}$-module. Let $\mathcal{U}$ be the category of unstable $\mathcal{A}$-modules. One observes that for any $n \geq 0$ there exists a unique (up to an isomorphism) unstable $\mathcal{A}$-module $F(n)$ such that for any $M^* \in \mathcal{U}$ one has a natural isomorphism

$$\mathrm{Hom}_{\mathcal{U}}(F(n), M^*) \cong M^n.$$

An unstable A-module M^* is called *nilpotent* if for any $x \in M$ there exists $k > 0$ such that $(Sq_0)^k x = 0$. Here $Sq_0 : M^n \to M^{2n}$ is given by $m \mapsto Sq^n m$ and $(Sq_0)^k$ denotes the k-th iteration of Sq_0. Let us notice that $H^*(X)$ is nilpotent as an unstable module if and only if $H^*(X)$ is nilpotent as a commutative algebra.

If V is a finite dimensional vector space, then by H_*V and H^*V we denote the homology and cohomology of V considered as an abelian group. Let us recall that one has the following isomorphisms of functors

$$H^*(V) \cong S^*(V^{\diamondsuit}[1]), \ H_*(V) \cong \Gamma^*(V[1]).$$

Here $V[1]$ indicates the fact that V is concentrated in degree 1. Since H^*V is an unstable module, any $\theta \in \mathcal{A}$ yields a morphism $S^n \to S^{n+|\theta|}$ and therefore it induces by duality a natural transformation $\Gamma^{n+|\theta|} \to \Gamma^n$. Hence for any $F \in \mathcal{F}$ the graded vector space $\mathbf{m}(F) = \bigoplus_n \mathrm{Hom}_{\mathcal{F}}(\Gamma^n, F)$ has a natural $\mathcal{A}$-module structure and $\mathbf{m}(F) \in \mathcal{U}$. In this way one obtains a functor $\mathbf{m} : \mathcal{F} \to \mathcal{U}$. It follows from the very definition that

$$\mathbf{m}J_V = \bigoplus_n \mathrm{Hom}_{\mathcal{F}}(\Gamma^n, J_V) \cong \bigoplus_n (\Gamma^n V)^{\diamondsuit} = S^*V.$$

The functor $\mathbf{m}$ has a left adjoint $\mathbf{f} : \mathcal{U} \to \mathcal{F}$. So one has an isomorphism

$$\mathrm{Hom}_{\mathcal{F}}(\mathbf{f}M^*, F) \cong \mathrm{Hom}_{\mathcal{U}}(M^*, \mathbf{m}F).$$

Putting $F = J_V$ one obtains $(\mathbf{f}(M^*)(V))^{\diamondsuit} \cong \mathrm{Hom}_{\mathcal{U}}(M^*, S^*V)$. One observes that if M^* is a finitely generated $\mathcal{A}$-module then $\mathrm{Hom}_{\mathcal{U}}(M^*, S^*V)$ is a finite dimensional vector space. Hence in this case $\mathbf{f}(M^*)(V)$ is isomorphic to $(\mathrm{Hom}_{\mathcal{U}}(M^*, S^*V))^{\diamondsuit}$. In general M^* is a union of finitely generated $\mathcal{A}$-modules and therefore $\mathrm{Hom}_{\mathcal{U}}(M^*, S^*V)$ has a structure of a profinite vector space, whose continuous dual is denoted by $\mathrm{Hom}_{\mathcal{U}}(M^*, S^*V)'$. Since $\mathbf{f}$ preserves colimits one has the following isomorphism

$$\mathbf{f}(M^*)(V) \cong \mathrm{Hom}_{\mathcal{U}}(M^*, S^*V)'.$$

For example, $\mathbf{f}(F(n)) = \Gamma^n$.

Theorem 1.2 ([16]). — *The functor* $\mathbf{f}$ *is exact and preserves tensor products. For any* $V \in \mathcal{V}$ *one has a natural isomorphism* $\mathbf{f}(S^*V) \cong J_V$. *Furthermore the values of* $\mathbf{f}$ *are analytic functors and for any analytic* F *the natural map* $\mathbf{f}(\mathbf{m}(F)) \to F$ *is an isomorphism. An object* $M^* \in \mathcal{U}$ *is nilpotent if and only if* $\mathbf{f}(M^*) = 0$.

Let $\mathcal{N}$ be the category of nilpotent unstable modules. Obviously it is a Serre subcategory in $\mathcal{U}$. Let us note that the theorem implies the equivalence of categories $\mathcal{U}/\mathcal{N} \cong \mathcal{F}_\omega$.

The original proof of this theorem used deep structural results on injective objects in the category of unstable modules due to Carlsson, Miller, Lannes, Zarati and others. Kuhn [20] gave a reasonably simple proof of Theorem 1.2. His proof can be divided in several steps. The first one is the following important result.

Theorem 1.3 ([20]). — *Let* $F \in \mathcal{F}$ *be a finite functor. Then one can embed* F *in a sum of the form* $\bigoplus_{i=1}^k S^{n_i}$.

Here a functor $T \in \mathcal{F}$ is called *finite* if it is of finite degree and has values in $\mathcal{V}$. The proof of Theorem 1.3 consists of several reductions. One observes that 2nd power map and multiplication in the symmetric algebra yield a natural transformation: $S^i \otimes S^j \to S^{i2^r+j}$, which is a monomorphism for $2^r > j$. Hence the class of functors for which Theorem 1.3 is true is closed with respect to tensor products. One observes also that F embeds in a finite sum of J_V, because $J_V, V \in \mathcal{V}$ are injective cogenerators. Since $J_V = J_K^{\otimes(\dim V)}$ and F has finite degree it suffices to consider the case when F is the image of S^m into $J_K, m \geq 1$ under the projection $S^*(V) \to S_2^*(V)$, because any sub-functor of finite degree of J_K is of such type. Using a combinatorial argument Kuhn was able to give an explicit embedding in this case (see [20]).

The second step in the proof of Theorem 1.2 is the following

Proposition 1.4. — *The full subcategory of* $\mathcal{F}$, *whose objects are* Γ^n, $n \geq 0$, *and the full subcategory of* $\mathcal{U}$, *whose objects are* $F(n)$, $n \geq 0$ *are isomorphic categories. The isomorphism takes* Γ^n *to* $F(n)$.

Sketch of the proof. — Let $a \in \mathcal{A}$ be a homogeneous element of degree m. It yields a natural transformation $a^* : S^n \to S^{n+m}$, thanks to the isomorphism $S^*(V[1]) \cong H^*(V^\diamondsuit)$. It is not too difficult to show that in this way we get a homomorphism

$$\mathrm{Hom}_{\mathcal{A}}(F(n+m), F(n)) \to \mathrm{Hom}_{\mathcal{F}}(S^n, S^{n+m})$$

and counting argument shows that this map is an isomorphism. □

The next step is the following variant of the Gabriel-Popescu theorem on abelian categories. Let us recall that a *ringoid* is a small category with abelian group structure on Hom-sets, such that the composition is biadditive. Clearly a ringoid with one object is the same ring considered as a category with one object. Moreover any small subcategory of an additive category is a ringoid. A *right module* over a ringoid $\mathbf{C}$ is a contravariant functor $M : \mathbf{C} \to Ab$ with the property that the maps $\mathrm{Hom}_{\mathbf{C}}(c, d) \to \mathrm{Hom}(M(d), M(c))$ given by $f \mapsto M(f)$ are homomorphisms of abelian groups, $c, d \in \mathbf{C}$. For a ringoid $\mathbf{C}$ we let mod-$\mathbf{C}$ be the category of all right $\mathbf{C}$-modules.

THEOREM 1.5 ([20]). — *Let* $\mathbf{A}$ *be an abelian category with exact directed colimits and let* $\mathbf{C} \subset \mathbf{A}$ *be a small subcategory, such that any object in* $\mathbf{A}$ *is a quotient of an object of the form* $\bigoplus c_i$*, where* $c_i \in \mathbf{C}$*. Let* $r : \mathbf{A} \to mod\text{-}\mathbf{C}$ *be the functor given by* $r(a)(c) = Hom_{\mathbf{A}}(c, a)$*. Then*

i) *The functor* r *has a left adjoint* $l : mod\text{-}\mathbf{C} \to \mathbf{A}$*.*

ii) *The functor* l *is exact,* r *is fully faithful and* $lra \to a$ *is an isomorphism for any* $a \in \mathbf{A}$*. Hence in this case* $mod\text{-}\mathbf{C}/Ker(l) \cong \mathbf{A}$*.*

iii) $\mathbf{A}$ *has enough injectives and* r *preserves injectives. Moreover if* I *and* J *are injective in* $\mathbf{A}$ *then* r *yields an isomorphism*

$$\mathrm{Hom}_{\mathbf{A}}(I, J) \cong \mathrm{Hom}_{mod-\mathbf{C}}(rI, rJ).$$

If additionally all objects from $\mathbf{C}$ *are small and projective in* $\mathbf{A}$ *then* r *(and hence* l*) is an equivalence of categories.*

Now we are in the position to give a sketch of the proof of Theorem 1.2. First one takes $\mathbf{A} = \mathcal{U}$ and $\mathbf{C} = \{F(n)\}_{n\geq 0}$. Here $\{F(n)\}_{n\geq 0}$ is the full subcategory of $\mathcal{U}$ whose objects are $F(n)$, $n \geq 0$. The last statement of Theorem 1.5 shows that $\mathcal{U} \cong mod\text{-}\{F(n)\}_{n\geq 0}$. Then we put $\mathbf{A} = \mathcal{F}_\omega$ and take $\mathbf{C} = \{\Gamma^n\}_{n\geq 0}$, where $\{\Gamma^n\}_{n\geq 0}$ is the full subcategory of $\mathcal{F}_\omega$ whose objects are $\Gamma^n, n \geq 0$. By Proposition 1.4 one has an equivalence of categories $mod\text{-}\mathbf{C} \cong \mathcal{U}$. It follows from Theorem 1.3 that the condition of Theorem 1.5 holds. Applying Theorem 1.5 we obtain an adjoint pair of functors (l, r) between the category $\mathcal{U}$ and $\mathcal{F}_\omega$. One checks readily that $r = \mathbf{m}$ and we get all statements of Theorem 1.2 except the statement about the tensor product and the characterization of nilpotent modules. We refer to [20] and [34] for the proof of these facts.

Remark 1.6. — By Proposition 1.4 one can reconstruct the category $\mathcal{U}$ and the Steenrod algebra (without the Bokstein operations) from the endomorphism ring $End_{\mathcal{F}}(S^*, S^*)$, which has a meaning in a far more general setting. Based on this observation an appropriate notion of "Steenrod algebra over a finite field K" was developed in [20] in such a way that the whole material in 1.3 works for any finite field.

Remark 1.7. — Among other applications of Theorem 1.2 we would like to mention the important article of Henn-Lannes-Schwartz [17] and also the work of Kuhn [23] and Schwartz [35] on topological realizability of unstable modules over the Steenrod algebra.

1.4. Artinian conjecture

The following conjecture was posed by L. Schwartz.

If K is a finite field then the category $\mathcal{F}$ is locally noetherian.

Here is another formulation:

If K is a finite field and $V \in \mathcal{V}$ then the injective object J_V is artinian.

It is known that the conjecture is true when $\dim V = 1$. As was proved by Powell [31] the conjecture is also true when $K = \mathbf{F}_2$ and $\dim V = 2$.

The conjecture implies the following assertions

Any injective object in $\mathcal{F}_\omega$ is also injective in $\mathcal{F}$ and the functor $\mathbf{f} : \mathcal{U} \to \mathcal{F}$ respects injective objects.

2. THE CATEGORY $\mathcal{P}$

In this section K is a field. We give several equivalent interpretations of the category of *strict polynomial functors*. We start with recalling the definition of Schur algebras.

2.1. Polynomial representations of GL_n and Schur algebras.

Let $GL_n = GL_{n,K}$ be the general linear group over K considered as an algebraic group. We recall that the coordinate ring $K[GL_n]$ is the quotient of the polynomial algebra $K[x_{ij}, y]_{1\leq i,j\leq n}$ by the ideal generated by $(y\, det(x_{ij}) - 1)$. It has a Hopf algebra structure, whose comultiplication is given by $\Delta(x_{ij}) = \sum_{k=1}^{n} x_{ik} \otimes x_{kj}$. Moreover we let M_n be the algebraic monoid of $n \times n$-matrices. The coordinate ring $K[M_n]$ is the polynomial algebra $K[x_{ij}]_{1\leq i,j\leq n}$. Clearly it is a subbialgebra of $K[GL_n]$. Comodules over $K[GL_n]$ are called *rational representations* of Gl_n. A rational representation of GL_n is called *polynomial* if it is also a comodule over $K[M_n]$. One observes that for each $d \geq 0$ the subspace $A(n,d) \subset k[M_n]$ of homogeneous polynomials of degree d is a subcoalgebra of $k[M_n]$. The linear dual $A(n,d)^{\diamondsuit}$ of $A(n,d)$ is known as *Schur algebra* and it is denoted by $S(n,d)$ (see [15]). A polynomial representation of GL_n is said to be *homogeneous of degree d* if it is a comodule over $A(n,d)$. Clearly the category of finite dimensional homogeneous polynomial representations of degree d is equivalent to the category of finite dimensional modules over the Schur algebra $S(n,d)$. We refer the reader to [25] and references there for extensive information on Schur algebras.

2.2. Polynomial laws and the category $\mathcal{P}$

For a vector space V we let $\underline{V}$ be the functor from the category $-$ K/alg of commutative K-algebras to the category of sets given by $R \mapsto R \otimes V$. A *polynomial law* from a vector space V to a vector space W is by definition a natural transformation of functors $\underline{V} \to \underline{W}$ (see [32]). Let Pol be the category of vector spaces and polynomial laws. A polynomial law $f \in \mathrm{Hom}_{Pol}(V, W)$ is called *homogeneous of degree* d if for any algebra R, $r \in R$ and $x \in R \otimes V$ one has $f_R(rx) = r^d f_R(x)$.

Following Friedlander and Suslin [14] *a strict polynomial functor* $F : \mathcal{V} \to \mathcal{V}$ is a rule which associates to each finite dimensional vector space V a vector space $F(V)$ and to each pair (V, W) of finite dimensional vector spaces a polynomial law

$$F(V, W) \in \mathrm{Hom}_{Pol}(\mathrm{Hom}(V, W), \mathrm{Hom}(F(V), F(W))$$

in such a way that for any $V \in \mathcal{V}$ one has $F(V, V)(1_V) = 1_{F(V,V)}$ and for a any $U, V, W \in \mathcal{V}$ the following diagram of polynomial laws commutes:

$$\begin{array}{ccc}
\mathrm{Hom}(V, W) \times \mathrm{Hom}(U, V) & \to & \mathrm{Hom}(U, W) \\
\downarrow & & \downarrow \\
\mathrm{Hom}(FV, FW) \times \mathrm{Hom}(FU, FV) & \to & \mathrm{Hom}(FU, FW),
\end{array}$$

where the vertical arrows are induced by F, while the horizontal arrows come from the composition.

One says that a strict polynomial functor F is of finite degree if the degrees of the polynomial laws $F(X, Y)$ are bounded above. A strict polyno mial functor F is called homogeneous of degree n if for any $V, W \in \mathcal{V}$ the polynomial $F(V, W)$ is homogeneous of degree n. We let $\mathcal{P}$ (resp. $\mathcal{P}_d$) be the category of strict polynomial functors of finite degree (resp. strict homogeneous polynomial functors of degree d). Since any polynomial law is a sum of homogeneous ones, it follows that $\mathcal{P} \cong \bigoplus_d \mathcal{P}_d$.

2.3. The category $\Gamma^d\mathcal{V}$

The vector space S^dV is the module of coinvariants of $V^{\otimes d}$ under the action of the symmetric group Σ_d on d symbols. Similarly Γ^dV is the module of invariants $\Gamma^d(V) = (V^{\otimes d})^{\Sigma_d}$. For any vector space V one has a map $\gamma^d : V \to \Gamma^d(V)$, which is given by $x \mapsto \gamma^d(x) := x^{\otimes n} \in (V^{\otimes n})^{\Sigma_n}$. Moreover for any vector spaces V and W there exists a unique linear map

$$\mu : \Gamma^dV \otimes \Gamma^dW \to \Gamma^d(V \otimes W)$$

with the property $\mu(\gamma^d(x) \otimes \gamma^d(y)) = \gamma^d(x \otimes y)$. One can use the transformation μ to define the composition in the category $\Gamma^d\mathcal{V}$, whose objects are those of $\mathcal{V}$ while morphisms are $\mathrm{Hom}_{\Gamma^d\mathcal{V}}(V, W) = \Gamma^d(\mathrm{Hom}(V, W))$. The identity arrows in $\Gamma^d\mathcal{V}$ are $\gamma^d(1_V)$. Clearly $\Gamma^d\mathcal{V}$ is a K-linear category, that is a category whose set of morphisms between two objects has a vector space structure and the composition is bilinear.

PROPOSITION 2.1. — *The following categories are equivalent:*

- *the category of K-linear functors $\Gamma^d\mathcal{V} \to \mathcal{V}$,*
- *the category of homogeneous strict polynomial functors of degree d,*
- *the category of homogeneous polynomial representations of degree d of GL_n, provided $n \geq d$,*
- *the category of finite dimensional modules over $S(n, d)$, provided $n \geq d$.*

Proof. — Clearly $\underline{V}$ is isomorphic to the functor $\mathrm{Hom}_{K/alg}(S^*(V^\diamondsuit), -)$ as soon as $V \in \mathcal{V}$. Hence by the Yoneda lemma one has $\mathrm{Hom}_{Pol}(V, W) \cong S^*(V^\diamondsuit) \otimes W \cong \bigoplus_d \mathrm{Hom}(\Gamma^d V, W)$. Elements of $\mathrm{Hom}(\Gamma^d V, W)$ correspond to homogeneous polynomial laws of degree d. This shows that a homogeneous strict polynomial functor of degree d is the same as a collection of $F(V) \in \mathcal{V}, V \in \mathcal{V}$ together with linear maps $\Gamma^d \mathrm{Hom}(V, W) \otimes F(V) \to F(W)$ satisfying associativity and unity conditions. Therefore the first two categories are equivalent. The last two categories are also equivalent by the discussion in 2.1. It remains to show the equivalence between the first and the third category. For each $m \geq 0$ we let $\Gamma^{d,m} \in \mathcal{P}_d$ be the functor given by $V \mapsto \Gamma^d(\mathrm{Hom}(K^m, V))$. By the Yoneda lemma one has an isomorphism $\mathrm{Hom}_{\mathcal{P}}(\Gamma^{d,m}, F) \cong F(K^m)$ for any $F \in \mathcal{P}_d$. So $\Gamma^{d,m},\ m \geq 0$ are small projective generators in the abelian category $\mathcal{P}_d$. By looking at cross-effects of $\Gamma^{d,m}$ it is clear that each $\Gamma^{d,m}$ is a direct summand of $\Gamma^{d,n}$ provided $m \leq n$ and each $\Gamma^{d,m}$ is a direct summand of an object of the form $\bigoplus_{k=1}^d \Gamma^{d,n}$ provided $n \geq d$. Therefore $\Gamma^{d,n}$ is a projective generator. Thus the category of K-linear functors $\Gamma^d \mathcal{V} \to \mathcal{V}$ is equivalent to the category of finite dimensional modules over $End_{\mathcal{P}}(\Gamma^{d,n}) = \Gamma^d(End(K^n))$. But the last algebra is isomorphic to the Schur algebra $S(n, d)$ and hence the result. $\square$

2.4. Elementary properties of $\mathcal{P}$

Clearly $f \mapsto \gamma^d(f)$ defines the (nonlinear) functor $\gamma^d : \mathcal{V} \to \Gamma^d \mathcal{V}$. The precomposition with γ^d yields the functor $\mathcal{P}_d \to \mathcal{F}_d$, which is a full embedding provided the field K contains at least d elements. By abuse of notation we denote the image of F under this functor by the same letter F.

The natural transformation $\Gamma^{d+l} \to \Gamma^d \otimes \Gamma^l$ coming from the Hopf algebra structure on the divided power algebra can be used to define the tensor product of strict homogeneous polynomial functors. It yields the functor $\mathcal{P}_d \times \mathcal{P}_l \to \mathcal{P}_{k+l}$, which corresponds to the obvious tensor functor in $\mathcal{F}$.

The dual of the natural transformation $S^d(S^l V) \to S^{dl}(V)$ is a transformation $\Gamma^{dl} \to \Gamma^d \circ \Gamma^l$, which can be used to define the composition of strict homogeneous polynomial functors. One observes also that the dual of a strict homogeneous polynomial functor is a strict homogeneous polynomial functor. Since Γ^d carries obvious structure of a strict homogeneous polynomial functor of degree d, we see that $T^d, S^d, \Lambda^d \in \mathcal{P}_d$. As was observed in the proof of Proposition 2.1 the functors $\Gamma^{d,m}$ are projective generators of $\mathcal{P}_d$. Therefore $S^{d,m} := D\Gamma^{d,m}$ are injective cogenerators of $\mathcal{P}_d, m \geq 0$. Another system of projective generators is $\Gamma^{d_1} \otimes \cdots \otimes \Gamma^{d_k}$, $d_1 + \cdots + d_k = d$.

2.5. Frobenius twist

In representation theory of algebraic groups over fields of characteristic $p > 0$ the Frobenius twist plays an important rôle (see [18]). By Proposition 2.1 to the Frobenius twist on polynomial representations corresponds a similar operation on strict polynomial functors, which can be described as follows. Let K be a field of characteristic $p > 0$. For a vector space V we let $V^{(1)}$ be the vector space obtained by extending scalars via the Frobenius homomorphism $f : K \to K$ given by $f(\lambda) = \lambda^p$. In other words $V^{(1)}$ is a vector space generated by $v^{(1)}$, $v \in V$ modulo the following relations

$$(v+w)^{(1)} = v^{(1)} + w^{(1)}, \ \ (\lambda v)^{(1)} = \lambda^p v^{(1)}, \ \lambda \in K.$$

One defines $V^{(r)}, r \geq 1$ by induction: $V^{(r+1)} = (V^{(r)})^{(1)}$.

The map $V^{(1)} \to S^p V$ given by $v^{(1)} \mapsto v^p$ is linear. By duality one obtains the natural transformation $\Gamma^p \to I^{(1)}$, where $I^{(1)}(V) = V^{(1)}$. Using this transformation one readily checks that $I^{(1)} \in \mathcal{P}_p$. For any $F \in \mathcal{P}_d$ we put $F^{(1)} = F \circ I^{(1)}$ and then by induction one defines $F^{(r+1)} = (F^{(r)})^{(1)}, r \geq 0$. Clearly $F^{(r)} \in \mathcal{P}_{dp^r}$ if $F \in \mathcal{P}_d$.

Let us note that for finite prime fields the functors $I^{(1)}$ and I are isomorphic in $\mathcal{F}$. Here $I(V) = V$. However in $\mathcal{P}$ these functors are different, they even have different degrees: $I \in \mathcal{P}_1$, while $I^{(1)} \in \mathcal{P}_p$. For finite fields the Frobenius twist induces an equivalence of categories $(-)^{(1)} : \mathcal{F}_d \to \mathcal{F}_d$.

2.6. $\mathrm{Ext}_{\mathcal{P}}$ and cohomology of algebraic groups

The following theorem was proved by Donkin (see Section 3 of [14]).

THEOREM 2.2 ([6, 7]). — *Let A and B be finite dimensional polynomial representations. Then*

$$\mathrm{Ext}^*_{K[M_n]}(A,B) \to \mathrm{Ext}^*_{K[GL_n]}(A,B)$$

is an isomorphism. Here Ext *is taken in the category of comodules.*

By Proposition 2.1 we also have

COROLLARY 2.3. — *Let $F, T \in \mathcal{P}_d$. Then*

$$\mathrm{Ext}^*_{\mathcal{P}}(F,T) \to \mathrm{Ext}^*_{K[GL_n]}(F(K^n), T(K^n))$$

is an isomorphism provided $n \geq d$.

3. FUNCTOR COHOMOLOGY

In this section we deal with Ext-groups in the categories $\mathcal{F}$ and $\mathcal{P}$. Calculations in $\mathcal{P}$ are easier, because each piece $\mathcal{P}_d$ of $\mathcal{P}$ is of finite global dimension thanks to a result of Donkin [6], who proved this result in terms of Schur algebras (see also Totaro [36] for an explicit formula for the global dimension of $\mathcal{P}_d$). Calculation in $\mathcal{P}$ is easier also because the functors $S^{n_1} \otimes \cdots \otimes S^{n_k}$ (resp. $\Gamma^{n_1} \otimes \cdots \otimes \Gamma^{n_k}$) are injective (resp. projective) in $\mathcal{P}$. Of course S^n is no longer injective in $\mathcal{F}$, but the following result shows that if one inverts the Frobenius $S^n \to S^{np}$ then one gets an injective object in $\mathcal{F}_\omega$.

Lemma 3.1 ([22]). — *Assume $F \in \mathcal{F}$ has a projective resolution of finite type. Then, for any $k > 0$, one has*

$$colim \ \mathrm{Ext}^k_{\mathcal{F}}(T, S^{p^n}) = 0,$$

where S^n denotes the n-th symmetric power and the limit is considered with respect to the Frobenius maps $\Phi : S^{p^n} \to S^{p^{n+1}}$.

Proof. — (see [12]). Let P_* be a projective resolution of F which is of finite type. Then DP_* is an injective resolution of DF. Hence

$$\mathrm{Ext}^k_{\mathcal{F}}(F, S^{2^n}) = H^k(Hom_{\mathcal{F}}(P_*, S^{2^n})) \cong H^k((\mathbf{m}(DP_*))^{2^n}) \cong (H^k(\mathbf{m}(DP_*)))^{2^n}$$

and it is enough to show that $H^k(\mathbf{m}(DP_*)) \in \mathcal{N}$. Since

$$\mathbf{f}(H^k(\mathbf{m}(DP_*))) = H^k(\mathbf{f}(\mathbf{m}(DP_*))) = H^k(DP_*) = 0, \ k > 0$$

Theorem 1.2 implies the expected inclusion. □

Lemma 3.1 together with Lemma 3.3 plays a crucial rôle in this section.

Let $\mathbf{C}$ and $\mathbf{D}$ be categories and let $l : \mathbf{C} \to \mathbf{D}$ and $r : \mathbf{D} \to \mathbf{C}$ be functors. If l is a left adjoint to r we will say that (l, r) is an adjoint pair from $\mathbf{C}$ to $\mathbf{D}$. Moreover for a small category $\mathbf{C}$ we let $\mathbf{C}$-*mod* be the category of all functors $\mathbf{C} \to \mathbf{Vect}$.

Lemma 3.2. — *Let (l, r) be an adjoint pair from $\mathbf{C}$ to $\mathbf{D}$. Assume $\mathbf{C}$ and $\mathbf{D}$ are small categories. Then for any $F : \mathbf{C} \to \mathbf{Vect}$ and $G : \mathbf{D} \to \mathbf{Vect}$ one has an isomorphism*

$$\mathrm{Ext}^*_{\mathbf{D}-mod}(F \circ r, G) \cong \mathrm{Ext}^*_{\mathbf{C}-mod}(F, G \circ l).$$

Lemma 3.3 ([27]). — *Let $A \in \mathcal{F}$ be an additive functor and let $B : \mathcal{V} \times \mathcal{V} \to \mathbf{Vect}$ be a bifunctor with the property $B(X, 0) = 0 = B(0, X)$ for all $X \in \mathcal{V}$. Then $\mathrm{Ext}^*_{\mathcal{F}}(A, B^d) = 0 = \mathrm{Ext}^*_{\mathcal{F}}(B^d, A)$, where $B^d(V) = B(V, V)$.*

Proof. — Apply the previous lemma to the adjoint pairs $(i_k, p_k) : \mathcal{V} \to \mathcal{V} \times \mathcal{V}$, $k = 1, 2, 3$, where $i_1(V) = (V, 0), i_2(V) = (0, V), i_3(V) = (V, V), p_k(V_1, V_2) = V_k, k = 1, 2, p_3(V_1, V_2) = V_1 \bigoplus V_2$ and use the fact that $A \circ p_3 = A \circ p_1 \bigoplus A \circ p_2$. □

A strict polynomial functor T is called additive if T is additive as an object in $\mathcal{F}$. The previous argument shows that Lemma 3.3 is still true in the framework of $\mathcal{P}$. Actually the proof of Lemma 3.3 gives a bit more:

$$\mathrm{Ext}^*_{\mathcal{E}}(K, T_1 \otimes \cdots \otimes T_n) = \ 0 = \ \mathrm{Ext}^*_{\mathcal{E}}(T_1 \otimes \cdots \otimes T_n, K),$$

where $degK < n$, $T_1(0) = \cdots = T_n(0) = 0$ and $\mathcal{E}$ is $\mathcal{P}$ or $\mathcal{F}$.

Let us see how to use these facts for explicit calculations. First we consider calculations in the category $\mathcal{P}$. Since $I \in \mathcal{P}$ is projective we have $\mathrm{Ext}^i_{\mathcal{P}}(I, -) = 0$ if $i > 0$. More interesting is $\mathrm{Ext}^*_{\mathcal{P}}(I^{(r)}, -)$ for $r \geq 1$.

For simplicity we restrict ourselves to the case $p = 2$ and refer the reader to [12], [14] and [13] for the case $p > 2$. We follow ideas of [12].

It is not too hard to show that one has an exact sequence

$$0 \to S^{2^h(1)} \to S^{2^h} \to \cdots \to S^{2^h-i} \otimes S^i \to \cdots \to S^{2^h} \to 0. \tag{3.1}$$

Here the first nontrivial map is given by the Frobenius, and the rest are the compositions $S^k \otimes S^l \to S^{k-1} \otimes S^1 \otimes S^l \to S^{k-1} \otimes S^{l+1}$, where first map (resp. second) is induced by the comultiplication (resp. multiplication) in the Hopf algebra $S^*(V)$. By the $\mathcal{P}$-version of Lemma 3.3 $\mathrm{Ext}^*_{\mathcal{P}}(I^{(r)}, S^k \otimes S^l) = 0$ if $k, l \geq 1$ and a fortiori $\mathrm{Hom}^*_{\mathcal{P}}(I^{(r)}, S^k \otimes S^l) = 0$. Let us note that $\mathrm{Hom}_{\mathcal{P}}(I^{(h)}, S^{2^h}) = \mathrm{Hom}_{\mathcal{F}}(I^{(h)}, S^{2^h})$ is one dimensional spanned by the iterated Frobenius map. Hence it follows from the injective resolution (3.1) that $\mathrm{Ext}^i_{\mathcal{P}}(I^{(h+1)}, S^{2^h(1)}) = K$, if $i = 0$ and $i = 2^{h+1}$ and $\mathrm{Ext}^i_{\mathcal{P}}(I^{(h+1)}, S^{2^h(1)}) = 0$ otherwise. Let us apply the degreewise action by the Frobenious twist on (3.1). Then one obtains a non-injective resolution of $S^{2^h(2)}$, which gives rise to a hypercohomology spectral sequence

$$E^1_{st} = \mathrm{Ext}^t_{\mathcal{P}}(I^{(h+2)}, S^{2^{h+1}-s(1)} \otimes S^{s(1)}) \Longrightarrow \mathrm{Ext}^{s+t}_{\mathcal{P}}(I^{(h+2)}, S^{2^h(2)}).$$

By the $\mathcal{P}$-version of Lemma 3.3 the spectral sequence has only two nontrivial columns corresponding to $s = 0$ and $s = 2^{h+1}$. Moreover the previous calculation shows that in both columns there are only two nontrivial terms corresponding to $t = 0$ and $t = 2^{h+2}$, so there is no space for differentials and we get $\mathrm{Ext}^i_{\mathcal{P}}(I^{(h+2)}, S^{2^h(2)}) = K$, if $i = 0, i = 2^{h+1}, 2^{h+2}, 2^{h+2} + 2^{h+1}$ and $\mathrm{Ext}^i_{\mathcal{P}}(I^{(h+2)}, S^{2^h(2)}) = 0$ otherwise. By iteration one obtains:

Proposition 3.4 ([14]). — *For any* $0 \leq j \leq r$ *one has* $\mathrm{Ext}^q_{\mathcal{P}}(I^{(r)}, S^{2^{r-j}(j)}) = K$ *if* $q \equiv 0 \bmod 2^{r-j+1}$ *and* $q < 2^{r+1}$ *and* $\mathrm{Ext}^q_{\mathcal{P}}(I^{(r)}, S^{2^{r-j}(j)}) = 0$ *otherwise.*

Such type of results was used in [14] to prove the following theorem:

The rational cohomology of any finite group scheme is a finitely generated algebra.

Now we do similar calculations in $\mathcal{F}$.

Theorem 3.5 ([12]). — $\mathrm{Ext}^i_{\mathcal{F}}(I^{(h)}, S^{2^h}) = K$ *if* $i \equiv$ *mod* 2^{h+1} *and* $= 0$ *otherwise.*

Proof. — The exact sequence (3.1) together with Lemma 3.3 yields the exact sequence

$$\cdots \to \mathrm{Ext}^{k-2^h}_{\mathcal{F}}(I^{(h)}, S^{2^h}) \to$$

$$\to \mathrm{Ext}^k_{\mathcal{F}}(I^{(h)}, S^{2^{h-1}}) \to \mathrm{Ext}^k_{\mathcal{F}}(I^{(h)}, S^{2^h}) \to \mathrm{Ext}^{k-2^h+1}_{\mathcal{F}}(I^{(h)}, S^{2^h}) \to \cdots$$

From this sequence it follows that $\mathrm{Ext}^k_{\mathcal{F}}(I^{(h)}, S^{2^{h-1}}) \to \mathrm{Ext}^k_{\mathcal{F}}(I^{(h)}, S^{2^h})$ is an isomorphism for $k < 2^h$. It is known that any finite functor admits a projective resolution of finite type (see [12]). Thus by Lemma 3.1 one has $\mathrm{Ext}^k_{\mathcal{F}}(I^{(h)}, S^{2^h}) = 0$ as soon as $0 < k < 2^{h+1}$. Now one can use this information in the same exact sequence to get $\mathrm{Ext}^k_{\mathcal{F}}(I^{(h)}, S^{2^h}) = 0$ for $2^{h+1} < k < 2^{h+2}$. By iterating this process one gets the result. □

It is worth to mention that Franjou, Friedlander, Scorichenko and Suslin give in [13] explicit calculations of $\mathrm{Ext}^*(\Gamma^{d(h)}, S^n)$, $\mathrm{Ext}^*(\Gamma^{d(h)}, \Lambda^n)$, $\mathrm{Ext}^*(\Gamma^{d(h)}, \Gamma^n)$ and $\mathrm{Ext}^*(\Lambda^{d(h)}, \Lambda^n)$ in the $\mathcal{F}$ and $\mathcal{P}$ framework. Let us also note that the groups $\mathrm{Ext}^*_{\mathcal{F}}(I, -)$ are known as Mac Lane cohomology (see [19]) and they are dual to the topological Hochschild homology (see [29]). The previous theorem for $h = 0$ was first obtained by Breen [4] in the framework of "Extensions du groupe additif" and Bökstedt [3] in the framework of topological Hochschild homology.

Returning to Proposition 3.4 one sees that the Frobenius twist induces monomorphisms

$$\mathrm{Ext}^i_{\mathcal{P}}(I^{(s+j)}, S^{2^s(j)}) \to \mathrm{Ext}^i_{\mathcal{P}}(I^{(s+j+1)}, S^{2^s(j+1)})$$

which stabilize for each i. This phenomenon is a particular case of the well known fact on cohomology of algebraic groups (see [18], page 347), which by virtue of Corollary 2.3 yields the following

Theorem 3.6. — *Let $F, T \in \mathcal{P}_d$. Then the Frobenius twist gives a monomorphism $\mathrm{Ext}^i_{\mathcal{P}}(F, T) \to \mathrm{Ext}^i_{\mathcal{P}}(F^{(1)}, T^{(1)})$. Moreover, for all sufficiently large m the map*

$$\mathrm{Ext}^i_{\mathcal{P}}(F^{(m)}, T^{(m)}) \to \mathrm{Ext}^i_{\mathcal{P}}(F^{(m+1)}, T^{(m+1)})$$

is an isomorphism.

By [13] the last assertion holds as soon as $i \leq 2^{m+1} - 1$.

We let $\mathrm{Ext}^i_{Gen}(F, T)$ be the common value of $\mathrm{Ext}^i_{\mathcal{P}}(F^{(m)}, T^{(m)})$, $m \gg 0$. Since the Frobenius twist in $\mathcal{F}_d$ is an equivalence of categories we see that the canonical map $\mathrm{Ext}^*_{\mathcal{P}}(F, T) \to \mathrm{Ext}^*_{\mathcal{F}}(F, T)$ gives rise to the homomorphism $\mathrm{Ext}^*_{Gen}(F, T) \to \mathrm{Ext}^*_{\mathcal{F}}(F, T)$. Now comparing the Proposition 3.4 and Theorem 3.5 one sees that the homomorphism $\mathrm{Ext}^*_{Gen}(I^{(h)}, S^{2^h}) \to \mathrm{Ext}^*_{\mathcal{F}}(I^{(h)}, S^{2^h})$ is an isomorphism. However in general $\mathrm{Ext}^*_{Gen}(F, T) \to \mathrm{Ext}^*_{\mathcal{F}}(F, T)$ is not an isomorphism even when $* = 0$. The following rather surprising result of Franjou, Friedlander, Scorichenko and Suslin gives a condition when it is an isomorphism.

Theorem 3.7 ([13]). — *Let $T, F \in \mathcal{P}_d$ be strict polynomial functors, then*

$$\mathrm{Ext}^*_{\mathcal{P}}(F, T) \to \mathrm{Ext}^*_{\mathcal{F}}(F, T)$$

is a monomorphism. Moreover if the field K contains at least d elements then

$$\mathrm{Ext}^*_{Gen}(F, T) \to \mathrm{Ext}^*_{\mathcal{F}}(F, T)$$

is an isomorphism.

The proof of the theorem is quite long and can be divided essentially into two parts. In the first part they prove a weaker version (Proposition 3.8) and then they use a base change argument in a very clever way to finish the proof. Here we prove only Proposition 3.8 and refer the interested reader to the original paper for the second part of the proof of Theorem 3.7.

Proposition 3.8. — *Let $F, T \in \mathcal{P}_d$ and let $m \geq 0$. Assume K has at least $2^m d$ elements. Then*
$$\mathrm{Ext}^i_{\mathcal{P}}(F^{(m)}, T^{(m)}) \to \mathrm{Ext}^i_{\mathcal{F}}(F, T)$$
is an isomorphism provided $i \leq 2^{m-d+1} - 2$.

The rest of this section is devoted to the proof of Proposition 3.8.

Definition 3.9 ([11]). — *A graded object $E^* = \bigoplus_n E^n \in \mathcal{E}$, where $\mathcal{E} = \mathcal{F}$ or $\mathcal{P}$ is called exponential if there exist natural isomorphisms $E^0(V) = K, E^*(V \bigoplus W) \cong E^*(V) \otimes E^*(W)$ and the values of E^n lie in $\mathcal{V}$.*

If E^* is an exponential functor, then $E^n(0) = 0$, for $n > 0$ and $deg(E^n) \leq n$. Typical examples of exponential functors are given by the symmetric algebra, the exterior algebra and the divided power algebra. If E^* and $\bar{E}^*$ are exponential functors then $E^* \otimes \bar{E}^*$ is also an exponential functor.

The same argument as in the proof of Lemma 3.3 gives also the following

Lemma 3.10 ([11, 13]). — *Let E^* be an exponential functor in $\mathcal{E}$. Then for any functors $B, C \in \mathcal{E}$ and any $i \geq 0$ one has isomorphisms of graded vector spaces*
$$\mathrm{Ext}^*_{\mathcal{E}}(E^n, B \otimes C) \cong \bigoplus_{i+j=n} \mathrm{Ext}^*_{\mathcal{E}}(E^i, B) \otimes \mathrm{Ext}^*_{\mathcal{E}}(E^j, C)$$
$$\mathrm{Ext}^*_{\mathcal{E}}(B \otimes C, E^n) \cong \bigoplus_{i+j=n} \mathrm{Ext}^*_{\mathcal{E}}(B, E^i) \otimes \mathrm{Ext}^*_{\mathcal{E}}(C, E^j).$$

Lemma 3.11 ([11, 13]). — *For any exponential functor E^* and any injective $J \in \mathcal{P}_n$, $n \equiv 0 \ mod\ 2^h$ one has $\mathrm{Ext}^i_{\mathcal{F}}(E^d, J) = 0$ provided $0 < i \leq 2^{n-d+2} - 2$.*

Proof. — First we exploit the fact that $S^n \in \mathcal{F}$ behaves like an injective with respect to functors of degree $\ll n$. Assume $n = \epsilon_0 + 2\epsilon_1 + \cdots + 2^k \epsilon_k$, $\epsilon_i = 0, 1$ is the 2-adic expansion of a natural number n and let $\alpha(n)$ be the number of nonzero elements of the set $\{\epsilon_0, \cdots, \epsilon_k, \}$. Then S^n is a retract of $(S^1)^{\otimes \epsilon_0} \otimes \cdots \otimes (S^{2^k})^{\otimes \epsilon_k}$. Hence a discussion after Lemma 3.3 shows that $\mathrm{Ext}^*_{\mathcal{F}}(K, S^n \otimes S^m) = 0$ as soon as $degK < \alpha(n) + \alpha(m)$. Based on this fact and using the exact sequence (3.1) it is not too difficult to prove that the Frobenius transformation yields an isomorphism $\mathrm{Ext}^i_{\mathcal{F}}(K, S^{2^n}) \to \mathrm{Ext}^i_{\mathcal{F}}(K, S^{2^{n+1}})$ provided $degK \leq d$ and $i \leq 2^{n-d+2} - 2$. Since any finite functor has a projective resolution of finite type (see [12]), it follows from Lemma 3.1 that for such K one has $\mathrm{Ext}^i_{\mathcal{F}}(K, S^{2^n}) = 0$ provided $degK \leq d$ and $0 < i \leq 2^{n-d+2} - 2$. Lemma 3.10 together with the fact that the functors $S^{n_1} \otimes \cdots \otimes S^{n_k}$, $n_1 + \cdots + n_k = n$ are injective generators in $\mathcal{P}_n$ can be used to finish the proof. $\square$

Proof of Proposition 3.8. — It consists of several reductions. Using a hypercohomology spectral sequence one can prove that it suffices to restrict ourselves to the case when F is a projective object. So we assume that $F = \Gamma^{d_1} \otimes \cdots \otimes \Gamma^{d_k}$, $d_1 + \cdots + d_k = d$. Let $T^{(m)} \to J^*$ be an injective resolution in $\mathcal{P}_{d2^m}$. Then one has a hypercohomology spectral sequence

$$E_1^{ij} = \mathrm{Ext}^i_{\mathcal{F}}(F^{(m)}, J^j) \Longrightarrow \mathrm{Ext}^{i+j}_{\mathcal{F}}(F^{(m)}, T^{(m)}).$$

Since $F^{(m)}$ is a direct summand of the exponential functor $(\Gamma^* \otimes \cdots \otimes \Gamma^*)^{(m)}$ one can use Lemma 3.11 to show that $E_1^{ij} = 0$ for all $0 < i \leq 2^{m-d+1} - 2$. So for all $i \leq 2^{m-d+1} - 2$ one has

$$\begin{aligned} &\mathrm{Ext}^i_{\mathcal{F}}(F^{(m)}, T^{(m)}) = H^i(\mathrm{Hom}_{\mathcal{F}}(F^{(m)}, J^*)) = \\ &= H^i(\mathrm{Hom}_{\mathcal{P}}(F^{(m)}, J^*)) = \mathrm{Ext}^i_{\mathcal{P}}(F^{(m)}, T^{(m)}) \end{aligned}$$

and hence the proposition. □

4. STABLE K-THEORY FOR FINITE FIELDS

4.1. Stable K-theory

Stable K-theory gives the possibility of reducing calculation of homology of the general linear groups with twisted coefficients to the homology of the general linear group with constant coefficients. This trick goes back to Waldhausen. Let R be a ring and let $\mathcal{V}_R$ be the category of finitely generated free R-modules. Moreover we let $\mathcal{B}$ be the category of all functors $\mathcal{V}_R^{op} \times \mathcal{V}_R \to Ab$. For any such functor D the abelian group $D(R^n, R^n)$ has a natural GL_nR-action, so one can consider the corresponding homology groups $H_*(GL_n(R), D(R^n, R^n))$. These groups form a direct system and we let $H_*(GL(R), D)$ be the corresponding limit when $n \to \infty$. It is a classical theorem of Dwyer that this system always stabilizes as soon as R satisfies some finiteness condition and D is of finite degree with respect to both variables (see [9]). Let us denote $H_0(GL(R), D)$ by $K_0^{st}(R, D)$. The functor

$$K_0^{st}(R, -) : \mathcal{B} \to Ab$$

is a right exact functor, whose left derived functors are denoted by $K_*^{st}(R, -)$ and called the *stable K-theory* of R. Similarly one can introduce the *cohomological version of stable K-theory* $K^*_{st}(R, -) : \mathcal{B} \to Ab$ as a left derived functor of the functor $K^0_{st}(R, -)$. Here $K^0_{st}(R, D) = H^0(GL(R), D)$ for any $D \in \mathcal{B}$. It was proved in [2] that in this way one recovers the original definition of Waldhausen for some class of rings including all fields and for such rings one has the following direct sum decomposition (see [2])

$$H_n(GL(R), D) \cong \bigoplus_{i+j=n} H_i(GL(R), K_j^{st}(R, D)),$$

where **the action of $GL(R)$ on K^{st} is trivial**. It was conjectured in [2] that the groups $K_*^{st}(R,D)$ are isomorphic to the homology $H_*(\mathcal{V}_R,D)$ of the category $\mathcal{V}_R$ with the coefficients in the bifunctor D provided D is of finite degree with respect to both variables. This conjecture is much stronger compared to a previous conjecture from [28] (see also [19], page 293), which was proved by Dundas and McCarthy [8]. We refer the reader to [29] for an explicit definition of $H_*(\mathcal{V}_R,D)$. For purposes of these notes it suffices to note that $H_*(\mathcal{V}_R,D)\cong(\mathrm{Ext}^*_{\mathcal{F}}(T,F))^{\diamondsuit}$ if $D(X,Y)=\mathrm{Hom}_R(FX,TY)$. Here $T,F\in\mathcal{F}(R):=(R{-}mod)^{\mathcal{V}_R}$ and the values of F lie in $\mathcal{V}_R$. By [2] the conjecture is true for a given ring R if and only if it is true for such bifunctors. Quillen proved in [33] that for a finite field K the (co)homology of $GL(K)$ with coefficients in K is trivial. Thus for a finite field K one has

$$K_*^{st}(K,D)\cong H_*(GL(K),D),\quad K^*_{st}(K,D)\cong H^*(GL(K),D)$$

and hence the following theorem proved by Betley and Suslin independently solves our conjecture for finite fields.

THEOREM 4.1. — *Let K be a finite field and let F and T be strict homogeneous functors of degree d. Then for each $i\geq 0$ the natural map*

$$\mathrm{Ext}^i_{\mathcal{F}}(F,T)\to\mathrm{Ext}^i_{GL_n(K)}(F(K^n),T(K^n))$$

is an isomorphism provided n is big enough.

We give here a sketch of the proof following Suslin (Appendix of [13]), see also [1] for a different argument.

First let us consider the case when $F=T=I$. In this case the theorem follows from the result of Dundas and McCarthy [8]. Following Friedlander and Suslin it can be proved also using the famous result of [5]. In our situation it says that for a fixed $m\geq 0$ there exist numbers $t(m)$ and $d(m)$ such that if $r\geq t(m)$ and $[K:\mathbf{F}_p]\geq d(m)$ then

$$\mathrm{Ext}^i_{\mathcal{P}}(I^{(r)},I^{(r)})\to H^i(GL_n(K),M_n(K))$$

is an isomorphism for all $i\leq m$. Here $M_n(K)$ is the adjoint representation of $GL_n(K)$. Now comparing this result to Theorem 3.7 we see that for a fixed m Theorem 4.1 is true for $F=T=I$, $i\leq m$ if the field is big enough. In order to handle small fields one uses the following trick. Since the $GL_n(K)$-module $M_n(K)$ is selfdual, one can pass to homology because $H_i(GL_n(K),M_n(K))\cong H^i(GL_n(K),M_n(K))^{\diamondsuit}$. Now in homology we have a possibility to change a field as follows.

Let $K\to L$ be an extension of finite fields of degree e. Then it yields a canonical homomorphism

$$can: H_*(GL(K),M(K))\to H_*(GL(L),M(L)),$$

where $M(K) = colim M_n(K)$. Choosing a basis for L over K one gets an embedding $L \subset M_e(L)$ and hence a homomorphism $H_*(GL_n(L), M_n(L)) \to H_*(GL_{en}(K), M_{en}(K))$ and therefore

$$Tr : H_*(GL(L), M(L)) \to H_*(GL(K), M(K)).$$

PROPOSITION 4.2. — *The composition*

$$Tr \circ can : H_*(GL(K), M(K)) \to H_*(GL(K), M(K))$$

coincides with the multiplication by $e = [L : K]$*, while the composition*

$$can \circ Tr : H_*(GL(L), M(L)) \to H_*(GL(L), M(L))$$

coincides with $\sum_{\sigma \in Gal(L/K)} \sigma_*$.

In order to finish the case $F = T = I$ we choose d sufficiently big with respect to m and prime to p and consider an extension $K \to L$ of degree d. Then for L the theorem is true in dimensions $i \leq m$. By Theorem 3.5 (for $h = 0$) we have $H_i(GL_n(L), M_n(L)) = 0$ for odd $i \leq m$ and $H_i(GL_n(L), M_n(L)) = L$ for even $i \leq m$. Based on Proposition 4.2 it is not too difficult to see that the same relations hold for $H_i(GL_n(K), M_n(K))$. Hence Theorem 4.1 is true for $F = T = I$.

Following Suslin, for a functor $F \in \mathcal{F}$ we let $aF \in \mathcal{F}$ be the functor which is given on objects by

$$(aF)(V) = \bigoplus_{W \subset V} F(V/W).$$

If $f : V \to V'$ is a linear map, then the W'-component of the induced map

$$(aF)(f) : \bigoplus_{W \subset V} F(V/W) \to \bigoplus_{W' \subset V'} F(V'/W')$$

is zero on all W except $W = f^{-1}(W')$. In this case the corresponding component of $(aF)(f)$ is given by $f_* : F(V/f^{-1}(W')) \to F(V'/W')$.

PROPOSITION 4.3. — *Assume $F \in \mathcal{F}$ has a projective resolution of finite type and let $T \in \mathcal{F}$ be a finite functor. Then one has a natural isomorphism*

$$\mathrm{Ext}^*_{GL(K)}(F, T) \to \mathrm{Ext}^*_{\mathcal{F}}(F, aT).$$

Proof. — By a hypercohomology spectral sequence argument it suffices to consider only the case when $F = P_V$ is a standard projective and $T = J_W$ is a standard injective. One can prove the claim in dimension zero by direct considerations, while in dimensions > 0 it follows from the fact that the corresponding bifunctor $D(X, Y) = \mathrm{Hom}(FX, TY)$ is injective in $\mathcal{B}$ and hence $\mathrm{Ext}^i_{GL(K)}(F, T) \cong K^i_{st}(K, D) = 0$, $i > 0$. □

The canonical homomorphism $T \to aT$ yields a homomorphism $\sigma(F, T)$: $\mathrm{Ext}^*_{\mathcal{F}}(F, T) \to \mathrm{Ext}^*_{\mathcal{F}}(F, aT)$. Thanks to Proposition 4.3 this map is an isomorphism if and only if Theorem 4.1 is true for the pair (F, T). We will also need the following

LEMMA 4.4. — *Let E^* be an exponential functor and let $B_1, \cdots, B_n \in \mathcal{F}$ be functors with the property $B_i(0) = 0$, $i = 1, \cdots, n$. Assume further that the natural homomorphisms $\sigma(E^s, B_i)$ are isomorphisms for all $s < m$. Then $\sigma(A^m, B_1 \otimes \cdots \otimes B_n)$ is an isomorphism as well.*

Proof of Theorem 4.1. — By the first part of the proof we know that $\sigma(I, I)$ is an isomorphism and we have to show that $\sigma(F, T)$ is an isomorphism provided F and T satisfy the conditions of the Theorem. This can be done by induction on the degree of functors based on Lemma 4.4 as follows. First one uses the exact sequence (3.1) (here we start to assume that $p = 2$, for odd p one needs to use De Rham and Koszul complexes). This sequence remains exact after applying the exact functor $a : \mathcal{F} \to \mathcal{F}$. By induction one readily proves that $\sigma(I, S^n)$ is an isomorphism for all n. Indeed it suffices to note that the induction assumption and Lemma 4.4 imply $\mathrm{Ext}^*_{\mathcal{F}}(I, a(S^i \otimes S^j)) = 0$, for $i + j < n$ and one can use the same method as in Section 3. By duality one obtains that $\sigma(\Gamma^n, I)$ is an isomorphism as well. Having these facts in mind and repeating the previous argument one sees that $\sigma(\Gamma^n, S^m)$ is an isomorphism for all n, m too. Now Lemma 4.4 yields that $\sigma(F, T)$ is an isomorphism when F is projective in $\mathcal{P}$ and T is an injective, and the hypercohomology spectral sequence gives the result. $\square$

REFERENCES

[1] S. BETLEY. *Stable K-theory of finite fields.* K-Theory 17 (1999), no. 2, 103–111.

[2] S. BETLEY & T. PIRASHVILI. *Stable K-theory as a derived functor.* J. Pure Appl. Algebra 96 (1994), no. 3, 245–258.

[3] M. BÖKSTEDT. *On the topological Hochschild homology of* $\mathbf{Z}$ *and* $\mathbf{Z}/p\mathbf{Z}$. Bielefeld (1987).

[4] L. BREEN. *Extensions du groupe additif.* Publ. Ihes 48 (1978), 39–125.

[5] E. CLINE, B. PARSHALL, L. SCOTT & W. VAN DER KALLEN. *Rational and generic cohomology.* Invent. Math. 39 (1977), 143–163.

[6] S. DONKIN. *On Schur algebras and related algebras. I.* J. Algebra 104 (1986), 310–328.

[7] S. DONKIN. *On Schur algebras and related algebras. II.* J. Algebra 111 (1987), 354–364.

[8] B. DUNDAS & R. MCCARTHY. *Stable K-theory and topological Hochschild homology,* Ann. of Math. 140 (1994), 685-701; erratum: Ann. of Math. 142 (1995), 425-426.

[9] W. DWYER. *Twisted homological stability for general linear groups.* Ann. of Math. 111 (1980), 239-251.

[10] S. Eilenberg. & S. Mac Lane. *On the groups $H(\pi, n)$. II.* Ann. of Math. 60 (1954), 49-139.

[11] V. Franjou. *Extensions entre puissances extérieures et entre puissances symétriques.* J. Algebra 179 (1996), 501–522.

[12] V. Franjou, J. Lannes & L. Schwartz. *Autour de la cohomologie de Mac Lane des corps finis.* Invent. Math. 115 (1994), 513–538.

[13] V. Franjou, E. Friedlander, A. Scorichenko & A. Suslin. *General linear and functor cohomology over finite fields.* Ann. of Math. (2) 150 (1999), 663–728.

[14] E. Friedlander & A. Suslin. *Cohomology of finite group schemes over a field.* Invent. Math. 127(1997), 209-270.

[15] J.A. Green. *Polynomial representations of* Gl_n. Lecture Notes in Mathematics 830. Springer-Verlag, (1980).

[16] H-W. Henn, J. Lannes & L. Schwartz. *The categories of unstable modules and unstable algebras over the Steenrod algebra modulo nilpotent objects.* Amer. J. Math. 115 (1993), 1053–1106.

[17] H-W. Henn, J. Lannes & L. Schwartz. *Localizations of unstable A-modules and equivariant mod p cohomology.* Math. Ann. 301 (1995), 23–68.

[18] J.C. Jantzen. *Representations of algebraic groups.* Pure and Applied Mathematics 131. Academic Press, Inc., Boston, Ma (1987).

[19] M. Jibladze & T. Pirashvili. *Cohomology of algebraic theories.* J. Algebra 137 (1991), 253–296.

[20] N.J. Kuhn. *Generic representations of the finite general linear groups and the Steenrod algebra. I.* Amer. J. Math. 116 (1994), 327–360.

[21] N.J. Kuhn. *Generic representations of the finite general linear groups and the Steenrod algebra. II.* K-Theory 8 (1994), 395–428.

[22] N.J. Kuhn. *Generic representations of the finite general linear groups and the Steenrod algebra. III.* K-Theory 9 (1995), 273–303.

[23] N.J. Kuhn. *On topologically realizing modules over the Steenrod algebra.* Ann. of Math. (2) 141 (1995), 321–347.

[24] I.G. Macdonald. *Symmetric functions and Hall polynomials.* Second edition. Oxford University Press (1995).

[25] S. Martin. *Schur algebras and representation theory.* Cambridge University Press (1993).

[26] T. Pirashvili. *Polynomial functors.* (Russian) Trudy Tbiliss. Mat. Inst. Razmadze Akad. Nauk Gruzin. Ssr 91 (1988), 55–66.

[27] T. Pirashvili. *Higher additivizations.* (Russian) Trudy Tbiliss. Mat. Inst. Razmadze Akad. Nauk Gruzin. Ssr 91 (1988), 44–54.

[28] T. Pirashvili. *New homology and cohomology of rings.* (Russian) Soobshch. Akad. Nauk Gruzin. Ssr 133 (1989), no. 3, 477–480.

[29] T. Pirashvili & F. Waldhausen. *Mac Lane homology and topological Hochschild homology.* J. Pure Appl. Algebra 82 (1992), no. 1, 81–98.

[30] L. Piriou & L. Schwartz. *Extensions de foncteurs simples.* K-Theory 15 (1998), 269–291.

[31] G. Powell. *The Artinian conjecture for* $I^{\otimes 2}$. J. Pure Appl. Algebra 128 (1998), 291–310.

[32] N. Roby. *Lois polynômes et lois formelles en théorie des modules.* Ann. Sci. Éc. Norm. Sup. (3) 80 (1963) 213–348.

[33] D. Quillen. *On the cohomology and K-theory of the general groups over a finite field*, Ann. of Math. 96 (1972), 552-586.

[34] L. Schwartz. *Unstable modules over the Steenrod algebra and Sullivan's fixed point set conjecture.* University of Chicago Press (1994).

[35] L. Schwartz. *À propos de la conjecture de non-réalisation due à N. Kuhn.* Invent. Math. 134 (1998), no. 1, 211–227.

[36] B. Totaro. *Projective resolutions of representations of* $\mathrm{Gl}(n)$. J. Reine Angew. Math. 482 (1997), 1–13.

Teimuraz PIRASHVILI

A.M. Razmadze Math. Institute
Aleksidze str. 1
Tbilisi, 380093, Georgia
E-mail : pira@rmi.acnet.ge

Séminaire BOURBAKI Juin 2000
52e année, 1999-2000, no 878, p. 389 à 409

FAITHFUL LINEAR REPRESENTATIONS OF THE BRAID GROUPS

by **Vladimir TURAEV**

The braid group on n strings B_n can be defined as the group generated by $n-1$ generators $\sigma_1, ..., \sigma_{n-1}$ with defining relations

$$\sigma_i \sigma_j = \sigma_j \sigma_i \tag{0.1}$$

if $|i-j| \geq 2$ and

$$\sigma_i \sigma_{i+1} \sigma_i = \sigma_{i+1} \sigma_i \sigma_{i+1} \tag{0.2}$$

for $i = 1, ..., n-2$. This group was introduced by Emil Artin in 1926. It has various interpretations, specifically, as the group of geometric braids in $\mathbf{R}^3$, as the mapping class group of an n-punctured disc, as the fundamental group of the configuration space of n points on the plane, etc. The algebraic properties of B_n have been studied by many authors. To mention a few, note the solution of the conjugacy problem in B_n given by F. Garside, the papers of N. Ivanov and J. McCarthy who proved that the mapping class groups and in particular B_n satisfy the "Tits alternative", and the work of P. Dehornoy establishing the existence of a left-invariant total order on B_n (see [Ga], [Iv2], [Ka]).

One of the most intriguing problems in the theory of braids is the question of whether B_n is linear, i.e., whether it admits a faithful representation into a group of matrices over a commutative ring. This question has its origins in a number of interrelated facts and first of all in the discovery by W. Burau [Bu] of an n-dimensional linear representation of B_n which for a long time had been considered as a candidate for a faithful representation. However, as it was established by J. Moody in 1991 this representation is not faithful for $n \geq 9$. Later it was shown to be unfaithful for $n \geq 5$. Thus, the question of the linearity of B_n remained open.

In 1999/2000 there appeared a series of papers of D. Krammer and S. Bigelow who proved that B_n is linear for all n. First there appeared a paper of Krammer [Kr1] in which he constructed a homomorphism from B_n to $GL(n(n-1)/2, R)$ where $R = \mathbf{Z}[q^{\pm 1}, t^{\pm 1}]$ is the ring of Laurent polynomials on two variables. He proved that this homomorphism is injective for $n = 4$ and conjectured that the same is true for all

n. Soon after that, Bigelow [Bi2] gave a beautiful topological proof of this conjecture. Another proof based on different ideas was obtained by Krammer [Kr2] independently.

The ring $\mathbf{Z}[q^{\pm 1}, t^{\pm 1}]$ can be embedded in the field of real numbers by assigning to q, t any algebraically independent non-zero real values. Therefore B_n embeds in $GL(n(n-1)/2, \mathbf{R})$. As an application, note that the linearity of B_4 implies the linearity of the group $\mathrm{Aut}(F_2)$, where F_n is a free group of rank n (see [DFG]). It is known that $\mathrm{Aut}(F_n)$ is not linear for $n \geq 3$, see [FP].

The representation of B_n considered by Krammer and Bigelow is one of a family of representations introduced earlier by R. Lawrence [La]. Her work was inspired by a study of the Jones polynomial of links and was concerned with representations of Hecke algebras arising from the actions of braids on homology of configuration spaces.

The same representation of B_n arises from a study of the so-called Birman-Murakami-Wenzl algebra C_n. This algebra is a quotient of the group ring $\mathbf{C}[B_n]$ by certain relations inspired by the theory of link polynomials. The irreducible finite dimensional representations of C_n were described by H. Wenzl in terms of Young diagrams. These representations yield irreducible finite dimensional representations of B_n. One of them was shown by M. Zinno [Zi] and independently by V. Jones to be equivalent to the Krammer representation which is henceforth irreducible.

The aim of this paper is to present these results. In Sect. 1 we consider the Burau representation and explain why it is not faithful. In Sect. 2 we outline Bigelow's approach following [Bi2]. In Sect. 3 we discuss the work of Krammer [Kr2]. Finally in Sect. 4 we discuss the Birman-Murakami-Wenzl algebras and the work of Zinno.

The author is indebted to S. Bigelow for useful comments on his work and to Ch. Kassel for careful reading of a preliminary version of this paper.

1. THE BURAU REPRESENTATION OF THE BRAID GROUP

1.1. Mapping class groups

It will be convenient for us to view the braid group as the mapping class group of a punctured disc. We recall here the definition and a few simple properties of the mapping class groups.

Let Σ be a connected oriented surface. By a *self-homeomorphism* of Σ we mean an orientation preserving homeomorphism $\Sigma \to \Sigma$ which fixes $\partial\Sigma$ pointwise. Two such homeomorphisms are *isotopic* if they can be included in a continuous one-parameter family of self-homeomorphisms of Σ. The mapping class group $\mathrm{Homeo}(\Sigma)$ of Σ is the group of isotopy classes of self-homeomorphisms of Σ with the group operation determined by composition.

Each self-homeomorphism of Σ induces an automorphism of the abelian group $H = H_1(\Sigma; \mathbf{Z})$. This is a "homological" representation $\mathrm{Homeo}(\Sigma) \to \mathrm{Aut}(H)$. The action of homeomorphisms preserves the skew-symmetric bilinear form $H \times H \to \mathbf{Z}$ determined by the algebraic intersection number. The value $[\alpha] \cdot [\beta] \in \mathbf{Z}$ of this form on the homology classes $[\alpha], [\beta] \in H$ represented by oriented loops α, β on Σ is computed as follows. Assume that α and β lie in a generic position so that they meet each other transversely in a finite set of points which are not self-crossings of α or β. Then

$$[\alpha] \cdot [\beta] = \sum_{p \in \alpha \cap \beta} \varepsilon_p \,,$$

where $\varepsilon_p = +1$ if the tangent vectors of α, β at p form a positively oriented basis and $\varepsilon_p = -1$ otherwise.

An example of a self-homeomorphism of Σ is provided by the Dehn twist τ_α about a simple closed curve $\alpha \subset \Sigma$. It is defined as follows. Identify a regular neighborhood of α in Σ with the cylinder $S^1 \times [0,1]$ so that $\alpha = S^1 \times (1/2)$. We choose this identification so that the product of the counterclockwise orientation on $S^1 = \{z \in \mathbf{C} \,|\, |z| = 1\}$ and the right-handed orientation on $[0,1]$ corresponds to the given orientation on Σ. The Dehn twist $\tau_\alpha : \Sigma \to \Sigma$ is the identity map outside $S^1 \times [0,1]$ and sends any $(x, s) \in S^1 \times [0,1]$ to $(e^{2\pi i s} x, s) \in S^1 \times [0,1]$. To compute the action of τ_α in homology, we orient α in an arbitrary way. The effect of τ_α on an oriented curve transversal to α is to insert $\alpha^{\pm 1}$ at each crossing of α with this curve, where ± 1 is the sign of the crossing. Therefore for any $g \in H$,

$$(\tau_\alpha)_*(g) = g + ([\alpha] \cdot g)[\alpha]. \tag{1.1}$$

Note that τ_α and its action on H do not depend on the choice of orientation on α.

A similar construction applies to arcs in Σ whose endpoints are punctures. Assume that Σ is obtained from another surface Σ' by puncturing, i.e., by removing a finite set of points lying in $\mathrm{Int}(\Sigma')$. These points will be called *punctures*. Let α be an embedded arc in Σ' whose endpoints are punctures x_1, x_2 and whose interior lies in Σ. One can define the Dehn "half-twist" $\tau_\alpha : \Sigma \to \Sigma$ which is the identity map outside a regular neighborhood of α in Σ' and which exchanges x_1 and x_2. This homeomorphism is obtained by the isotopy of the identity map of Σ' rotating α about its midpoint to the angle of π in the direction provided by the orientation of Σ. Restricting the resulting homeomorphism of $\Sigma' \to \Sigma'$ to Σ we obtain τ_α. To compute the action of τ_α on $H = H_1(\Sigma)$ we orient α from x_1 to x_2 and associate to α a loop α' in Σ as follows. Choose a point $z \in \alpha$ and for $i = 1, 2$ denote by μ_i the loop in Σ beginning at z and moving along α until coming very closely to x_i, then encircling x_i in the direction determined by the orientation of Σ and finally moving back to z along α. Set $\alpha' = \mu_1^{-1} \mu_2$. The homotopy class of the loop α' on Σ does not depend on the choice of z. The effect of τ_α on an oriented curve transversal to α is to insert $(\alpha')^{\pm 1}$ at each crossing of α with this curve. Thus for any $g \in H$, we have

$$(\tau_\alpha)_*(g) = g + ([\alpha] \cdot g)[\alpha'] \tag{1.2}$$

where $[\alpha] \cdot g = -g \cdot [\alpha] \in \mathbf{Z}$ is the algebraic intersection number of g with the 1-dimensional homology class $[\alpha]$ of Σ "modulo infinity" represented by α.

In general, the action of $\mathrm{Homeo}(\Sigma)$ on $H = H_1(\Sigma)$ is not faithful. We point out one source of non-faithfulness. If $\alpha, \beta \subset \Sigma$ are simple closed curves with $[\alpha] \cdot [\beta] = 0$ then formula (1.1) implies that $(\tau_\alpha)_*$ and $(\tau_\beta)_*$ commute in $\mathrm{Aut}(H)$. The Dehn twists τ_α, τ_β themselves commute if and only if α is isotopic to a simple closed curve disjoint from β, see for instance [Iv1]. It is easy to give examples of simple closed curves $\alpha, \beta \subset \Sigma$ which are not disjoint up to isotopy but have zero algebraic intersection number. Then the commutator $[\tau_\alpha, \tau_\beta]$ lies in the kernel of the homological representation. Using (1.2), one can similarly derive elements of the kernel from embedded arcs with endpoints in punctures.

1.2. Braid groups

Let $D = \{z \in \mathbf{C} \,|\, |z| \leq 1\}$ be the unit disc with counterclockwise orientation. Fix a set of $n \geq 1$ distinct punctures $X = \{x_1, ..., x_n\} \subset \mathrm{Int}(D)$. We shall assume that $x_1, ..., x_n \in (-1, +1) = \mathbf{R} \cap \mathrm{Int}(D)$ and $x_1 < x_2 < ... < x_n$. Set $D_n = D \backslash X$. The group $\mathrm{Homeo}(D_n)$ is denoted B_n and called the *n-th braid group*. An element of B_n is an isotopy class of a homeomorphism $D_n \to D_n$ which fixes $\partial D_n = \partial D = S^1$ pointwise. Such a homeomorphism uniquely extends to a homeomorphism $D \to D$ permuting $x_1, ..., x_n$. This defines a group homomorphism from B_n onto the symmetric group S_n. We can equivalently define B_n at the group of isotopy classes of homeomorphisms $D \to D$ which fix ∂D pointwise and preserve X as a set.

For $i = 1, ..., n-1$, the linear interval $[x_i, x_{i+1}] \subset (-1, +1) \subset \mathbf{R}$ is an embedded arc in D with endpoints in the punctures x_i, x_{i+1}. The corresponding Dehn half-twist $D_n \to D_n$ is denoted by σ_i. It is a classical fact that B_n is generated by $\sigma_1, ..., \sigma_{n-1}$ with defining relations (0.1), (0.2). The image of σ_i in S_n is the permutation $(i, i+1)$.

Another definition of B_n can be given in terms of braids. A (geometric) *braid on n strings* is an n-component one-dimensional manifold $E \subset D \times [0, 1]$ such that E meets $D \times \{0, 1\}$ orthogonally along the set $(X \times 0) \cup (X \times 1)$ and the projection on $[0, 1]$ maps each component of E homeomorphically onto $[0, 1]$. The braids are considered up to isotopy in $D \times [0, 1]$ constant on the endpoints. The group operation in the set of braids is defined by glueing one braid on the top of the other one and compressing the result into $D \times [0, 1]$.

The equivalence between these two definitions of B_n is established as follows. Any homeomorphism $h : D \to D$ which fixes ∂D pointwise is related to the identity map id_D by an isotopy $\{h_s : D \to D\}_{s \in [0,1]}$ such that $h_0 = \mathrm{id}_D$ and $h_1 = h$. If $h(X) = X$ then the set $\cup_{s \in [0,1]} (h_s(X) \times s) \subset D \times [0, 1]$ is a braid. Its isotopy class depends only on the element of B_n represented by h. This establishes an isomorphism between B_n and the group of braids on n strings. The generator $\sigma_i \in B_n$ corresponds to the i-th "elementary" braid represented by a plane diagram consisting of n linear intervals which are disjoint except at one intersection point where the i-th interval goes over the $(i+1)$-th interval.

1.3. The Burau representation

Let Λ denote the ring $\mathbf{Z}[t,t^{-1}]$. The Burau representation $B_n \to GL_n(\Lambda)$ sends the i-th generator $\sigma_i \in B_n$ into the matrix

$$I_{i-1} \oplus \begin{pmatrix} 1-t & t \\ 1 & 0 \end{pmatrix} \oplus I_{n-i-1}\,,$$

where I_k denotes the identity $(k\times k)$-matrix and the non-trivial (2×2)-block appears in the i-th and $(i+1)$-th rows and columns (see [Bu]). Substituting $t=1$, we obtain the standard representation of the symmetric group S_n by permutation matrices or equivalently the homological action of B_n on $H_1(D_n)=\mathbf{Z}^n$. The Burau representation is reducible: it splits as a direct sum of an $(n-1)$-dimensional representation and the trivial one-dimensional representation.

The Burau representation is known to be faithful for $n\leq 3$, see [Bir]. J. Moody [Mo] proved in 1991 that it is not faithful for $n\geq 9$. D. Long and M. Paton [LP] extended Moody's argument to $n\geq 6$. Recently, S. Bigelow [Bi1] proved that this representation is not faithful for $n=5$. The case $n=4$ remains open.

The geometric idea allowing to detect non-trivial elements in the kernel of the Burau representation is parallel to the one at the end of Sect. 1.1. We first give a homological description of the Burau representation. Observe that $H_1(D\backslash x_i)=\mathbf{Z}$ is generated by the class of a small loop encircling x_i in the counterclockwise direction. Each loop in $D\backslash x_i$ represents k times the generator where k is the *winding number* of the loop around x_i. Consider the homomorphism $H_1(D_n)\to\mathbf{Z}$ sending the homology class of a loop to its *total winding number* defined as the sum of its winding numbers around $x_1,...,x_n$. Let $\tilde{D}_n \to D_n$ be the corresponding regular covering. The group of covering transformations of $\tilde{D}_n$ is $\mathbf{Z}$ which we write as a multiplicative group with generator t. The group $H_1(\tilde{D}_n)$ acquires thus the structure of a Λ-module. It is easy to check that this is a free Λ-module of rank $n-1$.

Fix a basepoint $d\in\partial D$. Any homeomorphism $h: D_n\to D_n$ representing an element of B_n lifts uniquely to a homeomorphism $\tilde{h}:\tilde{D}_n\to\tilde{D}_n$ which fixes the fiber over d pointwise. This induces a Λ-linear automorphism $\tilde{h}_*$ of $H_1(\tilde{D}_n)$. The map $h\mapsto\tilde{h}_*$ defines a representation $B_n\to\mathrm{Aut}(H_1(\tilde{D}_n))$ equivalent to the $(n-1)$-dimensional Burau representation.

Now we extend the algebraic intersection to arcs and refine it so that it takes values in Λ. Let α,β be two embedded oriented arcs in D_n with endpoints in the punctures. (We assume that all four endpoints of α,β are distinct so that $n\geq 4$). Let $\tilde{\alpha},\tilde{\beta}$ be lifts of α,β to $\tilde{D}_n$, respectively. Set

$$\langle\alpha,\beta\rangle=\sum_{k\in\mathbf{Z}}(t^k\tilde{\alpha}\cdot\tilde{\beta})\,t^k\in\Lambda$$

where $t^k\tilde{\alpha}\cdot\tilde{\beta}\in\mathbf{Z}$ is the algebraic intersection number of the arcs $t^k\tilde{\alpha},\tilde{\beta}$ in $\tilde{D}_n$. This finite sum is only defined up to multiplication by a power of t depending on the choice of $\tilde{\alpha},\tilde{\beta}$. This will not be important for us since we are only interested in whether or not $\langle\alpha,\beta\rangle=0$. To compute $\langle\alpha,\beta\rangle$ explicitly one deforms α in general position with respect to β. Then $\langle\alpha,\beta\rangle=\sum_{p\in\alpha\cap\beta}\varepsilon_p\,t^{k_p}$ where $\varepsilon_p=\pm$ is the intersection sign at p and $k_p\in\mathbf{Z}$. The exponents $\{k_p\}_p$ are determined by the following condition: if $p,q\in\alpha\cap\beta$, then k_p-k_q is the total winding number of the loop going from p to q along α and then from q to p along β. Note that

$$\langle\beta,\alpha\rangle=\sum_{k\in\mathbf{Z}}(t^k\tilde{\beta}\cdot\tilde{\alpha})\,t^k=\sum_{k\in\mathbf{Z}}(\tilde{\beta}\cdot t^{-k}\tilde{\alpha})\,t^k=-\sum_{k\in\mathbf{Z}}(t^{-k}\tilde{\alpha}\cdot\tilde{\beta})\,t^k=-\overline{\langle\alpha,\beta\rangle}$$

where the overline denotes the involution in Λ sending any t^k to t^{-k}. Hence $\langle\alpha,\beta\rangle=0$ if and only if $\langle\beta,\alpha\rangle=0$.

We claim that if $\langle\alpha,\beta\rangle=0$, then the automorphisms $(\tilde{\tau}_\alpha)_*,(\tilde{\tau}_\beta)_*$ of $H_1(\tilde{D}_n)$ commute. Observe that the loops α',β' on D_n associated to α,β as in Sect. 1.1 have zero total winding numbers and therefore lift to certain loops $\tilde{\alpha}',\tilde{\beta}'$ in $\tilde{D}_n$. The effect of $\tilde{\tau}_\alpha$ on any oriented loop γ in $\tilde{D}_n$ is to insert a lift of $(\alpha')^{\pm1}$ at each crossing of γ with the preimage of α in $\tilde{D}_n$. Thus

$$(\tilde{\tau}_\alpha)_*([\gamma])=[\gamma]+\lambda_\gamma\,[\tilde{\alpha}']$$

for a certain Laurent polynomial $\lambda_\gamma\in\Lambda$. The coefficients of λ_γ are the algebraic intersection numbers of γ with lifts of α to $\tilde{D}_n$. By $\langle\alpha,\beta\rangle=0$, any lift of α has algebraic intersection number zero with any lift of β and hence with any lift of β'. Therefore, $\lambda_{\tilde{\beta}'}=0$ and $(\tilde{\tau}_\alpha)_*([\tilde{\beta}'])=[\tilde{\beta}']$. Similarly, $(\tilde{\tau}_\beta)_*([\gamma])=[\gamma]+\mu_\gamma\,[\tilde{\beta}']$ with $\mu_\gamma\in\Lambda$ and $(\tilde{\tau}_\beta)_*([\tilde{\alpha}'])=[\tilde{\alpha}']$. We conclude that

$$(\tilde{\tau}_\alpha\tilde{\tau}_\beta)_*([\gamma])=[\gamma]+\lambda_\gamma\,[\tilde{\alpha}']+\mu_\gamma\,[\tilde{\beta}']=(\tilde{\tau}_\beta\tilde{\tau}_\alpha)_*([\gamma]).$$

To show that the Burau representation is not faithful it remains to provide an example of oriented embedded arcs α,β in D_n with endpoints in distinct punctures such that $\langle\alpha,\beta\rangle=0$ and $\tau_\alpha\tau_\beta\neq\tau_\beta\tau_\alpha$ in B_n. For $n\geq6$, the simplest known example (see [Bi1]) is provided by the pair $\alpha=\varphi_1([x_3,x_4])$, $\beta=\varphi_2([x_3,x_4])$ where

$$\varphi_1=\sigma_1^2\sigma_2^{-1}\sigma_5^{-2}\sigma_4,\quad \varphi_2=\sigma_1^{-1}\sigma_2\sigma_5\sigma_4^{-1}.$$

To compute $\langle\alpha,\beta\rangle$ one can draw α,β and use the recipe above. To prove that the braids $\tau_\alpha=\varphi_1\sigma_3\varphi_1^{-1}$ and $\tau_\beta=\varphi_2\sigma_3\varphi_2^{-1}$ do not commute, one can use the solution of the word problem in B_n or the methods of the Thurston theory of surfaces (cf. Sect. 2.2). Thus the commutator $[\tau_\alpha,\tau_\beta]$ lies in the kernel of the Burau representation. This commutator can be represented by a word of length 44 in the generators $\sigma_1,...,\sigma_5$.

Similar ideas apply in the case $n=5$, although one has to extend them to arcs relating the punctures to the base point $d\in\partial D$. The shortest known word in the generators $\sigma_1,...,\sigma_4$ representing an element of the kernel has length 120.

2. THE WORK OF BIGELOW

2.1. A representation of B_n

We use the notation $D, D_n, X = \{x_1, ..., x_n\}$ introduced in Sect. 1. Let C be the space of all unordered pairs of distinct points in D_n. This space is obtained from $(D_n \times D_n)\backslash$diagonal by the identification $\{x, y\} = \{y, x\}$ for any distinct $x, y \in D_n$. It is clear that C is a connected non-compact 4-manifold with boundary. It has a natural orientation induced by the counterclockwise orientation of D_n. Set $d = -i \in \partial D$ where $i = \sqrt{-1}$ and $d' = -i\, e^{\varepsilon \pi i 2} \in \partial D$ with small positive ε. We take $c_0 = \{d, d'\}$ as the basepoint for C.

A closed curve $\alpha : [0,1] \to C$ can be written in the form $\alpha(s) = \{\alpha_1(s), \alpha_2(s)\}$ where $s \in [0,1]$ and α_1, α_2 are arcs in D_n such that $\{\alpha_1(0), \alpha_2(0)\} = \{\alpha_1(1), \alpha_2(1)\}$. The arcs α_1, α_2 are either both loops or can be composed with each other. They form thus a closed oriented one-manifold mapped to D_n. Let $a(\alpha) \in \mathbf{Z}$ be the total winding number of this one-manifold around the punctures $\{x_1, ..., x_n\}$. Composing the map $s \mapsto (\alpha_1(s) - \alpha_2(s))/|\alpha_1(s) - \alpha_2(s)| : [0,1] \to S^1$ with the projection $S^1 \to \mathbf{RP}^1$ we obtain a loop in RP^1. The corresponding element of $H_1(RP^1) = \mathbf{Z}$ is denoted by $b(\alpha)$. The formula $\alpha \mapsto q^{a(\alpha)} t^{b(\alpha)}$ defines a homomorphism, ϕ, from $H_1(C)$ to the (multiplicatively written) free abelian group with basis q, t. Let $R = \mathbf{Z}[q^{\pm 1}, t^{\pm 1}]$ be the group ring of this group.

Let $\tilde{C} \to C$ be a regular covering corresponding to the kernel of ϕ. The generators q, t act on $\tilde{C}$ as commuting covering transformations. The homology group $H_2(\tilde{C}) = H_2(\tilde{C}; \mathbf{Z})$ becomes thus an R-module.

Any self-homeomorphism h of D_n induces by $h(\{x, y\}) = \{h(x), h(y)\}$ a homeomorphism $C \to C$ also denoted h. It is easy to check that $h(c_0) = c_0$ and the action of h on $H_1(C)$ commutes with ϕ. Therefore this homeomorphism $C \to C$ lifts uniquely to a map $\tilde{h} : \tilde{C} \to \tilde{C}$ which fixes the fiber over c_0 pointwise and commutes with the covering transformations. Consider the representation $B_n \to \mathrm{Aut}(H_2(\tilde{C}))$ sending the isotopy class of h to the R-linear automorphism $\tilde{h}_*$ of $H_2(\tilde{C})$.

2.2. Theorem (S. Bigelow [Bi2]). — *The representation $B_n \to \mathrm{Aut}(H_2(\tilde{C}))$ is faithful for all $n \geq 1$.*

We outline below the main ideas of Bigelow's proof. The proof uses almost no information about the structure of the R-module $H_2(\tilde{C})$. The only thing needed is the absence of R-torsion or more precisely the fact that multiplication by a non-zero polynomial of type $q^a t^b - 1$ has zero kernel in $H_2(\tilde{C})$. In fact, $H_2(\tilde{C})$ is a free R-module of rank $n(n-1)/2$, as it was essentially shown in [La].

We shall use one well-known fact concerning isotopies of arcs on surfaces. Let N, T be embedded arcs in D_n with distinct endpoints lying either in the punctures or on ∂D_n. Assume that the interiors of N, T do not meet ∂D_n, and that N intersects T

transversely (in a finite number of points). A *bigon* for the pair (N,T) is an embedded disc in $\mathrm{Int}(D_n)$ whose boundary is formed by one subarc of N and one subarc of T and whose interior is disjoint from N and T. It is clear that in the presence of a bigon there is an isotopy of T constant on the endpoints and decreasing $\#(N\cap T)$ by two. Thurston's theory of surfaces implies the converse: if there is an isotopy of T (rel endpoints) decreasing $\#(N\cap T)$ then the pair (N,T) has at least one bigon, cf. [FLP, Prop. 3.10].

2.3. Noodles and forks

We need the following notation. For arcs $\alpha,\beta:[0,1]\to D_n$ such that $\alpha(s)\neq\beta(s)$ for all $s\in[0,1]$, we denote by $\{\alpha,\beta\}$ the arc in C given by $\{\alpha,\beta\}(s)=\{\alpha(s),\beta(s)\}$. We fix once and forever a point $\tilde c_0\in\tilde C$ lying over $c_0\in C$.

A *noodle* in D_n is an embedded arc $N\subset D_n$ with endpoints d and d'. For a noodle N, the set $\Sigma_N=\{\{x,y\}\in C\,|\,x,y\in N, x\neq y\}$ is a surface in C containing c_0. This surface is homeomorphic to a triangle with one edge removed. We orient N from d to d' and orient Σ_N as follows: at a point $\{x,y\}=\{y,x\}\in\Sigma_N$ such that x is closer to d along N than y, the orientation of Σ_N is the product of the orientations of N at x and y in this order. Let $\tilde\Sigma_N$ be the lift of Σ_N to $\tilde C$ containing $\tilde c_0$. The orientation of Σ_N lifts to $\tilde\Sigma_N$ in the obvious way. Clearly, $\tilde\Sigma_N$ is a proper surface in $\tilde C$ in the sense that $\tilde\Sigma_N\cap\partial\tilde C=\partial\tilde\Sigma_N$.

A *fork* in D_n is an embedded tree $F\subset D$ formed by three edges and four vertices d,x_i,x_j,z such that $F\cap\partial D=d, F\cap X=\{x_i,x_j\}$ and z is a common vertex of all 3 edges. The edge, H, relating d to z is called the *handle* of F. The union, T, of the other two edges is an embedded arc with endpoints $\{x_i,x_j\}$. This arc is called the *tines* of F. Note that in a small neighborhood of z, the handle H lies on one side of T which distinguishes a side of T. We orient T so that its distinguished side lies on its right. The handle H also has a distinguished side determined by d'. Pushing slightly the graph $F=T\cup H$ to the distinguished side (fixing the vertices x_i,x_j and pushing d to d') we obtain a "parallel copy" $F'=T'\cup H'$. The graph F' is a fork with handle H', tines T', and vertices d',x_i,x_j,z' where $z'=T'\cap H'$ lies on the distinguished side of both T and H. We can assume that F' meets F only in common vertices $\{x_i,x_j\}=T\cap T'$ and in one point lying on $H\cap T'$ close to z,z'.

For a fork F, the set $\Sigma_F=\{\{y,y'\}\in C\,|\,y\in T\backslash\{x_i,x_j\}, y'\in T'\backslash\{x_i,x_j\}\}$ is a surface in C homeomorphic to $(0,1)^2$. Let α_0 be an arc from d to z along H and let α_0' be an arc from d' to z' along H'. Consider the arc $\{\alpha_0,\alpha_0'\}$ in C and denote by $\tilde\alpha$ its lift to $\tilde C$ which starts in $\tilde c_0$. Let $\tilde\Sigma_F$ be the lift of Σ_F to $\tilde C$ which contains the lift $\tilde\alpha(1)$ of the point $\{z,z'\}\in\Sigma_F$. The surfaces Σ_F and $\tilde\Sigma_F$ have a natural orientation determined by the orientation of T and the induced orientation of T'.

We shall use the surfaces $\tilde\Sigma_N,\tilde\Sigma_F$ to establish a duality between noodles and forks. More precisely, for any noodle N and any fork F we define an element $\langle N,F\rangle$ of R as follows. By applying a preliminary isotopy we can assume that N intersects T

transversely in $m \geq 0$ points $z_1, ..., z_m$ (the numeration is arbitrary; the intersection of N with H may be not transversal). We choose the parallel fork $F' = T' \cup H'$ as above so that T' meets N transversely in m points $z'_1, ..., z'_m$ where each pair z_i, z'_i is joined by a short arc in N which lies in the narrow strip bounded by $T \cup T'$ and meets no other z_j, z'_j. Then the surfaces Σ_F and Σ_N intersect transversely in m^2 points $\{z_i, z'_j\}$ where $i, j = 1, ..., m$. Therefore for any $a, b \in \mathbf{Z}$, the image $q^a t^b \tilde{\Sigma}_N$ of $\tilde{\Sigma}_N$ under the covering transformation $q^a t^b$ meets $\tilde{\Sigma}_F$ transversely. Consider the algebraic intersection number $q^a t^b \tilde{\Sigma}_N \cdot \tilde{\Sigma}_F \in \mathbf{Z}$ and set

$$\langle N, F \rangle = \sum_{a,b \in \mathbf{Z}} (q^a t^b \tilde{\Sigma}_N \cdot \tilde{\Sigma}_F)\, q^a t^b.$$

The sum on the right-hand side is finite (it has $\leq m^2$ terms) and thus defines an element of R.

2.4. Lemma. — *$\langle N, F \rangle$ is invariant under isotopies of N and F in D_n constant on the endpoints.*

Proof. — We first compute $\langle N, F \rangle$ explicitly. Let $N \cap T = \{z_1, ..., z_m\}$ and $N \cap T' = \{z'_1, ..., z'_m\}$ as above. For every pair $i, j \in \{1, ..., m\}$, there exist unique integers $a_{i,j}, b_{i,j} \in \mathbf{Z}$ such that $q^{a_{i,j}} t^{b_{i,j}} \tilde{\Sigma}_N$ intersects $\tilde{\Sigma}_F$ at a point lying over $\{z_i, z'_j\} \in C$. Let $\varepsilon_{i,j} = \pm 1$ be the sign of that intersection. Then

$$\langle N, F \rangle = \sum_{i=1}^{m} \sum_{j=1}^{m} \varepsilon_{i,j}\, q^{a_{i,j}} t^{b_{i,j}}. \tag{2.1}$$

The numbers $a_{i,j}, b_{i,j}$ can be computed as follows. Let α_0 be an arc from d to z along H and let α'_0 be an arc from d' to z' along H'. Let β_i be an arc from z to z_i along T and let β'_j be an arc from z' to z'_j along T'. Finally, let $\gamma_{i,j}$ and $\gamma'_{i,j}$ be disjoint arcs in N connecting the points z_i, z'_j to the endpoints of N. Note that $\delta_{i,j} = \{\alpha_0, \alpha'_0\}\{\beta_i, \beta'_j\}\{\gamma_{i,j}, \gamma'_{i,j}\}$ is a loop in C. Then

$$q^{a_{i,j}} t^{b_{i,j}} = \phi(\delta_{i,j}). \tag{2.2}$$

Indeed, we can lift $\delta_{i,j}$ to a path $\tilde{\alpha}\tilde{\beta}\tilde{\gamma}$ in $\tilde{C}$ beginning at $\tilde{c}_0$ where $\tilde{\alpha}$, $\tilde{\beta}$, $\tilde{\gamma}$ are lifts of $\{\alpha_0, \alpha'_0\}$, $\{\beta_i, \beta'_j\}$, $\{\gamma_{i,j}, \gamma'_{i,j}\}$, respectively. By definition of $\tilde{\Sigma}_F$, the point $\tilde{\alpha}(1) = \tilde{\beta}(0)$ lies on $\tilde{\Sigma}_F$. Hence the lift $\tilde{\beta}$ of $\{\beta_i, \beta'_j\}$ lies on $\tilde{\Sigma}_F$. The path $\tilde{\alpha}\tilde{\beta}\tilde{\gamma}$ ends at $\phi(\delta_{i,j})(\tilde{c}_0) = \tilde{\gamma}(1)$. Hence the lift $\tilde{\gamma}$ of $\{\gamma_{i,j}, \gamma'_{i,j}\}$ lies on $\phi(\delta_{i,j})\tilde{\Sigma}_N$. Therefore the point $\tilde{\beta}(1) = \tilde{\gamma}(0)$ lying over $\{z_i, z'_j\}$ belongs to both $\tilde{\Sigma}_F$ and $\phi(\delta_{i,j})\tilde{\Sigma}_N$. This yields (2.2).

Note that the residue $b_{i,j} \pmod 2$ depends on whether the two points of D_n forming a point in C switch places moving along the loop $\delta_{i,j}$. This is determined by which of z_i, z'_j lies closer to d along N.

To compute $\varepsilon_{i,j}$ we observe that $\varepsilon_{i,j}$ is determined by the signs of the intersections of N and T at z_i, z_j and by which of z_i, z'_j lies closer to d on N. The sign of the intersection of N and T at z_i is $+$ if N crosses T from the left to the right and is $-$ otherwise. By our choice of orientations on N and T, this sign is $+$ if z_i lies closer to d along N than z_i' and is $-$ otherwise. Hence, this sign is $(-1)^{b_{i,i}}$. Therefore $\varepsilon_{i,j}$ is determined by $b_{i,i} + b_{j,j} + b_{i,j} \pmod 2$. A precise computation shows that

$$\varepsilon_{i,j} = -(-1)^{b_{i,i}+b_{j,j}+b_{i,j}}. \tag{2.3}$$

Now we can prove the lemma. It suffices to fix F and to prove that $\langle N, F\rangle$ is invariant under isotopies of N. A generic isotopy of N in D_n can be split into a finite sequence of local moves of two kinds:

(i) isotopies keeping N transversal to T,

(ii) a move pushing a small subarc of N across a subarc of $T\backslash z$.

It is clear from the discussion above that the move (i) does not change $\langle N, F\rangle$. The move (ii) adds two new intersection points z_{m+1}, z_{m+2} to the set $N \cap T = \{z_1, ..., z_m\}$. It follows from definitions and the discussion above that for any $i = 1, ..., m + 2$,

$$a_{i,m+1} = a_{i,m+2}, \;\; b_{i,m+1} = b_{i,m+2}, \;\; \varepsilon_{i,m+1} = -\varepsilon_{i,m+2}.$$

Hence for any $i = 1, ..., m+2$, the terms $\varepsilon_{i,m+1}q^{a_{i,m+1}}t^{b_{i,m+1}}$ and $\varepsilon_{i,m+2}q^{a_{i,m+2}}t^{b_{i,m+2}}$ cancel each other. Similarly, for any $j = 1, ..., m$, the terms $\varepsilon_{i,j}q^{a_{i,j}}t^{b_{i,j}}$ with $i = m + 1, m + 2$ cancel each other. Therefore $\langle N, F\rangle$ is the same before and after the move.

2.5. Lemma. — *$\langle N, F\rangle = 0$ if and only if there is an isotopy $\{T(s)\}_{s\in[0,1]}$ of the tines $T = T(0)$ of F in D_n (rel endpoints) such that $T(1)$ is disjoint from N.*

Proof. — Any isotopy $\{T(s)\}_{s\in[0,1]}$ of $T = T(0)$ extends to an ambient isotopy of D_n constant on ∂D_n and therefore extends to an isotopy $\{F(s)\}_{s\in[0,1]}$ of the fork $F = F(0)$. If $T(1)$ is disjoint from N then by Lemma 2.4, $\langle N, F\rangle = \langle N, F(1)\rangle = 0$.

The hard part of the lemma is the opposite implication. By applying a preliminary isotopy, we can assume that T intersects N transversely at a *minimal* number of points $z_1, ..., z_m$ with $m \geq 0$. We assume that $m \geq 1$ and show that $\langle N, F\rangle \neq 0$. To this end we use the lexicographic ordering on monomials $q^a t^b$. Namely we write $q^a t^b \geq q^{a'} t^{b'}$ with $a, b, a', b' \in \mathbf{Z}$ if either $a > a'$ or $a = a'$ and $b \geq b'$. We say that the ordered pair (i, j) with $i, j \in \{1, ..., m\}$ is *maximal* if $q^{a_{i,j}}t^{b_{i,j}} \geq q^{a_{k,l}}t^{b_{k,l}}$ for any $k, l \in \{1, ..., m\}$. We claim that

($*$) *if the pair (i, j) is maximal, then $b_{i,i} = b_{j,j} = b_{i,j}$.*

This claim and (2.3) imply that all entries of the maximal monomial, say $q^a t^b$, in (2.1) occur with the same sign $-(-1)^b$. Hence $\langle N, F\rangle \neq 0$.

To prove ($*$) we first compute $a_{i,j}$ for any i, j (not necessarily maximal). Let ξ_i be the loop obtained by moving from d to z_i along F then back to d along N. Let a_i be the total winding number of ξ_i around all n punctures. Let ξ be the loop obtained by moving from d to d' along N, and then moving clockwise along ∂D back to d. Let a be the total winding number of ξ. We claim that

$$a_{i,j} = a_i + a_j + a. \tag{2.4}$$

Indeed, if $b_{i,j}$ is even then the paths $\alpha_0\beta_i\gamma_{i,j}$ and $\alpha'_0\beta'_j\gamma'_{i,j}$ (in the notation of Lemma 2.4) are loops and $a_{i,j}$ is the sum of their total winding numbers. These loops are homotopic in D_n to ξ_i and $\xi_j\xi$, respectively. This implies (2.4). If $b_{i,j}$ is odd then $a_{i,j}$ is the total winding number of the loop $\alpha_0\beta_i\gamma_{i,j}\alpha'_0\beta'_j\gamma'_{i,j}$. This loop is homotopic in D_n to $\xi_i\xi\xi_j$ which implies (2.4) in this case.

Suppose now that the pair (i,j) is maximal. Then $a_{i,j}$ is maximal among all the integers $a_{k,l}$. By (2.4), it follows that $a_i = a_j$ is maximal among all the integers a_k. (Although we shall not need it, observe that then $a_{i,i} = a_{j,j} = a_{i,j}$.)

We now show that $b_{i,i} = b_{i,j}$. By the maximality of (i,j), we have that $b_{i,i} \leq b_{i,j}$. Suppose, seeking a contradiction, that $b_{i,i} < b_{i,j}$. Let α be an embedded arc from z'_i to z'_j along T'. Let β be an embedded arc from z'_j to z'_i along N.

If β does not pass through the point z_i, then we denote by w the winding number of the loop $\alpha\beta$ around z_i. Observe that $b_{i,j} - b_{i,i} = 2w$. To see this, consider the loop $\delta_{i,j} = \{\alpha_0, \alpha'_0\}\{\beta_i, \beta'_j\}\{\gamma_{i,j}, \gamma'_{i,j}\}$ appearing in (2.2). Clearly, $\beta'_j \sim \beta'_i\alpha$, where $\sim$ denotes homotopy of paths in $D_n \backslash z_i$ constant on the endpoints. The assumption that β does not pass through z_i implies that $\gamma_{i,j} = \gamma_{i,i}$ and $\gamma'_{i,j} \sim \beta\gamma'_{i,i}$. Then

$$\delta_{i,j} = \{\alpha_0, \alpha'_0\}\{\beta_i, \beta'_j\}\{\gamma_{i,j}, \gamma'_{i,j}\} \sim \{\alpha_0, \alpha'_0\}\{\beta_i, \beta'_i\}\{z_i, \alpha\beta\}\{\gamma_{i,i}, \gamma'_{i,i}\}$$

where z_i denotes the constant path in z_i. This implies that $b_{i,j} - b_{i,i} = 2w$.

If β passes through z_i, we first modify β in a small neighborhood of z_i so that z_i lies to its left. Let w be the winding number of the loop $\alpha\beta$ around z_i. A little more difficult but similar argument shows that $b_{i,j} - b_{i,i} = 2w - 1$. In either case $w > 0$.

Let $D_0 = D\backslash z_i$ and $p : \hat{D}_0 \to D_0$ be the universal (infinite cyclic) covering. Let $\hat{\alpha}$ be a lift of α to $\hat{D}_0$. Let $\hat{\beta}$ be the lift of β to $\hat{D}_0$ which starts at $\hat{\beta}(1)$. Consider a small neighborhood $V \subset D$ of the short arc in N connecting z_i to z'_i such that V meets $\alpha\beta$ only at z'_i. Let γ be a generic loop in $V\backslash z_i$ based at z'_i which winds w times around z_i in the clockwise direction. Let $\hat{\gamma}$ be the lift of γ to $\hat{D}_0$ beginning at $\hat{\beta}(1)$ and ending at $\hat{\alpha}(0)$. We can assume that $\hat{\gamma}$ is an embedded arc meeting $\hat{\alpha}\hat{\beta}$ only at the endpoints. Let $\hat{z}'_k$ be the first point of $\hat{\alpha}$ which lies also on $\hat{\beta}$ (possibly $\hat{z}'_k = \hat{\alpha}(1)$). Then $p(\hat{z}'_k) = z'_k$ for some $k = 1, ..., n$. Let $\hat{\alpha}'$ be the initial segment of $\hat{\alpha}$ ending at $\hat{z}'_k$. Let $\hat{\beta}'$ be the final segment of $\hat{\beta}$ starting at $\hat{z}'_k$. Set $\hat{\delta} = \hat{\alpha}'\hat{\beta}'\hat{\gamma}$. It is clear that $\hat{\delta}$ is a simple closed curve in $\hat{D}_0$. By the Jordan curve theorem it bounds a disc, $B \subset \hat{D}_0$, which lies either on the left or on the right of $\hat{\gamma}$. Since γ passes clockwise around z_i, the component of $D_0\backslash \mathrm{Im}(\gamma)$ adjacent to z_i lies on the right of γ. The set $p(B) \subset D_0$ being compact can not pass non-trivially over this component. Hence B lies on the left of $\hat{\gamma}$. Therefore $\hat{\delta}$ passes counterclockwise around B.

The number $a_k - a_i$ is equal to the total winding number of the loop $p(\hat{\alpha}')p(\hat{\beta}')$ in D_n around the punctures $x_1, ..., x_n$. Since γ is contractible in D_n, the loop $p(\hat{\alpha}')p(\hat{\beta}')$ is homotopic in D_n to $p(\hat{\alpha}')p(\hat{\beta}')\gamma = p(\hat{\delta})$. Therefore $a_k - a_i$ is equal to the number of points in $B \cap p^{-1}(X)$. Since a_i is maximal, we must have $a_k = a_i$ and $B \cap p^{-1}(X) = \varnothing$, so that $p(B) \subset D_n \backslash z_i$. Then we can isotop T so as to have fewer points of intersection with N. To see this we shall construct a bigon for the pair (N, T). If B meets $p^{-1}(N \cup T)$ only along $\hat{\alpha}'\hat{\beta}'$ then the projection $p|_{\mathrm{Int}(B)} : \mathrm{Int}(B) \to D_n \backslash z_i$ must be injective. It follows that $w = 1$ and the union of $p(B)$ with the small disc bounded by γ in V is a bigon for (N, T). Assume that $\mathrm{Int}(B) \cap p^{-1}(N \cup T) \neq \varnothing$. Note that $p^{-1}(N)$ (resp. $p^{-1}(T)$) is an embedded one-manifold in $\hat{D}_0$ whose components are non-trivially permuted by any covering transformation. If $p^{-1}(N)$ intersects $\mathrm{Int}(B)$ then this intersection consists of a finite set of disjoint arcs with endpoints on $\hat{\alpha}'$. At least one of these arcs bounds together with a subarc of $\hat{\alpha}'$ a disc, $B_0 \subset B$, whose interior does not meet $p^{-1}(N)$. If $p^{-1}(N)$ does not meet $\mathrm{Int}(B)$ then we set $B_0 = B$. Applying the same construction to the intersection of B_0 with $p^{-1}(T)$ we obtain a bigon $B_{00} \subset B_0$ for the pair $(p^{-1}(N), p^{-1}(T))$. The restriction of p to B_{00} is injective and yields a bigon for (N, T). Hence the intersection $N \cap T$ is not minimal. This contradicts our choice of N, T. Therefore, the assumption $b_{i,i} < b_{i,j}$ must have been false. So, $b_{i,i} = b_{i,j}$. Similarly, $b_{j,j} = b_{i,j}$. This completes the proof of $(*)$ and of Lemma 2.5.

2.6. Lemma. — *If a self-homeomorphism h of D_n represents an element of* $\mathrm{Ker}(B_n \to \mathrm{Aut}(H_2(\tilde{C})))$ *then for any noodle N and any fork F, we have* $\langle N, h(F)\rangle = \langle N, F\rangle$.

Proof. — Let $\{U_i \subset \mathrm{Int}(D)\}_{i=1}^m$ be disjoint closed disc neighborhoods of the points $\{x_i\}_{i=1}^m$, respectively. Let U be the set of points $\{x, y\} \in C$ such that at least one of x, y lies in $\cup_{i=1}^m U_i$. Let $\tilde{U} \subset \tilde{C}$ be the preimage of U under the covering map $\tilde{C} \to C$. Observe that the surface $\tilde{\Sigma}_F$ is an open square such that for a sufficiently big concentric closed subsquare $S \subset \tilde{\Sigma}_F$ we have $\tilde{\Sigma}_F \backslash S \subset \tilde{U}$. Hence $\tilde{\Sigma}_F$ represents a relative homology class $[\tilde{\Sigma}_F] \in H_2(\tilde{C}, \tilde{U})$. The boundary homomorphism $H_2(\tilde{C}, \tilde{U}) \to H_1(\tilde{U})$ maps $[\tilde{\Sigma}_F]$ into $[\partial S] \in H_1(\tilde{U})$. A direct computation in $\pi_1(U)$ (see [Bi2]) shows that $(q-1)^2(qt+1)[\partial S] = 0$. Therefore $(q-1)^2(qt+1)[\tilde{\Sigma}_F] = j(v_F)$ where j is the inclusion homomorphism $H_2(\tilde{C}) \to H_2(\tilde{C}, \tilde{U})$ and $v_F \in H_2(\tilde{C})$. Deforming if necessary N, we can assume that $N \cap (\cup_i U_i) = \varnothing$. Then $\tilde{\Sigma}_N \cap \tilde{U} = \varnothing$ and therefore

$$(q-1)^2(qt+1)\langle N, F\rangle = \sum_{a,b \in \mathbf{Z}} (q^a t^b \tilde{\Sigma}_N \cdot v_F)\, q^a t^b$$

where $q^a t^b \tilde{\Sigma}_N \cdot v_F$ is the (well-defined) algebraic intersection number between a properly embedded surface and a 2-dimensional homology class. (This number does not depend on the choice of v_F as above.)

Any self-homeomorphism h of D_n is isotopic to a self-homeomorphism of D_n preserving the set U. Therefore $v_{h(F)} = \tilde{h}_*(v_F)$. If $\tilde{h}_* = \mathrm{id}$, then

$$q^a t^b \tilde{\Sigma}_N \cdot v_F = q^a t^b \tilde{\Sigma}_N \cdot \tilde{h}_*(v_F) = q^a t^b \tilde{\Sigma}_N \cdot v_{h(F)}.$$

This implies that $(q-1)^2(qt+1)\langle N, F\rangle = (q-1)^2(qt+1)\langle N, h(F)\rangle$ and therefore $\langle N, F\rangle = \langle N, h(F)\rangle$.

2.7. Deduction of Theorem 2.2 from the lemmas

We shall prove that a self-homeomorphism h of D_n representing an element of $\mathrm{Ker}(B_n \to \mathrm{Aut}(H_2(\tilde{C})))$ is isotopic to the identity map rel ∂D_n. We begin with the following assertion.

(∗∗) *An embedded arc T in D_n with endpoints in (distinct) punctures can be isotopped off a noodle N if and only if $h(T)$ can be isotopped off N.*

Indeed, we can extend T to a fork F so that T is the tines of F. By Lemma 2.6, $\langle N, h(F)\rangle = 0$ if and only if $\langle N, F\rangle = 0$. Now Lemma 2.5 implies (∗∗).

We shall apply (∗∗) to the following arcs and noodles. For $i = 1, ..., n-1$, denote by T_i the embedded arc $[x_i, x_{i+1}] \subset (-1, +1) \subset D$ and denote by N_i the i-th "elementary" noodle obtained by rushing from d towards x_i, encircling x_i in the clockwise direction and then moving straight to d'. It is clear that T_i can be isotopped off N_j if and only if $j \neq i, i+1$. This and (∗∗) imply that h induces the identity permutation on the punctures of D_n.

Since T_1 is disjoint from N_3, we can isotop h rel ∂D_n so that $h(T_1)$ is disjoint from N_3. Similarly, $h(T_1)$ can be made disjoint from N_4. As it was explained in Sect. 2.2, this can be done by a sequence of isotopies eliminating bigons for the pair $(N_4, h(T_1))$. Since N_4 and $h(T_1)$ do not meet N_3, neither do these bigons. Hence our isotopies do not create intersections of $h(T_1)$ with N_3. Repeating this argument, we can assume that $h(T_1)$ is disjoint from all N_i with $i = 3, 4, ..., n-1$. By applying one final isotopy we can make $h(T_1) = T_1$. Applying the same procedure to T_2 we can ensure that $h(T_2) = T_2$ while keeping $h(T_1) = T_1$. Continuing in this way, we can assume that $h(T_i) = T_i$ for all $i = 1, ..., n-1$. Such a homeomorphism h is isotopic to a k-th power ($k \in \mathbf{Z}$) of the Dehn twist about a circle in $\mathrm{Int}(D_n)$ going very closely to ∂D_n. This Dehn twist acts on $H_2(\tilde{C})$ by multiplication by $q^{2n}t^2$. Since by assumption h acts trivially on $H_2(\tilde{C})$, we must have $k = 0$ so that h is isotopic to the identity rel ∂D_n.

2.8. Remarks

The proof of Lemma 2.6 shows that each fork F determines (a priori non-uniquely) a certain homology class $v_F \in H_2(\tilde{C})$. It follows from the computations in [Bi2] that this class is in fact well-determined by F. Thus, the forks yield a nice geometric way of representing elements of $H_2(\tilde{C})$ (this was implicit in [Kr1]). For instance, for any $1 \leq i < j \leq n$ we can consider the fork consisting of three linear segments connecting the point $-\sqrt{-1}/2$ to d, x_i, x_j. The corresponding classes $\{v_{i,j} \in H_2(\tilde{C})\}_{i,j}$ form a basis of the free R-module $H_2(\tilde{C})$. The action of the braid generators $\sigma_1, ..., \sigma_{n-1}$ on this basis can be described by explicit formulas (see [Bi2], cf. Sect. 3.1).

3. THE WORK OF KRAMMER

3.1. A representation of B_n

Following [Kr2], we denote by Ref $=$ Ref$_n$ the set of pairs of integers (i,j) such that $1 \le i < j \le n$. Clearly, card(Ref) $= n(n-1)/2$.

Let R be a commutative ring with unit and $q, t \in R$ be two invertible elements. Let $V = \oplus_{s \in \mathrm{Ref}} R x_s$ be the free R-module of rank $n(n-1)/2$ with basis $\{x_s\}_{s \in \mathrm{Ref}}$. Krammer [Kr2] defines an R-linear action of B_n on V by

$$\sigma_k(x_{i,j}) = \begin{cases} x_{i,j} & if\ \text{k<i-1}or\text{j<k}, \\ x_{i-1,j} + (1-q)x_{i,j} & if\ \text{k=i-1}, \\ tq(q-1)x_{i,i+1} + qx_{i+1,j} & if\ \text{k=i <j-1}, \\ tq^2 x_{i,j} & if\ \text{k=i =j-1}, \\ x_{i,j} + tq^{k-i}(q-1)^2 x_{k,k+1} & if\ \text{i<k<j-1}, \\ x_{i,j-1} + tq^{j-i}(q-1)x_{j-1,j} & if\ \text{i<k=j-1}, \\ (1-q)x_{i,j} + qx_{i,j+1} & if\ \text{k=j}, \end{cases}$$

where $1 \le i < j \le n$ and $k = 1, ..., n-1$. That the action of σ_k is invertible and that relations (0.1), (0.2) are satisfied should be verified by a direct computation. For $R = \mathbf{Z}[q^{\pm 1}, t^{\pm 1}]$, this representation is equivalent to the one considered in Sect. 2. In terms of the basis $\{v_{i,j} \in H_2(\tilde{C})\}_{i,j}$ mentioned in Sect. 2.8, the equivalence is given by

$$v_{i,j} = x_{i,j} + (1-q) \sum_{i<k<j} x_{k,j}, \quad x_{i,j} = v_{i,j} + (q-1) \sum_{i<k<j} q^{k-1-i} v_{k,j}.$$

3.2. Theorem (D. Krammer [Kr2]). — *Let $R = \mathbf{R}[t^{\pm 1}]$, $q \in \mathbf{R}$, and $0 < q < 1$. Then the representation $B_n \to \mathrm{Aut}(V)$ defined in Sect. 3.1 is faithful for all $n \ge 1$.*

This theorem implies Theorem 2.2: if a representation over $\mathbf{Z}[q^{\pm 1}, t^{\pm 1}]$ becomes faithful after assigning a real value to q, then it is faithful itself.

Below we outline the main ideas of Krammer's proof.

3.3. Positive braids and the set $\Omega \subset B_n$

We recall a few facts about the braid group B_n, see [Ch], [Ga], [Mi]. For $i = 1, 2, ..., n-1$ denote by s_i the transposition $(i, i+1) \in S_n$. The set $\{s_1, ..., s_{n-1}\}$ generates the symmetric group S_n. Let $|\ | : S_n \to \mathbf{Z}$ be the length function with respect to this generating set: for $x \in S_n$, $|x|$ is the smallest natural number k such that x is a product of k elements of the set $\{s_1, ..., s_{n-1}\}$. The canonical projection $B_n \to S_n$ has a unique set-theoretic section $r : S_n \to B_n$ such that $r(s_i) = \sigma_i$ for $i = 1, ..., n-1$ and $r(xy) = r(x)\, r(y)$ whenever $|xy| = |x| + |y|$. The group B_n admits a presentation by generators $\{r(x)\,|\, x \in S_n\}$ and relations $r(xy) = r(x)\, r(y)$ for all $x, y \in S_n$ such that $|xy| = |x| + |y|$. Set $\Omega = r(S_n) \subset B_n$. Note that $1 = r(1) \in \Omega$ and $\sigma_i = r(s_i) \in \Omega$ for all i.

The *positive braid monoid* B_n^+ is the submonoid of B_n generated by $\{\sigma_1, ..., \sigma_{n-1}\}$. The elements of B_n^+ are called *positive braids*. Clearly, $\Omega \subset B_n^+$.

For any $x \in B_n^+$ there is a unique longest $x' \in S_n$ such that $x \in r(x')B_n^+$. We denote $r(x') \in \Omega$ by $LF(x)$ where LF stands for the leftmost factor. Observe that

$$LF(xy) = LF(x\, LF(y)) \tag{3.1}$$

for any $x, y \in B_n^+$. This implies that the map $B_n^+ \times \Omega \to \Omega$ defined by $(x, y) \mapsto LF(xy)$ is an action of the monoid B_n^+ on Ω.

3.4. Half-permutations

A set $A \subset \mathrm{Ref}$ is called a *half-permutation* if whenever $(i, j), (j, k) \in A$, one has $(i, k) \in A$. Each half-permutation A determines an ordering $<_A$ on the set $\{1, 2, ..., n\}$ by $i <_A j \Leftrightarrow (i, j) \in A$ (and vice versa).

To state deeper properties of half-permutations we consider the set 2^{Ref} of all subsets of Ref and define a map $L : S_n \to 2^{\mathrm{Ref}}$ by

$$L(x) = \{(i, j) \,|\, 1 \leq i < j \leq n,\ x^{-1}(i) > x^{-1}(j)\} \subset \mathrm{Ref}.$$

Note that the set $L(x)$ is a half-permutation and $\mathrm{card}(L(x)) = |x|$. It is obvious that the map $L : S_n \to 2^{\mathrm{Ref}}$ is injective.

The key property of half-permutations is the following assertion ([Kr2, Lemma 4.3]): for every half-permutation $A \subset \mathrm{Ref}$ there is a greatest set $A' \subset A$ (with respect to inclusion) such that $A' = L(x)$ for a certain $x \in S_n$. The corresponding braid $r(x) \in \Omega$ is denoted by $GB(A)$ where GB stands for the greatest braid. This defines a map GB from the set of half-permutations to Ω. In particular, for any $x \in S_n$ we have $GB(L(x)) = r(x)$.

3.5. Actions of B_n of 2^{Ref} and on half-permutations

Let $R = \mathbf{R}[t^{\pm 1}]$, $q \in \mathbf{R}$, and $0 < q < 1$. The action of B_n on V defined in Sect. 3.1 has the following property: for any positive braid $x \in B_n^+$ the entries of the matrix of the map $v \mapsto xv : V \to V$ belong to $\mathbf{R}_{\geq 0} + t\mathbf{R}[t]$. (This is obvious for the generators $\sigma_1, ..., \sigma_{n-1}$ of B_n^+.) Therefore the action of B_n^+ preserves the set

$$W = \bigoplus_{s \in \mathrm{Ref}} (\mathbf{R}_{\geq 0} + t\mathbf{R}[t])\, x_s \subset V.$$

For a set $A \subset \mathrm{Ref}$, define W_A as the subset of W consisting of vectors $\sum_{s\in\mathrm{Ref}} k_s x_s$ with $k_s \in \mathbf{R}_{\geq 0} + t\mathbf{R}[t]$ such that $k_s \in t\mathbf{R}[t] \Leftrightarrow s \in A$. Clearly, W is the disjoint union of the sets W_A corresponding to various $A \subset \mathrm{Ref}$. For any $x \in B_n^+$ and $A \subset \mathrm{Ref}$ there is a unique $B \subset \mathrm{Ref}$ such that $xW_A \subset W_B$. We denote this set B by xA. This defines an action of B_n^+ on 2^{Ref}. By [Kr2, Lemma 4.2], this action maps half-permutations to half-permutations. This defines an action of B_n^+ on the set of half-permutations. Finally, Krammer observes that the map GB from this set to Ω is B_n^+-equivariant. Thus, for any positive braid $x \in B_n^+$ and a half-permutation $A \subset \mathrm{Ref}$, we have

$$GB(xA) = LF(x\, GB(A)). \tag{3.2}$$

The rest of the argument is contained in the following two lemmas.

3.6. Lemma. — *Let B_n act on a set U. Suppose we are given non-empty disjoint sets $\{C_x \subset U\}_{x\in\Omega}$ such that $xC_y \subset C_{LF(xy)}$ for all $x, y \in \Omega$. Then the action of B_n on U is faithful.*

Proof. — Denote by ℓ the group homomorphism $B_n \to \mathbf{Z}$ mapping $\sigma_1, ..., \sigma_{n-1}$ to 1. We check first that the inclusion $xC_y \subset C_{LF(xy)}$ holds for all $x \in B_n^+$ and $y \in \Omega$. Clearly, $\ell(B_n^+) \geq 0$ so that we can use induction on $\ell(x)$. If $\ell(x) = 0$ then $x = 1$ and the inclusion follows from the equality $LF(y) = y$. Let $\ell(x) \geq 1$. Then $x = \sigma_i v$ where $i = 1, ..., n-1$ and $v \in B_n^+$. Clearly, $\ell(v) = \ell(x) - 1$. We have

$$xC_y = \sigma_i v C_y \subset \sigma_i(C_{LF(vy)}) \subset C_{LF(\sigma_i LF(vy))} = C_{LF(\sigma_i vy)} = C_{LF(xy)}.$$

Here the first inclusion follows from the induction hypothesis, the second inclusion follows from the assumptions of the lemma, the middle equality follows from (3.1).

It is known that for any $b \in B_n$ there are $x, y \in B_n^+$ such that $b = xy^{-1}$. Therefore to prove the lemma it suffices to show that, if two elements $x, y \in B_n^+$ act in the same way on U, then $x = y$. We will show this by induction on $\ell(x) + \ell(y)$. If $\ell(x) + \ell(y) = 0$, then $x = y = 1$. Assume that $\ell(x) + \ell(y) > 0$. By assumption, C_1 is non-empty; choose any $u \in C_1$. We have $xu \in xC_1 \subset C_{LF(x)}$ and similarly $yu \in C_{LF(y)}$. Hence $xu = yu \in C_{LF(x)} \cap C_{LF(y)}$. By the disjointness assumption, this is possible only if $LF(x) = LF(y)$. Write $z = LF(x) \in \Omega$ and consider $x', y' \in B_n^+$ such that $x = zx'$ and $y = zy'$. We have $z \neq 1$ since otherwise $x = y = 1$. Then

$\ell(z) > 0$ and $\ell(x') + \ell(y') < \ell(x) + \ell(y)$. The induction assumption yields that $x' = y'$. Therefore $x = y$. This proves the inductive step and the lemma.

3.7. Lemma. — *For $x \in \Omega$, set*

$$C_x = \bigcup_{A \in GB^{-1}(x)} W_A \ \subset V$$

where A runs over all half-permutations such that $GB(A) = x$. Then the sets $\{C_x\}_{x\in\Omega}$ satisfy all the conditions of Lemma 3.6. *Therefore the action of B_n on V is faithful.*

Proof. — It is obvious that the sets $\{C_x\}_{x\in\Omega}$ are disjoint. The set C_x is non-empty because $\varnothing \neq W_{L(r^{-1}(x))} \subset C_x$.

To prove the inclusion $xC_y \subset C_{LF(xy)}$ for $x, y \in \Omega$, it suffices to prove that $xW_A \subset C_{LF(xy)}$ whenever A is a half-permutation such that $GB(A) = y$. This follows from the inclusions

$$xW_A \subset W_{xA} \subset C_{GB(xA)} = C_{LF(xy)}.$$

Here the first inclusion follows from the definition of the B_n^+-action of Ref. The second inclusion follows from the definition of $C_{GB(xA)}$. The equality follows from (3.2).

3.8. More about the representation

Let $\ell_\Omega : B_n \to \mathbf{Z}$ be the length function with respect to the generating set Ω: for $x \in B_n$, $\ell_\Omega(x)$ is the minimal natural number k such that $x = x_1 ... x_k$ where $x_i \in \Omega$ or $x_i^{-1} \in \Omega$ for each $i = 1, ..., k$. Among other related results, Krammer gives an explicit computation of ℓ_Ω in terms of his representation. Namely, take the Laurent polynomial ring $R = \mathbf{Z}[q^{\pm 1}, t^{\pm 1}]$ as the ground ring and denote by ρ Krammer's representation $B_n \to \mathrm{Aut}(V)$ defined in Sect. 3.1. For $x \in B_n$, consider the Laurent expansion $\rho(x) = A_k t^k + A_{k+1} t^{k+1} + ... + A_l t^l$ where $\{A_i\}_{i=k}^{l}$ are $(m \times m)$-matrices over $\mathbf{Z}[q^{\pm 1}]$ and $A_k \neq 0, A_l \neq 0$. Then

$$\ell_\Omega(x) = \max(l - k, l, -k).$$

This formula yields another proof of the faithfulness of ρ: If $x \in \mathrm{Ker}\rho$, then $k = l = 0$ so that $\ell_\Omega(x) = 0$ and $x = 1$.

The length function ℓ_Ω was first considered by Charney [Ch] who proved that the formal power series $\sum_{x\in B_n} z^{\ell_\Omega(x)} \in \mathbf{Z}[[z]]$ is a rational function. It is unknown whether the similar formal power series determined by the length function with respect to the generators $\sigma_1, ..., \sigma_{n-1}$ is rational.

4. BMW-ALGEBRAS AND REPRESENTATIONS OF B_n

4.1. Hecke algebras and representations of B_n

A vast family of representations of B_n including the Burau representation arise from a study of Hecke algebras. The Hecke algebra $H_n(\alpha)$ corresponding to $\alpha \in \mathbf{C}$ can be defined as the quotient of the complex group ring $\mathbf{C}[B_n]$ by the relations $\sigma_i^2 = (1-\alpha)\sigma_i + \alpha$ for $i = 1, ..., n-1$. This family of finite dimensional $\mathbf{C}$-algebras is a one-parameter deformation of $\mathbf{C}[S_n] = H_n(1)$. For α sufficiently close to 1, the algebra $H_n(\alpha)$ is isomorphic to $\mathbf{C}[S_n]$. For such α, the algebra $H_n(\alpha)$ is semisimple and its irreducible representations are indexed by the Young diagrams with n boxes. Their decomposition rules and dimensions are the same as for the irreducible representations of S_n.

Each representation of $H_n(\alpha)$ yields a representation of B_n via the natural projection $B_n \subset \mathbf{C}[B_n] \to H_n(\alpha)$. This gives a family of irreducible finite dimensional representations of B_n indexed by the Young diagrams with n boxes. These representations were extensively studied by V. Jones [Jo]. In particular, he observed that the $(n-1)$-dimensional Burau representation of B_n appears as the irreducible representation associated with the two-column Young diagram whose columns have $n-1$ boxes and one box, respectively.

4.2. Birman-Murakami-Wenzl algebras and their representations

Jun Murakami [Mu] and independently J. Birman and H. Wenzl [BW] introduced a two-parameter family of finite dimensional $\mathbf{C}$-algebras $C_n(\alpha, l)$ where α and l are non-zero complex numbers such that $\alpha^4 \neq 1$ and $l^4 \neq 1$. For $i = 1, ..., n-1$, set

$$e_i = (\alpha + \alpha^{-1})^{-1}(\sigma_i + \sigma_i^{-1}) - 1 \in \mathbf{C}[B_n].$$

The algebra $C_n(\alpha, l)$ is the quotient of $\mathbf{C}[B_n]$ by the relations

$$e_i\sigma_i = l^{-1}e_i, \quad e_i\sigma_{i-1}^{\pm 1}e_i = l^{\pm 1}e_i$$

for all i. (The original definition in [BW] involves more relations; for the shorter list given above, see [We].) The algebra $C_n(\alpha, l)$ admits a geometric interpretation in terms of so-called Kauffman skein classes of tangles in Euclidean 3-space. This family of algebras is a deformation of an algebra introduced by R. Brauer [Br] in 1937.

The algebraic structure and representations of $C_n(\alpha, l)$ were studied by Wenzl [We], who established (among other results) the following three facts.

(i) *For generic α, l, the algebra $C_n(\alpha, l)$ is semisimple.*

Here "generic" means that α is not a root of unity and $\sqrt{-1}\, l$ is not an integer power of $-\sqrt{-1}\,\alpha$. (The latter two numbers correspond to r and q in Wenzl's notation). In the sequel we assume that α, l are generic in this sense. We denote the number of boxes in a Young diagram λ by $|\lambda|$.

(ii) *The irreducible finite dimensional $C_n(\alpha, l)$-modules are indexed by the Young diagrams λ such that $|\lambda| \leq n$ and $|\lambda| \equiv n \,(\mathrm{mod}\, 2)$.*

The irreducible $C_n(\alpha, l)$-module corresponding to λ will be denoted by $V_{n,\lambda}$. Composing the natural projection $B_n \subset \mathbf{C}[B_n] \to C_n(\alpha, l)$ with the action of $C_n(\alpha, l)$ on $V_{n,\lambda}$ we obtain an irreducible representation of B_n.

Observe that the inclusion $B_{n-1} \hookrightarrow B_n$ sending each $\sigma_i \in B_{n-1}$ with $i = 1, ...,$ $n-2$ to $\sigma_i \in B_n$ induces an inclusion $C_{n-1}(\alpha, l) \hookrightarrow C_n(\alpha, l)$ for all $n \geq 2$.

(iii) *The $C_n(\alpha, l)$-module $V_{n,\lambda}$ decomposes as a $C_{n-1}(\alpha, l)$-module into a direct sum $\oplus_\mu V_{n-1,\mu}$ where μ ranges over all Young diagrams obtained by removing or (if $|\lambda| < n$) adding one box to λ. Each such μ appears in this decomposition with multiplicity 1.*

4.3. The Bratelli diagram for the BMW-algebras

The assertions (ii) and (iii) in Sect. 4.2 allow us to draw the Bratelli diagram for the sequence $C_1(\alpha, l) \subset C_2(\alpha, l) \subset ...$ On the level $n = 1, 2, ...$ of the Bratelli diagram one puts all Young diagrams λ such that $|\lambda| \leq n$ and $|\lambda| \equiv n \,(\mathrm{mod}\, 2)$. Then one connects by an edge each λ on the n-th level to all Young diagrams on the $(n-1)$-th level obtained by removing or (if $|\lambda| < n$) adding one box to λ. For instance, the $n = 1$ level consists of a single Young diagram with one box corresponding to the tautological one-dimensional representation of $C_1(\alpha, l) = \mathbf{C}$. The $n = 2$ level contains the empty Young diagram and two Young diagrams with two boxes. All three are connected by an edge to the diagram on the level 1. Note that every Young diagram λ appears on the levels $|\lambda|, |\lambda| + 2, |\lambda| + 4, ...$

The Bratelli diagram yields a useful method of computing the dimension of $V_{n,\lambda}$ where λ is a Young diagram on the n-th level. It is clear from (iii) that $\dim(V_{n,\lambda})$ is the number of paths on the Bratelli diagram leading from λ to the only diagram on the level 1. Here by a path we mean a path with vertices lying on consecutively decreasing levels. We give three examples of computations based on (iii).

(a) Let λ_n be the Young diagram represented by a column of n boxes. There is only one path from λ_n, positioned on the level n, to the top of the Bratelli diagram. Hence, $\dim(V_{n,\lambda_n}) = 1$ for all $n \geq 1$.

For $n \geq 2$, the algebra $C_n(\alpha, l)$ has two one-dimensional representations. In both of them all e_i act as 0 and all σ_i act as multiplication by one and the same number equal either to α or to α^{-1}. We choose the correspondence between the irreducible $C_n(\alpha, l)$-modules and the Young diagrams so that all σ_i act on V_{n,λ_n} as multiplication by α. If λ_n^T is the Young diagram represented by a row of n boxes, then similarly to (a) we have that $\dim(V_{n,\lambda_n^T}) = 1$ and all σ_i act on V_{n,λ_n^T} as multiplication by α^{-1}.

(b) For $n \geq 2$, let λ'_n be the two-column Young diagram whose columns have $n-1$ boxes and one box, respectively. For $n \geq 3$, the diagram λ'_n, positioned on the level n, is connected to only two Young diagrams on the previous level, namely, to λ'_{n-1}

and λ_{n-1}. Hence

$$\dim(V_{n,\lambda'_n}) = \dim(V_{n-1,\lambda'_{n-1}}) + \dim(V_{n-1,\lambda_{n-1}}) = \dim(V_{n-1,\lambda'_{n-1}}) + 1.$$

We have $\lambda'_2 = \lambda_2^T$ so that $\dim(V_{2,\lambda'_2}) = 1$. Hence $\dim(V_{n,\lambda'_n}) = n-1$ for all $n \geq 2$.

(c) For $n \geq 2$, consider the module $V_{n,\lambda_{n-2}}$ corresponding to the Young diagram λ_{n-2} positioned on the level n. If $n \geq 3$, then this diagram is connected to three Young diagrams on the previous level, namely, to $\lambda_{n-1}, \lambda_{n-3}, \lambda'_{n-1}$. Hence

$$\begin{aligned}\dim(V_{n,\lambda_{n-2}}) &= \dim(V_{n-1,\lambda_{n-1}}) + \dim(V_{n-1,\lambda_{n-3}}) + \dim(V_{n-1,\lambda'_{n-1}}) \\ &= \dim(V_{n-1,\lambda_{n-3}}) + n - 1.\end{aligned}$$

We gave $\lambda_0 = \varnothing$ and by (iii) above, $\dim(V_{2,\lambda_0}) = \dim(V_{1,\lambda_1}) = 1$. Thus for all $n \geq 2$,

$$\dim(V_{n,\lambda_{n-2}}) = n(n-1)/2,$$

i.e., $V_{n,\lambda_{n-2}}$ has the same dimension as the Krammer representation of B_n. We now rescale the representation $B_n \to \mathrm{Aut}(V_{n,\lambda_{n-2}})$ by dividing the action of each σ_i by α.

4.4. Theorem (M. Zinno [Zi]). — *The Krammer representation corresponding to* $q = -\alpha^{-2}$ *and* $t = \alpha^3 l^{-1}$ *is isomorphic to the rescaled representation* $B_n \to \mathrm{Aut}(V_{n,\lambda_{n-2}})$.

The proof given in [Zi] goes by a direct comparison of both actions of B_n on certain bases. Theorem 4.4 implies that the Krammer-Bigelow representation considered in Sect. 2 and 3 is irreducible.

REFERENCES

[Bi1] S. Bigelow – *The Burau representation is not faithful for $n \geq 5$*, Preprint `math.GT/9904100`.

[Bi2] S. Bigelow – *Braid groups are linear*, Preprint `math.GR/0005038`.

[Bir] J. S. Birman – *Braids, links, and mapping class groups*, Ann. of Math. Stud., vol. 82, Princeton Univ. Press, Princeton, N.J., 1974.

[BW] J. S. Birman, H. Wenzl – *Braids, link polynomials and a new algebra*, Trans. Amer. Math. Soc. **313** (1989), 249–273.

[Br] R. Brauer – *On algebras which are connected with the semisimple continuous groups*, Ann. of Math. **38** (1937), 857–872.

[Bu] W. Burau – *Über Zopfgruppen und gleichsinnig verdrillte Verkettungen*, Abh. Math. Semin. Hamburg. Univ. **11** (1935), 179-186.

[Ch] R. Charney – *Geodesic automation and growth functions for Artin groups of finite type*, Math. Ann. **301** (1995), 307–324.

[DFG] J. L. Dyer, E. Formanek, E. K. Grossman – *On the linearity of automorphism groups of free groups*, Arch. Math. (Basel) **38** (1982), 404–409.

[FLP] A. Fathi, F. Laudenbach, V. Poenaru – *Travaux de Thurston sur les surfaces*, Astérisque **66–67** (1991), Soc. Math. France, Paris.

[FP] E. Formanek, C. Procesi – *The automorphism group of a free group is not linear*, J. Algebra **149** (1992), 494–499.

[Ga] F. A. Garside – *The braid group and other groups*, Quart. J. Math. Oxford **20** (1969), 235–254.

[Iv1] N. V. Ivanov – *Automorphisms of Teichmüller modular groups*, Topology and geometry – Rohlin Seminar, Lecture Notes in Math., vol. 1346, Springer, Berlin-New York (1988), 199–270.

[Iv2] N. V. Ivanov – *Subgroups of Teichmüller modular groups*, Transl. of Math. Monographs, vol. 115. Amer. Math. Soc., Providence, RI, 1992.

[Jo] V. F. R. Jones – *Hecke algebra representations of braid groups and link polynomials*, Ann. of Math. **126** (1987), 335–388.

[Ka] C. Kassel – *L'ordre de Dehornoy sur les tresses*, Séminaire Bourbaki, exposé 865 (novembre 1999). This volume.

[Kr1] D. Krammer – *The braid group B_4 is linear*, Preprint, Basel (1999).

[Kr2] D. Krammer – *Braid Groups are linear*, Preprint, Basel (2000).

[La] R. J. Lawrence – *Homological representations of the Hecke algebra*, Comm. Math. Phys. **135** (1990), 141–191.

[LP] D. D. Long, M. Paton – *The Burau representation is not faithful for $n \geq 6$*, Topology **32** (1993), 439–447.

[Mi] J. Michel – *A note on words in braid monoids*, J. Algebra **215** (1999), 366–377.

[Mo] J. A. Moody – *The Burau representation of the braid group B_n is unfaithful for large n*, Bull. Amer. Math. Soc. (N.S.) **25** (1991), 379–384.

[Mu] J. Murakami – *The Kauffman polynomial of links and representation theory*, Osaka J. Math. **24** (1987), 745–758.

[We] H. Wenzl – *Quantum groups and subfactors of type B, C, and D*, Comm. Math. Phys. **133** (1990), 383–432.

[Zi] M. G. Zinno – *On Krammer's Representation of the Braid Group*, Preprint `math.RT/0002136`.

Vladimir TURAEV

Institut de Recherche Mathématique Avancée
Université Louis Pasteur et C.N.R.S.
7 rue René Descartes
F-67084 STRASBOURG Cedex
E-mail : `turaev@math.u-strasbg.fr`

Séminaire BOURBAKI Juin 2000
52^{e} année, 1999-2000, n^{o} 879, p. 411 à 433

RANDOM MATRICES AND PERMUTATIONS, MATRIX INTEGRALS AND INTEGRABLE SYSTEMS

by **Pierre VAN MOERBEKE**

Contents

The support of a National Science Foundation grant # DMS-98-4-50790, a Nato, a FNRS and a Francqui Foundation grant is gratefully acknowledged.

0. INTRODUCTION

The purpose of this lecture is to give a survey of recent interactions between the theory of random matrices, the theory of random permutations and so-called integrable models. The latter are described either by integrals over tangent spaces to symmetric spaces, or by integrals over classical groups.

Going back in time, t'Hooft was led, in the 70's, to so-called matrix models in order to understand the behavior of quantum gauge theories with gauge group $SU(N)$ when N gets large. In their pioneering work, Bessis, Itzykson and Zuber [8] considered the integral $Z_N^{(\varepsilon)}$ taken over the space $\mathcal{H}_N$ of Hermitian matrices of size N, whose log has the following expansion:

$$\log Z_N^{(\varepsilon)} = \log \frac{\int_{\mathcal{H}_N} dM\, e^{-\frac{1}{2}trM^2 - \frac{\varepsilon}{N}trM^4}}{\int_{\mathcal{H}_N} dM\, e^{-\frac{1}{2}trM^2}} = \sum_{g=0}^{\infty} N^{2-2g} \sum_{k=0}^{\infty} (-\varepsilon)^k W(g,k)$$

where $W(g,n)$ is the number of graphs with k vertices drawn on a surface of genus g, or in a dual language, the number of ways to cover a surface of genus g with k squares. Replacing $\frac{\varepsilon}{N}\operatorname{tr} M^4$ with $\sum_{i\geq 1} t_i \operatorname{tr} M^i$ leads to coverings of surfaces of genus g by n-gons of various n. The sequence of such integrals over N is a solution of the standard Toda lattice equations.

Two-matrix integrals (Chadha-Mahoux-Mehta [11], Itzykson-Zuber [20])

$$\int\int_{\mathcal{H}_N \times \mathcal{H}_N} dM_1 dM_2 e^{-tr(M_1^2 + M_2^2 - cM_1M_2 + \alpha M_1^4 + \beta M_2^4)}$$

provide a solution to the 2d-Toda lattice. They describe an Ising model on a random lattice, with spin $\sigma = \pm 1$ at each site with nearest neighbor interaction; the interaction between two sites depends on whether the spins are equal or not.

Kontsevich [25] proves that $Z_N(t)$ is a τ-function for the KdV equation for N large, with expansion (set $t_i = -c_i \operatorname{tr} Z^{-i}$ for appropriate integer c_i's)

$$\begin{aligned}
\log Z_N(t) &= \log \frac{\int_{\mathcal{H}_N} dX e^{-\frac{1}{2}tr\, X^2 Z + \frac{1}{6}tr\, X^3}}{\int_{\mathcal{H}_N} dX\, e^{-\frac{1}{2}tr\, X^2 Z}} \\
&= \sum_{\Gamma \in G_N} \frac{(\frac{1}{2})\#\text{vertices in } \Gamma}{\#\operatorname{Aut}\Gamma} \prod_{\text{ribbons } \alpha,\, \beta} \frac{2}{z_\alpha + z_\beta} \\
&= \sum_{n\geq 1, d_1,\dots,d_n \geq 0} \frac{t_{2d_1+1}\dots t_{2d_n+1}}{n!} \int_{\bar{\mathcal{M}}_{g,n}} \prod_{i=1}^{n} c_1(\mathcal{L}_i)^{d_i},
\end{aligned}$$

where G_N are all non-equivalent ribbon graphs Γ with N distinct loops obtained by picking vertices from which emerge 3 ribbons, interconnecting them and associating a z_α with each edge $(1 \leq \alpha \leq N)$. Each ribbon has two edges, one going with some z_α and another going with some z_β $(1 \leq \alpha, \beta \leq N)$. The product in the formula above is taken over all such ribbons of the graph. The second formula involves Chern classes $c_1(\mathcal{L}_i)$ of line bundles on the Deligne-Mumford smooth compactification $\overline{\mathcal{M}}_{g,n}$ of the moduli space $\mathcal{M}_{g,n}$ of smooth Riemann surfaces of genus g with n distinct marked points $x_1, \ldots, x_n$; the fibers of the line bundle are given by $T^*_{x_i}\mathcal{C}$ at each point $(\mathcal{C}, x_1, \ldots, x_n) \in \overline{\mathcal{M}}_{g,n}$.

Recently, other integrals have arisen, like

$$\int_{O(n)} e^{xtrM} dM, \quad \int_{U(n)} e^{\sqrt{x}\,\mathrm{tr}(M+\bar{M})} dM, \quad \int_{U(n)} \det(I+M)^k e^{-x\ tr\bar{M}} dM,$$

whose expansion in x relate to problems in combinatorics. The integrals over the space of Hermitian, symmetric and symplectic matrices, namely

$$\int_{\mathcal{H}_n,\ \mathcal{S}_n \text{ or } \mathcal{T}_{2n}} e^{-trV(M)} dM,$$

play a prominent role in theory of random matrices. These two sets of integrals provide the main thrust of this lecture. Sequences of such integrals (in n) provide solutions to integrable lattices. For a more detailed overview on these questions, see the MSRI lectures [39]. The last section contains a table listing these connections between matrix integrals, moment matrices and integrable lattices.

1. LARGEST INCREASING SEQUENCES IN RANDOM PERMUTATIONS

Let S_N be the group of permutations π_N, equipped with the uniform probability distribution:

$$P(\pi_N) = 1/N!. \tag{1.0.1}$$

An *increasing subsequence* of $\pi_N \in S_N$ is a sequence $1 \le j_1 < \ldots < j_k \le N$, such that $\pi(j_1) < \ldots < \pi(j_k)$. Define

$$L_N(\pi_N) = \text{ length of the longest increasing subsequence of } \pi_N\ . \tag{1.0.2}$$

Problem. — Find the probability $P(L_N \le n)$.

Examples. — For $\pi_7 = (\underline{3}, 1, \underline{4}, 2, \underline{6}, \underline{7}, 5)$, we have $L(\pi_7) = 4$. For $\pi_5 = (5, \underline{1}, \underline{4}, 3, 2)$, we have $L(\pi_5) = 2$.

1.1. Robinson-Schensted-Knuth correspondence and symmetric functions

I give here a very brief summary of well-known, but necessary facts to approach the problem:

– A *Young diagram* λ is a finite sequence of non-increasing, non-negative integers $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_\ell \geq 0$; also called a *partition* of $n = |\lambda| := \lambda_1 + ... + \lambda_\ell$, with $|\lambda|$ being the weight. It can be represented by a diagram, having λ_1 boxes in the first row, λ_2 boxes in the second row, etc., all aligned to the left. A *dual Young diagram* $\hat{\lambda} = (\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq ...)$ is the diagram obtained by flipping the diagram λ about its diagonal; thus $\hat{\lambda}_1 = \ell$ =length of first column of λ.

– A *Young tableau* of shape λ is an array of positive integers a_{ij} (at place (i,j) in the Young diagram) placed in the Young diagram λ, which are non-decreasing from left to right *and* strictly increasing from top to bottom.

– A *standard Young tableau* of shape λ is an array of integers $1,...,n$ placed in the Young diagram, which are strictly increasing from left to right *and* from top to bottom. The number of Young tableaux of a given shape $\lambda = (\lambda_1 \geq ... \geq \lambda_m)$ is given by a number of formulae (for the Schur polynomial s_λ, see below)[(1)]

$$\begin{aligned} f^\lambda &= \#\{\text{standard tableaux of shape } \lambda\} \\ &= \text{coefficient of } x_1 x_2 \dots x_n \text{ in } s_\lambda(x) \\ &= \frac{|\lambda|!}{\prod_{\text{all } i,j} h^\lambda_{ij}} = |\lambda|! \det\left(\frac{1}{(\lambda_i - i + j)!}\right) \\ &= |\lambda|! \prod_{1\leq i<j\leq m} (h_i - h_j) \prod_1^m \frac{1}{h_i!}, \quad \text{with } h_i := \lambda_i - i + m,\ m := \hat{\lambda}_1. \end{aligned} \tag{1.1.1}$$

– The *Schur polynomial* s_λ associated with a Young diagram λ is a symmetric function in the variables $x_1, x_2, ...$ (finite or infinite), defined by

$$s_\lambda(x_1, x_2, ...) := \sum_{\{a_{ij}\} \text{ tableaux of } \lambda} \prod_{ij} x_{a_{ij}}. \tag{1.1.2}$$

– The linear *space* Λ_n *of symmetric polynomials* in $x_1, ..., x_n$ with rational coefficients comes equipped with an inner product, which can also be expressed as an integral over the unitary group $U(n)$ for Haar measure dM:

$$\begin{aligned} \langle f, g\rangle &= \frac{1}{n!} \int_{(S_1)^n} f(z_1, ..., z_n) g(\bar{z}_1, ..., \bar{z}_n) \prod_{1\leq k<\ell\leq n} |z_k - z_\ell|^2 \prod_1^n \frac{dz_k}{2\pi i z_k} \\ &= \int_{U(n)} f(M) g(\bar{M}) dM. \end{aligned} \tag{1.1.3}$$

– An *orthonormal basis of the space* Λ_n is given by the Schur polynomials above $s_\lambda(x_1, ..., x_n)$, in which the numbers a_{ij} are restricted to $1, ..., n$; therefore we have the following "*Fourier series*":

$$f(x_1, ..., x_n) = \sum_{\substack{\lambda \text{ with} \\ \hat{\lambda}_1 \leq n}} \langle f, s_\lambda\rangle\, s_\lambda(x_1, ..., x_n), \text{ with } \langle s_\lambda, s_{\lambda'}\rangle = \delta_{\lambda\lambda'}. \tag{1.1.4}$$

[(1)] $h^\lambda_{ij} := \lambda_i + \hat{\lambda}_j - i - j + 1$ is the *hook length* of the i,jth box in the Young diagram.

In particular, one computes the following Fourier series:

$$(x_1 + \cdots + x_n)^k = \sum_{\substack{|\lambda|=k \\ \hat{\lambda}_1 \leq n}} f^\lambda s_\lambda. \tag{1.1.5}$$

If $\lambda = (\lambda_1 \geq ... \geq \lambda_\ell)$, with[(2)] $\hat{\lambda}_1 = \ell > n$, then obviously $s_\lambda = 0$.

– *Robinson-Schensted-Knuth correspondence*: There is a 1-1 correspondence

$$S_n \longrightarrow \left\{ \begin{array}{l} (P, Q), \text{two standard Young} \\ \text{tableaux from } 1, \ldots, n, \text{ where} \\ P \text{ and } Q \text{ have the same shape} \end{array} \right\}. \tag{1.1.6}$$

Given a permutation $i_1, ..., i_n$, the correspondence constructs two standard Young tableaux P, Q having the same shape λ. This construction is inductive. Namely, having obtained two equally shaped Young diagrams P_k, Q_k from $i_1, ..., i_k$, with the numbers $(i_1, ..., i_k)$ in the boxes of P_k and the numbers $(1, ..., k)$ in the boxes of Q_k, one creates a new diagram Q_{k+1}, by putting the *next number* i_{k+1} *in the first row of* P, according to the following rule:

(i) if $i_{k+1} \geq$ all numbers appearing in the first row of P_k, then one creates a new box with i_{k+1} in that box to the right of the first column,

(ii) if not, place i_{k+1} in the box (of the first row) with the smallest higher number. That number then gets pushed down to the second row of P_k according to the rule (i) or (ii), as if the first row had been removed.

The diagram Q is a bookkeeping device; namely, add a box (with the number $k + 1$ in it) to Q_k exactly at the place, where the new box has been added to P_k. This produces a new diagram Q_{k+1} of same shape as P_{k+1}.

The inverse of this map is constructed essentially by reversing the steps above.

Example. — $\pi = (5, 1, 4, 3, 2) \in S_5$,

$$\begin{array}{ccccc}
\begin{array}{ll} 5 & \\ & \\ & \\ & \end{array} &
\begin{array}{ll} 1 & \\ 5 & \\ & \\ & \end{array} &
\begin{array}{ll} 1 & 4 \\ 5 & \\ & \\ & \end{array} &
\begin{array}{ll} 1 & 3 \\ 4 & \\ 5 & \\ & \end{array} &
\begin{array}{ll} 1 & 2 \\ 3 & \\ 4 & \\ 5 & \end{array} \\[2ex]
\begin{array}{ll} 1 & \\ & \\ & \\ & \end{array} &
\begin{array}{ll} 1 & \\ 2 & \\ & \\ & \end{array} &
\begin{array}{ll} 1 & 3 \\ 2 & \\ & \\ & \end{array} &
\begin{array}{ll} 1 & 3 \\ 2 & \\ 4 & \\ & \end{array} &
\begin{array}{ll} 1 & 3 \\ 2 & \\ 4 & \\ 5 & \end{array}
\end{array}$$

(2) Remember, from the definition of the dual Young diagram, that $\hat{\lambda}_1$ is the length of the first column of λ.

Hence

$$\pi \longrightarrow (P(\pi), Q(\pi)) = \left(\begin{pmatrix} 1 & 2 \\ 3 & \\ 4 & \\ 5 & \end{pmatrix}, \begin{pmatrix} 1 & 3 \\ 2 & \\ 4 & \\ 5 & \end{pmatrix} \right)$$

and $L_5(\pi) = 2 = \#$columns of P or Q.

The Robinson-Schensted-Knuth correspondence has the following properties
– $\pi \mapsto (P, Q)$, then $\pi^{-1} \mapsto (Q, P)$
– length (longest increasing subsequence of π) = # (columns in P)
– length (longest decreasing subsequence of π) = # (rows in P)
– $\pi^2 = I$, then $\pi \mapsto (P, P)$
– $\pi^2 = I$, with k fixed points, then P has exactly k columns of odd length.

1.2. Plancherel measure, integrals over $U(n)$ and Toeplitz matrices

A next set of ideas is due to Vershik & Kerov [40], Diaconis & Shashahani [13], Biane [9], Rains [33, 34], Baik & Rains [7]. For a nice state-of-the-art account, see Aldous & Diaconis [5]. The uniform probability measure (see (1.1.1) for f^λ) (1.0.1) on S_N induces on Young diagrams, via the RSK bijection (1.1.6), the "Plancherel" probability measure

$$P_N(\lambda) = \frac{(f^\lambda)^2}{N!}, \quad |\lambda| = N, \tag{1.2.1}$$

and so from (1.2.1) and the RSK bijection, one deduces a number of formulae below; notice, equality (ii) follows from (1.1.1) and (iii) follows from (i) and (1.1.5):

$$\begin{aligned}
P_N(L(\pi_N) \le n) &= \frac{1}{N!} \# \left\{ \begin{array}{c} (P,Q),\ \text{standard Young tableaux each of} \\ \text{arbitrary shape } \lambda, \text{ with } |\lambda| = N,\ \lambda_1 \le n \end{array} \right\} \\
&\overset{\text{(i)}}{=} \frac{1}{N!} \sum_{\substack{|\lambda|=N \\ \lambda_1 \le n}} (f^\lambda)^2 = \frac{1}{N!} \sum_{\substack{|\lambda|=N \\ \hat\lambda_1 \le n}} (f^\lambda)^2, \quad \text{by duality of Young diagrams} \\
&\overset{\text{(ii)}}{=} N! \sum_{m=1}^{n} \frac{1}{m!} \sum_{\substack{h \in \mathbb{Z}^m, h_i \ge 1 \\ \sum h_j = N + \frac{m(m-1)}{2}}} \prod_{1 \le i < j \le m} (h_i - h_j)^2 \prod_{1}^{m} \frac{1}{h_j!^2} \\
&\overset{\text{(iii)}}{=} \frac{1}{N!} \sum_{\substack{|\lambda|=|\mu|=N \\ \hat\lambda_1, \hat\mu_1 \le n}} f^\lambda f^\mu \langle s_\lambda, s_\mu \rangle \\
&= \frac{1}{N!} \langle (x_1 + \cdots + x_n)^N, (x_1 + \cdots + x_n)^N \rangle \\
&\overset{\text{(iv)}}{=} \frac{1}{N!} \int_{U(n)} |\operatorname{tr} M|^{2N} dM
\end{aligned} \tag{1.2.2}$$

$$= \frac{1}{N!}\begin{pmatrix} 2N \\ N \end{pmatrix}^{-1} \int_{U(n)} (\operatorname{tr}(M+\bar{M}))^{2N} dM,$$

using in the last equality the fact that $\int_{U(n)}((\operatorname{tr} M)^j(\overline{\operatorname{tr} M})^j)dM = 0$ for $i \neq j$. In 1990, Gessel [17] considered the generating function below and showed that it equals a Toeplitz determinant (determinant of a matrix, whose (i,j)th entry depends on $i-j$ only).

$$\begin{aligned}
\tau_n(x) &:= \sum_{N=0}^{\infty} \frac{x^N}{N!} P(L(\pi_N) \leq n) \\
&\overset{\text{(i)}}{=} \int_{U(n)} e^{\sqrt{x}\ tr(M+\bar{M})} dM \\
&\overset{\text{(ii)}}{=} \frac{1}{n!}\int_{(S^1)^n} \Delta_n(z)\Delta_n(\bar{z}) \prod_{k=1}^{n} \left(e^{\sqrt{x}(z_k+\bar{z}_k)} \frac{dz_k}{2\pi i z_k} \right) \\
&\overset{\text{(iii)}}{=} \det\left(\int_0^{2\pi} e^{2\sqrt{x}\cos\theta} e^{i(k-\ell)\theta} d\theta \right)_{1\leq k,\ell\leq n} \quad \text{(Toeplitz determinant)} \qquad (1.2.3) \\
&\overset{\text{(iv)}}{=} \prod_{k=0}^{n-1} h_k(x)^{-1}, \quad \text{where } h_k(x) := \frac{\tau_{k+1}}{\tau_k} = \int_{S^1} |p_k(z)|^2 e^{\sqrt{x}(z+\bar{z})} dz, \\
&\overset{\text{(v)}}{=} e^x \prod_{k=n}^{\infty} h_k(x)^{-1},
\end{aligned}$$

where $p_k(z)$ are monic orthogonal polynomials on the circle S^1 for the weight $e^{\sqrt{x}(z+\bar{z})}dz$:

$$\int_{S^1} p_n(z)\overline{p_m(z)} e^{\sqrt{x}(z+\bar{z})} \frac{dz}{2\pi i z} = h_n(x)\delta_{mn}. \qquad (1.2.4)$$

Equality (1.2.3)(i) follows from (1.2.2)(iv); moreover (1.2.3)(ii) uses Haar measure on $U(n)$. Identity (1.2.3) (iii) follows from the fact that the product of the two Vandermonde appearing in the integral (ii) can be expressed as sum of determinants:

$$\begin{aligned}
\Delta_n(u)\Delta_n(v) &= \sum_{\sigma\in S_n} \det\left(u_{\sigma(k)}^{\ell-1} v_{\sigma(k)}^{k-1} \right)_{1\leq \ell,k\leq n} \\
&= \det\left(p_{\ell-1}^{(1)}(u_k) \right)_{1\leq \ell,k\leq n} \det\left(p_{\ell-1}^{(2)}(v_k) \right)_{1\leq \ell,k\leq n}.
\end{aligned}$$

The fact above that $\Delta_n(u)\Delta_n(v)$ can also be expressed as a product of two determinants, involving monic polynomials, implies (iv). Finally, by Szegö's strong limit theorem for Toeplitz determinants, we have $\lim_{n\to\infty} \tau_n = e^x$, thus leading to (v).

1.3. Virasoro constraints, integrable systems and Painlevé V equation

THEOREM 1.1. — *For every $\ell \geq 0$, the Gessel generating function*

$$\sum_{N=0}^{\infty} \frac{x^N}{N!} P(L(\pi_N) \leq \ell) = \int_{U(\ell)} e^{\sqrt{x}\,\mathrm{tr}(M+\bar{M})} dM \tag{1.3.1}$$

$$= \exp \int_0^x \log\left(\frac{x}{u}\right) g_\ell(u) du,$$

where g_ℓ is the unique solution to the initial value problem:

$$\begin{cases} g'' - \dfrac{g'^2}{2}\left(\dfrac{1}{g-1} + \dfrac{1}{g}\right) + \dfrac{g'}{u} + \dfrac{2}{u} g(g-1) - \dfrac{\ell^2}{2u^2}\dfrac{g-1}{g} = 0 \\ \\ g_\ell(u) = 1 - \dfrac{u^\ell}{(\ell!)^2} + O(u^{\ell+1}), \ \ \textit{near } u = 0. \quad \textbf{(Painlevé V)}. \end{cases} \tag{1.3.2}$$

Theorem 1.1 is due to Hisakado [19], Tracy-Widom [36], by methods of functional analysis and Adler-van Moerbeke [3], by integrable methods. A similar statement holds for the set S_{2n}^0 of fixed-point free involutions π^0(i.e., $(\pi^0)^2 = I$ and $\pi^0(k) \neq k$ for $1 \leq k \leq 2n$). Put the uniform distribution on S_{2n}^0 :

$$P(\pi_{2n}^0) = \frac{1}{(2n-1)!!} = \frac{2^n n!}{(2n)!}. \tag{1.3.3}$$

Then, in the following statement, due to Adler-van Moerbeke [3] and Baik-Rains [7], $O(n)_\pm$ refers to orthogonal matrices, with determinant $= \pm 1$.

THEOREM 1.2. — *The generating function*

$$\begin{aligned} 2\sum_{k=0}^{\infty} \frac{(x^2/2)^k}{k!} P(L(\pi_{2k}^0) \leq \ell + 1) &= E_{O(\ell+1)_-} e^{x\ trM} + E_{O(\ell+1)_+} e^{x\ trM} \\ &= \exp\left(\int_0^x \frac{f_\ell^-(u)}{u} du\right) + \exp\left(\int_0^x \frac{f_\ell^+(u)}{u} du\right), \end{aligned} \tag{1.3.4}$$

where $f = f_\ell^\pm$ is the unique solution to the initial value problem:

$$\begin{cases} f''' + \dfrac{1}{u} f'' + \dfrac{6}{u} f'^2 - \dfrac{4}{u^2} f f' - \dfrac{16u^2 + \ell^2}{u^2} f' + \dfrac{16}{u} f + \dfrac{2(\ell^2 - 1)}{u} = 0 \\ \\ f_\ell^\pm(u) = u^2 \pm \dfrac{u^{\ell+1}}{\ell!} + O(u^{\ell+2}), \ \ \textit{near } u = 0. \quad \textbf{(Painlevé V)} \end{cases} \tag{1.3.5}$$

Proof of Theorem 1.1. — A brief outline will be given here, because of its interesting connection with an integrable system, called the Toeplitz lattice (see [3]). Inserting times t_i, with $i = ..., -2, -1, 0, 1, 2, ...,$ and $t_0 = 0$, in the integrals (1.2.3)(ii) over $(S^1)^n$, one obtains, setting $t := (..., t_{-1}, t_0, t_1, ...)$:

$$I_n(t) = \int_{(S^1)^n} |\Delta_n(z)|^2 \prod_{k=1}^n e^{\sum_{i\neq 0} t_i z_k^i} \frac{dz_k}{2\pi i z_k}, \quad n \geq 0. \tag{1.3.6}$$

Virasoro constraints [3]. — The I_n's satisfy a set of three linear partial differential equations, forming an sl(2,$\mathbb{Z}$)-algebra (Virasoro constraints), explicitly given by the three equations:

$$\begin{aligned} &\left(\sum_{i\neq\mp1,0}(i\pm1)t_{i\pm1}\frac{\partial}{\partial t_i}+n\left(t_{\pm1}+\frac{\partial}{\partial t_{\mp1}}\right)\right)I_n(t)=0\\ \text{and}\quad &\left(\sum_{i\neq0}it_i\frac{\partial}{\partial t_i}\right)I_n(t)=0. \end{aligned}\tag{1.3.7}$$

Toeplitz lattice. — The t-dependent semi-infinite moment matrix,

$$m_n(t):=\left(\langle z^k,z^\ell\rangle_t\right)_{0\le k,\ell\le n-1}=\left(\oint_{S^1}\frac{\rho(z)dz}{2\pi iz}z^{k-\ell}e^{\sum_{i\neq0}(t_iz^i)}\right)_{0\le k,\ell\le n-1}\tag{1.3.8}$$

is a Toeplitz matrix and satisfies the simple differential equations:

$$\frac{\partial\mu_{ij}}{\partial t_k}=\mu_{i+k,j},\qquad\frac{\partial\mu_{ij}}{\partial t_{-k}}=\mu_{i,j+k},\quad\text{for}\ \ k\ge1.\tag{1.3.9}$$

In analogy with (1.2.3)(iii), define $\tau_n(t)$, instead of $\tau_n(x)$:

$$\tau_n(t):=\det m_n(t).$$

The Borel factorization $m_\infty(t)=S_1^{-1}(t)S_2(t)$ with a lower-triangular matrix S_1 (with 1's along the diagonal) and an upper-triangular matrix S_2, enables one to define[(3)] $L_1:=S_1\Lambda S_1^{-1}$ and $L_2:=S_2\Lambda^\top S_2^{-1}$, which combined with the Toeplitz nature of m_∞ has a peculiar "rank 2"-structure, where x_i and y_i are certain integrals over $U(n)$ and where $h_i=\tau_{i+1}/\tau_i$, $h_i/h_{i-1}=1-x_iy_i$, $h:=\operatorname{diag}(h_0,h_1,...)$ and $x_0=y_0=1$:

$$L_1=h\begin{pmatrix}-x_1y_0&1-x_1y_1&0&0&\\-x_2y_0&-x_2y_1&1-x_2y_2&0&\\-x_3y_0&-x_3y_1&-x_3y_2&1-x_3y_3&\\-x_4y_0&-x_4y_1&-x_4y_2&-x_4y_3&\\&&&&\ddots\end{pmatrix}h^{-1}$$

and

$$L_2=\begin{pmatrix}-x_0y_1&-x_0y_2&-x_0y_3&-x_0y_4&\\1-x_1y_1&-x_1y_2&-x_1y_3&-x_1y_4&\\0&1-x_2y_2&-x_2y_3&-x_2y_4&\\0&0&1-x_3y_3&-x_3y_4&\\&&&&\ddots\end{pmatrix}.\tag{1.3.10}$$

(3) where Λ is the shift matrix $(\Lambda v)_n=v_{n+1}$, $n\ge0$, i.e., Λ is the semi-infinite matrix with all 0 entries, except for 1's just above the diagonal.

The **2d-Toda Lattice** hierarchy[4]

$$\frac{\partial L_i}{\partial t_n} = [(L_1^n)_+ \ , L_i] \quad \text{and} \quad \frac{\partial L_i}{\partial t_{-n}} = [(L_2^n)_- \ , L_i] \ , \quad i = 1, 2, n \geq 1, \tag{1.3.11}$$

maintains the "rank 2" nature of the semi-infinite matrices L_1 and L_2. The first equation in this hierarchy is equivalent to the **discrete sinh-Gordon equation**

$$\frac{\partial^2 q_n}{\partial t_1 \partial t_{-1}} = -e^{q_n - q_{n-1}} + e^{q_{n+1} - q_n}, \quad \text{where} \ \ q_n = \log \frac{\tau_{n+1}}{\tau_n}.$$

The equations (1.3.11) induce on the x_i and y_i the **Toeplitz lattice** equations; see [3].

Using the three equations (1.3.7) and $\partial/\partial t_1$ of the first equation of (1.3.7), and setting all $t_i = 0$, enable one to extract various t-partials, like $\partial^2 \log \tau(t)/\partial t_{-1} \partial t_1$ and $\partial^2 \log \tau(t)/\partial t_{-2} \partial t_1$, in terms of pure partials in $\partial^k/\partial x^k$, where $x = t_1 t_{-1}$. Substituting these expressions in the integrable Toeplitz lattice equations leads to Painlevé V (1.3.2). The integrable system associated with the integral (1.3.4) in Theorem 1.2 (with t_i's inserted) is the standard Toda lattice, instead of the Toeplitz lattice; the integral satisfies the Virasoro constraints associated with the Toda lattice; see [3].

1.4. Random permutations π_N for large N

Around 1960 and based on Monte-Carlo methods, Ulam [38] conjectured that

$$\lim_{n \to \infty} \frac{E(L_n)}{\sqrt{n}} = c \ \ \text{exists.}$$

An argument of Erdös & Szekeres [14], dating back from 1935 showed that $E(L_n) \geq \frac{1}{2}\sqrt{n-1}$, and thus $c \geq 1/2$. In '72, Hammersley [18] showed rigorously that the limit exists. Logan and Shepp [23] showed the limit $c \geq 2$, and finally Vershik and Kerov [40] that $c = 2$. The next major contribution was due to Johansson [22] and Baik-Deift-Johansson [6]:

Theorem 1.3 ("Law of large numbers" and "Central limit theorem")

One has:

(1.4.1)

$$\lim_{n \to \infty} \frac{L_n}{2\sqrt{n}} = 1, \quad \textit{and} \quad \lim_{n \to \infty} P\left(\frac{L_n - 2\sqrt{n}}{n^{1/6}} \leq x\right) = \exp\left(-\int_x^\infty (u-x) g^2(u) du\right),$$

where $g(x)$ is a solution of (a version of) Painlevé II,

$$\begin{cases} g'' = xg + 2g^3 \quad (\textbf{Painlevé II}) \\ g(x) \cong A(x) := \int_{-\infty}^{\infty} e^{ixy - y^3/3} dy \cong \frac{e^{-\frac{2}{3}x^{\frac{3}{2}}}}{2\sqrt{\pi} x^{1/4}} \ \textit{for} \ x \nearrow \infty. \end{cases} \tag{1.4.2}$$

[4] $(\ )_+$ denotes the upper-triangular part of $(\ )$, including the diagonal, whereas $(\ )_-$ denotes the strictly lower-triangular part of $(\ )$.

The point $\gamma = \frac{2\sqrt{x}}{n} = 1$ corresponds to a "*phase transition*" for the partition function

$$e^{-x}\tau_n(x) = e^{-x}\int_{U(n)} e^{\sqrt{x}\ tr(M+\bar{M})}dM = \prod_{k=n}^{\infty} h_k(x)^{-1} = e^{-x}\sum_{n=0}^{\infty}\frac{x^N}{N!}P(L(\pi_N)\leq n)$$

On the one hand, Johansson shows, for all $\varepsilon > 0$, that there exist C and $\delta > 0$ such that
(1.4.3)

$$e^{-x}\tau_n(x) \leq Ce^{-\delta x},\ \text{for}\ \frac{2\sqrt{x}}{n} > 1+\varepsilon\ \ \text{and}\ \ e^{-x}\tau_n(x) \geq 1-\frac{C}{n},\ \text{for}\ \frac{2\sqrt{x}}{n} < 1-\varepsilon.$$

On the other hand, setting $P_{n,N} := P(L(\pi_N) \leq n)$, Johansson's de-Poissonization Lemma goes as follows: given any $\alpha > 3$, there are constants $C = C(\alpha)$ and $N_0 = N_0(\alpha)$ such that the following holds for $N \geq N_0$ and all $0 \leq n \leq N$, with $m(\alpha) = (\alpha^2 - 2\alpha - 1)/2$:

$$(1.4.4)\quad e^{-x}\sum_{n=0}^{\infty}\frac{x^N}{N!}P_{n,N}\big|_{x=N+\alpha\sqrt{N\log N}} - \frac{C}{N^{m(\alpha)}} \leq P_{n,N}$$
$$\leq e^{-x}\sum_{n=0}^{\infty}\frac{x^N}{N!}P_{n,N}\big|_{x=N-\alpha\sqrt{N\log N}} + \frac{C}{N^{m(\alpha)}}.$$

The two relations (1.4.3) and (1.4.4) combined lead to the law of large numbers in (1.4.1). To prove the "central limit theorem" (1.4.1), one uses the Riemann-Hilbert approach to obtain asymptotics. Indeed, as a first observation, setting $\rho_x(z) := e^{\sqrt{x}(z+z^{-1})}$, the only solution Y_{n+1} (2×2 matrix) to the following Riemann-Hilbert problem *à la* Fokas, Its and Kitaev [15], but adapted to S^1:

(1) $Y(z)$ holomorphic in $\mathbb{C}\backslash S^1$

(2) $Y_-(z) = Y_+(z)\begin{pmatrix} 1 & \frac{\rho_x(z)}{z^{n+1}} \\ 0 & 1 \end{pmatrix}$

(3) $Y(z)\begin{pmatrix} z^{-(n+1)} & 0 \\ 0 & z^{n+1} \end{pmatrix} = \big(I + O(z^{-1})$, when $z \to \infty$ is given by the matrix

$$(1.4.5)\qquad Y_{n+1}(z) = \begin{pmatrix} p_{n+1}(z) & \int_{S^1} p_{n+1}(u)\frac{\rho_x(u)}{u-z}\frac{du}{2\pi i u^{n+1}} \\ \\ -h_n^{-1}z^n\overline{p_n(1/\bar{z})} & -h_n^{-1}\int_{\mathbb{R}}\overline{p_n(1/\bar{u})}\frac{\rho_x(u)}{u-z}\frac{du}{2\pi i u^{n+1}} \end{pmatrix},$$

where $p_k(z)$ are the monic orthogonal polynomials (1.2.4) on the circle S^1 for the weight $\rho_x(z)dz$.

The Riemann-Hilbert formulation is an efficient tool to find the asymptotics of $(Y_{k+1}(0))_{21} = -h_k^{-1}(x)$. Then, summing up, $\sum_n^{\infty}\log h_k^{-1}(x) = \log(e^{-x}\tau_n(x))$, Baik-Deift-Johansson [6] show the following result: Define u such that $\frac{2\sqrt{x}}{n+1} =$

$1 - \frac{u}{2^{1/3}(n+1)^{2/3}}$, with $-M < u < M$ for a given $M > 0$. Then the estimate below holds, where g is the solution to (1.4.2):

$$\left| \log(e^{-x}\tau_n(x)) - \int_u^\infty 2if(y)dy \right| = \frac{C(M)}{n^{1/3}} + Ce^{-1/4M^{3/2}}$$

with $2if(y) = -\int_y^\infty g^2(\alpha)d\alpha$. This estimate combined with Johansson's de-Poissonization Lemma leads to (1.4.1) and (1.4.2).

2. THE SPECTRUM OF RANDOM MATRICES

Random matrix theory deals with an ensemble of matrices M, having some symmetry condition to guarantee the reality of the spectrum, e.g., the Hermitian ensemble $\mathcal{H}_n$, the symmetric ensemble $\mathcal{S}_n$ or the symplectic ensemble $\mathcal{T}_{2n}$. Define a probability measure, based on a Haar measure, which in "polar coordinates" is expressed in terms of the Vandermonde with a power β:

$$c_n e^{-trV(M)}dM = c_n|\Delta_n(z)|^\beta \prod_1^n e^{-V(z_k)}dz_k dU,$$

with $V' = \frac{g}{f}$ rational, $dM =$ Haar measure. The three ensembles $\mathcal{H}_n,\ \mathcal{S}_n,\ \mathcal{T}_{2n}$ appear very naturally as the tangent spaces (at the identity) to the simplest symmetric spaces:

non-compact symmetric space G/K	$\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{p}$ with $\mathfrak{p}$=tangent space to G/K at I	space $\mathfrak{p} = \{M \in \mathfrak{g}, \text{with...}\}$	Haar measure on $\mathfrak{p}$
$SL(n,\mathbb{C})/SU(n)$	$sl(n,\mathbb{C}) = su(n) \oplus \mathcal{H}_n$	$M = \bar{M}^\top$	$\beta = 2$
$SL(n,\mathbb{R})/SO(n)$	$sl(n,\mathbb{R}) = so(n) \oplus \mathcal{S}_n$	$M = M^\top$	$\beta = 1$
$SU^*(2n)/USp(n)$	$su^*(2n) = (sp(n,\mathbb{C}) \cap u(2n)) \oplus \mathcal{T}_{2n}$	$M = \bar{M}^\top,\ M = J\bar{M}J^{-1}$	$\beta = 4$

Question. — What is the statistics of the spectrum of M ?

For $\mathcal{H}_n$ and $\mathcal{S}_n$, if the probability $P(M \in dM)$ satisfies the following two requirements: (i) invariance under conjugation by unitary transformations $M \mapsto UMU^{-1}$, (ii) the random variables $M_{ii},\ \Re M_{ij},\ \Im M_{ij},\ 1 \le i < j \le n$ are independent, then $V(z)$ is quadratic (Gaussian ensemble) ([28]).

2.1. Virasoro constraints, Toda and Pfaff lattices and KP equations

Consider weights of the form $\rho(z)dz := e^{-V(z)}dz$ on an interval $F = [A, B] \subseteq \mathbb{R}$, with rational logarithmic derivative and subjected to the following boundary conditions:

$$-\frac{\rho'}{\rho} = \frac{g}{f} = \frac{\sum_0^\infty b_i z^i}{\sum_0^\infty a_i z^i}, \quad \lim_{z \to A,B} f(z)\rho(z)z^k = 0 \text{ for all } k \ge 0, \tag{2.1.1}$$

and a disjoint union of intervals $E = \bigcup_1^{2r}\ [x_{2i-1}, x_{2i}] \subset F \subset \mathbb{R}$.

THEOREM 2.1 (Adler-van Moerbeke [1])

The vector of integrals $I(t,x;\beta) = (I_0 = 1, I_1(t,x;\beta), \ldots)$, *with* $t := (t_1, t_2, \ldots)$ *and* $x := (x_1, \ldots, x_{2r})$ *and*

$$I_n(t,x;\beta) := \int_{E^n} |\Delta_n(z)|^\beta \prod_{k=1}^n \left(e^{\sum_1^\infty t_i z_k^i} \rho(z_k) dz_k \right) \quad \textit{for} \quad n > 0 \tag{2.1.2}$$

satisfies the following Virasoro constraints(5) *for all* $k \geq -1$*:*

$$\left(-\sum_1^{2r} x_i^{k+1} f(x_i) \frac{\partial}{\partial x_i} + \sum_{i \geq 0} \left(a_i \ {}^\beta\mathbb{J}_{k+i}^{(2)}(t,n) - b_i \ {}^\beta\mathbb{J}_{k+i+1}^{(1)}(t,n) \right) \right) I(t,x;\beta) = 0,$$

in terms of the coefficients a_i, b_i *of the rational function* $(-\log\rho)'$ *and the end points* x_i *of the subset* E, *as in (2.1.1). The* ${}^\beta\mathbb{J}_k^{(2)}$ *and* ${}^\beta\mathbb{J}_k^{(1)}$ *form a Virasoro and a Heisenberg algebra respectively, interacting as follows*

$$\begin{aligned} \left[{}^\beta\mathbb{J}_k^{(2)}, {}^\beta\mathbb{J}_\ell^{(2)} \right] &= (k-\ell) \ {}^\beta\mathbb{J}_{k+\ell}^{(2)} + c \left(\frac{k^3-k}{12} \right) \delta_{k,-\ell} \\ \left[{}^\beta\mathbb{J}_k^{(2)}, {}^\beta\mathbb{J}_\ell^{(1)} \right] &= -\ell \ {}^\beta\mathbb{J}_{k+\ell}^{(1)} + c' k(k+1) \delta_{k,-\ell}. \\ \left[{}^\beta\mathbb{J}_k^{(1)}, {}^\beta\mathbb{J}_\ell^{(1)} \right] &= \frac{k}{\beta} \delta_{k,-\ell}, \end{aligned} \tag{2.1.3}$$

with central charge $c = 1 - 6 \left(\left(\frac{\beta}{2} \right)^{1/2} - \left(\frac{\beta}{2} \right)^{-1/2} \right)^2$ *and* $c' = \left(\frac{1}{\beta} - \frac{1}{2} \right)$.

Moreover:

(i) $\tau_n(t) := \frac{1}{n!} I_n(t,x,\beta)|_{\beta=2}$ *form the* τ*-functions of the standard Toda lattice; in particular, each* τ_n *satisfies the KP equation:*

$$\left(\left(\frac{\partial}{\partial t_1} \right)^4 + 3 \left(\frac{\partial}{\partial t_2} \right)^2 - 4 \frac{\partial^2}{\partial t_1 \partial t_3} \right) \log \tau_n + 6 \left(\frac{\partial^2}{\partial t_1^2} \log \tau_n \right)^2 = 0.$$

(ii) $\tau_{2n}(t) := \begin{cases} \frac{1}{(2n)!} I_{2n}(t,x;\beta)|_{\beta=1} \\ \frac{1}{n!} I_n(2t,x;\beta)|_{\beta=4} \end{cases}$ *form the* τ*-functions of the Pfaff lattice (i.e., they are Pfaffians, rather than determinants); in particular, they satisfy the Pfaff-KP equation:*

$$\left(\left(\frac{\partial}{\partial t_1} \right)^4 + 3 \left(\frac{\partial}{\partial t_2} \right)^2 - 4 \frac{\partial^2}{\partial t_1 \partial t_3} \right) \log \tau_{2n} + 6 \left(\frac{\partial^2}{\partial t_1^2} \log \tau_{2n} \right)^2 = 12 \frac{\tau_{2n-2}\tau_{2n+2}}{\tau_{2n}^2}.$$

(5) When E equals the whole range F, then the $\partial/\partial x_i$'s are absent in the formulae (2.1.7).

2.2. The Gaussian ensemble: PDE's for the statistics of the spectrum

THEOREM 2.2. — *For the Gaussian ensemble, the probabilities*[(6)]*: ($\beta = 2, 1, 4$)*

$$
\begin{aligned}
P_n(E) &:= P_n(\text{ all spectral points of } M \in E) \\
&= \frac{\int_{\mathcal{H}_n(E),\ \mathcal{S}_n(E)\ or\ \mathcal{T}_n(E)} e^{-tr\ M^2} dM}{\int_{\mathcal{H}_n(\mathbb{R}),\ \mathcal{S}_n(\mathbb{R})\ or\ \mathcal{T}_n(\mathbb{R})} e^{-tr\ M^2} dM} \qquad (2.2.1)\\
&= \frac{\int_{E^n} |\Delta_n(z)|^\beta \prod_{k=1}^n e^{-z_k^2} dz_k}{\int_{\mathbb{R}^n} |\Delta_n(z)|^\beta \prod_{k=1}^n e^{-z_k^2} dz_k}
\end{aligned}
$$

satisfy the following PDE's in the x_i's ($\mathcal{B}_k = \sum_1^{2r} x_i^{k+1} \frac{\partial}{\partial x_i}$, $\delta_{1,4}^\beta = 1$ for $\beta = 1, 4$ and 0 otherwise, $f_n(x) = \frac{d}{dx} \log P_n$):[(7)]

$$
\delta_{1,4}^\beta Q \left(\frac{P_{n-{}^2_1} P_{n+{}^2_1}}{P_n^2} - 1 \right) \qquad \begin{cases} 2 & \textit{and } n \textit{ even, for } \beta = 1 \\ 1 & \textit{and } n \textit{ arbitrary, for } \beta = 4 \end{cases}
$$

$$
= \begin{cases} \left(\mathcal{B}_{-1}^4 + (Q_2 + 6\mathcal{B}_{-1}^2 F)\mathcal{B}_{-1}^2 + (2 - \delta_{1,4}^\beta) \dfrac{4}{\beta} (3\mathcal{B}_0^2 - 4\mathcal{B}_{-1}\mathcal{B}_1 + 6\mathcal{B}_0) \right) \log P_n \\ \hfill \textit{for } E = \bigcup_1^{2r} [x_{2i-1}, x_{2i}] \\ f_n''' + 6f_n'^2 + \left(\dfrac{4x^2}{\beta} (\delta_{1,4}^\beta - 2) + Q_2 \right) f_n' - \dfrac{4x}{\beta} (\delta_{1,4}^\beta - 2) f_n, \quad \textit{for } E = [-\infty, x]. \end{cases}
$$

Note that for $\beta = 2$ and for $E = [-\infty, x]$, the equation above takes on the simple form:

$$
f_n''' + 6f_n'^2 + 8nf_n' + 4xf_n = 0. \qquad \textbf{(Painlevé IV)} \qquad (2.2.2)
$$

Equation (2.2.2) was obtained by Tracy-Widom [35] and the rest of Theorem 2.2 is due to M. Adler, T. Shiota and P. van Moerbeke ([4] and [1]).

2.3. Infinite Hermitian matrix ensembles

Consider probability (2.2.1) for $\beta = 2$ and let the size n of the matrices go to ∞. To perform this limit, one uses a different representation; namely,

$$
P_n(E) := \frac{\int_{\mathcal{H}_n(E)} e^{-tr\ M^2} dM}{\int_{\mathcal{H}_n} e^{-tr\ M^2} dM} = \frac{1}{n!} \int_{E^n} \det(K_n(z_k, z_\ell))_{1 \le k, \ell \le n} \prod_1^n \rho(dz_i), \qquad (2.3.1)
$$

(6) $\mathcal{H}_n(E)$, $\mathcal{S}_n(E)$ or $\mathcal{T}_n(E)$ denotes the subset of matrices with spectrum in $E \subseteq \mathbb{R}$.

(7) Also, define the *invariant* polynomials

$$
Q = 12n \left(n + 1 - \frac{2}{\beta} \right) \quad \text{and} \quad Q_2 = 4(1 + \delta_{1,4}^\beta) \left(2n + \delta_{1,4}^\beta (1 - \frac{2}{\beta}) \right).
$$

can be represented in terms of a reproducing kernel

(2.3.2)
$$K_n(y,z) := \sum_{j=1}^{n} \frac{p_{j-1}(y)}{\sqrt{h_{j-1}}}\frac{p_{j-1}(z)}{\sqrt{h_{j-1}}}, \text{ where } \int_{\mathbb{R}} K_n(y,z)K_n(z,u)\rho(dz) = K_n(y,u).$$

The reproducing property follows from the orthogonality of the *monic orthogonal polynomials* $p_k = p_k(z)$ for the Gaussian weight $e^{-z^2}dz$ on $\mathbb{R}$, and the L^2-norms $h_k = \int_{\mathbb{R}} p_k^2(z)e^{-z^2}dz$ of the p_k's. We also have the following Fredholm determinant formula

$$P(\text{exactly } k \text{ eigenvalues } \in [x_1,x_2]) \\ = \frac{(-1)^k}{k!}\left(\frac{\partial}{\partial\lambda}\right)^k \det(I - \lambda K_n(y,z)I_{[x_1,x_2]}(z))\Big|_{\lambda=1}.$$

When the size n of the Hermitian matrices tends to ∞, the following holds:

– *Wigner's semi-circle law*: For this ensemble and for very large n, the density of eigenvalues tends to the semi-circle distribution on the interval $[-\sqrt{2n}, \sqrt{2n}]$.

– *Bulk scaling limit*: From the formula above, it follows that the average number of eigenvalues per unit length near $z = 0$ ("the bulk") is given by $\sqrt{2n}/\pi$ and thus the average distance between two consecutive eigenvalues is given by $\pi/\sqrt{2n}$. Upon using this rescaling, one shows ([24, 27, 29, 32, 21])

$$\lim_{n\nearrow\infty} \frac{\pi}{\sqrt{2n}}K_n\left(\frac{\pi y}{\sqrt{2n}}, \frac{\pi z}{\sqrt{2n}}\right) = \frac{\sin\pi(y-z)}{\pi(y-z)} =: K(y,z) \qquad \text{(Sine kernel)}$$

with

$$P(\text{no eigenvalues } \in [0,x]) = \det(I - K(y,z)I_{[0,x](z)}) = \exp\int_0^{\pi x}\frac{f(x)}{x}dx, \tag{2.3.3}$$

where $f(x,\lambda)$ is a solution to the **Painlevé V** differential equation, (see Jimbo, Miwa, Mori, Sato [21]), $('= \partial/\partial x)$

$$(xf'')^2 - 4(xf'-f)(-f'^2 - xf' + f) = 0 \text{ with } f(x;\lambda) \cong -\frac{x}{\pi}, \text{ for } x \simeq 0.$$

– *Edge scaling limit*: I will particularly concentrate on the "edge" $\sqrt{2n}$ of the Wigner semi-circle. There the scaling is $\sqrt{2}n^{1/6}$, which makes the problems considerably subtler; this will be used in (2.3.7) (see [10, 16, 28, 35]). The main result can be stated as follows:

THEOREM 2.3. — *Given the spectrum $z_1 \geq z_2 \geq \ldots$ of the large random Hermitian matrix M, define the "eigenvalues" in a new scale:*

$$u_i = 2n^{\frac{2}{3}}\left(\frac{z_i}{\sqrt{2n}} - 1\right) \quad \textit{for } n \nearrow \infty \tag{2.3.4}$$

then the statistics of the largest "eigenvalue" u_1 *(in the new scale) is given by the same probability distribution as the length of the longest increasing sequence:*

$$P(u_1 \le x) = \exp\left(-\int_x^\infty (\alpha - x) g^2(\alpha) d\alpha\right), \tag{2.3.5}$$
$$\textit{with} \begin{cases} g'' = xg + 2g^3 \quad (\textbf{Painlevé II}) \\ g(x) \cong A(x) \cong -\frac{e^{-\frac{2}{3}x^{\frac{3}{2}}}}{2\sqrt{\pi}x^{1/4}} \textit{ for } x \nearrow \infty. \end{cases}$$

More generally, $P(\textit{no points } u_i \in [x_1, x_2])$ *satisfies the partial differential equation*

$$\left(\mathcal{B}_{-1}^3 - 4(\mathcal{B}_0 - \frac{1}{2})\right) f + 6(\mathcal{B}_{-1} f)^2 = 0, \ \textit{with } f := \sum_1^2 \frac{\partial \log P}{\partial x_i}, \quad \mathcal{B}_k = \sum_1^2 x^{k+1} \frac{\partial}{\partial x_i}. \tag{2.3.6}$$

$\square$

Tracy-Widom [35] have first obtained result (2.3.5) by methods of functional analysis. In [4], equations (2.3.5) and (2.3.6) were derived using integrable systems and Virasoro constraints.

Proof. — Setting $z = \sqrt{2n} + \frac{u}{\sqrt{2}n^{1/6}}$, in the kernel $K_n(x, y)$ of (2.3.2), one proves

$$\lim_{n \nearrow \infty} \frac{1}{\sqrt{2}n^{1/6}} K_n \left(\sqrt{2n} + \frac{u}{\sqrt{2}n^{1/6}}, \sqrt{2n} + \frac{v}{\sqrt{2}n^{1/6}}\right) = K(u, v), \tag{2.3.7}$$

where the Airy kernel K is defined in terms of the Airy function:

$$K(u, v) = \int_0^\infty A(x + u) A(x + v) dx, \tag{2.3.8}$$
$$A(u) = \int_{-\infty}^\infty e^{iux - x^3/3} dx \cong \frac{e^{-(2/3)u^{3/2}}}{2\sqrt{\pi}u^{1/4}}, \quad \text{for } u \nearrow \infty.$$

To find the differential equations (2.3.6), one proceeds (sketchily) as follows: as a first step, notice that, setting $L^2 := (\frac{\partial}{\partial t_1})^2 + q(t)$, $t := (t_1, t_2, ...)$, the solution $q(t)$ and $\Psi(t; z)$ (with asymptotics $\Psi(t; z) = e^{\sum_1^\infty t_i z^i}(1 + O(z^{-1}))$, $z \in \mathbb{C}$, $z \to \infty$) to the initial value problem:

$$\begin{cases} (L^2 - z^2)\Psi(t; z) = 0, \quad \frac{\partial L^2}{\partial t_n} = [(L^n)_+, L^2], \quad \frac{\partial \Psi}{\partial t_n} = (L^n)_+ \Psi, \\ q(t^{(0)}) = -x, \quad \text{at } t = t^{(0)} := (x, 0, \frac{2}{3}, 0, 0, ...), \end{cases} \tag{2.3.9}$$

is given by Kontsevich's integral (Z diagonal and N arbitrary), which itself is intimately related to the Airy function $A(x)$,

$$\begin{cases} q(t) &= 2\dfrac{\partial^2}{\partial t_1^2} \log \tau(t), \\ \Psi(t; z) &= e^{\sum_1^\infty t_i z^i} \tau(t - [z^{-1}])/\tau(t), \\ \Psi(t^{(0)}; z) &= z^{1/2} A(x + z^2) = e^{xz + \frac{2}{3}z^3}(1 + O(z^{-1})), \quad z \to \infty. \end{cases} \tag{2.3.10}$$

with

$$\tau(t) = \frac{\int_{\mathcal{H}_N} dX e^{-tr(X^3/3+X^2Z)}}{\int_{\mathcal{H}_N} dX e^{-tr(X^2Z)}}, \quad t_n := -\frac{1}{n} trZ^{-n} + \frac{2}{3}\delta_{n3}$$

and

$$[z^{-1}] = (z^{-1}, z^{-2}/2, z^{-3}/3, ...) \in \mathbb{C}^\infty$$

These data define a kernel

$$K_t(z^2, z'^2) := \frac{1}{z^{\frac{1}{2}} z'^{\frac{1}{2}}} \int_0^\infty \Psi(t; z)\Psi(t; z')dt_1, \text{ where } t = (t_1, t_2, ...) \tag{2.3.11}$$

which flows off the Airy kernel $K(z^2, z'^2)$, defined in (2.3.8), upon using (2.3.10).

Then, one shows that both, Kontsevich's integral $\tau(t)$ and the product $\tau(t, E) := \tau(t)$.
$\det(I - K_t(\lambda, \lambda')I_E(\lambda'))$, with $E = \bigcup_1^{2r} [x_{2i-1}, x_{2i}] \subset \mathbb{R}$, satisfy Virasoro constraints, of which the two first ones read (after the time shift $t_3 \mapsto t_3 + 2/3$, in view of (2.3.9)):

$$\begin{aligned} \sum_{i=1}^{2r} \frac{\partial}{\partial x_i} \log \tau(t, E) &= \left(\frac{\partial}{\partial t_1} + \frac{1}{2}\sum_{i\geq 3} it_i \frac{\partial}{\partial t_{i-2}}\right) \log \tau(t, E) + \frac{t_1^2}{4} \\ \sum_{i=1}^{2r} x_i \frac{\partial}{\partial x_i} \log \tau(t, E) &= \left(\frac{\partial}{\partial t_3} + \frac{1}{2}\sum_{i\geq 1} it_i \frac{\partial}{\partial t_i}\right) \log \tau(t, E) + \frac{1}{16}, \end{aligned} \tag{2.3.12}$$

Both, $\tau(t)$ and $\tau(t, E)$, also satisfy the Korteweg-de Vries equation (KP equation, depending on odd t_i's only)

$$\left(\left(\frac{\partial}{\partial t_1}\right)^4 - 4\frac{\partial^2}{\partial t_1 \partial t_3}\right) \log \tau + 6\left(\frac{\partial^2}{\partial t_1^2} \log \tau\right)^2 = 0. \quad \textbf{(KdV)} \tag{2.3.13}$$

Differentiating (2.3.12) in t_1 and t_3, and setting all $t_i = 0$ enable one to express the t-partials $\frac{\partial^2}{\partial t_1^2} \log \tau\Big|_{t=0}$, $\frac{\partial^4}{\partial t_1^4} \log \tau\Big|_{t=0}$, $\frac{\partial^2}{\partial t_1 \partial t_3} \log \tau\Big|_{t=0}$ appearing in the KdV equation in terms of partials $\partial/\partial x_i$.

Setting these expressions in the the KdV-equation at $t = 0$, implies that the statistics of the scaled eigenvalues u_i (for the Airy kernel (2.3.8))

$$P(E^c) := P(\text{all "eigenvalues", } u_i \in E^c) = \det(I - KI_E) = \frac{\tau(t, E)}{\tau(t)}\Big|_{t=0},$$

satisfies the partial differential equation (2.3.6) with

$$f := \sum_1^{2r} \frac{\partial \log P}{\partial x_i}, \quad \mathcal{B}_k = \sum_1^{2r} x^{k+1}\frac{\partial}{\partial x_i}.$$

When $E = (x, \infty)$, the equation (2.3.6) for $f = \partial \log P((-\infty, x])/\partial x$ becomes a 3rd order ODE (Chazy-type equation)

$$f''' - 4xf' + 2f + 6f'^2 = 0. \tag{2.3.14}$$

According to [12] , this equation has a first integral, which is a Painlevé II equation

$$f''^2 + 4f'(f'^2 - xf' + f) = 0. \tag{2.3.15}$$

and this equation has a solution given by $f := g'^2 - xg^2 - g^4$ and $f' = -g^2$, which establishes (2.3.5) and (2.3.6).

3. LARGE RANDOM MATRICES AND PERMUTATIONS: A DIRECT CONNECTION VIA ENUMERATIVE GEOMETRY

For large random Hermitian matrices, whose entries have Gaussian real and imaginary part of variance $= 1/2$, the Wigner semi-circle has support on $(-2\sqrt{n}, 2\sqrt{n})$, instead of $[-\sqrt{2n}, \sqrt{2n}]$. Then the (slightly modified) scaling (2.3.4) is given by

$$u_i = n^{2/3}\left(\frac{z_i}{2n^{1/2}} - 1\right) \quad \text{for } n \nearrow \infty. \tag{3.1}$$

Given the set of partitions $\lambda = (\lambda_1 \geq \lambda_2 \geq ...)$ of n, in view of (1.4.1), consider

$$v_i = n^{1/3}\left(\frac{\lambda_i}{2n^{1/2}} - 1\right) \quad \text{for } n \nearrow \infty. \tag{3.2}$$

Theorems 1.3 and 2.3 show that the variables u_1 and v_1 have the same probability distribution. Okounkov [31] shows *in a direct way* that the joint distribution of all u_i and v_i coincide; more specifically:

Theorem 3.1 (Okounkov, [31]). — *For any $m \geq 1$, any $\xi_1, ..., \xi_m > 0$, the following holds:*

$$E\Big(\sum_{i=1}^{\infty} e^{\xi_1 u_i} ... \sum_{i=1}^{\infty} e^{\xi_m u_i}\Big) = E\Big(\sum_{i=1}^{\infty} e^{\xi_1 v_i} ... \sum_{i=1}^{\infty} e^{\xi_m v_i}\Big), \quad \textit{when} \quad n \nearrow \infty.$$

The left hand side, which relates to random permutations, hinges on the following expressions:

$$\left(\frac{1}{2\sqrt{n}}\right)^{k_1+...+k_m} \frac{n^{m/2}}{n!} tr(X_1^{k_1})...X_m^{k_m}))$$
$$= \left(\frac{1}{2}\right)^{k_1+...+k_m} \sum_{\mathbf{S}} n^{\frac{1}{2}(\chi(\mathbf{S})-m)} \left|\mathrm{Cov}_{\mathbf{S}}(k_1, ..., k_m)\right|,$$

In this expression, the X_i are certain matrices, such that e.g. $\lambda_j - j$ is an eigenvalue of X_1, if (j, λ_j) is a corner of the Young diagram (i.e., $\lambda_j > \lambda_{j+1}$). Also $\mathrm{Cov}_{\mathbf{S}}(k_1, ..., k_m)$ are coverings of S^2 ramified according to a rule, given by the numbers $k_1, ..., k_m$.

The right hand side relates to random matrices and has an interpretation as a "map" on a surface. It hinges on the following formula for $M \in U(n)$:

$$\left(\frac{1}{2\sqrt{n}}\right)^{k_1+\ldots+k_m} E(tr(M^{k_1})\ldots tr(M^{k_m})) = \left(\frac{1}{2}\right)^{k_1+\ldots+k_m} \sum_{\mathbf{S}} n^{\chi(\mathbf{S})-m} \left|\mathrm{Map}_{\mathbf{S}}(k_1,\ldots,k_m)\right|,$$

where the sum is taken over all the homeomorphism classes of orientable, not necessarily connected surfaces $\mathbf{S}$; $\chi(\mathbf{S})$ is its Euler number and $\mathrm{Map}_{\mathbf{S}}(k_1,\ldots,k_m)$ denotes the ways to cover the surface $\mathbf{S}$ with m polygons (a k_1-gon, a k_2-gon,..., a k_m-gon) by pairwise gluing the sides of different or the same k_i-gons. Each polygon has a marked vertex, to distinguish a k-gon from its $k-1$ rotations.

The main proposition, established by Okounkov and leading to Theorem 3.1, is the following: for connected surfaces, as $k_i \to \infty$,

$$\mathrm{Cov}_{\mathbf{S}}(k_1,\ldots,k_m) \cong \mathrm{Map}_{\mathbf{S}}(k_1,\ldots,k_m).$$

4. INTEGRALS, MOMENT MATRICES AND INTEGRABLE SYSTEMS

The first column contains a list of matrix integrals and a Fredholm determinant. After perturbing the integral by inserting times t_i's and possibly s_i's, the new integral thus obtained satisfies linear PDE's (Virasoro constraints). The integral can be represented as a determinant of a "moment" matrix, (defined by an appropriate inner-product) or a Pfaffian, if the moment matrix is skew. Performing an appropriate "Borel factorization" of this associated moment matrix, one shows that each of these integrals, as a function of the t_i and s_i's, satisfies an integrable lattice or an integrable PDE. In all the cases, combining both systems of equations leads to ODE's or PDE's for the corresponding integral, in x or in the boundary of E. The last column lists the connection with probability. For more information on the 6th integral and the double matrix integral, see [37, 3, 2].

τ-functions satisfying Virasoro constraints, after inserting t_i's	= determinant or Pfaffian of m_ℓ of the form:	underlying integrable lattice or PDE	connecting with
$\int_{\mathcal{H}_\ell(E)} e^{-trV(M)}dM$	Hänkel	Toda lattice	spectrum of random Hermitian matrices
$\int_{\mathcal{S}_\ell(E)} e^{-trV(M)}dM$	skew-symmetric	Pfaff lattice	spectrum of random symmetric matrices
$\int_{\mathcal{T}_\ell(E)} e^{-trV(M)}dM$	skew-symmetric	Pfaff lattice	spectrum of random symplectic matrices
$\int_{U(\ell)} e^{\sqrt{x}\ tr(M+\bar{M})}dM$	Toeplitz	Toeplitz lattice	longest increasing sequence in random permutations
$\int_{O(\ell)_\pm} e^{x\ trM}dM$	Hänkel	Toda lattice	longest increasing sequence in random involutions
$\int_{U(\ell)} \det(I+M)^k e^{-xtr\bar{M}}dM$	Toeplitz	Toeplitz lattice	longest increasing sequence in random words
$\det\big(I-K(y,z)I_{(x,\infty)}(z)\big)$ (K sine or Airy kernel)	Fredholm determinant	KdV equation	spectrum of infinite random Hermitian matrices (bulk or edge scaling limit)
$\displaystyle\int\int_{\mathcal{H}_\ell\times\mathcal{H}_\ell(E)} dM_1dM_2 \times e^{-tr(M_1^2+M_2^2-cM_1M_2)}$	general	2d-Toda lattice	coupled random matrices

REFERENCES

[1] M. Adler and P. van Moerbeke: *Symmetric random matrices and the Pfaff Lattice*, to appear in: Annals of Mathematics, sept 2000 (`solv-int/9903009`) and *The Hermitian, symmetric and symplectic ensembles and PDE's* (`math.-phys./009001`).

[2] M. Adler and P. van Moerbeke: *The spectrum of coupled random matrices*, Annals of Mathematics, **149**, 921–976 (1999).

[3] M. Adler and P. van Moerbeke: *Integrals over classical groups, random permutations, Toda and Toeplitz lattices*, Comm. Pure Appl. Math. (2000) (`math.CO/9912143`).

[4] M. Adler, T. Shiota and P. van Moerbeke: *Random matrices, vertex operators and the Virasoro algebra*, Phys. Lett. **A 208**, 101-112, (1995). And *Random matrices, Virasoro algebras and non-commutative KP*, Duke Math. J. **94**, 379-431 (1998).

[5] D. Aldous and P. Diaconis: *Longest increasing subsequences: From patience sorting to the Baik-Deift-Johansson theorem*, Bull. Am. Math. Soc. (new series) **36** (4), 413–432 (1999).

[6] J. Baik, P. Deift and K. Johansson: *On the distribution of the length of the longest increasing subsequence of random permutations*, Math. Archive, Journal Amer. Math. Soc. **12**, 1119-1178 (1999) (`math.CO/9810105`).

[7] J. Baik and E. Rains: *Algebraic aspects of increasing subsequences*, (1999), `math.CO/9905083`

[8] D. Bessis, Cl. Itzykson and J.-B. Zuber : *Quantum field theory techniques in graphical enumeration*, Adv. Appl. Math. **1**, 109-157 (1980).

[9] P. Biane: *Representations of symmetric groups and free probability*, preprint (1998).

[10] M. Bowick and E. Brézin: *Universal scaling of the tail of the density of eigenvalues in random matrix models*, Phys. Letters **B 268**, 21–28 (1991).

[11] S. Chadha, G. Mahoux, M.L. Mehta : *A method of integration over matrix variables* : II, J. Phys. A: Math. Gen. 14, 579-586 (1981).

[12] C.M. Cosgrove: *Chazy classes IX-XII of third-order differential equations*, Stud. Appl. Math. **104**, 3, 171-228 (2000).

[13] P. Diaconis, M. Shashahani: *On the eigenvalues of random matrices* J. Appl. Prob., suppl. in honour of Takàcs **31**A, 49-61 (1994).

[14] P. Erdös and G. Szekeres: *A combinatorial theorem in geometry*, Compositio Math., **2**, 463–470 (1935).

[15] A.S. Fokas, A.R. Its, A.V. Kitaev: *The isomonodromy approach to matrix models in 2d quantum gravity*, Comm. Math. Phys., **147**, 395–430 (1992).

[16] P.J. Forrester: *The spectrum edge of random matrix ensembles*, Nucl. Phys. B, (1993).

[17] I. M. Gessel: *Symmetric functions and P-recursiveness*, J. of Comb. Theory, Ser A, **53**, 257–285 (1990).

[18] J.M. Hammersley: *A few seedlings of research*, Proc. Sixth. Berkeley Symp. Math. Statist. and Probability, Vol. 1, 345–394, University of California Press (1972).

[19] M. Hisakado: *Unitary matrix models and Painlevé III*, Mod. Phys. Letters, **A 11** 3001–3010 (1996).

[20] Cl. Itzykson, J.-B. Zuber: *The planar approximation*, J. Math. Phys. **21**, 411–421 (1980).

[21] M. Jimbo, T. Miwa, Y. Mori and M. Sato: *Density matrix of an impenetrable Bose gas and the fifth Painlevé transcendent*, Physica 1D, 80–158 (1980).

[22] K. Johansson: *The Longest Increasing Subsequence in a Random Permutation and a Unitary Random Matrix Model*, Math. Res. Lett., **5**, no. 1-2, 63–82 (1998).

[23] B.F. Logan and L.A. Shepp: *A variational problem for random Young tableaux*, Advances in Math., **26**, 206–222 (1977).

[24] R.D. Kamien, H.D. Politzer, M.B. Wise: *Universality of random-matrix predictions for the statistics of energy levels* Phys. rev. letters **60**, 1995–1998 (1988).

[25] M. Kontsevich: *Intersection theory on the moduli space of curves and the matrix Airy function*, Comm. Math. Phys. **147**, 1-23 (1992).

[26] D. Knuth: "The Art of Computer programming, Vol III: Searching and Sorting", Addison-Wesley, Reading, MA, 1973.

[27] G. Mahoux, M.L. Mehta: *A method of integration over matrix variables: IV*, J. Phys. I (France) **1**, 1093–1108 (1991).

[28] M.L. Mehta: Random matrices, 2nd ed. Boston: Acad. Press, 1991.

[29] T. Nagao, M. Wadati: *Correlation functions of random matrix ensembles related to classical orthogonal polynomials*, J. Phys. Soc of Japan, **60** 3298–3322 (1991).

[30] A.M. Odlyzko: *On the distribution of spacings between the zeros of the zeta function* **48**, 273-308 (1987).

[31] A. Okounkov: *Random matrices and random permutations* (1999), `math.CO/9903176`

[32] L.A. Pastur: *On the universality of the level spacing distribution for some ensembles of random matrices*, Letters Math. Phys., **25** 259–265 (1992).

[33] E. M. Rains: *Topics in Probability on compact Lie groups*, Harvard University doctoral dissertation, (1995).
5B

[34] E. M. Rains: *Increasing subsequences and the classical groups*, Elect. J. of Combinatorics, **5**, R12 (1998).

[35] C.A. Tracy and H. Widom: *Level-Spacings distribution and the Airy kernel*, Commun. Math. Phys., **159**, 151–174 (1994).

[36] C.A. Tracy and H. Widom: *Random unitary matrices, permutations and Painlevé* (1999), `math.CO/9811154`

[37] C.A. Tracy and H. Widom: *On the distribution of the lengths of the longest monotone subsequences in random words* (1999), `math.CO/9904042`

[38] S.M. Ulam: *Monte Carlo calculations in problems of mathematical physics*, in *Modern Mathematics for the Engineers*, E.F. Beckenbach ed., McGraw-Hill, 261–281 (1961).

[39] P. van Moerbeke: *Integrable lattices: random matrices and permutations*, MSRI-volume on Random matrices and exactly solvable models, Eds.: P. Bleher, A. Its, Oxford University press, 2000 (`math.CO/00-10`).

[40] A.M. Vershik and S.V. Kerov: *Asymptotics of the Plancherel measure of the symmetric group and the limiting form of Young tableaux*, Soviet Math. Dokl., **18**, 527–531 (1977).

Pierre VAN MOERBEKE

Département de Mathématiques
Université de Louvain
B–1348 Louvain-la-Neuve (Belgium)
E-mail : `vanmoerbeke@geom.ucl.ac.be`

and

Brandeis University
Waltham, Mass 02454, USA
E-mail : `vanmoerbeke@math.brandeis.edu`

ASTÉRISQUE

2002

276. SÉMINAIRE BOURBAKI, volume 1999/2000, exposés 865-879

2001

275. J.-M. BISMUT, S. GOETTE – *Families torsion and Morse functions*
274. A. BONNET, G. DAVID – *Cracktip is a global Mumford-Shah minimizer*
273. K. NISHIYAMA, H. OCHIAI, K. TANIGUCHI, H. YAMASHITA, S. KATO – *Nilpotent orbits, associated cycles and Whittaker models for highest weight representations*
272. M. BOILEAU, J. PORTI – *Geometrization of 3-orbifolds of cyclic type (avec la collaboration de M. HEUSENER)*
271. M. KASHIWARA, P. SCHAPIRA – *Ind-Sheaves*
270. M. BONK, J. HEINONEN, P. KOSKELA – *Uniformizing Gromov hyperbolic spaces*
269. J.-L. WALDSPURGER – *Intégrales orbitales nilpotentes et endoscopie pour les groupes classiques non ramifiés*

2000

268. J. FRANCHETEAU, G. MÉTIVIER – *Existence de chocs faibles pour des systèmes quasi-linéaires hyperboliques multidimensionnels*
267. R. CERF – *Large Deviations for three Dimensional Supercritical Percolation*
266. SÉMINAIRE BOURBAKI, volume 1998/1999, exposés 850-864
265. O. BIQUARD – *Métriques d'Einstein asymptotiquement symétriques*
264. L. LIPSHITZ, Z. ROBINSON – *Rings of Separated Power Series and Quasi-Affinoid Geometry*
263. C. SABBAH – *Équations différentielles à points singuliers irréguliers et phénomène de Stokes en dimension 2*
262. A. VASY – *Propagation of singularities in three-body scattering*
261. Géométrie complexe et systèmes dynamiques, colloque en l'honneur d'Adrien Douady, Orsay 1995, M. FLEXOR, P. SENTENAC et J.-C. YOCCOZ, éditeurs

1999

260. S. D. CUTKOSKY – *Local monomialization and factorization of morphisms*
259. R. KRIKORIAN – *Réductibilité des systèmes produits-croisés à valeurs dans des groupes compacts*
258. Structure Theory of Set Addition, J.-M. DESHOUILLERS, B. LANDREAU and A.A. YUDIN, editors
257. J.-P. LABESSE – *Cohomologie, stabilisation et changement de base (avec la collaboration de L. BREEN et L. CLOZEL)*
256. F. MOREL – *Théorie homotopique des schémas*
255. R.E. KOTTWITZ, D. SHELSTAD – *Foundations of twisted endoscopy*
254. C.J. BUSHNELL, G. HENNIART – *Local tame lifting for* $\mathrm{GL}(n)$ *II : wildly ramified supercuspidals*
253. B. MAGNERON – *Involutions complexes et vecteurs sphériques associés pour les groupes de Lie nilpotents réels*

1998

252. SÉMINAIRE BOURBAKI, volume 1997/1998, exposés 835-849
251. Nombre et répartition de points de hauteur bornée, E. PEYRE, éditeur
250. C. BONATTI, R. LANGEVIN – *Difféomorphismes de Smale des surfaces (avec la collaboration de E. JEANDENANS)*
249. P. AUSCHER, P. TCHAMITCHIAN – *Square root problem for divergence operators and related topics*
248. P. COLMEZ – *Intégration sur les variétés p-adiques*
247. G. PISIER – *Non-commutative vector valued L_p-spaces and completely p-summing maps*

1997

246. V. TARASOV, A. VARCHENKO – *Geometry of q-hypergeometric functions, quantum affine algebras and elliptic quantum groups*
245. SÉMINAIRE BOURBAKI, volume 1996/1997, exposés 820-834
244. J.-M. BISMUT – *Holomorphic families of immersions and higher analytic torsion forms*
243. L. LAFFORGUE – *Chtoucas de Drinfeld et conjecture de Ramanujan-Petersson*
242. N. BURQ – *Pôles de diffusion engendrés par un coin*
241. SÉMINAIRE BOURBAKI, volume 1995/1996, exposés 805-819

1996

240. A. SÀ BARETTO, R. B. MELROSE, M. ZWORSKI – *Semilinear diffraction of conormal waves*
239. J.-L. VERDIER – *Des catégories dérivées des catégories abéliennes*
238. A. BROISE, F. DAL'BO, M. PEIGNÉ – *Méthodes des opérateurs de transfert : transformations dilatantes de l'intervalle et dénombrement de géodésiques fermées*
237. SÉMINAIRE BOURBAKI, volume 1994/1995, exposés 790-804
236. Hommage à P. A. MEYER et J. NEVEU
235. J.-P. OTAL – *Le théorème d'hyperbolisation pour les variétés fibrées de dimension 3*
234. A. GENESTIER – *Espaces symétriques de Drinfeld*

1995

233. I. KRIZ, J.P. MAY – *Operads, algebras modules and motives*
232. Recent advances in operator algebras (Orléans, 1992)
231. J.-C. YOCCOZ – *Petits diviseurs en dimension 1*
230. J.-Y. CHEMIN – *Fluides parfaits incompressibles*
229. B. PERRIN-RIOU – *Fonctions L p-adiques des représentations p-adiques*
228. Columbia University number theory seminar (New-York, 1992)
227. SÉMINAIRE BOURBAKI, volume 1993/1994, exposés 775-789

1994

226. K-theory (Strasbourg, 1992)
225. L. BREEN – *On the classification of 2-gerbes and 2-stacks*
224. P. SCHAPIRA, J.-P. SCHNEIDERS – *Index theorem for elliptic pairs*
223. Périodes p-adiques (séminaire de Bures, 1988)
222. Complex analytic methods in dynamical systems (IMPA, janvier 1992)
221. A. OGUS – *F-Crystals, Griffiths transversality, and the Hodge decomposition*
220. H.H. ANDERSEN, J.C. JANTZEN, W. SOERGEL – *Représentations of quantum groups at a p-th root of unity and of semisimple groups in characteristic p : independence of p*
219. H. RUBENTHALER – *Les paires duales dans les algèbres de Lie réductives*