332

ASTÉRISQUE

2010

SÉMINAIRE BOURBAKI
VOLUME 2008/2009
EXPOSÉS 997-1011

(997) *Dimères et surfaces aléatoires*

Raphaël CERF

SOCIÉTÉ MATHÉMATIQUE DE FRANCE
Publié avec le concours du CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

Séminaire BOURBAKI Novembre 2008
61e année, 2008-2009, no 997, p. 1 à 9

DIMÈRES ET SURFACES ALÉATOIRES
[d'après les travaux de Kenyon et d'Okounkov]

par **Raphaël CERF**

INTRODUCTION

Lorsqu'un système présente un phénomène de transition de phases, il est possible de forcer la coexistence de deux ou plusieurs phases distinctes. Ces phases sont délimitées par des surfaces de séparation. À l'échelle macroscopique, ces surfaces se stabilisent près d'une forme déterministe, mais elles fluctuent dans une échelle intermédiaire entre l'échelle microscopique et l'échelle macroscopique. La compréhension de la nature et la description quantitative de ces fluctuations constituent des problèmes extrêmement ardus. Par exemple, dans le modèle d'Ising en dimension trois, nous sommes encore très loin de posséder les outils nécessaires pour conduire une analyse rigoureuse et précise des fluctuations des interfaces aléatoires entre les deux phases du modèle. Plutôt que d'attaquer directement ce problème dont les difficultés semblent formidables, une approche plus modeste consiste à considérer des modèles effectifs d'interface. Ce sont des modèles de surfaces aléatoires qui sont construits en spécifiant explicitement la loi de la surface, par exemple le modèle SOS ou les modèles d'interface gaussiens. Les modèles de dimères étudiés par Kenyon et Okounkov donnent lieu à des surfaces aléatoires qui appartiennent en quelque sorte à une catégorie intermédiaire entre les interfaces générées par le modèle d'Ising et celles du modèle SOS. Ce sont des interfaces construites au-dessus d'un système probabiliste complexe, qui présente des transitions de phase, et dont l'analyse peut être conduite grâce à l'étude asymptotique de certains déterminants, qui donnent des formules exactes.

Dans cet exposé introductif, nous présentons brièvement le modèle de dimères, ainsi que certaines formules essentielles. Nous définissons les objets importants qui permettent l'analyse du modèle : la matrice de Kasteleyn, la courbe spectrale, l'amibe. Kenyon et Okounkov ont étudié en profondeur le lien entre les modèles de dimères et une certaine catégorie de courbes algébriques appelées les courbes de Harnack.

Nous énonçons ensuite le résultat de Okounkov, Kenyon et Sheffield qui permet de localiser les phases du modèle grâce à l'amibe. Nous terminons en présentant le début de l'étude des formes asymptotiques réalisée par Kenyon et Okounkov. En plus des articles originaux [**5, 6, 7, 8**], le lecteur pourra consulter les notes du cours de Saint-Flour de Kenyon [**4**].

1. CONFIGURATIONS DE DIMÈRES

Nous considérons un graphe planaire G qui est biparti et $\mathbb{Z}^2$ périodique. Biparti signifie qu'il est possible de colorier les sommets du graphe avec deux couleurs, par exemple en blanc et noir, de manière à ce que deux sommets voisins soient toujours de couleurs distinctes. La propriété $\mathbb{Z}^2$ périodique signifie que le graphe et son coloriage en blanc et noir sont préservés par les translations à coordonnées entières. Des exemples standards de tels graphes sont le réseau carré ou le réseau hexagonal. Une configuration de dimères sur le graphe G est un sous-ensemble d'arêtes de G tel que chaque sommet soit incident à exactement une arête de ce sous-ensemble. En théorie des graphes, un recouvrement par des dimères s'appelle aussi un couplage parfait. Nous notons $\mathcal{M}(G)$ l'ensemble des configurations de dimères sur le graphe G.

2. FONCTIONS DE HAUTEURS

À une configuration de dimères sur G est naturellement associé un flot unité qui va des sommets blancs vers les sommets noirs : chaque sommet blanc est une source de débit 1 et chaque sommet noir est un puits de débit 1. Fixons M_0 une configuration périodique de dimères sur G et notons ω_0 le flot associé. Si M est une autre configuration de dimères sur G et si ω est le flot associé, alors la différence des deux flots $\omega - \omega_0$ est un flot de divergence nulle. Fixons une face f_0. Soit γ un chemin dans le graphe dual G^* de G qui va de f_0 à une face f. Le flux total de $\omega - \omega_0$ à travers γ ne dépend pas du choix particulier de γ et est donc une fonction de la face f. Cette fonction s'appelle la fonction de hauteur associée à la configuration de dimères M. La fonction de hauteur dépend du choix de la face f_0 ainsi que de la configuration M_0. Cependant, la différence entre deux fonctions de hauteurs associées à deux configurations de dimères ne dépend plus de M_0 ni de f_0.

3. MESURES DE GIBBS

Une mesure de probabilité sur l'ensemble $\mathcal{M}(G)$ des configurations de dimères est construite de la manière suivante : on attribue à chaque arête e une énergie $\mathcal{E}(e)$. La fonction d'énergie d'une configuration de dimères est égale à la somme des énergies de ses arêtes :

$$\mathcal{E}(M) = \sum_{e \in M} \mathcal{E}(e).$$

Une mesure de Gibbs sur $\mathcal{M}(G)$ associée à l'énergie $\mathcal{E}$ est une mesure de probabilité qui vérifie la condition suivante : étant donné un sous-graphe fini F de G, la probabilité conditionnelle de la configuration M_F des dimères dans F connaissant la configuration des dimères en dehors de F est proportionnelle à $\exp -\mathcal{E}(M_F)$: le facteur de proportionnalité est choisi de manière à ce que la somme des probabilités fasse un et vaut donc

$$Z_F = \sum_{M_F \in \mathcal{M}(F)} \exp -\mathcal{E}(M_F).$$

4. MATRICE DE KASTELEYN

Soit F un graphe planaire fini biparti. Un poids de Kasteleyn pour F est un choix de signe pour chaque arête de F qui a la propriété suivante : une face qui a un nombre d'arêtes égal à 0 modulo 4 a un nombre impair de signes moins et une face qui a un nombre d'arêtes égal à 2 modulo 4 a un nombre pair de signes moins. Une matrice de Kasteleyn pour F est une matrice indexée par $B \times W$, où B (respectivement W) est l'ensemble des sommets noirs (respectivement blancs) de F et telle que $K(b, w) = 0$ s'il n'existe pas d'arête joignant b à w, et sinon

$$K(b, w) = \text{weight}(\langle b, w \rangle) \exp -\mathcal{E}(\langle b, w \rangle),$$

où $\text{weight}(\langle b, w \rangle)$ est le poids de Kasteleyn de l'arête joignant b à w. Kasteleyn **[2]** a prouvé que la fonction de partition Z_F était égale au déterminant d'une telle matrice de Kasteleyn :

$$|\det K| = Z_F.$$

Cette formule est fondamentale, car elle permet d'obtenir le comportement asymptotique de la fonction de partition dans la limite thermodynamique via l'étude du comportement asymptotique d'un déterminant.

5. GRAPHES SUR LE TORE

Nous supposons que le graphe planaire G est $\mathbb{Z}^2$ périodique. Notons alors G_1 le graphe quotient de G par l'action de $\mathbb{Z}^2$; c'est un graphe fini biparti sur le tore. Pour de tels graphes, nous pouvons encore construire une matrice de Kasteleyn comme ci-dessus, et aussi définir une mesure de Gibbs. La fonction de partition de la mesure de Gibbs s'exprime alors comme une somme de quatre déterminants

$$Z = \frac{1}{2}(-Z^{(00)} + Z^{(10)} + Z^{(01)} + Z^{(11)})$$

où $Z^{(\theta\tau)}$ est le déterminant de la matrice de Kasteleyn dans laquelle les signes des arêtes qui croisent un cycle dual horizontal fixé ont été multipliés par $(-1)^\theta$ et les signes des arêtes qui croisent un cycle dual vertical fixé ont été multipliés par $(-1)^\tau$.

6. POLYNÔME CARACTÉRISTIQUE

Soit K une matrice de Kasteleyn pour le graphe G_1 et soient z, w deux paramètres. Soit γ_x (respectivement γ_y) un chemin dans le graphe dual de G_1 qui fait un tour horizontal (respectivement vertical) autour du tore. Le poids de chaque arête traversée par γ_x est multiplié par z si le sommet noir de l'arête est à gauche de γ_x et par z^{-1} s'il est à droite. De même, le poids de chaque arête traversée par γ_y est multiplié par w si le sommet noir de l'arête est au-dessus de γ_y et par w^{-1} s'il est en dessous. Le polynôme caractéristique de G est alors défini par

$$P(z, w) = \det K(z, w).$$

Modulo les symétries $z \mapsto -z$ et $w \mapsto -w$, cette définition ne dépend pas du choix des signes utilisés dans la construction de la matrice de Kasteleyn K. Nous avons alors

$$P((-1)^\theta, (-1)^\tau) = Z^{(\theta\tau)}$$

et la fonction de partition s'exprime à l'aide du polynôme caractéristique

$$Z = \frac{1}{2}(-P(1,1) + P(1,-1) + P(-1,1) + P(-1,-1)).$$

En fait, le comportement à grande échelle du modèle de dimères se décrit grâce au polynôme caractéristique.

7. CHAMP MAGNÉTIQUE

Nous introduisons un champ magnétique dans le modèle de la manière suivante. Notons $(B_x, B_y) \in \mathbb{R}^2$ le champ magnétique. Soit γ_x un chemin dans le dual de G_1 qui

fait exactement un tour horizontal autour du tore. Ce chemin croise k arêtes de G_1. Pour chacune de ces k arêtes, nous ajoutons à son énergie B_x/k si le sommet de l'arête qui se trouve au-dessus de γ_x est noir et $-B_x/k$ si le sommet de l'arête qui se trouve au-dessus de γ_x est blanc. De même, soit γ_y un chemin dans le dual de G_1 qui fait exactement un tour vertical autour du tore. Ce chemin croise l arêtes de G_1. Pour chacune de ces l arêtes, nous ajoutons à son énergie B_y/l si le sommet de l'arête qui se trouve à gauche de γ_y est noir et $-B_y/l$ si le sommet de l'arête qui se trouve à gauche de γ_y est blanc. Nous obtenons ainsi une nouvelle fonction d'énergie qui dépend du champ magnétique $(B_x, B_y) \in \mathbb{R}^2$. Notons que la variation d'énergie d'une configuration de dimères due au champ magnétique vaut $B_x h_x + B_y h_y$, où (h_x, h_y) est le changement de hauteur de la configuration. En particulier, cette variation ne dépend pas du choix spécifique des chemins γ_x et γ_y.

8. LIMITE THERMODYNAMIQUE DE L'ÉNERGIE LIBRE

Notons maintenant G_n le graphe quotient de G par l'action de $n\mathbb{Z}^2$. Nous allons étudier le comportement asymptotique du modèle de dimères sur G_n lorsque n tend vers l'infini. Les polynômes caractéristiques associés au graphe G_n se calculent récursivement via la formule

$$P_n(z,w) \;=\; \prod_{z_0^n = z} \prod_{w_0^n = w} P(z_0, w_0).$$

Le logarithme de cette expression peut se réécrire comme une somme de Riemann sur le tore unité

$$\mathbb{T}^2 \;=\; \big\{\, (z,w) \in \mathbb{C}^2 : |z| = |w| = 1 \,\big\}.$$

Ceci, combiné avec un argument de sous-additivité standard, permet de montrer que

$$\lim_{n\to\infty} \frac{1}{n^2} \ln Z(G_n) \;=\; \frac{1}{(2\pi i)^2} \int_{\mathbb{T}^2} \ln |P(z,w)| \, \frac{dz}{z} \frac{dw}{w}.$$

La courbe spectrale du modèle de dimères est la courbe dans $\mathbb{C}^2$ d'équation $P(z,w) = 0$.

9. L'AMIBE DU MODÈLE DE DIMÈRES

La fonction de Ronkin du polynôme $P(z,w)$ est la fonction $\mathscr{R}(x,y)$ définie par

$$\mathscr{R}(x,y) \;=\; \frac{1}{(2\pi i)^2} \int_{\mathbb{T}^2} \ln |P(e^x z, e^y w)| \, \frac{dz}{z} \frac{dw}{w}.$$

L'amibe du polynôme $P(z,w)$ est l'image de la courbe complexe d'équation $P(z,w)=0$ par l'application $(z,w)\mapsto(\ln|z|,\ln|w|)$, i.e.

$$\mathbb{A}(P) \;=\; \{\,(\ln|z|,\ln|w|)\in\mathbb{R}^2 : P(z,w)=0\,\}.$$

La fonction de Ronkin de P est convexe dans $\mathbb{R}^2$ et linéaire sur chaque composante du complémentaire de $\mathbb{A}(P)$. L'amibe a des « tentacules » qui sont les régions où $z\to 0,\infty$ ou bien $w\to 0,\infty$. Chaque tentacule est asymptotique à une ligne d'équation

$$\alpha\ln|z|+\beta\ln|w|+\gamma=0.$$

Ces tentacules divisent le complémentaire de l'amibe en un certain nombre de composantes non bornées.

10. COURBES DE HARNACK

Les courbes spectrales qui proviennent d'un modèle de dimères sont des courbes particulières appelées courbes de Harnack. Une courbe de Harnack est une courbe C d'équation $P(z,w)=0$ telle que chaque point de l'amibe $\mathbb{A}(P)$ ait au plus deux antécédents par l'application

$$(z,w)\in C\mapsto(\ln|z|,\ln|w|)\in\mathbb{A}(P).$$

En fait, tous les points de l'amibe $\mathbb{A}(P)$ ont deux antécédents par cette application, sauf un nombre fini d'entre eux qui n'en ont qu'un. La connexion entre les modèles de dimères et les courbes de Harnack est étudiée dans **[6]**.

THÉORÈME 10.1 (**[6, 8]**). — *La courbe spectrale d'un modèle de dimères est une courbe de Harnack. Inversement, toute courbe de Harnack peut être réalisée comme la courbe spectrale d'un modèle de dimères avec poids sur un graphe biparti périodique.*

Les courbes de Harnack de genre zéro jouent un rôle spécial : ce sont les courbes spectrales des dimères isoradiaux étudiés dans **[3]**. Ce sont aussi les courbes de Harnack qui minimisent le volume sous la fonction de Ronkin associée à des conditions au bord fixées. Du point de vue probabiliste, cela signifie que les poids des dimères isoradiaux maximisent la fonction de partition à conditions au bord fixées.

11. LE DIAGRAMME DE PHASES

Les mesures de Gibbs ergodiques associées au modèle de dimères peuvent présenter différents types de comportement, décrits par trois phases : la phase gelée, la phase liquide et la phase gazeuse. Cette classification repose sur les fluctuations de la fonction

hauteur. Soient f et f' deux faces du graphe G ; considérons la différence de hauteurs $h(f) - h(f')$. Une mesure de Gibbs ergodique est dans la phase gelée si certaines différences de hauteurs sont déterministes. Elle est dans la phase gazeuse ou lisse si les fluctuations de la hauteur ont une variance bornée, i.e. la variance de la différence de hauteurs $h(f) - h(f')$ est bornée indépendamment de f et f'. Enfin, elle est dans la phase liquide ou rugueuse si les fluctuations de la hauteur n'ont pas une variance bornée.

Notons $\widetilde{\mu}_n(B_x, B_y)$ la mesure de Gibbs sur le graphe G_n soumise au champ magnétique $(B_x, B_y) \in \mathbb{R}^2$. Un théorème de Sheffield [**9**] garantit l'existence de la limite

$$\widetilde{\mu}(B_x, B_y) = \lim_{n\to\infty} \widetilde{\mu}_n(B_x, B_y).$$

Le résultat central de Kenyon, Okounkov et Sheffield [**8**] sur les phases de la mesure de Gibbs $\widetilde{\mu}(B_x, B_y)$ est le suivant.

THÉORÈME 11.1. — *La mesure de Gibbs $\widetilde{\mu}(B_x, B_y)$ se trouve respectivement dans les phases gelées, liquides ou gazeuses selon que le champ magnétique (B_x, B_y) se trouve respectivement dans la fermeture d'une composante non bornée du complémentaire de l'amibe $\mathbb{A}(P)$, dans l'intérieur de l'amibe $\mathbb{A}(P)$, ou dans la fermeture d'une composante bornée du complémentaire de l'amibe.*

Dans la phase liquide, les corrélations entre arêtes décroissent polynomialement, dans la phase gazeuse elles décroissent exponentiellement, et dans la phase gelée, certaines corrélations ne décroissent pas du tout.

12. FORME ASYMPTOTIQUE

Considérons les surfaces dans $\mathbb{R}^3$ qui s'obtiennent comme le bord d'une union de cubes unités disposés sur le réseau cubique $\mathbb{Z}^3$. Nous nous restreignons aux surfaces monotones, i.e. les surfaces qui se projettent bijectivement selon la direction $(1, 1, 1)$ sur un contour polygonal. Soit C une courbe dans $\mathbb{Z}^3$. Les surfaces du type décrit précédemment qui s'appuient sur la courbe C sont en bijection avec les pavages de la région Ω délimitée par la projection selon la direction $(1, 1, 1)$ de C, pavages qui utilisent trois types de losanges. Ils sont également en bijection avec les recouvrements par des dimères du graphe hexagonal. Soit maintenant C_n une suite de courbes dans $\mathbb{Z}^3$ telles que $n^{-1}C_n$ converge vers une courbe C dans $\mathbb{R}^3$. Choisissons au hasard selon la loi uniforme une surface qui s'appuie sur le contour C_n. Un résultat de loi des grands nombres [**1**] dit que, lorsque n tend vers l'infini, une telle surface aléatoire

converge vers une surface déterministe qui s'appuie sur la courbe limite C. Cette forme asymptotique est la solution du problème variationnel

$$\text{minimiser} \quad \int_\Omega \sigma(\nabla h(x,y))\, dx\, dy$$

sur l'ensemble des fonctions lipschitziennes h satisfaisant la condition au bord (la restriction du graphe de h à $\partial\Omega$ est C). La fonction σ est appelée la tension de surface. Pour un modèle de dimères avec poids sur un graphe biparti périodique, la tension de surface est la transformée de Legendre de la fonction de Ronkin de la courbe spectrale du modèle. La compréhension de la structure de la forme asymptotique nécessite d'étudier précisément la tension de surface, notamment sa régularité. En particulier, nous voudrions comprendre les mécanismes qui conduisent à la formation de faces sur la forme asymptotique. Les faces apparaissent lorsque la tension de surface n'est pas strictement convexe.

Dans l'article [**7**], Kenyon et Okounkov effectuent un changement de variables dans l'équation d'Euler–Lagrange associée au problème variationnel ci-dessus qui permet de se ramener à l'équation de Burgers complexe. Cette équation peut être résolue à l'aide d'une fonction holomorphe arbitraire en général. Cependant, pour un ensemble de conditions au bord dense, cette fonction holomorphe est algébrique. Ceci permet d'utiliser des outils de géométrie algébrique pour étudier les solutions du problème variationnel et la formation des singularités.

RÉFÉRENCES

[1] H. Cohn, R. Kenyon & J. Propp – A variational principle for domino tilings, *J. Amer. Math. Soc.* **14** (2001), p. 297–346.

[2] P. W. Kasteleyn – Graph theory and crystal physics, in *Graph Theory and Theoretical Physics*, Academic Press, 1967, p. 43–110.

[3] R. Kenyon – The Laplacian and Dirac operators on critical planar graphs, *Invent. Math.* **150** (2002), p. 409–439.

[4] ———, Lectures on dimers, notes du cours de Saint-Flour, 2007.

[5] R. Kenyon & A. Okounkov – What is ... a dimer ?, *Notices Amer. Math. Soc.* **52** (2005), p. 342–343.

[6] ———, Planar dimers and Harnack curves, *Duke Math. J.* **131** (2006), p. 499–524.

[7] ———, Limit shapes and the complex Burgers equation, *Acta Math.* **199** (2007), p. 263–302.

[8] R. Kenyon, A. Okounkov & S. Sheffield – Dimers and amoebae, *Ann. of Math.* **163** (2006), p. 1019–1056.

[9] S. Sheffield – Random surfaces, *Astérisque* **304** (2005).

Raphaël CERF

Université Paris XI
Département de Mathématiques
UMR 8628 du CNRS
Bâtiment 425
F-91405 Orsay Cedex
E-mail : rcerf@math.u-psud.fr

332

ASTÉRISQUE

2010

SÉMINAIRE BOURBAKI
VOLUME 2008/2009
EXPOSÉS 997-1011

(998) *Le groupe de Cremona et ses sous-groupes de type fini*

Charles FAVRE

SOCIÉTÉ MATHÉMATIQUE DE FRANCE
Publié avec le concours du CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

Séminaire BOURBAKI Novembre 2008
61e année, 2008-2009, n° 998, p. 11 à 43

LE GROUPE DE CREMONA ET SES SOUS-GROUPES DE TYPE FINI

par **Charles FAVRE**

INTRODUCTION

Le groupe de Cremona $\mathrm{Cr}(2)$ est le groupe constitué des applications birationnelles du plan projectif $\mathbb{P}^2(\mathbb{C})$. L'étude de ce groupe et de ses éléments fut initiée par E. De Jonquières, L. Cremona et M. Noether à la fin du XIXe siècle et leurs travaux ont largement inspiré les géomètres italiens du début du XXe siècle. Peu à peu cependant, le groupe de Cremona a quitté sa position centrale en géométrie algébrique. Comme nous le verrons, de nombreux aspects de la structure de $\mathrm{Cr}(2)$ restent encore mystérieux.

C'est un théorème dû à M. Noether que $\mathrm{Cr}(2)$ est engendré par $\mathrm{PGL}(3,\mathbb{C})$ et l'involution $[x:y:z]\mapsto[yz:zx:xy]$ dite involution de Cremona. Au début des années 80, M. Gizatullin [**51**] puis V.I. Iskhovskikh [**58**] ont donné une présentation de $\mathrm{Cr}(2)$ par générateurs et relations. Bien que fondamental, ce résultat est en pratique délicat à appliquer. On ne sait par exemple pas si $\mathrm{Cr}(2)$ est un groupe simple[(1)], ou même quelles sont les structures de groupe topologique sur $\mathrm{Cr}(2)$.

Un moyen pour appréhender la structure de $\mathrm{Cr}(2)$ est d'en étudier les sous-groupes. L'étude des sous-groupes finis de $\mathrm{Cr}(2)$ a été récemment complétée par I. Dolgachev et V.I. Iskhovskikh [**35**] et nous renvoyons à l'exposé de J.-P. Serre [**71**] à ce séminaire pour un rapport détaillé sur la question. De même, les sous-groupes continus de $\mathrm{Cr}(2)$ ont été décrits par L. Enriques [**39**] en 1893[(2)]. Un tel groupe est conjugué à un sous-groupe du groupe des automorphismes d'une surface torique (voir [**9**] pour plus de précisions).

(1) Le groupe des automorphismes polynomiaux du plan de jacobien 1 n'est pas simple, voir [**21**], et S. Cantat et S. Lamy ont annoncé l'analogue pour $\mathrm{Cr}(2)$.
(2) Voir [**24**] pour des résultats en dimension quelconque et [**76, 77**] pour une classification précise en dimension 3.

Ici nous nous intéresserons principalement aux sous-groupes G de $\mathrm{Cr}(2)$ infinis et *de type fini*. Nous allons décrire les résultats remarquables obtenus récemment par J. Deserti et S. Cantat sur ces groupes et les conséquences que l'on peut en tirer pour le groupe de Cremona lui-même.

Leurs travaux s'articulent autour de la question suivante : étant donné un groupe G de type fini, peut-on trouver un morphisme de groupe $\rho : G \to \mathrm{Cr}(2)$ qui soit injectif ? Pour traiter ce type de question, on dispose de deux méthodes. La première consiste à utiliser notre connaissance sur la structure des deux groupes G et $\mathrm{Cr}(2)$. Dans ce cas, les théorèmes de Noether, de Gizatullin et d'Iskhovskikh jouent un rôle crucial. La seconde consiste à faire agir le groupe d'arrivée sur un espace géométrique convenable, puis à analyser l'action induite par G sur cet espace. C'est cette méthode que nous allons mettre en œuvre ici, en nous reposant sur le fait établi par S. Cantat que $\mathrm{Cr}(2)$ agit par isométries sur un espace métrique hyperbolique au sens de Gromov. Ceci permet d'utiliser la théorie des actions de groupes sur de tels espaces, pour laquelle on dispose de nombreux outils, voir par exemple **[12, 49]**.

Afin de motiver l'introduction de cet espace, penchons-nous tout d'abord sur le cas de la dimension 1, c'est-à-dire sur le groupe des transformations de Möbius $\mathrm{PGL}(2,\mathbb{C})$. Ce groupe agit holomorphiquement sur l'espace projectif $\mathbb{P}^1(\mathbb{C})$, et par isométries sur l'espace hyperbolique de dimension 3 réelle $\mathbb{H}^3$.

Une application birationnelle $f \in \mathrm{Cr}(2)$ agit de même naturellement sur $\mathbb{P}^2(\mathbb{C})$. Cependant, cette action n'est pas holomorphe car f peut avoir des points d'indétermination. De plus, l'itération de f conduit en général à des phénomènes dynamiques extrêmement complexes, voir **[73]** : cette action n'est donc pas maniable pour le problème qui nous concerne.

Il a été progressivement réalisé que de nombreux aspects dynamiques d'une application birationnelle étaient contrôlés par son action sur la cohomologie $\mathrm{H}^2(X,\mathbb{R})$ d'un modèle birationnel adéquat X de $\mathbb{P}^2(\mathbb{C})$. La construction d'un tel modèle n'étant pas canonique, on est amené à considérer l'espace (de dimension infinie) de toutes les classes de cohomologie de tous les modèles birationnels de $\mathbb{P}^2(\mathbb{C})$, espace dont la construction précise a été donnée par Y.I. Manin **[66]**. Sa complétion pour la forme bilinéaire induite par le cup produit définit un espace de Hilbert réel que l'on notera $\mathrm{L}^2(\mathfrak{P})$ et qui est muni d'une forme d'intersection de type Minkowski[3] $(\alpha,\beta) \mapsto \alpha\cdot\beta$. Comme en dimension finie, l'une des deux nappes de l'ensemble $\{\alpha \in \mathrm{L}^2(\mathfrak{P}), \alpha\cdot\alpha = +1\}$ possède une métrique $d_{\mathbb{H}}$ qui en fait un espace hyperbolique au sens de Gromov. On notera $\mathbb{H}(\mathfrak{P})$ cet espace métrique.

FAIT. — *Le groupe de Cremona* $\mathrm{Cr}(2)$ *agit par isométries sur l'espace* $\mathbb{H}(\mathfrak{P})$.

[3] C'est-à-dire possédant un vecteur d'auto-intersection strictement positive, telle que la restriction de la forme sur l'orthogonal de ce vecteur soit définie négative.

Il existe une classification grossière des isométries d'un espace hyperbolique [**49**] : on parle d'isométrie hyperbolique, parabolique ou elliptique. On peut donc introduire la notion d'application birationnelle hyperbolique, parabolique ou elliptique suivant le type de son action induite sur $\mathbb{H}(\mathfrak{P})$.

Un résultat dû à J. Diller et moi-même [**31**] donne la caractérisation géométrique suivante de chacun de ces trois types. Une application birationnelle f est elliptique si elle fixe une classe $\alpha \in \mathbb{H}(\mathfrak{P})$. On montre que c'est le cas précisément lorsqu'un itéré de f induit un automorphisme sur une surface rationnelle minimale. Une application f est parabolique si elle n'est pas elliptique et si elle fixe une classe non-nulle $\alpha \in \mathrm{L}^2(\mathfrak{P})$ d'auto-intersection nulle. Dans ce cas, on montre que la classe α est déterminée par la fibre d'une fibration rationnelle ou elliptique f-invariante (dans un modèle birationnel adéquat). Enfin f est hyperbolique lorsque le rayon spectral $\lambda(f)$ de son action sur $\mathrm{L}^2(\mathfrak{P})$ vérifie $\lambda(f) > 1$. Dans ce cas, l'opérateur induit par f sur $\mathrm{L}^2(\mathfrak{P})$ possède deux vecteurs propres simples et d'auto-intersection nulle dont les valeurs propres associées sont $\lambda(f)$ et $\lambda(f)^{-1}$ respectivement. Nous renvoyons au théorème 2.7 ci-dessous pour un énoncé plus complet.

Définition. — *Un sous-groupe G de type fini de $\mathrm{Cr}(2)$ est dit élémentaire si G préserve une classe $\alpha \in \mathrm{L}^2(\mathfrak{P})$ non nulle telle que $\alpha \cdot \alpha \geq 0$.*

Les résultats principaux de S. Cantat [**17**] peuvent alors s'énoncer sous la forme suivante.

Théorème A. — *Soit G un sous-groupe de $\mathrm{Cr}(2)$ de type fini. Si G n'est pas élémentaire, alors il contient deux éléments hyperboliques qui engendrent un groupe libre non-abélien.*

Définition. — *Soit G un sous-groupe de type fini de $\mathrm{Cr}(2)$ élémentaire.*

- *G est dit hyperbolique s'il contient un élément hyperbolique.*
- *G est dit parabolique s'il contient un élément parabolique.*
- *G est dit elliptique s'il ne contient que des éléments elliptiques.*

Théorème B. — *Soit G un sous-groupe de $\mathrm{Cr}(2)$ de type fini et élémentaire.*

1. *Si G est de type hyperbolique, il existe un sous-groupe G_0 d'indice au plus 2 dans G, et un morphisme surjectif $\rho : G_0 \to \mathbb{Z}$ dont le noyau ne contient que des éléments elliptiques, et tel que $\rho(f) \neq 0$ si et seulement si f est hyperbolique.*
2. *Si G est de type parabolique, il existe une surface rationnelle X et une fibration rationnelle ou elliptique sur X qui est préservée par tous les éléments de G. De plus G ne contient aucun élément hyperbolique.*

3. *Si G est de type elliptique, alors soit il possède un sous-groupe distingué d'indice fini conjugué à un sous-groupe de* $\mathrm{PGL}(3,\mathbb{C})$ *; soit il existe une surface rationnelle X et une fibration rationnelle sur celle-ci préservée par tous les éléments du groupe.*

S. Lamy [**62**] avait précédemment donné une classification des sous-groupes du groupe $\mathrm{Aut}[\mathbb{C}^2]$ des automorphismes polynomiaux du plan, et ce sans hypothèse de type fini. Nous renvoyons au théorème 1.9 pour un énoncé précis.

Nous donnerons une preuve détaillée des théorèmes A et B à la section 3. Notons que ces résultats sont en fait vrais pour toute surface kählérienne : nous nous sommes restreint au plan projectif car les mêmes méthodes s'appliquent dans le cas général et sont plus faciles à mettre en œuvre. En effet, pour les surfaces de dimension de Kodaira non négative, une application biméromorphe est holomorphe sur son modèle minimal. De même, une application birationnelle d'une surface réglée non rationnelle préserve nécessairement le réglage.

Notons enfin que les méthodes que nous utiliserons étant purement algébriques, elles permettent aussi de traiter le cas du groupe de Cremona sur n'importe quel corps algébriquement clos de caractéristique nulle. Dans ce cadre, il faut remplacer l'action sur la cohomologie par l'action sur le groupe de Néron-Severi.

Avertissement

Afin de garder ce texte court, nous n'évoquerons pas les aspects concernant l'étude dynamique des itérés d'une application birationnelle, pour laquelle il existe une littérature importante, voir [**2, 38, 73**] et les références qui s'y trouvent. De même, nous n'aborderons pas le problème de construction d'automorphismes de type hyperbolique sur les surfaces rationnelles. Nous renvoyons à [**68**] ou aux récents travaux de E. Bedford et K. Kim [**3, 4**] sur ce sujet. Enfin, nous avons délibérément choisi de ne pas parler de la dimension supérieure : aucune des techniques présentées ici ne se généralise de manière évidente dans ce cadre.

Je remercie chaleureusement S. Cantat, J. Deserti, R. Dujardin, J.-F. Quint et A. Zuk pour leur aide dans la préparation de cet exposé.

1. APPLICATIONS

Commençons par quelques applications des théorèmes A et B. Nous verrons qu'il existe une analogie particulièrement frappante entre le groupe de Cremona et les

groupes linéaires[4], ce qui motive quelques questions que nous avons rassemblées à la fin de cette partie.

1.1. Groupes de Kazhdan et groupe de Cremona

La propriété (T) de Kazhdan pour un groupe topologique G localement compact est la propriété de rigidité suivante : toute action continue par isométries affines de G sur un espace de Hilbert possède un point fixe. Elle permet en un sens d'obtenir un contrôle sur la théorie des représentations de G. Nous renvoyons à la récente monographie [**5**] pour une discussion détaillée concernant cette propriété.

THÉORÈME 1.1 ([**17**]). — *Soit G un groupe discret vérifiant la propriété (T). Tout morphisme $\rho : G \to \mathrm{Cr}(2)$ d'image infinie est conjugué par un élément de $\mathrm{Cr}(2)$ à un morphisme à valeurs dans $\mathrm{PGL}(3,\mathbb{C})$.*

Démonstration. — Le groupe G vérifiant la propriété (T), il est de type fini, et toute fonction de type négatif[5] sur G est bornée. Comme la fonction $\phi(\alpha,\beta) = \log\cosh d_{\mathbb{H}}(\alpha,\beta)$ est de type négatif sur $\mathbb{H}(\mathfrak{P})$ (voir [**5**]), elle est bornée. Puisque $\mathbb{H}(\mathfrak{P})$ satisfait à l'inégalité de la médiane, on peut appliquer le lemme du centre [**56**, p. 37] qui implique l'existence d'un point $\alpha \in \mathbb{H}(\mathfrak{P})$ fixé par tous les éléments de $\rho(G)$. On en déduit que $\rho(G)$ est un sous-groupe de $\mathrm{Cr}(2)$ élémentaire de type elliptique. Par le théorème B, soit il est conjugué à un sous-groupe de $\mathrm{PGL}(3,\mathbb{C})$, soit il préserve une fibration rationnelle. Dans le second cas, l'action de G sur la base de la fibration (isomorphe à $\mathbb{P}^1(\mathbb{C})$) induit un morphisme du groupe vers $\mathrm{PGL}(2,\mathbb{C})$. Comme G possède la propriété (T), son image est nécessairement finie. On regarde maintenant l'action induite par G sur une fibre générique. À nouveau l'image dans $\mathrm{PGL}(2,\mathbb{C})$ est finie, donc $\rho(G)$ est fini, ce qui contredit notre hypothèse. □

Historiquement, les réseaux des groupes de Lie vérifiant la propriété (T) ont formé une classe très importante d'exemples. Rappelons qu'un réseau d'un groupe de Lie vérifie la propriété (T) si et seulement si le groupe de Lie la vérifie ; et qu'un groupe de Lie réel connexe et semi-simple ne vérifie pas la propriété (T) si et seulement s'il possède un facteur simple localement isomorphe à $\mathrm{SO}(n,1)$ ou $\mathrm{SU}(n,1)$. Le théorème suivant est une conséquence de [**25**] et [**17**].

---

[4] Pour p premier, le groupe $\langle a,b,c \mid c^p = [a,c] = [b,c] = 1, [a,b] = c\rangle$ ne se plonge pas dans $\mathrm{GL}_n(k)$ avec $p \geq n$ et k un corps de caractéristique distincte de p (Birkhoff). Dans $\mathrm{Cr}(2)$, on peut prendre $a = (\zeta x, y)$, $b = (x, xy)$, et $c = (x, \zeta^{-1}y)$ avec ζ une racine p-ième primitive de 1 (Cerveau-Deserti). Donc $\mathrm{Cr}(2)$ ne plonge dans aucun $\mathrm{GL}_n(k)$.

[5] Une fonction de type négatif est un moyen commode d'encoder une action affine sur un espace de Hilbert.

THÉORÈME 1.2. — *Soit G un groupe de Lie réel connexe simple vérifiant la propriété (T), et Γ un réseau de G muni d'un morphisme ρ dans $\mathrm{Cr}(2)$ d'image infinie. Alors à conjugaison près, l'image de ρ est incluse dans $\mathrm{PGL}(3,\mathbb{C})$. Si de plus $\rho(G)$ n'est pas relativement compacte dans $\mathrm{PGL}(3,\mathbb{C})$, alors on a $G=\mathrm{PGL}(3,\mathbb{R})$ ou $\mathrm{PGL}(3,\mathbb{C})$, et ρ est la restriction à Γ de $u\mapsto u$, $u\mapsto {}^tu^{-1}$, $u\mapsto \bar{u}$ ou $u\mapsto {}^t\bar{u}^{-1}$.*

Notons qu'il existe des réseaux de $\mathrm{SL}(3,\mathbb{R})$ munis de morphismes d'image dense dans $\mathrm{SU}(3)$. Fixons $d\geq 2$ un entier sans facteur carré, posons $K=\mathbb{Q}(\sqrt{d})$ et notons $\mathcal{O}$ son anneau des entiers. Soit L l'extension quadratique de K contenant $d^{1/4}$ et τ l'unique automorphisme de Galois de L sur K non trivial. Notons Γ le sous-groupe des matrices de $\mathrm{SL}(3,\mathcal{O})$ préservant la forme $(x,y)\in L^3\times L^3\mapsto \sum x_i\tau(y_i)\in L$. Le plongement canonique de K dans $\mathbb{R}^2$ induit alors un plongement de Γ dans $\mathrm{SL}(3,\mathbb{R})\times\mathrm{SU}(3)$. D'après le théorème de Borel-Harish-Chandra, l'image de Γ est un réseau dont la projection sur $\mathrm{SL}(3,\mathbb{R})$ est un réseau, et celle sur $\mathrm{SU}(3)$ est dense.

Démonstration. — Le réseau Γ hérite de la propriété (T) de G, donc par le théorème précédent, on peut supposer que ρ est à valeurs dans $\mathrm{PGL}(3,\mathbb{C})$. On applique alors les théorèmes de super-rigidité et d'arithméticité de Corlette-Gromov-Schoen en rang 1 et de Margulis en rang supérieur.

Par hypothèse $\rho(G)$ n'est pas relativement compact, donc ρ s'étend en un morphisme algébrique de G vers $\mathrm{PGL}(3,\mathbb{C})$ qui est injectif par simplicité de G. Les seuls sous-groupes de Lie de $\mathrm{PGL}(3,\mathbb{C})$ simples, non compacts et vérifiant la propriété (T) sont $\mathrm{PGL}(3,\mathbb{R})$ et $\mathrm{PGL}(3,\mathbb{C})$, comme on le voit en complexifiant G. On conclut en utilisant le fait classique que le groupe des automorphismes de $\mathrm{PGL}(3,\mathbb{C})$ est engendré par les automorphismes de corps de $\mathbb{C}$, les automorphismes intérieurs et la contragrédiente, voir **[30]**. □

Rappelons que $\mathrm{SL}(n,\mathbb{Z})$ possède la propriété (T) si et seulement si $n\geq 3$.

THÉORÈME 1.3 (**[25]**). — *Soit ρ un morphisme de $\mathrm{SL}(n,\mathbb{Z})$ dans $\mathrm{Cr}(2)$ d'image infinie. Alors $n\leq 3$. De plus, dans le cas $n=3$, le morphisme ρ est nécessairement injectif et est conjugué au plongement standard $\mathrm{SL}(3,\mathbb{Z})\subset\mathrm{PGL}(3,\mathbb{C})$ ou à la contragrédiente.*

Démonstration. — Au vu du théorème précédent, il faut démontrer que l'image de $\mathrm{SL}(n,\mathbb{Z})$ dans $\mathrm{PGL}(3,\mathbb{C})$ ne peut être relativement compacte. En utilisant les relations entre les générateurs standard de $\mathrm{SL}(n,\mathbb{Z})$, on démontre que ρ coïncide avec un morphisme de $\mathrm{SL}(n,\mathbb{C})$ dans $\mathrm{PGL}(3,\mathbb{C})$ sur un sous-groupe d'indice fini Γ dans $\mathrm{SL}(n,\mathbb{Z})$, voir **[74**, Theorem 6**]**. Cela implique $n=3$. En utilisant plus en détail les résultats de Steinberg, on montre que l'on peut prendre $\Gamma=\mathrm{SL}(3,\mathbb{Z})$.

On peut aussi donner une démonstration plus directe du fait que tout morphisme de $\mathrm{SL}(n,\mathbb{Z})$ dans $\mathrm{Cr}(2)$ est (modulo conjugaison) à valeurs dans $\mathrm{PGL}(3,\mathbb{C})$, en analysant l'image des sous-groupes de Heisenberg de $\mathrm{SL}(n,\mathbb{Z})$, voir **[25]**. □

Notons qu'il existe un corpus de conjectures dues à R. Zimmer qui concernent les morphismes des réseaux des groupes de Lie dans les groupes de difféomorphismes de variétés, voir **[45]** pour une discussion récente de ce type de questions. Les deux résultats précédents entrent naturellement dans le cadre de ces conjectures.

1.2. Propriétés structurelles de $\mathrm{Cr}(2)$

Déduisons maintenant des théorèmes A et B quelques résultats généraux sur $\mathrm{Cr}(2)$.

THÉORÈME 1.4 (**[27, 28]**). — *Tout endomorphisme de* $\mathrm{Cr}(2)$ *injectif est un automorphisme. De plus, le groupe des automorphismes de* $\mathrm{Cr}(2)$ *est engendré par les automorphismes de corps de* $\mathbb{C}$ *et les automorphismes intérieurs.*

Notons que l'approche initiale de **[27]** reposait sur l'étude des sous-groupes de $\mathrm{Cr}(2)$ abéliens non-dénombrables et maximaux pour l'inclusion, et que le cas du groupe $\mathrm{Aut}[\mathbb{C}^2]$ est traité dans **[26]**. Ici, nous suivons **[25]**.

Démonstration. — Soit φ un morphisme injectif de $\mathrm{Cr}(2)$. Par le théorème précédent, on peut supposer que φ est l'identité ou la contragrédiente sur $\mathrm{SL}(3,\mathbb{Z})$. En regardant les relations de commutations entre les applications affines du type $(x+ay+b, y+c)$ avec $a,b,c \in \mathbb{C}$, on montre que φ induit un morphisme algébrique de $\mathrm{PGL}(3,\mathbb{C})$ dans lui-même. Il n'est pas difficile de voir que $\varphi(1/x,1/y) = (a/x, b/y)$ pour un couple $a,b \in \mathbb{C}^*$ adéquat. Pour conclure, on utilise une des relations exhibées par M. Gizatullin pour sa description de $\mathrm{Cr}(2)$ par générateurs et relations. L'identité $(h\sigma)^3 = \mathrm{id}$ avec $\sigma[x:y:z] = [1/x:1/y:1/z]$ et $h[x:y:z] = [x:x-y:x-z]$ permet d'exclure la possibilité à φ d'être la contragrédiente sur $\mathrm{SL}(3,\mathbb{Z})$. □

THÉORÈME 1.5 (Alternative de Tits **[17]**). — *Soit* G *un groupe de type fini de* $\mathrm{Cr}(2)$. *Alors, soit* G *possède un sous-groupe libre non abélien, soit il contient un sous-groupe résoluble d'indice fini.*

L'alternative de Tits a été démontrée pour $\mathrm{Aut}[\mathbb{C}^2]$ par S. Lamy **[62]**.

Démonstration. — On vérifie que si G_1 et G_2 sont deux groupes vérifiant l'alternative de Tits, et s'il existe une suite exacte de groupes $1 \to G_1 \to G \to G_2 \to 1$, alors G vérifie aussi l'alternative. On applique maintenant les théorèmes A et B. Si G n'est pas élémentaire, il contient un sous-groupe libre. Si il est élémentaire elliptique ou parabolique, l'alternative résulte de la remarque précédente et de l'alternative pour les groupes $\mathrm{PGL}(3,\mathbb{C})$, $\mathrm{PGL}(2,\mathbb{C})$, $\mathrm{PGL}(2,\mathbb{C}(x))$ et pour le groupe des automorphismes d'une courbe elliptique sur $\mathbb{C}(x)$, voir **[75]**. Enfin si G est élémentaire hyperbolique,

c'est une extension de $\mathbb{Z}$ par un groupe G_0 ne contenant que des éléments elliptiques. Pour contourner le fait que G_0 n'est pas nécessairement de type fini, on remarque que le groupe dérivé $[G,G]$ est un sous-groupe de G_0 de type fini : il est donc élémentaire elliptique. On conclut en utilisant la suite $1 \to [G,G] \to G \to \mathrm{Ab}(G) \to 1$ où $\mathrm{Ab}(G)$ est l'abélianisé de G. □

Mentionnons enfin quelques résultats en vrac. Les deux premiers sont dus à S. Cantat [**17**], et le dernier à J. Deserti [**29**].

THÉORÈME 1.6 (Propriété de Burnside). — *Tout sous-groupe G de type fini de torsion de* $\mathrm{Cr}(2)$ *est fini.*

THÉORÈME 1.7. — *Soit G un sous-groupe de* $\mathrm{Cr}(2)$ *résoluble de type fini et sans torsion. Alors soit G est abélien, soit il préserve un feuilletage.*

THÉORÈME 1.8. — *Soit G un sous-groupe de* $\mathrm{Cr}(2)$ *nilpotent. Alors soit tous ses éléments sont de torsion, soit le groupe contient un sous-groupe d'indice fini dont le groupe dérivé est abélien.*

1.3. Exemples

Commençons par quelques constructions classiques.

Groupes de Nagata et de Coble. — Partons de deux cubiques lisses de $\mathbb{P}^2(\mathbb{C})$ s'intersectant en 9 points distincts $p_0,\dots,p_8$. La surface X obtenue par éclatement de ces 9 points est munie d'une fibration elliptique $\pi : X \to \mathbb{P}^1(\mathbb{C})$ possédant 9 sections $E_0,\dots,E_8$ (les diviseurs exceptionnels créés par les 9 éclatements). Munissons chaque fibre régulière F de π de la loi de groupe d'origine $E_0 \cap F$. Pour chaque $1 \leq i \leq 8$, on peut considérer l'application $\varphi_i : X \to X$ induit par l'addition $x \mapsto x + (F \cap E_i)$ dans la fibre F. Les applications φ_i sont des automorphismes de X, et on montre que, pour un choix générique des cubiques de départ, le groupe engendré par les φ_i est abélien libre de rang 8.

Plus généralement, prenons 9 points p_i dans le plan, et considérons la surface $X(p_i)$ obtenue en éclatant ceux-ci. Un entier $r \geq 1$ étant donné, on dit qu'une configuration de 9 points p_i est de type Halphen d'indice r, si le système linéaire $|-rK_{X(p_i)}|$ est un pinceau de courbes elliptiques. On montre que, pour un choix générique de 9 points de type Halphen d'indice r, le groupe d'automorphismes $\mathrm{Aut}(X(p_i))$ est isomorphe à $\mathbb{Z}^8$ si $r \geq 3$, et à $\mathbb{Z}/2 \rtimes \mathbb{Z}^8$ si $r = 1,2$, voir [**34**].

On dit qu'une configuration de 10 points p_i dans le plan est de type Coble si $|-2K_{X(p_i)}|$ possède une section. Dans ce cas, ces points se trouvent sur une courbe rationnelle de degré 6 de $\mathbb{P}^2(\mathbb{C})$ avec 10 points doubles (un pour chaque p_i). La dimension de l'espace de telles configurations modulo $\mathrm{PGL}(3,\mathbb{C})$ est égale à 9. Le groupe d'automorphismes $\mathrm{Aut}(X(p_i))$ agit naturellement par isométries sur l'orthogonal de la classe

canonique dans $\mathrm{NS}(X(p_i))$. On obtient donc un morphisme $\mathrm{Aut}(X(p_i)) \to \mathrm{O}(9,1)$. Pour une configuration générique, l'image de ce morphisme est un réseau de $\mathrm{O}(9,1)$, voir **[34]**.

Groupes de Wright. — Par le théorème de Jung, le groupe $\mathrm{Aut}[\mathbb{C}^2]$ est le produit amalgamé du groupe A des applications affines de $\mathbb{C}^2$, et du groupe E des automorphismes préservant la fibration rationnelle $\{x = \mathrm{cte}\}$. Par la théorie de Bass-Serre, il agit donc sur un arbre simplicial. On montre **[47]** qu'un élément de $\mathrm{Aut}[\mathbb{C}^2]$ peut être de type elliptique et dans ce cas il est conjugué à un élément de A ou de E ; ou être de type hyperbolique et, dans ce cas, il est conjugué à une application dite « de Hénon » c'est-à-dire à une composée d'applications de la forme $(x,y) \mapsto (y, ax + P(x))$ avec $a \in \mathbb{C}^*$ et $\deg(P) \geq 2$. S. Lamy **[62]** a donné la classification suivante des sous-groupes de $\mathrm{Aut}[\mathbb{C}^2]$ et ce, sans hypothèse de type fini.

THÉORÈME 1.9. — *Soit G un sous-groupe de $\mathrm{Aut}[\mathbb{C}^2]$. À conjugaison par un automorphisme polynomial près, on est dans l'un des quatre cas suivants* :

1. *G possède deux éléments hyperboliques engendrant un groupe libre non-abélien* ;
2. *G est un sous-groupe de A ou de E* ;
3. *G est de type fini élémentaire hyperbolique* ;
4. *tous les éléments de G sont elliptiques et on n'est pas dans le cas* (2).

Un groupe relevant du dernier cas est appelé groupe de Wright : les premiers exemples ont été donnés dans **[80]**. Un tel groupe G n'est jamais de type fini, mais est abélien, et est l'union croissante de groupes H_i cycliques et finis. Une description plus détaillée est donnée dans **[62**, Proposition 3.12**]**. Pour un exemple concret de groupe de Wright, on procède comme suit. Pour tout $k \geq 1$, notons $g_k(x,y) = (y, c_k y^{2^k+1} + x)$, avec $c_k \in \mathbb{C}^*$ arbitraire ; $\varphi_k = g_1^2 \circ g_2^2 \cdots \circ g_k^2$; $H_0 = \langle (-x,-y) \rangle$ et $H_k = \varphi_k \langle (\zeta_k x, \zeta_k y) \rangle \varphi_k^{-1}$ avec $\zeta_k = \exp(2i\pi/2^{k+1})$. Alors $H = \cup_{k \geq 0} H_k$ est un groupe de Wright.

Automorphismes de surfaces affines. — Le théorème 1.9 et plus généralement l'étude des propriétés algébriques de $\mathrm{Aut}[\mathbb{C}^2]$ par S. Lamy ont été une source importante d'inspiration pour l'étude des sous-groupes de type fini du groupe de Cremona. Notons que d'autres surfaces affines ont un groupe d'automorphismes dont la structure est très proche de celle de $\mathrm{Aut}[\mathbb{C}^2]$. C'est le cas pour les surfaces obtenues comme complémentaire d'une chaîne de courbes rationnelles portant un diviseur ample dans une surface projective **[52, 53]** ; ou des surfaces affines de $\mathbb{C}^3$ d'équations $xz = P(y)$ avec P de degré au moins 2 **[65]**. Le groupe d'automorphismes de toutes ces surfaces affines agit par isométries sur un arbre réel : le théorème 1.9 possède donc probablement un analogue dans toutes ces situations. Nous renvoyons à **[36]** pour une description d'une famille de générateurs de ces groupes.

Le cas des surfaces cubiques est particulièrement intéressant. Soit S une cubique lisse de $\mathbb{C}^3$ dont la complétion $\bar{S}$ dans $\mathbb{P}^3(\mathbb{C})$ est aussi lisse. On suppose de plus que le plan à l'infini P_∞ est un plan tri-tangent, c'est-à-dire que l'intersection $\bar{S} \cap P_\infty$ est l'union de trois droites L_x, L_y, L_z s'intersectant en trois points distincts $p_x = L_y \cap L_z$, $p_y = L_z \cap L_x$ et $p_z = L_x \cap L_y$. Prenons un point $p \in S$, et considérons la droite joignant p à p_x. Elle coupe S en un troisième point que l'on note $\sigma_x(p)$. On obtient ainsi une involution sur S. On construit de même deux autres involutions σ_y et σ_z. Le groupe $G = \langle \sigma_x, \sigma_y, \sigma_z \rangle$ est alors un sous-groupe du groupe des automorphismes de la surface affine S qui est isomorphe au produit libre $\mathbb{Z}/2 \star \mathbb{Z}/2 \star \mathbb{Z}/2$. En regardant l'action de G sur le cycle déterminé par les trois droites à l'infini, on montre de plus la trichotomie suivante pour l'action d'un élément $g \in G$: soit g est conjugué à σ_x et est de type elliptique ; soit il est conjugué à $\sigma_x \sigma_y$, de type parabolique et préserve une fibration rationnelle ; soit il est hyperbolique. En particulier, G est un groupe de type fini de $\mathrm{Cr}(2)$ non élémentaire.

Notons que le groupe G apparaît dans plusieurs contextes de nature très différente : dans l'étude de l'espace des représentations du groupe libre à trois générateurs à valeurs dans $\mathrm{SL}(2,\mathbb{C})$, voir **[8]** pour une description élémentaire ; dans l'étude des équations de Painlevé VI, voir **[37, 59]** ; et enfin dans la théorie des opérateurs de Schrödinger discrets, voir par exemple **[20]** et les références qui s'y trouvent.

Le lien entre surfaces cubiques et espace de représentations permet d'obtenir des morphismes du groupe modulaire de Teichmüller de la sphère privé de quatre points dans $\mathrm{Cr}(2)$ qui préservent le type[(6)]. De même, le lien avec les équations de Painlevé fournit des morphismes de $\pi_1(\mathbb{P}^1 \setminus \{0,1,\infty\})$ dans $\mathrm{Cr}(2)$ qui préservent le type.

Prenons le cas de la cubique singulière de Cayley S_{cay}, qui est obtenue comme quotient de $\mathbb{P}^1(\mathbb{C}) \times \mathbb{P}^1(\mathbb{C})$ par l'involution $(x,y) \mapsto (x^{-1}, y^{-1})$. L'action d'une application birationnelle mononomiale $(x,y) \mapsto (x^a y^b, x^c y^d)$ avec $a,b,c,d \in \mathbb{Z}$ vérifiant $ad - bc = +1$ descend sur S_{cay}. On obtient ainsi une action birationnelle de $\mathrm{SL}(2,\mathbb{Z})$ sur S_{cay}, et donc un morphisme de $\mathrm{SL}(2,\mathbb{Z})$ dans $\mathrm{Cr}(2)$ qui préserve le type. Une équation de S_{cay} dans $\mathbb{C}^3$ est $x^2 + y^2 + z^2 + xyz = 4$. On montre que l'action de $\mathrm{SL}(2,\mathbb{Z})$ se déforme en une action préservant le type sur les surfaces $\{x^2 + y^2 + z^2 + xyz = a\}$ avec $a \in \mathbb{C}$, voir **[54]**.

Nous renvoyons à **[16, 19]** pour plus de détails sur toutes ces constructions.

Groupe symplectique et groupe de Thompson. — Cet exemple est dû à A. Usnich **[78]**. On fixe des coordonnées homogènes $[x : y : z]$ dans $\mathbb{P}^2(\mathbb{C})$, et on note ω la 2-forme méromorphe donnée par $dx \wedge dy/xy$ dans la carte $[x : y : 1]$. Soit Symp le sous-groupe de $\mathrm{Cr}(2)$ préservant ω, c'est-à-dire le groupe des applications birationnelles f

(6) Le type d'un élément du groupe modulaire de Teichmüller est donné par le type de son action sur l'espace de Teichmüller.

telles que $f^*\omega = c\omega$ pour une constante $c \in \mathbb{C}^*$. Ce groupe contient $\mathrm{SL}(2,\mathbb{Z})$ identifié aux applications monomiales $(x,y) \mapsto (x^a y^b, x^c y^d)$ avec $ad - bc = +1$; le groupe $(x,y) \mapsto (\zeta x, \xi y)$ avec $(\zeta,\xi) \in \mathbb{C}^* \times \mathbb{C}^*$; ainsi que l'application $\varphi = (y, \frac{1+y}{x})$. Notons $\Delta = [x=0] + [y=0] + [z=0]$ le diviseur de $-\omega$, et Jf le diviseur déterminé par l'annulation du déterminant jacobien de f. Pour tout $f \in \mathrm{Symp}$, on a alors

$$f^*\Delta = \Delta + Jf. \tag{1}$$

Soit $M_{\mathbb{R}}$ l'espace des valuations monomiales sur $\mathbb{C}[x,y]$, c'est-à-dire des fonctions $\nu_{s,t} : \mathbb{C}[x,y] \to \mathbb{R}$ de la forme $\nu_{s,t}(\sum_{i,j} a_{ij} x^i y^j) = \min\{is+jt, a_{ij} \neq 0\}$ pour un couple $(s,t) \in \mathbb{R}^2$. Muni de la topologie de la convergence simple, cet espace s'identifie à $\mathbb{R}^2$ et contient naturellement le réseau $M_{\mathbb{Z}} \subset M_{\mathbb{R}}$ des valuations ne prenant que des valeurs entières.

Toute application birationnelle de $\mathbb{P}^2(\mathbb{C})$ agit sur l'espace des valuations de $\mathbb{C}[x,y]$. Lorsque $f \in \mathrm{Symp}$, on montre en utilisant (1) que cette action préserve $M_{\mathbb{R}}$ ainsi que le réseau $M_{\mathbb{Z}}$, et que cette action est affine par morceaux. Par exemple $\varphi_*(s,t) = (t, -s + \min\{0,t\})$. On obtient donc un morphisme ρ de Symp dans le groupe T des homéomorphismes de $\mathbb{R}^2$ qui sont affines par morceaux et préservent le réseau $\mathbb{Z}^2$. Ce groupe est appelé groupe de Thompson, voir [**13**]. En utilisant une présentation du groupe T astucieuse, A. Usnich montre que la restriction de ρ au sous-groupe de Symp engendré par $\mathrm{SL}(2,\mathbb{Z})$ et φ est surjective sur T. Notons que T ne peut être réalisé comme sous-groupe de $\mathrm{Cr}(2)$ car il ne vérifie pas l'alternative de Tits.

1.4. Quelques questions sur $\mathrm{Cr}(2)$

Les résultats précédents suggèrent quelques questions naturelles. Nombre d'entre elles sont déjà explicites dans [**17**].

Il est tout d'abord probable que le théorème B puisse être précisé.

QUESTION 1. — *Soit G un groupe élémentaire elliptique préservant une fibration rationnelle. Existe-t-il un modèle X dans lequel tous les éléments de G sont des automorphismes* ?

QUESTION 2. — *Soit G un groupe élémentaire hyperbolique. Le groupe G_0 donné par le théorème B est-il de type fini* ?

Nous avons décrit les morphismes des réseaux des groupes de Lie vérifiant la propriété (T) dans $\mathrm{Cr}(2)$. Il reste donc le cas des réseaux de $\mathrm{SO}(n,1)$ et $\mathrm{SU}(n,1)$. Peu d'exemples de morphismes injectifs de tels réseaux sont connus. Les automorphismes des surfaces de Coble génériques sont des réseaux de $\mathrm{SO}(9,1)$ comme on l'a vu à la section précédente.

QUESTION 3. — *Pour quels entiers n existe-t-il un réseau de* $\mathrm{SO}(n,1)$ *ou* $\mathrm{SU}(n,1)$ *s'injectant dans* $\mathrm{Cr}(2)$?

Il existe de nombreux morphismes du groupe libre vers le groupe de Cremona. Pour les groupes proches du groupe libre, il nous faut donc imposer des conditions supplémentaires sur les morphismes pour avoir une chance de les classer.

Prenons le cas du groupe fondamental $\pi_1(S)$ d'une surface réelle hyperbolique S de type fini éventuellement avec des points orbifolds. On a alors une notion naturelle de type d'un élément $\gamma \in \pi_1(S)$ suivant le type (hyperbolique, parabolique ou elliptique) de son action sur le disque de Poincaré.

QUESTION 4. — *Classer les morphismes injectifs* $\rho : \pi_1(S) \to \mathrm{Cr}(2)$ *préservant le type.*

Notons que des exemples de morphismes des groupes fondamentaux de surfaces compactes dans le groupe des automorphismes polynomiaux de $\mathbb{C}^2$ dont l'image contient au moins un élément hyperbolique ont été construits par S. Cantat et S. Lamy [**18**]. Le cas de $\mathrm{SL}(2,\mathbb{Z})$ paraît particulièrement intéressant au vu du théorème 1.3. Enfin on peut poser la même question pour les réseaux de $\mathrm{SO}(n,1)$ et $\mathrm{SU}(n,1)$ en adaptant la notion de type. Notons que le cas des représentations fidèles discrètes et préservant le type de $\mathrm{PSL}(2,\mathbb{Z})$ dans $\mathrm{PU}(2,1)$ est traité dans [**40**].

Le groupe de Cremona est à bien des égards proche des groupes linéaires. On peut donc se poser la question de savoir si les propriétés de Malcev et Selberg sont encore valables dans $\mathrm{Cr}(2)$.

QUESTION 5. — *Soit G un sous-groupe de type fini de* $\mathrm{Cr}(2)$. *Est-ce que G contient un sous-groupe d'indice fini et sans torsion ? Est-ce que G est résiduellement fini, c'est-à-dire est-ce que, pour tout $g \in G$, il existe un morphisme $\rho : G \to H$ avec H fini et $\rho(g) \neq 1$?*

Nous avons vu que les théorèmes de structure des groupes de type fini permettent de déduire de nombreuses propriétés du groupe de Cremona. On ne comprend cependant pas en détail la structure des sous-groupes abéliens.

QUESTION 6. — *Classer les sous-groupes abéliens de* $\mathrm{Cr}(2)$ *maximaux pour l'inclusion.*

Le cas non dénombrable a été étudié par J. Deserti [**27**] : un tel groupe préserve un feuilletage ou bien tous ses éléments sont de torsion. Le cas dénombrable semble plus délicat : un exemple d'un tel groupe isomorphe à $\mathbb{Z}^8$ est donné à la section 1.3. La classification des sous-groupes abéliens maximaux qui contiennent un élément hyperbolique est aussi ouverte.

QUESTION 7. — *Le centralisateur d'un élément hyperbolique de* $\mathrm{Cr}(2)$ *est-il une extension de* $\mathbb{Z}$ *par un groupe fini* ?

La réponse à cette question est positive pour les automorphismes du plan par S. Lamy **[62]**, ou sous certaines hypothèses de nature dynamique, voir **[17**, §5**]**.

Notons que, comme l'alternative de Tits est vérifiée pour $\mathrm{Cr}(2)$, tout sous-groupe de type fini moyennable du groupe de Cremona est résoluble. De plus tout sous-groupe résoluble de type fini de $\mathrm{Cr}(2)$ est élémentaire.

QUESTION 8. — *Tout sous-groupe résoluble de type fini de* $\mathrm{Cr}(2)$ *possède un sous-groupe d'indice fini dont le groupe dérivé est nilpotent.*

Inspiré par les résultats récents d'E. Breuillard sur l'alternative de Tits **[11]**, on peut enfin poser la question suivante.

QUESTION 9. — *L'alternative de Tits uniforme est-elle valide dans* $\mathrm{Cr}(2)$? *En d'autres termes, existe-t-il une constante* $C > 0$ *telle que, pour toute famille finie* $f_1, \dots, f_n \in \mathrm{Cr}(2)$ *engendrant un groupe non résoluble, on puisse trouver deux mots en les* f_i *de longueur au plus* C *et engendrant un groupe libre non abélien ?*

2. ACTION D'UNE APPLICATION RATIONNELLE DE SURFACE EN COHOMOLOGIE

Une application rationnelle $f : X \dashrightarrow X$ définie sur une variété projective induit une action linéaire sur la cohomologie de Dolbeault $\mathrm{H}^{p,q}(X)$ même lorsqu'elle admet des points d'indétermination. C'est un fait maintenant bien établi que cette action en cohomologie détermine de nombreux aspects de la dynamique de f. Par un théorème de M. Gromov **[55]**, et de T.-C. Dinh et N. Sibony **[32, 33]**, elle permet d'obtenir une borne supérieure sur l'entropie topologique. A. Russakovskii et B. Shiffman **[69]** ont aussi montré son importance pour les problèmes de distribution des sous-variétés itérées. C'est à ces auteurs que l'on doit la notion importante de degré dynamique que l'on rappellera ci-dessous.

On ne sait cependant exploiter l'information contenue dans l'action en cohomologie que sous certaines conditions portant sur l'ensemble d'indétermination de f. On est donc tenté de trouver un autre modèle birationnel dans lequel ces conditions sont satisfaites. Ce problème s'avère cependant très ardu[(7)], et l'idée de travailler simultanément avec tous les modèles birationnels est donc apparue naturellement.

[(7)] Seul le cas des surfaces a été traité systématiquement, et il n'est pas complètement résolu à ce jour, voir cependant **[31, 43, 44]**.

Nous allons tout d'abord expliquer comment parler de l'espace des classes de cohomologie de tous les modèles birationnels du plan projectif et en étudier la structure. En particulier, on y construira l'espace hyperbolique $\mathbb{H}(\mathfrak{P})$ mentionné dans l'introduction sur lequel $\mathrm{Cr}(2)$ agit par isométries. Ensuite nous utiliserons cet outil pour décrire en détail l'action d'une application birationnelle sur la cohomologie (théorème 2.7).

2.1. Classes dans la variété de Riemann-Zariski

Un modèle (lisse) de $\mathbb{P}^2$ est une surface projective lisse X munie d'une application birationnelle $\pi : X \dashrightarrow \mathbb{P}^2$. Deux modèles X, X' sont équivalents si l'application birationnelle induite $X \dashrightarrow X'$ est un isomorphisme. Un modèle X domine un autre X' si cette application est régulière. Dans ce cas, X est obtenue à partir de X' en éclatant un nombre fini de points (éventuellement infiniment proches). L'ensemble $\mathfrak{B}$ des classes d'isomorphismes de modèles forme un ensemble inductif : pour tout couple de modèles X, X' il existe un modèle X'' les dominant tous les deux.

Il n'existe pas de variété projective dominant tous les modèles de $\mathbb{P}^2$. Cependant on peut considérer la limite projective de tous les modèles. Muni de la topologie produit, on obtient un espace topologique[8] $\mathfrak{P}$ appelé variété de Riemann-Zariski. L'ensemble de tous les points de tous les modèles de $\mathbb{P}^2$ est naturellement inclus[9] dans $\mathfrak{P}$. Cet espace permet donc de s'abstenir de travailler avec un modèle en particulier : c'est pour cela qu'il a été introduit par Zariski [**81**] dans l'optique de démontrer des théorèmes de désingularisation. On ne travaillera pas directement sur cet espace, mais il est bon de garder à l'esprit que les espaces que l'on va maintenant introduire sont en quelque sorte les espaces de (co)homologie de $\mathfrak{P}$.

Soit X un modèle de $\mathbb{P}^2$. Considérons l'éclatement de X en un point $p \in X$: $\mu : X' \to X$, et notons E le diviseur exceptionnel, d'auto-intersection -1. Le morphisme naturel $\mu^* : \mathrm{H}^2(X, \mathbb{Z}) \to \mathrm{H}^2(X', \mathbb{Z})$ est alors injectif, et il n'est pas difficile de montrer que $\mathrm{H}^2(X', \mathbb{Z}) = \mathrm{H}^2(X, \mathbb{Z}) \oplus \mathbb{Z}[E]$. Cette décomposition est de plus orthogonale pour la forme d'intersection induite par le cup produit.

Comme $\mathrm{H}^2(\mathbb{P}^2, \mathbb{Z})$ est engendré par la classe L d'une droite pour laquelle $L \cdot L = +1$, pour tout modèle de $\mathbb{P}^2$ la forme d'intersection sur $\mathrm{H}^2(X, \mathbb{Z})$ est de type Minkowski, et possède une base orthogonale $\alpha_1, \dots, \alpha_k$ avec $\alpha_1^2 = +1$ et $\alpha_i^2 = -1$ pour $k \geq i \geq 2$.

Si X' domine X et $\mu : X' \to X$ est le morphisme induit, les applications linéaires naturelles $\mu^* : \mathrm{H}^2(X, \mathbb{R}) \to \mathrm{H}^2(X', \mathbb{R})$ et $\mu_* : \mathrm{H}^2(X', \mathbb{R}) \to \mathrm{H}^2(X, \mathbb{R})$ sont respectivement injective et surjective, et on a $\mu_* \mu^* = \mathrm{id}$ sur $\mathrm{H}^2(X, \mathbb{R})$.

---

[8] Celui-ci peut être muni d'une structure naturelle d'espace annelé.

[9] L'inclusion est stricte : l'ensemble de tous les points infiniment voisins, aussi appelé « bubble space » par Y.I. Manin [**66**, §35], a été utilisé par les géomètres italiens pour développer la théorie de l'intersection dans les surfaces, voir [**22**].

DÉFINITION 2.1. — *Une classe de Weil est un élément α de la limite projective $\varprojlim_{\mathfrak{B}} \mathrm{H}^2(X,\mathbb{R})$ selon les morphismes μ_*. En d'autres termes, c'est la donnée pour tout modèle X d'une classe $\alpha_X \in \mathrm{H}^2(X,\mathbb{R})$ telle que $\mu_*\alpha_{X'} = \alpha_X$ dès que X' domine X et $\mu : X' \to X$ est le morphisme induit.*

Une classe de Cartier est un élément α de la limite injective $\varinjlim_{\mathfrak{B}} \mathrm{H}^2(X,\mathbb{R})$ selon les morphismes μ^*. C'est-à-dire que α est une classe de Weil au sens précédent pour laquelle on peut trouver un modèle X_0 tel que $\alpha_X = \mu^*\alpha_{X_0}$ pour tout modèle X dominant X_0, avec $\mu : X \to X_0$ le morphisme induit.

Si α est une classe de Cartier, et X_0 est un modèle vérifiant la condition ci-dessus, on dira que α est déterminée dans X_0 ou que X_0 détermine α. Si α est une classe de Weil et X est un modèle arbitraire, la classe α_X est appelée incarnation de α dans X. Fixons un modèle X, et une classe $\alpha_0 \in \mathrm{H}^2(X,\mathbb{R})$. On peut alors lui associer la classe de Cartier α déterminée dans X par α_0. Ainsi on obtient un morphisme injectif de $\mathrm{H}^2(X,\mathbb{R})$ dans l'espace des classes de Cartier. L'ensemble des classes de Cartier est donc l'union des $\mathrm{H}^2(X,\mathbb{R})$ pour X parcourant l'ensemble des modèles de $\mathbb{P}^2$.

On notera $\mathrm{H}^2(\mathfrak{P})$ l'ensemble de toutes les classes de Cartier. De même, on notera $\mathrm{H}_2(\mathfrak{P})$ l'ensemble de classes de Weil. Ce sont des espaces vectoriels de dimension infinie, et on a $\mathrm{H}^2(\mathfrak{P}) \subset \mathrm{H}_2(\mathfrak{P})$.

Ces espaces ont été introduits pour la première fois par Y.I. Manin [**66**]. Ils sont ensuite apparus dans les travaux de V.V Shokurov [**72**] liés à la géométrie birationnelle des variétés projectives de dimension quelconque. Leur apparition en dynamique holomorphe est récente et est due simultanément à S. Cantat [**17**] et à S. Boucksom, M. Jonsson et moi-même [**10**].

2.2. Intersection

Nous suivons essentiellement [**10**] auquel nous renvoyons pour des preuves détaillées.

Soient α une classe de Weil, et β une classe de Cartier déterminée dans un modèle X. La quantité $\alpha_X \cdot \beta_X$ ne dépend alors pas du choix du modèle, et on la note $\alpha \cdot \beta \in \mathbb{R}$. On a donc une forme bilinéaire $\mathrm{H}_2(\mathfrak{P}) \times \mathrm{H}^2(\mathfrak{P}) \to \mathbb{R}$ qui est symétrique sur $\mathrm{H}^2(\mathfrak{P})$.

Pour bien comprendre la structure de cette forme, nous allons construire des bases de $\mathrm{H}_2(\mathfrak{P})$ et $\mathrm{H}^2(\mathfrak{P})$ respectivement. Introduisons l'ensemble $\mathcal{V}_X$ constitué des classes d'équivalence de couples (p,X) avec $p \in X$ et X un modèle dominant $\mathbb{P}^2$ pour la relation : $(p,X) \sim (p',X')$ si et seulement si le morphisme naturel $\mu : X \dashrightarrow X'$ est un isomorphisme au voisinage de p et qu'il envoie ce point sur p'. L'ensemble $\mathcal{V}_X$ est en bijection avec l'ensemble des anneaux de valuations discrètes du corps des fonctions

de $\mathbb{P}^2(\mathbb{C})$ dont le corps résiduel est une extension non triviale de $\mathbb{C}$ et dont le centre dans $\mathbb{P}^2$ est un point.

À tout élément $\nu = (p, X) \in \mathscr{V}_X$ est associée une classe de Cartier. Si $\mu : X' \to X$ est l'éclatement de X en p, on définit $\mathscr{E}(\nu) = \mathscr{E}(p, X)$ comme la classe de Cartier déterminée dans X' par la classe fondamentale du diviseur exceptionnel de μ. Notons de plus $\mathscr{L}$ la classe de Cartier déterminée par la classe d'une droite dans $\mathbb{P}^2$.

On a $\mathscr{L} \cdot \mathscr{L} = +1$, $\mathscr{E}(\nu) \cdot \mathscr{E}(\nu) = -1$, et $\mathscr{E}(\nu) \cdot \mathscr{E}(\nu') = \mathscr{E}(\nu) \cdot \mathscr{L} = 0$ pour tout $\nu \neq \nu' \in \mathscr{V}_X$.

THÉORÈME 2.2. — *L'application linéaire* $\theta : \mathrm{H}_2(\mathfrak{P}) \to \mathbb{R} \oplus \mathbb{R}^{\mathscr{V}_X}$ *définie par* $\theta(\alpha) = \alpha \cdot \mathscr{L} \oplus \{\nu \mapsto \alpha \cdot \mathscr{E}(\nu)\}$ *est un isomorphisme.*

Pour alléger les notations, pour toute collection a, $a_\nu \in \mathbb{R}$, on notera $a\mathscr{L} + \sum_{\nu \in \mathscr{V}_X} a_\nu \mathscr{E}(\nu)$ pour $(a, \{\nu \mapsto a_\nu\}) \in \mathbb{R} \oplus \mathbb{R}^{\mathscr{V}_X}$ et on l'identifiera à la classe de Weil correspondante.

Une classe $\alpha = a\mathscr{L} + \sum_{\nu \in \mathscr{V}_X} a_\nu \mathscr{E}(\nu)$ est de Cartier si et seulement si l'ensemble $\{\nu \in \mathscr{V}_X, a_\nu \neq 0\}$ est fini. Dans ce cas, on a $\alpha \cdot \alpha = a^2 - \sum_{\mathscr{V}_X} a_\nu^2$. On est donc amené à la définition suivante.

DÉFINITION 2.3. — *Une classe de Weil* $\alpha = a\mathscr{L} + \sum_{\nu \in \mathscr{V}_X} a_\nu \mathscr{E}(\nu)$ *est dite* L^2 *si* $\sum a_\nu^2 < +\infty$.

On peut donner une caractérisation intrinsèque des classes L^2. Si α est une classe de Weil, alors α est L^2 si et seulement si $\inf_{\mathfrak{B}} \{\alpha_X \cdot \alpha_X\} > -\infty$.

On note $\mathrm{L}^2(\mathfrak{P})$ l'espace des classes L^2. Il contient strictement $\mathrm{H}^2(\mathfrak{P})$ et est strictement contenu dans $\mathrm{H}_2(\mathfrak{P})$. On a une forme bilinéaire symétrique naturelle sur $\mathrm{L}^2(\mathfrak{P})$ envoyant $\alpha = a\mathscr{L} + \sum_{\nu \in \mathscr{V}_X} a_\nu \mathscr{E}(\nu)$, $\alpha' = a'\mathscr{L} + \sum_{\nu \in \mathscr{V}_X} a'_\nu \mathscr{E}(\nu)$ sur $\alpha \cdot \alpha' := aa' - \sum a_\nu a'_\nu$. C'est une forme non-dégénérée de type Minkowski. Notons par ailleurs que la forme quadratique $\alpha \mapsto 2(\alpha \cdot \mathscr{L}) - \alpha \cdot \alpha$ est définie positive, et que $\mathrm{L}^2(\mathfrak{P})$ est un espace complet pour la norme associée.

DÉFINITION 2.4. — *L'espace* $\mathbb{H}(\mathfrak{P})$ *est par définition le sous-ensemble des classes* $\alpha \in \mathrm{L}^2(\mathfrak{P})$ *telles que* $\alpha \cdot \alpha = +1$ *et* $\alpha \cdot \mathscr{L} > 0$.

Rappelons que, pour un espace vectoriel (V, q) de dimension finie munie d'une forme de type Minkowski, l'ensemble $\{\alpha \in V, q(\alpha) = +1\}$ est une hypersurface $\mathbb{H}(V)$ possédant deux composantes connexes, et que la restriction de q à cette hypersurface induit une métrique riemannienne pour laquelle $\mathbb{H}(V)$ devient le modèle hyperboloïde de l'espace hyperbolique de dimension $\dim(V) - 1$.

Dans notre situation, $\mathrm{L}^2(\mathfrak{P})$ est de dimension infinie, mais la situation est essentiellement la même. Posons $d_{\mathbb{H}}(\alpha, \beta) := (\cosh)^{-1}(\alpha \cdot \beta)$ avec $\cosh(x) = \frac{1}{2}(e^x + e^{-x})$.

On vérifie que $d_{\mathbb{H}}$ définit une métrique sur $\mathbb{H}(\mathfrak{P})$ pour laquelle cet espace devient hyperbolique au sens de Gromov.

Dans la suite, on notera $\mathscr{C}^*$ le cône convexe de toutes les classes $\alpha \in \mathrm{L}^2(\mathfrak{P})$ telles que $\alpha \cdot \alpha \geq 0$ et $\alpha \cdot \mathscr{L} > 0$. La réunion $\mathscr{C}^* \cup \{0\}$ est fermée pour la topologie faible duale de l'espace de Hilbert $\mathrm{L}^2(\mathfrak{P})$.

2.3. Positivité

Il existe plusieurs notions de positivité pour les classes de $\mathrm{H}^2(X, \mathbb{R})$ d'une variété kählérienne compacte X. Nous renvoyons à [**23**, §6] et à [**64**] pour une discussion détaillée.

Fixons X un modèle de $\mathbb{P}^2$ et α une classe dans $\mathrm{H}^2(X, \mathbb{R})$. À tout $\mathbb{R}$-diviseur $Z = \sum a_k C_k$ avec $a_k \in \mathbb{R}$ et C_k des courbes irréductibles de X est associée une classe notée $[Z]$ égale à $\sum a_k c_1(\mathscr{O}_X(C_k)) \in \mathrm{H}^2(X, \mathbb{R})$. La classe $c_1(\mathscr{O}_X(C_k))$ coïncide avec la classe fondamentale de C_k plongée dans X, et appartient au réseau $\mathrm{H}^2(X, \mathbb{Z})$. On dira que α est *effective* si et seulement si elle peut être représentée par un $\mathbb{R}$-diviseur effectif, c'est-à-dire dont tous les coefficients a_k sont positifs. L'ensemble des classes effectives est un cône convexe qui n'est pas fermé en général. Son adhérence est appelée le cône des classes pseudo-effectives et notée $\mathrm{Psef}(X) \subset \mathrm{H}^2(X, \mathbb{R})$. C'est un cône strict au sens où $\mathrm{Psef}(X) \cap -\mathrm{Psef}(X) = \{0\}$. On montre qu'une classe est pseudo-effective si et seulement si elle est représentée par un courant positif fermé de bidegré $(1,1)$.

Par dualité, on définit aussi la notion suivante : α est dite *nef* si et seulement si $\alpha \cdot \beta \geq 0$ pour toute classe β pseudo-effective. L'ensemble des classes nef est un cône convexe fermé strict contenu dans $\mathrm{Psef}(X)$. On le note $\mathrm{Nef}(X)$. On montre que l'intérieur du cône nef est constitué exactement des classes déterminées par les formes de Kähler sur X.

Définition 2.5. — *Une classe $\alpha \in \mathrm{H}_2(\mathfrak{P})$ est pseudo-effective* (resp. *nef*) *si et seulement si pour tout modèle X de $\mathbb{P}^2$ son incarnation α_X est pseudo-effective* (resp. *nef*).

Si α est une classe de Cartier déterminée dans X, il est facile de voir que α est psef (resp. nef) si et seulement si son incarnation dans X l'est. Notons enfin que toute classe nef est L^2 car la famille $\{\alpha_X \cdot \alpha_X\}$ est bornée inférieurement par 0.

2.4. Classification des éléments de $\mathrm{Cr}(2)$

On va maintenant s'intéresser à l'action d'une application birationnelle $f \in \mathrm{Cr}(2)$ sur la cohomologie. Deux ingrédients vont jouer un rôle important : la notion de degré dynamique, et l'action de f sur $\mathrm{L}^2(\mathfrak{P})$. Nous allons voir comment ceux-ci interagissent.

Commençons par le degré dynamique. Son origine est sans doute à chercher parmi les travaux de la fin des années 80 reliant entropie dynamique et croissance des volumes, qui ont culminé avec ceux de Gromov, Newhouse et Yomdin. Cette notion a

ensuite progressivement émergé après la publication du papier d'Arnol'd [**1**], voir par exemple [**7, 57, 79**], et a été finalement formalisée par Russakovskii et Shiffman [**69**].

En coordonnées homogènes $[x : y : z]$, une application birationnelle f de $\mathbb{P}^2$ est définie par trois polynômes homogènes sans facteur commun. Leur degré commun est appelé degré de f et on le note $\deg(f)$. Si L est une droite générique de $\mathbb{P}^2$, sa préimage $f^{-1}L$ est une courbe de degré $\deg(f)$. On vérifie facilement que $\deg(f\circ g)\leq \deg(f)\times\deg(g)$ pour tout couple $f,g\in \mathrm{Cr}(2)$. La suite $\deg(f^n)$ est donc sous-multiplicative, et on peut poser :

$$\lambda(f) := \lim_{n\to\infty} \deg(f^n)^{1/n} \ .$$

On appelle $\lambda(f)$ le degré dynamique de f, voir [**69**]. En utilisant la sous-multiplicativité, il est facile de voir que $\lambda(g^{-1}\circ f\circ g)=\lambda(f)$ pour tout $g\in\mathrm{Cr}(2)$.

La construction de l'action de f sur l'espace $\mathrm{L}^2(\mathfrak{P})$ nécessite plusieurs étapes. Prenons tout d'abord $f : X \to X'$ un morphisme (pas nécessairement birationnel) entre deux surfaces. On a alors deux applications linéaires f^*, f_* entre $\mathrm{H}^2(X,\mathbb{R})$ et $\mathrm{H}^2(X',\mathbb{R})$ préservant les réseaux $\mathrm{H}^2(X,\mathbb{Z})$ et $\mathrm{H}^2(X',\mathbb{Z})$ et qui vérifient $f^*\alpha\cdot\beta=\alpha\cdot f_*\beta$ et $f_*f^*\alpha = e(f)\times\alpha$ avec $e(f)$ le degré topologique de f. Ces constructions sont fonctorielles : pour deux morphismes f et g, on a $(f\circ g)_*=f_*\circ g_*$ et $(f\circ g)^*=g^*\circ f^*$.

Supposons que l'application $f : X \dashrightarrow X'$ soit seulement rationnelle, et prenons Γ une désingularisée du graphe de f. On a deux morphismes, l'un birationnel $\pi_1 : \Gamma \to X$ et l'autre $\pi_2 : \Gamma \to X'$ de degré topologique $e(f)$, tels que $f=\pi_2\circ\pi_1^{-1}$. On définit les applications linéaires $f_\# = \pi_{2*}\circ\pi_1^* : \mathrm{H}^2(X,\mathbb{R})\to\mathrm{H}^2(X',\mathbb{R})$ et $f^\# = \pi_{1*}\circ\pi_2^* : \mathrm{H}^2(X',\mathbb{R})\to\mathrm{H}^2(X,\mathbb{R})$.

Il est important de noter que ces constructions ne sont plus fonctorielles. Ainsi si f est une involution birationnelle de degré au moins 2 dans $\mathbb{P}^2$, les morphismes $f_\#$ et $f^\#$ s'identifient à la multiplication par $\deg(f)$ et on a $1=\deg(f\circ f^{-1})<\deg(f)\times\deg(f^{-1})$. On garde cependant toujours la propriété $f^\#\alpha\cdot\beta=\alpha\cdot f_\#\beta$ pour toutes classes α,β.

Fixons désormais $f\in\mathrm{Cr}(2)$, et soit α une classe de Weil. Pour tout modèle X, prenons un autre modèle X' tel que l'application $\phi : X' \to X$ induite par le relèvement de f soit holomorphe. On définit $f_*\alpha$ comme l'unique classe de Weil dont l'incarnation dans X est déterminée par $\phi_*(\alpha_{X'})$. Si α est une classe de Cartier déterminée dans X, on définit $f^*\alpha$ comme l'unique classe de Cartier déterminée dans X' par $f^*\alpha_X$.

On vérifie (voir [**10**]) que ces définitions sont cohérentes, que f_* préserve les classes de Cartier, et que f^* s'étend continûment aux classes de Weil (pour la topologie de la limite projective). On montre de plus que f^* et f_* induisent des opérateurs continus sur $\mathrm{L}^2(\mathfrak{P})$ et vérifient :

$$f^*\alpha\cdot f^*\beta=\alpha\cdot\beta \text{ et } f^*\alpha\cdot\beta=\alpha\cdot f_*\beta$$

pour toutes classes $\alpha, \beta \in \mathrm{L}^2(\mathfrak{P})$, voir [**10**].

Notons maintenant que, pour tout modèle X, on a :

$$\deg(f) = f^*\mathcal{L} \cdot \mathcal{L} = f^{\#}\mathcal{L}_X \cdot \mathcal{L}_X.$$

Lorsque le morphisme naturel $\mu : X \to \mathbb{P}^2$ est régulier, alors $\mathcal{L}_X = \mu^* L$ avec L la classe d'une droite dans le plan. En utilisant le fait que $f^{\#}$ préserve le cône convexe strict $\mathrm{Nef}(X)$ et que pour toutes classes nef α, β on a $2(\alpha \cdot \beta)\,\alpha \geq (\alpha \cdot \alpha)\,\beta$, on montre la

PROPOSITION 2.6. — *Soit X un modèle de $\mathbb{P}^2$. Fixons une norme $\|\cdot\|$ arbitraire sur $\mathrm{H}^2(X, \mathbb{R})$. Alors il existe une constante $C > 0$ telle que pour tout $f \in \mathrm{Cr}(2)$, on ait $C^{-1}\deg(f) \leq \|f^{\#}\| \leq C\deg(f)$.*

Cette proposition a été étendue en toute dimension et dans le cadre kählérien par Dinh et Sibony [**32**].

Nous pouvons maintenant décrire l'action d'une application birationnelle sur la cohomologie H^2. Ce résultat est essentiellement dû à J. Diller et moi-même [**31**] avec des contributions essentielles de M.H. Gizatullin [**50**] pour le cas parabolique (voir aussi [**6**]).

THÉORÈME 2.7. — *Soit $f \in \mathrm{Cr}(2)$. Alors le degré dynamique $\lambda(f)$ est un entier algébrique ≥ 1 dont les conjugués sur $\mathbb{Q}$ distincts de lui-même sont de norme ≤ 1. On est de plus dans un et un seul des trois cas suivants.*

1. *Soit $\lambda(f) > 1$: il existe alors deux classes nef α_+ et $\alpha_- \in \mathcal{C}^*$ telles que $f^*\alpha_+ = \lambda(f)\alpha_+$, et $f^*\alpha_- = \lambda(f)^{-1}\alpha_-$. L'isométrie induite par f sur $\mathbb{H}(\mathfrak{P})$ est de type hyperbolique, et $\deg(f^n) = c \cdot \lambda(f)^n + O(1)$ pour une constante $c > 0$.*
2. *Soit il existe une classe $\alpha \in \mathcal{C}^*$ telle que $f^*\alpha_0 = \alpha_0$ mais f^* ne possède aucun point fixe dans $\mathbb{H}(\mathfrak{P})$. L'isométrie induite par f sur $\mathbb{H}(\mathfrak{P})$ est donc de type parabolique. Dans ce cas, on a deux possibilités.*
   (a) *Soit $\deg(f^n) = c \cdot n + O(1)$ pour une constante $c > 0$ et il existe un modèle X muni d'une fibration rationnelle $\pi : X \to \mathbb{P}^1$ préservée par f.*
   (b) *Soit $\deg(f^n) = c \cdot n^2 + O(n)$ pour une constante $c > 0$ et il existe un modèle X muni d'une fibration elliptique $\pi : X \to \mathbb{P}^1$ préservée par f.*
3. *Soit l'application f^* fixe une classe dans $\mathbb{H}(\mathfrak{P})$. L'isométrie induite par f sur $\mathbb{H}(\mathfrak{P})$ est donc de type elliptique. Dans ce cas, on peut trouver un modèle X sur lequel f induise un automorphisme, et un itéré de f est conjugué à un automorphisme d'une surface rationnelle minimale.*

Démonstration. — La preuve s'effectue en plusieurs étapes.

Première étape. On introduit la notion suivante due à Fornæss et Sibony : un modèle X est dit adapté[10] à f si $(f^n)^\# = (f^\#)^n$ sur $\mathrm{H}^2(X,\mathbb{R})$ pour tout $n \geq 0$. C'est un fait simple mais important que X est un modèle adapté si et seulement s'il n'existe aucune courbe C contractée par f et envoyée par un itéré f^n sur un point d'indétermination de f pour un entier $n \geq 1$.

Par la proposition 2.6, si X est un modèle adapté, alors $\lambda(f)$ est le rayon spectral de l'application linéaire $f^\# : \mathrm{H}^2(X,\mathbb{R}) \to \mathrm{H}^2(X,\mathbb{R})$. Cependant, la construction d'un modèle adapté est en général très compliquée. Il existe même des applications rationnelles (non inversibles) qui n'admettent aucun modèle adapté, voir [**41**]. On a le résultat important suivant.

Théorème 2.8 ([**31**]). — *Soit $f : X \dashrightarrow X$ une application biméromorphe d'une surface complexe compacte lisse. Alors il existe un morphisme biméromorphe $\mu : X' \to X$ tel que la surface X' soit un modèle adapté pour l'application relevée $\mu^{-1} \circ f \circ \mu : X' \dashrightarrow X'$.*

Démonstration. — On écrit l'application biméromorphe f comme composée d'éclatements de points $X = X_0 \overset{\mu_1}{\dashrightarrow} X_1 \overset{\mu_2}{\dashrightarrow} \cdots \overset{\mu_n}{\dashrightarrow} X_n$ suivie par des contractions de courbes $X_n \overset{\mu_{n+1}}{\to} X_{n+1} \overset{\mu_{n+2}}{\to} \cdots \overset{\mu_{2n}}{\to} X_{2n} = X$. Pour simplifier les notations, pour un entier j arbitraire, on note $X_j = X_{j \bmod 2n}$ et $\mu_j = \mu_{j \bmod 2n}$. On définit N comme le nombre d'indices i entre $n+1$ et $2n$ pour lequel la courbe contractée par μ_i est envoyée par la composée $\mu_{j-1} \circ \cdots \circ \mu_{i+1} \circ \mu_i$ sur le point d'indétermination de μ_j (avec donc $j \bmod 2n$ entre 1 et n). Si X n'est pas un modèle adapté, alors $N \geq 1$ et on prend deux indices $i < j$ comme ci-dessus avec $|i-j|$ minimal. On éclate alors les $j-i+1$ points $p_i := \mu_i(C_i)$, $p_{i+1} := \mu_{i+1}(p_i)$, jusqu'à $p_j = \mu_j(p_{j-1})$. Ce faisant, on obtient un nouveau modèle birationnel de X et une nouvelle factorisation de l'application induite par f pour lequel le nombre N a chuté d'au moins une unité. En itérant ce procédé, on arrive finalement à un modèle adapté pour f. □

Deuxième étape. Supposons tout d'abord $\lambda(f) > 1$. On va étudier en détail les propriétés spectrales de $f^\#$ dans un modèle adapté. Si $\mu : X' \to X$ est l'éclatement de X en un point p, on a $\mu_*\mu^*\alpha = \alpha + (\alpha, E)E$ pour toute classe $\alpha \in \mathrm{H}^2(X',\mathbb{R})$ avec $E = \mu^{-1}(p)$. En itérant cette équation on arrive à la formule suivante.

Proposition 2.9 ([**31**]). — *Soit X un modèle adapté pour $f \in \mathrm{Cr}(2)$. Il existe une famille finie de diviseurs effectifs Z_k telle que pour toutes classes $\alpha, \beta \in \mathrm{H}^2(X,\mathbb{R})$ on ait :*

$$f^\#\alpha \cdot f^\#\beta = \alpha \cdot \beta + \sum_k (\alpha, Z_k)\,(\beta, Z_k)\ . \tag{2}$$

[10] On dit aussi que f est algébriquement stable dans X, voir [**46**].

De plus la réunion des supports des Z_k coïncide avec l'ensemble critique de f^{-1}.

Fixons un modèle adapté X pour f de sorte que le rayon spectral de $f^\# : \mathrm{H}^2(X,\mathbb{R}) \to \mathrm{H}^2(X,\mathbb{R})$ soit égal à $\lambda(f)$. Comme $f^\#$ préserve le réseau $\mathrm{H}^2(X,\mathbb{Z})$, $\lambda(f)$ est un entier algébrique. Le cône des classes nef est un cône convexe fermé strict et préservé par $f^\#$: le théorème de Perron-Frobenius fournit donc une classe nef α_X vérifiant $f^\#\alpha_X = \lambda(f)\alpha_X$. De (2) et du fait que la forme d'intersection est de type Minkowski, on montre que $\lambda(f)$ est une valeur propre simple de $f^\#$ et que toutes les autres valeurs propres sont de norme ≤ 1. Les conjugués de $\lambda(f)$ sont toutes des valeurs propres de $f^\#$ ce qui démontre la première assertion du théorème.

On cherche maintenant à construire une classe nef $\alpha_+ \in \mathrm{L}^2(\mathfrak{P})$ vérifiant $f^*\alpha_+ = \lambda(f)\alpha_+$. Le théorème 2.8 permet de choisir une famille de modèles adaptés $\{X_i\}_{i\in I}$ telle que tout modèle de $\mathbb{P}^2$ soit dominé par au moins un modèle X_i de cette famille. Pour alléger les notations, on note $f_i : X_i \dashrightarrow X_i$ l'application induite par f, et $f_i^\# : \mathrm{H}^2(X_i,\mathbb{R}) \to \mathrm{H}^2(X_i,\mathbb{R})$ l'opérateur associé. Pour tout i, on considère l'unique classe α_i vérifiant $f_i^\#\alpha_i = \lambda(f)\alpha_i$ normalisée de telle sorte que $\alpha_i \cdot \mathscr{L}_{X_i} = +1$. C'est une classe nef, et on vérifie que $\mu_*\alpha_j = \alpha_i$ dès que $\mu : X_j \to X_i$ est régulière. Il existe donc une classe de Weil α_+ dont l'incarnation dans chaque X_i est égale à α_i. Cette classe est nef (donc L^2) et elle vérifie $f^*\alpha_+ = \lambda(f)\alpha_+$.

En répétant la même construction avec f^{-1}, on trouve une classe de Weil nef α_- telle que $(f^{-1})^*\alpha_- = \lambda(f)\alpha_-$, et donc $f^*\alpha_- = \lambda(f)^{-1}\alpha_-$. Les deux classes α_+, α_- sont nécessairement distinctes et d'auto-intersection nulle, et la restriction de la forme d'intersection à l'orthogonal $\mathscr{H}$ au plan engendré par α_+ et α_- est définie négative. Comme f^* est une isométrie pour la forme d'intersection, sa restriction à $\mathscr{H}$ est un opérateur unitaire. Le spectre de f^* sur $\mathrm{L}^2(\mathfrak{P})$ est donc inclus dans la réunion du couple $\lambda(f)^{\pm 1}$ et du cercle unité. En écrivant $\deg(f^n) = f^{n*}\mathscr{L} \cdot \mathscr{L}$ et en utilisant les propriétés spectrales de f^* sur L^2 on montre sans problème que $\deg(f^n) = c\lambda(f)^n + O(1)$ avec $c = 1/(\alpha_+ \cdot \alpha_-)$.

Troisième étape. On suppose maintenant que $\lambda(f) = 1$. On choisit un modèle adapté X à f, et on construit comme précédemment une classe $\alpha \in \mathrm{H}^2(X,\mathbb{R})$ nef et $f^\#$-invariante. Soit H le sous-espace de $\mathrm{H}^2(X,\mathbb{R})$ engendré par les classes de courbes irréductibles et contractées par f^{-1}. De (2), on tire $\alpha \in H^\perp$.

Premier cas : H contient α. Dans ce cas, on montre que f préserve une fibration rationnelle et que l'on est dans le cas (2a) du théorème en produisant une courbe C rationnelle lisse d'auto-intersection nulle dont la classe est proportionnelle à α. Cette courbe est obtenue comme suit. On factorise f en une suite d'éclatements de points $\mu_1 : X' \to X$ composée avec une suite de contractions $\mu_2 : X' \to X$. On choisit X' minimale, et on pose $C = \mu_2(E)$ où E est une courbe rationnelle lisse d'auto-intersection -1 contractée par μ_1. Nous renvoyons à [**31**] pour les détails.

Second cas : H ne contient pas α. Les arguments qui suivent sont tirés de [**50**]. Le critère de Grauert permet de contracter toutes les courbes critiques pour f^{-1}, et on obtient un nouveau modèle sur lequel f devient un automorphisme. On peut en fait prendre ce modèle lisse car en dimension 2 tout automorphisme se relève sur la désingularisée minimale, voir [**63**]. Notons que, dans ce cas, $f^{\#}$ préserve la forme d'intersection sur le réseau $\mathrm{H}^2(X,\mathbb{Z})$.

Si la suite des degrés de f est bornée, alors $f^{\#}$ est d'ordre fini. Si $f^{k\#}=\mathrm{id}$ pour un entier k, alors l'application f^k descend comme automorphisme sur un modèle minimal de X. On est donc dans le cas (3) du théorème.

Sinon la suite des degrés tend vers l'infini, et pour toute classe de Kähler ω fixée, $\|f^{n\#}\omega\| \to +\infty$. On en déduit que tout point d'accumulation α de la suite $f^{n\#}\omega/\|f^{n\#}\omega\|$ est nef, vérifie $f^{\#}\alpha=\alpha$, et $\alpha\cdot\alpha=\alpha\cdot K_X=0$. La classe α est de plus entière, et on peut donc trouver un fibré en droites L tel que $\alpha=c_1(L)$. Le théorème de Hirzebruch-Riemann-Roch appliqué à L donne $h^0(L)=1+h^1(L)$. Si L possède au moins deux sections, alors il est associé à une fibration elliptique et l'équation $f^*\alpha=\alpha$ montre que celle-ci est préservée par f. Sinon $h^0(nL)=1$ pour tout $n\geq 0$ et l'on procède ainsi. Notons C le diviseur des zéros d'une section de L. On montre que $\mathcal{O}_C(L)$ n'est pas de torsion dans le groupe de Picard de la courbe (singulière et non réduite) C, en utilisant la suite exacte de restriction $0\to\mathcal{O}_X\to\mathcal{O}_X(C)\to\mathcal{O}_C(C)\to 0$. On regarde ensuite le morphisme $\rho:\mathrm{NS}(X)\to\mathrm{Pic}(C)$. La forme d'intersection est définie négative sur $\ker(\rho)$, donc quitte à remplacer f par un itéré, $f^{\#}|\ker(\rho)$ est l'identité. On montre alors que l'action de f sur $\mathrm{Pic}^0(C)$ est l'identité. Supposons pour simplifier que C soit irréductible, et prenons une classe $\alpha\in\mathrm{H}^2(X,\mathbb{R})$ telle que $\alpha\cdot C=0$. Alors $f^{\#}\alpha-\alpha\in\ker(\rho)$ et pour toute classe $\beta\in\ker(\rho)$, on a $(f^{\#}\alpha-\alpha)\cdot\beta=0$. Ceci implique $f^{\#}=\mathrm{id}$ sur $\mathrm{H}^2(X,\mathbb{R})$, et contredit notre hypothèse $\deg(f^n)\to+\infty$. Lorsque C est réductible, la preuve est analogue bien que plus délicate, voir [**50**, Proposition 6]. □

3. DÉMONSTRATION DES THÉORÈMES A ET B

Commençons par quelques remarques élémentaires. Rappelons que la forme quadratique $q(\alpha):=-\alpha\cdot\alpha+2(\alpha\cdot\mathscr{L})^2$ sur $\mathrm{L}^2(\mathfrak{P})$ est définie positive et munit $\mathrm{L}^2(\mathfrak{P})$ d'une structure d'espace de Hilbert réel. On notera $\|\alpha\|^2:=q(\alpha)$. On introduit aussi l'espace $\Delta(\mathfrak{P})=\{\alpha\in\mathrm{L}^2,\ \alpha\cdot\alpha\geq 0 \text{ et } \alpha\cdot\mathscr{L}=+1\}$, de telle sorte que $\mathscr{C}^*$ est un cône de base $\Delta(\mathfrak{P})$. Pour tout $\epsilon>0$, et toute classe $\alpha\in\Delta(\mathfrak{P})$ d'auto-intersection nulle, on pose $U(\alpha,\epsilon):=\{\beta\in\Delta(\mathfrak{P}),(\alpha\cdot\beta)<\epsilon\}$. C'est un voisinage ouvert de α dans

$\Delta(\mathfrak{P})$ de diamètre 2ϵ. Notons que, lorsque $f \in \mathrm{Cr}(2)$ est hyperbolique, l'opérateur f^* possède exactement deux vecteurs propres $\alpha_+ \neq \alpha_-$ d'auto-intersection nulle dans $\Delta(\mathfrak{P})$ associés aux valeurs propres $\lambda(f)$ et $\lambda(f)^{-1}$. En particulier une telle application ne préserve jamais de fibration. Dans la suite, on notera $\mathrm{Fix}(f) = \{\alpha_+, \alpha_-\}$. Enfin, par commodité, pour tout $f \in \mathrm{Cr}(2)$ on notera $\bar{f}$ l'action projective sur $\Delta(\mathfrak{P})$ induite par f^*.

On fixe désormais un sous-groupe G de $\mathrm{Cr}(2)$ de type fini.

Premier cas : il existe $f, g \in G$ hyperboliques tels que $\mathrm{Fix}(f) \cap \mathrm{Fix}(g) = \varnothing$. Dans ce cas, on montre que, pour un entier N assez grand, f^N et g^N engendrent un groupe libre non-abélien. C'est une simple application du lemme dit du ping-pong ou critère de Klein.

Deuxième cas : G contient un élément parabolique mais pas d'élément hyperbolique. Prenons $f \in G$ un élément parabolique. Il préserve une fibration (rationnelle ou elliptique) définie sur un certain modèle X. La classe de Cartier α_+ déterminée dans X par une fibre de cette fibration est nef et vérifie $f^*\alpha_+ = \alpha_+$. On va montrer que, pour tout $g \in G$ on a $g^*\alpha_+ = \alpha_+$, ce qui prouvera que l'on est dans le cas (2) du théorème B.

On peut normaliser α_+ de telle sorte que $\alpha_+ \in \Delta(\mathfrak{P})$, et on procède par contradiction. Prenons donc $\phi \in G$ tel que $\phi(\alpha_+) \neq \alpha_+$. Alors $g = \phi^{-1} f \phi$ est parabolique et fixe l'unique classe de Cartier $\beta_+ \in \Delta(\mathfrak{P})$ proportionnelle à $\phi(\alpha_+)$. On prend $\epsilon > 0$ de telle sorte que $U(\alpha_+, \epsilon) \cap U(\beta_+, \epsilon) = \varnothing$. Comme g est parabolique, pour N assez grand, $\bar{g}^N(U(\alpha_+, \epsilon))$ est inclus dans un voisinage arbitrairement petit de β_+. Quitte à prendre M assez grand, on aura de même que $\bar{f}^M \bar{g}^N(U(\alpha_+, \epsilon)) \subset U(\alpha_+, \epsilon/2) \subsetneq U(\alpha_+, \epsilon)$. Ceci n'est possible que si l'application $f^M g^N$ est hyperbolique, d'où la contradiction.

Troisième cas : G contient un élément hyperbolique mais on n'est pas dans le premier cas. On va montrer que l'on est dans la situation (1) du théorème B.

Fixons $f \in G$ hyperbolique. On regarde les orbites des éléments α_+ et α_- de $\mathrm{Fix}(f)$ sous l'action projective de G, c'est-à-dire les deux ensembles $S_\pm = \cup_{g \in G} \bar{g}\alpha_\pm \subset \Delta(\mathfrak{P})$. Si S_+ et S_- sont infinis, alors en conjuguant f par des éléments adéquats on trouve deux éléments hyperboliques f_1, f_2 tels que $\mathrm{Fix}(f_1) \cap \mathrm{Fix}(f_2) = \varnothing$, ce que nous avons exclu. On peut donc supposer que S_+ est fini. Comme $\bar{f}$ laisse fixe S_+, cet ensemble est soit réduit à α_+, soit égal à $\mathrm{Fix}(f)$. Dans le premier cas, le groupe G fixe projectivement α_+ et, dans le second, le sous-groupe G_0 des éléments de G fixant les deux classes α_+ et α_- est d'indice au plus deux.

En conclusion, quitte à remplacer G par un sous-groupe d'indice 2, on peut supposer que tous les éléments du groupe fixent projectivement la classe α_+. Considérons le

morphisme $\rho : G \to \mathbb{R}_+^*$ tel que $f^*\alpha_+ = \rho(f)\alpha_+$ pour tout $f \in G$. Pour conclure la preuve, il faut démontrer que l'image $\rho(G)$ est un sous-groupe discret de $\mathbb{R}_+^*$.

Pour cela, on va montrer que $\rho(G) \cap [1,2]$ est un ensemble fini. Prenons $f_1, \dots, f_n$ un système de générateurs de G, et notons K le corps de nombres engendré par les entiers algébriques $\lambda(f_1), \dots, \lambda(f_n)$. On note d le degré de K sur $\mathbb{Q}$. Prenons $\lambda \in \rho(G) \cap [1,2]$, et $a_1, \dots, a_d$ des entiers tels que $\lambda^d = \sum_{i=0}^{d-1} a_{d-i}\lambda^i$. Comme tous les conjugués de λ distincts de λ sont de norme ≤ 1, on a $|a_i| \leq 2\binom{d}{i}$. L'ensemble $\rho(G) \cap [1,2]$ est inclus dans l'ensemble des zéros d'au moins un polynôme $X^d - \sum_{i=0}^{d-1} a_{d-i}X^i$ avec $|a_i| \leq 2\binom{d}{i}$. C'est donc un ensemble fini.

QUATRIÈME CAS : G ne contient que des éléments elliptiques. C'est le cas le plus délicat. La démonstration repose sur le résultat clé suivant dont nous donnons une preuve ci-dessous. La classe $K_X \in \mathrm{H}^2(X, \mathbb{R})$ est celle déterminée par le fibré canonique de X.

PROPOSITION 3.1. — *Il existe une classe de Cartier α nef et G-invariante et vérifiant $\alpha_X \cdot K_X \leq 0$ pour tout modèle X.*

On fixe X un modèle dans lequel α est déterminée, et on considère l'ensemble $\mathscr{C}$ des courbes rationnelles de X contractées par au moins un élément de G. Notons $V \subset \mathrm{H}^2(X, \mathbb{R})$ le sous-espace linéaire engendré par les classes de courbe $C \in \mathscr{C}$. Montrons tout d'abord que $\alpha \in V^\perp$.

Prenons donc une courbe $C \in \mathscr{C}$, $g \in G$ une application qui contracte C, Γ une désingularisation du graphe de g de telle sorte que les deux projections $\pi_i : \Gamma \to X$, $i = 1, 2$, soient des morphismes birationnels et vérifient $g = \pi_2 \circ \pi_1^{-1}$. Si $\hat{C}$ est la transformée stricte de C par π_1, alors par hypothèse $\pi_2(\hat{C})$ est réduit à un point. Nous avons donc $[C] \cdot \alpha_X = \hat{C} \cdot \pi_1^*\alpha_X = \hat{C} \cdot \alpha_\Gamma = \hat{C} \cdot \pi_2^*\alpha_X = 0$, ce qui conclut la preuve.

a) Supposons que α n'appartienne pas à V. Dans ce cas, le critère de Grauert s'applique et l'on peut contracter toutes les courbes de $\mathscr{C}$. On obtient un modèle $\bar{X}$ sur lequel tout élément de G induit un automorphisme. On peut supposer ce modèle lisse. On regarde alors le morphisme naturel $\rho : G \to \mathrm{GL}(\mathrm{H}^2(\bar{X}, \mathbb{Z}))$. Comme tout élément de G est elliptique, tout élément de $\rho(G)$ est de torsion, donc $\rho(G)$ est fini. Le noyau de ρ est un sous-groupe d'indice fini qui descend comme sous-groupe du groupe des automorphismes d'un modèle minimal de $\bar{X}$. Si ce modèle n'est pas $\mathbb{P}^2(\mathbb{C})$, alors un sous-groupe d'indice 2 de G préserve une fibration rationnelle, ce qui conclut la preuve dans ce cas.

b) Supposons maintenant que α appartienne à V. Alors $\mathbb{R} \cdot \alpha$ est le noyau de la restriction de la forme d'intersection à V donc appartient à $\mathrm{H}^2(X, \mathbb{Z})$. On peut donc trouver un fibré en droites L tel que $c_1(L) = \alpha$. Par Hirzebruch-Riemann-Roch, on

a $h^0(nL) - h^1(nL) = 1 - nL \cdot K_X$ pour tout $n \geq 0$. Si $h^0(nL) \geq 2$ pour un entier n, alors L induit une fibration G-invariante. Celle-ci est rationnelle ou elliptique car $L \cdot K_X \leq 0$. Lorsqu'elle est elliptique, on peut considérer son modèle minimal sur lequel G induit un automorphisme. Les arguments de a) s'appliquent alors.

Il nous reste donc à traiter le cas où $h^0(nL) = 1$ pour tout entier $n \geq 0$, ce qui implique $L \cdot L = L \cdot K_X = 0$. Notons C le diviseur des zéros d'une section de L. On va construire un modèle X' tel que le morphisme naturel $X' \to X$ soit régulier et soit un isomorphisme au-dessus de $X \setminus C$, et pour lequel $G \subset \mathrm{Aut}(X')$. On pourra alors conclure comme précédemment.

La construction de X' repose sur l'analyse de l'action de G sur l'espace Γ_C des points infiniment proches au-dessus de C. Cette technique a été utilisée par M. Gizatullin et V.I. Danilov [**52, 53**] dans leur étude du groupe d'automorphismes de certaines surfaces affines. Elle a été systématisée par M. Jonsson et moi-même pour étudier la dynamique à l'infini des applications polynomiales de $\mathbb{C}^2$, voir [**43, 44**].

La construction de l'espace Γ_C suit celle de [**42**]. Soit $\mathfrak{B}_C$ l'ensemble des modèles (modulo isomorphisme) lisses, obtenus par éclatement de points au-dessus de C uniquement, et dont la préimage de C est à croisements normaux. Pour tout $\mu : Y \to X$ dans $\mathfrak{B}_C$, on note Γ_μ le graphe dual du diviseur $\mu^*(C)$ que l'on munit de la métrique suivante. Si E, E' sont des composantes irréductibles de $\mu^*(C)$ s'intersectant, on pose $d_\mu(E, E') = (\mathrm{ord}_E(\mu^* C)\mathrm{ord}_{E'}(\mu^* C))^{-1}$. On montre que ces métriques sont compatibles entre elles et induisent une métrique naturelle sur la limite inductive (l'union) $(\Gamma_C, d_C) := \varinjlim_{\mathfrak{B}_C} (\Gamma_C, d_C)$. On n'utilisera pas ce fait, mais on peut montrer que Γ_C s'identifie à l'espace des normes multiplicatives $|\cdot|$ sur $\mathbb{C}(X)$ centrées sur C et normalisées par $|C| = e^{-1}$ qui deviennent monomiales dans un système adéquat de coordonnées, voir [**42**, Chapitre 6].

Du fait que G préserve la classe $c_1(L)$ au sens de $\mathrm{H}^2(\mathfrak{P})$, G induit une action *isométrique* sur (Γ_C, d_C). Tout élément $g \in G$ est elliptique, donc se relève comme automorphisme d'un modèle $Y \in \mathfrak{B}_C$: en particulier, l'action de g sur Γ_C possède toujours au moins une orbite périodique. Comme $L \cdot L = L \cdot K_X = 0$, le genre de C est 0 ou 1 et deux cas se présentent : soit Γ_C se rétracte sur un cercle, soit c'est un arbre.

Dans le premier cas, G préserve ce cercle, et chacun de ses éléments agit par rotation d'angle rationnel sur celui-ci. Mais G est de type fini, donc les orbites des points du cercle sous l'action de G sont finies. Géométriquement, cela signifie que l'on peut trouver un modèle lisse $\mu : X' \to X$ dominant X et une famille finie de courbes irréductibles $C_1, \ldots, C_n$ de $\mu^* C$ permutées par l'action de tout élément de G. Après contraction de toutes les composantes irréductibles de $\mu^* C$ distinctes des C_i, on trouve un modèle (singulier puis lisse) sur lequel G agit par automorphismes.

Lorsque Γ_C est un arbre, on peut facilement adapter la preuve de [**70**, Corollaire 2, §6.5] pour montrer que G fixe globalement un point de Γ_C. Comme dans le cas précédent, on traduit géométriquement ce fait : on peut trouver un modèle lisse $\mu : X' \to X$ dominant X et une courbe irréductible C_0 de μ^*C fixée par l'action de tout élément de G. Après contraction de toutes les composantes irréductibles de μ^*C distinctes des C_0, on trouve un modèle sur lequel G agit par automorphismes.

Preuve de la proposition 3.1. — On montre tout d'abord que l'on peut trouver une classe G-invariante $\alpha \in \mathcal{C}^*$. Les arguments sont standard dans le contexte des actions de groupe par isométries sur les espaces hyperboliques. Comme précédemment, il est plus commode de travailler sur $\Delta(\mathfrak{P})$ que sur $\mathbb{H}(\mathfrak{P})$. On a deux métriques en concurrence : la métrique « euclidienne » $d_{\mathrm{L}^2}(\alpha,\beta) = \sqrt{-(\alpha-\beta)^2 + 2(\alpha-\beta)\cdot\mathcal{L}}$ définie sur $\Delta(\mathfrak{P})$; et la métrique hyperbolique $d_{\mathbb{H}}(\alpha,\beta)$ qui vérifie $\cosh d_{\mathbb{H}}(\alpha,\beta) = \alpha\cdot\beta/(\alpha^2)(\beta^2)$ et est définie sur $\Delta(\mathfrak{P})\cap\{\alpha^2>0\}$. Pour $g \in \mathrm{Cr}(2)$, on note $\bar g : \Delta(\mathfrak{P}) \to \Delta(\mathfrak{P})$ l'action naturelle.

Si l'orbite de $\mathcal{L}$ est un ensemble borné pour la métrique $d_{\mathbb{H}}$, le lemme du centre [**56**, p. 37] s'applique. Il existe une unique classe $\alpha \in \Delta(\mathfrak{P})$ telle que $\sup_{g\in G} d_{\mathbb{H}}(\alpha, \bar g\mathcal{L})$ soit minimale. Cette classe étant unique, elle est G-invariante. De plus, elle appartient à l'enveloppe convexe des classes $\bar g\mathcal{L}$ qui sont nef et vérifient $(\bar g\mathcal{L})_X \cdot K_X \le 0$ pour tout modèle : c'est donc aussi une classe nef intersectant négativement le fibré canonique de tout modèle.

Sinon on énumère les éléments de G de telle sorte que $d_{\mathbb{H}}(\bar g_n\mathcal{L},\mathcal{L})$ forme une suite croissante tendant vers l'infini. Du fait que tous les éléments de G sont elliptiques, on peut appliquer [**49**, Lemme 35], et donc trouver une constante $C>0$ telle que, pour tout $m \ge n$, on ait

$$d_{\mathbb{H}}(\bar g_m\mathcal{L}, \bar g_n\mathcal{L}) \le d_{\mathbb{H}}(\bar g_m\mathcal{L},\mathcal{L}) + C\ . \tag{3}$$

Bien que la preuve de [**49**] soit rédigée dans le cadre localement compact, on peut l'adapter à notre situation. Cette preuve se ramène en effet à l'existence d'une constante $C>0$ telle que, si l'on a $d_{\mathbb{H}}(\bar g^2\mathcal{L},\mathcal{L}) > d_{\mathbb{H}}(\bar g\mathcal{L},\mathcal{L}) + C$, alors g est nécessairement hyperbolique. Or, pour une constante C adéquate, l'hypothèse implique $d_{\mathbb{H}}(\bar g^n\mathcal{L}, \bar g^m\mathcal{L}) \ge C'|n-m| + C''$ pour tout couple d'entiers n, m, voir [**49**, Théorème 5.16]. En particulier, on a $\log\deg(g^n) \ge C'n + C''$ ce qui implique bien que g est hyperbolique[(11)].

[(11)] En particulier, on a le résultat suivant. Si $\deg(f^2) > 2^{18}\deg(f)$, alors f est de type hyperbolique. Pour $f \in \mathrm{Aut}[\mathbb{C}^2]$, on peut remplacer 2^{18} par 1. Dans ce cas, le résultat est dû à J.-P. Furter [**48**].

À partir de (3), un calcul élémentaire montre que, pour toute classe $\alpha \in \Delta(\mathfrak{P})$ avec $\alpha \cdot \alpha > 0$, on a :

$$\operatorname{diam}_{d_{\mathrm{L}^2}} \{\beta \in \Delta(\mathfrak{P}),\, d_{\mathbb{H}}(\beta,\alpha) \leq d_{\mathbb{H}}(\mathscr{L},\beta) + C \text{ et } d_{\mathbb{H}}(\beta,\mathscr{L}) \geq d_{\mathbb{H}}(\alpha,\mathscr{L})\} \leq \frac{2e^C}{\alpha \cdot \mathscr{L}}.$$

On en déduit que $\bar{g_n}\mathscr{L}$ forme une suite de Cauchy pour d_{L^2} et converge donc vers une classe non-nulle $\alpha \in \Delta(\mathfrak{P})$. On vérifie facilement que cette classe est G-invariante, qu'elle est nef et intersecte négativement le fibré canonique de tout modèle.

On va maintenant montrer que l'on peut « tronquer » de manière astucieuse la classe α pour la rendre Cartier tout en la gardant G-invariante. À cet endroit, on utilise de manière cruciale l'hypothèse de type fini sur G. On fixe donc un système de générateurs $f_1, \ldots, f_n$ de G et, pour chacun d'entre eux, on choisit un modèle X_i tel que f_i définisse un automorphisme de X_i. On choisit un modèle X qui les domine tous, et on note $\pi_i : X \to X_i$ les morphismes birationnels induits.

On utilise la représentation des classes donnée par le théorème 2.2. Pour chaque i, on écrit $\alpha_X = \alpha_{X_i} + \sum_{\mathscr{V}_{X_i}} a_{i,\nu}\mathscr{E}_\nu$ (ici on identifie α_X, α_{X_i} avec les classes de Cartier correspondantes). On note $\epsilon > 0$ le minimum des $|a_{i,\nu}|$ parmi ceux qui sont non-nuls. On écrit alors $\alpha = \alpha_X + \sum_{\mathscr{V}_X} a_\nu \mathscr{E}(\nu)$, et on pose $\hat{\alpha} := \alpha_X + \sum_{a_\nu \geq \epsilon} a_\nu \mathscr{E}(\nu)$. Comme α est nef, elle est en particulier L^2, donc l'ensemble $\{\nu,\, a_\nu \geq \epsilon\}$ est fini : la classe $\hat{\alpha}$ est donc de type Cartier. C'est l'incarnation de α dans un modèle (éventuellement singulier), donc $\hat{\alpha}$ est nef et intersecte négativement le fibré canonique de tout modèle.

Pour conclure, il nous faut démontrer que $\hat{\alpha}$ est G-invariante. Prenons une application f_i parmi les générateurs. Comme f_i induit un automorphisme sur X_i, pour tout $\nu \in \mathscr{V}_{X_i}$ il existe $\nu' \in \mathscr{V}_{X_i}$ tel que $f_i^*\mathscr{E}(\nu) = \mathscr{E}(\nu')$. Notons $V_i = \{\nu \in \mathscr{V}_{X_i},\, \alpha \cdot \mathscr{E}(\nu) < \epsilon\}$ de telle sorte que $\hat{\alpha} = \alpha - \sum_{V_i} (\alpha \cdot \mathscr{E}(\nu))\, \mathscr{E}(\nu)$. Alors f_i^* induit une bijection sur V_i et l'on a : $f_i^*\hat{\alpha} = f_i^*\alpha - \sum_{V_i} (\alpha \cdot \mathscr{E}(\nu))\, f_i^*\mathscr{E}(\nu) = \alpha - \sum_{V_i} (\alpha \cdot \mathscr{E}(\nu))\, \mathscr{E}(\nu) = \hat{\alpha}$, ce qui termine la preuve. □

4. VERS UNE DYNAMIQUE DES SOUS-GROUPES DE $\mathrm{Cr}(2)$?

Poursuivant la similitude entre $\mathrm{Cr}(2)$ et son analogue de dimension un $\mathrm{PGL}(2,\mathbb{C})$, on peut espérer développer une étude dynamique des sous-groupes non élémentaires de $\mathrm{Cr}(2)$ parallèle à la théorie des groupes kleiniens. Dans cette direction, nous avons le résultat suivant de S. Lamy.

THÉORÈME 4.1. — *Soit G un sous-groupe de $\mathrm{Aut}[\mathbb{C}^2]$ de type fini dont tous les éléments sont de type hyperbolique. Alors il existe un ouvert $U \subset \mathbb{C}^2$ non borné tel que l'intersection de la G-orbite de n'importe quel point de $\mathbb{C}^2$ avec U soit finie.*

En d'autres termes, un tel groupe possède toujours un ouvert de discontinuité non trivial. Il serait intéressant d'exhiber d'autres classes de sous-groupes de Cr(2) ayant un ouvert de discontinuité non trivial, ainsi que d'étudier la géométrie des espaces quotients que l'on peut obtenir ainsi.

RÉFÉRENCES

[1] V. I. Arnol'd – Dynamics of complexity of intersections, *Bol. Soc. Brasil. Mat. (N.S.)* **21** (1990), p. 1–10.

[2] E. Bedford & J. Diller – Energy and invariant measures for birational surface maps, *Duke Math. J.* **128** (2005), p. 331–368.

[3] E. Bedford & K. Kim – Dynamics of rational surface automorphisms : linear fractional recurrences, *J. Geom. Anal.* **19** (2009), p. 553–583.

[4] ———, Continuous families of rational surface automorphisms with positive entropy, prépublication arXiv:0804.2078.

[5] B. Bekka, P. De la Harpe & A. Valette – *Kazhdan's property (T)*, Mathematical Monographs, vol. 11, Cambridge Univ. Press, 2008.

[6] M. P. Bellon – Algebraic entropy of birational maps with invariant curves, *Lett. Math. Phys.* **50** (1999), p. 79–90.

[7] M. P. Bellon & C. Viallet – Algebraic entropy, *Comm. Math. Phys.* **204** (1999), p. 425–437.

[8] R. L. Benedetto & W. M. Goldman – The topology of the relative character varieties of a quadruply-punctured sphere, *Experiment. Math.* **8** (1999), p. 85–103.

[9] J. Blanc – Sous-groupes algébriques du groupe de Cremona, *Transform. Groups* **14** (2009), p. 249–285.

[10] S. Boucksom, C. Favre & M. Jonsson – Degree growth of meromorphic surface maps, *Duke Math. J.* **141** (2008), p. 519–538.

[11] E. Breuillard – A strong Tits alternative, prépublication arXiv:0804.1395.

[12] M. R. Bridson & A. Haefliger – *Metric spaces of non-positive curvature*, Grund. Math. Wiss., vol. 319, Springer, 1999.

[13] J. W. Cannon, W. J. Floyd & W. R. Parry – Introductory notes on Richard Thompson's groups, *Enseign. Math.* **42** (1996), p. 215–256.

[14] S. Cantat – Dynamique des automorphismes des surfaces $K3$, *Acta Math.* **187** (2001), p. 1–57.

[15] ———, Version kählérienne d'une conjecture de Robert J. Zimmer, *Ann. Sci. École Norm. Sup.* **37** (2004), p. 759–768.

[16] ———, Bers and Hénon, Painlevé and Schrödinger, *Duke Math. J.* **149** (2009), p. 411–460.

[17] ———, Sur les groupes de transformations birationnelles des surfaces, prépublication.

[18] S. Cantat & S. Lamy – Groupes d'automorphismes polynomiaux du plan, *Geom. Dedicata* **123** (2006), p. 201–221.

[19] S. Cantat & F. Loray – Holomorphic dynamics, Painlevé VI equation and character varieties, prépublication arXiv:0711.1579.

[20] D. Damanik – Substitution Hamiltonians with bounded trace map orbits, *J. Math. Anal. Appl.* **249** (2000), p. 393–411.

[21] V. I. Danilov – Non-simplicity of the group of unimodular automorphisms of an affine plane, *Mat. Zametki* **15** (1974), p. 289–293.

[22] P. Deligne – Intersections sur les surfaces régulières, in *Groupes de monodromie en géométrie algébrique (SGA 7)*, Lecture Notes in Math., vol. 340, Springer, 1973, p. 1–38.

[23] J.-P. Demailly – L^2 vanishing theorems for positive line bundles and adjunction theory, in *Transcendental methods in algebraic geometry (Cetraro, 1994)*, Lecture Notes in Math., vol. 1646, Springer, 1996, p. 1–97.

[24] M. Demazure – Sous-groupes algébriques de rang maximum du groupe de Cremona, *Ann. Sci. École Norm. Sup.* **3** (1970), p. 507–588.

[25] J. Déserti – Groupe de Cremona et dynamique complexe : une approche de la conjecture de Zimmer, *Int. Math. Res. Not.* **2006** (2006), art. ID 71701.

[26] ———, Sur le groupe des automorphismes polynomiaux du plan affine, *J. Algebra* **297** (2006), p. 584–599.

[27] ———, Sur les automorphismes du groupe de Cremona, *Compos. Math.* **142** (2006), p. 1459–1478.

[28] ———, Le groupe de Cremona est hopfien, *C. R. Math. Acad. Sci. Paris* **344** (2007), p. 153–156.

[29] ———, Sur les sous-groupes nilpotents du groupe de Cremona, *Bull. Braz. Math. Soc. (N.S.)* **38** (2007), p. 377–388.

[30] J. A. Dieudonné – *La géométrie des groupes classiques*, Ergebnisse Math. Grenzg., vol. 5, Springer, 1971.

[31] J. Diller & C. Favre – Dynamics of bimeromorphic maps of surfaces, *Amer. J. Math.* **123** (2001), p. 1135–1169.

[32] T.-C. Dinh & N. Sibony – Regularization of currents and entropy, *Ann. Sci. École Norm. Sup.* **37** (2004), p. 959–971.

[33] ———, Une borne supérieure pour l'entropie topologique d'une application rationnelle, *Ann. of Math.* **161** (2005), p. 1637–1644.

[34] I. Dolgachev – Infinite Coxeter groups and automorphisms of algebraic surfaces, in *The Lefschetz centennial conference, Part I (Mexico City, 1984)*, Contemp. Math., vol. 58, Amer. Math. Soc., 1986, p. 91–106.

[35] I. Dolgachev & V. A. Iskovskikh – Finite subgroups of the plane Cremona group., in *Algebra, arithmetic and geometry, Manin's Festschrift*, Progress in Math., Birkhäuser, 2008.

[36] A. Dubouloz & S. Lamy – Variations on log Sarkisov program for surfaces, prépublication arXiv:0802.2441.

[37] B. Dubrovin & M. Mazzocco – Monodromy of certain Painlevé-VI transcendents and reflection groups, *Invent. Math.* **141** (2000), p. 55–147.

[38] R. Dujardin – Laminar currents and birational dynamics, *Duke Math. J.* **131** (2006), p. 219–247.

[39] F. Enriques – Sui gruppi continui di transformazioni cremoniani nel piano, *Rend. Accad. Lincei* (1er sem., 1893).

[40] E. Falbel & J. R. Parker – The moduli space of the modular group in complex hyperbolic geometry, *Invent. Math.* **152** (2003), p. 57–88.

[41] C. Favre – Les applications monomiales en deux dimensions, *Michigan Math. J.* **51** (2003), p. 467–475.

[42] C. Favre & M. Jonsson – *The valuative tree*, Lecture Notes in Math., vol. 1853, Springer, 2004.

[43] ———, Eigenvaluations, *Ann. Sci. École Norm. Sup.* **40** (2007), p. 309–349.

[44] ———, Dynamical compactifications of $\mathbb{C}^2$, prépublication arXiv:0711.2770.

[45] D. Fisher – Groups acting on manifolds : around the Zimmer program, prépublication arXiv:0809.4849.

[46] J. E. Fornaess & N. Sibony – Complex dynamics in higher dimension. II, in *Modern methods in complex analysis (Princeton, NJ, 1992)*, Ann. of Math. Stud., vol. 137, Princeton Univ. Press, 1995, p. 135–182.

[47] S. Friedland & J. Milnor – Dynamical properties of plane polynomial automorphisms, *Ergodic Theory Dynam. Systems* **9** (1989), p. 67–99.

[48] J.-P. Furter – On the degree of iterates of automorphisms of the affine plane, *Manuscripta Math.* **98** (1999), p. 183–193.

[49] E. Ghys & P. De la Harpe (éds.) – *Sur les groupes hyperboliques d'après Mikhael Gromov*, Progress in Math., vol. 83, Birkhäuser, 1990.

[50] M. K. Gizatullin – Rational G-surfaces, *Izv. Akad. Nauk SSSR Ser. Mat.* **44** (1980), p. 110–144.

[51] ———, Defining relations for the Cremona group of the plane, *Izv. Akad. Nauk SSSR Ser. Mat.* **46** (1982), p. 909–970.

[52] M. K. Gizatullin & V. I. Danilov – Automorphisms of affine surfaces. I, *Izv. Akad. Nauk SSSR Ser. Mat.* **39** (1975), p. 523–565.

[53] ———, Automorphisms of affine surfaces. II, *Izv. Akad. Nauk SSSR Ser. Mat.* **41** (1977), p. 54–103.

[54] W. M. Goldman – The modular group action on real SL(2)-characters of a one-holed torus, *Geom. Topol.* **7** (2003), p. 443–486.

[55] M. Gromov – On the entropy of holomorphic maps, *Enseign. Math.* **49** (2003), p. 217–235.

[56] P. de la Harpe & A. Valette – La propriété (T) de Kazhdan pour les groupes localement compacts, *Astérisque* **175** (1989).

[57] J. Hietarinta & C. Viallet – Singularity confinement and degree growth, in *SIDE III—symmetries and integrability of difference equations (Sabaudia, 1998)*, CRM Proc. Lecture Notes, vol. 25, Amer. Math. Soc., 2000, p. 209–216.

[58] V. A. Iskovskikh – Proof of a theorem on relations in the two-dimensional Cremona group, *Uspekhi Mat. Nauk* **40** (1985), p. 255–256.

[59] K. Iwasaki & T. Uehara – An ergodic study of Painlevé VI, *Math. Ann.* **338** (2007), p. 295–345.

[60] S. Lamy – Problèmes de densité d'orbites pour des groupes d'automorphismes de $\mathbf{C}^2$, *C. R. Acad. Sci. Paris Sér. I Math.* **329** (1999), p. 807–810.

[61] ———, Dynamique des groupes paraboliques d'automorphismes polynomiaux de $\mathbb{C}^2$, *Bol. Soc. Brasil. Mat. (N.S.)* **32** (2001), p. 185–212.

[62] ———, L'alternative de Tits pour Aut$[\mathbb{C}^2]$, *J. Algebra* **239** (2001), p. 413–437.

[63] H. B. Laufer – *Normal two-dimensional singularities*, Annals of Math. Studies, vol. 71, Princeton Univ. Press, 1971.

[64] R. Lazarsfeld – *Positivity in algebraic geometry. I & II*, Ergebnisse Math. Grenzg., vol. 48 & 49, Springer, 2004.

[65] L. Makar-Limanov – On groups of automorphisms of a class of surfaces, *Israel J. Math.* **69** (1990), p. 250–256.

[66] Y. I. Manin – *Cubic forms*, 2e éd., North-Holland Mathematical Library, vol. 4, North-Holland Publishing Co., 1986.

[67] G. A. Margulis – *Discrete subgroups of semisimple Lie groups*, Ergebnisse Math. Grenzg., vol. 17, Springer, 1991.

[68] C. T. McMullen – Dynamics on blowups of the projective plane, *Publ. Math. I.H.É.S.* **105** (2007), p. 49–89.

[69] A. Russakovskii & B. Shiffman – Value distribution for sequences of rational mappings and complex dynamics, *Indiana Univ. Math. J.* **46** (1997), p. 897–932.

[70] J-P. Serre – Arbres, amalgames, SL_2, *Astérisque* **46** (1977).

[71] ———, Le groupe de Cremona et ses groupes finis, Séminaire Bourbaki 2008/09, exposé n° 1000, ce volume.

[72] V. V. Shokurov – Prelimiting flips, *Tr. Mat. Inst. Steklova* **240** (2003), p. 82–219.

[73] N. Sibony – Dynamique des applications rationnelles de $\mathbf{P}^k$, in *Dynamique et géométrie complexes (Lyon, 1997)*, Panor. Synthèses, vol. 8, Soc. Math. France, 1999, p. 97–185.

[74] R. Steinberg – Some consequences of the elementary relations in SL_n, in *Finite groups—coming of age (Montreal, Que., 1982)*, Contemp. Math., vol. 45, Amer. Math. Soc., 1985, p. 335–350.

[75] J. Tits – Free subgroups in linear groups, *J. Algebra* **20** (1972), p. 250–270.

[76] H. Umemura – On the maximal connected algebraic subgroups of the Cremona group. I, *Nagoya Math. J.* **88** (1982), p. 213–246.

[77] ———, On the maximal connected algebraic subgroups of the Cremona group. II, in *Algebraic groups and related topics (Kyoto/Nagoya, 1983)*, Adv. Stud. Pure Math., vol. 6, North-Holland, 1985, p. 349–436.

[78] A. Usnich – Symplectic automorphisms of CP^2 and the Thompson group T, prépublication arXiv:math/0611604.

[79] A. P. Veselov – Growth and integrability in the dynamics of mappings, *Comm. Math. Phys.* **145** (1992), p. 181–193.

[80] D. Wright – Abelian subgroups of $\mathrm{Aut}_k(k[X, Y])$ and applications to actions on the affine plane, *Illinois J. Math.* **23** (1979), p. 579–634.

[81] O. Zariski – The compactness of the Riemann manifold of an abstract field of algebraic functions, *Bull. Amer. Math. Soc.* **50** (1944), p. 683–691.

[Note ajoutée en avril 2010. Dans un travail récent, J. Blanc a muni Cr(2) d'une structure de groupe topologique connexe et montré qu'il était topologiquement simple.]

Charles FAVRE

CNRS et Université Paris VII
UFR de Mathématiques
Équipe Géométrie et Dynamique
UMR 7586 du CNRS
Case 7012
2 place Jussieu
F-75230 Paris cedex 05
E-mail : favre@math.jussieu.fr

332

ASTÉRISQUE

2010

SÉMINAIRE BOURBAKI
VOLUME 2008/2009
EXPOSÉS 997-1011

(999) *Convexes divisibles*

Jean-François QUINT

SOCIÉTÉ MATHÉMATIQUE DE FRANCE

Publié avec le concours du CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

Séminaire BOURBAKI Novembre 2008
61e année, 2008-2009, no 999, p. 45 à 73

CONVEXES DIVISIBLES
[d'après Yves Benoist]

par **Jean-François QUINT**

1. INTRODUCTION

Dans cet exposé, nous allons nous intéresser à des groupes discrets d'automorphismes projectifs de certains ouverts convexes de l'espace projectif.

De même que l'étude des groupes discrets d'automorphismes de l'espace euclidien peut s'interpréter en termes de pavages euclidiens périodiques, celle des groupes discrets d'automorphismes de convexes peut se comprendre comme celle de pavages projectifs périodiques. Quand l'ouvert convexe auquel on s'intéresse est un ellipsoïde, ces pavages sont des pavages hyperboliques et les situations que nous allons rencontrer peuvent être en grande partie considérées comme des généralisations de celle-ci.

Nous verrons comment la compréhension de ces groupes d'automorphismes fait appel à des théories variées comme les systèmes dynamiques hyperboliques, la théorie géométrique des groupes, les groupes algébriques, les représentations des groupes discrets, l'analyse non-linéaire...

Je remercie chaleureusement Yves Benoist pour sa relecture attentive de ce texte et ses nombreuses remarques et suggestions, ainsi que Frédéric Paulin pour ses corrections de la première version.

1.1. Convexes

Soit C un cône ouvert convexe dans un espace vectoriel réel de dimension finie V. On dit que C est saillant (ou proprement convexe) s'il ne contient pas de droite affine, ce qui revient à dire qu'il existe un hyperplan H de V tel que $\bar{C} - \{0\}$ soit contenu dans une des composantes connexes de $V - H$.

Soit Ω un ouvert de l'espace projectif $\mathbb{P}(V)$. On dit que Ω est convexe s'il existe un hyperplan H de V tel que $\Omega \cap \mathbb{P}(H) = \varnothing$ et que Ω soit convexe au sens usuel dans l'ouvert affine $\mathbb{P}(V) - \mathbb{P}(H)$. Si Ω est convexe, on dit qu'il est saillant s'il existe

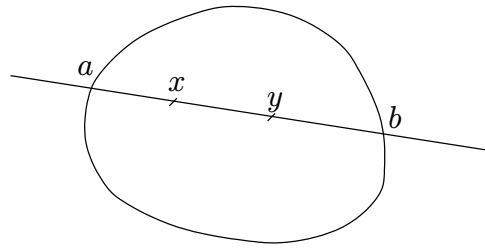


FIGURE 1. Distance de Hilbert

un hyperplan H de V tel que l'adhérence de Ω ne rencontre pas $\mathbb{P}(H)$. L'ouvert Ω est convexe (resp. convexe saillant) si et seulement s'il existe dans V un cône ouvert convexe différent de V (resp. convexe saillant) dont Ω soit la trace projective.

Un cône ouvert convexe peut toujours être vu comme un ouvert convexe de l'espace projectif $\mathbb{P}(V \oplus \mathbb{R})$.

On note $\operatorname{Aut} C$ (resp. $\operatorname{Aut}\Omega$) le sous-groupe fermé de $\mathrm{GL}(V)$ (resp. de $\mathrm{PGL}(V)$) constitué des éléments qui stabilisent le cône ouvert convexe C de V (resp. l'ouvert convexe Ω de $\mathbb{P}(V)$).

Soit Ω un ouvert convexe saillant de $\mathbb{P}(V)$. On peut définir (voir **[15]**) sur Ω une distance d_Ω invariante par $\operatorname{Aut}\Omega$, appelée distance de Hilbert, de la façon suivante. Si x et y sont deux points distincts de Ω, la droite projective engendrée par x et y intersecte la frontière de Ω en deux points distincts a et b. La distance de Hilbert $d_\Omega(x,y)$ entre x et y est alors la valeur absolue du birapport $[a,b,x,y]$ entre ces quatre points, c'est-à-dire que, pour un choix de paramétrisation projective de cette droite par $\mathbb{P}^1_{\mathbb{R}} = \mathbb{R} \cup \{\infty\}$, on a

$$d_\Omega(x,y) = \frac{1}{2}\left|\log\left(\frac{x-b}{x-a}\frac{y-a}{y-b}\right)\right|$$

(voir figure 1). La distance d_Ω est propre, c'est-à-dire que ses boules fermées sont compactes, et elle induit sur Ω la topologie usuelle de Ω. Le groupe $\operatorname{Aut}\Omega$ préserve la distance d_Ω et agit proprement sur Ω (voir lemme 2.11).

On dit qu'un cône ouvert convexe saillant C de V est divisible s'il existe un sous-groupe discret Λ de $\operatorname{Aut} C$ tel que $\Lambda\backslash C$ soit compact. De même, on dit qu'un ouvert convexe saillant Ω de $\mathbb{P}(V)$ est divisible s'il existe un sous-groupe discret Γ de $\operatorname{Aut}\Omega$ tel que $\Gamma\backslash\Omega$ soit compact. D'après le lemme de Selberg (voir **[3, 42]**), on peut toujours supposer que Λ et Γ sont sans torsion.

Exemple 1.1. — Soient $d \geq 3$ la dimension de V, q une forme quadratique de signature $(1, d-1)$ sur V et C_q une des deux composantes connexes de l'ensemble $\{x \in V | q(x) > 0\}$. Alors, C_q est un cône convexe saillant, homogène sous l'action

d'un sous-groupe d'indice 2 du groupe des similitudes de q. Comme ce groupe admet des sous-groupes discrets co-compacts, le cône C_q est divisible. De même, la trace projective Ω_q de C_q est un ouvert convexe saillant divisible : on l'appelle l'ellipsoïde de q. L'ouvert Ω_q s'identifie à l'espace hyperbolique réel de dimension $d-1$ et la distance de Hilbert est alors égale à la distance hyperbolique. En particulier, si Γ est un sous-groupe discret (sans torsion) de $\operatorname{Aut}\Omega_q$ qui divise Ω_q, l'espace $\Gamma\backslash\Omega_q$ est une variété hyperbolique.

Exemple 1.2. — Le groupe des matrices diagonales à coefficients >0 agit simplement transitivement sur le cône ouvert convexe saillant $(\mathbb{R}_+^*)^d \subset \mathbb{R}^d$. Si Γ est un groupe discret de telles matrices qui agit co-compactement sur $(\mathbb{R}_+^*)^d \subset \mathbb{R}^d$, le quotient est difféomorphe au tore $\mathbb{T}^d$.

Exemple 1.3. — Le groupe $\mathrm{GL}_d(\mathbb{R})$ agit transitivement sur le cône convexe saillant des matrices symétriques définies positives. Les réseaux de $\mathrm{GL}_d(\mathbb{R})$ divisent donc ce cône.

Dans tous ces exemples, les cônes qui apparaissent sont homogènes sous leur groupe d'automorphismes. Nous verrons plus loin qu'il existe de nombreux cônes convexes divisibles non homogènes. Les premiers exemples de tels cônes divisibles non homogènes ont été construits par Kac et Vinberg dans [**29**]. Auparavant, les cônes convexes homogènes avaient été classifiés par Vinberg dans [**46**] et [**47**].

L'étude systématique des convexes divisibles remonte aux travaux de Benzécri [**13**]. En particulier, Benzécri montre qu'un ouvert convexe saillant divisible Ω de $\mathbb{P}^2$ qui n'est pas un triangle, c'est-à-dire qui n'est pas conjugué par un élément de $\mathrm{PGL}_3(\mathbb{R})$ à la trace projective de $(\mathbb{R}_+^*)^3$, est strictement convexe et a un bord de classe $\mathscr{C}^1$. En outre, si la dérivée seconde de $\partial\Omega$ est définie et non nulle en un point, Ω est un ellipsoïde.

Dans [**45**], Vey montre que, si un cône ouvert convexe saillant C de V est divisé par un groupe Γ, la représentation de Γ dans V est semi-simple. Plus précisément, si $V = V_1 \oplus \ldots \oplus V_l$ est la décomposition de V en Γ-sous-modules irréductibles, il existe des cônes ouverts convexes saillants $C_1 \subset V_1, \ldots, C_l \subset V_l$ tels que $C = C_1 + \ldots + C_l$. Ce résultat implique en particulier que, si Ω est un ouvert convexe saillant d'un espace affine qui est divisé par un groupe de transformations affines, le convexe Ω est un cône. L'étude des ouverts convexes affines divisibles se ramène donc bien à celle des cônes convexes divisibles.

1.2. Structures localement homogènes

La théorie des convexes divisibles est fortement liée à celle des espaces localement homogènes. Rappelons que, si G est un groupe de Lie connexe agissant transitivement

sur une variété X, un (G,X)-atlas sur une variété M est un ensemble $\mathscr{T}$ de cartes (U,φ), où U est un ouvert de M et φ un difféomorphisme de U sur un ouvert de X, ayant la propriété que, si (U,φ) et (W,ψ) appartiennent à $\mathscr{T}$, il existe un élément g de G tel que, pour tout x dans $U\cap W$, on ait $\psi(x)=g\varphi(x)$. Une (G,X)-structure sur M consiste en la donnée d'un (G,X)-atlas maximal $\mathscr{T}$ sur M avec $\bigcup_{(U,\varphi)\in\mathscr{T}} U = M$. Les éléments de $\mathscr{T}$ sont alors appelés cartes compatibles avec la (G,X)-structure. Par définition, les cartes compatibles recouvrent M. Supposons donc M munie d'une telle (G,X)-structure et notons $\pi : \tilde{M} \to M$ un revêtement universel de M et Γ le groupe des automorphismes de ce revêtement. Il existe alors un difféomorphisme local $D : \tilde{M} \to X$ et un homomorphisme de groupes $h : \Gamma \to G$ ayant les propriétés suivantes : *(i)* pour tout ouvert W de $\tilde{M}$ tel que π induise un difféomorphisme de W sur $\pi(W)$ et pour toute carte compatible (U,φ) avec U connexe et $U\subset\pi(W)$, il existe un élément g de G tel que, pour tout x dans W avec $\pi(x)\in U$, on ait $D(x)=g\varphi(\pi(x))$; *(ii)* pour tous x dans $\tilde{M}$ et γ dans Γ, on a $D(\gamma x)=h(\gamma)D(x)$. Si (D',h') est un autre couple satisfaisant aux mêmes propriétés, il existe g dans G tel que, pour tout x dans X, on ait $D'(x)=gD(x)$ et que, pour tout γ dans Γ, on ait $h'(\gamma)=gh(\gamma)g^{-1}$. Par abus de langage, on dit que D est la développante de la (G,X)-structure et h son morphisme d'holonomie. La donnée de D et de h détermine complètement la (G,X)-structure.

Exemple 1.4. — Le groupe $\mathrm{GA}(V)$ des automorphismes affines de V agit transitivement sur V. Soient $C\subset V$ un cône ouvert convexe saillant et Γ un sous-groupe discret sans torsion de $\operatorname{Aut} C$. La $(\mathrm{GA}(V),V)$-structure naturelle de C induit une $(\mathrm{GA}(V),V)$-structure sur $\Gamma\backslash C$. Son application développante est l'injection naturelle $C\to V$ et son holonomie l'injection naturelle $\Gamma\to\mathrm{GA}(V)$.

Exemple 1.5. — Le groupe $\mathrm{PGL}(V)$ des automorphismes projectifs de V agit transitivement sur $\mathbb{P}(V)$. Soient $\Omega\subset\mathbb{P}(V)$ un ouvert convexe saillant et Λ un sous-groupe discret sans torsion de $\operatorname{Aut}\Omega$. La $(\mathrm{PGL}(V),\mathbb{P}(V))$-structure naturelle de Ω induit une $(\mathrm{PGL}(V),\mathbb{P}(V))$-structure sur $\Lambda\backslash\Omega$. Son application développante est l'injection naturelle $\Omega\to\mathbb{P}(V)$ et son holonomie l'injection naturelle $\Lambda\to\mathrm{PGL}(V)$.

Dans ces exemples, l'application développante est injective. Ce n'est pas toujours le cas comme on peut le voir, par exemple dans **[21]** ou **[44]**.

Une $(\mathrm{GA}(V),V)$-structure sera appelée structure affine plate. Une $(\mathrm{PGL}(V),\mathbb{P}(V))$-structure sera appelée structure projective plate. Supposons M munie d'une structure projective (resp. affine) plate. On appelle alors géodésiques de M les courbes tracées sur M dont les composantes connexes de l'intersection avec toute carte compatible sont des segments de droite projective (resp. affine). La structure projective (resp. affine) sera dite convexe si et seulement si toute courbe

tracée sur M est homotope, à extrémités fixées, à un segment géodésique. D'après [**21**, 3.1], la structure projective (resp. affine) de M est convexe si et seulement si sa développante est un difféomorphisme sur un ouvert convexe de $\mathbb{P}(V)$ (resp. de V), c'est-à-dire si M s'identifie au quotient de ce convexe par un groupe discret d'automorphismes et que la structure plate est celle considérée dans les exemples ci-dessus. En particulier, si M est compacte et si cet ouvert convexe est saillant, il est divisible et, si la structure est affine, d'après le théorème de Vey, l'image de sa développante est un cône.

1.3. Déformation de convexes divisibles

Soit M une variété compacte. D'après un théorème de Koszul [**32, 33**], une structure affine plate sur M est convexe à revêtement universel saillant si et seulement s'il existe une 1-forme fermée sur M dont la dérivée covariante (au sens de la connexion naturellement associée à la structure affine) est définie positive. En particulier, l'ensemble des structures affines plates convexes à revêtement universel saillant est ouvert dans l'ensemble des structures affines plates de M.

Soit C un cône ouvert convexe saillant de V et soit Γ un sous-groupe discret (sans torsion) de $\operatorname{Aut} C$ qui divise C. Soit $\rho_0 : \Gamma \to \mathrm{GL}(V)$ l'injection naturelle. D'après un théorème de Thurston, généralisant une idée originale de Weil [**16, 49**], tout morphisme $\rho : \Gamma \to \mathrm{GL}(V)$ suffisamment proche de ρ_0 est le morphisme d'holonomie d'une structure affine plate sur $M = \Gamma\backslash C$. En particulier, d'après le théorème de Koszul, si ρ est suffisamment proche de ρ_0, cette structure est convexe et son revêtement universel est saillant. Par conséquent, dans l'espace des homomorphismes de Γ dans $\mathrm{GL}(V)$, ceux dont l'image est discrète et divise un cône ouvert convexe saillant de V constituent une partie ouverte.

Étendons ces résultats au cas projectif. Notons $\mathrm{SL}^{\pm}(V)$ le groupe des automorphismes linéaires de V dont le déterminant a pour valeur absolue 1. Soient Ω un ouvert convexe saillant de $\mathbb{P}(V)$ et Λ un sous-groupe discret (sans torsion) de $\operatorname{Aut}\Omega$ qui divise Ω. Notons C l'un des deux cônes ouverts convexes de V dont l'image dans $\mathbb{P}(V)$ est Ω. Alors, le morphisme naturel de Λ dans $\mathrm{PGL}(V)$ se relève de manière unique en un morphisme de Λ dans $\mathrm{SL}^{\pm}(V)$ qui préserve C. Soit h l'homothétie de rapport 2 dans V : le groupe $h^{\mathbb{Z}}\Lambda$ divise le cône C. En utilisant le cas affine, on en déduit donc à nouveau que, dans l'espace des homomorphismes de Λ dans $\mathrm{PGL}(V)$, ceux dont l'image est discrète et divise un ouvert convexe saillant de $\mathbb{P}(V)$ constituent une partie ouverte.

Pour tout $n \geq 2$, les réseaux co-compacts de $\mathrm{SO}(1,n)$ divisent un ellipsoïde dans $\mathbb{P}^n$. Dans [**28**], Johnson et Millson construisent des déformations de certains de ces réseaux qui sont Zariski denses dans $\mathrm{SL}_{n+1}(\mathbb{R})$. Quand ces déformations sont suffisamment proches de la représentation hyperbolique, elles divisent donc un ouvert convexe

saillant de $\mathbb{P}(V)$. Il découle de la classification de Vinberg [**47**] que ces ouverts convexes saillants ne sont pas homogènes.

Si Λ est un groupe, on appelle centre virtuel de Λ le sous-groupe de Λ constitué des éléments dont le centralisateur est d'indice fini dans Λ. Dans [**10**], Benoist démontre le

THÉORÈME 1.6. — *Soit Λ un groupe de centre virtuel trivial. L'ensemble des homomorphismes fidèles et discrets de Λ dans $\mathrm{PGL}(V)$ dont l'image divise un ouvert convexe saillant de $\mathbb{P}(V)$ est fermé dans l'ensemble des homomorphismes de Λ dans $\mathrm{PGL}(V)$.*

Cet ensemble est donc alors une réunion de composantes connexes de l'ensemble des homomorphismes de Λ dans $\mathrm{PGL}(V)$. Nous donnerons des éléments de la démonstration de ce théorème dans la section 2.

Quand V est de dimension 3 et Λ sans torsion, l'hypothèse sur Λ implique que, si une représentation fidèle et discrète ρ de Λ dans $\mathrm{PGL}(V)$ divise un ouvert convexe saillant Ω de $\mathbb{P}(V)$, le quotient $\rho(\Lambda)\backslash\Omega$ est une surface fermée de genre ≥ 2. Dans ce cas, le théorème 1.6 avait été démontré par Choi et Goldman dans [**17**]. Les structures projectives plates convexes sur le tore de dimension 2 ont été classifiées par Kuiper dans [**36**] et [**37**] ; celles sur les surfaces de genre ≥ 2 l'ont été par Goldman dans [**22**].

1.4. Hyperbolicité

Un espace métrique (X, d) est dit géodésique si, pour tous x et y dans X, il existe une courbe isométrique $\gamma : [0, d(x, y)] \to X$ avec $\gamma(0) = x$ et $\gamma(d(x, y)) = y$. Une telle courbe est appelée géodésique de x à y. Bien que γ ne soit a priori pas unique, on note parfois $[x, y]$ l'ensemble $\gamma([0, d(x, y)])$. L'espace (X, d) est dit propre si ses boules fermées sont compactes. Si Ω est un ouvert convexe saillant de $\mathbb{P}(V)$, l'espace métrique (Ω, d_Ω) est géodésique et propre et ses segments projectifs sont des géodésiques.

Rappelons (voir [**20**]) que, si δ est un réel > 0, un espace métrique géodésique (X, d) est dit δ-hyperbolique au sens de Gromov si, pour tous x, y, z dans X, pour tout choix de géodésiques $[x, y]$, $[y, z]$ et $[x, z]$, pour tout u dans $[x, y]$, il existe v dans $[x, z] \cup [y, z]$ avec $d(u, v) \leq \delta$: on dit aussi que les triangles de X sont δ-fins.

Un groupe de type fini est dit hyperbolique si, pour une certaine partie génératrice, son graphe de Cayley, muni de sa distance naturelle, est un espace hyperbolique au sens de Gromov. Cette définition ne dépend pas de la partie génératrice choisie. Plus généralement, si un groupe de type fini Γ agit proprement et co-compactement par isométries sur un espace métrique géodésique et propre X, le groupe Γ est hyperbolique si et seulement X est hyperbolique (voir [**20**]).

Dans le cas des convexes divisibles, cet énoncé a été précisé par Benoist dans [**9**], sous la forme du

THÉORÈME 1.7. — *Soient Ω un ouvert convexe saillant de $\mathbb{P}(V)$ et Λ un sous-groupe discret de $\operatorname{Aut}\Omega$ divisant Ω. Les assertions suivantes sont équivalentes*

(i) *Le groupe Λ est hyperbolique.*
(ii) *Le bord de Ω est de classe $\mathcal{C}^1$.*
(iii) *Le convexe Ω est strictement convexe.*

Un ouvert convexe saillant Ω de $\mathbb{P}(V)$ est dit strictement convexe si son bord ne contient pas de segment de droite projective non réduit à un point. Il existe des ouverts strictement convexes à bord $\mathcal{C}^1$ dont la distance de Hilbert n'est pas hyperbolique (voir [**8**, 1.3]). Dans [**31**], Karlsson et Noskov ont montré qu'un ouvert convexe saillant dont le bord était de classe $\mathcal{C}^2$ à hessien défini positif avait une métrique de Hilbert hyperbolique. Ce critère a été amélioré par Benoist. Pour énoncer son résultat, introduisons une notion de régularité sur les fonctions convexes, inspirée de la notion de fonction quasi-symétrique apparaissant dans [**2**]. Si U est un ouvert convexe d'un espace vectoriel réel de dimension finie E et $f : U \to \mathbb{R}$ une fonction convexe de classe $\mathcal{C}^1$, nous dirons que f est quasi-symétrique s'il existe un réel $H \geq 1$ tel que, pour tous x dans U et h dans E tels que $x+h$ et $x-h$ appartiennent à U, on ait

$$f(x+h) - f(x) - \mathrm{d}f(x)h \leq H(f(x-h) - f(x) - \mathrm{d}f(x)(-h)).$$

Un ouvert convexe saillant Ω de $\mathbb{P}(V)$ sera dit quasi-symétriquement convexe s'il existe un recouvrement de son bord par des ouverts où, pour un système de coordonnées projectives convenables, le bord de Ω est le graphe d'une fonction quasi-symétriquement convexe. Dans [**8**], Benoist démontre le

THÉORÈME 1.8. — *Soient Ω un ouvert convexe saillant de $\mathbb{P}(V)$ et d_Ω sa distance de Hilbert. Alors l'espace métrique (Ω, d_Ω) est hyperbolique si et seulement si Ω est quasi-symétriquement convexe.*

Nous donnerons des éléments de la démonstration des théorèmes 1.7 et 1.8 à la section 3.

2. DÉFORMATION DE CONVEXES DIVISIBLES

2.1. Adhérence de Zariski des groupes qui divisent un convexe

L'étude des déformations de représentations divisant un cône convexe passe entre autres par une description précise de l'adhérence de Zariski dans $\mathrm{PGL}(V)$ d'un sous-groupe discret divisant un tel cône.

Si C est un cône ouvert convexe saillant de V, on dit que C est un cône produit s'il existe une décomposition non triviale en somme directe $V = V_1 \oplus V_2$ de V et des cônes

convexes ouverts saillants C_1 de V_1 et C_2 de V_2 tels que $C = C_1 + C_2$. Dans le cas contraire, on dit que C est irréductible. Dans le cas général, il existe, à permutation près, une unique décomposition $V = V_1 \oplus \ldots \oplus V_l$ et d'uniques cônes convexes ouverts saillants irréductibles $C_1 \subset V_1, \ldots, C_l \subset V_l$ tels que $C = C_1 + \ldots + C_l$. En particulier, $\mathrm{Aut}\, C$ permute les espaces $V_1, \ldots, V_l$ et la composante Zariski connexe de $\mathrm{Aut}\, C$ stabilise chacun de ces espaces. Dans [**7**], Benoist décrit l'adhérence de Zariski des groupes Zariski connexes qui divisent un cône :

THÉORÈME 2.1 ([**7**, Théorème 1.1 et Proposition 4.4]). — *Soient C un cône convexe ouvert saillant de V, $C = C_1 + \ldots + C_l$ sa décomposition en cônes irréductibles, $V = V_1 \oplus \ldots \oplus V_l$ la décomposition de V associée et Γ un sous-groupe discret et Zariski connexe de $\mathrm{Aut}\, C$ qui divise C. Alors l'adhérence de Zariski de Γ est un produit $G_1 \times \ldots \times G_l$ où, pour tout $1 \leq i \leq l$,*

(i) *le groupe G_i est un sous-groupe réductif de $\mathrm{GL}(V_i)$ contenant les homothéties ;*
(ii) *le groupe $S_i = \mathrm{SL}(V_i) \cap G_i$ est simple ;*
(iii) *si C_i est homogène, G_i est commensurable à $\mathrm{Aut}\, C_i$;*
(iv) *si C_i n'est pas homogène, il est cependant divisible et $G_i = \mathrm{GL}(V_i)$.*

En outre, le commutant H de Γ dans $\mathrm{GL}(V)$ est exactement le groupe des éléments dont la restriction à chacun des V_i, $1 \leq i \leq l$, est scalaire et le centre $H \cap \Gamma$ de Γ est un réseau de H.

Esquissons les grandes lignes de la démonstration de ce théorème : elle consiste essentiellement à montrer que, pour tout $1 \leq i \leq l$, si C_i n'est pas homogène, la projection de Γ dans $\mathrm{GL}(V_i)$ est Zariski dense. Notons que cette projection préserve le cône C_i, mais n'est pas nécessairement discrète. Pour l'étudier, nous utiliserons une nouvelle notion.

Si C est un cône ouvert convexe saillant de V, on dit qu'un sous-groupe Γ de $\mathrm{Aut}\, C$ balaie C s'il existe un compact K de C tel que $C = \Gamma K$. Il existe des cônes balayables non divisibles : il ressort en effet de la classification de Vinberg [**47**] que, si C est homogène, C est divisible si et seulement si $\mathrm{Aut}\, C$ est unimodulaire. La démonstration du théorème 2.1 repose sur la

PROPOSITION 2.2 ([**7**, Proposition 3.2]). — *Soient C un cône convexe ouvert saillant de V et Γ un sous-groupe discret et Zariski connexe de $\mathrm{Aut}\, C$ qui balaie C et qui agit irréductiblement sur V. Alors, si C n'est pas homogène, Γ est Zariski dense dans $\mathrm{GL}(V)$.*

Détaillons un point essentiel de la démonstration.

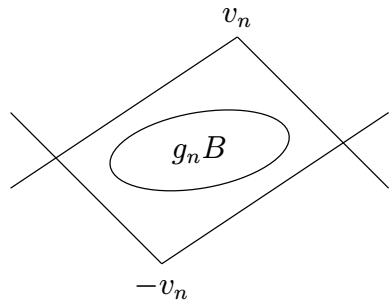


FIGURE 2. Démonstration du lemme 2.3

LEMME 2.3 ([**7**, Lemme 2.2]). — *Soient C un cône convexe ouvert saillant de V et Γ un sous-groupe discret et Zariski connexe de* $\mathrm{Aut}\, C$ *qui balaie C. Soient $\bar{C}$ l'adhérence de C dans V et $\bar{\Gamma}$ l'adhérence de Γ dans l'espace des endomorphismes de V. Alors, pour tout v dans $\bar{C}$, il existe g dans $\bar{\Gamma}$ et w dans C avec $gw = v$.*

Démonstration. — Soient (v_n) une suite d'éléments de C tendant vers v et (g_n) une suite d'éléments de Γ telle qu'il existe un compact K de C avec, pour tout n, $w_n = g_n^{-1} v_n \in K$. Nous allons montrer que (g_n) est bornée dans l'espace des endomorphismes de V, ce qui implique le résultat. Soit en effet B une boule de centre 0 pour une certaine norme sur V, de rayon suffisamment petit pour que $B + K \subset C$. Alors, pour tout n, on a $v_n + g_n B = g_n(w_n + B) \subset C$, donc $g_n B \subset -v_n + C$ et, par symétrie, $g_n B \subset v_n - C$. Le résultat en découle puisque, comme C est saillant et la suite (v_n) bornée, l'ensemble $\bigcup_{n\in\mathbb{N}}(v_n - C) \cap (-v_n + C)$ est borné. □

La stratégie de la démonstration de la proposition 2.2 consiste à montrer que la partie semi-simple de l'adhérence de Zariski G de Γ possède une orbite ouverte dans V, puis à conclure en utilisant des résultats de classification des représentations irréductibles de groupes semi-simples admettant une orbite ouverte dus à Kimura et Sato [**41**]. Commençons par reprendre des résultats de Vey sur la structure de G.

LEMME 2.4 ([**45**]). — *Soient C un cône convexe ouvert saillant de V et Γ un sous-groupe discret et Zariski connexe de* $\mathrm{Aut}\, C$ *qui balaie C et qui agit irréductiblement sur V. Alors, si G est l'adhérence de Zariski de Γ dans* $\mathrm{GL}(V)$ *et si $S = G \cap \mathrm{SL}(V)$, le groupe S est semi-simple et on a $G = \mathbb{R}^* S$.*

Démonstration. — Comme Γ agit irréductiblement sur V, son commutant dans l'algèbre des endomorphismes de V est une algèbre à division et G est réductif, donc, pour montrer que S est semi-simple, il suffit de montrer que cette algèbre à division est constituée uniquement des scalaires, c'est-à-dire que Γ ne préserve pas de structure

complexe sur V. Supposons, par l'absurde, qu'il existe une telle structure complexe J et donnons-nous un compact $K \subset C$ avec $\Gamma K = C$. Pour t suffisamment petit dans $\mathbb{R}$, on a $\exp(itJ)K \subset C$, donc

$$\exp(itJ)C = \exp(itJ)\Gamma K = \Gamma \exp(itJ)K \subset \Gamma C = C,$$

c'est-à-dire que $\exp(itJ)$ stabilise C. Comme ceci est vrai pour tout t dans un voisinage de 0, c'est vrai pour tout t dans $\mathbb{R}$ et donc, en particulier, $-C = \exp(i\pi J)C = C$, ce qui est contradictoire. Il en résulte que S est semi-simple.

À présent, pour montrer qu'on a $G = \mathbb{R}^* S$, il suffit de montrer qu'on a $\Gamma \not\subset \mathrm{SL}^{\pm}(V)$. Pour cela, rappelons que le cône dual de C est l'ensemble C^* des formes linéaires φ dans V^* telles que, pour tout x dans C, on ait $\varphi(x) > 0$. C'est un cône ouvert convexe saillant de V^* et on a $C^{**} = C$. Fixons une mesure de Lebesgue sur V^* et introduisons la fonction caractéristique ξ_C de C donnée par, pour tout x dans C,

$$\xi_C(x) = \int_{C^*} \exp(-\varphi(x))\mathrm{d}\varphi.$$

La convergence de l'intégrale découle immédiatement du caractère saillant de C^*. Pour tout g dans $\mathrm{Aut}\, C$, on a $\xi_C \circ g = |\det g|^{-1} \xi_C$. En particulier, la fonction ξ_C est homogène de degré $-\dim V$. Supposons qu'on a $\Gamma \subset \mathrm{SL}^{\pm}(V)$. Alors, ξ_C est Γ-invariante. Comme Γ balaie C, ξ_C est bornée sur C : ceci contredit le fait que ξ_C est homogène de degré $-\dim V$ et, donc, on a $\Gamma \not\subset \mathrm{SL}^{\pm}(V)$, ce qu'il fallait démontrer. □

La démonstration de la proposition 2.2 repose alors sur le

Lemme 2.5 ([**7**, Lemme 2.4]). — *Soient C un cône convexe ouvert saillant de V et Γ un sous-groupe discret et Zariski connexe de $\mathrm{Aut}\, C$ qui balaie C et qui agit irréductiblement sur V. Notons G l'adhérence de Zariski de Γ dans $\mathrm{GL}(V)$ et $S = G \cap \mathrm{SL}(V)$. Alors, si P est un polynôme homogène S-invariant non constant sur V, P est nul sur le bord ∂C de C.*

Démonstration. — Soit $d \geq 1$ le degré de P. Donnons-nous v dans ∂C. D'après le lemme 2.3, il existe g dans $\bar{\Gamma}$ et w dans C avec $gw = v$. Soit (g_n) une suite d'éléments de Γ tendant vers g. Écrivons, comme dans le lemme 2.4, $g_n = \lambda_n h_n$ avec $\lambda_n \in \mathbb{R}^*$ et $h_n \in S$. Alors on a

$$P(g_n w) = \lambda_n^d P(h_n w) = \lambda_n^d P(w).$$

Comme $P(g_n w) \xrightarrow[n\to\infty]{} P(v)$, si $P(v) \neq 0$ la suite (λ_n) ne tend pas vers 0 et, donc, g est inversible. Alors, on a $gC = C$ et donc $v \in C$, ce qui est contradictoire. Il vient bien $P(v) = 0$, ce qu'il fallait démontrer. □

Exemple 2.6. — Soit q une forme quadratique de signature $(1,n)$ sur V. L'algèbre des polynômes $\mathrm{SO}(q)$-invariants sur V est exactement $\mathbb{R}[q]$. Ces polynômes sont donc nuls sur le cône isotrope de q.

Esquissons rapidement la fin de la démonstration de la proposition 2.2.

Si S possède un polynôme homogène invariant non constant P, on montre que le cône C est homogène sous l'action de la composante neutre G° de G. En effet, d'après le lemme 2.5, si $Z = \{x \in V | P(x) = 0\}$, on a $\partial C \subset Z$ et, si Z' est l'ensemble des points réguliers de Z, pour des raisons de dimension, l'ensemble $Z' \cap \partial C$ est une réunion de composantes connexes de Z'. En particulier, cette intersection est stable par G°. Comme, par des raisonnements classiques de géométrie algébrique réelle (voir [**4**]), l'ensemble $Z - Z'$ est de codimension ≥ 2, $Z' \cap \partial C$ est dense dans ∂C et ∂C est invariant par la composante neutre de G°. Par conséquent G° préserve C et, comme Γ balaie C, G° agit co-compactement sur C. D'après le théorème d'Abels [**1**] de structure des actions propres de G°, on a $C = G/K \times_K F$ où K est un sous-groupe compact maximal de G° et F un fermé K-invariant de C et donc, comme G/K et C sont homéomorphes à des espaces euclidiens et que F est compact, F est réduit à un point.

Si S ne possède pas de polynôme invariant non constant, on montre que, d'après un théorème de Rosenlicht [**34**], ceci implique que S possède une orbite ouverte dans V et on en déduit, en utilisant la classification de ces situations par Kimura et Sato [**41**] et des résultats préalables de Benoist [**6**], que, dans ce cas, un sous-groupe Zariski dense de $\mathbb{R}^* S$ ne peut balayer un cône convexe saillant de V que si $S = \mathrm{SL}(V)$.

2.2. L'espace des représentations qui divisent un convexe

Dans ce paragraphe, nous allons donner des éléments de la démonstration du théorème 1.6. Commençons par revenir sur l'hypothèse algébrique faite sur le groupe Λ dans le théorème 1.6. Grâce au théorème 2.1, on peut préciser les situations où un sous-groupe discret de $\mathrm{PGL}(V)$ divisant un ouvert convexe saillant a un centre virtuel trivial :

Proposition 2.7 ([**10**, Corollaire 2.13]). — *Soient Ω un ouvert convexe saillant de $\mathbb{P}(V)$ et Λ un sous-groupe discret de $\mathrm{Aut}\,\Omega$ divisant Ω. Alors les assertions suivantes sont équivalentes :*

(i) *Le centre virtuel de Λ est trivial.*
(ii) *Le groupe Λ ne préserve pas de réunion finie de sous-espaces projectifs de $\mathbb{P}(V)$.*
(iii) *Le groupe Λ ne contient pas de sous-groupe distingué nilpotent infini.*

L'hypothèse du théorème 1.6 se trouve alors justifiée par le résultat suivant de Goldman et Millson :

LEMME 2.8 (Goldman & Millson [**23**, Théorème 1.1]). — *Soit Λ un groupe de type fini n'admettant pas de sous-groupe distingué nilpotent infini et soit G un groupe de Lie. Alors l'ensemble des morphismes fidèles et d'image discrète de Λ dans G est fermé dans l'ensemble des morphismes de Λ dans G.*

La démonstration repose sur le célèbre

LEMME 2.9 (Zassenhaus [**40**, Théorème 8.16]). — *Soit G un groupe de Lie. Il existe un voisinage ouvert U de l'élément neutre de G tel que tout sous-groupe discret de G engendré par des éléments de U soit nilpotent.*

Démonstration. — Cela résulte de ce que l'application $G\times G \to G, (g,h) \mapsto ghg^{-1}h^{-1}$ a une différentielle nulle en (e,e). □

Démonstration du lemme 2.8. — Soit (ρ_n) une suite de morphismes fidèles et d'image discrète de Λ dans G qui converge vers un morphisme ρ. Montrons que ρ est lui-même fidèle et d'image discrète. Pour cela, notons H la composante neutre de l'adhérence de $\rho(\Lambda)$ et $\Delta = \rho^{-1}(H)$. Il s'agit de montrer que Δ est trivial. Comme Δ est distingué, nous allons commencer par montrer qu'il est nilpotent. Remarquons que, si Δ' est un sous-groupe de type fini de Δ, il existe un entier n tel que $\rho_n(\Delta')$ soit contenu dans la composante neutre de G. En particulier, si Δ' est nilpotent, son degré de nilpotence est borné par une constante ne dépendant que de G. Pour montrer que Δ est nilpotent, il suffit donc de montrer que tous ses sous-groupes de type fini le sont. Donnons-nous un voisinage ouvert U de e comme dans le lemme 2.9. Comme H est connexe, les sous-groupes denses de H sont engendrés par leur intersection avec U et, donc, Δ est engendré par $\rho^{-1}(U)\cap\Delta$. Soient alors $\gamma_1,\dots,\gamma_r$ des éléments de $\rho^{-1}(U)\cap\Delta$. Il existe un entier n tel que, pour tout $1\leq i\leq r$, on ait $\rho_n(\gamma_i)\in U$. Alors, comme ρ_n est d'image discrète, le sous-groupe de G engendré par les $\rho_n(\gamma_i)$, $1\leq i\leq r$, est nilpotent et, comme ρ_n est fidèle, le sous-groupe de Δ engendré par les γ_i, $1\leq i\leq r$, est lui aussi nilpotent. Par conséquent, Δ est nilpotent et, donc, fini. En particulier, H est trivial. Soit p l'ordre de Δ. Il existe un voisinage V de e dans G ne contenant pas d'éléments de p-torsion différents de e. Pour n suffisamment grand, $\rho_n(\Delta)$ est inclus dans V, donc $\rho_n(\Delta)=e$ et $\Delta=e$. □

Donnons à présent quelques précisions sur la démonstration du théorème 1.6.

Pour cela, commençons par introduire une notion de théorie des groupes due à Serre [**43**]. Si Γ est un groupe de type fini sans torsion agissant proprement en préservant l'orientation sur un espace topologique X homéomorphe à un espace euclidien de dimension d, la cohomologie du quotient $\Gamma\backslash X$ est nulle en degré $>d$ et l'espace $\mathrm{H}^d(X)$ est non-trivial si et seulement si $\Gamma\backslash X$ est compact. Or, pour tout entier r, l'espace $\mathrm{H}^r(X)$ s'identifie naturellement au groupe de cohomologie $\mathrm{H}^r(\Gamma,\mathbb{R})$. Par conséquent,

si Γ possède une action propre préservant l'orientation et co-compacte sur un espace X homéomorphe à un espace euclidien, la dimension de X ne dépend que de Γ et est exactement le plus grand entier d tel que $\mathrm{H}^d(\Gamma, \mathbb{R}) \neq 0$. On appelle cet entier la dimension cohomologique de Γ et on le note $\operatorname{cd}\Gamma$. Si Γ agit proprement et co-compactement en préservant l'orientation sur un espace X homéomorphe à un espace euclidien, on a donc $\dim X \geq \operatorname{cd}\Gamma$ avec égalité si et seulement si l'action est co-compacte. Si Δ est un sous-groupe d'indice fini de Γ, on a $\operatorname{cd}\Delta = \operatorname{cd}\Gamma$. Par conséquent, si Γ est un groupe ayant des sous-groupes d'indice fini sans torsion, on appelle dimension cohomologique virtuelle de Γ et on note $\operatorname{vcd}\Gamma$ la dimension cohomologique de ces sous-groupes : elle ne dépend que de Γ. Ceci s'applique en particulier aux groupes linéaires de type fini, d'après le lemme de Selberg [**3, 42**].

Donnons un exemple d'application de cette notion dans notre contexte :

LEMME 2.10. — *Soient C un cône ouvert convexe saillant de V et Γ un sous-groupe discret de $\operatorname{Aut} C$. Alors Γ divise C si et seulement si Γ divise C^*.*

Grâce à la notion de dimension cohomologique, ce résultat repose essentiellement sur le lemme suivant dont nous avons jusqu'à présent différé la démonstration :

LEMME 2.11. — *Soit Ω un ouvert convexe saillant de $\mathbb{P}(V)$. Le groupe $\operatorname{Aut}\Omega$ agit proprement sur Ω.*

Démonstration. — Pour démontrer ce résultat, nous allons construire une application continue et $\operatorname{Aut}\Omega$-équivariante de Ω vers l'ensemble des ellipsoïdes pointés de $\mathbb{P}(V)$. Comme l'action de $\operatorname{Aut}\Omega$ sur cet ensemble s'identifie à son action sur un quotient de $\mathrm{PGL}(V)$ par un sous-groupe compact, cette action est propre et le résultat en découle.

Soit Ω^* le convexe dual de Ω, c'est-à-dire l'ensemble des droites de V^* engendrées par des formes linéaires φ dans V^* telles que, pour tout $v \neq 0$ dans V, si $\mathbb{R}v$ appartient à $\bar{\Omega}$, on ait $\varphi(v) \neq 0$. L'ensemble Ω^* est un ouvert convexe saillant de $\mathbb{P}(V^*)$ et on a $\operatorname{Aut}\Omega = \operatorname{Aut}\Omega^*$. Pour tout x dans Ω, l'ouvert Ω^* est convexe et borné dans l'ouvert affine $\mathbb{P}(V^*) - \mathbb{P}\left(x^\perp\right)$. On note $\xi_x \in \Omega^*$ le centre de gravité de Ω^* dans $\mathbb{P}(V^*) - \mathbb{P}\left(x^\perp\right)$. L'application $x \mapsto \xi_x$ est continue et $\operatorname{Aut}\Omega$-équivariante de Ω dans Ω^* (c'est même un homéomorphisme de Ω dans Ω^*, voir [**45**, § 1]).

Pour tout x dans Ω, considérons à présent Ω comme un ouvert convexe et borné de l'espace affine $\mathbb{P}(V) - \mathbb{P}\left(\xi_x^\perp\right)$. Alors, d'après [**27, 38**], il existe un unique ellipsoïde $E_x \subset \Omega$ de centre x et de volume maximal dans $\mathbb{P}(V) - \mathbb{P}\left(\xi_x^\perp\right)$. L'application $x \mapsto (x, E_x)$ est bien une application continue et $\operatorname{Aut}\Omega$-équivariante de Ω dans l'ensemble des ellipsoïdes pointés de $\mathbb{P}(V)$. □

Démonstration du lemme 2.10. — Si Γ divise C, comme, d'après le lemme 2.11, Γ agit proprement sur C, on a $\operatorname{vcd}\Gamma = \dim C = \dim V = \dim C^*$ et, donc, comme Γ agit proprement sur C^*, Γ divise C^*. La réciproque est immédiate, par dualité. □

Par ailleurs, rappelons qu'un élément g de $\mathrm{GL}(V)$ est dit proximal s'il possède une unique valeur propre complexe de module maximal et que l'espace propre associé est de dimension 1. Cette valeur propre est alors réelle. Nous dirons que g est positivement proximal si cette valeur propre est positive. On dit qu'un sous-groupe Γ de $\mathrm{GL}(V)$ est proximal s'il contient des éléments proximaux. Si Γ est proximal, on dit qu'il est positivement proximal si tous ses éléments proximaux sont positivement proximaux. On a la

PROPOSITION 2.12 ([**6**, Proposition 1.1]). — *Soit Γ un sous-groupe de $\mathrm{GL}(V)$ agissant irréductiblement sur V. Alors Γ préserve un cône convexe ouvert saillant de V si et seulement si Γ est positivement proximal.*

Donnons-nous alors un groupe Λ de centre virtuel trivial et une suite (ρ_n) de morphismes fidèles et d'image discrète de Λ dans $\mathrm{PGL}(V)$ dont les images divisent des ouverts convexes saillants de $\mathbb{P}(V)$. Supposons que la suite (ρ_n) converge vers un morphisme ρ de Λ dans $\mathrm{PGL}(V)$. Comme les ρ_n se relèvent de manière unique en des morphismes de Λ dans $\mathrm{SL}^{\pm}(V)$ qui préservent des cônes ouverts convexes saillants, on peut dorénavant supposer que les ρ_n et ρ sont à valeurs dans $\mathrm{SL}^{\pm}(V)$. En particulier, pour tout n, d'après la proposition 2.12, le groupe $\rho_n(\Lambda)$ est positivement proximal. En revanche, la représentation ρ n'est pas a priori proximale. Décrivons cependant une propriété qu'elle hérite du caractère positivement proximal des ρ_n. Nous dirons qu'un élément g de $\mathrm{GL}(V)$ est semi-proximal s'il admet une valeur propre réelle dont le module est égal à son rayon spectral et qu'il est positivement semi-proximal si cette valeur propre peut être choisie positive.

L'argument du théorème de Perron-Frobenius permet de démontrer le

LEMME 2.13 ([**10**, Lemme 3.2]). — *Soit C un cône ouvert convexe saillant de V. Les éléments de $\operatorname{Aut} C$ sont positivement semi-proximaux.*

Remarque 2.14. — Ces éléments peuvent parfois ne pas être proximaux : c'est le cas par exemple lorsque C est le cône du futur d'une forme quadratique q de signature $(1, \dim V - 1)$. En effet, si $\dim V \geq 3$, le groupe $\mathrm{SO}(q)$ contient des éléments unipotents.

Par conséquent, d'après le lemme 2.13, pour tout n, les éléments de $\rho_n(\Lambda)$ sont tous positivement semi-proximaux, donc ceux de $\rho(\Lambda)$ aussi.

Notons (ρ', V') la somme des sous-quotients irréductibles de (ρ, V). Le noyau de ρ' est fini. Cela vient de ce que, d'après le lemme 2.8, la représentation ρ est fidèle et discrète, et, d'après la proposition 2.7, les sous-groupes distingués nilpotents de Λ

sont finis. Par conséquent, on a $\operatorname{vcd}\rho'(\Lambda) = \operatorname{vcd}\Lambda = \dim V - 1 = \dim V' - 1$. Le point essentiel de la démonstration du théorème 1.6 consistera à montrer que $\rho'(\Lambda)$ préserve un cône ouvert convexe saillant de V'. En effet, comme $\operatorname{vcd}\rho'(\Lambda) = \dim V' - 1$, $\rho'(\Lambda)$ divise la trace projective de ce cône et, donc, comme les sous-groupes distingués nilpotents de $\rho'(\Lambda)$ sont finis, d'après la proposition 2.7, la représentation (ρ', V') est irréductible, si bien que $\rho = \rho'$ et $\rho(\Lambda)$ divise un ouvert convexe saillant de $\mathbb{P}(V)$.

En d'autres termes, le théorème 1.6 découle de la

PROPOSITION 2.15 ([**10**, Proposition 3.4]). — *Soit* Γ *un sous-groupe discret de type fini de* $\mathrm{GL}(V)$ *agissant de façon semi-simple sur* V*. On suppose que tous les éléments de* Γ *sont positivement semi-proximaux. Alors, on a* $\operatorname{vcd}(\Gamma) \leq \dim V$ *avec égalité si et seulement si* Γ *divise un cône ouvert convexe saillant de* V*.*

Nous allons à présent donner les grandes lignes de la démonstration de la proposition 2.15, qui utilise le langage général de la théorie des groupes algébriques réductifs. Pour cela, nous allons introduire le groupe algébrique qui est l'adhérence de Zariski de Γ dans $\mathrm{GL}(V)$. Quitte à remplacer Γ par un sous-groupe d'indice fini (qui agit encore de façon irréductible sur V, puisque Γ est totalement irréductible), on peut supposer que ce groupe est Zariski connexe ; comme Γ est irréductible, son adhérence de Zariski est réductive.

Soit donc $\mathbf{G}$ un groupe algébrique réel réductif. Soient $\mathbf{A}$ un tore $\mathbb{R}$-déployé maximal de $\mathbf{G}$, $\mathbf{Z}$ le centralisateur de $\mathbf{A}$ dans $\mathbf{G}$ et $\mathbf{M}$ le plus grand sous-groupe $\mathbb{R}$-anisotrope de $\mathbf{Z}$ de sorte que $\mathbf{Z} = \mathbf{MA}$. On note G (resp. A, resp. Z, resp. M) le groupe des points réels de $\mathbf{G}$ (resp. $\mathbf{A}$, resp. $\mathbf{Z}$, resp. $\mathbf{M}$). Le groupe M est compact, le groupe A est constitué de matrices simultanément diagonalisables sur $\mathbb{R}$ et on a $Z = MA$. Notons A^* l'ensemble des éléments de A dont le centralisateur dans $\mathbf{G}$ est exactement $\mathbf{Z}$: c'est le complémentaire dans A d'une réunion finie de sous-groupes de codimension 1. Un élément g de G est dit loxodromique s'il est conjugué à un élément de MA^*. Soit Γ un sous-groupe Zariski dense de G. D'après [**39**], Γ contient des éléments loxodromiques. On dit qu'une représentation rationnelle (ρ, V) de $\mathbf{G}$ est Γ-semi-proximale (resp. positivement Γ-semi-proximale) si tout élément loxodromique de Γ a une image semi-proximale (resp. positivement semi-proximale) par ρ. D'après [**10**, 5.1], une représentation irréductible Γ-semi-proximale est G-semi-proximale.

Exemple 2.16. — D'après [**28**], on peut construire, par déformation d'une représentation du groupe fondamental d'une surface, un sous-groupe discret et Zariski dense Γ de $\mathrm{SL}_3(\mathbb{R})$ divisant un cône convexe de $\mathbb{R}^3$. D'après le lemme 2.13, la représentation naturelle de $\mathrm{SL}_3(\mathbb{R})$ dans $\mathbb{R}^3$ est positivement Γ-semi-proximale, mais elle n'est pas positivement $\mathrm{SL}_3(\mathbb{R})$-semi-proximale.

La proposition 2.15 est alors une conséquence directe de la proposition suivante, qui est la formulation originale de [**10**, Proposition 3.4] :

PROPOSITION 2.17 ([**10**, Proposition 3.4]). — *Soient* Γ *un sous-groupe discret de type fini et Zariski dense de* G *et* (ρ, V) *une représentation rationnelle de* $\mathbf{G}$ *de noyau fini. On suppose que* (ρ, V) *est positivement* Γ*-semi-proximale. Alors, on a* $\operatorname{vcd}\Gamma \leq \dim V$ *avec égalité si et seulement si* Γ *divise un cône ouvert convexe saillant de* V*.*

Donnons une idée des méthodes employées dans la démonstration de la proposition 2.17. Fixons un sous-groupe compact maximal K de G. L'espace G/K est difféomorphe à un espace euclidien, de sorte que, si Γ est un sous-groupe discret de type fini de G, on a $\operatorname{vcd}\Gamma \leq \dim G/K$. Un examen précis de la liste des algèbres de Lie simples et de leurs représentations de petite dimension [**26**] permet de démontrer le

LEMME 2.18 ([**10**, Corollaire 5.10]). — *Supposons* $\mathbf{G}$ *semi-simple, connexe et simplement connexe. Soit* (ρ, V) *une représentation rationnelle de* $\mathbf{G}$ *de noyau fini et positivement* G*-semi-proximale. On a* $\dim G/K \leq \dim V$ *et, si* $\dim G/K = \dim V$*, la représentation* ρ *est positivement* G*-proximale et* G *agit proprement et transitivement sur un ouvert convexe saillant* G*-invariant de* $\mathbb{P}(V)$*.*

Le second point-clef de la démonstration de la proposition 2.17 s'obtient par une étude des caractères de $\mathbf{A}$ qui sont le plus haut poids restreint d'une représentation semi-proximale de $\mathbf{G}$:

LEMME 2.19 ([**10**, Lemme 5.8]). — *Supposons* $\mathbf{G}$ *semi-simple, connexe et simplement connexe. Soit* (ρ, V) *une représentation rationnelle de* $\mathbf{G}$ *irréductible et semi-proximale. Alors, si* ρ *n'est pas proximale, il existe une représentation rationnelle irréductible proximale* (ρ', V') *de* G *avec* $\dim V' < \dim V$ *et telle que, pour tout sous-groupe Zariski dense* Γ *de* G*,* (ρ', V') *soit positivement* Γ*-semi-proximale si et seulement si* (ρ, V) *l'est.*

Une des difficultés réside dans le fait que la représentation ρ' peut éventuellement être la représentation triviale ; un exemple de cette situation est donné dans [**10**, 5.6].

Pour conclure nous allons utiliser ces lemmes pour la

Démonstration de la proposition 2.17. — Établissons ce résultat quand ρ est irréductible et que le groupe dérivé de $\mathbf{G}$ est simple. Alors, d'après le lemme de Schur, l'image par ρ de la partie centrale de $\mathbf{A}$ est scalaire et, donc, comme ρ est de noyau fini, cette partie centrale est de dimension ≤ 1. Quitte à remplacer $\mathbf{G}$ par un revêtement fini, on peut supposer qu'on a $\mathbf{G} = \mathbf{S} \times \mathbf{T}$ où $\mathbf{S}$ est semi-simple, connexe et simplement

connexe et où $\mathbf{T}$ est un tore dont la composante $\mathbb{R}$-déployée est de dimension ≤ 1. On note S le groupe des points réels de $\mathbf{S}$.

Supposons que la représentation ρ est positivement G-semi-proximale. D'après le lemme 2.18, appliqué au groupe $\mathbf{S}$, on a

$$\operatorname{vcd}\Gamma \leq \dim G/K \leq \dim(S/(K\cap S)) + 1 \leq \dim V.$$

En cas d'égalité, d'une part, comme $\operatorname{vcd}\Gamma = \dim G/K$, l'espace $\Gamma\backslash G/K$ est compact et Γ est donc un réseau co-compact de G et, d'autre part, comme $\dim G/K = \dim V$, d'après le lemme 2.18, G agit transitivement sur un cône ouvert convexe saillant de V, si bien que Γ divise ce cône.

Supposons que la représentation ρ n'est ni positivement G-semi-proximale, ni G-proximale et donnons-nous (ρ', V') comme dans le lemme 2.19 appliqué au groupe $\mathbf{S}$. Étendons ρ' en une représentation de $\mathbf{G}$ à travers un caractère non-trivial de $\mathbf{T}$. Alors, ρ' est positivement Γ-proximale, mais n'est pas positivement G-proximale. En particulier, la restriction de ρ' à $\mathbf{S}$ n'est pas triviale et, donc, le noyau de ρ' est fini. Alors, d'après la proposition 2.12, le groupe $\rho'(\Gamma)$ préserve un cône ouvert convexe saillant de V' et on a donc

$$\operatorname{vcd}\Gamma \leq \dim V' < \dim V.$$

Enfin, supposons que ρ est G-proximale. Alors, d'après la proposition 2.12, le groupe $\rho(\Gamma)$ préserve un cône ouvert convexe saillant de V et on a

$$\operatorname{vcd}\Gamma \leq \dim V,$$

avec égalité si et seulement si Γ divise ce cône.

La démonstration du cas général repose sur des arguments analogues et sur une étude approfondie des valeurs propres des éléments des sous-groupes Zariski denses des groupes semi-simples. □

3. CONVEXES DIVISIBLES ET HYPERBOLICITÉ

3.1. Convexes divisibles strictement convexes

Nous nous intéressons à présent à la démonstration du théorème 1.7. Pour cela, commençons par établir la

PROPOSITION 3.1 ([**9**, Proposition 2.5]). — *Soient Ω un ouvert convexe saillant divisible de $\mathbb{P}(V)$ et d_Ω sa distance de Hilbert. L'espace métrique (Ω, d_Ω) est hyperbolique si et seulement si Ω est strictement convexe.*

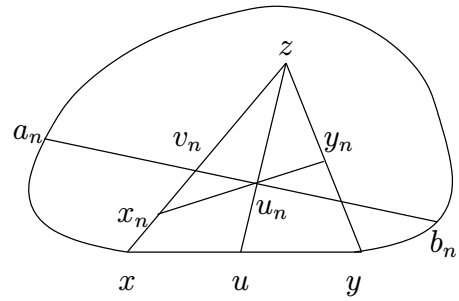


FIGURE 3. Un convexe hyperbolique est strictement convexe

Démonstration. — Supposons tout d'abord que Ω n'est pas strictement convexe et montrons que sa distance de Hilbert n'est pas hyperbolique. Cette partie ne nécessite pas que Ω soit divisible. Donnons-nous donc un segment maximal $[x, y]$ contenu dans le bord de Ω avec $x \neq y$, un point u de $]x, y[$ et un point z de Ω. Choisissons des suites (x_n), (y_n) et (u_n) de points de $]x, z]$, $]y, z]$ et $]u, z]$ qui convergent respectivement vers x, y et u. Supposons qu'il existe $\delta > 0$ tel que, pour tout n, il existe v_n dans $[x_n, z] \cup [y_n, z]$ avec $d_\Omega(u_n, v_n) \leq \delta$ et, quitte à extraire une sous-suite et à échanger les rôles de x et de y, supposons que v_n appartient à $]x, z]$. Alors, comme la distance d_Ω est propre, la suite (v_n) tend vers x. Soient a_n et b_n les points d'intersection de la droite projective engendrée par u_n et v_n avec le bord de Ω, de façon à ce que v_n appartienne au segment $[a_n, u_n]$ et que u_n appartienne au segment $[b_n, v_n]$. Cette situation est représentée par la figure 3. Comme le segment $[u_n, v_n]$ tend vers $[u, x]$ et que le segment $[x, y]$ est maximal dans le bord de Ω, la suite (a_n) tend vers x. De même, quitte à extraire une sous-suite, on peut supposer que la suite (b_n) tend vers un point b de $[u, y]$. Plongeons $\bar{\Omega}$ dans un ouvert affine et choisissons une forme affine φ qui ne soit pas constante sur $[x, y]$. Alors, pour n suffisamment grand, on a

$$d_\Omega(u_n, v_n) = \frac{1}{2} \left| \log \left(\frac{\varphi(u_n) - \varphi(b_n)}{\varphi(u_n) - \varphi(a_n)} \frac{\varphi(v_n) - \varphi(a_n)}{\varphi(v_n) - \varphi(b_n)} \right) \right|$$

et, donc, comme (a_n) et (v_n) tendent vers x tandis que (u_n) tend vers u et que (b_n) tend vers b, $d_\Omega(u_n, v_n) \xrightarrow[n\to\infty]{} \infty$, ce qui est contradictoire.

Réciproquement, supposons que Ω est strictement convexe, mais que sa distance de Hilbert n'est pas hyperbolique. Donnons-nous alors des suites (x_n), (y_n), (z_n) de points de Ω avec $u_n \in [x_n, y_n]$ et $d_\Omega(u_n, [x_n, z_n] \cup [y_n, z_n]) \xrightarrow[n\to\infty]{} \infty$. Comme Ω est divisible, quitte à remplacer, pour tout n, u_n par une de ses images par un élément de $\operatorname{Aut}\Omega$ et à extraire une sous-suite, on peut supposer que (u_n) converge vers un point u de Ω et que (x_n), (y_n) et (z_n) convergent vers des points x, y et z de $\bar{\Omega}$. Alors, si on avait $x \neq z$, comme Ω est strictement convexe, on aurait $]x, z[\subset \Omega$ et

$d_\Omega(u,]x,z[) = \infty$. Par conséquent, on a $x = z$. De même, on a $y = z$ et, donc, comme $u \in [x,y]$, $u = x = y$, ce qui est impossible. □

Rappelons qu'on a noté Ω^* le convexe dual de Ω. L'ensemble Ω^* est un ouvert convexe saillant de $\mathbb{P}(V^*)$ et le bord de Ω est de classe $\mathscr{C}^1$ si et seulement si Ω^* est strictement convexe. La démonstration du théorème 1.7 s'achève avec la

PROPOSITION 3.2 ([**9**, Proposition 2.7]). — *Soit Ω un ouvert convexe saillant divisible de $\mathbb{P}(V)$. L'ouvert Ω est strictement convexe si et seulement si Ω^* est strictement convexe.*

Démonstration. — Par dualité, il suffit de montrer une seule des implications. Soit Λ un sous-groupe discret sans torsion de $\operatorname{Aut}\Omega$ divisant Ω et préservant l'orientation. Alors, la dimension cohomologique de Λ est $\dim V - 1$ et, donc, la représentation adjointe de Λ divise Ω^*. Si Ω est strictement convexe, d'après la proposition 3.1, l'espace métrique (Ω, d_Ω) est hyperbolique, donc le groupe Λ est hyperbolique, donc l'espace métrique (Ω^*, d_{Ω^*}) est hyperbolique et, à nouveau d'après la proposition 3.1, le convexe Ω^* est strictement convexe. □

Pour terminer ce paragraphe, donnons des informations sur la régularité du bord des convexes divisibles strictement convexes. Munissons une fois pour toutes V d'une norme et $\mathbb{P}(V)$ de la distance associée d. Si M est une sous-variété de classe $\mathscr{C}^1$ de $\mathbb{P}(V)$ et x un point de M, notons $\mathrm{T}^{\mathbb{P}}_x M$ la sous-variété projective tangente en x à M. Si α est un élément de $]1,2[$, nous dirons que M est de classe $\mathscr{C}^\alpha$ si, pour tout compact K de M, il existe un réel $C \geq 0$ tel que, pour tous x et y dans K, on ait

$$d(y, \mathrm{T}^{\mathbb{P}}_x M) \leq C d(x,y)^\alpha$$

(cette notion est équivalente à la notion usuelle de régularité $\mathscr{C}^\alpha$, en vertu de propriétés classiques des fonctions höldériennes, voir [**9**, 4.2] ou [**35**]). Si β est un élément de $]2,\infty[$, nous dirons que M est β-convexe si, pour tout compact K de M, il existe un réel $C > 0$ tel que, pour tous x et y dans K, on ait

$$d(y, \mathrm{T}^{\mathbb{P}}_x M) \geq C d(x,y)^\beta.$$

Si Ω est un ouvert convexe saillant de $\mathbb{P}(V)$ et si $\frac{1}{\alpha} + \frac{1}{\beta} = 1$, on montre aisément que Ω est de classe $\mathscr{C}^\alpha$ si et seulement si Ω^* est β-convexe (voir [**9**, 4.5] et [**24**, 3.2]). Dans ce cas, on pose

$$\alpha_\Omega = \sup\{\alpha \in]1,2[\,|\, \partial\Omega \text{ est de classe } \mathscr{C}^\alpha\} \text{ et } \beta_\Omega = \inf\{\beta \in]2,\infty[\,|\, \partial\Omega \text{ est } \beta\text{-convexe}\},$$

de sorte que $\frac{1}{\alpha_\Omega} + \frac{1}{\beta_{\Omega^*}} = 1$.

Introduisons des objets reliés à ces quantités pour les éléments de $\mathrm{PGL}(V)$. Soit d la dimension de V. Si g est un élément de $\mathrm{GL}(V)$, on note $\lambda_1(g), \ldots, \lambda_d(g)$ les valeurs propres de g, comptées avec multiplicité et rangées de façon à ce que leurs modules

décroissent et $\ell_1(g) \geq \ldots \geq \ell_d(g)$ les logarithmes de ces modules. Ainsi, g est proximal si et seulement si on a $\ell_1(g) > \ell_2(g)$. On dit que g est bi-proximal si g et son inverse sont proximaux ou encore si on a $\ell_1(g) > \ell_2(g)$ et $\ell_{d-1}(g) > \ell_d(g)$. Dans ce cas, on pose

$$\alpha_g = \frac{\ell_1(g) - \ell_d(g)}{\ell_1(g) - \ell_{d-1}(g)} \text{ et } \beta_g = \frac{\ell_1(g) - \ell_d(g)}{\ell_1(g) - \ell_2(g)},$$

si bien que $\frac{1}{\alpha_g} + \frac{1}{\beta_{g^{-1}}} = 1$ et $\min(\alpha_g, \alpha_{g^{-1}}) \leq 2$. Les nombres α_g et β_g ne dépendent que de l'image de g dans $\mathrm{PGL}(V)$.

Dans le cas des convexes divisibles strictement convexes, nous pouvons améliorer le lemme 2.13 :

LEMME 3.3 ([**9**, Proposition 5.1]). — *Soient Ω un ouvert strictement convexe de $\mathbb{P}(V)$. Si Λ est un sous-groupe discret sans torsion de $\mathrm{Aut}\,\Omega$ qui divise Ω, les relevés dans $\mathrm{GL}(V)$ des éléments non triviaux de Λ sont proximaux.*

Dans ce cas, on pose $\alpha_\Lambda = \inf_{g \in \Lambda - \{e\}} \alpha_g \leq 2$ et $\beta_\Lambda = \sup_{g \in \Lambda - \{e\}} \beta_g \geq 2$. On a le

THÉORÈME 3.4 (Benoist [**9**, Corollaire 5.3], Guichard [**24**])

Soit Ω un ouvert strictement convexe divisible de $\mathbb{P}(V)$. Il existe $1 < \alpha < 2$ tel que $\partial\Omega$ soit de classe $\mathscr{C}^\alpha$ et on a $\alpha_\Omega = \alpha_\Lambda = \alpha_{\Omega^*}$ et $\beta_\Omega = \beta_\Lambda = \beta_{\Omega^*}$.

L'existence de $\alpha > 1$ tel que le bord de Ω soit $\mathscr{C}^\alpha$ découle d'une propriété des flots d'Anosov. En effet, on peut définir le flot géodésique de Ω sur le fibré en sphères $\mathrm{S}\Omega$ de la façon suivante : si x est un élément de Ω et v une direction tangente à $\mathbb{P}(V)$ en x, on lui associe l'unique segment de droite projective $[x, y_{x,v}[\subset \Omega$ où $y_{x,v}$ est le point de $\partial\Omega$ tel que, dans un ouvert affine contenant Ω, le vecteur $y_{x,v} - x$ soit dans la direction v ; alors, le flot géodésique au temps $t \geq 0$ envoie (x, v) sur l'unique point (z, w) de $\mathrm{S}\Omega$ tel que z appartienne à $[x, y_{x,v}[$, que $d_\Omega(x, z) = t$ et que $y_{x,v} = y_{z,w}$. On définit de même le flot géodésique en temps négatif. La régularité de ce flot est donc fortement liée à la régularité de $\partial\Omega$. Si g est un élément de $\mathrm{Aut}\,\Omega$, ce flot commute à l'action naturelle de g sur $\mathrm{S}\Omega$. On a la

PROPOSITION 3.5 ([**9**, Proposition 3.3]). — *Soient Ω un ouvert strictement convexe de $\mathbb{P}(V)$ et Λ un sous-groupe discret sans torsion de $\mathrm{Aut}\,\Omega$ qui divise Ω. Le flot géodésique de $\Lambda\backslash\mathrm{S}\Omega$ est un flot d'Anosov.*

Cette proposition entraîne alors (voir [**9**, Proposition 4.6]) que le bord de Ω est de classe $\mathscr{C}^\alpha$, pour un certain α. Ce résultat est à rapprocher des propriétés classiques de régularité höldérienne associées aux flots d'Anosov (voir [**25**, 19.1]).

Expliquons à présent une partie de la démonstration du théorème 3.4. Il s'agit de la

PROPOSITION 3.6 ([**9**, Corollaire 5.3]). — *Soit Ω un ouvert strictement convexe de $\mathbb{P}(V)$. Si Λ est un sous-groupe discret sans torsion de $\operatorname{Aut}\Omega$ qui divise Ω, on a $1 < \alpha_\Omega \leq \alpha_\Lambda \leq 2 \leq \beta_\Lambda \leq \beta_\Omega < \infty$.*

La réciproque de cette proposition est due à Guichard [**24**].

Démonstration. — Commençons par montrer qu'on a $\beta_\Lambda \leq \beta_\Omega$. Soit g dans $\Lambda - \{e\}$. Il s'agit de montrer qu'on a $\beta_g \leq \beta_\Omega$. Pour cela, notons x_g^+ la droite propre de valeur propre $\lambda_1(g)$ de g dans V. Pour tout x dans $\mathbb{P}(V)$, si x n'appartient pas à l'image projective de l'unique hyperplan H_g^- de V supplémentaire de x_g^+ et stable par g, on a $g^n x \xrightarrow[n\to\infty]{} x_g^+$ et, donc, x_g^+ appartient à $\bar{\Omega}$. Comme l'action de Λ sur Ω est propre, x_g^+ appartient à $\partial\Omega$. Par ailleurs, l'hyperplan H_g^- est le point fixe attracteur de l'action de g^* dans $\mathbb{P}(V^*)$. De même, si x_g^- est la droite propre de valeur propre $\lambda_d(g)$ dans V et H_g^+ son unique supplémentaire g-stable, alors H_g^+ est le point fixe attracteur de l'action de $(g^{-1})^*$ dans $\mathbb{P}(V^*)$, si bien que H_g^+ appartient à $\partial\Omega^*$, c'est-à-dire que l'hyperplan projectif $\mathbb{P}\left(H_g^+\right)$ est tangent à Ω. Comme cet hyperplan contient x_g^+ et que Ω est strictement convexe, cet hyperplan est exactement l'hyperplan tangent à $\partial\Omega$ en x_g^+.

Considérons dorénavant l'ouvert affine $W = \mathbb{P}(V) - \mathbb{P}\left(H_g^-\right)$ comme un espace vectoriel d'élément neutre x_g^+. Alors, l'espace $W_g^+ = \mathbb{P}\left(H_g^+\right) \cap W$ est un hyperplan de W et l'intersection D_g^- de W avec la droite projective engendrée par x_g^+ et par x_g^- est une droite vectorielle de W. Dans une base de W associée à la somme directe $W = W_g^+ \oplus D_g^-$, l'action de g se lit comme celle d'une matrice

$$A(g) = \frac{1}{\lambda_1(g)} \begin{pmatrix} B(g) & 0 \\ 0 & \lambda_d(g) \end{pmatrix}$$

où $B(g)$ est une matrice de rayon spectral $e^{\ell_2(g)}$. En particulier, pour tout x dans W, on a

$$d(g^n x, W_g^+) = O\left(e^{-n(\ell_1(g)-\ell_d(g))}\right),$$

tandis que, si x n'appartient pas au supplémentaire g-invariant dans W du sous-espace de W où toutes les valeurs propres de $B(g)$ sont de module $e^{\ell_2(g)-\ell_1(g)}$,

$$e^{-n(\ell_1(g)-\ell_2(g))} = O(d(g^n x, x_g^+)).$$

Comme $\partial\Omega$ contient des points x satisfaisant à ces hypothèses, si $\beta \geq 2$ est tel que $\partial\Omega$ soit β-convexe, on a

$$e^{-n\beta(\ell_1(g)-\ell_2(g))} = O\left(e^{-n(\ell_1(g)-\ell_d(g))}\right),$$

c'est-à-dire $\beta \geq \beta_g$, ce qu'il fallait démontrer.

Alors, comme Λ divise Ω^*, on a aussi $\beta_\Lambda \leq \beta_{\Omega^*}$ et, donc, par dualité, $\alpha_\Omega \leq \alpha_\Lambda$. □

COROLLAIRE 3.7 ([**9**, Proposition 6.1]). — *Soit Ω un ouvert strictement convexe divisible de $\mathbb{P}(V)$. Si $\alpha_\Omega = 2$, Ω est un ellipsoïde.*

Autrement dit, la régularité des ouverts strictement convexes divisibles qui ne sont pas homogènes ne peut pas être trop grande.

Démonstration. — Soit Λ un sous-groupe discret sans torsion de $\operatorname{Aut}\Omega$ qui divise Ω. D'après le théorème 2.1 et la classification de Vinberg des cônes convexes homogènes [**47**], si Ω n'est pas un ellipsoïde, Λ est Zariski dense dans $\mathrm{PGL}(V)$. Ce fait peut d'ailleurs se montrer directement (voir [**6**, 3.6]). Supposons donc que Ω n'est pas un ellipsoïde et que $\alpha_\Omega = 2$. Alors, d'après la proposition 3.6, pour tout g dans G, on a $\alpha_g = 2$. Or, comme Λ est Zariski dense, d'après [**5**, 1.2], si ℓ_Λ est le cône fermé engendré par l'image de l'ensemble $\{(\ell_1(g), \dots, \ell_d(g)) | g \in \Lambda\}$ dans le quotient de $\mathbb{R}^d$ par la droite engendré par le vecteur $(1, 1, \dots, 1)$, le cône ℓ_Λ est convexe et d'intérieur non vide, ce qui contredit le fait que $\alpha_g = 2$ pour tout g dans Λ, d'où le résultat. □

3.2. Convexes hyperboliques

Dans ce paragraphe, nous allons donner quelques indications sur les grandes étapes de la démonstration du théorème 1.8. Pour cela, notons X l'ensemble des ouverts convexes saillants de $\mathbb{P}(V)$, muni de l'action naturelle de $\mathrm{PGL}(V)$. On munit $\mathbb{P}(V)$ de la distance provenant d'une norme sur V et X de la distance de Hausdorff associée. L'espace X est alors localement compact. Pour tout $\delta > 0$, on note X^δ l'ensemble des Ω dans X tels que (Ω, d_Ω) soit δ-hyperbolique. Nous avons la

PROPOSITION 3.8 ([**8**, Proposition 2.10 et Corollaire 2.12]). — *L'ensemble X^δ est un fermé $\mathrm{PGL}(V)$-invariant de X dont tous les éléments sont strictement convexes et à bord $\mathcal{C}^1$. Réciproquement, si F est un fermé $\mathrm{PGL}(V)$-invariant de X dont tous les éléments sont strictement convexes (resp. à bord $\mathcal{C}^1$), il existe $\delta > 0$ tel que $F \subset X^\delta$.*

Le fait qu'il soit équivalent, pour un fermé $\mathrm{PGL}(V)$-invariant de X, d'avoir tous ses éléments strictement convexes ou tous ses éléments de classe $\mathcal{C}^1$, découle du résultat suivant de Benzécri :

LEMME 3.9 (Benzécri [**13**]). — *Soit Ω dans X. Si le bord de Ω n'est pas $\mathcal{C}^1$, l'adhérence de $\mathrm{PGL}(V)\Omega$ dans X contient un élément qui n'est pas strictement convexe. Si Ω n'est pas strictement convexe, l'adhérence de $\mathrm{PGL}(V)\Omega$ dans X contient un élément dont le bord n'est pas $\mathcal{C}^1$.*

Démonstration. — Il suffit de montrer la première de ces deux assertions, la seconde s'en déduisant par dualité. Nous démontrons le résultat quand $\dim V = 3$, le cas général s'en déduisant par des arguments généraux (voir [**8**, 2.7] ou [**13**, V.3]).

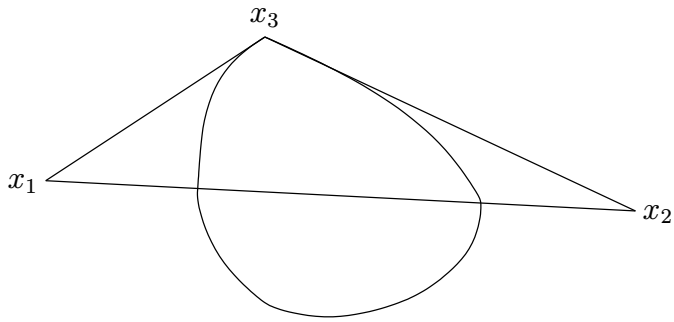


FIGURE 4. Un triangle dans une adhérence d'orbite

Soit donc Ω un ouvert convexe saillant de $\mathbb{P}(V)$ dont le bord n'est pas $\mathscr{C}^1$ et soient H_1 et H_2 des plans distincts de V tels que les droites projectives $\mathbb{P}(H_1)$ et $\mathbb{P}(H_2)$ soient toutes deux tangentes à Ω en un point x_3 de Ω. Choisissons des points $x_1 \in \mathbb{P}(H_1)$ et $x_2 \in \mathbb{P}(H_2)$ différents de x_3 et un élément g de $\mathrm{GL}(V)$ admettant x_3 pour droite propre de valeur propre 1 et x_1 et x_2 pour droites propres de valeur propre $\lambda > 1$. On vérifie aisément que la suite $(g^n\Omega)$ converge vers un triangle de sommets x_1, x_2 et x_3 (voir figure 4). Le résultat en découle. □

Introduisons un nouvel espace. On note X^* l'ensemble des couples (Ω, x) où Ω appartient à X et x appartient à Ω. La démonstration de la proposition 3.8 s'établit de manière analogue à celle de la proposition 3.1 en utilisant le

LEMME 3.10 (Benzécri [**13**, V.2], [**8**, Proposition 2.3]). — *L'action naturelle de* $\mathrm{PGL}(V)$ *sur* X^* *est propre et co-compacte.*

Démonstration. — La démonstration du fait que cette action est propre est analogue à celle du lemme 2.11. Le fait qu'elle est co-compacte provient alors de ce que, pour tous $\Omega_1 \subset \Omega_2$ dans X, l'ensemble $\{\Omega \in X | \Omega_1 \subset \Omega \subset \Omega_2\}$ est compact dans X. □

Démonstration de la proposition 3.8. — D'après le lemme 3.9, il suffit de montrer que, pour tout $\delta > 0$, l'ensemble X^δ est fermé et qu'une partie fermée et $\mathrm{PGL}(V)$-invariante de X ne contient que des éléments strictement convexes si et seulement s'il existe $\delta > 0$ avec $F \subset X^\delta$. Grâce au lemme 3.10, la démonstration de ces faits est analogue à celle de la proposition 3.1, en remplaçant l'ouvert Ω par une suite d'ouverts. □

Nous allons à présent utiliser la proposition 3.8 pour caractériser l'hyperbolicité d'un convexe en termes d'objets vivant sur le bord de ce convexe. Si Ω est un ouvert strictement convexe à bord $\mathscr{C}^1$ et a, b, c, d sont des points de $\partial\Omega$, on dit que le quadruplet (a, b, c, d) est harmonique si les droites projectives (ac) et (bd) se coupent en

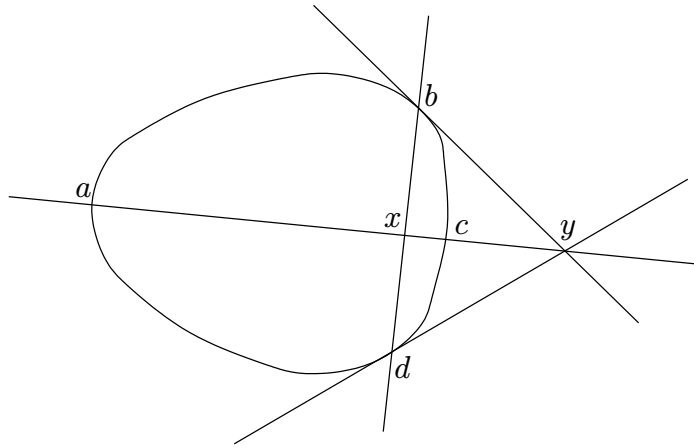


FIGURE 5. Un quadruplet harmonique

un point x de Ω et si la droite (ac) rencontre les hyperplans projectifs tangents en b et d à Ω en un même point y (voir figure 5). Dans ce cas, on pose

$$\psi_{\partial\Omega}(a,b,c,d) = -[x,y,a,c] = -\frac{c-x}{a-x}\frac{a-y}{c-y} > 0.$$

L'ensemble des quadruplets harmoniques de $\partial\Omega$ est noté $H(\partial\Omega)$ et on pose

$$\Delta_\Omega = \sup_{\xi\in H(\partial\Omega)} \psi_{\partial\Omega}(\xi).$$

Comme, si (a,b,c,d) est harmonique, (c,b,a,d) l'est aussi et

$$\psi_{\partial\Omega}(c,b,a,d) = \psi_{\partial\Omega}(a,b,c,d)^{-1},$$

on a $\Delta_\Omega \geq 1$.

PROPOSITION 3.11 ([**8**, Proposition 3.2]). — *Pour tout $\delta > 0$, il existe $\Delta \geq 1$ tel que, pour tout Ω dans X, si (Ω, d_Ω) est δ-hyperbolique, on ait $\Delta_\Omega \leq \Delta$. Pour tout $\Delta \geq 1$, il existe $\delta > 0$ tel que, pour tout Ω dans X, si $\Delta_\Omega \leq \Delta$, (Ω, d_Ω) soit δ-hyperbolique.*

En particulier, (Ω, d_Ω) est hyperbolique si et seulement si $\Delta_\Omega < \infty$.

Démonstration. — Montrons la première assertion. Supposons au contraire qu'il existe une suite (Ω_n) dans X^δ avec $\Delta_{\Omega_n} \xrightarrow[n\to\infty]{} \infty$. Choisissons alors, pour tout n, un quadruplet harmonique (a_n, b_n, c_n, d_n) dans $H(\partial\Omega_n)$ de façon à ce que $\psi_{\partial\Omega_n}(a_n, b_n, c_n, d_n) \xrightarrow[n\to\infty]{} \infty$ et notons x_n et y_n comme dans la définition. Alors, comme l'action de $\mathrm{PGL}(V)$ sur $\mathbb{P}(V)$ préserve les bi-rapports des points alignés, d'après le lemme 3.10, quitte à extraire une sous-suite, on peut supposer que la suite

(Ω_n) converge vers un élément Ω de X et que (x_n) converge vers un élément x de Ω. D'après la proposition 3.8, Ω est strictement convexe et à bord $\mathscr{C}^1$. Quitte à encore extraire une sous-suite, on peut aussi supposer que (a_n), (b_n), (c_n) et (d_n) convergent vers des points a, b, c et d de $\partial\Omega$ et que (y_n) converge vers un point y de $\mathbb{P}(V) - \Omega$. Comme x appartient à $\Omega \cap (ac) \cap (bd)$, on a $a \neq c$ et $b \neq d$ et, donc, comme y est dans l'intersection des hyperplans tangents à Ω en b et d et que Ω est strictement convexe, on a $y \notin \partial\Omega$. Alors, le quadruplet (a, b, c, d) est harmonique et on a $\psi_{\partial\Omega}(a, b, c, d) = \lim_{n\to\infty} \psi_{\partial\Omega_n}(a_n, b_n, c_n, d_n) = \infty$, ce qui est absurde.

La seconde assertion se démontre de manière analogue : on pose

$$E = \{\Omega \in X | \Omega \text{ est strictement convexe et à bord } \mathscr{C}^1 \text{ et } \Delta_\Omega \leq \Delta\}.$$

Alors, d'après la proposition 3.8, il s'agit de montrer que tout élément de l'adhérence de E dans X est strictement convexe. Ceci se montre par l'absurde, en utilisant le fait, découlant de la démonstration du lemme 3.9, que, si ce n'était pas le cas, il existerait des éléments de l'adhérence de E dont une section plane serait un triangle. □

La fin de la démonstration du théorème 1.7 consiste à montrer que, si Ω est un ouvert strictement convexe à bord $\mathscr{C}^1$, on a $\Delta_\Omega < \infty$ si et seulement si Ω est quasi-symétriquement convexe. La démonstration de ce fait découle d'une étude minutieuse de la notion de fonction quasi-symétriquement convexe.

4. EXEMPLES ET CONTRE-EXEMPLES

Nous mentionnons ici des exemples de convexes divisibles qui sont construits par Benoist dans [**11**] et [**12**]. Ces exemples sont tous construits sur le principe suivant : on se donne un système de Coxeter (S, M), c'est-à-dire un ensemble S et une famille $M = (m_{s,t})_{s,t\in S}$ d'éléments de $\mathbb{N} \cup \{\infty\}$ avec, pour tout s dans S, $m_{s,s} = 1$ et, pour tous $s \neq t$ dans S, $m_{s,t} = m_{t,s} \geq 2$. Le groupe de Coxeter W associé à ce système est le groupe défini par des générateurs $(g_s)_{s\in S}$, avec les relations $(g_s g_t)^{m_{s,t}} = e$, pour s, t dans S tels que $m_{s,t} < \infty$. En particulier, les éléments $(g_s)_{s\in S}$ sont des involutions. On va alors chercher à construire une représentation linéaire de W qui divise un cône convexe. Dans cette représentation, les $(g_s)_{s\in S}$ agiront comme des symétries hyperplanes. Il s'agit donc d'exhiber un espace vectoriel V et des éléments $(v_s)_{s\in S}$ de V et $(\varphi_s)_{s\in S}$ de V^* avec, pour tout s dans S, $\varphi_s(v_s) = 2$, de façon à ce que les symétries hyperplanes $\sigma_s : v \mapsto v - \varphi_s(v)v_s, V \to V$, $s \in S$, vérifient les mêmes relations que les $(g_s)_{s\in S}$. Des critères dus à Tits et à Vinberg (voir [**11**, 4.1], [**14**], [**48**]) permettent de garantir que cette représentation est fidèle et discrète et qu'elle divise un cône convexe. Par ailleurs, un critère dû à Moussong [**18, 19**] permet de déterminer si le groupe W est hyperbolique.

Appliquées à des exemples précis, ces notions permettent d'établir les résultats suivants.

PROPOSITION 4.1 ([**11**, Proposition 1.3]). — *Il existe un sous-groupe discret et Zariski dense de* $\mathrm{PGL}_4(\mathbb{R})$ *qui divise un ouvert convexe saillant non strictement convexe de* $\mathbb{P}^3$.

PROPOSITION 4.2 ([**12**, Proposition 3.1]). — *Il existe un sous-groupe discret et Zariski dense de* $\mathrm{PGL}_5(\mathbb{R})$ *qui divise un ouvert strictement convexe de* $\mathbb{P}^4$, *mais qui n'est pas isomorphe à un sous-groupe discret co-compact de* $\mathrm{SO}(4,1)$.

Ce dernier résultat a récemment été étendu par Kapovich [**30**] en toute dimension ≥ 4 par des méthodes différentes. Notons que la conjecture d'hyperbolisation de Thurston implique qu'un tel exemple ne peut pas exister en dimension 3.

RÉFÉRENCES

[1] H. ABELS – Parallelizability of proper actions, global K-slices and maximal compact subgroups, *Math. Ann.* **212** (1974/75), p. 1–19.

[2] L. V. AHLFORS – *Lectures on quasiconformal mappings*, 2e éd., University Lecture Series, vol. 38.

[3] R. C. ALPERIN – An elementary account of Selberg's lemma, *Enseign. Math.* **33** (1987), p. 269–273.

[4] R. BENEDETTI & J.-J. RISLER – *Real algebraic and semi-algebraic sets*, Actualités Mathématiques., Hermann, 1990.

[5] Y. BENOIST – Propriétés asymptotiques des groupes linéaires, *Geom. Funct. Anal.* **7** (1997), p. 1–47.

[6] ———, Automorphismes des cônes convexes, *Invent. Math.* **141** (2000), p. 149–193.

[7] ———, Convexes divisibles. II, *Duke Math. J.* **120** (2003), p. 97–120.

[8] ———, Convexes hyperboliques et fonctions quasisymétriques, *Publ. Math. I.H.É.S.* **97** (2003), p. 181–237.

[9] ———, Convexes divisibles. I, in *Algebraic groups and arithmetic*, Tata Inst. Fund. Res., 2004, p. 339–374.

[10] ———, Convexes divisibles. III, *Ann. Sci. École Norm. Sup.* **38** (2005), p. 793–832.

[11] ———, Convexes divisibles. IV. Structure du bord en dimension 3, *Invent. Math.* **164** (2006), p. 249–278.

[12] ———, Convexes hyperboliques et quasiisométries, *Geom. Dedicata* **122** (2006), p. 109–134.

[13] J.-P. Benzécri – Sur les variétés localement affines et localement projectives, *Bull. Soc. Math. France* **88** (1960), p. 229–332.

[14] N. Bourbaki – *Éléments de mathématique. Fasc. XXXVIII : Groupes et algèbres de Lie. Chapitres 4–6*, Hermann, 1975.

[15] H. Busemann & P. J. Kelly – *Projective geometry and projective metrics*, Academic Press Inc., 1953.

[16] R. D. Canary, D. B. A. Epstein & P. Green – Notes on notes of Thurston, in *Analytical and geometric aspects of hyperbolic space (Coventry/Durham, 1984)*, London Math. Soc. Lecture Note Ser., vol. 111, Cambridge Univ. Press, 1987, p. 3–92.

[17] S. Choi & W. M. Goldman – Convex real projective structures on closed surfaces are closed, *Proc. Amer. Math. Soc.* **118** (1993), p. 657–661.

[18] M. W. Davis – Nonpositive curvature and reflection groups, in *Handbook of geometric topology*, North-Holland, 2002, p. 373–422.

[19] ———, *The geometry and topology of Coxeter groups*, London Mathematical Society Monographs Series, vol. 32, Princeton Univ. Press, 2008.

[20] É. Ghys & P. De la Harpe (éds.) – *Sur les groupes hyperboliques d'après Mikhael Gromov*, Progress in Math., vol. 83, Birkhäuser, 1990.

[21] W. M. Goldman – Projective structures with Fuchsian holonomy, *J. Differential Geom.* **25** (1987), p. 297–326.

[22] ———, Convex real projective structures on compact surfaces, *J. Differential Geom.* **31** (1990), p. 791–845.

[23] W. M. Goldman & J. J. Millson – Local rigidity of discrete groups acting on complex hyperbolic space, *Invent. Math.* **88** (1987), p. 495–520.

[24] O. Guichard – Sur la régularité Hölder des convexes divisibles, *Ergodic Theory Dynam. Systems* **25** (2005), p. 1857–1880.

[25] B. Hasselblatt & A. Katok – *Introduction to the modern theory of dynamical systems*, Encyclopedia of Mathematics and its Applications, vol. 54, Cambridge Univ. Press, 1995.

[26] S. Helgason – *Differential geometry, Lie groups, and symmetric spaces*, Pure and Applied Mathematics, vol. 80, Academic Press Inc., 1978.

[27] F. John – Extremum problems with inequalities as subsidiary conditions, in *Studies and Essays Presented to R. Courant on his 60th Birthday, January 8, 1948*, Interscience Publishers, Inc., New York, N. Y., 1948, p. 187–204.

[28] D. Johnson & J. J. Millson – Deformation spaces associated to compact hyperbolic manifolds, in *Discrete groups in geometry and analysis (New Haven, Conn., 1984)*, Progr. Math., vol. 67, Birkhäuser, 1987, p. 48–106.

[29] V. G. Kac & È. B. Vinberg – Quasi-homogeneous cones, *Mat. Zametki* **1** (1967), p. 347–354.

[30] M. Kapovich – Convex projective structures on Gromov-Thurston manifolds, *Geom. Topol.* **11** (2007), p. 1777–1830.

[31] A. Karlsson & G. A. Noskov – The Hilbert metric and Gromov hyperbolicity, *Enseign. Math.* **48** (2002), p. 73–89.

[32] J.-L. Koszul – Variétés localement plates et convexité, *Osaka J. Math.* **2** (1965), p. 285–290.

[33] ———, Déformations de connexions localement plates, *Ann. Inst. Fourier (Grenoble)* **18** (1968), p. 103–114.

[34] H. Kraft, P. Slodowy & T. A. Springer (éds.) – *Algebraische Transformationsgruppen und Invariantentheorie*, DMV Seminar, vol. 13, Birkhäuser, 1989.

[35] S. G. Krantz – Lipschitz spaces, smoothness of functions, and approximation theory, *Exposition. Math.* **1** (1983), p. 193–260.

[36] N. H. Kuiper – Sur les surfaces localement affines, in *Géométrie différentielle. Colloques Internationaux du Centre National de la Recherche Scientifique, Strasbourg, 1953*, Centre National de la Recherche Scientifique, 1953, p. 79–87.

[37] ———, On convex locally-projective spaces, in *Convegno Internazionale di Geometria Differenziale, Italia, 1953*, Edizioni Cremonese, 1954, p. 200–213.

[38] D. Li & H. Queffélec – *Introduction à l'étude des espaces de Banach*, Cours Spécialisés, vol. 12, Soc. Math. France, 2004.

[39] G. Prasad – **R**-regular elements in Zariski-dense subgroups, *Quart. J. Math. Oxford Ser.* **45** (1994), p. 541–545.

[40] M. S. Raghunathan – *Discrete subgroups of Lie groups*, Ergebnisse Math. Grenzg., vol. 68, Springer, 1972.

[41] M. Sato & T. Kimura – A classification of irreducible prehomogeneous vector spaces and their relative invariants, *Nagoya Math. J.* **65** (1977), p. 1–155.

[42] A. Selberg – On discontinuous groups in higher-dimensional symmetric spaces, in *Contributions to function theory (Internat. Colloq. Function Theory, Bombay, 1960)*, Tata Institute of Fundamental Research, 1960, p. 147–164.

[43] J-P. Serre – Cohomologie des groupes discrets, in *Prospects in mathematics (Proc. Sympos., Princeton Univ., Princeton, N.J., 1970)*, Ann. of Math. Studies, vol. 70, Princeton Univ. Press, 1971, p. 77–169.

[44] D. Sullivan & W. Thurston – Manifolds with canonical coordinate charts : some examples, *Enseign. Math.* **29** (1983), p. 15–25.

[45] J. Vey – Sur les automorphismes affines des ouverts convexes saillants, *Ann. Scuola Norm. Sup. Pisa* **24** (1970), p. 641–665.

[46] È. B. Vinberg – The theory of homogeneous convex cones, *Trudy Moskov. Mat. Obšč.* **12** (1963), p. 303–358.

[47] ———, Structure of the group of automorphisms of a homogeneous convex cone, *Trudy Moskov. Mat. Obšč.* **13** (1965), p. 56–83.

[48] ———, Discrete linear groups that are generated by reflections, *Izv. Akad. Nauk SSSR Ser. Mat.* **35** (1971), p. 1072–1112.

[49] A. Weil – On discrete subgroups of Lie groups, *Ann. of Math.* **72** (1960), p. 369–384.

Jean-François QUINT

Université Paris XIII
Institut Galilée
LAGA
UMR 7539 du CNRS
99 avenue J.-B. Clément
F-93430 Villetaneuse
E-mail : quint@math.univ-paris13.fr

332

ASTÉRISQUE

2010

SÉMINAIRE BOURBAKI
VOLUME 2008/2009
EXPOSÉS 997-1011

(1000) *Le groupe de Cremona et ses sous-groupes finis*

Jean-Pierre SERRE

SOCIÉTÉ MATHÉMATIQUE DE FRANCE
Publié avec le concours du CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

Séminaire BOURBAKI Novembre 2008
61e année, 2008-2009, no 1000, p. 75 à 100

LE GROUPE DE CREMONA ET SES SOUS-GROUPES FINIS

par **Jean-Pierre SERRE**

Qu'est-ce que le *groupe de Cremona*? Pour un géomètre, c'est le groupe des transformations birationnelles du plan (affine ou projectif, peu importe) dans lui-même. Pour un algébriste, c'est le groupe des k-automorphismes du corps $k(t_1, t_2)$ où t_1, t_2 sont des indéterminées, et k est un corps (le plus souvent $\mathbf{C}$). Ces groupes ont été beaucoup étudiés dans les 150 dernières années. Le présent exposé se bornera à en décrire les *sous-groupes finis*, dans un cadre un peu plus large que celui où $k = \mathbf{C}$; nous nous intéresserons par exemple au cas où le corps k est égal à $\mathbf{Q}$ ou à un corps fini, cf. §§ 2-5. Le § 1 contient quelques définitions préliminaires, ainsi qu'un « théorème de fusion » pour le tore standard du groupe de Cremona.

1. TORE STANDARD ET FUSION

1.1. Définitions et exemples

Soit k un corps et soit n un entier ≥ 1. On note $\mathrm{Cr}_n(k)$ le groupe des k-automorphismes de $k(t_1, \ldots, t_n)$ où les t_i sont des indéterminées indépendantes ; c'est le groupe de Cremona de rang n (ou plutôt le groupe de ses k-points, cf. [De 70, § 1]).

On a par exemple $\mathrm{Cr}_1(k) = \mathbf{PGL}_2(k)$. Par la suite, on s'intéressera surtout[1] au cas $n = 2$ et l'on écrira alors Cr à la place de Cr_2.

Soit X une k-variété géométriquement irréductible réduite dont le corps des fonctions $k(X)$ soit isomorphe à $k(t_1, \ldots, t_n)$ (autrement dit une variété k-rationnelle). Choisissons un isomorphisme entre $k(X)$ et $k(t_1, \ldots, t_n)$. Tout élément de $\mathrm{Cr}_n(k)$ peut alors être identifié à un automorphisme birationnel $X \dashrightarrow X$, autrement dit à un « pseudoautomorphisme » de X au sens de [De 70, § 1.2]. Suivant ce que l'on veut

[1] On trouvera dans [Se 09, § 6] quelques conjectures (ou « questions ») sur le cas $n \geqslant 3$, mais je ne suis pas sûr qu'elles soient raisonnables.

faire, on prendra pour X l'espace affine $\mathbf{Aff}^n$, l'espace projectif $\mathbf{P}_n$ ou le tore standard $T = (\mathbf{G}_m)^n$, cf. § 1.2 ci-dessous.

Exemples de sous-groupes de $\mathrm{Cr}_n(k)$ pour $n = 2$

a) Prenons $X = \mathbf{P}_2$. On a $\mathrm{Aut}_k(X) = \mathbf{PGL}_3(k)$; comme un automorphisme birégulier est *a fortiori* birationnel, on obtient un plongement de $\mathbf{PGL}_3(k)$ dans $\mathrm{Cr}(k)$. (Du point de vue de [De 70], où l'on regarde Cr comme un foncteur en groupes, on a un plongement de $\mathbf{PGL}_3$ dans Cr ; on n'a pas besoin de mentionner k.)

b) Prenons $X = \mathbf{P}_1 \times \mathbf{P}_1$; on obtient ainsi un plongement de $\mathbf{PGL}_2 \times \mathbf{PGL}_2$ dans Cr.

c) (de Jonquières) Soit $a \in \mathbf{PGL}_2(k)$, et soit $b_t \in \mathbf{PGL}_2(k(t))$. Le couple (a, b_t) définit un élément g_{a,b_t} de $\mathrm{Cr}(k)$ par

$$(t_1,\ t_2) \quad \mapsto \quad (a(t_1),\ b_{t_1}(t_2)).$$

[*Exemple* : $(t_1,\ t_2) \ \mapsto (1/t_1,\ (t_1^3 + t_2)/(t_1 t_2 + 1))$.]

Les g_{a,b_t} forment un sous-groupe de $\mathrm{Cr}(k)$ isomorphe au produit semi-direct de $\mathbf{PGL}_2(k)$ par $\mathbf{PGL}_2(k(t))$. On peut l'interpréter comme le groupe des transformations birationnelles de $\mathbf{P}_1 \times \mathbf{P}_1$ qui sont compatibles avec la première projection : $\mathbf{P}_1 \times \mathbf{P}_1 \ \to \ \mathbf{P}_1$.

1.2. Le tore standard et son normalisateur

Soit T le produit de n exemplaires du groupe multiplicatif $\mathbf{G}_m$. Ce tore opère de façon évidente (par multiplication des coordonnées) sur $\mathbf{Aff}^n$, et aussi d'ailleurs sur $\mathbf{P}_n$, ce qui donne un plongement $T \to \mathrm{Cr}_n$. Nous dirons que T est le « tore standard » de Cr_n.

Le dual $T^* = \mathrm{Hom}(T, \mathbf{G}_m)$ de T est $\mathbf{Z}^n$. Le groupe d'automorphismes de T^* (qui est aussi celui de T) est le groupe $\Gamma = \mathbf{GL}_n(\mathbf{Z})$. Le produit semi-direct $N(k)$ de Γ par $T(k)$ opère sur T, donc se plonge dans $\mathrm{Cr}_n(k)$, ce que nous écrirons simplement $N = T.\Gamma \hookrightarrow \mathrm{Cr}_n$.

Lorsque k est algébriquement clos, on démontre que $T(k)$ est son propre centralisateur dans $\mathrm{Cr}_n(k)$ et que $N(k)$ est son normalisateur (cf. [De 70, prop. 10, cor. 5], dans un cadre un peu différent). Cette situation ressemble à celle que l'on a pour les groupes semi-simples, T jouant le rôle d'un tore maximal et Γ celui du groupe de Weyl (qui est toutefois infini si $n > 1$). Il y a aussi des analogues des racines (ce sont les éléments primitifs de T^*) et des sous-groupes additifs radiciels (il y en a une infinité pour chaque racine si $n > 1$, cf. [De 70, § 2]). On verra par la suite d'autres cas où l'analogie

$$\text{groupe de Cremona de rang } n \longleftrightarrow \text{ groupe semi-simple de rang } n$$

est à peu près justifiée — et aussi d'autres où elle ne l'est vraiment pas !

1.3. Le théorème de fusion

Dans un groupe fini, si un p-Sylow est commutatif, son normalisateur contrôle sa fusion (i.e. deux sous-groupes d'un p-Sylow qui sont conjugués dans le groupe le sont déjà dans le normalisateur du Sylow). De même, dans un groupe semi-simple, le groupe de Weyl contrôle la fusion du tore maximal. Ce sont là des résultats qui sont à la fois faciles à démontrer et très utiles (Borel, par exemple, en faisait grand usage).

Nous allons voir qu'il y a un résultat analogue pour les groupes de Cremona. Nous devrons toutefois faire l'hypothèse suivante :

(D) $$n \leqslant 2 \quad \text{ou} \quad \mathrm{car}(k) = 0.$$

[La lettre D est l'initiale de « désingularisation » : celle-ci intervient dans la démonstration, cf. fin du § 1.5.]

Voici l'énoncé du théorème :

THÉORÈME 1.1. — *Faisons l'hypothèse* (D) *ci-dessus. Soient A et B deux parties de $T(k)$ et soit $g \in \mathrm{Cr}_n(k)$ tel que $gAg^{-1} = B \subset \mathrm{Cr}_n(k)$. Il existe alors un élément γ de $\Gamma = \mathbf{GL}_n(\mathbf{Z})$ tel que $\gamma a \gamma^{-1} = gag^{-1}$ pour tout $a \in A$.*

(En particulier A et B sont conjugués dans $N(k)$.)

La démonstration sera donnée au § 1.6 ; comme on le verra, elle repose de façon essentielle sur des résultats de Reichstein et Youssin [RY 02].

Exemples. — a) Deux éléments de $T(k)$ qui sont conjugués dans $\mathrm{Cr}_n(k)$ le sont dans $N(k)$, cf. [Bl 06b, prop.6].

b) Soit $m \geqslant 1$ premier à $\mathrm{car}(k)$. Supposons que k contienne les racines m-ièmes de l'unité. Soit T_m le groupe des éléments $x \in T(k)$ tels que $x^m = 1$; ce groupe est abélien de type $(m, \ldots, m)$; on a $\mathrm{Aut}(T_m) = \mathbf{GL}_n(\mathbf{Z}/m\mathbf{Z})$. Le théorème 1.1, appliqué à $A = B = T_m$, dit que les seuls éléments de $\mathrm{Aut}(T_m)$ qui soient induits par des automorphismes intérieurs de $\mathrm{Cr}_n(k)$ sont ceux dont le déterminant dans $(\mathbf{Z}/m\mathbf{Z})^\times$ est égal à ± 1. Sous une forme un peu différente, ce résultat est dû à Reichstein et Youssin [RY 02].

PROBLÈME. — Y a-t-il une forme « infinitésimale » du théorème 1.1 ? Autrement dit, remplaçons dans l'énoncé l'hypothèse $A, B \subset T(k)$ par $A, B \subset \mathrm{Lie}(T)$; est-ce que le théorème est encore valable[(2)] ? Même le cas où A et B n'ont qu'un seul élément ne

[(2)] Il faut d'abord se convaincre qu'il a un sens. C'est clair du point de vue de [De 70], où $\mathrm{Lie}(\mathrm{Cr}_n)$ est défini comme le noyau de $\mathrm{Cr}_n(k[t]/(t^2)) \to \mathrm{Cr}_n(k)$. Plus concrètement, on peut voir un élément de $\mathrm{Lie}(T)$ comme une section rationnelle du fibré tangent à $\mathbf{Aff}^n$, et l'on peut donc le transformer par $g \in \mathrm{Cr}_n(k)$.

paraît pas évident. Lorsque $k = \mathbf{C}$, il est possible que la question puisse se traiter par voie analytique, en utilisant les feuilletages associés aux éléments non nuls de $\mathrm{Lie}(T)$.

1.4. Un lemme

Supposons k algébriquement clos.

LEMME 1.2. — *Soient V et V' deux sous-variétés fermées de T et soient V_o et V'_o des sous-ensembles denses de V et V'. Soit $g \in \mathrm{Cr}_n(k)$ tel que $g.V_o.g^{-1} = V'_o \subset \mathrm{Cr}_n(k)$. On a alors $g.V.g^{-1} = V'$ et l'application $v \mapsto g.v.g^{-1}$ est un isomorphisme de variétés de V sur V'.*

(Précisons que par « sous-variété » nous entendons « sous-schéma réduit, séparé et de type fini sur k » et que nous identifions une telle variété à l'ensemble de ses k-points. Cela donne une correspondance bijective entre parties fermées de $T(k)$ et sous-variétés fermées de T.)

Démonstration. — Identifions $\mathrm{Cr}_n(k)$ au groupe des automorphismes birationnels de T. L'application rationnelle $g : T \dashrightarrow T$ a un domaine de définition U qui est un ouvert dense de T ; elle définit un morphisme birationnel $g : U \to T$. D'autre part, si x est un élément de T on l'identifie à l'élément de $\mathrm{Cr}_n(k)$ donné par la translation τ_x de T : $\tau_x : z \mapsto z.x$. Si $v \in V_o$ et si $v' = gvg^{-1}$, on a $g \circ \tau_v = \tau_{v'} \circ g$ en tout point de $U \cap Uv^{-1}$. Soit U_2 l'ouvert de $V \times T$ formé des couples (v, z) tels que $z \in U$ et $z.v \in U$. Si $(v, z) \in U_2$, posons $F(v, z) = g(z.v).g(z)^{-1}$. L'application $F : U_2 \to T$ est un morphisme. De plus, si $v \in V_o$ et si $v' = gvg^{-1}$ comme ci-dessus, on a $F(v, z) = v'$ pour tout z tel que $(v, z) \in U_2$; or de tels points sont denses dans U_2 puisque V_o est dense dans V. Il en résulte, par continuité, que $F(U_2)$ est contenu dans V'. Ainsi, F se factorise en $U_2 \to V \to V'$. Comme la projection $U_2 \to V$ est surjective, on obtient ainsi un morphisme $f : V \to V'$, et par construction on a $f(v) = gvg^{-1}$ si $v \in V_o$. Le même argument, appliqué à V' donne un morphisme $f' : V' \to V$ qui est inverse de f. Cela démontre le lemme. □

1.5. Démonstration du théorème 1.1

Revenons à la démonstration du th. 1.1. On peut supposer que k est algébriquement clos, et que A et B sont des sous-groupes de $T(k)$. Soient $\overline{A}$ et $\overline{B}$ leurs adhérences pour la topologie de Zariski ; ce sont des sous-groupes algébriques réduits (donc lisses) du tore T.

D'après le lemme 1.2, on a $g\overline{A}g^{-1} = \overline{B}$. Il suffit donc de prouver le théorème pour $\overline{A}$ et $\overline{B}$; autrement dit, on peut supposer que A et B sont des sous-groupes algébriques lisses de T. De plus, le lemme 1.2 dit que l'application $f : a \mapsto gag^{-1}$ est

un isomorphisme du groupe algébrique A sur le groupe algébrique B. Soient A^* et B^* les duaux de A et de B. On a par définition

$$A^* = \mathrm{Hom}(A, \mathbf{G}_m) \quad \text{et} \quad B^* = \mathrm{Hom}(B, \mathbf{G}_m).$$

Les injections $A \to T$ et $B \to T$ donnent par dualité des surjections $\mathbf{Z}^n \to A^*$ et $\mathbf{Z}^n \to B^*$; de même f donne un isomorphisme $f^* : B^* \to A^*$. Tout revient à trouver un automorphisme de $\mathbf{Z}^n$ qui rende commutatif le diagramme

$$\begin{array}{ccc} \mathbf{Z}^n & \longrightarrow & B^* \\ \downarrow & & \downarrow f^* \\ \mathbf{Z}^n & \longrightarrow & A^*. \end{array}$$

Pour le prouver, on va utiliser les constructions et les résultats de [RY 02]. Si E est un groupe commutatif, engendrable par n éléments, le groupe $\wedge^n E$ est cyclique ; notons $I(E)$ l'ensemble de ses générateurs et notons $J(E)$ le quotient de $I(E)$ par la relation d'équivalence $x \sim \pm x'$. Si $E = \mathbf{Z}^n$, $I(E)$ a deux éléments et $J(E)$ en a un seul. La surjection $\mathbf{Z}^n \to B^*$ applique cet élément sur un certain élément j_B de $J(B^*)$. On définit de même $j_A \in J(A^*)$, et l'isomorphisme f^* transforme j_B en un élément $f^*(j_B)$ de $J(A^*)$.

Lemme 1.3. — *Pour qu'il existe un automorphisme de $\mathbf{Z}^n$ rendant commutatif le diagramme ci-dessus, il faut et il suffit que $f^*(j_B) = j_A$.*

Cela résulte du lemme 2.4 de [RY 02].

Nous sommes ainsi ramenés à montrer que les éléments j_A et $f^*(j_B)$ de $J(A^*)$ sont égaux. On considère pour cela les deux représentations linéaires r et r' de A données respectivement par $A \to T \to \mathbf{GL}_n$ et $A \to B \to T \to \mathbf{GL}_n$. Ce sont des représentations fidèles. Or, toute représentation linéaire fidèle ρ de A de degré n a un invariant $\mathrm{inv}(\rho)$ dans $J(A^*)$, qui se définit de la manière suivante (cf. [RY 02, § 7]) : on décompose ρ en somme directe de n représentations de degré 1, données par des caractères $\chi_1, \ldots, \chi_n$ et l'on forme le produit extérieur $\chi_1 \wedge \cdots \wedge \chi_n \in \wedge^n A^*$. Comme ρ est fidèle, cet élément est un générateur de $\wedge^n A^*$. Il est bien défini au signe près. Son image dans $J(A^*)$ est l'invariant $\mathrm{inv}(\rho)$ cherché.

On peut appliquer cette construction aux deux représentations r et r' définies ci-dessus ; on a évidemment $\mathrm{inv}(r) = j_A$ et $\mathrm{inv}(r') = f^*(j_B)$. Or l'élément g de $\mathrm{Cr}_n(k)$ transforme r en r' ; ces deux représentations sont donc birationnellement équivalentes. D'après [RY 02, th. 7.1], cela entraîne que $\mathrm{inv}(r) = \mathrm{inv}(r')$, i.e. $j_A = f^*(j_B)$, ce qui achève la démonstration du théorème.

Remarque. — En fait, Reichstein et Youssin, dans [RY 02], se limitent au cas où $\mathrm{car}(k) = 0$, car ils utilisent la résolution équivariante des singularités (cf. [BM 97],

[AW 97] et [Ko 07]). En caractéristique $p > 0$ un tel résultat n'est connu que si $n = 1$ ou 2 [méthode : itérer le processus « éclatement des points singuliers et normalisation »], et les démonstrations de [RY 02] sont valables sans changement (elles deviennent même plus simples). C'est pour cela que nous avons dû faire l'hypothèse (D) au § 1.3. On retrouvera cette restriction au § 2, dans l'énoncé du th. 2.1.

1.6. Complément : la topologie de Zariski de $\mathrm{Cr}_n(k)$

Continuons à supposer k algébriquement clos. Soit X une k-variété (au sens du § 1.5) et soit f un élément du groupe $\mathrm{Cr}_n(X)$, autrement dit un pseudo-automorphisme du X-schéma $X \times \mathbf{Aff}^n$, cf. [De 70, § 1]. La donnée de f définit une application $X(k) \to \mathrm{Cr}_n(k)$. Nous définirons la *topologie de Zariski de* $\mathrm{Cr}_n(k)$ comme la topologie la plus fine qui rende continues les applications $X(k) \to \mathrm{Cr}_n(k)$, quels que soient les choix de X et de $f \in \mathrm{Cr}_n(X)$. Une partie Y de $\mathrm{Cr}_n(k)$ est fermée pour cette topologie si et seulement si $f^{-1}(Y)$ est fermée dans $X(k)$ pour tout X et pour tout $f \in \mathrm{Cr}_n(X)$, cf. Bourbaki TG I.14. On définit de manière analogue la topologie de Zariski de $\mathrm{Cr}_n(k) \times \mathrm{Cr}_n(k)$ et l'on montre facilement que l'application produit $\mathrm{Cr}_n(k) \times \mathrm{Cr}_n(k) \to \mathrm{Cr}_n(k)$ est continue.

Exemples. — Le centralisateur dans $\mathrm{Cr}_n(k)$ d'une partie quelconque de $\mathrm{Cr}_n(k)$ est fermé. Une partie du tore $T(k)$ est fermée dans $\mathrm{Cr}_n(k)$ si et seulement si elle est fermée dans $T(k)$ pour la topologie de Zariski de $T(k)$. (L'injection $T(k) \to \mathrm{Cr}_n(k)$ est une immersion fermée.)

Problème. — Si $n > 2$, est-il vrai que $\mathrm{Cr}_n(k)$ est connexe ? C'est vrai pour $n = 2$ à cause du théorème de Max Noether suivant lequel $\mathrm{Cr}(k)$ est engendré par ses deux sous-groupes $\mathbf{PGL}_3(k)$ et $\mathbf{PGL}_2(k) \times \mathbf{PGL}_2(k)$.

Ce problème revient à déterminer un « H^0 ». On peut aller plus loin. Si C est un groupe abélien fini (d'ordre premier à la caractéristique, pour être prudent) on peut définir pour tout i le groupe de cohomologie étale $H^i_{et}(\mathrm{Cr}_n, C)$ comme la limite projective des $H^i_{et}(X, C)$, cette limite étant prise sur la catégorie des couples (X, f) comme ci-dessus. Le cas $i = 1$ est particulièrement intéressant car il équivaut à l'étude des « revêtements » finis abéliens de Cr_n. Hélas, je ne vois aucun moyen de calculer effectivement ces groupes.

2. SOUS-GROUPES FINIS (GÉNÉRALITÉS)

À partir de maintenant, on suppose que le corps de base k est un corps parfait, et l'on appelle « corps de fonctions » sur k toute extension K/k qui est de type fini et k-régulière (i.e. k est algébriquement fermé dans K).

2.1. Un résultat préliminaire

Nous allons nous intéresser aux sous-groupes finis du groupe de Cremona $\mathrm{Cr}_n(k)$ ou, ce qui revient au même aux groupes finis de k-automorphismes de $k(t_1,\dots,t_n)$. Commençons par un résultat général, qui permet de passer des corps de fonctions (simples à définir mais difficiles à contrôler) aux variétés projectives (plus concrètes). Si K est un corps de fonctions sur k, on appelle *modèle* de K une k-variété algébrique irréductible réduite séparée X dont le corps des fonctions est K. On peut voir X, à la Chevalley-Nagata, comme un ensemble de sous-anneaux locaux de K contenant k. Si un groupe fini G opère sur K, cela donne un sens à l'expression « X est stable par G ». De tels modèles existent, et on peut même leur imposer un certain nombre de propriétés agréables. C'est ce que dit le théorème suivant :

THÉORÈME 2.1. — *Soit K un corps de fonctions sur k et soit G un sous-groupe fini de $\mathrm{Aut}_k(K)$. Alors* :

a) *Il existe un modèle projectif normal X de K stable par G.*

b) *Soit $n = \mathrm{deg.tr}_k K$. Si $n \leqslant 2$ ou si $\mathrm{car}(k) = 0$, on peut en outre choisir X lisse sur k.*

Démonstration. — a) Soit K^G le sous-corps de K fixé par G. Choisissons une k-variété projective irréductible X' dont le corps des fonctions soit K^G (l'existence d'une telle variété est claire). Soit X la normalisée de X' dans l'extension K/K^G. La variété X répond à la question. Si l'on avait choisi X' normale (ce qui était évidemment possible), on aurait $X/G = X'$.

b) On choisit X comme dans a) et on la remplace par une désingularisation équivariante, ce qui est possible vu les hypothèses faites sur n ou sur $\mathrm{car}(k)$, comme on l'a dit au § 1.5. □

Si l'on applique ceci au cas $K = k(t_1,t_2)$, on voit que tout sous-groupe fini G de $\mathrm{Cr}(k)$ peut se réaliser comme sous-groupe de $\mathrm{Aut}(X)$, où X est une surface projective lisse k-rationnelle. Bien entendu, il n'y a pas en général unicité (on peut faire éclater tout sous-ensemble fini de $X(\overline{k})$ stable par G et par $\mathrm{Gal}(\overline{k}/k)$). Mais on va voir plus bas (th. 2.2) que l'ensemble des modèles X possibles possède des éléments ayant une forme spécialement simple.

2.2. Classe anticanonique et surfaces de Del Pezzo (rappels)

Dans ce qui suit, « surface » est une abréviation pour « k-surface projective lisse géométriquement connexe », et « surface k-rationnelle » signifie « surface dont le corps des fonctions est isomorphe à $k(t_1,t_2)$ ». Une surface est dite « géométriquement rationnelle » si elle devient rationnelle sur une clôture algébrique $\overline{k}$ de k.

Rappelons quelques définitions.

Classe anticanonique. — L'opposée de la classe canonique $K_X \in \mathrm{Pic}(X)$. C'est la classe du fibré inversible $\omega^{-1} = \det(\mathcal{T}_X)$, où $\mathcal{T}_X$ est le faisceau tangent de X et ω est le faisceau dualisant. Exemple : si $X = \mathbf{P}_2$, la classe anticanonique de X est $3L$, où L est la classe des droites ; le faisceau inversible correspondant est le faisceau $O(3)$.

Surface de Del Pezzo. — Surface telle que la classe anticanonique $-K_X$ soit ample. Références : [Ko 96, pp. 172-176], [De 80], [Do 07, chap. 8], [Ma 66, § 24]. Une telle surface a un *degré d*, défini par

$$d = K_X \,.\, K_X .$$

Il peut prendre n'importe quelle valeur entre 1 et 9. Sur $\overline{k}$, une surface de Del Pezzo de degré $\neq 8$ est isomorphe ([Ma 66, th. 24.4]) au plan projectif $\mathbf{P}_2$ dont on a éclaté $9 - d$ points en position suffisamment générale (pas 3 points sur une même droite, pas 6 points sur une même conique, et encore une autre condition lorsque $d = 1$, cf. [De 80, p. 27, th. 1]) ; pour $d = 8$ il y a deux possibilités : la quadrique $\mathbf{P}_1 \times \mathbf{P}_1$ et le plan projectif dont on a éclaté un point. En particulier, une surface de Del Pezzo est géométriquement rationnelle ; elle n'est pas toujours k-rationnelle.

Les propriétés géométriques et arithmétiques de ces surfaces sont fort intéressantes (on a écrit des volumes rien que sur le cas $d = 3$) ; ces propriétés interviennent de façon essentielle dans les démonstrations des §§ 3-4 ci-dessous (ce qui est d'ailleurs regrettable, car on ne peut pas espérer quelque chose d'aussi précis pour Cr_n si $n > 2$).

L'une des propriétés les plus utiles est la suivante : si X est une surface de Del Pezzo, le groupe $\mathrm{Aut}_k X$ des k-automorphismes de X se plonge dans une suite exacte :

$$1 \to \mathrm{A}^{\mathrm{o}}(k) \to \mathrm{Aut}_k X \to W,$$

où A^{o} est un groupe linéaire connexe (égal à 1 si $d < 6$) et W est le groupe de Weyl d'un système de racines de rang $9 - d$ (si $d < 7$), cf. [Ma 66, §§ 25-26] et [De 80]. Par exemple :

Si $d = 9$, X est une surface de Severi-Brauer (autrement dit une k-forme de $\mathbf{P}_2$) et A^{o} est de type A_2 : c'est une k-forme intérieure de $\mathbf{PGL}_3$; on a $W = 1$.

Si $d = 8$, X est, ou bien l'éclatée de $\mathbf{P}_2$ en un point rationnel, ou bien une k-forme de quadrique (surface de degré 2 dans une variété de Severi-Brauer de dimension 3) ; dans le second cas, A^{o} est un groupe adjoint de type $A_1.A_1$ et le groupe W est d'ordre 2.

Si $d = 6$, le groupe $T = \mathrm{A}^{\mathrm{o}}$ est un tore de dimension 2 ; il y a 6 courbes exceptionnelles sur $\overline{k}$; le graphe de leurs relations d'incidence est un hexagone ; le complémentaire dans X de leur réunion est un T-torseur et X est une variété toroïdale ; le groupe W est un groupe diédral d'ordre 12. [Pour une description explicite de ce cas, voir [CKM 07, § 4.2].]

Si $d = 5$, on a $A^o = 1$ et W est isomorphe au groupe symétrique $\mathfrak{S}_5$; il y a 10 courbes exceptionnelles sur $\overline{k}$; le graphe de leurs relations d'incidence est le graphe de Petersen[3] ; la flèche $\mathrm{Gal}(\overline{k}/k) \to W = \mathfrak{S}_5$ induit une correspondance bijective entre les classes de surfaces de Del Pezzo de degré 5 et les k-algèbres étales de rang 5, cf. Skorobogatov [Sk 01, § 3.1].

Si $d = 1, 2$ ou 3, on a $A^o = 1$ et W est un groupe de Weyl de type E_8, E_7 ou E_6 ; les classes dans $\mathrm{Pic}(X_{/\overline{k}})$ des courbes exceptionnelles correspondent aux poids $\neq 0$ d'une représentation irréductible de E_8, E_7 ou E_6 de dimension 248, 56 ou 27.

2.3. Sous-groupes finis et modèles minimaux

Soit G un groupe fini, et soit X une G-surface, c'est-à-dire une surface (au sens défini au début du § 2.2) munie d'un plongement de G dans $\mathrm{Aut}_k X$. On dit que X est « G-k-minimale » si les conditions équivalentes suivantes sont satisfaites :

(2.3.1) Tout morphisme birationnel $f : X \to X'$ de G-surfaces est un isomorphisme.

(2.3.2) Il n'existe pas de famille finie non vide (C_i) de courbes de $X_{/\overline{k}}$ ayant les trois propriétés suivantes :

i) les C_i sont deux à deux disjointes ;

ii) chaque C_i est une courbe exceptionnelle, i.e. lisse de genre 0 et de self-intersection -1 ;

iii) la famille (C_i) est stable à la fois par l'action de G et par celle de $\mathrm{Gal}(\overline{k}/k)$.

L'équivalence de ces deux conditions résulte des propriétés standard des courbes exceptionnelles, cf. par exemple [Ha 77, Chap.V, § 3 et § 5, th. 5.7].

Nous pouvons maintenant énoncer le théorème suivant, qui jouera un rôle essentiel dans toute la suite :

THÉORÈME 2.2. — *Avec les notations du th.* 2.1 *et en supposant* $n = 2$, *alors* :

a) *On peut choisir la G-surface X de telle sorte qu'elle soit G-k-minimale.*

b) *Si X est géométriquement rationnelle, et G-k-minimale, X est de l'un des deux types suivants* :

b1) *Il existe une courbe projective lisse C de genre* 0, *munie d'une action de G, et un morphisme $\pi : X \to C$ qui commute à l'action de G et fait de X un fibré en coniques*[4]. *On a* $\mathrm{Pic}(X)^G \cong \mathbf{Z}^2$.

b2) *X est une surface de Del Pezzo* ; *on a* $\mathrm{Pic}(X)^G \cong \mathbf{Z}$.

---

[3] Le graphe de Petersen s'obtient de la manière suivante : on se donne un ensemble Ω à 5 éléments, et l'on prend pour sommets du graphe les parties à 2 éléments de Ω ; deux sommets sont joints par une arête si les parties correspondantes de Ω sont disjointes. C'est un graphe de valence 3 ; son groupe d'automorphismes s'identifie au groupe des permutations de Ω (on peut reconstituer canoniquement Ω à partir du graphe).

[4] On dit que $\pi : X \to C$ est un *fibré en coniques* si la fibre générique de π est lisse de genre 0, et si les fibres spéciales sont, soit lisses, soit formées de deux courbes de genre 0 se coupant transversalement

Démonstration. — L'assertion a) est immédiate. Pour b), voir [DI 08, § 2]. □

REMARQUE. — Si X est k-rationnelle, alors, dans le cas b1), la courbe C est isomorphe à la droite projective, puisqu'elle a des points rationnels. C'est le cas le plus intéressant pour la suite.

Historique. — Le cas $G = 1$ et $k = \mathbf{C}$ est dû à Castelnuovo. Il entraîne par exemple que toute surface rationnelle minimale est, soit le plan projectif, soit une surface réglée (fibré en coniques sans fibres singulières). Il y a eu ensuite toute une série de progrès, dus notamment à la théorie de Mori, cf. [Mo 82] ; on en trouvera une description dans [DI 07, § 3.3]. Je signale en particulier le texte de Manin [Ma 67] qui traite à la fois le cas $G = 1$ (k étant quelconque) et le cas où G est commutatif et k algébriquement clos ; l'hypothèse de commutativité a été supprimée plus tard par Iskovskikh [Is 79]. Voir aussi [Zh 01] et [Ko 96, III.2.1].

3. SOUS-GROUPES FINIS (k ALGÉBRIQUEMENT CLOS)

Dans ce paragraphe, on suppose k algébriquement clos. La détermination (à conjugaison près, si possible) de tous les sous-groupes finis de $\mathrm{Cr}(k)$ est une question très ancienne, qui est maintenant à peu près résolue : le lecteur curieux pourra consulter les longues listes données dans [DI 07] et [Bl 06]. Je vais me borner à en résumer les aspects les plus saillants.

3.1. Un analogue d'un théorème de Jordan

Il s'agit du théorème qui dit qu'un sous-groupe fini d'un groupe réductif L ne peut être « gros » qu'à cause de sa partie abélienne. De façon plus précise, il existe un entier J_L tel que, pour tout sous-groupe fini G de $L(k)$, d'ordre premier à $\mathrm{car}(k)$, il existe un sous-groupe abélien normal A de G dont l'indice divise J_L. (*Démonstration.* Il suffit de le prouver lorsque $L = \mathbf{GL}_n$, et un argument simple de relèvement en caractéristique 0 permet de se ramener au cas $k = \mathbf{C}$, qui est connu depuis longtemps, cf. par exemple [Fr 11].)

Le même énoncé vaut pour le groupe de Cremona :

THÉORÈME 3.1 ([Se 09, § 5.4]). — *Il existe un entier $J > 0$ ayant la propriété suivante : tout sous-groupe fini de $\mathrm{Cr}(k)$ d'ordre premier à $\mathrm{car}(k)$ contient un sous-groupe abélien normal d'indice divisant J.*

---

en un point (conique dégénérant en deux droites distinctes). Noter que, lorsqu'il n'y a pas de fibre singulière, on a un véritable « fibré » au sens topologique.

La démonstration consiste à examiner les différents cas du th. 2.2. On obtient une valeur explicite pour J : par exemple on peut prendre $J = 2^{10}3^4 5^2 7$, et même un peu mieux.

3.2. Sous-groupes simples non abéliens

Soit G un groupe fini simple non abélien, plongeable dans $\mathrm{Cr}(k)$. Supposons d'abord que $|G|$ soit premier à $\mathrm{car}(k)$. Le th. 3.1 montre que $|G|$ divise J ; à isomorphisme près, il n'y a donc qu'un nombre fini de possibilités pour G. En fait, ces possibilités se réduisent à trois (et même à zéro si $\mathrm{car}(k) = 2$ ou 3, à une si $\mathrm{car}(k) = 5$, et à deux si $\mathrm{car}(k) = 7$) :

THÉORÈME 3.2. — *Le groupe G est isomorphe à $A_1(q) = \mathbf{PSL}_2(\mathbf{F}_q)$ avec $q = 5$, 7 ou 9.*

Autrement dit, G est d'ordre 60, 168 ou 360 ; noter que ce sont les plus petits ordres possibles pour un groupe simple non abélien. Noter aussi que ces groupes sont plongeables dans $\mathrm{Cr}(k)$, puisqu'ils le sont dans $\mathbf{PGL}_3(k)$, cf. e.g. [Mi 11].

Supposons maintenant que G soit d'ordre divisible par $p = \mathrm{car}(k)$, et ne soit isomorphe à aucun des trois groupes du th. 3.2. Alors :

THÉORÈME 3.3. — *Le groupe G est isomorphe, soit à un groupe simple de type*[(5)] *$A_1(p^m), A_2(p^m)$ ou ${}^2A_2(p^m)$, soit à l'un des groupes suivants* :

— *lorsque $p = 2$, le groupe ${}^2A_3(2)$;*

— *lorsque $p = 5$, le groupe alterné $\mathfrak{A}_7$.*

(Le groupe ${}^2A_3(2)$ est le groupe d'automorphismes de la surface cubique d'équation $x^3 + y^3 + z^3 + t^3 = 0$ en caractéristique 2. Il est isomorphe au sous-groupe $W(E_6)^+$ du groupe de Weyl $W(E_6)$; son ordre est $2^6 3^4 5$. Quant au groupe $\mathfrak{A}_7$, il est plongeable dans $\mathbf{PGL}_3(k)$ si $p = 5$.)

Démonstration. — On applique le th. 2.2, et l'on est ramené aux trois cas suivants :

i) X est un fibré en coniques : ce cas donne seulement les $A_1(p^m)$ et le groupe $\mathfrak{A}_5$;

ii) $X = \mathbf{P}_2$: on utilise la classification des sous-groupes finis de $\mathbf{PGL}_3(k)$, cf. [Bl 67] et [Mi 11] ;

iii) X est une surface de Del Pezzo de degré $1, 2, \ldots, 8$: ce cas ne donne aucun nouveau groupe à l'exception de ${}^2A_3(2)$ pour $p = 2$. □

(5) Rappelons que A_1, A_2, ${}^2A_2$ et ${}^2A_3$ désignent respectivement $\mathbf{PSL}_2$, $\mathbf{PSL}_3$, $\mathbf{PSU}_3$ et $\mathbf{PSU}_4$.

3.3. Sous-groupes abéliens élémentaires d'ordre premier à $\mathrm{car}(k)$

Supposons G fini et contenu dans $\mathrm{Cr}(k)$. Nous dirons que G est *toral* si l'un de ses conjugués est contenu dans le tore standard T.

THÉORÈME 3.4. — *Soit ℓ un nombre premier distinct de $\mathrm{car}(k)$.*

i) *Si $\ell > 5$, tout ℓ-sous-groupe G de $\mathrm{Cr}(k)$ est toral ; en particulier, G est commutatif de rang $\leqslant 2$.*

ii) *Pour $\ell = 2, 3$ ou 5, il existe des sous-groupes cycliques d'ordre ℓ de $\mathrm{Cr}(k)$ qui ne sont pas toraux. De plus, le rang maximum d'un ℓ-sous-groupe abélien est* 4 (resp. 3, resp. 2) *si $\ell = 2$* (resp. $\ell = 3$, resp. $\ell = 5$).

Démonstration. — Voir [Be 07] , [BB 04] et [Bl 06]. □

Je renvoie aussi à ces textes pour les descriptions des classes de conjugaison non torales citées dans ii). Le cas $\ell = 2$ est particulièrement intéressant, car certaines de ces classes (celles dites « de de Jonquières ») correspondent aux classes d'isomorphisme de courbes hyperelliptiques de genre quelconque $\geqslant 1$. (Si $y^2 = f(x)$ est l'équation d'une telle courbe on lui fait correspondre l'involution $(x, y) \mapsto (x, f(x)/y)$.) Ces classes ont donc à la fois un « module discret » (le genre) et des « modules continus » (ceux de la courbe — par exemple son invariant modulaire si c'est une courbe elliptique).

Il est intéressant de comparer au cas des groupes semi-simples (cf. [Se 99]) : les nombres 2, 3, 5 se comportent comme les nombres premiers dits « de torsion ». Il y a toutefois des différences importantes : dans un groupe semi-simple, il n'y a qu'un nombre fini de classes pour un ordre fixé, et tout sous-groupe cyclique d'ordre premier à la caractéristique est toral.

3.4. Éléments d'ordre une puissance de $p = \mathrm{car}(k)$

Supposons que la caractéristique p de k soit > 0. Le groupe additif W_n des vecteurs de Witt de longueur n est un espace affine de dimension n ; en faisant opérer W_n sur lui-même par translations, on obtient un plongement de W_n dans $\mathrm{Aut}(\mathbf{Aff}^n)$, donc aussi dans Cr_n. Cela montre que $\mathrm{Cr}_n(k)$ contient des éléments d'ordre p^n. On peut se demander s'il contient des éléments d'ordre p^{n+1}. La réponse est « non » lorsque $n = 1$ ou 2 : c'est clair pour $n = 1$ et le cas $n = 2$ vient d'être traité par Dolgachev :

THÉORÈME 3.5 ([Do 08]). — *Si $p = \mathrm{car}(k)$, le groupe $\mathrm{Cr}(k)$ ne contient pas d'élément d'ordre p^3.*

Vu le th. 2.2, il suffit de prouver que, ni un fibré en coniques, ni une surface de Del Pezzo n'a d'automorphisme d'ordre p^3. C'est immédiat pour les fibrés en coniques (utiliser le fait que $\mathbf{PGL}_2$ n'a pas d'élément d'ordre p^2). Il faut ensuite examiner les surfaces de Del Pezzo. Lorsque $p > 2$, c'est facile. Pour $p = 2$, il y a plusieurs cas non

évidents, le plus délicat étant celui du degré 1, où un calcul détaillé est nécessaire, cf. [Do 08, § 5].

REMARQUE. — Ici, l'analogie entre le groupe de Cremona et les groupes semi-simples ne fonctionne pas bien. En effet, si S est un groupe semi-simple de rang 2, et si $p > 5$, le groupe $S(k)$ ne contient pas d'élément d'ordre p^2.

3.5. Points fixes

Soit X une G-surface rationnelle et soit X^G la sous-variété des points fixes de G. Il est souvent utile d'avoir des renseignements sur la structure de X^G. En voici quelques-uns :

3.5.1. *Si G est cyclique, on a $X^G \neq \varnothing$.* — Plus généralement, soit V une variété projective lisse rationnelle, et soit $f : V \to V$ un endomorphisme de V ; supposons en outre $\dim V \leqslant 2$ ou $\mathrm{car}(k) = 0$. Alors f a un point fixe.

En effet, les hypothèses entraînent que $H^i(V, O_V)$ est 0 pour $i > 0$ et est égal à k si $i = 0$. Le nombre de Lefschetz de f en cohomologie cohérente est donc égal à 1. D'après la formule de Woods Hole ([Il 77, cor. 6.12]), cela entraîne que f a un point fixe.

3.5.2. *Si G est un p-groupe, et si $p = \mathrm{car}(k)$, alors $X^G \neq \varnothing$.* — En effet, X est p-acyclique (pour la cohomologie étale à coefficients dans $\mathbf{Z}/p\mathbf{Z}$) : cela résulte de la suite exacte d'Artin-Schreier

$$0 \to \mathbf{Z}/p\mathbf{Z} \to \mathbf{G}_a \xrightarrow{\wp} \mathbf{G}_a \to 0,$$

et de la nullité des $H^i(X, O_X)$ pour $i > 0$. On peut alors appliquer la théorie de Smith, sous la forme donnée dans [Se 08, § 7.4] ; on en déduit que X^G est p-acyclique, et en particulier est non vide.

Variante (dans le style de [Se 08]). — S'il existait un contre-exemple sur k, il en existerait aussi un sur $\overline{\mathbf{F}}_p$. Le couple (G, X) serait alors défini sur un corps $\mathbf{F}_q$, et le groupe G opérerait sans point fixe sur l'ensemble fini $X(\mathbf{F}_q)$, ce qui entraînerait $|X(\mathbf{F}_q)| \equiv 0 \pmod{p}$. Quitte à augmenter $\mathbf{F}_q$, on peut supposer que X est $\mathbf{F}_q$-rationnelle. On a alors $|X(\mathbf{F}_q)| \equiv 1 \pmod{q}$ car c'est vrai pour $X = \mathbf{P}_2$ et la classe de $|X(\mathbf{F}_q)| \pmod{q}$ ne change pas par éclatements. Contradiction.

3.5.3. *Si $|G|$ est premier à $\mathrm{car}(k)$, X^G est lisse.* — Cela résulte de la lissité de X, cf. [Fo 73] et [Iv 72, lemme 2.3].

3.5.4. *Invariance des composantes de X^G de genre > 0.* — Parmi les diverses composantes irréductibles de X^G, il peut y en avoir qui soient des courbes de genre > 0. Ces composantes *ne dépendent pas du modèle X choisi.* Cela provient de ce que éclatements et contractions ne peuvent créer ou détruire que des courbes de genre 0. Noter que, si G est toral, il n'y a aucune telle courbe. Cela donne une condition nécessaire (et souvent suffisante, cf. [Bl 06]) pour qu'un sous-groupe de $\mathrm{Cr}(k)$ soit toral.

3.5.5. *Invariance de X^G lorsque G est commutatif d'ordre premier à $\mathrm{car}(k)$.* — Soient (G, X) et (G, Y) deux G-surfaces et soit $f : X \dashrightarrow Y$ un G-isomorphisme birationnel (pas nécessairement un morphisme). Disons que deux points $x \in X$ et $y \in Y$ se correspondent si (x, y) est contenu dans l'adhérence du graphe de f. Alors, pour tout $y \in Y^G$, il existe $x \in X^G$ qui correspond à y (sous réserve que G soit commutatif d'ordre premier à $\mathrm{car}(k)$).

Démonstration. — Après éclatement de X, on peut supposer que f est un morphisme. On applique alors un résultat (bien plus général) de J. Kollár et E. Szabó, cf. [KS 00] et [RY 02, prop. 6.2]. □

Une conséquence de ce résultat est que, si G n'a pas de point fixe (sur un modèle quelconque), il n'est pas toral puisqu'un groupe toral a au moins 3 points fixes dans son action sur $\mathbf{P}_2$. Ceci s'applique par exemple aux sous-groupes de type (3,3) de $\mathbf{PGL}_3(k)$ en caractéristique $\neq 3$ dont l'image réciproque dans $\mathbf{SL}_3$ est non commutative : ces sous-groupes sont « non toraux » dans le groupe de Cremona.

3.6. Application : la dimension essentielle du groupe $\mathfrak{A}_6$

Supposons $\mathrm{car}(k) = 0$ (et k algébriquement clos, comme ci-dessus). Si G est un groupe fini, on définit (cf. par exemple [BR 97], [RY 00] ou [Se 03, § 5]) sa *dimension essentielle* relativement à k comme la dimension minimum d'un G-torseur versel. C'est un invariant de G que l'on note $\mathrm{ed}_k(G)$. Lorsque G est un groupe alterné $\mathfrak{A}_n$, on savait calculer $\mathrm{ed}_k(G)$ pour $n \leqslant 5$ alors que, pour $n = 6$, on savait seulement que $\mathrm{ed}_k(G)$ est égal à 2 ou à 3 (cf. [BR 97, § 6.4]). En fait :

THÉORÈME 3.6. — *On a* $\mathrm{ed}_k(\mathfrak{A}_6) = 3$.

Tout revient à montrer que $\mathrm{ed}_k(\mathfrak{A}_6)$ n'est pas égal à 2. Si c'était le cas, le corps de définition d'un $\mathfrak{A}_6$-torseur versel de dimension minimum serait le corps des fonctions d'une surface X. Le fait que ce torseur soit versel entraîne que X est unirationnelle, donc rationnelle vu les hypothèses faites sur k. Cela permet de supposer que X est, soit un fibré en coniques, soit une surface de Del Pezzo de degré $1, 2, \ldots$, ou 9. On élimine facilement toutes ces possibilités, sauf la dernière, en utilisant le fait que tout homomorphisme de $\mathfrak{A}_6$ dans $\mathbf{GL}_i(k)$ (resp. dans $\mathbf{PGL}_j(k)$) est trivial si $i < 5$ (resp. $j < 3$) [pour les cas de degrés 1,2,3, remarquer que $\mathfrak{A}_6$ opère sur $H^0(X, \omega^{-1})$, qui

est de dimension 2,3,4 respectivement]. Il reste seulement la possibilité $\deg(X) = 9$, i.e. $X = \mathbf{P}_2$ et $\mathfrak{A}_6 \subset \mathbf{PGL}_3(k)$. Mais dans ce cas on constate qu'un 3-Sylow de $\mathfrak{A}_6$ agit sans point fixe sur $\mathbf{P}_2$ (cf. fin du § 3.5.5), et ce n'est pas possible pour une action verselle d'après [RY 00, prop. 5.3].

Remarques. — 1) En utilisant des arguments analogues, A. Duncan a récemment déterminé tous les groupes finis G tels que $\mathrm{ed}_k(G) = 2$, où k est comme ci-dessus.

2) L'hypothèse $\mathrm{car}(k) = 0$ a été utilisée pour assurer que « unirationnel » entraîne « rationnel ». En caractéristique $p > 0$, il faudrait être capable de remplacer « unirationnel » par « séparablement unirationnel », mais je ne vois pas comment y parvenir.

4. SOUS-GROUPES FINIS (k PARFAIT)

4.1. Invariants cyclotomiques

Au lieu d'essayer de décrire tous les sous-groupes finis G possibles, on peut se borner, comme l'avait fait Minkowski dans [Mi 87] pour $\mathbf{GL}_n$, à donner une majoration multiplicative de $|G|$. Cela revient à majorer les ordres des ℓ-sous-groupes de Sylow de G, pour tout ℓ premier $\neq \mathrm{car}(k)$. Il se trouve (ce n'était nullement évident *a priori*) que le résultat ne dépend que de l'image $\mathrm{Im}(\chi)$ du ℓ-ième caractère cyclotomique $\chi : \mathrm{Gal}(\overline{k}/k) \to \mathbf{Z}_\ell^\times$ associé à ℓ et k. Pour exprimer le résultat, il est commode d'associer au couple (k, ℓ) les invariants numériques (t, m) définis de la façon suivante (cf. [Se 07, § 4] et [Se 09, § 1]) :

4.1.1. — Supposons $\ell > 2$. Le groupe $\mathbf{Z}_\ell^\times$ se décompose en produit direct

$$\mathbf{Z}_\ell^\times = C_{\ell-1} \times \{1 + \ell.\mathbf{Z}_\ell\},$$

où $C_{\ell-1} = (\mathbf{Z}/\ell\mathbf{Z})^\times$ est cyclique d'ordre $\ell - 1$ et $1 + \ell.\mathbf{Z}_\ell$ est un pro-ℓ-groupe isomorphe au groupe additif $\mathbf{Z}_\ell$. Comme $\ell - 1$ est premier à ℓ, tout sous-groupe fermé de $\mathbf{Z}_\ell^\times$ est produit direct de ses intersections avec les deux facteurs. En particulier, on a

$$\mathrm{Im}(\chi) = C_t \times \{1 + \ell^m.\mathbf{Z}_\ell\},$$

où t est un diviseur de $\ell - 1$, C_t est un groupe cyclique d'ordre t et m appartient à $\{1, 2, \ldots, \infty\}$, avec la convention que $\ell^\infty = 0$. Cela définit t et m sans ambiguïté ; si l'on désire préciser k et ℓ, on les note $t(k, \ell)$ et $m(k, \ell)$.

[Autre définition de t : on a $t = [k(z_\ell) : k]$ où z_ℓ est une racine primitive ℓ-ième de l'unité dans $\overline{k}$. Quant au groupe C_t, c'est $\mathrm{Gal}(k(z_\ell)/k)$.]

4.1.2. — Le cas $\ell = 2$. On pose $t = [k(i) : k]$ et l'on définit $m \in \{2, \ldots, \infty\}$ par la formule $\mathrm{Im}(\chi^2) = 1 + 2^{m+1}\mathbf{Z}_2$.

Exemples. — 1) Si $k = \mathbf{Q}$, on a $(t, m) = (\ell - 1, 1)$ si $\ell > 2$ et $(t, m) = (2, 2)$ si $\ell = 2$.

2) Si k est un corps fini à q éléments, et si $\ell > 2$, t est l'ordre de l'image de q dans $(\mathbf{Z}/\ell\mathbf{Z})^\times$ et m est égal à $v_\ell(q^t - 1) = v_\ell(q^{\ell-1} - 1)$; si $\ell = 2$, on a $m = v_2(q^2 - 1) - 1$, $t = 1$ si $q \equiv 1 \pmod 4$ et $t = 2$ si $q \equiv -1 \pmod 4$.

(Ici et dans toute la suite on note $v_\ell(x)$ la valuation ℓ-adique d'un entier x. Si A est un ensemble fini, on écrit $v_\ell(A)$ à la place de $v_\ell(|A|)$).

4.2. Le théorème principal

Fixons un nombre premier ℓ distinct de $\mathrm{car}(k)$, et soient t, m les entiers correspondants. Définissons $M(k, \ell) \in \{0, 1, \ldots, \infty\}$ par la recette suivante :

$$\text{Si } \ell = 2, \quad M(k,\ell) = 2m + 3.$$

$$\text{Si } \ell = 3, \quad M(k,\ell) = \begin{cases} 4 & \text{si } \ t = m = 1 \\ 2m + 1 & \text{sinon.} \end{cases}$$

$$\text{Si } \ell > 3, \quad M(k,\ell) = \begin{cases} 2m & \text{si } t = 1 \text{ ou } 2 \\ m & \text{si } t = 3, 4 \text{ ou } 6 \\ 0 & \text{si } t = 5 \text{ ou } t > 6. \end{cases}$$

THÉORÈME 4.1 ([Se 09, § 2.2]). — *Soit c un entier $\geqslant 0$. Pour qu'il existe un sous-groupe de $\mathrm{Cr}(k)$ d'ordre ℓ^c, il faut et il suffit que l'on ait $c \leqslant M(k, l)$.*

Démonstration. — Voir § 4.4 pour « il suffit » et § 4.5 pour « il faut ». □

COROLLAIRE. — *Il y a équivalence entre* :

a) $\mathrm{Cr}(k)$ *contient un élément d'ordre ℓ.*

b) *Le degré de l'extension cyclotomique $k(z_\ell)/k$ est égal à $1, 2, 3, 4$ ou 6.*

En effet, b) est équivalent à $M(k, \ell) > 0$.

En particulier $\mathrm{Cr}(\mathbf{Q})$ contient un élément d'ordre ℓ si et seulement si $\ell = 2, 3, 5$ ou 7.

EXERCICE. — Écrire explicitement des éléments de $\mathrm{Cr}(\mathbf{Q})$ d'ordre 2, 3, 5 et 7 (difficulté : trivial, facile, pas très difficile, difficile).

4.3. Cas particuliers

Disons que k est « petit » si ses invariants $t(k,\ell)$ et $m(k,\ell)$ ont les deux propriétés suivantes :

$$m(k,\ell) < \infty \text{ pour tout } \ell \neq \mathrm{car}(k) \tag{4.3.1}$$

$$t(k,\ell) \to \infty \text{ quand } \ell \to \infty. \tag{4.3.2}$$

Bien sûr, un corps fini est petit ; même chose pour $\mathbf{Q}$. Il en est de même d'un corps de nombres, et plus généralement de tout corps qui est extension de type fini d'un corps qui est petit. Tout corps local à corps résiduel petit est petit. Si k est petit, les $M(k,\ell)$ du § 4.2 sont 0 pour ℓ assez grand, et sont finis pour tout ℓ. On peut alors définir un entier $M(k)$ par la formule

$$M(k) = \prod_{\ell} \ell^{M(k,\ell)}, \tag{4.3.3}$$

et le th. 4.1 entraîne :

THÉORÈME 4.2. — *Supposons que k soit petit au sens ci-dessus. Alors les sous-groupes finis de* $\mathrm{Cr}(k)$ *d'ordre premier à* $\mathrm{car}(k)$ *sont d'ordre borné, et le ppcm de leurs ordres est égal à l'entier $M(k)$ défini par la formule* (4.3.3).

Exemples de calcul de $M(k)$. — a) $k = \mathbf{Q}$. On trouve :

$$M(\mathbf{Q}) = 2^7.3^3.5.7. \tag{4.3.4}$$

b) $k = \mathbf{F}_q$. On trouve (cf. [Se 09, § 2.5]) :

$$M(\mathbf{F}_q) = \begin{cases} 3.(q^4-1)(q^6-1) & \text{si } q \equiv 4 \text{ ou } 7 \pmod 9 \\ (q^4-1)(q^6-1) & \text{sinon.} \end{cases} \tag{4.3.5}$$

Par exemple :

$$M(\mathbf{F}_2) = 3^3.5.7\,; \quad M(\mathbf{F}_3) = 2^7.5.7.13\,; \quad M(\mathbf{F}_4) = 3^4.5^2.7.13.17.$$

REMARQUE. — Il y a une analogie frappante entre la formule (4.3.5) et la formule correspondante pour un groupe algébrique semi-simple déployé de rang 2 : si S est un tel groupe, le ppcm des ordres des sous-groupes de $S(\mathbf{F}_q)$ d'ordre premier à $\mathrm{car}(k)$ est $(q^a-1)(q^b-1)$, où les entiers (a,b) sont égaux à (2,2), (2,3), (2,4) ou (2,6) si S est de type $A_1.A_1, A_2, B_2$ ou G_2. Pour le groupe de Cremona, le couple (a,b) est égal à (4,6), et il y a un facteur exceptionnel égal à 3 lorsque k contient les racines 3-ièmes de l'unité mais pas les racines 9-ièmes. L'analogie n'est donc pas parfaite.

4.4. Construction de ℓ-sous-groupes de $\mathrm{Cr}(k)$

Pour prouver la partie « il suffit » du théorème 4.1, on doit construire, pour tout $c \leqslant M(k,\ell)$, un sous-groupe de $\mathrm{Cr}(k)$ d'ordre ℓ^c. On peut supposer que l'invariant $t = t(k,\ell)$ est égal à 1, 2, 3, 4 ou 6, car sinon on a $M(k,\ell) = 0$ d'où $c = 0$. Comme au § 4.1, notons C_t le groupe $\mathrm{Gal}(k(z_\ell)/k)$; c'est un groupe cyclique d'ordre t ; il est donc plongeable dans le groupe $\mathbf{GL}_2(\mathbf{Z}) = \mathrm{Aut}(T)$. Choisissons un tel plongement $i : C_t \to \mathbf{GL}_2(\mathbf{Z})$, en imposant la condition que, si t est pair, on a $-1 \in \mathrm{Im}(i)$. En composant

$$\mathrm{Gal}(\overline{k}/k) \to \mathbf{Z}_\ell^\times \to C_t \to \mathbf{GL}_2(\mathbf{Z}) = \mathrm{Aut}(T),$$

on obtient un homomorphisme $\varphi : \mathrm{Gal}(\overline{k}/k) \to \mathrm{Aut}(T)$. Soit T_φ le tore déduit de T par torsion galoisienne au moyen de φ. D'après un théorème de Voskresenskiĭ (cf. [Vo 98, § 4.9]) T_φ est une variété k-rationnelle. En faisant agir par translation T_φ sur elle-même, on obtient un plongement de T_φ dans Cr, donc de $T_\varphi(k)$ dans $\mathrm{Cr}(k)$. De plus, le groupe C_t opère sur T_φ par automorphismes. Cela donne un plongement du produit semi-direct $P = T_\varphi(k).C_t$ dans $\mathrm{Cr}(k)$. Lorsque $t = 1$ ou 2, on peut faire un peu mieux : le groupe diédral D_4 d'ordre 8 opère sur T_φ, et l'on obtient ainsi un plongement de $P' = T_\varphi(k).D_4$ dans $\mathrm{Cr}(k)$. Il ne reste plus qu'à vérifier que le groupe P (resp. P' si $t = 1$ ou 2) contient un sous-groupe d'ordre ℓ^c, ce qui ne présente pas de difficulté (cf. [Se 09, § 3]), sauf dans un cas particulier : celui où $\ell = 3, t = 1, m = 1, c = 4$; dans ce cas, la construction ci-dessus fournit au plus un groupe d'ordre 3^3 ; pour obtenir un groupe d'ordre 3^4, on doit utiliser un 3-Sylow du groupe d'automorphismes de la surface cubique d'équation

$$x^3 + y^3 + z^3 + t^3 = 0,$$

cf. [Se 09, § 3.1].

Exemple. — Prenons $k = \mathbf{Q}$ et $\ell = 7$. On a alors $t = 6$, et le tore T_φ se compactifie de façon T_φ-équivariante en une surface de Del Pezzo X de degré 6, dont les 6 courbes exceptionnelles sont permutées transitivement par le groupe C_6. Le groupe $T_\varphi(\mathbf{Q})$ contient un unique sous-groupe cyclique C_7 d'ordre 7, qui est normalisé par C_6 ; le produit semi-direct $C_7.C_6$ opère sur X.

Remarque. — Au lieu d'utiliser des tores, comme nous l'avons fait, on aurait pu se servir de surfaces de Del Pezzo, et même seulement de celles de degré $d = 3, 6, 8$ ou 9, suivant la valeur de l'invariant t :

— pour $t = 1, 2$, on prend $d = 8$ (sauf si $p = 3$, où l'on a besoin de $d = 6$ et de $d = 3$) ;

— pour $t = 3$ (resp. 4, resp. 6), on prend $d = 9$ (resp. 8, resp. 6).

4.5. Majoration de l'ordre d'un ℓ-sous-groupe de $\mathrm{Cr}(k)$

Pour achever la démonstration du th. 4.1, il faut prouver que, si G est un sous-groupe fini de $\mathrm{Cr}(k)$, on a $v_\ell(G) \leqslant M(k,\ell)$. On va faire un peu mieux : au lieu de ne s'intéresser qu'aux k-automorphismes du corps $k(t_1,t_2)$, on va considérer une k-forme L de $k(t_1,t_2)$, autrement dit un corps de fonctions sur k jouissant de la propriété suivante :

$$\overline{k} \otimes_k L \text{ est } \overline{k}\text{-isomorphe à } \overline{k}(t_1,t_2). \tag{4.5.1}$$

Une surface dont le corps des fonctions est L est géométriquement rationnelle, mais pas nécessairement k-rationnelle.

THÉORÈME 4.3 ([Se 09, § 4.1]). — *Soit G un sous-groupe fini de $\mathrm{Aut}_k(L)$, où L satisfait à (4.5.1). Soit ℓ un nombre premier $\neq \mathrm{car}(k)$. On a $v_\ell(G) \leqslant M(k,\ell)$.*

Démonstration. — Vu les th. 2.1 et 2.2, on peut supposer que G est contenu dans $\mathrm{Aut}_k(X)$ où X est, soit un fibré en coniques G-équivariant de base une courbe C de genre 0, soit une surface de Del Pezzo de degré 1, 2, ... ou 9. On examine ces dix cas les uns après les autres. En voici deux, à titre d'exemples (pour les autres voir [Se 09, §§ 4.3–4.12]) :

— *Le cas où X est un fibré en coniques.* Le groupe G opère sur la base C de la fibration, et définit donc un sous-groupe G_C de $\mathrm{Aut}_k(C)$. Le noyau N de $G \to G_C$ est un sous-groupe de $\mathrm{Aut}_{k(C)}(F)$, où F est la fibre générique de $X \to C$, et $k(C)$ est le corps des fonctions de C. Comme C (resp. F) est une courbe de genre 0, son groupe d'automorphismes est une k-forme (resp. une $k(C)$-forme) de $\mathbf{PGL}_2$. Cela permet, en appliquant par exemple [Se 07, th. 5], de majorer les ordres de G_C et de N en termes des invariants (t,m) [noter que k et $k(C)$ ont les mêmes invariants, car l'extension $k(C)/k$ est régulière, donc linéairement disjointe des extensions cyclotomiques de k]. On trouve ainsi :

$$v_\ell(G_C) \text{ et } v_\ell(N) \leqslant \begin{cases} m+1 & \text{si } \ell = 2, \\ m & \text{si } \ell > 2 \text{ et } t = 1 \text{ ou } 2, \\ 0 & \text{si } t > 2. \end{cases}$$

Comme $v_\ell(G) = v_\ell(G_C) + v_\ell(N)$, on en tire :

$$v_\ell(G) \leqslant \begin{cases} 2m+2 & \text{si } \ell = 2, \\ 2m & \text{si } \ell > 2 \text{ et } t = 1 \text{ ou } 2, \\ 0 & \text{si } t > 2. \end{cases}$$

En comparant avec la définition de $M(k,\ell)$, on constate que cette majoration entraîne $v_\ell(G) \leqslant M(k,\ell)$.

— *Le cas où X est une surface de Del Pezzo de degré* 3. La classe anticanonique est très ample ; elle donne un plongement de X dans $\mathbf{P}_3$ qui identifie X à une surface cubique. Le groupe G se plonge ainsi à la fois dans $\mathbf{GL}_4(k)$ (par son action sur $H^0(X, \omega^{-1})$) et dans le groupe de Weyl $W(E_6)$ (par son action sur $\mathrm{Pic}(X_{\overline{k}})$). L'ordre de $W(E_6)$ est $2^7 3^4 5$. Cela donne une borne pour $v_\ell(G)$ qui est $\leqslant M(k, \ell)$ sauf lorsque $\ell = 3$ et $t = 2$ (autrement dit k ne contient pas z_3) ; dans ce cas exceptionnel, on a en effet $M(k, \ell) = 3$. Il reste donc à prouver que, si $\ell = 3$ et si $|G| = 3^4$, le corps k contient z_3. Cela se démontre en considérant le plongement de G dans $\mathbf{GL}_4(k)$: si g est un élément non trivial du centre de G, on trouve que ses valeurs propres sont de la forme $\{1, z_3, z_3, z_3\}$, cf. [Se 09, § 4.10]. D'où $z_3 \in k$. □

5. DE LA CARACTÉRISTIQUE p À LA CARACTÉRISTIQUE 0

Dans le genre de questions que l'on vient de discuter, on a souvent envie de remplacer la caractéristique $p > 0$ par la caractéristique 0. En effet, dans cette dernière, on dispose d'outils plus puissants (méthodes analytiques, « vanishing theorem », etc.), et l'on a bien davantage de références. Il y a même des différences sérieuses en ce qui concerne les surfaces de Del Pezzo de degré 1 et 2. Ainsi, en caractéristique 0, toute surface de Del Pezzo de degré 2 est un revêtement quadratique du plan projectif (un « plan double ») dont le lieu de ramification (la « courbe de diramation ») est une quartique lisse ; par contre, en caractéristique 2, le lieu de ramification est une conique (comptée avec multiplicité 2), et cette conique peut ne pas être lisse : elle peut être formée de deux droites, et elle peut même être une droite double, cf. [CO 90].

En fait, tant que l'on ne s'intéresse qu'à des actions de groupes d'ordres premiers à la caractéristique, la théorie des « relèvements » (ou des « déformations ») de Grothendieck montre que tout ce qui existe en caractéristique $p > 0$ existe aussi en caractéristique 0. Ce n'est pas difficile à démontrer : il suffit de suivre pas à pas ce qu'il a expliqué au séminaire Bourbaki en 1959 ([Gr 59] — voir aussi [Il 05, th. 8.5.9 (b)]), en y incorporant l'action du groupe G. C'est ce que nous allons faire.

5.1. Énoncé du théorème

On se donne un anneau noethérien complet A de corps résiduel k, ainsi qu'une k-variété projective lisse X, géométriquement connexe, et un groupe fini G qui opère sur X. On s'intéresse à un « relèvement » de ces données sur A ; cela signifie un A-schéma X_A lisse et projectif sur A, muni d'une action de G, dont la fibre spéciale est G-isomorphe à X.

[Le cas le plus intéressant pour nous est celui où k est de caractéristique $p > 0$, A est l'anneau $W(k)$ des vecteurs de Witt, et X est une surface géométriquement rationnelle munie d'une action de G, cf. § 5.4 ci-dessous.]

Nous ferons les hypothèses suivantes :

(5.1.1) $|G|$ n'est pas divisible par $\mathrm{car}(k)$.

(5.1.2) On a $H^2(X, O_X) = 0$, où O_X est le faisceau des anneaux locaux de X.

(5.1.3) On a $H^2(X, \mathscr{T}_X) = 0$, où $\mathscr{T}_X$ est le faisceau tangent de X, i.e. le dual de Ω^1_X.

THÉORÈME 5.1. — *Si les trois hypothèses ci-dessus sont satisfaites, il existe un relèvement de (G, X) à A.*

5.2. Démonstration du théorème quand A est artinien

On procède par récurrence sur la longueur de A. Si cette longueur est 1, on a $A = k$ et l'on prend $X_A = X$. Si elle est > 1, on choisit un idéal non nul I de A tel que $\mathfrak{m}.I = 0$, où $\mathfrak{m}$ est l'idéal maximal de A. Vu l'hypothèse de récurrence, on dispose d'un relèvement $(G, X_{A/I})$ de (G, X) à l'anneau A/I. L'hypothèse (5.1.3) permet de relever $X_{A/I}$ à A, cf. [Gr 59, th. 9]. De plus, l'ensemble des classes de relèvements possibles a une structure naturelle d'espace affine (i.e. de torseur) sur le k-espace vectoriel $H^1(X, \mathscr{T}_X) \otimes_k I$ ([Gr 59, cor. 2 au th. 9]), et le groupe G opère de façon naturelle sur cet ensemble. Or un groupe fini d'automorphismes d'un espace affine, qui est d'ordre premier à la caractéristique, a un point fixe (prendre le barycentre d'une orbite). On peut donc choisir un relèvement X_A de $X_{A/I}$ qui soit G-invariant. Cela signifie que, pour tout $g \in G$, il existe un automorphisme de X_A qui relève l'action de g sur $X_{A/I}$. Notons E_g l'ensemble de ces automorphismes, et soit E la réunion des E_g ; l'ensemble E est un sous-groupe de $\mathrm{Aut}_A(X_A)$, et l'on a une suite exacte :

$$1 \longrightarrow N \longrightarrow E \longrightarrow G \longrightarrow 1,$$

où $N = \mathrm{Ker} : \mathrm{Aut}_A(X_A) \to \mathrm{Aut}_{A/I}(X_{A/I})$. Le groupe N est isomorphe à $H^0(X, \mathscr{T}_X) \otimes_k I$; il a une structure naturelle de k-espace vectoriel. Comme G est d'ordre premier à $\mathrm{car}(k)$, on a $H^i(G, N) = 0$ pour tout $i > 0$ et en particulier $H^2(G, N) = 0$. Cela montre que la suite exacte ci-dessus est scindée : elle admet une section $G \to E$. Cela permet de faire agir G sur X_A, ce qui démontre le théorème dans le cas considéré.

5.3. Fin de la démonstration du théorème

On applique ce que l'on vient de démontrer aux quotients $A/\mathfrak{m}^r$ pour $r = 1, 2, \ldots$. En passant à la limite sur r on obtient un A-schéma formel muni d'une action de G. L'hypothèse (5.1.2) entraîne d'après [Gr 59, th. 4] que ce schéma est algébrisable, et cela de façon unique (« GAGA formel »). Cela fournit le A-schéma X_A cherché, ainsi

que l'action de G sur X_A. Le fait que X_A soit projectif sur A résulte aussi de [Gr 59, th. 4].

5.4. Le cas des surfaces géométriquement rationnelles

Soit X une k-surface géométriquement rationnelle.

LEMME 5.2. — *Les conditions* (5.1.2) *et* (5.1.3) *sont satisfaites, autrement dit l'on a* $H^2(X, O_X) = 0$ *et* $H^2(X, \mathscr{T}_X) = 0$.

Soit $\omega = \det(\Omega^1)$ le faisceau dualisant de X. Par dualité, il suffit de voir que les deux espaces $\mathbf{T}_1(X) = H^0(X, \omega)$ et $\mathbf{T}_2(X) = H^0(X, \omega \otimes \Omega^1)$ sont 0. Quitte à faire une extension finie de k on peut supposer que X est une surface rationnelle. Or, de façon générale, si $\mathbf{T}$ est un foncteur tensoriel covariant (comme ici $E \mapsto \det(E)$ et $E \mapsto \det(E) \otimes E$), l'espace des sections de $\mathbf{T}(\Omega^1)$ est un invariant birationnel de la variété (projective lisse) considérée [lorsque $\mathbf{T}$ est le foncteur $\det^{\otimes m}$, c'est l'invariance birationnelle du m-ième plurigenre — le cas général se démontre de la même manière]. Il suffit donc de vérifier que $\mathbf{T}_1(X) = \mathbf{T}_2(X) = 0$ lorsque $X = \mathbf{P}_1 \times \mathbf{P}_1$, ce qui est immédiat.

Supposons maintenant que k soit de caractéristique $p > 0$. On peut appliquer le th. 5.1 à (G, X), où G est un sous-groupe de $\mathrm{Aut}(X)$ d'ordre premier à p. Choisissons pour A l'anneau $W(k)$ des vecteurs de Witt à coefficients dans k et soit $F = A[1/p]$ le corps des fractions de A. Le th. 5.1 fournit un couple (G, X_A) sur A, d'où par le changement de base $A \to F$ un couple (G, X_F) sur F.

LEMME 5.3. — *La surface* X_F *est géométriquement rationnelle.*

Comme F est de caractéristique 0, on peut appliquer à X_F le critère de Castelnuovo classique (cf. e.g. [Ko 96, III.2.4]). D'après ce critère, il y a deux choses à démontrer :

a) Que le F-espace vectoriel $V_F = H^0(X_F, \omega^{\otimes 2})$ est 0.

Posons $V_A = H^0(X_A, \omega^{\otimes 2})$; c'est un A-module de type fini, et l'on a $V_F = F \otimes_A V_A$; d'autre part, $V_A/p.V_A$ se plonge dans $H^0(X, \omega^{\otimes 2})$, qui est 0 puisque X est géométriquement rationnelle. D'après le lemme de Nakayama, on a donc $V_A = 0$, d'où $V_F = 0$.

b) Que le F-espace vectoriel $H^0(X_F, \Omega^1)$ est 0.

La démonstration est la même : on utilise le fait que $H^0(X, \Omega^1) = 0$.

[*Variante* : prouver b) en utilisant le fait que les genres arithmétiques de X et de X_F sont les mêmes.]

REMARQUE. — Si X est une surface de Del Pezzo, il en est de même de X_F, et ces deux surfaces ont le même degré.

APPLICATION. — Supposons le th. 4.3 (celui qui dit que $v_\ell(G) \leqslant M(k,\ell)$) démontré en caractéristique 0, et montrons qu'il est vrai en caractéristique $p > 0$, si $\ell \neq p$. On peut supposer que G est un ℓ-groupe. Par hypothèse, il existe une k-surface géométriquement rationnelle dont le groupe d'automorphismes contient G. Grâce à ce qui précède, on en déduit une F-surface géométriquement rationnelle qui a les mêmes propriétés. D'où $v_\ell(G) \leqslant M(F,\ell)$, et l'on conclut en remarquant que $M(F,\ell) = M(k,\ell)$, car les corps F et k ont les mêmes invariants cyclotomiques (m,t) pour tout $\ell \neq p$.

REMARQUE. — Inversement, si l'on admet le th. 4.3 pour les corps finis, on l'en déduit pour tous les corps (même ceux de caractéristique 0) par la méthode de « réduction mod p » due à Minkowski ([Mi 87]), combinée avec un théorème de densité à la Chebotarev, cf. [Se 07, § 6.5].

RÉFÉRENCES

[AW 97] D. ABRAMOVICH & J. WANG – Equivariant resolution of singularities in characteristic 0, *Math. Res. Lett.* **4** (1997), p. 427–433.

[Be 07] A. BEAUVILLE – p-elementary subgroups of the Cremona group, *J. Algebra* **314** (2007), p. 553–564.

[BB 04] A. BEAUVILLE & J. BLANC – On Cremona transformations of prime order, *C. R. Math. Acad. Sci. Paris* **339** (2004), p. 257–259.

[BM 97] E. BIERSTONE & P. D. MILMAN – Canonical desingularization in characteristic zero by blowing up the maximum strata of a local invariant, *Invent. math.* **128** (1997), p. 207–302.

[Bl 06] J. BLANC – Conjugacy classes of affine automorphisms of $\mathbb{K}^n$ and linear automorphisms of $\mathbb{P}^n$ in the Cremona groups, *Manuscripta Math.* **119** (2006), p. 225–241.

[Bl 06b] ———, Finite abelian subgroups of the Cremona group of the plane, *Univ. Genève, thèse* n° 3777 (2006). Voir aussi *C. R. Math. Acad. Sci. Paris* **344** (2007), p. 21–26.

[Bl 67] D. M. BLOOM – The subgroups of PSL(3, q) for odd q, *Trans. Amer. Math. Soc.* **127** (1967), p. 150–178.

[BR 97] J. BUHLER & Z. REICHSTEIN – On the essential dimension of a finite group, *Compositio Math.* **106** (1997), p. 159–179.

[CKM 07] J.-L. COLLIOT-THÉLÈNE, N. A. KARPENKO & A. S. MERKUR′EV – Rational surfaces and the canonical dimension of the group PGL_6, *Algebra i Analiz* **19** (2007), p. 159–178 (en russe) ; traduction anglaise : *St. Petersburg Math. J.* **19** (2008), p. 793–804.

[CO 90] P. Cragnolini & P. A. Oliverio – On the proof of Castelnuovo's rationality criterion in positive characteristic, *J. Pure Appl. Algebra* **68** (1990), p. 297–323.

[De 70] M. Demazure – Sous-groupes algébriques de rang maximum du groupe de Cremona, *Ann. Sci. École Norm. Sup.* **3** (1970), p. 507–588.

[De 80] ———, Surfaces de Del Pezzo, I-IV, in *Séminaire sur les Singularités des Surfaces*, Lecture Notes in Math., vol. 777, Springer, 1980, p. 21–69.

[Do 07] I. V. Dolgachev – Topics in classical algebraic geometry, part I, notes de cours, Univ. Michigan, Ann Arbor, 2007.

[Do 08] ———, On elements of order p^s in the plane Cremona group over a field of characteristic p, à paraître.

[DI 07] I. V. Dolgachev & V. A. Iskovskikh – Finite subgroups of the plane Cremona group, in *Algebra, Arithmetic and Geometry, Manin's Festschrift*, Progress in Math., vol. 269, Birkhäuser, 2008.

[DI 08] ———, On elements of prime order in the plane Cremona group over a perfect field, prépublication arXiv:0707.4305.

[Fo 73] J. Fogarty – Fixed point schemes, *Amer. J. Math.* **95** (1973), p. 35–51.

[Fr 11] F. G. Frobenius – Über den von L. Bieberbach gefundenen Beweis eines Satzes von C. Jordan, *Sitz. Kön. Preuss. Akad. Berlin* (1911), p. 241–248 (= *Ges. Abh.*, vol. III, 493–500).

[Gr 59] A. Grothendieck – Géométrie formelle et géométrie algébrique, Séminaire Bourbaki, vol. 1958/59, exposé n° 182, *Astérisque (hors série)* **5** (1995), p. 193–220 ; erratum p. 390.

[Ha 77] R. Hartshorne – *Algebraic geometry*, Springer, 1977, Graduate Texts in Mathematics, No. 52.

[Il 77] L. Illusie – Formule de Lefschetz pour les faisceaux cohérents, in *Cohomologie l-adique et fonctions L (SGA 5)*, Lecture Notes in Math., Springer, 1977, p. 73–137.

[Il 05] ———, Grothendieck's existence theorem in formal geometry, in *Fundamental algebraic geometry*, Math. Surveys Monogr., vol. 123, Amer. Math. Soc., 2005, p. 179–233.

[Is 79] V. A. Iskovskikh – Modèles minimaux de surfaces rationnelles sur des corps quelconques (en russe), *Izv. Akad. Nauk SSSR Ser. Mat.* **43** (1979), p. 19–43 ; traduction anglaise : *Math. USSR Izvestija* **14** (1980), p. 17–39.

[Is 96] ———, Factorisation des applications birationnelles de surfaces rationnelles du point de vue de la théorie de Mori (en russe), *Uspekhi Math. Nauk* **51** (1996), p. 3–72 ; traduction anglaise : *Russian Math. Surveys* **51** (1996), p. 585–652.

[Iv 72] B. Iversen – A fixed point formula for action of tori on algebraic varieties, *Invent. math.* **16** (1972), p. 229–236.

[Ko 96] J. Kollár – *Rational curves on algebraic varieties*, Ergebnisse Math. Grenzg., vol. 32, Springer, 1996.

[Ko 07] ———, *Lectures on resolution of singularities*, Annals of Math. Studies, vol. 166, Princeton Univ. Press, 2007.

[KS 00] J. Kollár & E. Szabó – Fixed points of group actions and rational maps, *Canadian J. Math.* **52** (2000), p. 1054–1056.

[Ma 66] Y. I. Manin – Les surfaces rationnelles sur les corps parfaits, *Publ. Math. IHÉS* **30** (1966), p. 415–475.

[Ma 67] ———, Les surfaces rationnelles sur les corps parfaits (en russe) II, *Mat. Sbornik* **72** (1967), p. 161–192 ; traduction anglaise : *Math. USSR Sbornik* **1** (1967), p. 141–168.

[Ma 86] ———, *Cubic forms*, 2e éd., North-Holland Mathematical Library, vol. 4, North-Holland Publishing Co., 1986.

[Mi 87] H. Minkowski – Zur Theorie der positiven quadratischen Formen, *J. Crelle* **101** (1887), p. 196–202 (= Ges. Abh., Band I, no VI).

[Mi 11] H. H. Mitchell – Determination of the ordinary and modular ternary linear groups, *Trans. Amer. Math. Soc.* **12** (1911), p. 207–242.

[Mo 82] S. Mori – Threefolds whose canonical bundles are not numerically effective, *Ann. of Math.* **116** (1982), p. 133–176.

[RY 00] Z. Reichstein & B. Youssin – Essential dimensions of algebraic groups and a resolution theorem for G-varieties, *Canad. J. Math.* **52** (2000), p. 1018–1056.

[RY 02] ———, A birational invariant for algebraic group actions, *Pacific J. Math.* **204** (2002), p. 223–246.

[Se 99] J-P. Serre – Sous-groupes finis des groupes de Lie, Séminaire Bourbaki, vol. 1998/99, exposé no 864, *Astérisque* **266** (2000), p. 415–430.

[Se 03] ———, Cohomological invariants, Witt invariants, and trace forms (Notes by Skip Garibaldi), in *Cohomological invariants in Galois cohomology*, Univ. Lecture Ser., vol. 28, Amer. Math. Soc., 2003, p. 1–100.

[Se 07] ———, Bounds for the orders of the finite subgroups of $G(k)$, in *Group representation theory*, EPFL Press, Lausanne, 2007, p. 405–450.

[Se 08] ———, How to use finite fields for problems concerning infinite fields, à paraître dans *Contemp. Math.* **487**, AMS, p. 183–193.

[Se 09] ———, A Minkowski-style bound for the orders of the finite subgroups of the Cremona group of rank 2 over an arbitrary field, *Moscow Math. J.* **9** (2009), p. 193–208.

[Sk 01] A. Skorobogatov – *Torsors and rational points*, Cambridge Tracts in Mathematics, vol. 144, Cambridge Univ. Press, 2001.

[Vo 98] V. E. Voskresenskiĭ – *Algebraic groups and their birational invariants*, Translations of Mathematical Monographs, vol. 179, Amer. Math. Soc., 1998.

[Zh 01] D.-Q. Zhang – Automorphisms of finite order on rational surfaces, *J. Algebra* **238** (2001), p. 560–589.

[Note ajoutée en avril 2010. Le problème posé au § 1.6 (connexité de $\mathrm{Cr}_n(k)$ pour la topologie de Zariski) a été résolu affirmativement par J. Blanc dans « Groupes de Cremona, connexité et simplicité », *Ann. Sci. E.N.S.* **43** (2010), 357–364.]

Jean-Pierre SERRE

Collège de France
3, rue d'Ulm
F–75231 Paris Cedex 05
E-mail : serre@noos.fr

332

ASTÉRISQUE

2010

SÉMINAIRE BOURBAKI
VOLUME 2008/2009
EXPOSÉS 997-1011

(1001) *Paradoxe de Scheffer-Shnirelman revu sous l'angle de l'intégration convexe*

Cédric VILLANI

SOCIÉTÉ MATHÉMATIQUE DE FRANCE
Publié avec le concours du CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

Séminaire BOURBAKI Novembre 2008
61e année, 2008-2009, no 1001, p. 101 à 134

PARADOXE DE SCHEFFER-SHNIRELMAN REVU SOUS L'ANGLE DE L'INTÉGRATION CONVEXE [d'après C. De Lellis et L. Székelyhidi]

par **Cédric VILLANI**

« Vous, les normaux*, dès qu'on vous met devant la moindre des choses sortant de l'ordinaire, vous vous mettez à poser des questions... »*
F. Brown, Paradoxe Perdu

Le paradoxe de Banach–Tarski est parfois qualifié d'« énoncé le plus surprenant de toutes les mathématiques ». Dans le domaine de la mécanique des fluides, si un théorème peut prétendre à ce titre, c'est à coup sûr le paradoxe de Scheffer–Shnirelman, selon lequel un fluide (non visqueux, incompressible) peut brusquement décider de s'agiter frénétiquement, sans qu'aucune force extérieure ne lui ait été appliquée. En fait cet énoncé est peut-être encore plus troublant que celui de Banach–Tarski, car il ne repose même pas sur l'axiome du choix...

Depuis sa découverte dans les années 1990, le paradoxe de Scheffer–Shnirelman a suscité l'étonnement, l'effroi ou le mépris, mais il s'agissait en tout cas d'un résultat singulier, pour ainsi dire isolé du courant de la recherche en équations aux dérivées partielles. La situation vient de changer radicalement avec les remarquables travaux de Camillo De Lellis et László Székelyhidi, qui réinterprètent et précisent les résultats de Scheffer et Shnirelman, tout en les replaçant au cœur de la recherche moderne en théorie des inclusions différentielles. Par la même occasion, ils révèlent un lien inattendu entre l'énoncé de Scheffer–Shnirelman et d'autres paradoxes célèbres, en particulier le théorème de plongement isométrique de Nash–Kuiper.

Cet exposé est consacré aux travaux de De Lellis et Székelyhidi, tels qu'on les trouve dans les articles [**15, 17**] ; les preuves des résultats principaux étant simples et élégantes, il me sera possible de les présenter dans leur intégralité. L'exposé se conclura par des prolongements, certains faisant l'objet de recherches en cours, d'autres extrêmement spéculatifs.

C'est un plaisir de remercier Camillo De Lellis et László Székelyhidi pour les innombrables échanges que nous avons eus au cours de la rédaction de ce texte. J'étends ces remerciements à Bernard Dacorogna et François Murat pour leurs explications

généreusement prodiguées ; et à Claude Bardos, Yann Brenier, Alessio Figalli, Uriel Frisch, Étienne Ghys, Emmanuel Giroux, Misha Gromov, Yann Ollivier et Jean-Claude Sikorav pour leurs commentaires sur ce manuscrit.

1. DEUX ÉNONCÉS PARADOXAUX

1.1. Paradoxe de Nash–Kuiper

Pour $n \in \mathbb{N} = \{1, 2, \ldots\}$ et $r > 0$ on note $B^n(r) = \{x \in \mathbb{R}^n;\ |x| < r\}$ et $S^{n-1}(r) = \{x \in \mathbb{R}^n;\ |x| = r\}$, où $|\cdot|$ désigne la norme euclidienne. On munit $B^n(r)$ et $S^{n-1}(r)$ de la structure de variété riemannienne induite par $\mathbb{R}^n$ euclidien.

THÉORÈME 1.1 (Nash 1954, Kuiper 1955). — *Pour tout $n \in \mathbb{N}$ et $r \in]0,1[$ il existe un plongement isométrique C^1 de $S^n(1)$ dans $B^{n+1}(r)$.*

En d'autres termes, il existe $\varphi \in C^1(S^n(1); B^{n+1}(r))$ réalisant une isométrie entre les variétés riemanniennes $S^n(1)$ et $\varphi(S^n(1))$, toutes deux munies de leur distance géodésique intrinsèque. En langage informel, le théorème 1.1 permet de « compacter » une sphère de rayon 1 à l'intérieur d'une boule minuscule, *sans faire de plis*. Si cet énoncé est choquant, c'est parce qu'il va à l'encontre du principe familier[(1)] selon lequel, pour $n \geq 2$, $S^n(1)$ admet un unique plongement isométrique dans $\mathbb{R}^{n+1}$, modulo bien sûr une isométrie de $\mathbb{R}^{n+1}$: c'est le théorème de rigidité de Cohn–Vossen, qui ici ne s'applique pas car il requiert plus de régularité ! (La régularité C^2 est suffisante au vu des travaux de Pogorelov [**42**].)

Même sans invoquer ce théorème de rigidité, on voit bien que le plongement de Nash–Kuiper semble contredire le « theorema egregium » de Gauss, selon lequel la courbure est invariante par isométrie. Soit en effet $r' < r$ minimal tel que $\varphi(S^n(1)) \subset \overline{B^n(r')}$, et soit x un point où $\varphi(S^n(1))$ est tangent (intérieurement) à $S^n(r')$. Un argument élémentaire de comparaison montre que les courbures sectionnelles de $\varphi(S^n(1))$ en x, définies en un sens généralisé et calculées extrinsèquement, sont minorées par $1/(r')^2 > 1$ — alors que les courbures de $S^n(1)$ sont toutes égales à 1 !

En fait, toute la saveur du théorème 1.1 réside dans l'hypothèse de régularité : en régularité lipschitzienne le résultat serait à peu près évident, alors qu'en régularité C^2 il serait faux. On peut d'ailleurs affiner : en régularité $C^{1,\alpha}$ $(0 < \alpha < 1)$ le théorème

(1) Pour être plus rigoureux, ce qui fait partie de notre expérience « tangible », c'est que l'on ne peut déformer isométriquement une sphère (ou plus généralement une surface de courbure gaussienne strictement positive), à moins de « forcer » (penser à une balle de ping-pong cabossée...). Il s'avère que le plongement de Nash–Kuiper peut s'obtenir par déformation isométrique C^1 de l'identité [**28**].

reste vrai pour α assez petit, et devient faux pour α assez grand, sans que l'on sache précisément quel est l'exposant optimal [**3, 4, 10**].

Le théorème 1.1 est démontré comme cas particulier d'un énoncé beaucoup plus général : *étant données* (M, g) *une variété riemannienne compacte de dimension* n, *et* (V, h) *une variété riemannienne de dimension* $n+1$, *s'il existe un plongement (resp. une immersion)* $\varphi \in C^1(M;V)$ *telle que* $\varphi^* h < g$ (i.e. φ réduit strictement les distances) *alors pour tout* $\varepsilon > 0$ *il existe un plongement (resp. une immersion) isométrique*[(2)] $\widetilde{\varphi} \in C^1(M;V)$ *telle que* $\|\widetilde{\varphi} - \varphi\|_{C^0} \leq \varepsilon$. Ce théorème est présenté par exemple dans les ouvrages de Gromov [**28**, section 2.4.9], Eliashberg et Mishachev [**23**, Chapitre 21], Székelyhidi [**49**]. Nash [**39**] prouvait ce résultat pour une variété V de dimension $n+2$, et Kuiper [**32**] parvenait peu après à affiner sa méthode pour aboutir à l'énoncé ci-dessus, optimal en un certain sens. Dans ces articles ainsi que dans [**33**] on pourra trouver d'autres résultats, s'appliquant par exemple à des variétés non compactes.

La stratégie de Nash, révolutionnaire à l'époque, consistait à partir de la « sous-solution stricte » φ, et à accroître progressivement les distances, « direction par direction », en introduisant des spirales ; Kuiper remplaçait ensuite les spirales par des « zigzags » convenablement régularisés.

FIGURE 1. Brique élémentaire des constructions de Nash et Kuiper. La spirale et le zigzag sont bien plus longs que le segment, tout en étant très proches de lui pour la topologie uniforme.

Quelques décennies plus tard, le procédé de Nash fut revu en profondeur et généralisé par Gromov [**28**], fondant ainsi la théorie de l'*intégration convexe*, qui a de nombreuses applications en géométrie [**23**, Partie IV] [**47**].

Entre les mains de spécialistes des équations aux dérivées partielles, la technique d'intégration convexe s'est transformée en une formidable machine à produire des contre-exemples, culminant avec la construction par Müller et Šverák [**37**] de systèmes elliptiques 2×2 « fortement quasiconvexes » admettant des solutions lipschitziennes, nulle part C^1. Un autre résultat marquant est la construction par Kirchheim et Preiss

(2) Rappelons la différence entre plongement C^1 et immersion : dans les deux cas la différentielle est supposée injective en tout point, mais le plongement est en outre supposé globalement injectif ; en clair, une immersion autorise l'auto-intersection. Une immersion φ est dite isométrique si pour tout x la différentielle $d_x\varphi$ préserve la norme des vecteurs tangents. Cela ne veut pas dire pour autant que φ induit une isométrie au sens des espaces métriques — sauf bien sûr si φ est un plongement !

d'une application lipschitzienne non affine $u : \mathbb{R}^2 \to \mathbb{R}^2$ dont le gradient ne prend que cinq valeurs différentes $A_1, \dots, A_5$, telle que u varie de manière « purement multidimensionnelle », au sens où $\mathrm{rang}\,(A_i - A_j) > 1$ pour tous indices i, j distincts.[3] Sur ces sujets et d'autres applications de l'intégration convexe en théorie des équations aux dérivées partielles on pourra consulter l'article de revue [**31**].

1.2. Paradoxe de Scheffer–Shnirelman

Dans sa version la plus simple (du moins la plus simple à énoncer), l'équation d'Euler incompressible modélise le mouvement d'un fluide non visqueux, de densité constante emplissant tout l'espace $\mathbb{R}^n$ ($n \in \mathbb{N}$, $n \geq 2$), observé sur un intervalle de temps $I \subset \mathbb{R}$:

$$\begin{cases} \dfrac{\partial v}{\partial t} + \nabla \cdot (v \otimes v) + \nabla p = f \\ \nabla \cdot v = 0, \end{cases} \tag{1}$$

où l'inconnue est le couple formé du champ de vitesses $v : \mathbb{R}^n \times I \to \mathbb{R}^n$ et de la pression $p : \mathbb{R}^n \times I \to \mathbb{R}$; et $f : \mathbb{R}^n \times I \to \mathbb{R}^n$ est un champ de forces, que l'on supposera très régulier (C^∞ à support compact). Dans (1) j'ai utilisé des notations traditionnelles : en notant $\partial_i = \frac{\partial}{\partial x^i}$, $1 \leq i \leq n$, on définit $(\nabla p)^i = \partial_i p$, $(v \otimes v)^{ij} = v^i\, v^j$, $(\nabla \cdot M)^j = \sum_i \partial_i M^{ij}$ (divergence composante par composante de la matrice M), et $\nabla \cdot v = \sum_i \partial_i v^i$. L'hypothèse de divergence nulle, $\nabla \cdot v = 0$, exprime l'incompressibilité du fluide et permet de réécrire $\nabla \cdot (v \otimes v) = (v \cdot \nabla) v$, où $(v \cdot \nabla v)^i = \sum_j v^j \partial_j v^i$, ce qui est la dérivée convective de la vitesse du fluide.

Le modèle d'Euler, qui vient tout juste de fêter son deux cent cinquantième anniversaire, est l'une des pierres angulaires de la mécanique des fluides. Pour en savoir plus sur ce sujet on pourra consulter [**1**], [**2**], [**27**], [**34**], [**36**], et les nombreuses références qui s'y trouvent.

La régularité des solutions de l'équation (1), pour des conditions initiales régulières, est un formidable problème ouvert en dimension $n \geq 3$. Il existe divers résultats de régularité « conditionnels », par exemple un théorème de Constantin, Fefferman et Majda [**9**], qui s'applique pour des champs de vitesses bornés dont le rotationnel varie de manière lipschitzienne.

De toute façon, il est a priori légitime de vouloir appliquer l'équation d'Euler à des données physiquement réalistes, non lisses. On est donc naturellement amené à la définition suivante, où l'on note $\mathscr{D}'$ l'espace des distributions :

[3] Si ∇u vaut respectivement A_i et A_j de part et d'autre d'un ensemble rectifiable, il est facile de vérifier que $\mathrm{rang}\,(A_i - A_j) = 1$ et que u se comporte comme $x \to |x_1|$: la variation de gradient se produit « selon une direction ». Au contraire, dans l'exemple de Kirchheim et Preiss, les variations font intervenir toutes les directions à la fois.

DÉFINITION 1.2. — *On dit que* (v,p) *est solution faible de* (1) *si*

$$v \in C(I;\mathscr{D}'(\mathbb{R}^n))^n \cap L^2_{\mathrm{loc}}(\mathbb{R}^n \times I)^n,\ p \in L^1_{\mathrm{loc}}(\mathbb{R}^n \times I),$$

et l'identité (1) *a lieu au sens des distributions.*

Un bref commentaire sur l'hypothèse $v \in L^2_{\mathrm{loc}}(dx\,dt)$: elle est naturelle du point de vue de la théorie des distributions pour donner un sens à (1) puisqu'elle implique $v \otimes v \in L^1_{\mathrm{loc}}$; elle est également naturelle du point de vue physique, car $\frac{1}{2}\,|v(x,t)|^2\,dx$ est la densité (instantanée) d'énergie cinétique du fluide.

Jusqu'ici il n'y a rien à redire, tout est cohérent avec la théorie générale des équations aux dérivées partielles, et le modèle semble convaincant. C'est pourquoi le résultat suivant créa un certain choc.

THÉORÈME 1.3 (Scheffer 1993, Shnirelman 1997). — *Il existe une solution faible non nulle de l'équation d'Euler incompressible en dimension 2, sans forçage* ($f \equiv 0$)*, à support compact en espace-temps.*

Du point de vue physique, cet énoncé est « évidemment » absurde : il décrit un fluide initialement au repos, qui tout à coup se met à s'agiter spontanément, sans qu'aucune force ait été exercée sur lui ; après quoi il revient au repos de lui-même, violant outrageusement le principe de conservation de l'énergie.[(4)]

La preuve de Scheffer **[43]**, très dense et longue, n'était accompagnée d'aucune explication, si ce n'est l'inspiration procurée par l'écoute d'un séminaire de Mandelbrot... On y trouve de fait une répétition « autosimilaire » d'un motif à peu près constant par morceaux. La démonstration proposée quelques années plus tard par Shnirelman **[44]**, considérablement plus lisible quoiqu'encore très complexe, fut donc accueillie avec soulagement, même si les résultats de Shnirelman étaient légèrement plus faibles (équation posée dans $\mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$ plutôt que dans $\mathbb{R}^2$ tout entier), et sa construction basée sur des techniques très différentes (analyse fréquentielle plutôt que spatiale).

Les solutions de Scheffer et Shnirelman étant jusqu'à nouvel ordre irréalistes, on peut douter de l'intérêt du théorème 1.3. Force est de constater cependant qu'il nous oblige à reconsidérer le modèle (1), ou tout au moins la définition des solutions. La méthode la plus simple et la plus radicale pour éliminer ces solutions paradoxales consiste à imposer la conservation de l'énergie. Cependant, pour des raisons recevables liées à la physique, on ne souhaite pas forcément le faire ; et même du point de vue mathématique, il n'est pas clair que cela soit suffisant pour sélectionner des solutions physiquement réalistes, nous en reparlerons dans la section 6. À ce jour nous ne

(4) Un tel fluide ferait l'envie des ingénieurs hydroélectriciens... Cependant, comme nous le verrons plus loin, son champ de vitesses oscille si violemment que l'on ne pourrait peut-être même pas convertir son énergie en travail mécanique !

savons pas, même en dimension 2, quelle est « la bonne » notion de solution faible de l'équation d'Euler incompressible : le paradoxe de Scheffer–Shnirelman nous a au moins permis de mieux prendre la mesure de notre ignorance dans le domaine de la mécanique des fluides.

1.3. Énoncé précisé par De Lellis et Székelyhidi

Dix ans après le travail de Shnirelman, De Lellis et Székelyhidi proposent une nouvelle démonstration du théorème 1.3. Leur construction, élégante et simple, fournit de nombreuses informations supplémentaires, renforçant encore le caractère paradoxal des solutions. Dans l'énoncé suivant, je noterai C_c^∞ l'espace des fonctions C^∞ à support compact, L^2_w l'espace de Lebesgue L^2 muni de sa topologie faible, et H^{-1} le dual distributionnel de l'espace de Sobolev $H^1 = W^{1,2}$ (moralement H^{-1} est fait de fonctions dont la primitive est L^2).

THÉORÈME 1.4 (De Lellis et Székelyhidi 2007, 2008). — *Soient Ω un ouvert de $\mathbb{R}^n$, $T > 0$, et $\overline{e}$ une fonction uniformément continue $\Omega\times]0,T[\to]0,+\infty[$, avec $\overline{e} \in L^\infty(]0,T[; L^1(\Omega))$. Alors pour tout $\eta > 0$ il existe une solution faible (v,p) de l'équation d'Euler* (1)*, sans forçage ($f \equiv 0$), telle que*

(i) $v \in C(\mathbb{R}; L^2_w(\mathbb{R}^n))^n$;

(ii) $v(x,t) = 0$ *si* $(x,t) \notin \Omega\times]0,T[$; *en particulier* $v(\,\cdot\,,0) = v(\,\cdot\,,T) \equiv 0$;

(iii) $\dfrac{|v(x,t)|^2}{2} = -\dfrac{n}{2}\, p(x,t) = \overline{e}(x,t)$ *pour tout* $t \in]0,T[$ *et presque tout* $x \in \Omega$;

(iv) $\sup\limits_{0\le t\le T} \|v(\,\cdot\,,t)\|_{H^{-1}(\mathbb{R}^n)} \le \eta.$

En outre,

(v) $(v,p) = \lim\limits_{k\to\infty} (v_k,p_k)$ *dans* $L^2(dx\,dt)$,

où chaque (v_k,p_k) est un couple de fonctions C^∞ à support compact, solution classique de l'équation d'Euler (1) *avec un forçage $f_k \in C_c^\infty(\mathbb{R}^n\times\mathbb{R};\mathbb{R}^n)$ bien choisi, $f_k \longrightarrow 0$ au sens des distributions.*

Quelques commentaires s'imposent sur cet énoncé :

— La continuité à valeurs dans L^2_w est un résultat beaucoup plus précis que la continuité à valeurs dans $\mathcal{D}'$; combinée avec l'estimation (iii), elle implique en fait la continuité à valeurs dans L^2 pour $0 < t < T$, et pour tous les temps si $\|\overline{e}(\,\cdot\,,t)\|_{L^1(\Omega)} \to 0$ quand $t \to 0$ et $t \to T$. (Rappelons que si $v_k \to v$ dans L^2_w et $\|v_k\|_{L^2} \to \|v\|_{L^2}$, alors $v_k \to v$ dans L^2.) Ainsi le théorème 1.4 renforce le théorème 1.3 en couvrant l'espace fonctionnel naturellement associé à la conservation de l'énergie cinétique.

— La condition (iii) montre que l'on peut prescrire *arbitrairement* la norme ponctuelle du champ de vitesses à l'intérieur de Ω. Choisissons $\Omega = B^n(1)$, $T = 1$, $\overline{e} \equiv 1/2$,

et examinons une solution v fournie par le théorème 1.4 : à $t = 0$ et $t = 1$ toutes les particules sont au repos ; mais pour $t \in]0, 1[$ toutes les particules se déplacent à vitesse 1 à l'intérieur de Ω, tandis qu'à l'extérieur de Ω tout le monde reste au repos — bien que la frontière de Ω ne soit matérialisée par aucune limite physique. Quant à la pression, elle est égale à une constante à l'intérieur de Ω, ce qui correspond à une absence d'interaction entre particules[(5)]... Autant d'énoncés physiquement invraisemblables !

— En corollaire, le théorème 1.4 fournit des solutions d'énergie finie (et même de vitesse uniformément bornée !), réfutant ainsi une objection que l'on avait coutume d'élever contre le théorème de Scheffer–Shnirelman. Ces solutions ne sont en revanche pas régulières : comme on le verra dans la section 6, c'est ce défaut de régularité qui autorise la non-conservation de l'énergie.

— La condition (iv) dit que le champ de vitesses v oscille extrêmement vite, de sorte que l'on n'observe presque aucun mouvement à échelle « macroscopique ».

— Enfin l'énoncé (v), déjà prouvé par Shnirelman dans le cas périodique, montre que les solutions pathologiques ne sont pas « isolées », mais s'obtiennent comme limites de solutions « irréprochables » avec un terme de forçage (tout aussi irréprochable) bien choisi. En d'autres termes, l'adhérence (pour une topologie suffisamment grossière) de l'espace des solutions « raisonnables » contient des solutions paradoxales.

Au-delà de ces améliorations notables, le grand mérite de De Lellis et Székelyhidi est d'avoir réinterprété les solutions paradoxales en termes d'inclusion différentielle, faisant ainsi entrer ce sujet dans un cadre déjà bien exploré qui combine l'analyse oscillatoire de Tartar, l'intégration convexe de Gromov et les arguments de catégorie de Baire, pour produire des solutions irrégulières d'équations aux dérivées partielles. En ce sens le paradoxe de Scheffer–Shnirelman et celui de Nash–Kuiper se retrouvent réunis ; on notera d'ailleurs a posteriori la similitude (ténue) entre les énoncés des théorèmes 1.1 et 1.4 : une obstruction (rigidité liée à la conservation de la courbure ou de l'énergie) est contournée grâce à un défaut de régularité (C^1 ou L^∞), et l'on s'autorise en outre la petitesse dans un espace de régularité encore inférieure (C^0 ou H^{-1}). Le lien avec l'intégration convexe était déjà présent en filigrane dans l'article de Scheffer (qui dans sa thèse en 1974 avait réalisé un travail pionnier sur les « configurations T_k », maintenant couramment utilisées par les spécialistes d'intégration convexe **[30, 31]**) ; cependant les articles de De Lellis et Székelyhidi rendent ce lien limpide. Pour caricaturer, ce qui dans la démonstration du théorème 1.4 joue le rôle des « zigzags » utilisés pour le théorème 1.1, ce sont des solutions particulières de l'équation d'Euler, oscillant rapidement entre deux valeurs constantes du champ de vitesses.

(5) L'« interaction » est en fait concentrée au bord de Ω, là où la pression varie brusquement.

Les trois sections à venir sont consacrées à la preuve du théorème 1.4, les deux suivantes à des compléments et développements.

2. ANALYSE GÉOMÉTRIQUE DES ÉQUATIONS D'EULER INCOMPRESSIBLES

Cette section basée sur [**15, 17**] est consacrée à l'analyse qualitative de l'équation d'Euler incompressible dans le formalisme des inclusions différentielles.

2.1. Inclusion différentielle

De nombreuses équations aux dérivées partielles non linéaires peuvent se réécrire comme un système d'équations *linéaires* couplé à une « relation constitutive » non linéaire, souvent appelée *inclusion différentielle* (même quand la relation ne fait pas explicitement intervenir de dérivée). Ce principe, d'usage courant, a été systématiquement étudié et formalisé par Tartar [**50, 51**] pour des problèmes principalement issus de la physique, et par Gromov [**28**] pour des applications en géométrie. Il n'y a pas en général unicité de la réécriture sous forme d'inclusion différentielle ; trouver « la bonne » reformulation peut demander une réflexion spécifique sur le problème étudié.

La proposition ci-dessous [**17**] reformule l'équation d'Euler incompressible ; je noterai I_n la matrice identité $n \times n$, et $\mathcal{S}_0^n$ l'espace des matrices symétriques réelles $n \times n$ de trace nulle.

PROPOSITION 2.1. — *Les deux systèmes d'équations suivants sont équivalents : d'une part,*

$$\begin{cases} \dfrac{\partial v}{\partial t} + \nabla \cdot (v \otimes v) + \nabla p = 0 \\ \nabla \cdot v = 0, \end{cases} \tag{2}$$

où $(v, p) \in L^\infty(\mathbb{R}^n \times \mathbb{R}; \mathbb{R}^n \times \mathbb{R}^n)$ *; et d'autre part,*

$$\begin{cases} \dfrac{\partial v}{\partial t} + \nabla \cdot M + \nabla q = 0 \\ \nabla \cdot v = 0, \end{cases} \tag{3}$$

$$M = v \otimes v - \frac{|v|^2}{n}\, I_n \qquad \textit{presque partout}, \tag{4}$$

où $(v, M, q) \in L^\infty(\mathbb{R}^n \times \mathbb{R}; \mathbb{R}^n \times \mathcal{S}_0^n \times \mathbb{R})$.

La preuve de l'équivalence entre (2) et (3)–(4) est immédiate : il suffit de poser $q = p + |v|^2/n$, ou $p = q - |v|^2/n$. On notera que (2) décrit l'équation d'Euler incompressible sans forçage, et pour tous les temps.

Pour simple qu'elle soit, la proposition 2.1 a le mérite de séparer clairement

— d'une part, le système augmenté (3), dont la *linéarité* permettra une analyse ondulatoire ;

— d'autre part, la contrainte (4), qui invite à étudier le *graphe*

$$K = \Big\{(v, M) \in \mathbb{R}^n \times \mathscr{S}_0^n; \quad M = v \otimes v - \frac{|v|^2}{n} I_n\Big\}. \tag{5}$$

La clé de l'analyse à venir réside dans le comportement de K vis-à-vis des directions de résonance du système linéaire.

Dans la suite, il sera commode de considérer des « tranches » du graphe K, en se limitant à des vecteurs de norme fixée :

Notation 2.2. — Étant donné $r > 0$ on pose

$$\begin{aligned} K_r &= \big\{(v, M) \in K;\ |v| = r\big\} \\ &= \Big\{(v, M) \in \mathbb{R}^n \times \mathscr{S}_0^n;\ \ M = v \otimes v - \frac{|v|^2}{n} I_n,\ |v| = r\Big\}. \end{aligned}$$

2.2. Cône d'onde de Tartar

Étant donné un système du premier ordre $\sum_i A^i \partial_i z = 0$ (où z est une fonction à valeurs vectorielles et les A^i sont des matrices), l'identification des solutions oscillantes dirigées selon un vecteur a (dans l'espace d'arrivée) passe par la résolution de la relation de dispersion $\sum_i \xi_i A^i a = 0$, où ξ_i est la fréquence d'oscillation dans la direction i (dans l'espace de départ). L'ensemble des directions a pour lesquelles on peut trouver une solution $\xi \neq 0$ est appelé **cône d'onde** du système ; c'est en particulier Tartar qui a rendu son usage systématique dans divers problèmes d'équations aux dérivées partielles. L'énoncé suivant est tiré de **[17]** ; je note $a \cdot b$ le produit scalaire euclidien de deux vecteurs a et b.

Proposition 2.3. — *Le cône d'onde associé au système* (3) *est égal à*

$$\Lambda = \left\{(v, M, q) \in \mathbb{R}^n \times \mathscr{S}_0^n \times \mathbb{R}; \quad \det \begin{bmatrix} M + qI_n & v \\ v & 0 \end{bmatrix} = 0\right\}. \tag{6}$$

En particulier,

(a) *Pour tout* $(v, M) \in \mathbb{R}^n \times \mathscr{S}_0^n$ *il existe* $q \in \mathbb{R}$ *tel que* $(v, M, q) \in \Lambda$ *;*

(b) *Pour tout* $v_0 \in \mathbb{R}^n$ *il existe* $p_0 \in \mathbb{R}$ *et* $\xi \in \mathbb{R}^n \setminus \{0\}$ *tel que pour toute fonction* $h \in L^1_{\rm loc}(\mathbb{R})$ *et tout* $\varepsilon > 0$, *le couple*

$$(v,p)(x,t) = (v_0,p_0)\, h\left(\frac{\xi\cdot x}{\varepsilon}\right)$$

soit solution de l'équation (2).

Preuve de la proposition 2.3. — Soit U la matrice réelle $(n+1)\times(n+1)$ définie par $U = \begin{bmatrix} M + qI_n & v \\ v & 0 \end{bmatrix}$. (Je n'ai pas fait de différence notationnelle entre v et son transposé.) Le système (3) se réécrit simplement $\mathrm{div}_{x,t} U = 0$ (la divergence en espace-temps de U est nulle), et la relation de dispersion associée est simplement $U\,[\xi, c] = 0$, où $(\xi, c) \in \mathbb{R}^n \times \mathbb{R}$ (ξ la fréquence spatiale, c la fréquence temporelle). Cette équation admet une solution non nulle si et seulement si U est de déterminant nul, d'où la formule (6).

L'énoncé (a) découle facilement : si $\xi \in v^\perp$ et $c = 0$ l'équation devient $M\xi + q\xi = 0$, il suffit donc de choisir q égal à une valeur propre de $-M$, considéré comme un opérateur symétrique sur $v^\perp$. Enfin, pour obtenir (b) on pose $M = v\otimes v - (|v|^2/n)\, I_n$, $c = 0$, $p = q - |v|^2/n$. □

La proposition 2.3 indique que le cône d'onde de l'équation d'Euler incompressible est très gros ; en conséquence, on peut construire des solutions oscillantes en espace-temps — et même stationnaires — colinéaires à n'importe quelle direction fixée. Cependant, dans la suite il sera commode de se limiter à certaines directions de Λ, formées de matrices de rang 2, qui ne sont pas stationnaires en temps mais qui en contrepartie sont associées à une pression constante.

Lemme 2.4. — *Soient* $a, b \in \mathbb{R}^n$ *tels que* $|a| = |b|$ *et* $a \neq b$, *et soit*

$$W(a,b) = (a-b,\, a\otimes a - b\otimes b) \in \mathbb{R}^n \times \mathscr{S}^n_0.$$

Alors $(W(a,b), 0) \in \Lambda$.

Preuve du lemme 2.4. — Il suffit de remarquer que si $z \in (a-b)^\perp$ et $c = z\cdot a = z\cdot b$, alors $(z,-c) \in \ker \begin{bmatrix} a\otimes a - b\otimes b & a-b \\ a-b & 0 \end{bmatrix}$. □

Notation 2.5. — Avec les mêmes notations que dans la proposition 2.3 et le lemme 2.4, étant donné $r > 0$ on pose

$$\Lambda_r = \Big\{ t\, W(a,b);\quad |a| = |b| = r,\ a \neq \pm b,\ t \geq 0 \Big\}.$$

Le cas échéant on identifiera Λ_r à $\Lambda_r \times \{0\} \subset \Lambda$.

2.3. Enveloppe Λ-convexe de K

La suite du programme consiste à identifier l'enveloppe Λ-convexe de K, que l'on peut définir intuitivement comme l'ensemble de tous les états accessibles par « mélange » des états de K au moyen d'oscillations rapides dans les directions de Λ.

Dans la théorie de la **compacité par compensation**, fondée par Tartar **[50]** et Murat **[38]**, puis appliquée aux systèmes de lois de conservation par Tartar **[51]** et DiPerna **[19]**, la situation favorable est celle où l'enveloppe Λ-convexe est aussi petite que possible, car cela limite les possibilités de perte de compacité. Au contraire, dans le cadre de l'intégration convexe, une grande enveloppe Λ-convexe sera synonyme de grande marge de manœuvre.

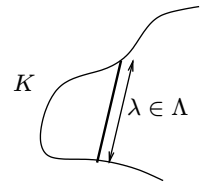


FIGURE 2. Construction de l'enveloppe Λ-convexe de K

Dans sa version la plus restreinte, l'enveloppe Λ-convexe de K est définie simplement comme le plus petit ensemble K' contenant K qui soit Λ-convexe (c'est-à-dire : $\forall a, b \in K',\ b - a \in \Lambda \Longrightarrow [a, b] \subset K'$) ; c'est un cas particulier de l'« enveloppe P-convexe » considérée par Gromov **[28]**. Dans de nombreux exemples issus des équations aux dérivées partielles on est forcé de considérer une notion plus faible d'enveloppe Λ-convexe, à savoir le complémentaire de l'ensemble des points qui peuvent être séparés de K par une fonction Λ-convexe (convexe dans les directions de Λ). À ce sujet on pourra consulter **[31]** pour des explications et références.(6)

Pour l'équation d'Euler incompressible, le cône d'onde est particulièrement gros, de sorte que la notion restreinte d'enveloppe Λ-convexe nous suffira ; et suffira même à engendrer la totalité de l'enveloppe *convexe*, au sens usuel, du graphe K. C'est le contenu de la proposition suivante **[15, 17]** ; si A et B sont deux compacts de l'espace euclidien je noterai $\mathrm{dist}\,(A, B) = \inf\{|a - b|;\ a \in A,\ b \in B\}$.

PROPOSITION 2.6. — *Avec les notations 2.2 et 2.5,*

(6) Notons, en rapport avec le sujet de cet exposé, que l'enveloppe Λ-convexe généralisée semble remonter à la thèse non publiée de Scheffer !

(i) *Pour tout $r > 0$, l'enveloppe Λ-convexe de K_r coïncide avec l'enveloppe convexe de K_r, et vaut*

$$K_r^{\rm co} = \Big\{(v, M) \in \mathbb{R}^n \times \mathcal{S}_0^n; \quad |v| \leq r \quad et \quad v \otimes v - M \leq \frac{r^2}{n}\, I_n\Big\}. \tag{7}$$

En outre $K_r = K_r^{\rm co} \cap \{|v| = r\}$.

(ii) *Il existe une constante $C = C(n) > 0$ telle que pour tout $r > 0$ et tout $z = (v, M)$ dans l'intérieur de $K_r^{\rm co}$ il existe $\lambda = (\overline{v}, \overline{M}) \in \Lambda_r$ tel que $[z - \lambda, z + \lambda] \subset \mathrm{int}\, K_r^{\rm co}$ et*

$$|\overline{v}| \geq \frac{C}{r}\,(r^2 - |v|^2); \qquad \mathrm{dist}\Big([z - \lambda, z + \lambda],\, \partial K_r^{\rm co}\Big) \geq \frac{1}{2}\,\mathrm{dist}\,\big(z, \partial K_r^{\rm co}\big). \tag{8}$$

Comme corollaire important de la proposition 2.6 (i), la solution nulle appartient à l'intérieur de l'enveloppe Λ-convexe de K. C'est ce qui permettra de construire des solutions paradoxales à support compact, en partant de 0. La dernière partie de l'énoncé indique que K_r, vu selon Λ_r, n'est pas « trop pointu » : dès que $z \in K_r^{\rm co}$ est quelque peu éloigné du bord, on peut l'écrire comme milieu d'un Λ_r-segment de taille minorée, inclus dans $K_r^{\rm co}$.

Remarque 2.7. — Du point de vue technique, la convexité de l'enveloppe Λ-convexe est un soulagement. En effet, contrairement à d'autres notions classiquement utilisées en intégration convexe (enveloppe rang-1-convexe, enveloppe quasiconvexe), l'enveloppe convexe est stable par perturbation, ce qui nous permettra d'échapper complètement au problème de l'« approximation par l'intérieur » [**31**], pensum classique dans le sujet. (Cette propriété d'approximation est liée à la « propriété de relaxation » considérée dans [**12**].)

Preuve de la proposition 2.6. — Soit C_r l'ensemble apparaissant au membre de droite de (7) ; il est évident que $K_r \subset C_r$. On va montrer que (a) C_r est convexe ; (b) C_r est compact ; (c) K_r contient tous les points extrémaux de C_r. Le théorème de Krein–Milman impliquera alors $K_r^{\rm co} = C_r$.

Preuve de (a) : Si $\sigma_{\max}(S)$ désigne la plus grande valeur propre de la matrice symétrique S, et $\langle x, y\rangle = x \cdot y$ est le produit scalaire de x et y, on a

$$\Phi(v, M) := \sigma_{\max}(v \otimes v - M) = \max_{\xi \in S^{n-1}(1)} \big(\langle \xi, v\rangle^2 - \langle M\xi, \xi\rangle\big);$$

la fonction Φ, supremum de fonctions convexes, est donc convexe sur $\mathbb{R}^n \times \mathcal{S}_0^n$. Il s'ensuit que $C_r = \Phi^{-1}([0, r^2/n]) \cap \{|v| \leq r\}$ est bien convexe. (Noter que pour tout $S \in \mathcal{S}_0^n$ on a $\sigma_{\max}(S) \geq 0$.)

Preuve de (b) : Si $(v, M) \in C_r$ on a $M \geq v \otimes v - (r^2/n)\, I_n \geq -(r^2/n)\, I_n$, et la condition de trace nulle entraîne donc $M \leq (1 - 1/n)\, r^2\, I_n$. Ces bornes impliquent la compacité de C_r.

Preuve de (c) : Soit $(v, M) \in C_r$. Quitte à effectuer un changement de base on peut supposer que $v \otimes v - M = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$ avec $r^2/n \geq \lambda_1 \geq \dots \geq \lambda_n$. On distingue alors trois cas :

— Si $|v| = r$ alors $M - (v \otimes v - \frac{|v|^2}{n} I_n) \geq 0$, et comme c'est une matrice à trace nulle elle est identiquement nulle, d'où $(v, M) \in K_r$.

— Si $\lambda_n = r^2/n$, alors $v \otimes v - M = \frac{r^2}{n} I_n$, en prenant la trace on trouve $|v|^2 = r^2$, d'où $(v, M) \in K_r$.

— Si $|v| < r$ et $\lambda_n < r^2/n$, écrivons $v = \sum v^i e_i$, où les e_i $(1 \leq i \leq n)$ sont les vecteurs de base ; on pose alors $\overline{v} = e_n$, $\overline{M} = \sum_{i=1}^{n-1} v^i (e_i \otimes e_n + e_n \otimes e_i) = v \otimes e_n + e_n \otimes v - 2v^n e_n \otimes e_n$, de sorte que pour tout $t \in \mathbb{R}$,

$$(v + t\overline{v}) \otimes (v + t\overline{v}) - (M + t\overline{M}) = (v \otimes v - M) + (2t\, v^n + t^2)\, e_n \otimes e_n.$$

Comme $\lambda_n < r^2/n$, cette identité implique $\Phi(v + t\overline{v}, M + t\overline{M}) \leq r^2/n$ pour $|t|$ assez petit ; condition qui implique aussi $|v + t\overline{v}| < r$. Donc (v, M) n'est pas un point extrémal de C_r, ce qui achève la démonstration de (c). Au passage, on a montré que $C_r \cap \{|v| = r\} \subset K_r$, et il s'ensuit que $K_r = C_r \cap \{|v| = r\}$.

Continuons la preuve de la proposition 2.6. Si l'on prouve (ii), alors pour z et λ comme dans l'énoncé, on peut prolonger le segment $[z - \lambda, z + \lambda]$ jusqu'à ce qu'il rencontre ∂C_r ; on aura ainsi z barycentre de (z_0, z_1) avec $z_0 - z_1 \in \Lambda_r$, et chaque $z_i = (v_i, M_i)$ vérifiera soit $|v_i| = r$ (donc $z_i \in K_r$), soit $|v_i| < r$ et $\lambda_1(v_i, M_i) = r^2/n$, où $\lambda_1(v, M) \geq \dots \geq \lambda_n(v, M)$ désignent les valeurs propres ordonnées de $v \otimes v - M$. Dans ce dernier cas on appliquera à nouveau (ii) et un passage à la limite facile pour écrire z_i comme barycentre de $(z_{i,0}, z_{i,1})$ avec $z_{i,0} - z_{i,1} \in \Lambda_r$, $|z_{i,0} - z_{i,1}| \geq$ const. $(r^2 - |v_i|^2)/r > 0$, et chaque $z_{i,j} = (v_{i,j}, M_{i,j})$ vérifiera soit $|v_{i,j}| = r$, soit $\lambda_1(v_{i,j}, M_{i,j}) = \lambda_2(v_{i,j}, M_{i,j}) = r^2/n$. On continuera de même, jusqu'à avoir exprimé z comme « Λ_r-barycentre » d'éléments (v, M) vérifiant tous $|v| = r$ ou $\lambda_n(v, M) = r^2/n$, et appartenant donc tous à K_r ; ceci achèvera de prouver (i).

Établissons donc (ii). Soit $z = (v, M) \in \operatorname{int} K_r^{\mathrm{co}}$, où int désigne l'intérieur topologique. Par le théorème de Carathéodory, z est combinaison convexe d'au plus $N + 1$ éléments de K_r, où $N = \dim(\mathbb{R}^n \times \mathscr{S}_0^n) = n(n+3)/2 - 1$: autrement dit, il existe $m \leq N + 1$, $(z_i)_{1 \leq i \leq m}$ dans K_r, et $(\mu_i)_{1 \leq i \leq m}$ dans $[0, 1]$, tels que $z = \sum_i \mu_i z_i$ et $\sum_i \mu_i = 1$.

Comme z est dans l'intérieur de K_r^{co}, quitte à réduire m, on peut supposer que tous les coefficients μ_i sont dans l'intervalle ouvert $]0, 1[$. Écrivons $z_i = (v_i,\, v_i \otimes v_i - \frac{r^2}{n} I_n)$, $|v_i| = r$. Quitte à perturber les z_i (ce qui est possible tout en restant dans l'intérieur de K_r^{co}), on peut supposer que $v_i \neq \pm v_j$ pour $i \neq j$. Sans perte de généralité

$\mu_1 = \max_i \mu_i$. Soit $j > 1$ tel que $\mu_j \, |v_j - v_1| = \max_{i \geq 2} \mu_i \, |v_i - v_1|$; on pose

$$\overline{z} = (\overline{v}, \overline{M}) = \frac{1}{2} \mu_j \left(v_j - v_1, \, v_j \otimes v_j - v_1 \otimes v_1 \right) = \frac{\mu_j}{2} \, (z_j - z_1).$$

Pour tout $t \in [-1,1]$, on a $z + t\overline{z} = \sum_i \mu_i' \, z_i$, où $\mu_i' = \mu_i$ pour $i \notin \{1, j\}$, $\mu_1' = \mu_1 - t\mu_j/2$, $\mu_j' = \mu_j + t\mu_j/2$. On vérifie aisément que tous ces coefficients restent dans $]0,1[$ pour $t \in [-1,1]$, et que $\sum \mu_j' = \sum \mu_j = 1$; on a donc bien $[z - \overline{z}, z + \overline{z}] \subset \operatorname{int} K_r^{\mathrm{co}}$. Enfin,

$$\begin{aligned} \frac{1}{4r\,N}(r^2 - |v|^2) = \frac{1}{4r\,N}(r + |v|)(r - |v|) \leq \frac{r - |v|}{2N} = \frac{|v_1| - |v|}{2N} \\ \leq \frac{|v_1 - v|}{2N} = \frac{\left| \sum \mu_i \, (v_i - v_1) \right|}{2N} \leq \frac{\mu_j \, |v_j - v_1|}{2} = |\overline{v}|, \end{aligned}$$

ce qui établit la première inégalité dans (8). Pour obtenir la seconde, il suffit de diviser λ par 2 et d'utiliser la convexité de K_r^{co}. □

3. SOLUTIONS OSCILLANTES LOCALISÉES

Notre but maintenant est de construire des solutions du système linéaire (3), *localisées en espace-temps*, qui oscillent entre deux états de K_r^{co} selon une direction $\lambda \in \Lambda$ donnée : quelque chose qui ressemble à $\lambda \, h(\frac{x \cdot \xi + ct}{\varepsilon})$, où h est à valeurs réelles comme dans la proposition 2.3.

Du fait de la contrainte de localisation (support compact), on ne peut espérer construire des solutions dont l'image soit *exactement* dirigée selon λ ; mais on va voir que l'on peut leur imposer de prendre leurs valeurs dans un voisinage arbitraire de $[-\lambda, \lambda]$.

Cette construction est la brique élémentaire des preuves de De Lellis et Székelyhidi. On peut la réaliser pour n'importe quel $\lambda \in \Lambda$ [**17**, proposition 2.2], mais il est commode de se restreindre aux $\lambda \in \Lambda_r$ [**15**] : cette restriction permettra d'imposer une contrainte supplémentaire de pression constante.

Localiser une onde à valeurs vectorielles, dont les composantes sont tenues de vérifier certaines relations, est délicat. En revanche, localiser une fonction à valeurs réelles est facile : il suffit de multiplier par une fonction plateau. Pour localiser les solutions oscillantes de (3), nous allons d'abord encoder une onde à valeurs vectorielles par une onde scalaire, que nous appellerons *potentiel*. Voici une analogie : si l'on veut localiser un champ de gradients, $F = \nabla \psi$, plutôt que d'appliquer une transformation matricielle et différentielle compliquée à $\nabla \psi$ il sera beaucoup plus efficace de considérer $\nabla(\chi \psi)$, où χ est une fonction plateau.

3.1. Potentiel

On rappelle la notation 2.5.

PROPOSITION 3.1. — *Pour tout $r > 0$ et tout $\lambda \in \Lambda_r$ il existe un opérateur différentiel matriciel homogène d'ordre 3, à coefficients constants,*

$$A(\nabla) : \ C_c^\infty(\mathbb{R}^{n+1}) \to C_c^\infty(\mathbb{R}^{n+1}; \mathbb{R}^n \times \wp_0^n),$$

et $(\xi, c) \in \mathbb{R}^n \times \mathbb{R}$, $\xi \neq 0$, tels que

(i) *pour tout $\varphi \in C_c^\infty(\mathbb{R}^{n+1})$, $A(\nabla)\varphi$ est solution du système linéaire (3) avec $q = 0$;*
(ii) *si $\psi \in C^\infty(\mathbb{R})$ et $\varphi(x,t) = \psi(\xi \cdot x + c\,t)$, alors*

$$[A(\nabla)\varphi](x,t) = \lambda\, \psi'''(\xi \cdot x + c\,t).$$

On peut en fait exhiber une formule explicite (ingénieuse) pour l'opérateur A. Dans [**15**] cela est fait via l'identification $(v, M) \to \begin{bmatrix} M & v \\ v & 0 \end{bmatrix}$; ici je travaillerai directement sur v et M, au risque de renforcer l'apparence miraculeuse des formules. (On trouvera dans [**17**] une formule plus simple pour la dimension 2.)

Preuve de la proposition 3.1. — Sans perte de généralité on suppose $\lambda = W(a,b)$, avec $a, b \in \mathbb{R}^n$, $|a| = |b| = r > 0$, $b \neq \pm a$. Notons $\partial_i = \partial/\partial x^i$, $\partial_t = \partial/\partial t$. On définit alors $A = (A_v, A_M)$, avec, pour $1 \leq i, j \leq n$,

$$\begin{cases} A_v^i(\nabla) = \sum_{k,\ell} (a^i b^k - b^i a^k)\, \partial_{k\ell\ell} \\ A_M^{ij}(\nabla) = \sum_k (b^i a^k - a^i b^k)\, \partial_{tkj} + \sum_k (b^j a^k - a^j b^k)\, \partial_{tki}, \end{cases}$$

où les indices k et ℓ varient dans $\{1, \dots, n\}$.

Soient maintenant $f \in C^\infty(\mathbb{R}^{n+1})$, $v_f = A_v(\nabla) f$, $M_f = A_M(\nabla) f$. Par calcul direct,

- $\operatorname{tr} M = 2 \sum_{i,k} (b^i a^k - a^i b^k)\, \partial_{tki} f = 0$ (par échange de i et k) ;
- $(\partial_t v + \nabla \cdot M)^j =$
$\sum_{\ell,k} (a^j b^k - b^j a^k)\, \partial_{kt\ell\ell} f + \sum_{i,k} (b^i a^k - a^i b^k)\, \partial_{ktij} f + \sum_{i,k} (b^j a^k - a^j b^k)\, \partial_{ktii} f = 0$;
- $\nabla \cdot v = \sum_{i,k,\ell} (a^i b^k - b^i a^k)\, \partial_{k\ell\ell i} f = 0.$

Ceci prouve (i). Ensuite, si $f(x,t) = \psi(\frac{(a+b)\cdot x + Kt}{\varepsilon})$, alors par calcul explicite

$$(v_f, M_f)(x,t) = \varepsilon^{-3} \left(2(a-b)\, s^2,\ -2(a \otimes a - b \otimes b)\, sK \right) \psi''' \left(\frac{(a+b)\cdot x + Kt}{\varepsilon} \right),$$

où $s = |a+b|^2/2 = r^2 + a \cdot b$. Il suffit de choisir $K = -s$, $\varepsilon = (2s^2)^{1/3}$, $\xi = (a+b)/\varepsilon$, $c = K/\varepsilon$ pour obtenir (ii). □

3.2. Localisation

On désigne par $|A|$ la mesure de Lebesgue d'un ensemble borélien $A \subset \mathbb{R}^n$.

PROPOSITION 3.2. — *Soient O un ouvert borné non vide de $\mathbb{R}^n$, $I =]t_0, t_1[\subset \mathbb{R}$, $r > 0$, $\lambda = (\overline{v}, \overline{M}) \in \Lambda_r$, et $\mathcal{V}$ un voisinage de $[-\lambda, \lambda] \subset \mathbb{R}^n \times \mathcal{S}_0^n$. Soient O' un ouvert non vide de $\mathbb{R}^n$ tel que $\overline{O'} \subset O$, $\theta \in]0, (t_1 - t_0)/2[$ et $I_\theta = [t_0 + \theta, t_1 - \theta]$. Alors pour tout $\eta > 0$ il existe $z = (v, M) \in C_c^\infty(O \times I; \mathcal{V})$ tel que*

- *$(v, M, 0)$ est solution de (3) ;*
- *$\sup_{t \in I} \|v(\,\cdot\,, t)\|_{H^{-1}(O)} \leq \eta$;*
- *$\inf_{t \in I_\theta} \dfrac{1}{|O'|} \displaystyle\int_{O'} |v(x,t)|^2 \, dx \geq \dfrac{|\overline{v}|^2}{3}$.*

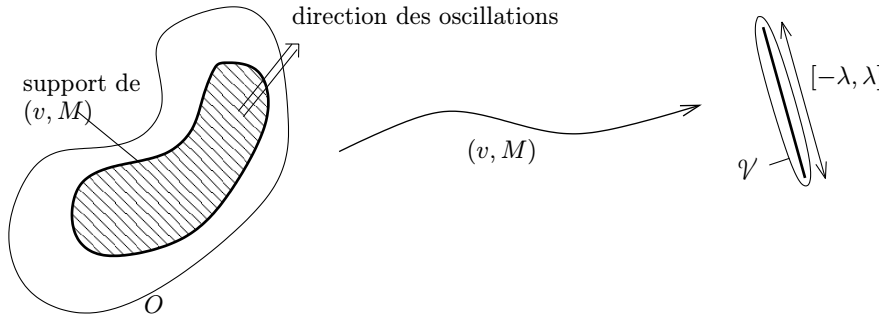


FIGURE 3. Ondes localisées à grande fréquence, d'après la proposition 3.2.

Preuve de la proposition 3.2. — Étant donné λ, soient A et (ξ, c) comme dans l'énoncé de la proposition 3.1. Soit $\varphi = \varphi(x,t)$ une fonction plateau C^∞ à support compact dans $O \times I$, $0 \leq \varphi \leq 1$, φ identiquement égale à 1 sur $O' \times I_\theta$. On pose (avec un léger abus de notation)

$$z_\varepsilon(x,t) = (v_\varepsilon, M_\varepsilon)(x,t) = A(\nabla) \left[\varepsilon^3 \, \varphi(x,t) \, \cos\left(\frac{\xi \cdot x + c\,t}{\varepsilon} \right) \right].$$

Cette fonction est bien sûr à support compact dans $O \times I$, et par la proposition 3.1 (i), $(v_\varepsilon, M_\varepsilon, 0)$ résout (3). Par la formule de Leibniz,

$$z_\varepsilon(x,t) = \varepsilon^3 \, \varphi \, A(\nabla) \, \cos\left(\frac{\xi \cdot x + c\,t}{\varepsilon} \right) + \sum_k \varepsilon^3 \, B_k(\nabla) \, \varphi \; C_k(\nabla) \, \cos\left(\frac{\xi \cdot x + c\,t}{\varepsilon} \right),$$

où les B_k et C_k sont des opérateurs matriciels, les C_k étant homogènes de degré 0, 1 ou 2. D'après la proposition 3.1 (ii), $A(\nabla)\cos(\frac{\xi\cdot x+ct}{\varepsilon}) = \lambda\,\varepsilon^{-3}\sin(\frac{\xi\cdot x+ct}{\varepsilon})$; on conclut que

$$\left\| z_\varepsilon(x,t) - \lambda\,\varphi(x,t)\,\sin\left(\frac{\xi\cdot x + c\,t}{\varepsilon}\right)\right\|_{L^\infty(O\times I)} = O(\varepsilon).$$

En choisissant ε assez petit, on est assuré que z_ε prend ses valeurs dans le voisinage $\mathscr{V}$ de $[-\lambda,\lambda]$.

Par ailleurs, $z_\varepsilon(x,t) = \lambda\,\sin(\frac{\xi\cdot x+ct}{\varepsilon})$ dès que $(x,t)\in O'\times I_\theta$; donc

$$\frac{1}{|O'|}\int_{O'} |v_\varepsilon(x,t)|^2\,dx = |\overline{v}|^2\left(\frac{1}{|O'|}\int_{O'}\sin^2\left(\frac{\xi\cdot x+c\,t}{\varepsilon}\right)\,dx\right).$$

Par un calcul classique, le terme entre parenthèses converge vers $1/2$ quand $\varepsilon\to 0$, *uniformément en* t. On conclut que pour ε assez petit, $\frac{1}{|O'|}\int_{O'}|v(x,t)|^2\,dx \geq |\overline{v}|^2/3$.

Soit maintenant $\zeta\in H^1(O)$; en utilisant la définition de z_ε et en intégrant une fois par parties dans les variables x^i, on voit que *pour chaque* t, on a

$$\left|\int z_\varepsilon(x,t)\zeta(x)\,dx\right| \leq O(\varepsilon)\int|\nabla\zeta|\,dx \leq O(\varepsilon)\,\|\zeta\|_{H^1(O)}.$$

On en déduit que $\|z_\varepsilon\|_{H^{-1}(O)} = O(\varepsilon)$. Finalement, $z = z_\varepsilon$ remplit toutes les conditions requises pour ε assez petit. □

4. INTÉGRATION CONVEXE

Le principe général de l'intégration convexe, développée pour les équations aux dérivées partielles par Gromov **[28]**, et pour les équations différentielles ordinaires par Filippov **[25]**, se résume ainsi : étant donnée une équation non linéaire $\mathscr{E}(u)$,

(i) on réécrit $\mathscr{E}(u)$ sous la forme $\mathscr{L}(z)\wedge(z\in K)$, où $\mathscr{L}$ est une équation linéaire ;

(ii) on introduit un z_0 sous-solution stricte du système, c'est-à-dire vérifiant $\mathscr{L}(z_0)$ et $z_0\in\operatorname{int}K^{\mathrm{co}}$, où K^{co} désigne l'enveloppe convexe, en un sens adéquat (par exemple l'enveloppe Λ-convexe) de K ;

(iii) on construit une suite $(z_k)_{k\in\mathbb{N}}$, définie récursivement par des opérations linéaires, qui s'approche de K tout en restant dans $\operatorname{int}K^{\mathrm{co}}$;

(iv) on passe à la limite, après avoir éventuellement modifié la suite (z_k) pour assurer sa convergence en un sens approprié.

De nombreux exemples sont traités dans les ouvrages **[28, 47, 49]**.

Dans notre cas, l'étape (i) a été réalisée dans la sous-section 2.1, et la proposition 2.6 nous autorise à choisir $z_0 = 0$. Il reste à définir précisément l'espace des sous-solutions, construire une suite s'approchant du bord de cet espace, et passer à la limite ; c'est ce que nous allons faire dans cette section.

4.1. Espace de sous-solutions

Avec les notations du théorème 1.4, et la notation 2.2, posons

$$X_0 = \Big\{ z = (v, M) \in C_c^\infty\big(\Omega\times]0,T[\,;\mathbb{R}^n \times \mathscr{S}_0^n\big), \quad \text{solution de (3) avec } q = 0\,; \\ \forall (x,t) \in \Omega\times]0,T[\,,\ z(x,t) \in \operatorname{int} K^{\rm co}_{\sqrt{2\,\overline{e}(x,t)}} \Big\}. \tag{9}$$

Ce sera l'espace des sous-solutions strictes. Notons que $z \in X_0$ sera considéré comme défini *pour tous les temps*, même s'il est bien sûr nul en dehors de $]0,T[$.

On définit en outre, pour $\tau \in]0,T/2[$ et Ω_0 un ouvert d'adhérence compacte incluse dans Ω,

$$J_{\tau,\Omega_0}(v) = \sup_{\tau\le t\le T-\tau} \int_{\Omega_0} \left[\overline{e}(x,t) - \frac{|v(x,t)|^2}{2}\right] dx. \tag{10}$$

Cette fonctionnelle (ou plutôt cette famille de fonctionnelles, quand τ et Ω_0 varient) quantifie l'éloignement de z du bord de X_0. On note que $J_{\tau,\Omega_0}(v) \ge 0$ si $z = (v, M) \in X_0$.

Dans la suite, si X est un espace métrique, je désignerai par $C(]0,T[\,;X)$ l'espace des fonctions continues $]0,T[\to X$, muni de la topologie de la convergence *localement* uniforme sur $]0,T[$; et par $L^2_{\rm loc}$ l'espace des fonctions localement carré-intégrables, muni de sa topologie naturelle.

PROPOSITION 4.1. — *Soient Ω un ouvert de $\mathbb{R}^n$, et $T > 0$; soit X_0 défini par* (9)*. Alors*

(i) *Si $z = (v, M)$ appartient à X_0 et $p = -|v|^2/n$, alors (v, p) est solution de l'équation d'Euler* (1) *avec forçage $f = \nabla\cdot(v\otimes v - \frac{|v|^2}{n} I_n - M) \in C_c^\infty(\Omega\times]0,T[\,;\mathbb{R}^n)$.*

(ii) *Soit $(v_k, M_k)_{k\in\mathbb{N}}$ une suite à valeurs dans X_0, convergeant dans $\mathscr{D}'(\mathbb{R}^n\times\mathbb{R})$ et dans $C(]0,T[\,;L^2_{\rm loc}(\Omega))$ vers $z = (v, M)$, telle que $J_{\tau,\Omega_0}(v_k) \to 0$ pour tout $\tau > 0$ et pour tout ouvert Ω_0 d'adhérence compacte incluse dans Ω. Alors v est une solution faible de* (2)*, et vérifie*

$$\begin{cases} v \in C(\mathbb{R}; L^2_w(\mathbb{R}^n)); \\ \dfrac{|v(x,t)|^2}{2} = \overline{e}(x,t) = -\dfrac{n}{2}\, p(x,t) \qquad \text{pour tout } t \in]0,T[\text{ et presque tout } x \in \Omega; \\ v \text{ est identiquement nulle en dehors de } \Omega\times]0,T[. \end{cases}$$

Preuve de la proposition 4.1. — L'énoncé (i) est évident.

Avant de démontrer (ii), notons que, d'après la définition de K_r et la preuve de la proposition 2.6, on a $|v_k(x,t)|^2 \le 2\,\overline{e}(x,t)$ et $|M_k(x,t)| \le C(n)\,\overline{e}(x,t)\,I_n$ presque partout ; $(z_k)_{k\in\mathbb{N}}$ est donc globalement borné dans $L^\infty(\mathbb{R}; L^2(\mathbb{R}^n;\mathbb{R}^n) \times L^1(\mathbb{R}^n;\mathscr{S}_0^n))$. On déduit de l'équation (3) que (v_k) est précompacte dans $C(\mathbb{R}; W^{-1,1}(\mathbb{R}^n))$; quitte

à extraire une sous-suite on a donc $v_k \to v$ dans cet espace, et $M_k \to M$ au sens des distributions. En particulier, $v \in C(\mathbb{R}; \mathscr{D}'(\mathbb{R}^n)) \cap L^\infty(\mathbb{R}; L^2(\mathbb{R}^n))$, et il s'ensuit que $v \in C(\mathbb{R}; L^2_w(\mathbb{R}^n))$.

Maintenant, pour prouver que v est solution faible de l'équation d'Euler, on utilise la proposition 2.1 ; le passage à la limite dans (3) ne pose pas de problème car les équations sont linéaires, il suffit donc de vérifier la condition (4).

Pour tout $k \in \mathbb{N}$ on a $|v_k(x,t)|^2 \leq 2\,\overline{e}(x,t)$ en tout $(x,t) \in \Omega\times]0,T[$. Comme $v_k \longrightarrow v$ dans $C(]0,T[\,; L^2_{\mathrm{loc}}(\Omega))$ et que $\overline{e}$ est continue, on déduit de résultats classiques sur la convergence faible que $|v(x,t)|^2 \leq 2\,\overline{e}(x,t)$ pour tout $t \in]0,T[$ et presque tout $x \in \Omega$.

De même, si l'on pose $\Phi(v,M) = \sigma_{\max}(v\otimes v - M)$ comme dans la preuve de la proposition 2.6, la convexité de Φ entraîne $\Phi(v,M) \leq \liminf \Phi(v_k, M_k)$, d'où $\Phi(v,M) \leq \overline{e}$, c'est-à-dire $(v,M) \in K^{\mathrm{co}}_{\sqrt{2\,\overline{e}}}$, pour presque tout $(x,t) \in \Omega\times]0,T[$.

Soient $\tau \in]0,T/2[$ et Ω_0 un ouvert compactement inclus dans Ω. Par hypothèse, pour tout $t \in]\tau, T-\tau[$,

$$\int_{\Omega_0} \left[\overline{e}(x,t) - \frac{|v_k(x,t)|^2}{2}\right] dx \xrightarrow[k\to\infty]{} 0,$$

et par convergence forte (dans $C([\tau, T-\tau]; L^2(\Omega_0))$) on peut passer à la limite pour trouver $\frac{|v(x,t)|^2}{2} = \overline{e}(x,t)$ pour presque tout $x \in \Omega_0$. Grâce à la proposition 2.6 (i) on a donc $(v,M)(x,t) \in K_{\sqrt{2\,\overline{e}(x,t)}}$. Ceci étant vrai pour tout choix de τ et Ω_0, on a $(v,M)(x,t) \in K$ pour tout $t \in]0,T[$ et presque tout $x \in \Omega$.

Par ailleurs, $v(x,t)$ s'annule identiquement pour t en dehors de $[0,T]$. La convergence dans $C(\mathbb{R}; L^2_w(\mathbb{R}^n))$ garantit que v est continue en temps à valeurs distributions, d'où $v(\,\cdot\,,0) = v(\,\cdot\,,T) = 0$. Soit $t \in]0,T[$; alors, comme $|v(x,t)|^2 = 2\,\overline{e}(x,t)$ pour presque tout $x \in \Omega$, on a $\int_\Omega |v(x,t)|^2\,dx \leq \int |v(x,t)|^2\,dx \leq \liminf |v_k(x,t)|^2\,dx \leq 2\int_\Omega \overline{e}(x,t)\,dx = \int_\Omega |v(x,t)|^2\,dx$, ce qui force $v(x,t) = 0$ presque partout en dehors de Ω. (Il était clair a priori que v s'annulait identiquement en dehors de $\overline{\Omega}$, mais nous voulions une information plus précise.)

On conclut que (quitte à le redéfinir sur un ensemble négligeable) (v,M) est partout à valeurs dans K, et par la proposition 2.1, v est bien une solution de l'équation d'Euler. Ceci achève la preuve. □

4.2. Amélioration

Soit $z \in X_0$; nous allons voir que l'on peut l'approcher au sens faible par un z' qui est « plus près du bord de X_0 » que z.

PROPOSITION 4.2 (Étape d'amélioration). — *Avec les notations précédentes, soient* $z = (v, M) \in X_0$, $\tau \in]0, T/2[$ *et* Ω_0 *un ouvert d'adhérence compacte incluse dans* Ω. *On suppose que*

$$J_{\tau,\Omega_0}(v) \geq \alpha > 0$$

et on se donne $\eta > 0$. *Alors il existe* $z' = (v', M') \in X_0$ *tel que*

- $\|z' - z\|_{C([0,T];H^{-1}(\Omega))} \leq \eta$;
- $J_{\tau,\Omega_0}(v') \leq J_{\tau,\Omega_0}(v) - \beta(\alpha)$,

où $\beta(\alpha) = \beta_{\tau,\Omega_0}(\alpha)$ *est une fonction positive strictement croissante de* α.

Remarque 4.3. — Comme on va le voir, on peut donner une formule assez explicite pour la fonction $\beta(\alpha)$, qui se comporte comme α^2 pour de petites valeurs de α **[15]**. Cette information est inutile pour la preuve du théorème 1.4, mais pourrait s'avérer importante pour l'étude quantitative des solutions construites par ce théorème.

Preuve de la proposition 4.2. — L'idée consiste à identifier les endroits où $\overline{e} - \frac{|v|^2}{2}$ est grand, et à y ajouter des oscillations rapides dans des directions autorisées, de façon à faire augmenter $\frac{|v|^2}{2}$ en moyenne. Sans perte de généralité on supposera $\alpha \in]0, 1]$.

On divise $\Omega_0 \times [\tau, T - \tau]$ selon un pavage régulier dans $\mathbb{R}^n \times \mathbb{R}$, avec décalage d'un demi-carreau d'une colonne à l'autre, de sorte que chaque ligne $\{t = \text{constante}\}$ ne rencontre que la moitié des jointures. (Voir la Figure 4 ; c'est à dessein que je reste informel, les formules précises auraient pour résultat probable d'emplir le lecteur de confusion.) Oubliant ce qui se passe près des bords, je ferai comme si $\Omega_0 \times [\tau, T - \tau]$ était une union de cubes C de côté ε.

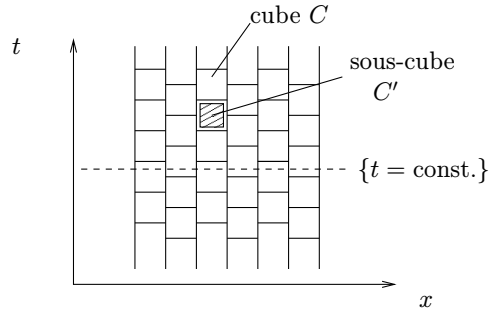


FIGURE 4. Pavage de $\Omega_0 \times [\tau, T - \tau]$

Dans chaque cube C on définit un sous-cube C', de même centre que C, de côté $(0,9)^{1/n}\varepsilon$. Comme v et $\overline{e}$ sont uniformément continues sur $\Omega_0 \times [\tau, T - \tau]$, l'oscillation de $\overline{e} - |v|^2/2$ sur chaque cube C est majorée par $\alpha/10$ pour ε assez petit. De plus,

toute ligne $\{t = t_0\}$ intersecte (aux bords près...) au moins un sur deux des cubes C' tels que C intersecte $\{t = t_0\}$, soit une proportion supérieure à (disons) $1/4$ des $x \in \Omega_0$. Un argument de type « somme de Riemann » montre alors que pour ε assez petit (dépendant de α),

$$\int_{\Omega_0} \left[\overline{e}(x,t_0) - \frac{|v(x,t_0)|^2}{2}\right] dx \leq 4 \sum_{C' \cap \{t=t_0\} \neq \varnothing} |C'| \left(\overline{e}_C - \frac{|v_C|^2}{2}\right) + \frac{\alpha}{10}, \tag{11}$$

où $\overline{e}_C$ (resp. v_C) est la valeur de $\overline{e}$ (resp. v) en le centre du cube C. Notons que $|C'| = O(\varepsilon^n)$ pour $\varepsilon \to 0$.

Pour tout $c > 0$, la contribution au membre de droite de (11) des cubes C tels que $\overline{e}_C - \frac{|v_C|^2}{2} \leq c\,\alpha$ est grossièrement majorée par $4|C'|\mathscr{N} c\alpha$, où $\mathscr{N} = O(1/\varepsilon^n)$ est le nombre maximal de cubes intersectant $\{t = t_0\}$; en choisissant $c > 0$ (indépendant de ε) tel que $c \leq 1/(40|C'|\mathscr{N})$ on obtient

$$\int_{\Omega_0} \left[\overline{e}(x,t_0) - \frac{|v(x,t_0)|^2}{2}\right] dx \leq 4 \sum_C \left\{|C'| \left(\overline{e}_C - \frac{|v_C|^2}{2}\right);\ \overline{e}_C - \frac{|v_C|^2}{2} \geq c\,\alpha\right\} + \frac{\alpha}{5}.$$

En particulier,

$$\sum_C \left\{|C'| \left(\overline{e}_C - \frac{|v_C|^2}{2}\right);\ \overline{e}_C - \frac{|v_C|^2}{2} \geq c\,\alpha\right\} \geq \frac{\alpha}{5}. \tag{12}$$

On cherche maintenant un couple (v', M') qui fasse mieux que (v, M), sous la forme

$$(v', M') = (v, M) + \sum_C (\widetilde{v}_C, \widetilde{M}_C),$$

où chaque $z_C = (\widetilde{v}_C, \widetilde{M}_C)$ est à support compact dans C. Si $\overline{e}_C - \frac{|v_C|^2}{2} \leq c\,\alpha$, on pose $\widetilde{z}_C = 0$. Le gain au niveau de la fonctionnelle J_{τ,Ω_0} (au temps t_0) sera donc

$$\begin{aligned}
\Delta &:= \int_{\Omega_0} \left[\overline{e}(x,t_0) - \frac{|v'(x,t_0)|^2}{2}\right] dx - \int_{\Omega_0} \left[\overline{e}(x,t_0) - \frac{|v(x,t_0)|^2}{2}\right] dx \\
&= \int_{\Omega_0} \left(\frac{|v'(x,t_0)|^2}{2} - \frac{|v(x,t_0)|^2}{2}\right) dx \\
&= \sum_{\overline{e}_C - \frac{|v_C|^2}{2} \geq c\,\alpha} \left(\int_C \frac{|\widetilde{v}_C(x,t_0)|^2}{2}\, dx + \int_C v(x,t_0) \cdot \widetilde{v}_C(x,t_0)\, dx\right) \\
&\geq \sum_{\overline{e}_C - \frac{|v_C|^2}{2} \geq c\,\alpha} \left(\int_{C'} \frac{|\widetilde{v}_C(x,t_0)|^2}{2}\, dx - \|v(\,\cdot\,,t_0)\|_{H^1}\, \|\widetilde{v}_C(\,\cdot\,,t_0)\|_{H^{-1}(C)}\right).
\end{aligned}$$

On imposera $\|\widetilde{v}_C(\,\cdot\,,t_0)\|_{H^{-1}(C)} \leq \delta/(\mathscr{N}\,\|v(\,\cdot\,,t_0)\|_{H^1})$, de sorte que

$$\Delta \geq \sum_{\overline{e}_C - \frac{|v_C|^2}{2} \geq c\,\alpha} \int_{C'} \frac{|\widetilde{v}_C(x,t_0)|^2}{2}\, dx - \delta, \tag{13}$$

avec δ aussi petit que l'on souhaite.

Si l'on peut imposer également

$$\frac{1}{|C'|}\int_{C'}\frac{|\widetilde{v}_C(x,t_0)|^2}{2}\,dx \geq \frac{\beta}{\alpha}\left(\overline{e}_C - \frac{|v_C|^2}{2}\right),$$

alors en reportant dans (13) et en choisissant δ suffisamment petit, on obtiendra la borne $\Delta \geq \sum \frac{|C'|}{2}\,\frac{\beta}{\alpha}\,(\overline{e}_C - \frac{|v_C|^2}{2})$, d'où $\Delta \geq \beta/10$ par (12).

Récapitulons : étant donné un cube C, tel que $\overline{e}_C - \frac{|v_C|^2}{2} \geq c\,\alpha$, il suffit de construire $(\widetilde{v}_C, \widetilde{M}_C)$ à support compact dans C, tel que

(a) $(v+\widetilde{v}_C, M+\widetilde{M}_C) \in \operatorname{int} K^{\rm co}_{\sqrt{2\overline{e}}}$ en tout $(x,t)\in C$;

(b) $\dfrac{1}{|C'|}\displaystyle\int_{C'}\frac{|\widetilde{v}_C(x,t_0)|^2}{2}\,dx \geq \frac{\beta}{\alpha}\left(\overline{e}_C - \frac{|v_C|^2}{2}\right)$;

(c) $\sup_t \|\widetilde{v}_C(\,\cdot\,,t)\|_{H^{-1}} \leq \eta$, arbitrairement petit.

Soit $\rho = \inf\left\{\operatorname{dist}\left(z(x,t),\,\partial K^{\rm co}_{\sqrt{2\overline{e}(x,t)}}\right);\ (x,t)\in\Omega_0\times[\tau,T-\tau]\right\} > 0$. La proposition 2.6 (ii) fournit $\lambda = (\overline{v},\overline{M})$ tel que $z_C + [-\lambda,\lambda]$ est inclus dans $K^{\rm co}_{\sqrt{2\overline{e}_C}}$, à une distance au moins $\rho/2$ du bord, et $|\overline{v}| \geq$ const. $(\overline{e}_C - |v_C|^2/2)^2/\|\overline{e}\|_\infty$. Comme z et $K^{\rm co}_{\sqrt{2\overline{e}}}$ varient uniformément continûment en fonction de (x,t), quitte à réduire encore ε on peut supposer que $z(x,t)+[-\lambda,\lambda]$ est inclus dans l'intérieur de $K^{\rm co}_{\sqrt{2\overline{e}(x,t)}}$ pour tout $(x,t)\in C$. Il existe donc un voisinage $\mathcal{V}$ de $[-\lambda,\lambda]$ tel que $z(x,t)+\mathcal{V} \subset \operatorname{int} K^{\rm co}_{\sqrt{2\overline{e}(x,t)}}$ pour tout $(x,t)\in C$. On construit alors par la proposition 3.2 une onde $\widetilde{z}_C$ localisée, à support dans C, à valeurs dans $\mathcal{V}$, telle que

$$\frac{1}{|C'|}\int_{C'}|\widetilde{v}_C(x,t)|^2\,dx \geq \frac{|\overline{v}|^2}{3} \geq \text{const.}\left(\overline{e}_C - \frac{|v_C|^2}{2}\right)^2 \geq \text{const.}\,\alpha\left(\overline{e}_C - \frac{|v_C|^2}{2}\right),$$

et $\sup_t \|\widetilde{v}_C(\,\cdot\,,t)\|_{H^{-1}}$ est arbitrairement petit.

Ceci achève la preuve avec $\beta = \text{const.}\,\alpha^2$. □

Avec une démonstration presque identique, on prouve la variante suivante de la proposition 4.2 :

PROPOSITION 4.4 (Amélioration sur plusieurs sous-ouverts)

Soient $z=(v,M)\in X_0$, $(\tau_\ell)_{1\leq\ell\leq L}$ et $(\Omega_\ell)_{1\leq\ell\leq L}$ vérifiant $0<\tau_{\ell+1}<\tau_\ell<T/2$ et $\overline{\Omega}_\ell \subset \Omega_{\ell+1} \subset \overline{\Omega}_{\ell+1} \subset \Omega$ pour tout $\ell\in\{1,\dots,L\}$. On suppose que

$$J_{\tau_\ell,\Omega_\ell}(v) \geq \alpha_\ell > 0,$$

et on se donne $\eta>0$. Alors il existe $z'=(v',M')\in X_0$ tel que

- $\|z'-z\|_{C([0,T];H^{-1}(\Omega))} \leq \eta$;
- $\forall \ell\in\{1,\dots,L\},\ \ J_{\tau_\ell,\Omega_\ell}(v') \leq J_{\tau_\ell,\Omega_\ell}(v) - \beta_\ell(\alpha_\ell)$,

où $\beta_\ell(\alpha) = \beta_{\tau_\ell,\Omega_\ell}(\alpha)$ est une fonction positive strictement croissante de α.

4.3. Itération

On peut maintenant démontrer le théorème 1.4, en utilisant de manière répétée la proposition 4.2 pour construire une suite $(z_k)_{k\in\mathbb{N}}$ qui s'approche du bord de X_0. Pour assurer la convergence (localement) *forte* on séparera les échelles en régularisant dans le style de Nash. (Le formalisme est proche de [**37**, Section 3.3].)

Preuve du théorème 1.4. — Soit $(\rho_\varepsilon)_{\varepsilon>0}$ un noyau régularisant en espace-temps, construit selon le procédé habituel : $\rho_\varepsilon(x,t) = \varepsilon^{-(n+1)}\,\psi(x/\varepsilon)\,\chi(t/\varepsilon)$, où ψ et χ sont des fonctions positives C^∞ d'intégrale unité, respectivement sur $\mathbb{R}^n$ et $\mathbb{R}$, à support dans la boule unité. Si f est une fonction $\mathbb{R}^{n+1}\to\mathbb{R}$, on notera $f*\rho_\varepsilon$ le produit de convolution de f par ρ_ε sur $\mathbb{R}^{n+1}$.

On se donne $(\tau_j)_{j\in\mathbb{N}}$ une suite de nombres positifs dans $]0,T/2[$, décroissante et tendant vers 0 ; ainsi qu'une suite $(\Omega_j)_{j\in\mathbb{N}}$ d'ouverts bornés tels que $\overline{\Omega}_j \subset \Omega_{j+1}$ et $\bigcup \Omega_j = \Omega$. On note $\beta_j = \beta_{\tau_j,\Omega_j}$ la fonction strictement croissante apparaissant dans la proposition 4.4.

On pose $z_0 = 0$, $\varepsilon_0 = 1$. Ayant construit $z_j = (v_j, M_j) \in X_0$ et $\varepsilon_j > 0$ pour $j \le k-1$, on applique la proposition 4.4 pour trouver $z_k = (v_k, M_k)$ satisfaisant

- $J_{\tau_j,\Omega_j}(v_k) \le J_{\tau_j,\Omega_j}(v_{k-1}) - \beta_j(J_{\tau_j,\Omega_j}(v_{k-1}))$ pour tout $j \le k$;
- $\sup_t \|(z_k - z_{k-1})*\rho_{\varepsilon_j}\|_{L^2(\Omega)} < 2^{-k}$ pour tout $j \le k-1$;
- $\sup_t \|z_k - z_{k-1}\|_{H^{-1}(\Omega)} < \eta\, 2^{-k}$.

Puis on choisit ε_k tel que $\|z_j - z_j*\rho_{\varepsilon_k}\|_{L^2} < 2^{-k}$ pour tout $j \le k$.

La suite (z_k) est globalement bornée dans $L^2(\mathbb{R}^n\times\mathbb{R})$, et donc (à extraction près) converge au sens des distributions vers $z \in L^2(\mathbb{R}^n\times\mathbb{R})$. En particulier, $z_\ell * \rho_{\varepsilon_k} \to z*\rho_{\varepsilon_k}$ quand $\ell\to\infty$, k étant fixé.

Il est clair que $\sup_t \|z\|_{H^{-1}(\Omega)} \le \sum_k \sup_t \|z_k - z_{k-1}\|_{H^{-1}(\Omega)} \le \eta$.

Soit maintenant j fixé. Pour tout $k \ge j$ on a, dans $C([\tau_j, T-\tau_j]; L^2(\Omega_j))$,

$$\|z_k - z\| \le \|z_k - z_k*\rho_{\varepsilon_k}\| + \|z_k*\rho_{\varepsilon_k} - z*\rho_{\varepsilon_k}\| + \|z*\rho_{\varepsilon_k} - z\|. \tag{14}$$

Par construction, le premier terme du membre de droite de (14) est majoré par 2^{-k} ; quant au dernier, il converge vers 0 quand $k\to\infty$. Pour estimer le second, on écrit

$$\big\|z_k*\rho_{\varepsilon_k} - z*\rho_{\varepsilon_k}\big\| = \lim_{\ell\to\infty}\big\|z_k*\rho_{\varepsilon_k} - z_\ell*\rho_{\varepsilon_k}\big\| \le \sum_{m\ge k}\big\|(z_m - z_{m+1})*\rho_{\varepsilon_k}\big\| \le 2^{-k}.$$

Il s'ensuit que z_k converge vers z dans $C([\tau_j, T-\tau_j]; L^2(\Omega_j))$, et donc dans $C(]0,T[\,; L^2_{\mathrm{loc}}(\Omega))$ (convergence localement uniforme en t) puisque j est arbitraire.

Soit $\ell_j = \limsup J_{\tau_j,\Omega_j}(z_k)$; si $\ell_j \ge \alpha > 0$, alors en passant à la limite dans l'inégalité d'amélioration de J on trouve $\ell_j \le \ell_j - \beta_j(\alpha) < \ell_j$, ce qui est absurde. On a donc $\lim_{k\to\infty} J_{\tau_j,\Omega_j}(z_k) = 0$. Ceci étant vrai pour tout j, on a $\lim_{k\to\infty} J_{\tau,\Omega_0}(z_k) = 0$ pour tout $\tau > 0$ et tout ouvert Ω_0 compactement inclus dans Ω. Il suffit maintenant d'appliquer la proposition 4.1 pour conclure la preuve du théorème 1.4. □

5. ARGUMENT DE CATÉGORIE DE BAIRE

La théorie de Baire est le point de départ d'une méthode de résolution des inclusions différentielles, alternative à l'intégration convexe. Après les travaux pionniers de Cellina [**6**], De Blasi et Pianigiani [**14**], Bressan et Flores [**5**], cette méthode a été amenée à un haut degré de sophistication par Dacorogna et Marcellini (voir [**11**] et les nombreux travaux qui ont suivi). Le livre de Dacorogna et Marcellini [**12**] présente un exposé complet de la théorie, en partie motivée par le désir d'aller au-delà des limitations de la méthode des solutions de viscosité.

Les deux stratégies (intégration convexe et théorie de Baire) fournissent des résultats comparables pour quantité de problèmes, et comportent en fait de nombreux points communs ; ceci est particulièrement clair dans la présentation de Kirchheim [**29, 30**] (réécrite dans [**13**]), basée sur l'étude des points de continuité d'une fonction de première catégorie au sens de Baire (i.e. limite ponctuelle d'une suite de fonctions continues). Les liens entre ces deux méthodes sont également analysés de manière synthétique par Sychev [**48**].

Certains énoncés d'intégration convexe, comme le théorème de Nash–Kuiper, ont jusqu'à présent résisté aux tentatives d'approche topologique « à la Baire » ; il en va de même de certains problèmes dans lesquels les considérations topologiques (au sens géométrique) ou la régularité jouent un rôle important.[7]

En revanche, le théorème de Scheffer–Shnirelman s'y prête fort bien, et c'est d'ailleurs la méthode privilégiée par De Lellis et Székelyhidi [**15, 17**]. Je vais expliquer le cœur de l'argument :

Ébauche de preuve topologique du théorème 1.4. — Soit X l'adhérence de X_0 pour la topologie $C(\mathbb{R}; \mathscr{D}'(\Omega)) \cap C([0,T[\,; L^2_{\rm loc}(\Omega))$. Pour tous $\tau > 0$ et Ω' ouvert d'adhérence compacte dans Ω, la fonctionnelle $J_{\tau,\Omega'}$ est semi-continue supérieurement sur X ; on en déduit que ses points de continuité forment un ensemble gras au sens de Baire (i.e. contenant une intersection dénombrable d'ouverts denses). Soit z un tel point de continuité : si l'on n'avait pas $z \in K$ pour tout $t \in]\tau, T-\tau[$ et presque tout $x \in \Omega'$, on pourrait utiliser la proposition 4.2 pour construire une suite (z_k), convergeant vers z dans X, telle que $\limsup J_{\tau,\Omega'}(z_k) < J_{\tau,\Omega'}(z)$, ce qui serait absurde.

Soient maintenant $(\tau_j)_{j\in\mathbb{N}}$ une suite de nombres positifs décroissant vers 0, et $(\Omega_j)_{j\in\mathbb{N}}$ une suite croissante exhaustive d'ouverts dans Ω. L'ensemble des points de

[7] Comme le remarque Gromov, il est intéressant de comparer la méthode de Baire et l'intégration convexe dans le problème de la construction d'applications isométriques lipschitziennes entre variétés de dimension égale : la première méthode fournit une application dont la différentielle est presque partout une isométrie, tandis que la seconde fournit une « vraie » isométrie pour la distance intrinsèque. Même dans cet énoncé où la régularité n'est pas meilleure que lipschitzienne, l'intégration convexe fournit donc en un certain sens plus de régularité.

continuité de toutes les fonctionnelles J_{τ_j,Ω_j} est encore gras ; et si $z = (v, M)$ est un tel point de continuité, le raisonnement précédent montre que z est à valeurs dans K, pour tout t et presque tout x. Par la proposition 2.1, v est une solution de l'équation d'Euler (4). □

Par rapport à la technique itérative de la sous-section 4.3, l'argument topologique a l'avantage de fournir automatiquement *une infinité* de solutions distinctes ; en outre il est légèrement plus simple à mettre en œuvre. En contrepartie, l'intégration convexe est plus constructive et laisse la porte ouverte à une étude qualitative et quantitative plus fine.

6. PROLONGEMENTS ET PERSPECTIVES

6.1. Unicité

L'unicité de la solution d'une équation aux dérivées partielles, pour une condition initiale donnée, est une question essentielle. Pour l'équation d'Euler incompressible, le principal résultat en la matière est le célèbre théorème de Yudovich : unicité en dimension 2 dans la classe des solutions de tourbillon (= rotationnel) borné.

Les solutions paradoxales de Scheffer et Shnirelman montrent que l'on ne peut espérer l'unicité de la solution des équations d'Euler incompressible pour une donnée initiale irrégulière (disons L^2 ou L^∞) à moins d'imposer des conditions supplémentaires. Mais quel critère de sélection utiliser ? Au vu de considérations physiques et de divers exemples d'équations aux dérivées partielles, il est naturel d'imposer, au choix et selon son tempérament :

— la préservation de la borne d'énergie : $\int |v(x,t)|^2\,dx \le \int |v(x,0)|^2\,dx$ (la condition la plus faible) ;

— la conservation globale de l'énergie : $\int |v(x,t)|^2\,dx = \int |v(x,0)|^2\,dx$;

— la décroissance globale de l'énergie : $\int |v(x,t)|^2\,dx \le \int |v(x,s)|^2\,dx \quad (t \ge s)$;

— la décroissance locale de l'énergie : $\partial_t \dfrac{|v|^2}{2} + \nabla \cdot \left[\left(\dfrac{|v|^2}{2} + p\right) v\right] \le 0$ (condition imposée par Shnirelman dans **[45]**) ;

— la conservation locale de l'énergie : $\partial_t \dfrac{|v|^2}{2} + \nabla \cdot \left[\left(\dfrac{|v|^2}{2} + p\right) v\right] = 0$ (la condition la plus forte).

En exploitant leur construction, De Lellis et Székelyhidi montrent qu'*aucun* de ces critères ne garantit l'unicité. Je vais exposer les grandes lignes de leur argumentation.

Tout d'abord, de même que dans le théorème de Nash–Kuiper on peut partir de n'importe quelle contraction stricte pour la déformer en une isométrie, dans le théorème 1.4 on peut partir de n'importe quelle « sous-solution stricte » de l'équation d'Euler pour la déformer en une vraie solution. Ici une sous-solution désigne un triplet (v_0, M_0, q_0), solution du système linéaire (3), de classe C^∞, à support strictement inclus dans Ω pour tout $t \in]0,T[$, et tel que $v_0 \otimes v_0 - M_0 < (2\,\overline{e}/n) I_n$ dans $\Omega \times]0,T[$. (Dans le théorème 1.4 nous nous contentions de choisir $(v_0, M_0, q_0) = 0$.)

Donnons-nous maintenant une fonction positive e sur $\mathbb{R}$, paire, telle que $e(0) = 1/2$ et e décroît pour $t > 0$ jusqu'à s'annuler en $t = 1$. On pose $\Omega = B^n(1)$. En modifiant légèrement la démonstration du théorème 1.4, on peut construire de telles sous-solutions, à support dans $\{-1 \leq t \leq 1\}$, telles que $|v_0(x,0)| = 1$ presque partout dans Ω, $(v,M)(x,t) \in \operatorname{int} K^{\rm co}_{\sqrt{2\,e(t)}}$ pour $t \neq 0$, et (v,M) est *lisse* en dehors de $\{t = 0\}$. (Le défaut de régularité ne survient que lorsqu'on touche le bord du convexe, ici en $t = 0$.)

Choisissons alors $v_0(\,\cdot\,,0)$ comme donnée initiale, et donnons-nous un nouveau profil d'énergie $\overline{e} \geq e$. En appliquant encore une démonstration similaire à celle du théorème 1.4, et l'argument de catégorie de Baire, on peut construire une infinité de solutions définies sur $\mathbb{R}^n \times \mathbb{R}$ (ou $\mathbb{R}^n \times I$, I étant un intervalle contenant 0), qui partent de la donnée initiale v_0 en $t = 0$, et vérifient $|v(x,t)|^2 = 2\,\overline{e}(x,t)$ pour tout $t > 0$ et presque tout $x \in \Omega$; et $v = 0$ en dehors de Ω. On a beaucoup de marge pour choisir le profil $\overline{e}$: la seule contrainte est qu'il soit strictement au-dessus de e pour $t \neq 0$ (à chaque étape de la preuve de De Lellis et Székelyhidi, l'énergie ne peut qu'augmenter ; de même qu'au cours de la preuve du théorème de Nash–Kuiper, les distances ne peuvent que croître). Ce profil $\overline{e}$ peut donc être constant, ou strictement décroissant ; on peut même choisir $\overline{e} = 1/2$ pour $|t| < 1$, et imposer $v \equiv 0$ en $|t| = 1$, puis répéter par périodicité, de façon à construire une solution d'Euler qui s'annule périodiquement...

Ces solutions mettent facilement en défaut tous les critères de sélection proposés ci-dessus. Par exemple, si l'on choisit $\overline{e} \equiv 1/2$, on a $|v(x,t)| = 1$ et $p(x,t) = -1/n$ pour tout t et presque tout $x \in \Omega$, donc $\partial_t \frac{|v|^2}{2} = 0$, $\nabla \cdot [(\frac{|v|^2}{2} + p)v] = \nabla \cdot [(\frac{|v|^2}{2} + p)(1_{x \in \Omega})v] = ((n-2)/(2n))\nabla \cdot v = 0$, et la conservation *locale* de l'énergie est vérifiée !

En fait, parmi tous les critères de sélection habituellement utilisés en mécanique des fluides, le seul dont on puisse encore espérer qu'il s'applique ici est celui de Dafermos, qui postule que la solution pertinente est celle qui fait décroître l'énergie le plus vite possible. Rien cependant ne laisse penser que ce critère s'appliquera tel quel à l'équation d'Euler incompressible, et il faut peut-être se résoudre à abandonner la propriété d'unicité pour des données initiales dans l'espace d'énergie — ou à rechercher cette unicité pour des quantités « moyennées ».

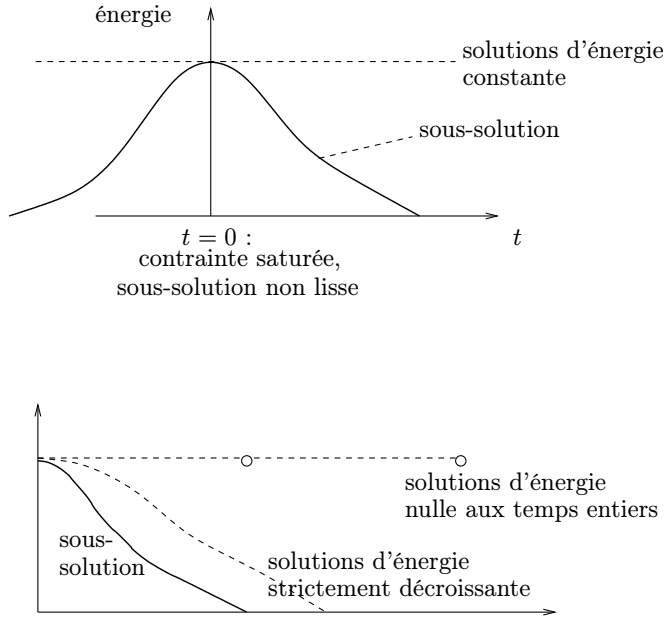


FIGURE 5. Construction de contre-exemples à l'unicité

6.2. Régularité, dissipation et conjecture d'Onsager

Dans un célèbre article sur la turbulence [**41**], Onsager suggérait que des solutions suffisamment irrégulières de l'équation d'Euler incompressible en dimension 3 pourraient dissiper de l'énergie. Cette conjecture, si elle était vérifiée, suggérerait la possibilité d'utiliser l'équation d'Euler pour modéliser la turbulence et les cascades d'énergie à la Kolmogorov — plutôt que l'équation de Navier–Stokes dans la limite de faible viscosité, ou grand nombre de Reynolds.

Onsager suggérait même une valeur exacte pour la régularité critique : 1/3 de dérivée. Ce nombre est lié à d'autres « constantes magiques » $(4/3, 4/5, 4/15, 5/3, \ldots)$ apparaissant en théorie statistique de la turbulence, obtenues par des raisonnements heuristiques et des hypothèses ad hoc (possibilité de prendre des moyennes sur l'ensemble des solutions, hypothèse « ergodique » de Taylor, hypothèse de dissipation non nulle de Kolmogorov...). Pour en savoir plus sur les théories de turbulence, parmi un très grand nombre de sources (parfois contradictoires) on pourra consulter les ouvrages [**26**] et [**36**, Chapitre 7], la synthèse [**2**] ou les articles de recherche [**22, 24**].

Par la suite, Constantin, E et Titi [**8**] ont effectivement prouvé qu'une solution d'Euler périodique en dimension 3, admettant plus d'1/3 de dérivée (au sens technique où cette solution appartient à un espace de Besov $B^{(\alpha,\infty),3}(\mathbb{T}^3)$ pour $\alpha > 1/3$) vérifie

la loi de conservation de l'énergie. Ce résultat a été précisé dans [**21, 22**] et [**7**], entre autres.

En contrepartie, ce n'est qu'en 2000 que Shnirelman [**45**] construisit des solutions distributionnelles de l'équation d'Euler tridimensionnelle dont l'énergie est *strictement* décroissante. Le théorème de De Lellis et Székelyhidi permet maintenant de renforcer ces résultats en imposant un profil arbitraire à l'énergie cinétique en fonction du temps.

Les données initiales construites par De Lellis et Székelyhidi (selon le schéma présenté dans la sous-section 6.1) sont forcément « très irrégulières » : elles n'admettent pas de solution régulière, même en temps petit ! En effet, les solutions d'Euler vérifiant la préservation de la borne d'énergie sont des solutions dissipatives au sens de Lions [**34**] [**15**, Section 8], et donc coïncident automatiquement avec la solution régulière s'il en existe une (dans ce cas nécessairement unique). Ici, « régulière » veut dire que le champ de vitesses v appartient à $C([0,T];L^2(\mathbb{R}^n))$ et $\nabla v + (\nabla v)^T \in L^1([0,T];L^\infty(\mathbb{R}^n))$.

Si tout cela semble corroborer l'intuition d'Onsager, il faut cependant noter que la construction de De Lellis et Székelyhidi (de même que les résultats de Constantin, E et Titi) s'applique en dimension 2 aussi bien qu'en dimension 3, alors que la suggestion d'Onsager était spécifiquement liée à la dimension 3. En fait, on peut soutenir qu'il n'existe à ce jour aucun argument vraiment convaincant, basé sur l'analyse des équations d'Euler ou de Navier–Stokes incompressibles, reflétant la différence — sur laquelle s'accordent les physiciens — entre les turbulences 2-dimensionnelle et 3-dimensionnelle ! Un nouveau rappel, s'il en était besoin, de notre ignorance en la matière.

Indépendamment de toute interprétation physique, ce serait remarquable si l'exposant $1/3$ proposé par Onsager était effectivement optimal, au sens où pour tout $\alpha < 1/3$ on pourrait construire des solutions « dérivables α fois » d'énergie strictement décroissante. (Onsager énonçait cette conjecture dans les espaces de Hölder, mais il n'y a pas vraiment d'argument pour les préférer à d'autres familles d'espaces fonctionnels quantifiant le degré de régularité, tels que les espaces de Sobolev ou de Besov...) Pour l'instant, même la continuité des solutions construites par De Lellis et Székelyhidi reste un rêve lointain ; en dimension 2 elle semble utopique à cause de la petitesse de K, mais en dimension plus grande cette question n'est peut-être pas désespérée.

On notera au passage que la même problématique se pose pour les isométries de Nash–Kuiper : on sait d'après Borisov [**3, 4**] que (au moins pour des surfaces analytiques de courbure strictement positive) l'on peut construire des isométries « paradoxales » en régularité $C^{1,\alpha}$ pour $\alpha < 1/7$, alors qu'il y a rigidité pour $\alpha > 2/3$. Pour diverses raisons, Borisov suggère que l'exposant $1/2$ est un bon candidat pour être

optimal, mais ce problème reste grand ouvert. Dans un travail en cours **[10]**, Conti, De Lellis et Székelyhidi simplifient et améliorent les résultats de Borisov, ce qui constitue peut-être un premier pas dans cette direction. Au passage, leurs travaux suggèrent l'exposant $1/3$ comme valeur plausible de la régularité critique !

6.3. h-principe pour l'équation d'Euler

La notion de h-principe (h = « homotopie »), popularisée par Gromov **[28]**, s'applique à des équations aux dérivées partielles, ou plus généralement des inclusions différentielles, pour lesquelles les seules obstructions sont de nature « globale ». Par exemple, dans le théorème de Nash–Kuiper, l'existence d'une immersion de M dans V (contrainte globale, topologique, souple) implique l'existence d'immersions isométriques (condition locale, en apparence rigide), que l'on peut obtenir par déformation à partir de la première. Raffinée pour tenir compte au mieux de la topologie, l'intégration convexe a constitué l'une des bases — mais pas la seule — de la théorie du h-principe, forte de nombreux résultats frappants. Citons le paradoxe du retournement de la sphère **[46]**, l'approximation par des métriques à courbure de Ricci négative **[35]**, la déformation de champs d'hyperplans en feuilletages **[52]**, et de nombreuses avancées en géométrie symplectique **[20, 28]**. Le livre d'Eliashberg et Mishachev **[23]** fournit une introduction au h-principe.

Comme on l'a vu, en un certain sens l'équation d'Euler vérifie ce principe : toute « sous-solution stricte », au sens expliqué dans la sous-section 6.1, peut être déformée en une solution exacte. Dans le cadre du problème de Cauchy, on est naturellement amené à la question non triviale : *pour quelles données initiales peut-on trouver une sous-solution stricte de l'équation d'Euler ?* Comme on l'a déjà noté, ces données initiales sont nécessairement très *irrégulières*, en un sens mal défini. Pour ce qui est de caractériser, ou tout au moins de décrire plus précisément ces solutions, en termes analytiques et/ou géométriques, nous n'avons pour l'heure essentiellement aucun indice.

7. ÉPILOGUE

Je conclurai cet exposé par des considérations peu mathématiques. De manière générale, pour résoudre un paradoxe, les deux principales attitudes consistent soit à accepter et réinterpréter l'énoncé paradoxal, soit à rejeter le modèle et les axiomes qui y ont conduit. Pour exclure Banach–Tarski, on pourra ainsi, selon ses goûts, s'interdire de considérer des ensembles non mesurables, ou bien rejeter l'axiome du choix.

L'attitude à adopter face aux plongements de Nash–Kuiper n'est pas évidente. On pourrait chercher des critères « extrinsèques » interdisant ces plongements ; on peut

aussi, au contraire, accepter ces plongements comme réalistes. Après tout, les plongements lipschitziens ne sont pas inaccessibles. Considérons par exemple une sphère élastique que l'on comprime et que l'on laisse se détendre ; des simulations numériques mèneront parfois à une position finale qui n'est pas une sphère standard, mais bien une sphère cabossée [**40**].[(8)]

Pour ce qui est du paradoxe de Scheffer–Shnirelman, on a pu penser un temps que la clé de la résolution résidait dans la définition des solutions faibles. Après les travaux de De Lellis et Székelyhidi, ce point de vue semble difficile à défendre ; en revanche on pourrait maintenant remettre en cause le modèle d'Euler lui-même, du moins quand il est appliqué à des données initiales trop irrégulières. Du point de vue physique, cela signifierait que l'équation d'Euler est impuissante à décrire des échelles très fines. En revanche, tant que l'on reste dans le cadre de la théorie de Yudovich, c'est-à-dire avec un tourbillon borné en dimension 2, le modèle semble inattaquable ; le cas des solutions faibles de Delort [**18**] (dont le tourbillon peut être une mesure positive) restant à approfondir.

Ce n'est d'ailleurs pas la première fois que l'équation d'Euler est contestée : le paradoxe de D'Alembert [**36**, Appendice 1.4], selon lequel un oiseau ne peut voler dans une atmosphère inviscide, incompressible et irrotationnelle, a déjà ébréché sa crédibilité il y a bien longtemps. On peut répondre au paradoxe de D'Alembert en invoquant la présence de vorticité, elle-même engendrée par la viscosité au bord (près de l'aile de l'oiseau...) [**36**, Section 1.4]. Ceci revient au fond à modifier le modèle, ou tout au moins à restreindre son champ d'application ; faut-il faire de même pour répondre à De Lellis et Székelyhidi ?

RÉFÉRENCES

[1] V. I. Arnold & B. A. Khesin – *Topological methods in hydrodynamics*, Applied Mathematical Sciences, vol. 125, Springer, 1998.

(8) Cette expérience rappelle Lebesgue froissant son mouchoir pour démontrer qu'une surface développable n'est pas forcément réglée...

[2] C. Bardos & E. S. Titi – Euler equations for an ideal incompressible fluid, *Uspekhi Mat. Nauk* **62** (2007), p. 5–46 ; traduction anglaise : *Russian Math. Surveys* **62** (2007), p. 409–451.

[3] Y. F. Borisov – $C^{1,\alpha}$-isometric immersions of Riemannian spaces, *Dokl. Akad. Nauk SSSR* **163** (1965), p. 11–13 ; traduction anglaise : *Soviet. Math. Dokl.* **6** (1965), p. 869–871.

[4] ———, Irregular surfaces of the class $C^{1,\beta}$ with an analytic metric, *Sibirsk. Mat. Zh.* **45** (2004), p. 25–61 ; traduction anglaise : *Siberian Math. J.* **45** (2004), 19–52.

[5] A. Bressan & F. Flores – On total differential inclusions, *Rend. Sem. Mat. Univ. Padova* **92** (1994), p. 9–16.

[6] A. Cellina – On the differential inclusion $x' \in [-1, +1]$, *Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Natur.* **69** (1980), p. 1–6.

[7] A. Cheskidov, P. Constantin, S. Friedlander & R. Shvydkoy – Energy conservation and Onsager's conjecture for the Euler equations, *Nonlinearity* **21** (2008), p. 1233–1252.

[8] P. Constantin, W. E & E. S. Titi – Onsager's conjecture on the energy conservation for solutions of Euler's equation, *Comm. Math. Phys.* **165** (1994), p. 207–209.

[9] P. Constantin, C. Fefferman & A. J. Majda – Geometric constraints on potentially singular solutions for the 3-D Euler equations, *Comm. Partial Differential Equations* **21** (1996), p. 559–571.

[10] S. Conti, C. De Lellis & L. Székelyhidi – Travail en cours.

[11] B. Dacorogna & P. Marcellini – General existence theorems for Hamilton–Jacobi equations in the scalar and vectorial cases, *Acta Math.* **178** (1997), p. 1–37.

[12] ———, *Implicit partial differential equations*, Progress in Nonlinear Differential Equations and their Applications, vol. 37, Birkhäuser, 1999.

[13] B. Dacorogna & G. Pisante – A general existence theorem for differential inclusions in the vector valued case, *Port. Math. (N.S.)* **62** (2005), p. 421–436.

[14] F. S. De Blasi & G. Pianigiani – A Baire category approach to the existence of solutions of multivalued differential equations in Banach spaces, *Funkcial. Ekvac.* **25** (1982), p. 153–162.

[15] C. De Lellis & L. Székelyhidi – On admissibility criteria for weak solutions of the Euler equations, *Arch. Ration. Mech. Anal.* **195** (2010), p. 225–260.

[16] ———, Communication personnelle.

[17] ———, The Euler equations as a differential inclusion, à paraître dans *Ann. of Math.*, consultable en ligne à arXiv:math/0702079.

[18] J.-M. Delort – Existence de nappes de tourbillon en dimension deux, *J. Amer. Math. Soc.* **4** (1991), p. 553–586.

[19] R. J. DiPerna – Compensated compactness and general systems of conservation laws, *Trans. Amer. Math. Soc.* **292** (1985), p. 383–420.

[20] S. K. Donaldson – Symplectic submanifolds and almost-complex geometry, *J. Differential Geom.* **44** (1996), p. 666–705.

[21] J. Duchon & R. Robert – Dissipation d'énergie pour des solutions faibles des équations d'Euler et Navier–Stokes incompressibles, *C. R. Acad. Sci. Paris Sér. I Math.* **329** (1999), p. 243–248.

[22] ———, Inertial energy dissipation for weak solutions of incompressible Euler and Navier–Stokes equations, *Nonlinearity* **13** (2000), p. 249–255.

[23] Y. Eliashberg & N. Mishachev – *Introduction to the h-principle*, Graduate Studies in Math., vol. 48, Amer. Math. Soc., 2002.

[24] G. L. Eyink – Local 4/5-law and energy dissipation anomaly in turbulence, *Nonlinearity* **16** (2003), p. 137–145.

[25] A. F. Filippov – Classical solutions of differential equations with multi-valued right-hand side, *SIAM J. Control* **5** (1967), p. 609–621.

[26] U. Frisch – *Turbulence : The legacy of A. N. Kolmogorov*, Cambridge Univ. Press, 1995.

[27] P. Gérard – Résultats récents sur les fluides parfaits incompressibles bidimensionnels (d'après J.-Y. Chemin et J.-M. Delort), Séminaire Bourbaki, vol. 1991/92, exposé n° 757, *Astérisque* **206** (1992), p. 411–444.

[28] M. Gromov – *Partial differential relations*, Ergebnisse Math. Grenzg., vol. 9, Springer, 1986.

[29] B. Kirchheim – Deformations with finitely many gradients and stability of quasiconvex hulls, *C. R. Acad. Sci. Paris Sér. I Math.* **332** (2001), p. 289–294.

[30] ———, Rigidity and geometry of microstructures, thèse d'habilitation, université de Leipzig, 2003.

[31] B. Kirchheim, S. Müller & V. Šverák – Studying nonlinear PDE by geometry in matrix space, in *Geometric analysis and nonlinear partial differential equations*, Springer, 2003, p. 347–395.

[32] N. H. Kuiper – On C^1-isometric imbeddings. I, II, *Nederl. Akad. Wetensch. Proc. Ser. A.* **58** = *Indag. Math.* **17** (1955), p. 545–556, 683–689.

[33] ———, Isometric and short imbeddings, *Nederl. Akad. Wetensch. Proc. Ser. A. 62 = Indag. Math.* **21** (1959), p. 11–25.

[34] P.-L. Lions – *Mathematical topics in fluid mechanics. Vol. 1 : Incompressible models*, Oxford Lecture Series in Mathematics and its Applications, vol. 3, The Clarendon Press Oxford Univ. Press, 1996.

[35] J. Lohkamp – Curvature h-principles, *Ann. of Math.* **142** (1995), p. 457–498.

[36] C. Marchioro & M. Pulvirenti – *Mathematical theory of incompressible non-viscous fluids*, Applied Mathematical Sciences, vol. 96, Springer, 1994.

[37] S. Müller & V. Šverák – Convex integration for Lipschitz mappings and counterexamples to regularity, *Ann. of Math.* **157** (2003), p. 715–742.

[38] F. Murat – Compacité par compensation, *Ann. Scuola Norm. Sup. Pisa Cl. Sci.* **5** (1978), p. 489–507, partie II : Proceedings of the International Meeting on Recent Methods in Nonlinear Analysis (Rome, 1978), Pitagora, Bologna, 1979, p. 245–256.

[39] J. Nash – C^1 isometric imbeddings, *Ann. of Math.* **60** (1954), p. 383–396.

[40] Y. Ollivier – Communication personnelle.

[41] L. Onsager – Statistical hydrodynamics, *Nuovo Cimento* **6**, Supplements (1949), p. 279–287.

[42] A. V. Pogorelov – *Extrinsic geometry of convex surfaces*, Translations of Mathematical Monographs, vol. 35, Amer. Math. Soc., 1973.

[43] V. Scheffer – An inviscid flow with compact support in space-time, *J. Geom. Anal.* **3** (1993), p. 343–401.

[44] A. Shnirelman – On the nonuniqueness of weak solution of the Euler equation, *Comm. Pure Appl. Math.* **50** (1997), p. 1261–1286.

[45] ———, Weak solutions with decreasing energy of incompressible Euler equations, *Comm. Math. Phys.* **210** (2000), p. 541–603.

[46] S. Smale – A classification of immersions of the two-sphere, *Trans. Amer. Math. Soc.* **90** (1958), p. 281–290.

[47] D. Spring – *Convex integration theory, solutions to the h-principle in geometry and topology*, Monographs in Math., vol. 92, Birkhäuser, 1998.

[48] M. A. Sychev – A few remarks on differential inclusions, *Proc. Roy. Soc. Edinburgh Sect. A* **136** (2006), p. 649–668.

[49] L. Székelyhidi – Differential inclusions, ouvrage en cours de rédaction.

[50] L. Tartar – Compensated compactness and applications to partial differential equations, in *Nonlinear analysis and mechanics : Heriot–Watt Symposium, Vol. IV*, Res. Notes in Math., vol. 39, Pitman, 1979, p. 136–212.

[51] ———, The compensated compactness method applied to systems of conservation laws, in *Systems of nonlinear partial differential equations (Oxford, 1982)*, NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci., vol. 111, Reidel, 1983, p. 263–285.

[52] W. P. Thurston – Existence of codimension-one foliations, *Ann. of Math.* **104** (1976), p. 249–268.

Cédric VILLANI

École Normale Supérieure de Lyon
U.M.P.A.
UMR 5669 du CNRS
46 allée d'Italie
F–69364 Lyon Cedex 07
E-mail : cvillani@umpa.ens-lyon.fr

332

ASTÉRISQUE

2010

SÉMINAIRE BOURBAKI
VOLUME 2008/2009
EXPOSÉS 997-1011

(1002) *La conjecture de Weinstein en dimension* 3

Denis AUROUX

SOCIÉTÉ MATHÉMATIQUE DE FRANCE
Publié avec le concours du CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

Séminaire BOURBAKI Mars 2009
61e année, 2008-2009, no 1002, p. 135 à 159

LA CONJECTURE DE WEINSTEIN EN DIMENSION 3
[d'après C. H. Taubes]

par **Denis AUROUX**

Étant donnée une variété fermée munie d'une forme de contact a, la conjecture de Weinstein [**22**] affirme l'existence d'orbites périodiques du champ de Reeb (le champ de vecteurs qui engendre le noyau de da). Cette conjecture a été prouvée fin 2006 par C. H. Taubes [**18**] pour toute variété de contact de dimension 3. Le présent exposé décrit le résultat de Taubes et les principaux ingrédients de la preuve, en particulier les équations de Seiberg-Witten en dimension 3.

1. LA CONJECTURE DE WEINSTEIN

Soit M une variété compacte orientée de dimension $2n+1$, équipée d'une forme de contact a, c'est-à-dire une 1-forme telle que $a\wedge(da)^n$ soit une forme volume compatible avec l'orientation de M. Le champ d'hyperplans $\xi = \mathrm{Ker}(a)$ définit alors une structure de contact sur M.

La 2-forme da est de rang $2n$, et son noyau définit une famille de droites dans le fibré tangent TM, transverses aux hyperplans de contact ; le champ de Reeb est le champ de vecteurs v qui engendre ces droites, normalisé par la condition $a(v) = 1$.

Définition 1.1. — *Le champ de Reeb est l'unique champ de vecteurs v sur M tel que $da(v,\cdot) = 0$ et $a(v) = 1$.*

Par exemple, en dimension 3, il existe des coordonnées locales (x, y, z) telles que $a = dz - y\,dx$; le champ de Reeb est alors $v = \partial/\partial z$.

La conjecture de Weinstein [**22**], formulée il y a trente ans, affirme l'existence d'orbites périodiques du champ de Reeb sur les variétés compactes :

Conjecture 1.2 (Weinstein). — *Le champ de Reeb v possède au moins une courbe intégrale fermée.*

La conjecture de Weinstein était motivée par un résultat de Rabinowitz [**12**] affirmant l'existence d'orbites périodiques dans le cas particulier où M est une hypersurface étoilée dans $\mathbb{R}^{2n}$; ce résultat a ensuite été étendu par Viterbo [**21**] au cas des hypersurfaces de type contact dans $\mathbb{R}^{2n}$. Plus récemment, des progrès spectaculaires ont été accomplis grâce à la théorie des courbes pseudo-holomorphes. Ainsi, le lien entre orbites périodiques du champ de Reeb sur la variété de contact M et courbes pseudo-holomorphes dans la symplectisation $(M \times \mathbb{R}, d(e^t a))$ a permis à Hofer [**3**] d'établir la conjecture de Weinstein pour la sphère S^3 munie d'une forme de contact arbitraire, ainsi que pour toutes les structures de contact vrillées en dimension 3, et pour toutes les variétés de dimension 3 telles que $\pi_2(M) \neq 0$ (voir également l'exposé de F. Laudenbach au Séminaire Bourbaki [**10**]).

Divers autres cas particuliers de la conjecture de Weinstein en dimension 3 ont également été résolus relativement récemment : voir par exemple les travaux d'Abbas, Cieliebak et Hofer [**1**] pour le cas des variétés qui admettent un livre ouvert planaire, et de Gay [**2**] pour les variétés qui contiennent un tore de torsion non nulle.

Dans son article [**18**], Taubes établit la conjecture de Weinstein pour toutes les variétés de dimension 3. Son résultat peut être formulé de façon plus précise. Pour ceci, on rappelle que la forme da définit une structure symplectique sur le champ d'hyperplans de contact $\xi = \mathrm{Ker}(a)$. On peut donc définir la classe de Chern $c_1(\xi) \in H^2(M, \mathbb{Z})$. Comme toute variété compacte orientée de dimension 3 est spin, la seconde classe de Stiefel-Whitney du fibré tangent $TM = \xi \oplus \mathbb{R}v$ s'annule. Ceci implique que $c_1(\xi)$ est divisible par 2.

THÉORÈME 1.3 (Taubes [**18**]). — *Soit $e \in H^2(M, \mathbb{Z})$ une classe telle que $2e + c_1(\xi)$ soit un élément de torsion de $H^2(M, \mathbb{Z})$. Alors il existe une collection non vide de courbes intégrales fermées γ_i du champ de Reeb, et des multiplicités entières $m_i \geq 1$, telles que $\sum m_i[\gamma_i] \in H_1(M, \mathbb{Z})$ représente le dual de Poincaré de e.*

La preuve de ce théorème repose sur une étude détaillée du comportement des solutions des équations de Seiberg-Witten sur M vis-à-vis d'une certaine perturbation des équations. Taubes montre que, si une suite de solutions des équations perturbées satisfait certaines estimées lorsque le paramètre de perturbation tend vers l'infini, alors il existe une sous-suite pour laquelle le lieu d'annulation d'une des composantes de la solution converge vers une collection d'orbites fermées du champ de Reeb. Ce résultat [**18**, théorème 2.1] est l'analogue en dimension 3 d'un théorème précédemment établi par Taubes, qui relie les équations de Seiberg-Witten sur les variétés symplectiques de dimension 4 aux courbes pseudo-holomorphes [**15**] (voir également les articles [**14, 16**] et l'exposé de D. Kotschick au Séminaire Bourbaki [**8**]). Nous tenterons d'esquisser les grandes lignes de l'argument au §2.

Le reste de la preuve du théorème 1.3 consiste alors à établir l'existence de solutions satisfaisant les conditions requises. Ceci repose sur les travaux de Kronheimer et Mrowka concernant l'homologie de Seiberg-Witten-Floer [**9**], et sur des estimées précises concernant le flot spectral et l'énergie des solutions [**18**]. Les principaux ingrédients de la preuve seront présentés au §3.

La relation qu'établit Taubes entre solutions des équations de Seiberg-Witten et orbites du champ de Reeb permet en fait d'obtenir des résultats beaucoup plus précis que le théorème 1.3. Taubes a récemment annoncé [**20**] un isomorphisme entre l'homologie de Seiberg-Witten-Floer déjà mentionnée, qui est l'homologie d'un complexe engendré par les solutions des équations de Seiberg-Witten, et l'homologie de contact plongée définie par M. Hutchings [**4, 5, 6**], qui est l'homologie d'un complexe engendré par des collections d'orbites fermées du champ de Reeb.

THÉORÈME 1.4 (Taubes [**20**]). — *Soit M une variété compacte orientée de dimension 3, munie d'une forme de contact a générique. Alors la cohomologie de Seiberg-Witten-Floer de M est isomorphe à l'homologie de contact plongée de (M, a).*

L'énoncé de ce théorème sera clarifié au §4, où nous donnerons une brève description de l'homologie de contact plongée.

Entre autres applications, la relation entre cohomologie de Seiberg-Witten-Floer et homologie de contact plongée permet à Taubes de construire des orbites fermées du champ de Reeb représentant des classes d'homologie plus générales que celles garanties par le théorème 1.3 [**19**]. Elle permet également à Hutchings et Taubes [**7**] de généraliser la conjecture de Weinstein aux variétés de dimension 3 munies d'une « structure hamiltonienne stable », c'est-à-dire une 1-forme a et une 2-forme ω telles que $d\omega = 0$, $a \wedge \omega > 0$, et $\mathrm{Ker}(\omega) \subset \mathrm{Ker}(da)$ (les formes de contact correspondent au cas particulier $\omega = da$).

2. ÉQUATIONS DE SEIBERG-WITTEN ET ORBITES DU CHAMP DE REEB

2.1. Les équations de Seiberg-Witten en dimension 3

Soit M une variété compacte orientée de dimension 3, et soit g une métrique riemannienne sur M. Le fibré des repères orthonormés directs de (TM, g) est un fibré principal au-dessus de M, de groupe structural $SO(3)$. D'autre part, l'action du groupe unitaire $U(2)$ par rotations de la droite projective $\mathbb{CP}^1 \cong S^2$ induit un morphisme surjectif $\pi : U(2) \to SO(3)$, dont le noyau est le centre de $U(2)$, isomorphe à S^1.

DÉFINITION 2.1. — *Une* structure spin^c *sur M est la donnée d'un relèvement (via la projection π) du fibré des repères en un fibré principal P de groupe structural $U(2)$. Le* fibré des spineurs $\mathbb{S} = P \times_{U(2)} \mathbb{C}^2$ *est le fibré vectoriel complexe de rang 2 associé à P via la représentation standard de $U(2)$.*

L'ensemble des structures spin^c sur M est un espace principal homogène pour $H^2(M,\mathbb{Z})$: étant donnés une structure spin^c et son fibré des spineurs $\mathbb{S}$, le fibré des spineurs pour une autre structure spin^c est de la forme $\mathbb{S}\otimes E$, où E est un fibré en droites complexes au-dessus de M.

Le fibré des spineurs est naturellement muni d'une métrique hermitienne, et d'une action du fibré cotangent T^*M par endomorphismes antihermitiens, appelée *multiplication de Clifford*, telle que $cl(u)\,cl(v) + cl(v)\,cl(u) = -2g(u,v)\,\mathrm{Id}$ pour tous $u,v \in T^*M$, et $cl(e_1)cl(e_2)cl(e_3) = \mathrm{Id}$ lorsque (e_1,e_2,e_3) est un repère orthonormé direct.

Une *connexion spin^c* sur le fibré des spineurs est une connexion hermitienne induite par une connexion sur le fibré principal P qui relève la connexion de Levi-Civita sur le fibré des repères. La multiplication de Clifford vérifie alors la règle de Leibniz. Une connexion spin^c est déterminée uniquement par la connexion hermitienne qu'elle induit sur le fibré déterminant $\det(\mathbb{S}) = \wedge^2\mathbb{S}$: une connexion hermitienne $\mathbb{A}$ sur $\det(\mathbb{S})$ et la connexion de Levi-Civita déterminent une connexion spin^c sur $\mathbb{S}$, que l'on notera $\nabla^{\mathbb{A}}$.

L'*opérateur de Dirac* associé à une connexion spin^c $\nabla^{\mathbb{A}}$ est l'opérateur elliptique autoadjoint du premier ordre $D_{\mathbb{A}} : \Gamma(\mathbb{S}) \to \Gamma(\mathbb{S})$ obtenu en composant la dérivation covariante $\nabla^{\mathbb{A}} : \Gamma(\mathbb{S}) \to \Gamma(T^*M\otimes\mathbb{S})$ et la multiplication de Clifford $T^*M\otimes\mathbb{S}\to\mathbb{S}$.

Les équations de Seiberg-Witten sont des équations aux dérivées partielles concernant une section $\psi\in\Gamma(\mathbb{S})$ du fibré des spineurs et une connexion hermitienne $\mathbb{A}$ sur le fibré en droites $\det(\mathbb{S})$. Elles peuvent s'écrire sous la forme :

$$\begin{cases} *F_{\mathbb{A}} = \psi^*\tau\psi \\ D_{\mathbb{A}}\psi = 0 \end{cases} \tag{1}$$

où $F_{\mathbb{A}} \in \Omega^2(M,i\mathbb{R})$ est la courbure de la connexion $\mathbb{A}$, et $\psi^*\tau\psi \in \Omega^1(M,i\mathbb{R})$ est le dual pour la métrique g de la forme linéaire $\langle\psi, cl(\cdot)\psi\rangle : T^*M \to i\mathbb{R}$. Comme cl induit un isomorphisme compatible avec la métrique entre 1-formes à valeurs imaginaires pures et endomorphismes hermitiens de trace nulle de $\mathbb{S}$, la forme $\psi^*\tau\psi$ est caractérisée par

$$cl(\psi^*\tau\psi) = -2\psi^*\otimes\psi + |\psi|^2\,\mathrm{Id}. \tag{2}$$

Les équations (1) sont invariantes sous l'action du *groupe de jauge* $\mathcal{G} = C^\infty(M,S^1)$, qui agit par $u : (\mathbb{A},\psi) \mapsto (\mathbb{A} - 2u^{-1}du, u\psi)$. L'*espace des modules* est le quotient de l'espace des solutions par l'action du groupe de jauge.

Comme on le verra au §3, il est important de distinguer entre solutions *irréductibles*, pour lesquelles ψ n'est pas identiquement nul et l'action du groupe de jauge est libre, et solutions *réductibles*, de la forme $(\mathbb{A}, 0)$ où $\mathbb{A}$ est une connexion plate sur $\det(\mathbb{S})$, sur lesquelles $\mathcal{G}$ agit avec stabilisateur S^1 (les transformations de jauge constantes).

Soit a une forme de contact sur une variété compacte orientée M de dimension 3. Comme $a \wedge da > 0$ est une forme volume, il est possible de choisir une métrique riemannienne g telle que $|a| = 1$ et $da = 2 * a$ en tout point de M. La distribution de contact $\xi = \mathrm{Ker}(a)$ est alors munie d'une structure complexe J telle que $g_{|\xi} = da(\cdot, J\cdot)$, ce qui en fait un fibré en droites complexes au-dessus de M.

Etant donnée une structure spin^c sur M, la multiplication de Clifford par a vérifie $cl(a)^2 = -\mathrm{Id}$, et les sous-espaces propres associés aux valeurs propres $\pm i$ décomposent le fibré des spineurs en la somme directe de deux fibrés en droites complexes. Il existe une structure spin^c dite *canonique*, pour laquelle le sous-fibré associé à la valeur propre $+i$ est topologiquement trivial, et le fibré des spineurs est $\mathbb{S}_0 = \underline{\mathbb{C}} \oplus \xi$; de plus il existe une connexion spin^c ∇^0 pour laquelle la section $(1, 0)$ du fibré $\mathbb{S}_0 = \underline{\mathbb{C}} \oplus \xi$ appartient au noyau de l'opérateur de Dirac. Pour une structure spin^c quelconque, le fibré des spineurs est de la forme

$$\mathbb{S} = \mathbb{S}_0 \otimes E = E \oplus (\xi \otimes E),$$

où E est un fibré en droites complexes sur M. Le fibré déterminant est alors $\det(\mathbb{S}) = \xi \otimes E^{\otimes 2}$, et la structure spin^c est déterminée de façon unique par la classe de Chern $c_1(E) \in H^2(M, \mathbb{Z})$. Une connexion spin^c sur $\mathbb{S}$ est déterminée par la donnée d'une connexion hermitienne A sur E et de la connexion ∇^0 sur $\mathbb{S}_0$. La connexion induite sur le fibré déterminant est alors de la forme $\mathbb{A} = \mathbb{A}_0 + 2A$, où $\mathbb{A}_0$ est la connexion induite par ∇^0 sur $\xi = \det(\mathbb{S}_0)$. On peut donc réécrire les équations de Seiberg-Witten en termes de la connexion A sur E.

Plus précisément, Taubes considère une modification des équations de Seiberg-Witten, qui dépend d'un paramètre réel $r \geq 1$:

Définition 2.2. — *Les équations de Seiberg-Witten perturbées par la forme de contact a s'écrivent sous la forme :*

$$\begin{cases} *F_A = r(\psi^* \tau \psi - ia) + *(i\, d\mu - \frac{1}{2} F_\xi) \\ D_A \psi = 0. \end{cases} \tag{3}$$

Dans ces équations :

- *les inconnues sont une connexion hermitienne A sur E et une section ψ du fibré des spineurs ;*
- *les paramètres sont un nombre réel $r \geq 1$ et une 1-forme $\mu \in \Omega^1(M, \mathbb{R})$;*
- *$F_\xi \in \Omega^2(M, i\mathbb{R})$ est l'unique forme harmonique telle que $\frac{i}{2\pi}[F_\xi] = c_1(\xi)$;*
- *D_A est l'opérateur de Dirac associé à la connexion spin^c ∇^A induite par A.*

Le rôle de la 1-forme μ est purement technique : il s'agit d'assurer une propriété de transversalité pour les solutions par un choix générique de μ.

2.2. L'existence d'orbites fermées du champ de Reeb

Le principal ingrédient de la preuve de la conjecture de Weinstein est le résultat suivant, qui constitue le théorème 2.1 de **[18]** :

THÉORÈME 2.3 (Taubes). — *Etant donnée une structure spin^c sur M, soit $(r_n)_{n\in\mathbb{N}}$ une suite de nombres réels tendant vers $+\infty$, et soit $(A_n,\psi_n)_{n\in\mathbb{N}}$ une suite de solutions des équations de Seiberg-Witten perturbées* (3) *pour μ fixé et $r=r_n$. S'il existe des nombres réels δ et $\mathcal{E}$ (indépendants de n) tels que les deux conditions*

$$\sup_M(1-|\psi_n|)\geq\delta>0, \tag{4}$$

$$E(A_n)=i\int_M a\wedge F_{A_n}\leq\mathcal{E} \tag{5}$$

soient satisfaites, alors il existe une collection non vide de courbes intégrales fermées γ_i du champ de Reeb, et des multiplicités entières $m_i\geq 1$, telles que $\sum m_i[\gamma_i]\in H_1(M,\mathbb{Z})$ représente le dual de Poincaré de $c_1(E)$.

De façon plus précise, la décomposition $\mathbb{S}=E\oplus(\xi\otimes E)$ permet d'écrire $\psi_n=(\alpha_n,\beta_n)$. Le résultat est alors que, sous les hypothèses du théorème, les sous-ensembles $\alpha_n^{-1}(0)$ de M où la première composante de ψ_n s'annule admettent une sous-suite qui converge au sens des courants vers une collection d'orbites fermées du champ de Reeb.

La conjecture de Weinstein est alors ramenée au problème de l'existence de solutions des équations de Seiberg-Witten perturbées vérifiant les conditions (4) et (5). La stratégie utilisée par Taubes pour construire ces solutions sera l'objet du §3 ci-dessous.

La preuve du théorème 2.3 repose sur des estimées *a priori* pour les solutions de (3) lorsque $r\to+\infty$. Dans ce qui suit, pour ne pas alourdir inutilement les notations, lorsqu'il n'y a pas d'ambiguïté on utilisera la même notation ∇^A pour les dérivations covariantes induites par A et $\mathbb{A}_0$ sur les sous-fibrés E et $\xi\otimes E$ de $\mathbb{S}$.

LEMME 2.4 (Taubes **[18]**, lemmes 2.2 et 2.3). — *Pour μ fixé, il existe une constante c_0 telle que, pour tout $r\geq 1$, si (A,ψ) est une solution de* (3) *alors les composantes*

(α, β) de ψ vérifient les estimées :

$$|\alpha| \leq 1 + \frac{c_0}{r} \tag{6}$$

$$|\beta|^2 \leq \frac{c_0}{r}(1 - |\alpha|^2) + \frac{c_0}{r^2} \tag{7}$$

$$|\nabla^A \alpha| \leq c_0 r^{1/2} \tag{8}$$

$$|\nabla^A \beta| \leq c_0. \tag{9}$$

Idée de démonstration. — D'après la formule de Lichnerowicz-Weitzenböck,

$$D_A^2 \psi = (\nabla^A)^* \nabla^A \psi + \frac{R}{4}\psi - \frac{1}{2} cl(*F_{\mathbb{A}})\psi,$$

où R est la courbure scalaire de la métrique g, et $F_{\mathbb{A}} = F_{\mathbb{A}_0} + 2F_A$ est la courbure de la connexion $\mathbb{A}$ sur $\det(\mathbb{S})$. En utilisant (3) et (2), on obtient donc l'identité

$$\langle \psi, D_A^2 \psi \rangle = d^* d \frac{|\psi|^2}{2} + |\nabla^A \psi|^2 + \frac{R}{4}|\psi|^2 + r|\psi|^4 + ir\, \psi^* cl(a)\psi + \psi^* cl(*u)\psi = 0,$$

où $u = -i\, d\mu + \frac{1}{2}F_\xi - \frac{1}{2}F_{\mathbb{A}_0}$. Par conséquent,

$$d^* d \frac{|\psi|^2}{2} + r\left(|\psi|^2 - 1 - \frac{c}{r}\right)|\psi|^2 \leq 0$$

(où c est une constante qui borne $|\frac{R}{4}| + |u|$). Le principe du maximum implique alors que $|\psi|^2 \leq 1 + \frac{c}{r}$ en tout point, ce qui donne (6).

Si l'on considère séparément les composantes de l'identité $D_A^2 \psi = 0$ selon les sous-fibrés E et $\xi \otimes E$, on obtient des équations de la forme

$$(\nabla^A)^* \nabla^A \alpha - r(1 - |\alpha|^2 - |\beta|^2)\alpha + f_0(\nabla^A \alpha) + f_1(\nabla^A \beta) + f_2 \alpha + f_3 \beta = 0, \tag{10}$$

$$(\nabla^A)^* \nabla^A \beta + r(1 + |\alpha|^2 + |\beta|^2)\beta + f_0'(\nabla^A \alpha) + f_1'(\nabla^A \beta) + f_2' \alpha + f_3' \beta = 0, \tag{11}$$

où les quantités f_i et f_i' dépendent uniquement de la métrique g et de la forme u, et sont donc bornées indépendamment de r, α et β. En considérant une combinaison linéaire bien choisie des équations (10) et (11), Taubes en déduit une inégalité de la forme

$$d^* d\left(|\beta|^2 - \frac{c_1}{r}(1 - |\alpha|^2) - \frac{c_2}{r^2}\right) + r\left(|\beta|^2 - \frac{c_1}{r}(1 - |\alpha|^2) - \frac{c_2}{r^2}\right) \leq 0$$

où c_1 et c_2 sont des constantes indépendantes de r, α, β ; ceci implique (7) par le principe du maximum.

Une fois les bornes sur $|\alpha|$ et $|\beta|$ établies, des techniques standard d'estimation elliptique permettent de borner $|\nabla^A \psi|$ par un multiple de $r^{1/2}$, ce qui implique (8). Enfin, Taubes obtient la borne (9) en considérant l'équation (11) et en utilisant les estimées obtenues précédemment. □

La preuve du théorème 2.3 fait également appel à une comparaison avec les solutions des équations vortex :

DÉFINITION 2.5. — *Les* équations vortex *sur* $\mathbb{C}$ *sont les équations*

$$\begin{cases} *d\mathscr{A} = -i(1-|\sigma|^2), \\ \overline{\partial}_{\mathscr{A}}\sigma = 0, \end{cases} \tag{12}$$

où σ *est une fonction sur* $\mathbb{C}$ *à valeurs complexes,* $\mathscr{A}$ *est une 1-forme sur* $\mathbb{C}$ *à valeurs imaginaires pures, et* $\overline{\partial}_{\mathscr{A}} = \overline{\partial} + \mathscr{A}^{0,1}$.

La géométrie des équations (12) est bien comprise (voir par exemple le §2b de [**17**]) :

LEMME 2.6. — *Soit* $(\mathscr{A},\sigma)$ *une solution de* (12) *telle que* $\int_{\mathbb{C}}(1-|\sigma|^2) < \infty$. *Alors :*

1. $|\sigma| \leq 1$ *en tout point de* $\mathbb{C}$;
2. *il existe un entier naturel* m *tel que* $\int_{\mathbb{C}}(1-|\sigma|^2) = 2\pi m$;
3. *les zéros de* σ *sont isolés, et sont au nombre de* m *(en comptant avec multiplicités) ;*
4. $|\sigma|$ *est proche de* 1 *partout sauf au voisinage des zéros :* $1-|\sigma|$ *est majoré par une quantité qui décroît exponentiellement avec la distance à* $\sigma^{-1}(0)$.

De plus, l'espace des modules $\mathfrak{C}_m$ *des solutions d'énergie* $\int_{\mathbb{C}}(1-|\sigma|^2) = 2\pi m$ *s'identifie naturellement avec le produit symétrique* $\mathrm{Sym}^m(\mathbb{C})$, *via* $\sigma \mapsto \sigma^{-1}(0)$.

Nous pouvons maintenant esquisser la preuve du théorème 2.3. Soit (A,ψ) une solution de l'équation de Seiberg-Witten perturbée (3). L'équation de Dirac $D_A\psi = 0$ peut s'écrire en termes des composantes (α,β) de ψ sous la forme suivante :

$$D_A\begin{pmatrix}\alpha\\ \beta\end{pmatrix} = \begin{pmatrix} i\nabla_v^A\alpha + 2i\partial_\xi^A\beta \\ -i\nabla_v^A\beta + 2i\overline{\partial}_\xi^A\alpha\end{pmatrix} + \begin{pmatrix} f_0\alpha + f_1\beta \\ f_0'\alpha + f_1'\beta\end{pmatrix} = 0, \tag{13}$$

où les quantités f_i et f_i' sont déterminées par la métrique et uniformément bornées, ∇_v^A est la dérivation covariante dans la direction du champ de Reeb v, et ∂_ξ^A et $\overline{\partial}_\xi^A$ sont respectivement les parties $\mathbb{C}$-linéaire et $\mathbb{C}$-antilinéaire de la dérivation covariante dans la direction des plans de contact $\xi = \mathrm{Ker}(a)$. Plus précisément, la restriction de $\nabla^A\alpha$ aux plans de contact définit un homomorphisme $\mathbb{R}$-linéaire $\nabla_\xi^A\alpha \in \mathrm{Hom}_{\mathbb{R}}(\xi,E)$, dont la partie $\mathbb{C}$-antilinéaire pour la structure complexe J sur ξ est une section de $\mathrm{Hom}_{\mathbb{C}}(\overline{\xi},E) \cong \xi\otimes E$, que l'on note $\overline{\partial}_\xi^A\alpha$. La section $\partial_\xi^A\beta$ de $\mathrm{Hom}_{\mathbb{C}}(\xi,\xi\otimes E) \cong E$ est définie de façon analogue.

L'équation (13) et les estimées du lemme 2.4 impliquent l'existence d'une constante c_0' indépendante de r telle que

$$|\nabla_v^A\alpha| \leq c_0' \quad \text{et} \quad |\overline{\partial}_\xi^A\alpha| \leq c_0'. \tag{14}$$

D'autre part, on déduit de (3) que

$$(15)\quad ia \wedge F_A = r(|\beta|^2 - |\alpha|^2 + 1)\, vol_g + ia \wedge (i\, d\mu - \tfrac{1}{2} F_\xi) = r(1 - |\alpha|^2)\, vol_g + \mathcal{O}(1).$$

Ces estimées impliquent que, dans la direction des plans de contact, pour r suffisamment grand les solutions de (3) peuvent être approximées par des solutions des équations vortex (12) après une homothétie d'un facteur $r^{1/2}$.

Plus précisément, soit $p \in M$ un point où $|\alpha| \leq 1 - \frac{\delta}{2}$, où $\delta > 0$ est la constante qui apparaît dans l'énoncé du théorème 2.3. L'hypothèse (4) et l'inégalité (7) impliquent l'existence d'un tel point lorsque (A, ψ) est l'une des solutions (A_n, ψ_n) considérées dans l'énoncé du théorème 2.3 et r est suffisamment grand.

On identifie un voisinage de p (centré sur la courbe intégrale du champ de Reeb passant par p) avec un cylindre de la forme $D \times I$, où D est un disque muni de la coordonnée complexe $x + iy$ et I est un intervalle muni de la coordonnée réelle z, de telle sorte que le champ de Reeb s'identifie à $\partial/\partial z$, le point p correspond à l'origine, et les disques $D \times \{z\}$ sont tangents à la distribution de contact à l'origine (de telles coordonnées seront dites *adaptées*). Taubes montre alors que, pour z fixé, si l'on munit $D \times \{z\}$ de la coordonnée complexe $r^{1/2}(x + iy)$, la restriction de (A, α) peut être approximée au voisinage de l'origine par une solution de (12) [**18**, lemme 6.1].

Soit $D_r \subset D$ le disque de rayon $cr^{-1/2}$, avec $c > 0$ une constante bien choisie, et soit

$$e(z) = r \int_{D_r \times \{z\}} (1 - |\alpha|^2).$$

La borne uniforme sur $|\overline{\partial}^A_\xi \alpha|$ et l'inégalité $|\alpha(p)| \leq 1 - \frac{\delta}{2}$ permettent de montrer que $e(0)$ est minoré par une constante $\epsilon > 0$ indépendante de r. D'autre part, la borne uniforme sur $|\nabla^A_v \alpha|$ donne une borne uniforme sur de/dz. Il existe donc une constante ℓ indépendante de r telle que $e(z) \geq \epsilon/2$ pour tout $z \in [0, \ell]$. En utilisant (15), on en déduit que la contribution du cylindre $C(p) = D_r \times [0, \ell]$ (un voisinage de rayon $cr^{-1/2}$ et de longueur ℓ de la courbe intégrale du champ de Reeb passant par p) à l'énergie $E(A) = i \int a \wedge F_A$ est minorée par une constante indépendante de r.

D'autre part, la borne $e(\ell) \geq \epsilon/2$ implique (en comparant $\alpha_{|D \times \{\ell\}}$ à une solution de (12) et en utilisant les propriétés de ces solutions) l'existence d'un point p' de $D_r \times \{\ell\}$ où $|\alpha| \leq 1 - \frac{\delta}{2}$. Il est donc possible de répéter la construction précédente en partant du point p', et ainsi de suite : ceci donne une collection de cylindres $C(p_i)$ « empilés » le long d'une courbe intégrale du champ de Reeb (avec un décalage d'au plus $cr^{-1/2}$ entre deux cylindres consécutifs).

Comme la contribution de chaque cylindre $C(p_i)$ à l'énergie $E(A)$ est minorée de façon indépendante de r, l'hypothèse $E(A) \leq \mathcal{E}$ permet de majorer le nombre de cylindres disjoints par une constante N indépendante de r. Ceci implique que la courbe intégrale γ du champ de Reeb issue de p revient dans un voisinage de taille

$\mathcal{O}(r^{-1/2})$ d'elle-même en un temps borné par $N\ell$. On en déduit également que $\alpha^{-1}(0)$ est contenu dans un voisinage $\mathcal{V} = \mathcal{V}(A,\psi)$ de taille $\mathcal{O}(r^{-1/2})$ d'une collection de telles courbes intégrales « presque fermées » du champ de Reeb, de longueur totale bornée par $N\ell$.

Considérons maintenant la suite de solutions (A_n,ψ_n) qui apparaît dans l'énoncé du théorème 1.3 : la compacité de M et l'existence d'une borne indépendante de n sur la taille des voisinages $\mathcal{V}_n = \mathcal{V}(A_n,\psi_n)$ permettent de prouver l'existence d'une sous-suite pour laquelle $\mathcal{V}_n \supset \alpha_n^{-1}(0)$ converge vers une collection d'orbites fermées du champ de Reeb. Ceci achève la preuve du théorème 2.3.

3. HOMOLOGIE DE SEIBERG-WITTEN-FLOER ET ESTIMÉES

3.1. L'homologie de Seiberg-Witten-Floer, d'après Kronheimer-Mrowka [9]

En première approximation, l'homologie de Seiberg-Witten-Floer est l'homologie de Morse ($\mathcal{G}$-équivariante) d'une fonctionnelle dont les points critiques sont les solutions des équations de Seiberg-Witten. Plus précisément, soit $\mathcal{C} = \mathrm{Conn}(\det(\mathbb{S})) \times \Gamma(\mathbb{S})$ l'espace des configurations pour les équations de Seiberg-Witten, et soit $\mathbb{A}_{\mathrm{ref}}$ une connexion hermitienne de référence sur $\det(\mathbb{S})$. On pose

$$\mathfrak{a}_0(\mathbb{A},\psi) = -\frac{1}{8}\int_M (\mathbb{A}-\mathbb{A}_{\mathrm{ref}}) \wedge (F_{\mathbb{A}} + F_{\mathbb{A}_{\mathrm{ref}}}) + \frac{1}{2}\int_M \psi^* D_{\mathbb{A}}\psi.$$

Les points critiques de $\mathfrak{a}_0$ sont exactement les solutions de (1). De plus, lorsque $c_1(\det(\mathbb{S}))$ est une classe de torsion, la fonctionnelle $\mathfrak{a}_0$ est invariante sous l'action du groupe de jauge $\mathcal{G}$; dans le cas général, $\mathfrak{a}_0$ n'est invariante qu'à multiples de $2\pi^2$ près.

Le hessien de $\mathfrak{a}_0$ en un point critique $(\mathbb{A},\psi)$ est décrit par l'opérateur auto-adjoint $\mathcal{L}_{(\mathbb{A},\psi)} : \Omega^1(M,i\mathbb{R}) \oplus \Gamma(\mathbb{S}) \to \Omega^1(M,i\mathbb{R}) \oplus \Gamma(\mathbb{S})$ obtenu en linéarisant les équations (1). Plus précisément, à cause de l'invariance dans la direction des orbites de $\mathcal{G}$, on considère la restriction de $\mathcal{L}_{(\mathbb{A},\psi)}$ à $\mathrm{Ker}(d^*) \oplus \Gamma(\mathbb{S})$. Le spectre de cet opérateur est discret et non borné ; si l'on fixe un opérateur de référence $\mathcal{L}_0$, le flot spectral d'une famille d'opérateurs reliant $\mathcal{L}_0$ à $\mathcal{L}_{(\mathbb{A},\psi)}$ permet de définir l'indice d'un point critique non dégénéré de $\mathfrak{a}_0$: on pose $\deg(\mathbb{A},\psi) = -sf(\mathcal{L}_0, \mathcal{L}_{(\mathbb{A},\psi)})$.

Lorsque $c_1(\det(\mathbb{S}))$ est une classe de torsion, la situation est compliquée par la présence de solutions réductibles $(\mathbb{A},0)$, sur lesquelles le groupe de jauge agit avec stabilisateur S^1 (les transformations de jauge constantes). La solution utilisée par Kronheimer et Mrowka pour définir l'homologie de Seiberg-Witten-Floer consiste à modifier l'espace des configurations par un éclatement réel le long des configurations

réductibles [**9**]. Concrètement, il s'agit de remplacer $\mathscr{C}/\mathscr{G}$ par $\mathscr{C}^\sigma/\mathscr{G}$, où

$$\mathscr{C}^\sigma = \{(\mathbb{A}, \phi, s) \in \mathrm{Conn}(\det(\mathbb{S})) \times \Gamma(\mathbb{S}) \times \mathbb{R}_{\geq 0},\ \|\phi\|_{L^2} = 1\}.$$

La projection $\pi^\sigma : \mathscr{C}^\sigma \to \mathscr{C}$ définie par $\pi^\sigma(\mathbb{A}, \phi, s) = (\mathbb{A}, s\phi)$ permet de relever l'action de $\mathscr{G}$ sur $\mathscr{C}$ en une action libre sur $\mathscr{C}^\sigma$.

La fonctionnelle $\mathfrak{a}_0$ se relève en une fonctionnelle $\mathfrak{a}_0^\sigma$ sur $\mathscr{C}^\sigma$, dont les points critiques sont de deux types : d'une part, les solutions irréductibles de (1), et d'autre part, des points critiques réductibles où $\mathbb{A}$ est une connexion plate, ϕ est un vecteur propre de $D_\mathbb{A}$, et $s = 0$. De plus, le gradient de $\mathfrak{a}_0^\sigma$ est tangent au bord ($s = 0$) de $\mathscr{C}^\sigma/\mathscr{G}$.

Étant donnée une fonction de Morse f sur une variété à bord B, telle que le gradient de f soit tangent à ∂B et $f_{|\partial B}$ soit également une fonction de Morse, Kronheimer et Mrowka définissent trois versions du complexe de Morse, $\check{C}_*(f)$, $\hat{C}_*(f)$, et $\bar{C}_*(f)$. Ces complexes sont construits à partir de C^0_* engendré par les points critiques de f dans l'intérieur de B, C^s_* engendré par les points critiques *stables* de $f_{|\partial B}$, c'est-à-dire ceux où le hessien de f est positif dans la direction normale à ∂B, et C^u_* engendré par les points critiques *instables* de $f_{|\partial B}$, c'est-à-dire ceux où le hessien est négatif dans la direction normale au bord (l'indice de ces points critiques varie donc de 1 selon qu'on les considère comme points critiques de f ou de $f_{|\partial B}$). Le complexe $\bar{C}_*(f) = (C^s_* \oplus C^u_{*+1}, \bar{\partial})$ est alors le complexe de Morse de $f_{|\partial B}$, tandis que $\check{C}_*(f) = (C^0_* \oplus C^s_*, \check{\partial})$ et $\hat{C}_*(f) = (C^0_* \oplus C^u_*, \hat{\partial})$, où les différentielles $\check{\partial}$ et $\bar{\partial}$ sont définies par

$$\check{\partial} = \begin{pmatrix} \partial^0_0 & \partial^u_0 \bar{\partial}^s_u \\ \partial^0_s & \partial^s_s + \partial^u_s \bar{\partial}^s_u \end{pmatrix} \quad \text{et} \quad \hat{\partial} = \begin{pmatrix} \partial^0_0 & \partial^u_0 \\ \bar{\partial}^s_u \partial^0_s & \bar{\partial}^u_u + \bar{\partial}^s_u \partial^u_s \end{pmatrix}$$

où, pour $i, j \in \{0, s, u\}$, $\partial^i_j : C^i \to C^j$ compte les trajectoires du gradient de f contenues dans l'intérieur de B tandis que $\bar{\partial}^i_j : C^i \to C^j$ compte celles contenues dans ∂B. Les homologies des trois complexes sont reliées par une suite exacte :

$$\cdots \to \bar{H}_*(f) \to \hat{H}_*(f) \to \check{H}_*(f) \to \bar{H}_{*-1}(f) \to \cdots$$

qui coïncide avec la suite exacte pour l'homologie relative de $(B, \partial B)$.

En traitant de façon similaire les points critiques de $\mathfrak{a}_0^\sigma$ sur $\mathscr{C}^\sigma/\mathscr{G}$, Kronheimer et Mrowka construisent trois variantes de l'homologie de Seiberg-Witten-Floer pour une variété M munie d'une structure spin^c $\mathfrak{s}$, notées $\overline{HM}_*(M, \mathfrak{s})$, $\widetilde{HM}_*(M, \mathfrak{s})$ et $\widehat{HM}_*(M, \mathfrak{s})$. Dans le cas où $c_1(\det(\mathbb{S}))$ est une classe de torsion, ces groupes sont $\mathbb{Z}$-gradués ; sinon, ils sont $\mathbb{Z}/N$-gradués, où N est la divisibilité de $c_1(\det(\mathbb{S}))$. On a une suite exacte

$$\cdots \to \overline{HM}_*(M, \mathfrak{s}) \to \widehat{HM}_*(M, \mathfrak{s}) \to \widetilde{HM}_*(M, \mathfrak{s}) \to \overline{HM}_{*-1}(M, \mathfrak{s}) \to \cdots .$$

Le résultat suivant joue un rôle crucial dans la preuve du théorème 1.3 :

THÉORÈME 3.1 (Kronheimer-Mrowka [**9**], corollaire 35.1.4)

Pour toute variété M et toute structure spin^c $\mathfrak{s}$ telle que $c_1(\det(\mathbb{S}))$ soit une classe de torsion, l'ensemble des degrés $d \in \mathbb{Z}$ pour lesquels $\widehat{HM}_d(M,\mathfrak{s}) \neq 0$ est infini et n'admet pas de borne inférieure.

Ce résultat découle d'un théorème de structure pour $\overline{HM}_*(M,\mathfrak{s})$ [**9**, théorème 35.1.1] qui repose sur une analyse détaillée de la contribution des points critiques réductibles de $\mathfrak{a}_0^\sigma$, c'est-à-dire les points critiques de la forme $(\mathbb{A},\phi,0)$, où $\mathbb{A}$ est une connexion plate sur $\det(\mathbb{S})$ (dont l'existence est garantie par l'hypothèse sur la classe de Chern) et ϕ est un vecteur propre de l'opérateur de Dirac $D_{\mathbb{A}}$. (Les propriétés de compacité des équations de Seiberg-Witten impliquent que les points critiques irréductibles sont en nombre fini et n'affectent donc pas le résultat.)

3.2. Perturbation de contact et flot spectral

En présence d'une forme de contact a, les équations de Seiberg-Witten perturbées considérées par Taubes sont également les équations des points critiques d'une fonctionnelle $\mathfrak{a}$ sur l'espace des configurations. Comme au §2, on identifie l'espace des connexions spin^c avec l'espace des connexions hermitiennes sur le fibré en droites E ; l'espace des configurations est donc $\mathscr{C} = \mathrm{Conn}(E) \times \Gamma(\mathbb{S})$. On se restreint au cas où $c_1(\det(\mathbb{S})) = 2c_1(E) + c_1(\xi)$ est une classe de torsion ; il existe alors une connexion hermitienne A_E sur E dont la courbure vérifie $F_{A_E} = -\frac{1}{2}F_\xi$.

DÉFINITION 3.2. — *Pour un nombre réel $r \geq 1$ et une 1-forme $\mu \in \Omega^1(M,\mathbb{R})$ fixés, on associe à $(A,\psi) \in \mathrm{Conn}(E) \times \Gamma(\mathbb{S})$ les quantités suivantes :*

$$E(A) = \frac{i}{2}\int_M a \wedge F_A, \tag{16}$$

$$\mathfrak{cs}(A) = -\int_M (A - A_E) \wedge (F_A - F_{A_E}), \tag{17}$$

$$\mathfrak{a}(A,\psi) = \frac{1}{2}(\mathfrak{cs}(A) - rE(A)) + i\int_M \mu \wedge F_A + r\int_M \psi^* D_A \psi. \tag{18}$$

L'homologie de Seiberg-Witten-Floer peut être construite en utilisant la fonctionnelle $\mathfrak{a}$, dont les points critiques sont les solutions de (3), au lieu de $\mathfrak{a}_0$. L'invariance de la théorie vis-à-vis des perturbations implique que le théorème 3.1 reste valide pour la version perturbée de l'homologie de Seiberg-Witten-Floer.

Le théorème 3.1 implique l'existence de solutions des équations de Seiberg-Witten perturbées (3). En fait, il existe toujours des solutions réductibles, de la forme $(A,0)$ où $A = A_E - i\frac{r}{2}a + i\mu + i\theta$ avec $\theta \in \Omega^1(M,\mathbb{R})$ une 1-forme fermée. Toutefois ces solutions ne permettent pas de construire des orbites fermées du champ de Reeb ; en fait, elles ont une énergie $E(A) = \frac{r}{2}\int_M a \wedge da + \mathcal{O}(1) = r\,\mathrm{vol}(M,g) + \mathcal{O}(1)$ qui tend vers l'infini lorsque $r \to +\infty$, ce qui ne permet pas d'appliquer le théorème 2.3.

Il est donc nécessaire de prouver l'existence de solutions *irréductibles*. Taubes accomplit ceci en estimant le flot spectral d'une famille d'opérateurs de Dirac lorsque la connexion spin^c varie d'une connexion de référence à la connexion A considérée ; ceci lui permet d'estimer le flot spectral pour la linéarisation des équations (3), ce qui détermine le degré d'une solution donnée de (3) dans le complexe de Seiberg-Witten-Floer.

PROPOSITION 3.3 (Taubes [**18**], proposition 5.1). — *Pour $\mu \in \Omega^1(M,\mathbb{R})$ fixé, il existe une constante κ telle que, pour tout $r \geq 1$, le degré du générateur du complexe de Seiberg-Witten-Floer correspondant à une solution non dégénérée (A,ψ) de* (3) *vérifie*

$$\left|\deg(A,\psi) + \tfrac{1}{4\pi^2}\mathfrak{cs}(A)\right| \leq \kappa r^{31/16}. \tag{19}$$

Ce résultat implique en particulier que le degré des solutions réductibles (pour lesquelles $\mathfrak{cs}$ est de l'ordre de r^2) tend vers $-\infty$ lorsque $r \to \infty$; pour r suffisamment grand, les représentants d'une classe d'homologie de Seiberg-Witten-Floer de degré fixé sont donc toujours des solutions *irréductibles*.

Le point clé de la preuve consiste à montrer que le flot spectral pour une famille d'opérateurs de Dirac qui relie D_{A_E} à D_A est de l'ordre de $\frac{1}{8\pi^2}\mathfrak{cs}(A)$ [**18**, proposition 5.5]. L'argument repose sur une analyse détaillée du noyau de la chaleur.

Nous nous contenterons d'illustrer ce résultat en considérant le cas particulier de la famille de connexions $A = A_E - ira/2$. Soit $\psi = (\alpha,\beta)$ un vecteur propre de D_A, associé à une valeur propre λ telle que $|\lambda| \ll r^{1/2}$, et normalisé de telle sorte que $\|\psi\|_{L^2} = 1$.

Comme dans la preuve du lemme 2.4, la formule de Lichnerowicz-Weitzenböck permet d'écrire l'identité $D_A^2\psi = \lambda^2\psi$ sous la forme

$$\begin{aligned}(\nabla^A)^*\nabla^A\alpha - r\alpha + f_0(\nabla^A\alpha) + f_1(\nabla^A\beta) + f_2\alpha + f_3\beta &= \lambda^2\alpha,\\ (\nabla^A)^*\nabla^A\beta + r\beta + f_0'(\nabla^A\alpha) + f_1'(\nabla^A\beta) + f_2'\alpha + f_3'\beta &= \lambda^2\beta,\end{aligned}$$

où f_i et f_i' sont bornés indépendamment de r et ψ. Comme $|\lambda| \ll r^{1/2}$, la seconde équation permet de montrer que $\|\beta\|_{L^2}$ est borné par un multiple constant de $r^{-1/2}$, tandis que $\|\alpha\|_{L^2}$ est proche de 1. Si l'on considère maintenant ψ et λ comme des fonctions de r (que l'on suppose localement lisses), en dérivant l'égalité $D_A\psi = \lambda\psi$ par rapport à r on obtient :

$$-\frac{i}{4}cl(a)\psi + D_A\dot{\psi} = \dot{\lambda}\psi + \lambda\dot{\psi}. \tag{20}$$

D'autre part, la normalisation de ψ implique que $\langle\psi,\dot{\psi}\rangle_{L^2} = 0$. De même, $\langle\psi, D_A\dot{\psi}\rangle_{L^2} = \langle D_A\psi,\dot{\psi}\rangle_{L^2} = \lambda\langle\psi,\dot{\psi}\rangle_{L^2} = 0$. En prenant le produit scalaire de

(20) avec ψ, il vient donc

$$\dot{\lambda} = -\frac{i}{4}\int_M \psi^* cl(a)\psi = \frac{1}{4}\int_M |\alpha|^2 - |\beta|^2 \approx \frac{1}{4}.$$

Pour Λ fixé, la variation du flot spectral entre $r-\Lambda$ et $r+\Lambda$ est donc approximativement le nombre de valeurs propres de D_A telles que $|\lambda| \leq \Lambda/4$. L'estimée sur le flot spectral découle alors d'une estimation du nombre de telles valeurs propres par une quantité de l'ordre de $r\,\mathrm{vol}(M,g)\,\Lambda$, qui peut s'expliquer en utilisant (13) pour comparer l'équation de Dirac à l'équation $\overline{\partial}^A_\xi \alpha = 0$. Comme la courbure de la connexion A dans la direction des plans de contact est de l'ordre de $-i\frac{r}{2}\,da$, on peut construire des solutions approchées de l'équation de Dirac dont l'expression en coordonnées locales adaptées est de la forme $\exp(-\frac{r}{8}(|x|^2+|y|^2))\,h(x+iy)\,\chi(x+iy,z)$, où h est une fonction holomorphe et χ est une fonction *cut-off* qui s'annule à une distance de l'ordre de $r^{-1/2}$ dans la direction des plans de contact et de l'ordre de Λ^{-1} dans la direction du champ de Reeb. La dimension de l'espace des sections ainsi construites est de l'ordre de $r\,\mathrm{vol}(M)\,\Lambda$.

3.3. Estimation des fonctionnelles $\mathfrak{a}$, E et $\mathfrak{cs}$

La proposition 3.3 permet de contrôler la fonctionnelle $\mathfrak{cs}(A)$ pour une solution de (3) représentant une classe d'homologie de degré donné en homologie de Seiberg-Witten-Floer. Toutefois, le théorème 2.3 requiert un contrôle sur $E(A)$; pour ce faire, Taubes prouve le résultat suivant (cf. [**18**, corollaire 4.7]) :

Proposition 3.4 (Taubes). — *Étant donnée une structure spin^c telle que $c_1(\det(\mathbb{S}))$ soit une classe de torsion, soit $r \mapsto (A(r),\psi(r))$ une famille de solutions de* (3) *dont la dépendance en r est différentiable sauf pour un ensemble discret de valeurs de r pour lesquelles il peut y avoir des discontinuités. Si $\mathfrak{a}(r) = \mathfrak{a}(A(r),\psi(r))$ est une fonction continue de r, alors il existe une suite $r_n \to +\infty$ telle que $A_n = A(r_n)$ possède l'une des deux propriétés (mutuellement exclusives) suivantes :*

1. *il existe une constante $\epsilon > 0$ telle que $E(A_n) \geq \epsilon r_n$ et $\mathfrak{cs}(A_n) \geq \epsilon r_n^2$,*
2. *il existe une constante $\mathcal{E}$ telle que $E(A_n) \leq \mathcal{E}$ pour tout n.*

La preuve fait appel au résultat technique suivant (cf. [**18**, lemme 2.4]) :

Lemme 3.5 (Taubes). — *Il existe une constante κ telle que, si (A,ψ) est une solution de* (3)*, alors il existe une transformation de jauge $u : M \to S^1$ telle que*

$$|(A - u^{-1}du) - A_E| \leq \kappa r^{2/3}(|E(A)|+1)^{1/3} \tag{21}$$

et

$$|\mathfrak{cs}(A - u^{-1}du)| \leq \kappa r^{2/3}(|E(A)|+1)^{4/3}. \tag{22}$$

Idée de démonstration. — Le choix de la transformation de jauge u permet d'assurer que la 1-forme $\hat{a} = A - u^{-1}du - A_E$ vérifie $d^*\hat{a} = 0$; on peut alors décomposer $\hat{a}$ en la somme d'une forme harmonique h et d'une forme co-exacte b. De plus, quitte à modifier u par une transformation de jauge harmonique, on peut borner h par une constante dépendant uniquement de la métrique. Par ailleurs, $b = d^*G(F_A - F_{A_E})$, où G est l'opérateur de Green qui permet d'inverser le laplacien sur les 2-formes. Les estimées standard sur la fonction de Green impliquent alors une borne de la forme

$$|\hat{a}(x)| \leq c\left(\int_M \frac{|F_A|}{\mathrm{dist}(x,.)^2} + 1\right)$$

pour tout $x \in M$. Pour $\rho > 0$ fixé, l'intégrale sur le complémentaire de la boule de rayon ρ est bornée par $\rho^{-2}\int_M |F_A|$. L'équation (3) et les estimées du lemme 2.4 impliquent que $|F_A| = r(1 - |\alpha|^2) + \mathcal{O}(1)$; en utilisant (15), on a donc $\int_M |F_A| = E(A) + \mathcal{O}(1)$. D'autre part, comme $|F_A| = \mathcal{O}(r)$, l'intégrale sur la boule de rayon ρ centrée en x est bornée par un multiple de $r\rho$. On a donc une borne de la forme

$$|\hat{a}(x)| \leq c\rho^{-2}(E(A) + c_1) + c_2\, r\rho + c,$$

pour tout $\rho > 0$.

En choisissant ρ de façon optimale, c'est-à-dire $\rho = r^{-1/3}(E(A)+c_1)^{1/3}$, on obtient la borne voulue sur $|\hat{a}|$. La borne sur $\mathfrak{cs}(A - u^{-1}du) = -\int_M \hat{a}\wedge(F_A - F_{A_E})$ se déduit alors de la borne sur $|\hat{a}|$ et de l'estimée $\int_M |F_A| = E(A) + \mathcal{O}(1)$. □

Démonstration de la proposition 3.4. — Pour des raisons de simplicité, on se limitera au cas $\mu = 0$. Si l'on pose $\mathfrak{cs}(r) = \mathfrak{cs}(A(r))$ et $E(r) = E(A(r))$, la définition de $\mathfrak{a}$ implique alors que

$$\mathfrak{a}(r) = \frac{1}{2}(\mathfrak{cs}(r) - rE(r)).$$

En utilisant l'invariance de $\mathfrak{cs}$ par transformations de jauge lorsque $c_1(\det(\mathbb{S}))$ est une classe de torsion, le lemme 3.5 donne une borne sur $\mathfrak{cs}(r)$. Comme la proposition est clairement satisfaite lorsque $E(r) \leq 1$ pour une infinité de valeurs de r, on peut se restreindre au cas où $E(r) \geq 1$ pour $r \gg 1$, ce qui permet de réécrire (22) sous la forme

$$|\mathfrak{cs}(r)| \leq c_0 r^{2/3}E(r)^{4/3}. \tag{23}$$

On considère deux cas :

Cas 1. Supposons pour commencer qu'il existe $\eta > 0$ tel que $|\mathfrak{cs}(r)| \geq \eta r E(r)$ pour une infinité de valeurs de r. Alors, pour ces valeurs de r, on déduit de (23) que

$$r^{2/3}E(r)^{4/3} \geq \frac{1}{c_0}|\mathfrak{cs}(r)| \geq \frac{\eta}{c_0} rE(r),$$

d'où $E(r)^{1/3} \geq (\eta/c_0)\, r^{1/3}$, soit $E(r) \geq (\eta/c_0)^3\, r$, et donc $|\mathfrak{cs}(r)| \geq \eta(\eta/c_0)^3\, r^2$. On est donc dans le premier cas de l'alternative proposée par l'énoncé.

Cas 2. Sinon, il existe $\eta < \frac{1}{5}$ et r_0 tels que $|\mathfrak{cs}(r)| \leq \eta r E(r)$ pour tout $r \geq r_0$. Soit

$$\mathcal{V}(r) = -\frac{2\mathfrak{a}(r)}{r} = E(r) - \frac{1}{r}\mathfrak{cs}(r).$$

Pour $r \geq r_0$, on a $\mathcal{V}(r) \geq (1-\eta)E(r)$, et donc

$$|\mathfrak{cs}(r)| \leq \epsilon r \mathcal{V}(r) \tag{24}$$

où $\epsilon = \frac{\eta}{1-\eta} < \frac{1}{4}$. Lorsque $r \mapsto (A(r), \psi(r))$ est différentiable, le fait que $(A(r), \psi(r))$ est un point critique de $\mathfrak{a}$ implique que

$$\frac{d\mathfrak{a}(r)}{dr} = d\mathfrak{a}\left(\frac{dA(r)}{dr}, \frac{d\psi(r)}{dr}\right) + \frac{\partial \mathfrak{a}}{\partial r}(A(r), \psi(r)) = \frac{\partial \mathfrak{a}}{\partial r}(A(r), \psi(r)),$$

où

$$\frac{\partial \mathfrak{a}}{\partial r}(A, \psi) = -\frac{1}{2}E(A) + \int_M \psi^* D_A \psi = -\frac{1}{2}E(A).$$

On a donc

$$\frac{d\mathcal{V}}{dr} = \frac{2\mathfrak{a}}{r^2} - \frac{2}{r}\frac{d\mathfrak{a}}{dr} = \left(\frac{\mathfrak{cs}}{r^2} - \frac{E}{r}\right) + \frac{E}{r} = \frac{\mathfrak{cs}}{r^2}. \tag{25}$$

En combinant (25) et (24) on obtient l'inégalité $d\mathcal{V}/dr \leq \epsilon \mathcal{V}/r$, d'où l'on déduit par intégration une borne de la forme $|\mathcal{V}(r)| \leq C\, r^\epsilon$, et donc $E(r) \leq \frac{1}{1-\eta}\mathcal{V}(r) \leq C' r^\epsilon$.

La borne (23) implique alors que $|\mathfrak{cs}(r)| \leq c_0 C'^{4/3}\, r^{(2+4\epsilon)/3}$; on déduit de (25) que

$$\left|\frac{d\mathcal{V}}{dr}\right| = \frac{|\mathfrak{cs}|}{r^2} \leq \frac{c_0 C'^{\frac{4}{3}}}{r^{\frac{4}{3}(1-\epsilon)}}.$$

Puisque $\frac{4}{3}(1-\epsilon) > 1$, cette borne implique que $\mathcal{V}$ est une fonction bornée de r, et donc $E(r) \leq \frac{1}{1-\eta}\mathcal{V}(r)$ est également bornée. □

3.4. La preuve du théorème 1.3

La preuve du théorème 1.3 combine les ingrédients qui ont été introduits dans les sections précédentes. Elle peut être divisée en cinq étapes.

Étape 1. — Soit $\mathfrak{s}$ la structure spin^c sur M telle que $c_1(E) = e$. Le théorème 3.1 garantit l'existence d'une infinité de degrés $d \in \mathbb{Z}$ tels que $\widehat{HM}_d(M, \mathfrak{s}) \neq 0$. De plus, la proposition 3.3 garantit que, pour d fixé, il existe $r_0 = r_0(d)$ tel que, pour tout $r \geq r_0$, les solutions réductibles de (3) sont d'indice strictement inférieur à d.

Étape 2. — Soit $\theta \neq 0$ une classe d'homologie de Seiberg-Witten-Floer de degré d fixé ; si $c_1(E) = 0$, on impose $d \neq 0$ (la raison de cette condition sera expliquée à l'étape 4). Il n'est en général pas possible d'associer à θ une famille différentiable de solutions de (3). Le problème est le même qu'en théorie de Morse : lorsque r varie,

il peut y avoir des annulations entre points critiques de $\mathfrak{a}$. Toutefois, Taubes montre qu'il existe une famille de solutions $(A(r), \psi(r))$ qui est différentiable en dehors d'un ensemble discret de discontinuités, et telle que les valeurs critiques de $\mathfrak{a}$ correspondant à ces solutions dépendent continûment de r. L'idée est la suivante : pour r et μ fixés, étant donné un cycle η qui représente la classe θ, de la forme $\eta = \sum_i n_i p_i$ où p_i sont des points critiques de $\mathfrak{a}$, on pose $\hat{\mathfrak{a}}(\eta) = \max_i \mathfrak{a}(p_i)$. On définit alors

$$\mathfrak{a}_\theta(r) = \min\{\hat{\mathfrak{a}}(\eta) : \eta \text{ représente } \theta\}.$$

Taubes montre que, pour μ générique, $\mathfrak{a}_\theta$ est une fonction continue de r pour $r \geq r_0(d)$ [**18**, proposition 4.2]. Plus précisément, en choisissant μ de façon générique on peut s'assurer que, en dehors d'un ensemble discret de valeurs de r, les points critiques de $\mathfrak{a}$ qui nous intéressent sont non dégénérés, varient de façon différentiable avec r, et correspondent à des valeurs critiques distinctes. Comme en théorie de Cerf, on doit considérer divers types d'accidents (apparition ou disparition de points critiques, saut dans la différentielle en présence d'une trajectoire d'indice 0, égalité entre deux valeurs critiques) ; Taubes montre que la continuité de $\mathfrak{a}_\theta$ est préservée dans tous les cas. L'argument est en fait similaire à celui qui permet la construction d'invariants spectraux en théorie de Floer [**11, 13**]. La famille de solutions $(A(r), \psi(r))$ est alors obtenue en considérant les points critiques qui réalisent la valeur critique $\mathfrak{a}_\theta(r)$.

Étape 3. — Puisque les solutions $(A(r), \psi(r))$ correspondent à des points critiques de degré d fixé, l'estimation du flot spectral (proposition 3.3) implique que $|\mathfrak{cs}(A(r))|$ croît moins rapidement que r^2. Ceci élimine la première possibilité dans la proposition 3.4. Il existe donc une suite de valeurs $r_n \to +\infty$ telle que $E(A(r_n))$ soit bornée. Les solutions $(A_n, \psi_n) = (A(r_n), \psi(r_n))$ vérifient alors la condition (5) du théorème 2.3.

Étape 4. — Comme au §2, soient (α_n, β_n) les composantes de ψ_n. Lorsque $c_1(E)$ est non nul, l'existence d'un point où α_n s'annule implique immédiatement (en utilisant le lemme 2.4 pour borner β_n) que la condition (4) du théorème 2.3 est satisfaite.

Lorsque $c_1(E) = 0$ (ce qui ne peut arriver que si $c_1(\xi)$ est une classe de torsion), il existe toujours une classe d'équivalence de solutions de (3) pour laquelle la condition (4) n'est pas satisfaite. Si l'on choisit μ de telle sorte que $i\,d\mu = \frac{1}{2}F_\xi$, une telle solution est donnée par A la connexion triviale sur $E = \underline{\mathbb{C}}$ (la connexion spin^c sur $\mathbb{S} = \mathbb{S}_0$ est alors la connexion de référence ∇^0) et $\psi = (1, 0)$. Toutefois, Taubes montre que ces solutions particulières correspondent à des points critiques de $\mathfrak{a}$ de degré nul [**18**, lemme 5.4]. Le choix d'un degré $d \neq 0$ garantit donc que les solutions (A_n, ψ_n) vérifient (4).

Étape 5. — Les conditions (4) et (5) étant satisfaites, le théorème 2.3 s'applique et garantit l'existence d'orbites fermées du champ de Reeb.

4. HOMOLOGIE DE CONTACT PLONGÉE ET COHOMOLOGIE DE SEIBERG-WITTEN-FLOER

4.1. L'homologie de contact plongée

Soit M une variété compacte orientée de dimension 3, munie d'une forme de contact a. L'homologie de contact plongée $ECH_*(M,a)$ est l'homologie d'un complexe dont les générateurs sont des collections d'orbites fermées du champ de Reeb, muni d'une différentielle qui compte des courbes pseudo-holomorphes plongées dans la variété symplectique $\mathbb{R}\times M$; voir [**4**] et [**5**, §11].

Plus précisément, soit γ une orbite périodique du champ de Reeb v associé à la forme de contact a. Étant donné un point p de γ, la linéarisation du flot de Reeb induit sur le plan de contact $\xi_p = \mathrm{Ker}(a_p)$ une application de premier retour $P_\gamma \in \mathrm{End}(\xi_p)$, qui préserve la forme symplectique da. Un choix générique de la forme de contact garantit que toutes les orbites périodiques du champ de Reeb sont *non dégénérées*, c'est-à-dire que $P_\gamma - \mathrm{id}$ est inversible. L'orbite γ est dite *elliptique* si les valeurs propres de P_γ sont complexes conjuguées, *hyperbolique* si elles sont réelles.

Définition 4.1. — *Pour $\Gamma \in H_1(M,\mathbb{Z})$ fixé, on note $ECC_*(M,a,\Gamma)$ le groupe abélien libre engendré par tous les ensembles finis de la forme $\{(\gamma_i,m_i)\}$, où les γ_i sont des orbites périodiques plongées distinctes du champ de Reeb, les m_i sont des entiers positifs tels que $m_i=1$ lorsque γ_i est hyperbolique, et $\sum m_i[\gamma_i]=\Gamma$.*

Soit J une structure presque complexe sur $\mathbb{R}\times M$ (où on note s la coordonnée dans la direction réelle) invariante par translation dans la direction réelle et telle que $J(\frac{\partial}{\partial s}) = v$, $J(\xi)=\xi$, et $da(u,Ju)\geq 0$ pour tout $u\in\xi$. De plus, on choisit J de façon générique parmi les structures presque complexes qui possèdent ces propriétés, afin d'assurer de bonnes propriétés de transversalité pour les courbes pseudo-holomorphes.

Si γ est une orbite périodique de v, alors le cylindre $\mathbb{R}\times\gamma\subset\mathbb{R}\times M$ est pseudo-holomorphe. On considère des courbes pseudo-holomorphes dont le comportement pour $s\to\pm\infty$ est asymptotique à une collection de revêtements de tels cylindres : on dit qu'une courbe pseudo-holomorphe possède une extrémité positive (resp. négative) de multiplicité k le long de l'orbite γ si lorsque $s\to+\infty$ (resp. $s\to-\infty$) elle est asymptotique à un revêtement de degré k de $\mathbb{R}\times\gamma$. Il peut y avoir plusieurs extrémités asymptotiques à une même orbite.

Étant donnés deux générateurs $\underline{\gamma}=\{(\gamma_i,m_i)\}$ et $\underline{\gamma}'=\{(\gamma_i',m_i')\}$, on note $\mathcal{M}(\underline{\gamma},\underline{\gamma}')$ l'espace des modules des courbes pseudo-holomorphes dans $\mathbb{R}\times M$ (pas nécessairement connexes) dont les extrémités positives (resp. négatives) sont asymptotiques aux orbites γ_i (resp. γ_i'), et telles que la somme des multiplicités des extrémités positives (resp. négatives) le long de γ_i (resp. γ_i') soit égale à m_i (resp. m_i').

La théorie des courbes pseudo-holomorphes associe à une telle courbe C un indice

$$\mathrm{ind}(C) = -\chi(C) + 2c_1(\xi_{|C}, \tau) + \sum_{i,k} CZ_\tau(m_{i,k}\gamma_i) - \sum_{i,k} CZ_\tau(m'_{i,k}\gamma'_i),$$

où τ est un choix de trivialisation du champ de plans de contact ξ le long des orbites γ_i et γ'_i ; $c_1(\xi_{|C}, \tau)$ est le degré du fibré en droites complexes $\xi_{|C}$ relativement à la trivialisation τ aux extrémités de C ; les entiers $m_{i,k}$ (resp. $m'_{i,k}$), dont la somme vaut m_i (resp. m'_i), sont les multiplicités des diverses extrémités de C le long de γ_i (resp. γ'_i) ; enfin, CZ_τ est l'indice de Conley-Zehnder, qui mesure la rotation du flot de Reeb linéarisé par rapport à la trivialisation τ le long d'une orbite fermée. Pour J générique, au voisinage d'une courbe C dont aucune composante n'est un revêtement multiple, l'espace des modules $\mathcal{M}(\underline{\gamma}, \underline{\gamma}')$ est une variété lisse de dimension $\mathrm{ind}(C)$.

Hutchings définit également un second indice

$$I(C) = c_1(\xi_{|C}, \tau) + Q_\tau(C) + \sum_i \sum_{\ell=1}^{m_i} CZ_\tau(\ell\gamma_i) - \sum_i \sum_{\ell=1}^{m'_i} CZ_\tau(\ell\gamma'_i),$$

où Q_τ est une forme d'intersection en homologie relative, définie par rapport à la trivialisation τ : c'est le nombre d'intersection algébrique entre une surface plongée S qui représente la même classe relative que C et une déformation S' choisie de telle sorte que, pour $|s|$ suffisamment grand, les tresses obtenues au voisinage de chaque orbite de Reeb en intersectant S et S' avec $\{s\} \times M$ ont un nombre d'entrelacement nul vis-à-vis de la trivialisation τ.

Soit $H_2(M, \underline{\gamma}, \underline{\gamma}')$ l'espace affine modelé sur $H_2(M, \mathbb{Z})$ dont les éléments sont les classes d'homologie relative représentées par les chaînes Z telles que $\partial Z = \sum m_i\gamma_i - \sum m'_i\gamma'_i$. Alors la projection sur M d'une courbe pseudo-holomorphe $C \in \mathcal{M}(\underline{\gamma}, \underline{\gamma}')$ définit une classe $[C] \in H_2(M, \underline{\gamma}, \underline{\gamma}')$. Hutchings [**4**] montre que $I(C)$ ne dépend que de $\underline{\gamma}$, $\underline{\gamma}'$ et $[C]$. De plus, $I(C) - I(C') = \langle c_1(\xi) + 2\Gamma, [C] - [C'] \rangle$. Si on note N la divisibilité de la classe $c_1(\xi) + 2\Gamma$, cette propriété et l'additivité de I permettent de définir un *degré relatif* sur $ECC_*(M, a, \Gamma)$, à valeurs dans $\mathbb{Z}/N$, tel que

$$\deg(\underline{\gamma}) - \deg(\underline{\gamma}') \equiv I(C) \mod N$$

pour tout $C \in \mathcal{M}(\underline{\gamma}, \underline{\gamma}')$. De plus, Hutchings montre le résultat suivant :

THÉORÈME 4.2 (Hutchings [**4**]). — *Pour tout $C \in \mathcal{M}(\underline{\gamma}, \underline{\gamma}')$ dont aucune composante n'est un revêtement multiple, on a l'inégalité*

$$\mathrm{ind}(C) \leq I(C) - 2\delta(C), \tag{26}$$

où $\delta(C)$ est le nombre de points doubles de C. De plus, cette inégalité est une égalité si et seulement si C vérifie une condition « de partition », qui détermine les multiplicités individuelles des extrémités de C le long de chaque orbite γ_i ou γ'_i.

Hutchings montre également que si J est générique, alors toutes les courbes pseudo-holomorphes vérifient $I(C) \geq 0$, et $I(C) = 0$ si et seulement si C est une union de cylindres de la forme $\mathbb{R} \times \gamma_i$ ou de revêtements de tels cylindres, tandis que $I(C) = 1$ si et seulement si C est constituée d'une courbe plongée lisse C' telle que $I(C') = 1$ et qui vérifie la condition de partition, et d'une collection de cylindres disjoints de C'. Après passage au quotient par les translations dans la direction de $\mathbb{R}$, ces courbes sont des points isolés de l'espace des modules.

DÉFINITION 4.3. — *Étant donné un générateur $\underline{\gamma} \in ECC_*(M, a, \Gamma)$, on définit*

$$\partial(\underline{\gamma}) = \sum_{\underline{\gamma}'} \sum_{\substack{C \in \mathcal{M}(\underline{\gamma}, \underline{\gamma}')/\mathbb{R} \\ I(C)=1}} \varepsilon(C)\, \underline{\gamma}',$$

où $\varepsilon(C) \in \{\pm 1\}$ est un signe déterminé par l'orientation de l'espace des modules.

L'existence de l'homologie de contact plongée $ECH_*(M, a, \Gamma)$ découle alors du résultat suivant :

THÉORÈME 4.4 (Hutchings-Taubes [**6**]). — *∂ est bien définie, et $\partial^2 = 0$.*

4.2. Homologie de contact plongée et cohomologie de Seiberg-Witten-Floer

Soit M une variété compacte orientée de dimension 3, munie d'une forme de contact a pour laquelle les orbites périodiques du champ de Reeb sont non dégénérées. Étant donnée une classe d'homologie $\Gamma \in H_1(M, \mathbb{Z})$, soit E le fibré en droites complexes tel que $c_1(E) \in H^2(M, \mathbb{Z})$ soit la classe duale de Poincaré de Γ, et soit $\mathfrak{s}_\Gamma$ la structure spin^c dont le fibré des spineurs est $\mathbb{S}_\Gamma = \mathbb{S}_0 \otimes E = E \oplus (\xi \otimes E)$.

La *cohomologie de Seiberg-Witten-Floer* $\widehat{HM}^*(M, \mathfrak{s}_\Gamma)$ est définie en considérant le dual du complexe qui détermine l'homologie $\widehat{HM}_*(M, \mathfrak{s}_\Gamma)$ décrite au §3.1. En d'autres termes, c'est la *cohomologie* de Morse de la fonctionnelle $\mathfrak{a}_0$ (avec un traitement spécifique des points critiques réductibles, comme on l'a vu au §3.1). De façon équivalente, comme un changement d'orientation de M change le signe de $\mathfrak{a}_0$, on a un isomorphisme

$$\widehat{HM}^*(M, \mathfrak{s}_\Gamma) \simeq \widehat{HM}_{-*}(-M, \mathfrak{s}_\Gamma)$$

(cf. [**9**, corollaire 22.5.10]). On rappelle également qu'en général le degré n'est défini que de façon relative (la *différence* entre le degré de deux générateurs), et modulo un entier N, à savoir la divisibilité de $c_1(\det(\mathbb{S}_\Gamma)) = c_1(\xi) + 2\Gamma$.

Taubes a récemment annoncé [**20**] l'existence d'un isomorphisme entre homologie de contact plongée et cohomologie de Seiberg-Witten-Floer, qui change le signe du degré :

THÉORÈME 4.5 (Taubes). — *$ECH_*(M, a, \Gamma) \simeq \widehat{HM}^{-*}(M, \mathfrak{s}_\Gamma)$.*

Le texte [**20**] donne une description relativement détaillée de la preuve de ce théorème ; toutefois, l'argument repose sur une série d'articles actuellement en préparation.

Dans les grandes lignes, il s'agit de construire une application linéaire du complexe $ECC_*(M,a,\Gamma)$ vers le complexe de Seiberg-Witten-Floer, puis de montrer qu'elle est compatible avec les différentielles et induit un isomorphisme en (co)homologie. La construction repose sur le lien géométrique déjà mentionné au §2.2 entre solutions des équations de Seiberg-Witten perturbées (3) et orbites périodiques du champ de Reeb.

Plusieurs détails viennent compliquer la situation. Tout d'abord, la partie du complexe de Seiberg-Witten-Floer qui correspond aux solutions irréductibles est toujours de rang fini (même si ce rang peut augmenter avec r), alors que le complexe qui détermine l'homologie de contact plongée est en général de rang infini. Taubes considère une filtration naturelle sur le complexe $ECC_*(M,a,\Gamma)$: étant donné un générateur $\underline{\gamma} = \{(\gamma_i, m_i)\}$, on note $\ell(\underline{\gamma}) = \sum m_i \int_{\gamma_i} a$. La condition $da(\cdot, J\cdot) \geq 0$ imposée sur la structure presque complexe implique que, si $\mathcal{M}(\underline{\gamma}, \underline{\gamma}')$ est non vide, alors $\ell(\underline{\gamma}') \leq \ell(\underline{\gamma})$. Pour $L > 0$ fixé, on note $ECC_*^L(M,a,\Gamma)$ le sous-complexe de rang fini engendré par les générateurs tels que $\ell(\underline{\gamma}) \leq L$: on a alors

$$ECH_*(M,a,\Gamma) = \varinjlim_{L\to\infty} H_*(ECC_*^L(M,a,\Gamma), \partial).$$

Un examen de la preuve du théorème 2.3 suggère que les collections d'orbites du champ de Reeb telles que $\ell(\underline{\gamma}) \leq L$ correspondent aux limites pour $r \to \infty$ de familles de solutions irréductibles de (3) telles que $E(A) = i\int_M a \wedge F_A \leq \mathcal{E}(L) = 2\pi L$.

D'autre part, la construction d'une solution des équations de Seiberg-Witten à partir d'une collection d'orbites de Reeb est grandement simplifiée si l'on dispose d'un modèle local adapté au voisinage de ces orbites. Taubes montre [**20**, proposition 2.5] que, pour L fixé, il est possible de déformer la forme de contact a et la structure presque complexe J, sans affecter les orbites fermées du champ de Reeb de longueur au plus L ni la différentielle sur $ECC_*^L(M,a,\Gamma)$, de manière à garantir que le flot du champ de Reeb et la structure presque complexe soient décrits par un modèle local explicite au voisinage de chaque orbite fermée de longueur au plus L. Le modèle local souhaité est déterminé complètement par le type elliptique ou hyperbolique de l'orbite et par le nombre de rotation du flot de Reeb linéarisé le long de l'orbite ; voir la condition (2.11) de [**20**].

Pour L fixé, une fois a et J modifiés de la façon décrite ci-dessus, et pour r suffisamment grand, Taubes construit une bijection Φ^r entre les générateurs de $ECC_*^L(M,a,\Gamma)$ et les solutions de (3) telles que $E(A) \leq 2\pi L$: voir [**20**, théorème 4.2].

La bijection Φ^r est obtenue en plusieurs étapes. Tout d'abord, étant donnés une orbite fermée γ du champ de Reeb et un entier $m \geq 1$ (avec $m = 1$ si γ est hyperbolique), Taubes construit une solution approchée des équations (3) au voisinage

de γ à partir d'une famille de solutions d'énergie $2\pi m$ des équations vortex (12), $c_\gamma : S^1 \to \mathfrak{C}_m$ (cf. lemme 2.6). Plus précisément, étant donnée une famille de solutions $t \mapsto c(t) = (\mathcal{A}(t), \sigma(t))$ des équations vortex, on construit au voisinage de γ une connexion A sur E et une section ψ du fibré des spineurs dont les restrictions à une famille de disques transverses à γ, munis de la coordonnée complexe $r^{1/2}(x+iy)$, sont respectivement $\mathcal{A}(t)$ et $(\sigma(t), 0)$. Taubes montre alors que (A, ψ) est une solution approchée de (3) si et seulement si la famille de configurations $c(t)$ satisfait la condition

$$\frac{dc}{dt} = X_h(c(t)), \tag{27}$$

où X_h est le champ de vecteurs hamiltonien sur $\mathfrak{C}_m \simeq \mathrm{Sym}^m(\mathbb{C})$ associé à une fonction $h : \mathfrak{C}_m \to \mathbb{R}$ déterminée explicitement par le modèle local pour le flot de Reeb au voisinage de γ. Taubes montre de plus que, pour les modèles locaux définis par [**20**, condition (2.11)], l'homologie de Floer associée au hamiltonien h est

$$\begin{cases} \mathbb{Z} & \text{lorsque } m = 1, \\ \mathbb{Z} & \text{lorsque } m > 1 \text{ et } \gamma \text{ est elliptique}, \\ 0 & \text{lorsque } m > 1 \text{ et } \gamma \text{ est hyperbolique}. \end{cases}$$

Le dernier cas est exclu par les hypothèses.

Il est en fait facile de décrire la solution c_γ de (27) qui correspond au générateur de $HF_*(h)$: il s'agit de la trajectoire constante telle que, pour tout t, $c_\gamma(t)$ est l'unique solution d'énergie $2\pi m$ de (12) pour laquelle $\sigma^{-1}(0) = \{0\}$.

L'étape suivante consiste à recoller les solutions approchées de (3) ainsi construites au voisinage des différentes orbites γ_i qui composent $\underline{\gamma}$ avec une configuration de référence définie sur le reste de la variété M, pour laquelle la connexion A est plate et la section ψ est de norme constante égale à 1. Ceci permet d'associer à $\underline{\gamma}$ une solution approchée de (3). Taubes montre enfin que, pour r suffisamment grand, cette solution approchée peut être perturbée en une solution de (3), que l'on note $\Phi^r(\underline{\gamma})$.

L'injectivité de Φ^r découle immédiatement de la construction ; la surjectivité, quant à elle, est essentiellement une conséquence du théorème 2.3.

Une fois les générateurs des deux complexes identifiés par la bijection Φ^r, il reste à identifier les espaces de modules qui définissent les différentielles dans les deux théories. Ceci fait l'objet des théorèmes 4.3 et 4.4 de [**20**]. D'une part, étant donnés deux générateurs $\underline{\gamma}, \underline{\gamma}'$ de $ECC^L_*(M, a, \Gamma)$, Taubes montre que pour r suffisamment grand les courbes pseudo-holomorphes d'indice $I = 1$ qui relient $\underline{\gamma}$ à $\underline{\gamma}'$ sont en bijection avec les trajectoires d'indice 1 du gradient de la fonctionnelle $\mathfrak{a}$ qui relient $\Phi^r(\underline{\gamma})$ à $\Phi^r(\underline{\gamma}')$ [**20**, théorème 4.3]. La construction de cette bijection est similaire à celle utilisée dans l'article [**17**], mais présente des complications supplémentaires dues à la

non-compacité de $\mathbb{R} \times M$. D'autre part, Taubes montre également que, pour r suffisamment grand, la différentielle en cohomologie de Seiberg-Witten-Floer préserve le sous-groupe engendré par les solutions de (3) pour lesquelles $E(A) \leq 2\pi L$ [**20**, théorème 4.4] ; ceci est en fait une conséquence des estimées sur $\mathfrak{cs}$, $\mathfrak{a}$ et E qui ont été décrites au §3. Le théorème 4.5 est alors obtenu à partir de ces résultats par passage à la limite lorsque $L \to \infty$.

Entre autres conséquences remarquables du théorème 4.5, on notera celle-ci :

COROLLAIRE 4.6 (Taubes). — *L'homologie de contact plongée $ECH_*(M, a, \Gamma)$ ne dépend pas de la forme de contact, mais uniquement de la classe de cohomologie $c_1(\xi) + 2\Gamma$. De plus, l'homologie de contact plongée est de rang fini en chaque degré.*

RÉFÉRENCES

[1] C. ABBAS, K. CIELIEBAK & H. HOFER – The Weinstein conjecture for planar contact structures in dimension three, *Comment. Math. Helv.* **80** (2005), p. 771–793.

[2] D. T. GAY – Four-dimensional symplectic cobordisms containing three-handles, *Geom. Topol.* **10** (2006), p. 1749–1759.

[3] H. HOFER – Pseudoholomorphic curves in symplectizations with applications to the Weinstein conjecture in dimension three, *Invent. Math.* **114** (1993), p. 515–563.

[4] M. HUTCHINGS – An index inequality for embedded pseudoholomorphic curves in symplectizations, *J. Eur. Math. Soc. (JEMS)* **4** (2002), p. 313–361.

[5] M. HUTCHINGS & M. SULLIVAN – Rounding corners of polygons and the embedded contact homology of T^3, *Geom. Topol.* **10** (2006), p. 169–266.

[6] M. HUTCHINGS & C. H. TAUBES – Gluing pseudoholomorphic curves along branched covered cylinders. I, *J. Symplectic Geom.* **5** (2007), p. 43–137.

[7] ———, The Weinstein conjecture for stable Hamiltonian structures, *Geom. Topol.* **13** (2009), p. 901–941.

[8] D. Kotschick – The Seiberg-Witten invariants of symplectic four-manifolds (after C. H. Taubes), Séminaire Bourbaki, vol. 1995/96, exposé n° 812, *Astérisque* **241** (1997), p. 195–220.

[9] P. Kronheimer & T. Mrowka – *Monopoles and three-manifolds*, New Mathematical Monographs, vol. 10, Cambridge Univ. Press, 2007.

[10] F. Laudenbach – Orbites périodiques et courbes pseudo-holomorphes, application à la conjecture de Weinstein en dimension 3 (d'après H. Hofer et al.), Séminaire Bourbaki, vol. 1993/94, exposé n° 786, *Astérisque* **227** (1995), p. 309–333.

[11] Y.-G. Oh – Construction of spectral invariants of Hamiltonian paths on closed symplectic manifolds, in *The breadth of symplectic and Poisson geometry*, Progr. Math., vol. 232, Birkhäuser, 2005, p. 525–570.

[12] P. H. Rabinowitz – Periodic solutions of Hamiltonian systems, *Comm. Pure Appl. Math.* **31** (1978), p. 157–184.

[13] M. Schwarz – On the action spectrum for closed symplectically aspherical manifolds, *Pacific J. Math.* **193** (2000), p. 419–461.

[14] C. H. Taubes – The Seiberg-Witten and Gromov invariants, *Math. Res. Lett.* **2** (1995), p. 221–238.

[15] ———, SW $\Rightarrow$ Gr : from the Seiberg-Witten equations to pseudo-holomorphic curves, *J. Amer. Math. Soc.* **9** (1996), p. 845–918.

[16] ———, The geometry of the Seiberg-Witten invariants, in *Surveys in differential geometry, Vol. III (Cambridge, MA, 1996)*, Int. Press, Boston, MA, 1998, p. 299–339.

[17] ———, Gr $\Longrightarrow$ SW : from pseudo-holomorphic curves to Seiberg-Witten solutions, *J. Differential Geom.* **51** (1999), p. 203–334.

[18] ———, The Seiberg-Witten equations and the Weinstein conjecture, *Geom. Topol.* **11** (2007), p. 2117–2202.

[19] ———, The Seiberg-Witten equations and the Weinstein conjecture. II. More closed integral curves of the Reeb vector field, *Geom. Topol.* **13** (2009), p. 1337–1417.

[20] ———, Embedded contact homology and Seiberg-Witten Floer cohomology I, prépublication arXiv:0811.3985.

[21] C. Viterbo – A proof of Weinstein's conjecture in $\mathbf{R}^{2n}$, *Ann. Inst. H. Poincaré Anal. Non Linéaire* **4** (1987), p. 337–356.

[22] A. Weinstein – On the hypotheses of Rabinowitz' periodic orbit theorems, *J. Differential Equations* **33** (1979), p. 353–358.

Denis AUROUX

Massachusetts Institute of Technology
Department of Mathematics
77 Massachusetts Avenue
Cambridge, MA 02139
États-Unis
E-mail : auroux@math.mit.edu

332

ASTÉRISQUE

2010

SÉMINAIRE BOURBAKI
VOLUME 2008/2009
EXPOSÉS 997-1011

(1003) *Le théorème de la sphère différentiable*

Gérard BESSON

SOCIÉTÉ MATHÉMATIQUE DE FRANCE

Publié avec le concours du CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

Séminaire BOURBAKI Mars 2009
61e année, 2008-2009, no 1003, p. 161 à 181

LE THÉORÈME DE LA SPHÈRE DIFFÉRENTIABLE [d'après Brendle-Schoen]

par **Gérard BESSON**

INTRODUCTION

En 1951, H.E. Rauch ([**40**]) prouve qu'une variété riemannienne complète dont la courbure sectionnelle est positive et varie entre deux bornes dont le rapport est proche de 1 a un revêtement universel homéomorphe à une sphère. Plus précisément, pour $\delta > 0$, nous dirons qu'une variété riemannienne compacte M est *ponctuellement* δ-pincée lorsqu'en chaque point le rapport de la plus petite courbure sectionnelle à la plus grande est supérieur ou égal à δ. Nous dirons qu'elle est δ-pincée si c'est le rapport du minimum global (sur tout M) au maximum global qui vérifie cette inégalité. Enfin, elle est dite strictement δ-pincée (ponctuellement ou non) si cette inégalité est stricte. Cette notion peut s'utiliser aussi bien lorsque la courbure est négative que lorsqu'elle est positive. Dans ce texte, toute les variétés riemanniennes seront de courbure sectionnelle *strictement positive* ; pour des exemples de résultats en courbure négative on pourra consulter les articles [**27**] et [**43**]. Le théorème de Rauch affirme donc qu'une variété riemannienne simplement connexe δ-pincée, pour un δ explicite et proche de 1, est homéomorphe à une sphère. Par ailleurs, un calcul standard (voir [**21**] section 3.D.2, p. 149) montre que les projectifs complexes, de dimension complexe supérieure ou égale à 2, ont une courbure qui varie entre 1 et 4, pour un bon choix de leur normalisation ; c'est aussi le cas pour les autres espaces riemanniens symétriques compacts de rang un. Le théorème de Rauch ne peut donc être vrai pour une variété riemannienne de courbure positive 1/4-pincée (non strictement). Ce sont les travaux de M. Berger [**1**] et W. Klingenberg [**29**], dans les années 1960, qui conduisent au résultat optimal : une variété riemannienne simplement connexe *strictement* 1/4-pincée est homéomorphe à une sphère. Le lecteur peut consulter [**4**] (section 12.2.2.1, p. 552) pour une description de la technique ainsi qu'un historique plus précis.

Notons que ces résultats qui ressortissent à la relation courbure-topologie n'excluent pas que la variété soit une sphère exotique. C'est d'ailleurs une question intéressante de savoir si une sphère exotique peut porter une métrique riemannienne de courbure strictement positive. Un article récent ([**39**]), posté sur la Toile, propose un exemple pour lequel la réponse est positive. Le but de ce texte est de décrire le remarquable théorème suivant, dû à S. Brendle et R. Schoen.

THÉORÈME 0.1 (S. Brendle-R. Schoen [**10, 11**]). — *Soit M une variété riemannienne de courbure sectionnelle strictement positive et ponctuellement strictement $1/4$-pincée ; alors M admet une métrique de courbure sectionnelle constante. Elle est donc difféomorphe au quotient d'une sphère par un sous-groupe de $O(n)$.*

Une conséquence est donc qu'aucune sphère exotique n'admet de métrique strictement 1/4-pincée. Le résultat est important mais la méthode l'est tout autant. Elle repose sur l'utilisation du flot de la courbure de Ricci, introduit par R. Hamilton dans [**22**]. L'idée est de construire une déformation de la métrique riemannienne qui converge vers une métrique de courbure constante. Cette situation idéale ne se produit que lorsque la métrique de départ possède de bonnes propriétés. Rappelons que l'article séminal de R. Hamilton a connu des développements exceptionnels qui ont culminé en les travaux de G. Perelman ([**35, 36, 37**], voir aussi [**5**]) prouvant la conjecture de géométrisation. Dans [**22**], R. Hamilton prouve par cette méthode le théorème suivant.

THÉORÈME 0.2 (R. Hamilton [**22**]). — *Soit M une variété riemanienne compacte de dimension 3 qui possède une métrique de courbure de Ricci strictement positive ; alors elle possède une métrique de courbure constante.*

Ensuite, la même technique étendue à la dimension 4 permet à R. Hamilton de prouver les résultats contenus dans [**23, 26**].

C'est une remarquable généralisation aux dimensions quelconques qui constitue le cœur du travail que nous tentons de décrire et elle est due à Ch. Böhm et B. Wilking ([**6**]). On dira qu'un opérateur de courbure est 2-positif si la somme de ses deux plus petites valeurs propres est strictement positive. C'est une notion qui est apparue dans l'article [**14**] de H. Chen.

THÉORÈME 0.3 (Ch. Böhm-B. Wilking [**6**]). — *Soit M une variété riemannienne compacte d'opérateur de courbure 2-positif ; alors M admet une métrique de courbure constante.*

Les méthodes analytiques pour aborder 0.1 ne sont pas nouvelles. Elles sont présentes dans [**31**] où les auteurs utilisent la théorie des applications harmoniques et

dans [**23**], [**30**], [**28**] et [**34**] avec le flot de Ricci. Notons que dans [**30**], Ch. Margerin étudie une condition de courbure naturelle et donne un résultat optimal. Il s'agit d'analyse géométrique et un élément essentiel est le principe du maximum pour les systèmes paraboliques qui permet de réduire la question à un problème algébrique et de répondre à l'interrogation : quelles sont les propriétés (inégalités) de la courbure qui sont préservées par le flot de Ricci ?

Dans le texte qui suit, nous nous limiterons aux dimensions supérieures ou égales à 4. Après quelques rappels sur les différentes notions de courbure, nous tenterons de décrire les principales étapes de la preuve. La première consiste à utiliser le principe du maximum mentionné ci-dessus afin de réduire le problème à l'étude d'une équation différentielle ordinaire sur l'espace des endomorphismes de courbure. La seconde consiste à exhiber des inégalités satisfaites par la courbure qui sont invariantes par ce système dynamique. Enfin, des arguments géométriques standards permettent de conclure. Les références générales sont [**21**] pour les bases de la géométrie riemannienne, [**4**] pour un panorama exhaustif de ses réussites et [**15, 16, 17**] pour le flot de Ricci. Ces dernières références sont exhaustives et la méthode de Ch. Böhm et B. Wilking est décrite avec tous les détails dans [**16**], chapite 11. Il ne nous a donc pas semblé utile de les reproduire. Ce texte est à utiliser comme un guide pour la lecture des articles originaux, très informatifs, et de ces ouvrages. Nous conseillons également le récent survol de S. Brendle et R. Schoen [**12**], ainsi que le livre [**9**].

Je tiens à remercier chaleureusement Ch. Böhm, S. Brendle, L. Ni et H. Seshadri pour leurs réponses patientes à mes questions naïves, ainsi que M. Berger, P. Bérard, L. Bessières, J-.P Bourguignon, Z. Djadli, S. Gallot, H. Nguyen et Th. Richard pour des échanges fructueux.

1. RAPPELS DE GÉOMÉTRIE RIEMANNIENNE

Dans ce qui suit, (M, g) est une variété riemannienne compacte. Nous désignerons indifféremment la métrique riemannienne par g ou $\langle ., . \rangle$. On note ∇ la dérivation covariante. Soient X, Y, Z et T des champs de vecteurs ; on définit le tenseur de courbure de type $(0, 4)$ (voir [**21**]) par

$$R(X, Y, Z, T) = \langle \nabla_Y \nabla_X Z - \nabla_X \nabla_Y Z - \nabla_{[Y,X]} Z, T \rangle .$$

On montre que ce tenseur est antisymétrique par rapport à X et Y, ainsi que par rapport à Z et T et qu'il vérifie $R(X, Y, Z, T) = R(Z, T, X, Y)$. Il définit donc un endomorphisme symétrique de $\Lambda^2(TM)$ sur lui-même que nous noterons également R et qui s'appelle l'opérateur de courbure. Les conventions que nous utilisons ici sont celles de [**21**], le lecteur pourra les comparer à celles de [**16**]. Le tenseur de courbure

vérifie également la première identité de Bianchi, qui est l'analogue de l'identité de Jacobi vérifiée par le crochet de Lie d'une algèbre de Lie. Plus précisément, on a

$$R(X,Y,Z,T) + R(Y,Z,X,T) + R(Z,X,Y,T) = 0\,.$$

La courbure sectionnelle d'un 2-plan $P \subset T_mM$ tangent en m à M est

$$K(P) = \langle R(x \wedge y), x \wedge y\rangle\,,$$

où (x,y) est une base orthonormée de P. On note que K est la valeur de R calculée sur les vecteurs décomposés alors que R est défini sur tout $\Lambda^2(TM)$.

Si $\{e_i\}$ est une base orthonormée de T_mM, on munit $\Lambda^2(T_mM)$ du produit scalaire tel que $\{e_i \wedge e_j\}_{i<j}$ en est une base orthonormée. Nous noterons $R_{ijkl} = \langle R(e_i \wedge e_j), e_k \wedge e_l\rangle$. La courbure de Ricci en m est une forme bilinéaire symétrique sur T_mM que nous verrons également comme un endomorphisme symétrique de $\mathbf{R}^n$ (nous ferons l'abus de langage qui consiste à donner le même nom à ces deux objets). Elle est définie par

$$\mathrm{Ric}_{ij} = \mathrm{Ric}(e_i,e_j) = \langle \mathrm{Ric}(e_i), e_j\rangle = \sum_k R_{ikjk}\,.$$

C'est un tenseur de même nature que la métrique. Enfin, la courbure scalaire est la trace de la courbure de Ricci, c'est-à-dire, pour $m \in M$

$$\mathrm{scal}(m) = \sum_i \mathrm{Ric}_{ii} = \sum_{i,k} R_{ikik} = 2\,\mathrm{trace}\,R\,.$$

Avec ces conventions l'opérateur de courbure de la sphère unité est l'identité, ses courbures sectionnelles sont toutes égales à 1, sa courbure de Ricci est égale à $(n-1)g$ et sa courbure scalaire est constante égale à $n(n-1)$.

1.1. Opérateurs de courbure algébriques

Pour $m \in M$, le choix d'une base orthonormée $\{e_i\}$ de T_mM permet de l'identifier à l'espace euclidien $\mathbf{R}^n$. Le produit scalaire permet aussi d'identifier $\Lambda^2(\mathbf{R}^n)$ avec l'algèbre de Lie $\mathfrak{so}(n)$ en associant au vecteur unitaire $e_i \wedge e_j$ l'endomorphisme de rang 2 qui est la rotation d'angle $\pi/2$ dans le plan engendré par e_i et e_j. Via cette identification $\langle A,B\rangle = -1/2\,\mathrm{trace}(AB)$, pour $A,B \in \Lambda^2(\mathbf{R}^n)$. L'espace des endomorphismes symétriques (que nous identifierons aux formes bilinéaires symétriques) de $\Lambda^2(\mathbf{R}^n)$ est noté $\mathscr{S}^2(\mathfrak{so}(n))$. Cet espace encode les trois premières relations satisfaites par le tenseur de courbure. Nous appelons opérateur de courbure algébrique un élément de $\mathscr{S}^2(\mathfrak{so}(n))$ qui vérifie de plus la première identité de Bianchi ; l'espace des opérateurs de courbures algébriques est noté $\mathscr{S}^2_B(\mathfrak{so}(n))$ (voir [**16**], p. 81).

1.2. Opérations algébriques sur $\mathscr{S}^2(\mathfrak{so}(n))$

Soient A et B deux endomorphismes symétriques de $\mathbf{R}^n$; on définit un endomorphisme symétrique de $\mathfrak{so}(n) \simeq \Lambda^2(\mathbf{R}^n)$ par (voir **[16]**, p. 74) :

$$(A \wedge B)(v \wedge w) = \frac{1}{2}(A(v) \wedge B(w) + B(v) \wedge A(w)) .$$

On vérifie aisément que $A \wedge B \in \mathscr{S}^2(\mathfrak{so}(n))$ et que la relation $A \wedge B = B \wedge A$ est satisfaite.

Soient maintenant R et S deux éléments de $\mathscr{S}^2(\mathfrak{so}(n))$; on définit $R\sharp S \in \mathscr{S}^2(\mathfrak{so}(n))$ par la relation :

$$\langle (R\sharp S)(h), h\rangle = \frac{1}{2}\sum_{\alpha,\beta}\langle [R(\omega_\alpha), S(\omega_\beta], h\rangle \,.\, \langle [\omega_\alpha, \omega_\beta], h\rangle ,$$

où $h \in \mathfrak{so}(n)$ et $\{\omega_\alpha\}$ est une base orthonormée de $\mathfrak{so}(n)$. On vérifie que la définition est indépendante du choix de la base et que $R\sharp S = S\sharp R$ (voir **[6]** et **[16]**, p. 72). Enfin nous noterons $R^\sharp = R\sharp R$. Cette opération introduite dans les travaux de R. Hamilton (voir par exemple **[23]**) jouera un rôle fondamental dans la suite.

1.3. Composantes irréductibles sous l'action de $O(n)$

Le groupe $O(n)$ agit par changement de base sur l'espace des opérateurs de courbure algébriques. On peut décomposer l'espace $\mathscr{S}^2_B(\mathfrak{so}(n))$ en composantes irréductibles sous cette action qui sont au nombre de trois. Pour un opérateur de courbure R, notons Ric_0 la partie sans trace de son tenseur de Ricci, c'est-à-dire

$$\langle \mathrm{Ric}(R)(e_i), e_i\rangle = \sum_{k=1}^{n} R_{ikik} \quad \text{et} \quad \mathrm{Ric}_0(R) = \mathrm{Ric}(R) - \frac{\mathrm{scal}(R)}{n} I_{\mathbf{R}^n} ,$$

où $\mathrm{scal}(R) = 2\,\mathrm{trace}(R)$. Alors, on a

$$\mathscr{S}^2_B(\mathfrak{so}(n)) = \mathbf{R} I_{\mathscr{S}^2_B} \oplus \langle \mathrm{Ric}_0\rangle \oplus \langle W\rangle .$$

Le premier facteur est constitué des opérateurs de courbure algébriques multiples de l'identité, le second des multiples d'opérateurs du type $A \wedge I_{\mathbf{R}^n}$ où A est un opérateur symétrique de $\mathbf{R}^n$ de trace nulle et le troisième est le noyau de l'application $R \mapsto \mathrm{Ric}(R)$. Ce dernier est l'espace des tenseurs de Weyl, qui constituent la composante de R la plus difficile à comprendre. Rappelons que la composante de Weyl du tenseur de courbure d'une métrique riemannienne est nulle si et seulement si celle-ci est localement conformément plate. Pour un tenseur de courbure R, nous noterons R_I, R_{Ric_0} et R_W ses différentes composantes et nous noterons indifféremment I l'identité de $\mathscr{S}^2_B(\mathfrak{so}(n))$ comme celle de $\mathbf{R}^n$. Nous avons donc (voir **[16]**, p. 87, pour les détails),

$$R = \frac{\mathrm{scal}(R)}{n(n-1)} I + \frac{2}{n-2}\,\mathrm{Ric}_0(R) \wedge I + W .$$

1.4. Une nouvelle identité

Ch. Böhm et B. Wilking démontrent dans [**6**] une nouvelle identité satisfaite par tout tenseur de courbure algébrique. Elle est remarquable de simplicité.

PROPOSITION 1.1 ([**6**]). — *Pour tout $R \in \mathcal{S}^2_B(\mathfrak{so}(n))$, on a*

$$R + R\sharp I = (n-1)R_I + \frac{n-2}{2} R_{\mathrm{Ric}_0} = \mathrm{Ric} \wedge I\,.$$

Nous en déduirons ci-dessous la décomposition selon les composantes irréductibles de l'expression quadratique en R, qui apparaît dans l'équation d'évolution (le long du flot de Ricci) de l'opérateur de courbure. Il est remarquable que de nouvelles identités puissent être découvertes sur un espace aussi étudié.

1.5. Endomorphismes $0(n)$-invariants de $\mathcal{S}^2_B(\mathfrak{so}(n))$

Soit ℓ une application linéaire auto-adjointe de $\mathcal{S}^2_B(\mathfrak{so}(n))$ dans lui-même, équivariante pour l'action de $O(n)$ sur cet espace. Elle est diagonalisable et ses espaces propres sont $O(n)$ invariants ; elle préserve donc les composantes irréductibles décrites ci-dessus. De plus, elle est un multiple de l'identité en restriction à chacune d'elle (voir [**16**], p. 89). Dans [**6**], les auteurs s'intéressent à de telles applications qui préservent la composante de Weyl. Elles s'écrivent sous la forme générale

$$\ell_{a,b}(R) = R + 2(n-1)aR_I + (n-2)bR_{\mathrm{Ric}_0} = (1+2(n-1)a)R_I + (1+(n-2)b)R_{\mathrm{Ric}_0} + R_W\,,$$

où a et b sont des nombres réels.

2. LE FLOT DE RICCI

C'est une équation différentielle ordinaire sur l'espace des métriques inventée par R. Hamilton dans l'article séminal [**22**]. On cherche une famille $g(t)$ de métriques lisses, dépendant de manière C^∞ de t, et solution du problème de Cauchy suivant :

$$\begin{cases} \dfrac{\partial g}{\partial t} &= -2\,\mathrm{Ric}_{g(t)} \\ g(0) &= g_0. \end{cases}$$

Une version normalisée peut être écrite, pour laquelle le volume de la métrique qui évolue est fixé ; il suffit, en effet, de remplacer la première ligne de l'équation précédente par :

$$\frac{\partial g}{\partial t} = -2\,\mathrm{Ric}_{g(t)} + \Big(\frac{2}{n}\frac{1}{\mathrm{vol}(M, g(t))}\int_M \mathrm{scal}(x,t)dv_{g(t)}(x)\Big)g(t).$$

On peut consulter [**5**] ainsi que [**17**] pour une description détaillée des vertus de cette théorie. Les théorèmes classiques permettent de montrer que les solutions existent en temps petit pour toute donnée initiale C^∞.

L'idée est de prouver que sous les hypothèses des théorèmes 0.1 et 0.3 le flot de Ricci normalisé converge vers une métrique de courbure constante.

2.1. Évolution des courbures

Un calcul maintenant classique permet de décrire l'évolution des différents types de courbure. Nous nous limiterons à la courbure scalaire et à l'opérateur de courbure. On a alors

$$\frac{\partial \operatorname{scal}}{\partial t} + \Delta_{g(t)} \operatorname{scal} = 2|\operatorname{Ric}_{g(t)}|^2_{g(t)},$$

où $\Delta_{g(t)}$ désigne le laplacien, agissant sur les fonctions, défini par la métrique $g(t)$. La convention adoptée est celle dite « des géomètres » (en dimension 1, c'est $-d^2/dx^2$). La norme du tenseur de Ricci est prise pour la métrique $g(t)$.

De même on a

$$\frac{\partial R}{\partial t} + \Delta R = 2(R^2 + R^\sharp).$$

Nous utilisons ici le laplacien brut, c'est-à-dire l'opposé de la trace de la dérivée covariante seconde du tenseur R. La notation R^2 désigne le carré de l'endomorphisme R et $R^\sharp$ désigne l'expression quadratique définie dans la section précédente. L'évolution du tenseur de courbure ne s'écrit aussi simplement que si nous utilisons une astuce due à K. Uhlenbeck que nous décrivons ci-dessous.

2.2. Le principe du maximum

C'est l'outil indispensable pour l'étude des solutions de l'équation de la chaleur. Il faut en donner ici une version vectorielle, c'est-à-dire une version pour les systèmes paraboliques. C'est ce que fait R. Hamilton dans [**22**] et [**23**]. Le lecteur intéressé peut également consulter [**17**] et [**16**].

Considérons une équation aux dérivées partielles du type $\partial s/\partial t + \Delta_t s = F(s)$ et l'équation différentielle ordinaire $ds/dt = F(s)$. Les E.D.P. ci-dessus sont dites de *réaction-diffusion*, le terme de diffusion est donné par le laplacien ; si $F = 0$, il s'agit d'une équation de la chaleur qui « étale » les solutions. Le terme non-linéaire $F(s)$ est le terme de réaction qui, en l'absence de laplacien, conduit (souvent) à l'explosion en temps fini des solutions (convergence de certaines normes vers $+\infty$). La question est de savoir qui de la réaction ou de la diffusion l'emportera, dans une situation donnée. Le principe du maximum résulte d'une comparaison entre le comportement des solutions de l'équation aux dérivées partielles et la situation extrême donnée par l'équation différentielle ordinaire.

L'équation satisfaite par la courbure scalaire peut s'écrire $\partial R/\partial t + \Delta R \geq 0$. Le principe du maximum scalaire le plus simple conduit alors au fait que le minimum de la courbure scalaire augmente le long du flot de Ricci.

Nous décrivons maintenant une version vectorielle de ce principe. Soit M munie d'une famille C^∞ de métriques $g(t)$, pour $t \in [0,T]$, et soit $\pi : \mathcal{E} \longrightarrow M$ un fibré vectoriel muni d'une métrique *fixe* et d'une famille C^∞ de connexions compatibles, ∇_t. Ces données permettent de définir un laplacien agissant sur les sections de $\mathcal{E}$, qui dépend de t et que nous noterons simplement Δ. Considérons une fonction $F : \mathcal{E} \times [0,T] \longrightarrow \mathcal{E}$ de classe C^∞ (pour simplifier) telle que, pour t donné, $F(.,t)$ préserve les fibres. Soit $\mathcal{K}$ un fermé de $\mathcal{E}$ que nous supposons invariant par le transport parallèle de ∇_t, pour tout $t \in [0,T]$, et tel que $\mathcal{K}_m = \mathcal{K} \cap \pi^{-1}(m)$ soit fermé et convexe. L'hypothèse clé est une relation entre $\mathcal{K}$ et l'équation différentielle ordinaire $\frac{du}{dt} = F(u)$, définie dans chaque fibre $\mathcal{E}_m$ de $\mathcal{E}$; nous supposons que toute solution de celle-ci telle que $u(0) \in \mathcal{K}_m$ reste dans $\mathcal{K}_m$ pour tout $t \in [0,T]$.

THÉORÈME 2.1 ([**23**] [**22**] ou [**17**], théorème 4.8). — *Avec les hypothèses ci-dessus, soit $s(t)$ une solution de l'E.D.P.*

$$\frac{\partial s}{\partial t} + \Delta s = F(s),$$

telle que $s(o) \in \mathcal{K}$; alors, pour tout $t \in [0,T]$, $s(t) \in \mathcal{K}$.

Pour appliquer ceci à l'opérateur de courbure il faut noter que, bien que la métrique du fibré $\Lambda^2(T^*M)$ dépende de t, une astuce, due à K. Uhlenbeck (voir ci-dessous et [**17**], section 6.1), permet de se ramener à une métrique fixe sur un fibré fixe avec toutefois une connexion qui dépend du temps.

L'ensemble $\mathcal{K}$ est à penser comme une version géométrique d'inégalités sur la courbure comme par exemple la positivité de l'opérateur de courbure.

Exemple : Rappelons le cas de la dimension 3, pour lequel ce principe du maximum a joué un rôle important dans les travaux de G. Perelman ; il montre en effet que la courbure scalaire contrôle tout le tenseur de courbure. En chaque point $m \in M$ et $t \in [0,T]$ l'endomorphisme R se diagonalise et possède trois valeurs propres notées $\lambda(m,t) \geq \mu(m,t) \geq \nu(m,t)$. Une particularité de la dimension 3 est que ces nombres sont toujours des courbures sectionnelles. On montre que les deux expressions R^2 et $R^\sharp$ se diagonalisent dans la même base et ont pour valeurs propres respectives $(\lambda^2, \mu^2, \nu^2)$ et $(\mu\nu, \lambda\nu, \lambda\mu)$. L'équation différentielle ordinaire prend donc une forme très simple

$$\begin{cases} \frac{d\lambda}{dt} &= \lambda^2 + \mu\nu \\ \frac{d\mu}{dt} &= \mu^2 + \lambda\nu \\ \frac{d\nu}{dt} &= \nu^2 + \lambda\mu. \end{cases}$$

En général nous aurons donc à considérer l'équation différentielle ordinaire (E.D.O.)

$$\frac{dR}{dt} = R^2 + R^\sharp = Q(R)\,,$$

définie sur $\mathcal{S}^2_B(\mathfrak{so}(n))$ (nous omettrons le facteur 2 qui ne joue aucun rôle).

2.3. Constructions de $\mathcal{K}$

Il est maintenant utile de rappeler l'astuce due à K. Uhlenbeck afin de se ramener à un fibré vectoriel fixe. Considérons une solution $(M, g(t))$ du flot de Ricci, définie sur un intervalle $[0, T)$. On peut construire une isométrie de fibrés vectoriels dépendant du temps

$$\iota(t) : (TM, g(0)) \longrightarrow (TM, g(t))$$

en résolvant l'équation

$$\frac{d\iota}{dt}(t) = \mathrm{Ric}_{g(t)} \circ \iota(t) \quad \text{ et } \iota(0) = id\,,$$

où Ric est ici vu comme un endomorphisme. Elle permet de tout ramener à un fibré vectoriel métrique fixe. En particulier la connexion de Levi-Civita rappelée dépend du temps. Cette isométrie s'étend à tous les fibrés vectoriels naturels construits à partir de TM comme le fibré des tenseurs de courbure $\mathcal{E}$ dont la fibre type est isométrique à $\mathcal{S}^2_B(\mathfrak{so}(n))$. Alors, en considérant $\iota(t)^*(R_g(t))$, on écrit l'équation d'évolution de l'opérateur de courbure sur ce fibré fixe, ce qui permet d'appliquer le principe du maximum énoncé ci-dessus.

Pour construire des ensembles $\mathcal{K}$ comme ci-dessus dans $\mathcal{E}$, nous pouvons procéder comme suit. Considérons un fermé convexe $F \subset \mathcal{S}^2_B(\mathfrak{so}(n))$. Nous supposerons de plus que F est invariant par l'action de $O(n)$ sur les tenseurs de courbure. Nous pouvons alors transporter F sur chaque fibre $\mathcal{E}_m$ par l'isométrie induite par le choix d'une base orthonormée de T_mM ; l'ensemble image est $\mathcal{K}_m$. L'invariance par $O(n)$ de F garantit que cette construction ne dépend pas du choix de la base. De même le transport parallèle associé à la connexion (dépendant du temps) induit une isométrie entre les fibres de $\mathcal{K}$; l'invariance de F par $O(n)$ assure l'invariance de $\mathcal{K}$ par le transport parallèle par toutes ces connexions. Si de plus F est invariant par l'E.D.O., l'ensemble $\mathcal{K}$ ainsi construit vérifiera l'hypothèse du théorème 2.1. Un tel ensemble correspond à des conditions de courbure préservées par le flot.

3. LA MÉTHODE DE BÖHM ET WILKING

Le problème est donc maintenant ramené à l'étude du terme quadratique $Q(R)$, le second membre de l'équation différentielle ordinaire, qu'il faut interpréter comme la valeur en R d'un champ de vecteurs sur $\mathscr{S}^2_B(\mathfrak{so}(n))$. Une condition de courbure invariante peut être définie par un cône convexe fermé $O(n)$-invariant, par exemple le cône C des opérateurs de courbure 2-positifs ou nuls (la somme de leurs deux plus petites valeurs propres est positive ou nulle). Pour montrer l'invariance de cette condition par l'E.D.O., il faut montrer que si une trajectoire qui part de l'intérieur du cône atteint le bord elle ne peut en sortir ; pour cela il suffit de prouver que le champ de vecteurs qui régit cette E.D.O. « repousse » la trajectoire vers l'intérieur ou, dans le pire des cas, la fait évoluer sur le bord du cône. Il faut donc prouver que le champ de vecteurs $Q(R)$ est dans le cône tangent T_RC à C en tout point de ∂C. Soit donc C un tel cône, invariant par l'E.D.O. ; on peut en construire d'autres en prenant l'image de C par les applications $\ell_{a,b}$ définies ci-dessus, pour des choix convenables de a et b. Les applications linéaires $\ell_{a,b}$ envoyant bord sur bord et cône tangent sur cône tangent, l'ensemble $\ell_{a,b}(C)$ est invariant par l'E.D.O., c'est-à-dire que C est invariant par $\ell_{a,b}^{-1} \circ Q \circ \ell_{a,b}$ si

$$X_{a,b}(R) = \ell_{a,b}^{-1}((\ell_{a,b}(R))^2 + (\ell_{a,b}(R))^\sharp)$$

est dans le cône tangent T_RC à C en tout point de ∂C. Comme c'est le cas de $Q(R)$ il suffit, par convexité de C, de le montrer pour $D_{a,b}(R) = X_{a,b}(R) - Q(R)$. Les applications $\ell_{a,b}$ préservent la composante de Weyl et on montre qu'en conséquence $D_{a,b}(R)$ ne dépend que de la composante de Ricci de l'opérateur de courbure R. Les formules deviennent alors plus simples et les calculs possibles.

Grâce à cette idée lumineuse, Ch. Böhm et B. Wilking construisent une famille de cônes convexes $C(s)_{s\in[0,1]}$, $O(n)$-invariants, invariants par l'E.D.O. dont l'intersection est constituée des opérateurs de courbure λI, où $\lambda > 0$, c'est-à-dire de l'opérateur de courbure de la sphère à normalisation près. Si la théorie fonctionne, les trajectoires de l'E.D.P. seront « piégées » par cette famille et contraintes de converger vers l'opérateur de courbure de la sphère. L'idée sous-jacente est d'essayer d'obtenir une fonction de Lyapunov associée à l'E.D.O. (en restriction et projection sur la sphère unité de $\mathscr{S}^2_B(\mathfrak{so}(n))$) dont le minimum serait l'opérateur de courbure de la sphère afin de montrer que les trajectoires convergent vers ce point fixe. Une telle fonction n'est pas vraiment nécessaire ; en effet, ce sont plutôt ses ensembles de niveau qui jouent le rôle clé. Les cônes ci-dessus que nous décrirons plus loin généralisent cette idée. Ils ne sont pas nécessairement inclus les uns dans les autres mais peuvent être utilisés de la même manière.

Précisons quelques points.

3.1. Propriétés de $Q(R)$ et $D_{a,b}(R)$

Nous listons un certain nombre de propriétés de ces deux quantités. Le lecteur est renvoyé à [**6**] et [**16**] pour les détails.

Pour $R \in \mathscr{S}^2_B(\mathfrak{so}(n))$, on montre que $Q(R) \in \mathscr{S}^2_B(\mathfrak{so}(n))$ (voir [**16**] p. 88). Une observation de G. Huisken montre que l'E.D.O. que nous considérons est le flot de gradient de la fonction

$$P(R) = \frac{1}{3}\operatorname{trace}(R^3 + RR^\sharp)\,,$$

qui est malheureusement trop difficile à étudier pour servir de fonction de Lyapunov. Enfin, on peut calculer les composantes dans la décomposition irréductible de $Q(R)$. C'est là que la nouvelle identité est utile.

Le théorème suivant est la clé de l'étude de l'invariance des cônes $\ell_{a,b}(C)$.

THÉORÈME 3.1 (Ch. Böhm-B. Wilking [**6**]). — *Pour a et b des nombres réels,*

$$\begin{aligned} D_{a,b}(R) =& ((n-2)b^2 - 2(a-b))\operatorname{Ric}_0 \wedge \operatorname{Ric}_0 + 2a \operatorname{Ric} \wedge \operatorname{Ric} + 2b^2 \operatorname{Ric}_0^2 \wedge I \\ &+ \frac{\operatorname{trace}(\operatorname{Ric}_0^2)}{n + 2n(n-1)a}(nb^2(1-2b) - 2(a-b)(1-2b+nb^2))I\,. \end{aligned}$$

Le fait remarquable, prévisible au vu de la construction, est que $D_{a,b}(R)$ est indépendant de la composante de Weyl de R. Par exemple, on peut calculer ses valeurs propres en fonction de celles de Ric. Il suffit en effet de choisir une base orthonormée de $\mathbf{R}^n$ qui diagonalise Ric, on en déduit une base orthonormée de $\mathfrak{so}(n)$ dans laquelle $D_{a,b}(R)$ se diagonalise. Ses valeurs propres dépendent alors du choix de a et b.

3.2. Construction de la famille de cônes

Rappelons que $n \geq 4$. Nous posons $\bar{b} = \frac{\sqrt{2n(n-2)+4}-2}{n(n-2)}$ et $a_0(b) = b + \frac{(n-2)}{2}b^2$. Ces valeurs sont telles que $D_{a_0(b),b}$ soit strictement positif pour $b \in]0, \bar{b}]$. Les propositions suivantes donnent le point de départ de la construction (voir [**6**] et [**16**], p. 113) :

PROPOSITION 3.2 (R. Hamilton [**25**]). — *Le cône $\mathfrak{C}$ des opérateurs de courbure algébriques 2-positifs ou nuls est préservé par l'E.D.O.*

Notons que ce résultat est valable en dimension 3 également. Dans ce cas, l'hypothèse signifie que la courbure de Ricci est positive ou nulle (voir [**23**]). Soit $\{\omega_i\}$ une base orthonormée de $\mathfrak{so}(n)$ constituée de vecteurs propres de R. Montrer que le cône est préservé revient à prouver que $\langle Q(R)\omega_i, \omega_i\rangle + \langle Q(R)\omega_j, \omega_j\rangle \geq 0$ pour tout couple $i < j$ tel que $\langle R\omega_i, \omega_i\rangle + \langle R\omega_j, \omega_j\rangle = 0$; c'est-à-dire que le champ de vecteurs $Q(R)$ « repousse » la trajectoire vers l'intérieur du cône (au sens large). On utilise ici la description précise de $Q(R)$.

3.2.1. *Première étape.* —

PROPOSITION 3.3. — *Soit $\mathfrak{C}$ le cône des opérateurs de courbure 2-positifs ou nuls pour $n \geq 4$ et $b \in]0, \bar{b}]$; alors on a les propriétés suivantes :*

(i) $\mathfrak{C}_b = \ell_{a_0(b),b}(\mathfrak{C})$ *est préservé par l'E.D.O.*
(ii) *Pour $b > 0$, le champ de vecteurs $Q(R)$ est transverse au bord de $\ell_{a_0(b),b}(\mathfrak{C})$ aux points $R \neq 0$.*
(iii) $\ell_{a_0(\bar{b}),\bar{b}}(\mathfrak{C} \setminus \{0\})$ *est inclus dans le cône des opérateurs de courbure positifs.*

C'est pour prouver cette proposition que l'on utilise la description de $D_{a,b}(R)$ en suivant le schéma mentionné ci-dessus. On montre essentiellement que $D_{a,b}(R)$ est positif ou nul. Ceci nous fournit une première famille de cônes qui relie $\mathfrak{C}$ à un cône inclus dans l'ensemble des opérateurs de courbure strictement positifs.

3.2.2. *Deuxième étape.* — En suivant le même schéma et le même type de preuves, on construit maintenant la famille suivante. Pour $b \in [0, 1/2]$ posons

$$a(b) = \frac{(n-2)b^2 + 2b}{2 + 2(n-2)b^2} \quad \text{et} \quad p(b) = \frac{(n-2)b^2}{1 + (n-2)b^2}\,.$$

Alors,

$$\mathfrak{C}'_b = \ell_{a(b),b}\Big(\big\{R \in \mathscr{P}^2_B(\mathfrak{so}(n))|\ R \geq 0, \mathrm{Ric} \geq p(b)\frac{\mathrm{trace}(\mathrm{Ric})}{n}\big\}\Big)$$

est un cône convexe fermé $O(n)$-invariant qui est de plus invariant par l'E.D.O. ([**6**], lemme 3.4). Il relie le cône des opérateurs de courbure positifs ou nuls à $\mathfrak{C}'_{\frac{1}{2},\frac{1}{2}}$.

3.2.3. *Troisième étape.* — Enfin, pour b fixé égal à $1/2$ et $s \geq 0$, on pose

$$a(s) = \frac{1+s}{2} \quad \text{et} \quad p(s) = 1 - \frac{4}{n+2+4s}\,.$$

On définit alors

$$\mathfrak{C}''_s = \ell_{a(s),\frac{1}{2}}\Big(\Big\{R \in \mathscr{P}^2_B(\mathfrak{so}(n))|\ R \geq 0, \mathrm{Ric} \geq p(s)\frac{\mathrm{trace}(\mathrm{Ric})}{n}\Big\}\Big)$$

qui possède les mêmes propriétés que précédemment. Notons que

$$\lim_{s \to +\infty} \frac{1}{a(s)} \ell_{a(s),\frac{1}{2}}(R) = 2(n-1)R_I\,,$$

en conséquence la famille $\mathfrak{C}''_s$ converge vers $\mathbf{R}^+ I$ lorsque s tend vers l'infini. Ces deux dernières étapes consistent à inclure dans la construction des ensembles convexes invariants un pincement de la courbure de Ricci. Les inégalités qui définissent ces familles fournissent en effet un contrôle de la courbure de Ricci si l'opérateur de courbure est normalisé pour que sa plus grande valeur propre soit 1, par exemple.

Il suffit alors de concaténer les trois familles suivantes : $\mathfrak{C}_b$, $\mathfrak{C}'_b \cap \mathfrak{C}_{\bar{b}}$ et $\mathfrak{C}''_s \cap \mathfrak{C}_{\bar{b}}$ pour obtenir, après changement de paramètre, une collection de cônes convexes fermés

$O(n)$-invariants et invariants par l'E.D.O. qui relie les opérateurs de courbure 2-positifs ou nuls aux multiples de l'identité, que nous noterons $C(s)_{s\in[0,1]}$. Une telle famille est appelée dans **[6]** « pinching family ».

3.3. Ensemble de pincement

Cette construction n'est pas vraiment suffisante pour la conclusion souhaitée, c'est-à-dire que les solutions de l'E.D.P. convergent après renormalisation vers l'opérateur de courbure de la sphère. En effet, il se pourrait que l'opérateur de courbure tende vers l'infini dans ces cônes sans qu'il se rapproche d'un multiple de l'identité. On peut dire que les cônes sont trop ouverts à l'infini. Alors on construit à partir de $C(s)$ un ensemble F que nous appellerons ensemble de pincement. C'est une notion qui a été introduite par R. Hamilton dans **[23]** et Ch. Böhm et B. Wilking en donnent une généralisation.

THÉORÈME 3.4 (Ch. Böhm-B. Wilking **[6]**). — *Pour $\epsilon \in]0,1[$ et $h_0 > 0$, il existe un ensemble convexe fermé $O(n)$-invariant $F \subset \mathcal{S}^2_B(\mathfrak{so}(n))$ tel que*

(i) *F soit préservé par l'E.D.O.,*
(ii) *$C(\epsilon) \cap \{R : \mathrm{trace}(R) \leq h_0\} \subset F \subset C(\epsilon)$,*
(iii) *l'adhérence de $F \setminus C(s)$ soit compacte pour tout $s \in [\epsilon, 1[$.*

Ce théorème se prouve en montrant que l'intersection de tous les ensembles convexes fermés $O(n)$-invariants qui vérifient (i) et (ii) convient. Dans **[6]**, on trouve un théorème plus général et qui peut s'appliquer à toute une variété de familles de cônes. Une façon intuitive de comprendre F est de penser à une parabole qui coupe tous les cônes centrés à l'origine ; le cône à l'infini d'un tel objet est réduit à l'axe vertical. Aller à l'infini dans cet ensemble force à se rapprocher de cet axe après normalisation.

On peut transporter cet ensemble sur le fibré des opérateurs de courbure sur M comme nous l'avons indiqué ci-dessus pour construire un ensemble $\mathcal{F}$ défini et invariant par le transport parallèle.

3.4. Conclusion

Pour conclure la preuve de 0.3, on procède comme suit. Soit (M, g_0) une variété riemannienne compacte d'opérateur de courbure 2-positif (strictement) et soit $g(t)$ le flot de Ricci de donnée initiale g_0 défini sur un intervalle maximal $t \in [0, T[$. Par compacité de M, il existe $\epsilon \in]0,1[$ et $h_0 > 0$ tels que l'opérateur de courbure de g_0 soit en tout point $m \in M$ dans $C(\epsilon) \cap \{R : \mathrm{trace}(R) \leq h_0\}$, après identification de T_mM à $\mathbf{R}^n$. Il existe alors un ensemble de pincement F vérifiant les propriétés ci-dessus. Soit $\mathcal{F}$ le sous-ensemble du fibré des opérateurs de courbure obtenu. Le principe du

maximum affirme que le tenseur de courbure de la métrique $g(t)$ est dans $\mathcal{G}$ pour tout $t \geq 0$.

Il est aisé de vérifier que la courbure scalaire d'une métrique d'opérateur de courbure 2-positif est elle-même strictement positive. C'est alors un fait classique [**17**] que le flot de Ricci a une singularité en temps fini, c'est-à-dire que $T < +\infty$ et qu'il existe (m_i, t_i), avec $t_i \to T$ tel que

$$Q_i = |R(m_i, t_i)| = \max_{M \times [0, t_i]} |R(m, t)| \longrightarrow +\infty\,.$$

Ici nous avons noté $R(m,t)$ l'opérateur de courbure de $g(t)$ au point m, la norme est celle des endomorphismes. On construit alors une famille de flot de Ricci sur M, $g_i(t) = Q_i g(t_i + Q_i^{-1} t)$. Cette opération s'appelle une renormalisation parabolique (voir [**5**]) et produit une autre solution de l'équation du flot de Ricci. Un théorème de compacité dû à R. Hamilton ([**24**]) permet de montrer que la suite de flots $(M, g_i(t), m_i)$ converge pour la topologie de la convergence pointée vers un flot de Ricci complet $(M_\infty, g_\infty(t), m_\infty)$. On peut utiliser ici un résultat de G. Perelman (voir [**35**]) qui affirme que le rayon d'injectivité en m_i de la métrique $g_i(0)$ est uniformément minoré. Par construction $|R_\infty(m_\infty, 0)| = 1$, où R_∞ désigne l'opérateur de courbure de l'espace limite. Comme $F \backslash C(s)$ est compact donc borné pour tout $s \in [\epsilon, 1[$ alors, pour tout s supérieur à ϵ, il existe i assez grand tel que $R(m_i, t_i) \in C(s)$. Ceci implique que $Q_i^{-1} R(m_i, t_i) \in C(s)$. Par passage à la limite, on a donc $R_\infty(m_\infty, 0) \in \cap_{s \in [\epsilon, 1[} C(s)$, c'est-à-dire $R_\infty(m_\infty, 0) \in \mathbf{R}_+ I$. La courbure sectionnelle de $g_\infty(0)$ est constante positive (ne dépend pas du 2-plan) en m_∞. Par continuité, pour $p \in M_\infty$ proche de p_∞, on a $|R_\infty(p, 0)| > 1/2$ et donc, pour une suite de points $p_i \in M$ qui converge vers p, $|R(p_i, t_i)| \longrightarrow +\infty$. Le même argument s'applique donc en p montrant que la courbure sectionnelle en p de $g_\infty(0)$ est constante positive. De proche en proche, on montre qu'en chaque point de M_∞ la courbure sectionnelle est constante et positive (la constante pouvant dépendre du point) ; on en déduit que M_∞ est compacte et la convergence n'est plus pointée mais globale et donc M_∞ est difféomorphe à M. En dimension supérieure ou égale à 3 la constance en chaque point de la courbure sectionnelle implique la constance globale. On peut raffiner et également montrer que la convergence est exponentielle dans toutes les topologies C^k (voir [**28**], [**30**] et [**34**]). La preuve nécessiterait plus de détails qui sont laissés au lecteur (voir [**6**] et [**16**], p. 130).

4. LES TRAVAUX DE BRENDLE ET SCHOEN

Le schéma de preuve qui précède est suffisamment souple pour s'appliquer à d'autres situations. C'est cette approche qui est utilisée dans [**10, 11**]. La condition

de courbure utilisée précédemment, la 2-positivité du tenseur de courbure est remplacée par la notion de courbure isotrope positive. Par commodité, nous utiliserons l'acronyme *PIC* (positive isotropic curvature) pour la désigner. Elle est apparue pour la première fois dans l'article de M. Micallef et J. Moore ([**31**]). Rappelons sa définition.

Soit (M, g) une variété riemannienne dont l'endomorphisme de courbure en m est noté R. La métrique riemanienne peut être étendue à $T_mM \otimes \mathbf{C}$ comme une forme bilinéaire complexe. De même R est étendue en une application linéaire complexe. Pour un 2-plan $P \subset T_mM \otimes \mathbf{C}$, de base $\{z, w\}$, on définit sa courbure sectionnelle complexe

$$K(P) = \frac{\langle R(z \wedge w), \bar{z} \wedge \bar{w} \rangle}{\langle z \wedge w, \bar{z} \wedge \bar{w} \rangle}\,.$$

Un sous-espace $V \subset T_mM \otimes \mathbf{C}$ est dit *totalement isotrope* si tout vecteur $z \in V$ est isotrope, c'est-à-dire si $\langle z, z \rangle = 0$.

DÉFINITION 4.1. — *On dit que M est de courbure isotrope strictement positive si $K(P) > 0$ pour tout 2-plan $P \subset T_mM \otimes \boldsymbol{C}$ totalement isotrope.*

Cette condition ne peut être non vide que pour $n \geq 4$. C'est une notion qui apparaît lors de l'étude de la stabilité des surfaces minimales dans une variété riemannienne. C'est justement en étudiant la stabilité de sphères de dimension 2 harmoniques et conformes dans M que M. Micallef et J. Moore montrent qu'une variété compacte simplement connexe, de dimension supérieure ou égale à 4 et dont la courbure sectionnelle isotrope est positive (PIC), est homéomorphe à une sphère. Ce résultat permet de donner une preuve différente du théorème de pincement de M. Berger et W. Klingenberg grâce à la remarque suivante :

PROPOSITION 4.2. — *Une variété riemannienne strictement $1/4$-pincée est PIC.*

Pour un 2-plan P totalement isotrope on peut trouver quatre vecteurs orthonormés $\{e_1, e_2, e_3, e_4\}$ de sorte que $\{e_1 + ie_2, e_3 + ie_4\}$ soit une base de P. Un calcul immédiat donne alors que la condition PIC peut s'écrire

$$R_{1313} + R_{1414} + R_{2323} + R_{2424} - 2R_{1234} > 0\,.$$

Maintenant si la variété est strictement $1/4$-pincée, par exemple si sa courbure sectionnelle varie dans l'intervalle $]1/4, 1]$, tous les termes R_{ijij}, qui sont des courbures sectionnelles, sont supérieurs à $1/4$. Une inégalité due à M. Berger (voir [**2**], p. 69) montre que $|R_{1234}| < 1/2$, ce qui prouve la proposition ci-dessus.

Remarquons que l'énoncé de M. Micallef et J. Moore impose à la variété d'être simplement connexe. Cette restriction est nécessaire car la métrique canonique sur $S^{n-1} \times S^1$ est PIC, on ne peut même pas espérer prouver que la condition PIC

implique que le revêtement universel est compact. La question de décrire le groupe fondamental des variétés PIC est intéressante et on trouve dans [**11**] des références utiles. Dans [**26**], R. Hamilton étudie l'évolution du flot de Ricci dont la donnée initiale est PIC, en dimension 4 et montre que des singularités doivent apparaître. Il introduit alors l'idée de la chirurgie utilisée par la suite dans les travaux de G. Perelman ; comme dans ceux-ci les singularités ont des voisinages qui peuvent ressembler à des ouverts de $S^3 \times S^1$. Bien que la méthode décrite dans [**26**] contienne une erreur, cet article est d'une très grande importance (voir [**13**]). Pour notre propos, il montre que la méthode de Böhm et Wilking ne peut pas fonctionner telle qu'elle est décrite ; il faut tenir compte d'éventuelles singularités et donc envisager une chirurgie métrique comparable à celle qui est utilisée dans les travaux de Perelman. C'est une question largement ouverte.

Dans [**11**], c'est une idée brillante qui permet de contourner ce problème. Soit (M, g_0) une variété riemannienne compacte ; notons $\tilde{R}$ le tenseur de courbure de $M \times \mathbf{R}$ et $\hat{R}$ celui de $M \times \mathbf{R}^2$. S. Brendle et R. Schoen montrent les propriétés suivantes :

1) La condition « R est PIC » est préservée par le flot.
2) La condition « $\tilde{R}$ est PIC » est également préservée par le flot et est plus restrictive que la précédente. En effet, elle implique, par exemple, que la courbure de Ricci de R est strictement positive.
3) L'opérateur de courbure $\hat{R}$ ne peut pas être PIC à cause du facteur plat de dimension 2, comme on peut aisément le vérifier. En revanche, pour $\epsilon > 0$, considérons l'opérateur $R_\epsilon = R - \epsilon I$ et son extension $\hat{R}_\epsilon$ sur $M \times \mathbf{R}^2$. Dire que $\hat{R}_\epsilon$ a une courbure isotrope positive ou nulle (nous dirons faiblement PIC) est une restriction supplémentaire. Cela implique, par exemple, que R est PIC et que sa courbure sectionnelle est strictement positive.

Le schéma de la preuve de Ch. Böhm et B. Wilking peut alors s'appliquer avec ces notions plus restrictives. Plus précisément, on a

THÉORÈME 4.3 (S. Brendle-R. Schoen [**11**], theorem 2). — *Soit (M, g_0) une variété riemannienne compacte telle que, pour ϵ assez petit, $\hat{R}_\epsilon$ soit faiblement PIC. Alors, le flot de Ricci normalisé de donnée initiale g_0 converge vers une métrique de courbure constante (dans la topologie C^∞).*

Le théorème principal 0.1 est alors une conséquence de

PROPOSITION 4.4 ([**11**]). — *Soit (M, g_0) une variété riemannienne compacte de courbure sectionnelle strictement positive et strictement $1/4$-pincée ; alors, pour ϵ assez petit, $\hat{R}_\epsilon$ est faiblement PIC. Ici R est l'opérateur de courbure de (M, g_0).*

Idées de la preuve. — Les preuves suivent maintenant le schéma de Böhm et Wilking. La première étape est la plus difficile et consiste à montrer que la condition faiblement PIC est préservée par le flot. Comme précédemment, en utilisant les notations ci-dessus, cela se ramène à prouver que, si $R_{1313} + R_{1414} + R_{2323} + R_{2424} - 2R_{1234} = 0$, alors

$$Q(R)_{1313} + Q(R)_{1414} + Q(R)_{2323} + Q(R)_{2424} - 2Q(R)_{1234} \geq 0\,.$$

Ce calcul fastidieux est fait avec beaucoup de clarté et plusieurs astuces dans [**11**]. Signalons que ce résultat a été obtenu indépendamment par un jeune mathématicien australien, Huy Nguyen (Australian National University), dans sa thèse ([**33**]).

La seconde étape consiste à construire la famille de cônes convexes qui se contracte sur $\mathbf{R}_+ I$. Le point de départ est

$$\hat{C} = \left\{R :\ \hat{R} \text{ faiblement } PIC\right\}.$$

La construction utilise comme précédemment les applications $\ell_{a,b}$ pour des choix convenables de a et b. Enfin on montre la proposition par des calculs du type de ceux utilisés pour la preuve de 4.2 bien que sensiblement plus sophistiqués. □

Dans un travail récent, S. Brendle développe avec succès le même formalisme pour une métrique riemannienne telle que $(M, g_0) \times \mathbf{R}$ ait un opérateur de courbure PIC (c'est-à-dire pour $\tilde{R}$ avec les notations ci-dessus) (voir [**8**]).

5. QUELQUES EXTENSIONS

Dans les travaux de M. Berger et W. Klingenberg, le pincement non strict est abordé, c'est-à-dire : une variété simplement connexe 1/4-pincée (de courbure sectionnelle positive) qui n'est pas homéomorphe à une sphère est difféomorphe à un espace symétrique compact de rang un. L'analogue où l'on obtient un difféomorphisme avec la sphère est prouvé par S. Brendle et R. Schoen dans [**10**].

Dans [**3**], M. Berger montre que, pour chaque dimension n, il existe un nombre $\varepsilon(n)$ avec $0 < \varepsilon(n) < \frac{1}{4}$ tel que toute variété complète, simplement connexe M de dimension n qui admet une métrique $\varepsilon(n)$-pincée soit homéomorphe à S^n ou difféomorphe à un espace riemannien symétrique compact de rang un. L'analogue dans l'esprit des travaux ci-dessus est prouvé par P. Petersen et T. Tao dans [**38**].

Pour plus de résultats sur les conséquences géométriques et topologiques de la condition PIC, on peut consulter les références récentes suivantes : [**7**], [**32**], [**18**], [**19**], [**41**], [**20**] (et beaucoup d'autres certainement).

Dans ce contexte, la nouvelle frontière est certainement d'apprendre à pratiquer des chirurgies sur des flots de Ricci de dimension supérieure ou égale à 4.

6. CONCLUSION

La quasi-coïncidence entre les travaux brièvement décrits ci-dessus et ceux de G. Perelman montre de manière claire que le flot de Ricci arrive à pleine maturité. On constate également des progrès sensibles dans la compréhension du tenseur de courbure d'une variété riemannienne ; en effet, il est remarquable de découvrir de nouvelles identités générales comme c'est le cas dans le travail de Böhm et Wilking. On peut espérer que l'approfondissement des recherches dans ces directions produise à l'avenir de nouveaux résultats profonds.

Enfin, il faut insister sur l'omniprésence des travaux de R. Hamilton qui en plus de construire l'édifice a ouvert pratiquement toutes les pistes.

RÉFÉRENCES

[1] M. Berger – Les variétés Riemanniennes (1/4)-pincées, *Ann. Scuola Norm. Sup. Pisa* **14** (1960), p. 161–170.

[2] ———, Sur quelques variétés riemanniennes suffisamment pincées, *Bull. Soc. Math. France* **88** (1960), p. 57–71.

[3] ———, Sur les variétés riemanniennes pincées juste au-dessous de 1/4, *Ann. Inst. Fourier (Grenoble)* **33** (1983), p. 135–150.

[4] ———, *A panoramic view of Riemannian geometry*, Springer, 2003.

[5] G. Besson – Preuve de la conjecture de Poincaré en déformant la métrique par la courbure de Ricci (d'après G. Perel′man), Séminaire Bourbaki, vol. 2004/2005, exposé n° 947, *Astérisque* **307** (2006), p. 309–347.

[6] C. Böhm & B. Wilking – Manifolds with positive curvature operators are space forms, *Ann. of Math.* **167** (2008), p. 1079–1097.

[7] S. Brendle – Einstein manifolds with nonnegative isotropic curvature are locally symmetric, prépublication arXiv:0812.0335, 2008.

[8] ———, A general convergence result for the Ricci flow in higher dimensions, *Duke Math. J.* **145** (2008), p. 585–601.

[9] ———, *Ricci flow and the sphere theorem*, Graduate Studies in Math., vol. 111, Amer. Math. Soc., 2010.

[10] S. Brendle & R. M. Schoen – Classification of manifolds with weakly 1/4-pinched curvatures, *Acta Math.* **200** (2008), p. 1–13.

[11] ———, Manifolds with 1/4-pinched curvature are space forms, *J. Amer. Math. Soc.* **22** (2009), p. 287–307.

[12] ———, Sphere theorems in geometry, in *Surveys in differential geometry. Vol. XIII. Geometry, analysis, and algebraic geometry : forty years of the Journal of Differential Geometry*, Surv. Differ. Geom., vol. 13, Int. Press, Somerville, MA, 2009, p. 49–84.

[13] B.-L. Chen & X.-P. Zhu – Ricci flow with surgery on four-manifolds with positive isotropic curvature, *J. Differential Geom.* **74** (2006), p. 177–264.

[14] H. Chen – Pointwise $\frac{1}{4}$-pinched 4-manifolds, *Ann. Global Anal. Geom.* **9** (1991), p. 161–176.

[15] B. Chow *et al.* – *The Ricci flow : techniques and applications. Part I*, Math. Surveys and Monographs, vol. 135, Amer. Math. Soc., 2007.

[16] ———, *The Ricci flow : techniques and applications. Part II*, Math. Surveys and Monographs, vol. 144, Amer. Math. Soc., 2008.

[17] B. Chow & D. Knopf – *The Ricci flow : an introduction*, Math. Surveys and Monographs, vol. 110, Amer. Math. Soc., 2004.

[18] A. M. Fraser – Fundamental groups of manifolds with positive isotropic curvature, *Ann. of Math.* **158** (2003), p. 345–354.

[19] A. M. Fraser & J. Wolfson – The fundamental group of manifolds of positive isotropic curvature and surface groups, *Duke Math. J.* **133** (2006), p. 325–334.

[20] S. Gadgil & H. Seshadri – On the topology of manifolds with positive isotropic curvature, *Proc. Amer. Math. Soc.* **137** (2009), p. 1807–1811.

[21] S. Gallot, D. Hulin & J. Lafontaine – *Riemannian geometry*, 3e éd., Universitext, Springer, 2004.

[22] R. S. Hamilton – Three-manifolds with positive Ricci curvature, *J. Differential Geom.* **17** (1982), p. 255–306.

[23] ———, Four-manifolds with positive curvature operator, *J. Differential Geom.* **24** (1986), p. 153–179.

[24] ———, A compactness property for solutions of the Ricci flow, *Amer. J. Math.* **117** (1995), p. 545–572.

[25] ———, The formation of singularities in the Ricci flow, in *Surveys in differential geometry, Vol. II (Cambridge, MA, 1993)*, Int. Press, Cambridge, MA, 1995, p. 7–136.

[26] ———, Four-manifolds with positive isotropic curvature, *Comm. Anal. Geom.* **5** (1997), p. 1–92.

[27] L. Hernández – Kähler manifolds and 1/4-pinching, *Duke Math. J.* **62** (1991), p. 601–611.

[28] G. Huisken – Ricci deformation of the metric on a Riemannian manifold, *J. Differential Geom.* **21** (1985), p. 47–62.

[29] W. Klingenberg – Über Riemannsche Mannigfaltigkeiten mit positiver Krümmung, *Comment. Math. Helv.* **35** (1961), p. 47–54.

[30] C. Margerin – A sharp characterization of the smooth 4-sphere in curvature terms, *Comm. Anal. Geom.* **6** (1998), p. 21–65.

[31] M. J. Micallef & J. D. Moore – Minimal two-spheres and the topology of manifolds with positive curvature on totally isotropic two-planes, *Ann. of Math.* **127** (1988), p. 199–227.

[32] M. J. Micallef & M. Y. Wang – Metrics with nonnegative isotropic curvature, *Duke Math. J.* **72** (1993), p. 649–672.

[33] H. Nguyen – Isotropic curvature and the Ricci flow, prépublication, 2007.

[34] S. Nishikawa – Deformation of Riemannian metrics and manifolds with bounded curvature ratios, in *Geometric measure theory and the calculus of variations (Arcata, Calif., 1984)*, Proc. Sympos. Pure Math., vol. 44, Amer. Math. Soc., 1986, p. 343–352.

[35] G. Perelman – The entropy formula for the Ricci flow and its geometric applications, prépublication arXiv:math.DG/0211159, 2002.

[36] ———, Finite extinction time for the solutions to the Ricci flow on certain three-manifolds, prépublication arXiv:math.DG/0307245, 2003.

[37] ———, Ricci flow with surgery on three-manifolds, prépublication arXiv:math.DG/0303109, 2003.

[38] P. Petersen & T. Tao – Classification of almost quarter-pinched manifolds, *Proc. Amer. Math. Soc.* **137** (2009), p. 2437–2440.

[39] P. Petersen & F. Wilhelm – An exotic sphere with positive sectional curvature, prépublication arXiv:0805.0812, 2008.

[40] H. E. Rauch – A contribution to differential geometry in the large, *Ann. of Math.* **54** (1951), p. 38–55.

[41] H. Seshadri – Manifolds with nonnegative isotropic curvature, prépublication arXiv:0707.3894, 2008.

[42] A. Besse – *Einstein manifolds*, Ergebnisse Math. Grenzg. (3), vol. 10, Springer, 1987.

[43] S.-T. Yau & F. Zheng – Negatively $\frac{1}{4}$-pinched Riemannian metric on a compact Kähler manifold, *Invent. Math.* **103** (1991), p. 527–535.

Gérard BESSON

Université Joseph Fourier (Grenoble I)
Institut Fourier
B.P. 74
38402 SAINT MARTIN D'HÈRES Cedex
E-mail : G.Besson@ujf-grenoble.fr

332

ASTÉRISQUE

2010

SÉMINAIRE BOURBAKI
VOLUME 2008/2009
EXPOSÉS 997-1011

(1004) *Sur la géométrie et la dynamique des transformations de contact*

Emmanuel GIROUX

SOCIÉTÉ MATHÉMATIQUE DE FRANCE
Publié avec le concours du CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

Séminaire BOURBAKI Mars 2009
61e année, 2008-2009, no 1004, p. 183 à 220

SUR LA GÉOMÉTRIE ET LA DYNAMIQUE DES TRANSFORMATIONS DE CONTACT [d'après Y. Eliashberg, L. Polterovich *et al.*]

par **Emmanuel GIROUX**

Au commencement étaient les transformations de contact...

Étant donné une variété M, les 1-jets des germes de fonctions $M \to \mathbf{R}$ forment une variété J^1M naturellement fibrée au-dessus de M. Parmi toutes les sections locales de ce fibré, les 1-jets des fonctions f définies sur un ouvert de M sont nommées sections holonomes et constituent une classe élue. Si $M = \mathbf{R}^n$ par exemple, $J^1M = \mathbf{R}^n \times \mathbf{R}^n \times \mathbf{R}$ fibre par projection sur le premier facteur et une section

$$x \in \mathbf{R}^n \longmapsto \big(x, \varphi(x), f(x)\big) \in \mathbf{R}^n \times \mathbf{R}^n \times \mathbf{R}$$

est holonome si et seulement si $\varphi = df$. Cette condition traduit en fait une propriété de contact, à savoir que la 1-forme $dt - \varphi$ sur $M \times \mathbf{R}$, $t \in \mathbf{R}$, s'annule identiquement sur l'espace tangent au graphe de f.

La condition d'holonomie est ainsi l'origine de la géométrie de contact. Lorsque S. Lie introduit les transformations de contact (Berührungstransformationen), c'est pour étudier les symétries des équations différentielles. Or une telle équation (en une variable) ou, plus généralement, une équation aux dérivées partielles du premier ordre sur les fonctions $M \to \mathbf{R}$ s'écrit $\Phi(x, df(x), f(x)) = 0$ et est donc donnée par une hypersurface $\{\Phi = 0\}$ dans l'espace J^1M. Une symétrie de l'équation est une transformation de J^1M qui préserve à la fois l'hypersurface $\{\Phi = 0\}$ et la condition d'holonomie, de manière à envoyer solution sur solution.

Une forme de contact sur J^1M est une 1-forme différentielle non singulière dont le rappel par toute section holonome locale est nul. De fait, il en existe une et une seule à multiplication près par une fonction. Son expression locale, c'est-à-dire pour $M = \mathbf{R}^n$, est

$$u(x, y, t)\Big(dt - \sum_1^n y_i dx_i\Big), \qquad (x, y, t) \in \mathbf{R}^n \times \mathbf{R}^n \times \mathbf{R}.$$

Le noyau des formes de contact est donc un champ d'hyperplans ξ bien défini qui admet pour sous-variétés intégrales tous les graphes des sections holonomes locales. Ainsi, une transformation de J^1M préserve la condition d'holonomie si et seulement si elle préserve ce champ ξ des hyperplans dits de contact. Un exemple emblématique est, sur $J^1\mathbf{R}^n$, la transformation de Legendre

$$(x, y, t) \in J^1\mathbf{R}^n = \mathbf{R}^n \times \mathbf{R}^n \times \mathbf{R} \longmapsto \Big(-y, x, t - \sum_1^n x_j y_j\Big).$$

Ce n'est apparemment que vers 1950, sans doute à l'instigation de S. S. Chern, que le champ d'hyperplans ξ reçoit le nom de structure de contact (étant la structure associée au pseudo-groupe des transformations de contact) et prend peu à peu le pas sur les transformations de contact pour devenir l'objet central de la géométrie de contact.

1. INTRODUCTION

Le travail présenté ici est dû à Yakov Eliashberg et Leonid Polterovich, ainsi qu'à Sang Seon Kim, Mohan Bhupal et Sheila Sandon. Les références en sont **[2, 12, 13, 24]**, le dernier de ces articles redémontrant les résultats principaux du second par une méthode différente. L'intérêt majeur de ce travail est, à mes yeux, de mettre en lumière les rapports de force subtils qui existent, en géométrie de contact, entre flexibilité et rigidité.

Bien que la géométrie de contact soit de nos jours clairement identifiée comme une partie de la géométrie symplectique, elle a des particularités très fortes. On connaît par exemple peu d'invariants qui prennent un continuum de valeurs et ils sont de nature dynamique plutôt que géométrique. La raison en est que les variétés symplectiques attachées aux variétés de contact ont un volume et des capacités infinis. Ce caractère discret des invariants de contact apparaît nettement dans la suite, se traduisant en particulier par de brusques changements dans le comportement géométrique des transformations de contact.

A. Géométrie des transformations de contact

Dans $\mathbf{R}^{2n} = T^*\mathbf{R}^n = \mathbf{C}^n$ muni de sa structure symplectique standard, on considère la boule et le cylindre ouverts d'aire r, à savoir respectivement

$$B(r) = B^{2n}(r) = \{z \in \mathbf{C}^n \mid \pi|z|^2 < r\}$$
$$\text{et} \quad C(r) = C^{2n}(r) = \{z \in \mathbf{C}^n \mid \pi|z_1|^2 < r\}.$$

Pour une simple raison de volume, $B(r_1)$ ne se plonge symplectiquement dans $B(r_2)$ que si $r_1 \leq r_2$. Mieux, le « Non-Squeezing Theorem » de M. Gromov, un des nombreux résultats de [**16**] révélateurs de la rigidité en géométrie symplectique, montre que $B(r_1)$ ne se plonge symplectiquement dans $C(r_2)$ que si $r_1 \leq r_2$.

Ces contraintes n'existent pas dans $\mathbf{R}^{2n+1} = J^1\mathbf{R}^n$ muni de sa structure de contact standard. Pour $r > 0$ petit par exemple, la transformation de contact $(x, y, t) \in \mathbf{R}^{2n+1} \mapsto (\sqrt{r}x, \sqrt{r}y, rt)$ envoie tout compact donné dans une boule arbitrairement petite et une isotopie de contact à support compact peut faire le même travail. Cependant, comme Y. Eliashberg, S. S. Kim et L. Polterovich le montrent dans [**12**], la situation est plus subtile dans la variété $J^1(\mathbf{R}^n, \mathbf{T})$ des 1-jets (des germes) de fonctions à valeurs dans le cercle $\mathbf{T} = \mathbf{R}/\mathbf{Z}$: on ne peut tasser n'importe quel domaine dans n'importe quel autre par des transformations de contact. Pour énoncer le résultat principal, on précise la terminologie.

DÉFINITION 1.1 ([**12**]). — *Étant donné des domaines (ouverts connexes) U_1 et U_2 dans une variété de contact (V, ξ), on dit qu'on peut* tasser U_1 dans U_2 *si on peut trouver une isotopie de plongements de contact $\psi_s \colon \operatorname{Adh} U_1 \to V$, $s \in [0, 1]$, entre l'inclusion ψ_0 et un plongement ψ_1 qui envoie $\operatorname{Adh} U_1$ dans U_2. On dit que le tassement se fait au sein d'un domaine $U_0 \subset V$ si $\psi_s(\operatorname{Adh} U_1) \subset U_0$ pour tout $s \in [0, 1]$. Noter que, si U_1 est relativement compact, toute isotopie qui tasse U_1 dans U_2 au sein de U_0 se prolonge en une isotopie de contact ambiante à support compact dans U_0.*

Pour tout domaine U de $\mathbf{R}^{2n}$, on notera $\widehat{U}$ le produit $U \times \mathbf{T} \subset \mathbf{R}^{2n} \times \mathbf{T} = J^1(\mathbf{R}^n, \mathbf{T})$. Le théorème principal qu'on veut démontrer ici est le suivant :

THÉORÈME 1.2 ([**12**]). — (a) *En dimension $2n + 1 > 3$, n'importe quel domaine qui est relativement compact dans $\widehat{C}(1)$ peut être tassé dans $\widehat{B}(d)$ pour tout $d > 0$.*

(b) *Si $r_2 \leq k \leq r_1$ pour un certain entier k, on ne peut pas plonger $\operatorname{Adh} \widehat{B}(r_1)$ dans $\widehat{C}(r_2)$ par une transformation de contact ambiante de $\mathbf{R}^{2n} \times \mathbf{T}$ à support compact et on ne peut donc pas tasser $\widehat{B}(r_1)$ dans $\widehat{C}(r_2)$.*

En dimension $2n + 1 = 3$, si $r_2 < r_1$ (et même si $r_1 < 1$), les résultats de [**9**] (voir aussi [**15**]) montrent qu'on ne peut pas tasser $\widehat{B}(r_1)$ dans $\widehat{B}(r_2)$. En dimension supérieure, par contre, on ne sait pas s'il est possible de tasser $\widehat{B}(r_1)$ dans $\widehat{B}(r_2)$ ou même dans $\widehat{C}(r_2)$ quand $k - 1 < r_2 < r_1 < k$ pour un certain entier $k > 1$.

On peut reformuler en partie l'énoncé précédent en regardant, pour tout domaine borné $U \subset \mathbf{R}^{2n}$, les quantités suivantes :

- $\mathrm{w}_-(U)$, le supremum des nombres r pour lesquels $B(r)$ se plonge symplectiquement dans U ;
- $\mathrm{w}_+(U)$, l'infimum des nombres r tels que U se plonge symplectiquement dans $C(r)$;

– $\widehat{\mathrm{w}}(U)$, l'infimum des nombres r tels que le domaine $r^{-1/2}U \times \mathbf{T} \subset \mathbf{R}^{2n} \times \mathbf{T}$ puisse être tassé dans $\widehat{B}(r')$ quel que soit $r' > 0$.

L'encadrement ci-dessous traduit fidèlement le point (b) du théorème 1.2 et l'impossibilité de tasser $\widehat{B}(1)$ dans $\widehat{B}(r)$ pour r arbitrairement petit :

COROLLAIRE 1.3 ([**12**]). — *Pour tout domaine borné $U \subset \mathbf{R}^{2n}$, $n \geq 2$,*

$$\mathrm{w}_-(U) \leq \widehat{\mathrm{w}}(U) \leq \mathrm{w}_+(U).$$

Un autre phénomène intéressant mis à jour par Eliashberg, Kim et Polterovich est que le tassement, quand il est possible, nécessite de la place. Par exemple, si $r_2 \leq 1/m \leq r_1 < 1$ pour un certain entier m, on peut tasser $\widehat{B}(r_1)$ dans $\widehat{B}(r_2)$ (théorème 1.2-(b)) mais pas au sein de $\widehat{B}(1/(m-1))$. De façon plus générale :

THÉORÈME 1.4 ([**12**]). — *Si $r_2 \leq k/m \leq r_1 < k/(m-1)$ pour des entiers k et m, on ne peut pas plonger* $\mathrm{Adh}\,\widehat{B}(r_1)$ *dans $\widehat{B}(r_2)$ par une transformation de contact de $\mathbf{R}^{2n} \times \mathbf{T}$ qui préserve $\widehat{B}(k/(m-1))$. En particulier, on ne peut pas tasser $\widehat{B}(r_1)$ dans $\widehat{B}(r_2)$ au sein de $\widehat{B}(k/(m-1))$.*

Dans [**12**], pour démontrer les théorèmes 1.2-(b) et 1.4, Eliashberg, Kim et Polterovich élaborent une homologie de Floer généralisée pour les structures hamiltoniennes stables. Cette théorie leur permet d'interpréter l'homologie symplectique convenablement filtrée d'un ouvert étoilé générique $U \subset \mathbf{R}^{2n}$ [**7, 14, 33**] comme un invariant de contact du domaine $\widehat{U} = U \times \mathbf{T} \subset \mathbf{R}^{2n} \times \mathbf{T}$, invariant du type de ceux issus de la théorie symplectique des champs [**11**].

Dans un article tout récent [**24**], S. Sandon a présenté une approche plus spécifique mais aussi efficace en pratique et bien plus économique. La méthode, fondée sur la théorie des fonctions génératrices, produit également des invariants homologiques ainsi qu'un invariant de type capacité. Crise oblige, c'est cette approche qu'on suivra ici.

B. Ordonnancement des transformations de contact

Le point de départ du travail d'Eliashberg et Polterovich [**12, 13**] est la recherche d'une géométrie sur le groupe des transformations de contact et l'étude d'une relation de préordre naturelle sur le revêtement universel de celui-ci. Une conclusion inattendue de leur travail est que le comportement de ce préordre est intimement lié aux phénomènes de tassement discutés dans la section ci-dessus.

Soit (K, η) une variété de contact, $\mathscr{G} = \mathscr{G}(K, \eta)$ la composante neutre du groupe $\mathscr{D}_c(K, \eta)$ des transformations de contact de (K, η) à support compact et $\widetilde{\mathscr{G}} = \widetilde{\mathscr{G}}(K, \eta)$ le revêtement universel de $\mathscr{G}$.

On dit qu'un chemin ϕ_t dans $\mathscr{G}(K,\eta)$ est *positif* (resp. *strictement positif*) si le vecteur $\dot{\phi}_t(q)$, à tout instant t et pour tout point q, est dans le demi-espace positif fermé (resp. ouvert) de $T_{\phi_t(q)}K$ délimité par l'hyperplan coorienté $\eta_{\phi_t(q)}$. Noter que, si ϕ_t et ψ_t, $t\in[0,1]$, sont deux chemins positifs (resp. strictement positifs), le chemin $\phi_t\circ\psi_t$ est positif (resp. strictement positif).

Étant donné $\widetilde{\phi}_0,\widetilde{\phi}_1\in\widetilde{\mathscr{G}}(K,\eta)$, on dit que $\widetilde{\phi}_1$ domine $\widetilde{\phi}_0$ et on écrit $\widetilde{\phi}_1\succeq\widetilde{\phi}_0$ si l'élément $\widetilde{\phi}_1\circ\widetilde{\phi}_0^{-1}$ se représente par un chemin positif ϕ_t entre ϕ_0 et ϕ_1 dans $\mathscr{G}(K,\eta)$. La relation $\succeq$ est clairement réflexive et transitive mais pas toujours antisymétrique. Elle définit donc en général seulement un préordre sur $\widetilde{\mathscr{G}}$. Ce préordre est par ailleurs invariant au sens où, si $\widetilde{\phi}_1\succeq\widetilde{\phi}_0$ et $\widetilde{\psi}_1\succeq\widetilde{\psi}_0$, alors $\widetilde{\phi}_1\widetilde{\psi}_1\succeq\widetilde{\phi}_0\widetilde{\psi}_0$.

DÉFINITION 1.5 ([**12**]). — *On dit qu'une variété de contact (K,η) est* de type ordonnable *ou* non-ordonnable *selon que la relation $\succeq$ est antisymétrique — et définit donc un ordre partiel sur $\widetilde{\mathscr{G}}(K,\eta)$ — ou pas.*

Le théorème suivant, dont on précise la genèse, dans la remarque 1.7 ci-après, regroupe les principaux résultats connus sur cette notion :

THÉORÈME 1.6 ([**2, 6, 8, 12, 22, 24**]). — (a) *Les espaces de jets $J^1\mathbf{R}^n$ et $J^1(\mathbf{R}^n,\mathbf{T})$ sont de type ordonnable.*

(b) *La sphère de contact standard $(\mathbf{S}^{2n-1},\eta_0)$ de dimension au moins* 3 *est de type non-ordonnable.*

(c) *L'espace des éléments de contact coorientés sur une variété close M est de type ordonnable.*

Remarque 1.7. — Ce théorème est essentiellement dû à Eliashberg, Kim et Polterovich pour les parties (b) et (c) et à M. Bhupal et S. Sandon — qui en ont traité les deux cas indépendamment — pour la partie (a). Cependant, comme on le verra plus loin, on peut déduire la partie (b) du travail de G. I. Olshanskii [**22**] bien antérieur à [**12**]. En outre, la partie (c) est démontrée dans [**12**] (via l'homologie de Floer pour les structures hamiltoniennes) sous l'hypothèse que le groupe fondamental de M est fini ou contient une infinité de classes de conjugaison. En pratique, cette hypothèse est anodine dans la mesure où on ne connaît aucun groupe de présentation finie qui ne la vérifie pas mais elle semble incongrue. Grâce à la théorie des fonctions génératrices, V. Chernov et S. Nemirovski s'en affranchissent dans [**6**] en regardant comment les isotopies de contact positives opèrent sur les sous-variétés legendriennes, notamment sur les fibres de l'espace des éléments de contact. L'idée d'étudier cette action est en fait due à V. Colin, E. Ferrand et P. Pushkar qui, dans [**8**], prouvent implicitement la partie (c) du théorème 1.6 dans le cas où M est revêtue par un ouvert de l'espace euclidien.

On démontrera ci-après les parties (a) et (b) du résultat ci-dessus mais pas la partie (c).

C. Dynamique des transformations de contact

Eliashberg et Polterovich dégagent dans [**13**] deux notions qui semblent très intéressantes pour l'étude dynamique des transformations de contact et qui permettent de formuler de multiples questions. Ces notions sont celles de *prédominance* et de *nombre de progression relatif*. Avant de les appliquer aux transformations de contact, on les définit dans le cadre abstrait des groupes ordonnés.

Soit $(G, \leq)$ un groupe ordonné, c'est-à-dire un groupe G muni d'un ordre partiel $\leq$ invariant au sens où son cône positif

$$G^+ = \{g \in G \mid g \geq 1\}$$

est stable par produit et invariant par tout automorphisme intérieur. Concrètement, si $g_0 \leq g_1$ et $h_0 \leq h_1$, alors $g_0 h_0 \leq g_1 h_1$.

DÉFINITION 1.8 ([**13**]). — *On dit qu'un élément $g \in G$ est* prédominant *si $g \geq 1$ et si, pour tout $h \in G$, il existe $p \in \mathbf{N}$ tel que $g^p \geq h$.*

L'ensemble G^{++} des éléments prédominants est une sorte d'intérieur de G^+ (voir la proposition 1.11). Un fait général [**1**] est que $G^+ G^{++} = G^{++} G^+ = G^{++}$. En effet, soit $f \in G^+$ et $g \in G^{++}$. Pour tout $h \in G$, il existe $p \in \mathbf{N}$ tel que $g^p \geq h$ et

$$\begin{aligned} g^p h^{-1} \in G^+ &\Rightarrow g h^{-1} g^{p-1} \in G^+ \Rightarrow f g h^{-1} g^{p-1} \in G^+ \\ &\Rightarrow g^{p-1} f g h^{-1} \in G^+ \Rightarrow g^{p-2} f g f g h^{-1} \in G^+ \Rightarrow (fg)^p h^{-1} \in G^+. \end{aligned}$$

Soit maintenant $g \in G^{++}$ un élément prédominant et $h \in G$ un élément quelconque. On pose

$$\gamma(h/g) = \lim_{k \to \infty} \frac{1}{k} \gamma_k(h/g)$$
$$\text{où} \quad \gamma_k(h/g) = \inf\{p \in \mathbf{Z} \mid g^p \geq h^k\}, \quad k \in \mathbf{N}.$$

Pour voir que la limite existe, on note d'abord que l'infimum $\gamma_k(h/g) = \gamma_k$ est fini. En effet, soit $q \geq 0$ un entier tel que $g^q \geq h^{-1}$. Si $g^p \geq h^k$, alors $h^{-k} \geq g^{-p}$, donc $g^{kq} \geq g^{-p}$, de sorte que $p \geq -kq$. Par suite, $\gamma_k \geq -kq > -\infty$. D'autre part, comme les inégalités $g^{\gamma_k} \geq h^k$ et $g^{\gamma_l} \geq h^l$ entraînent $g^{\gamma_k + \gamma_l} \geq h^{k+l}$, la suite γ_k est sous-additive. Un argument classique garantit alors l'existence de la limite $\gamma(h/g)$.

DÉFINITION 1.9 ([**13**]). — *Le nombre $\gamma(h/g)$ est appelé* nombre de progression de h relatif à g.

Des manipulations simples **[1, 13]** montrent que ce nombre a les propriétés suivantes :

$$\begin{aligned}
\gamma(h/g) &= \inf\{p/q \mid g^p \geq h^q,\ p \in \mathbf{Z},\ q \in \mathbf{N}\}, \qquad g \in G^{++},\ h \in G,\\
\gamma(g^m/g) &= m, \qquad g \in G^{++},\ m \geq 1,\\
\gamma(h/g) &> 0 \quad \text{et} \quad \gamma(g/h)\gamma(h/g) \geq 1, \qquad g, h \in G^{++}.
\end{aligned}$$

Exemple 1.10 (**[13]**). — Soit $\mathscr{G} = \mathscr{G}(\mathbf{T})$ le groupe des difféomorphismes croissants du cercle, la première variété de contact qui soit. Le revêtement universel $\widetilde{\mathscr{G}}$ de $\mathscr{G}$ est le groupe des difféomorphismes croissants de $\mathbf{R}$ qui commutent avec la translation $\epsilon\colon x \mapsto x+1$ et l'ordre $\preceq$ sur $\widetilde{\mathscr{G}}$ est donné par

$$\phi_0 \preceq \phi_1 \quad \text{si et seulement si} \quad \phi_0(x) \leq \phi_1(x) \quad \text{pour tout } x \in \mathbf{R}.$$

Ainsi, ϵ est représenté par le lacet strictement positif $x \mapsto x+t$, $t \in \mathbf{T}$, et est prédominant. L'affirmation qu'on laisse à vérifier au lecteur est que, si

$$\tau(\phi) = \lim_{q\to\infty} \frac{\phi^q(x) - x}{q}$$

désigne le nombre de translation de $\phi \in \widetilde{\mathscr{G}}$, alors

$$\gamma(\phi/\epsilon) = \tau(\phi) = 1/\gamma(\epsilon/\phi),$$

où la première identité vaut pour tout $\phi \in \widetilde{\mathscr{G}}$ et la seconde pour ceux qui sont prédominants.

Dans le cadre de la géométrie de contact, les notions ci-dessus s'appliquent au revêtement universel $\widetilde{\mathscr{G}}$ du groupe $\mathscr{G} = \mathscr{G}(K, \eta)$ lorsque la variété de contact (K, η) est de type ordonnable.

PROPOSITION 1.11 (**[13]**). — *Soit (K, η) une variété de contact close et de type ordonnable. Tous les éléments de $\widetilde{\mathscr{G}}$ qui se représentent par un chemin strictement positif dans $\mathscr{G}$ sont prédominants dans $(\widetilde{\mathscr{G}}, \preceq)$.*

Démonstration. — Soit $\widetilde{\phi} \in \widetilde{\mathscr{G}}$ un élément représenté par un chemin ϕ_t strictement positif. Grâce au lemme 2.4, on peut supposer que le champ de vecteurs qui engendre l'isotopie ϕ_t est 1-périodique en temps, de sorte que ϕ_t est défini pour tout $t \in \mathbf{R}$.

Soit d'autre part $\widetilde{\psi} \in \widetilde{\mathscr{G}}$ un élément quelconque représenté par un chemin ψ_t, $t \in [0, 1]$. Pour tout entier m positif assez grand, le chemin $\phi_{mt} \circ \psi_{1-t}$ est strictement positif et représente $\widetilde{\phi}^m \widetilde{\psi}^{-1}$. □

Dans toute variété de contact close et de type ordonnable, on dispose donc d'éléments prédominants par rapport auxquels tout élément de $\widetilde{\mathscr{G}}$ a un nombre de progression bien défini. Cela dit, le calcul de ce nombre semble délicat. Pour la variété des éléments de contact coorientés sur le tore, Eliashberg et Polterovich donnent un résultat [**13**, théorème 1.5.A] grâce à la notion de contour homologique, notion dont le nom vient de [**9**] et l'idée de [**27**]. Plutôt que de présenter ce résultat, on laisse Eliashberg et Polterovich décrire leur motivation :

> C'est une idée classique en théorie des systèmes dynamiques de mesurer la vitesse de rotation des trajectoires d'un flot autour d'un cycle donné dans la variété. Une telle mesure peut être réalisée rigoureusement et s'est avérée utile dans des situations variées comprenant la dynamique hamiltonienne et celle des difféomorphismes du cercle. En particulier, elle est étroitement liée aux propriétés asymptotiques de l'ensemble des périodes de certaines orbites fermées du flot. Nous adoptons un point de vue différent en considérant un flot comme une courbe dans le groupe des difféomorphismes (voir [**23**] pour quelques applications de ce point de vue dans le contexte de la géométrie de Hofer). Notre suggestion est de mesurer la rotation de cette courbe autour d'un cycle dans le groupe ! Grâce à la notion de nombre de progression relatif, nous pouvons mettre en œuvre rigoureusement cette idée pour le groupe des transformations de contact. Comme nous le verrons ci-dessous dans quelques exemples, les résultats des deux mesures (celle que nous suggérons et la mesure classique) ont une grande parenté.

On renvoie le lecteur à [**13**] pour les quelques exemples en question mais on l'invite surtout à en étudier lui-même d'autres.

2. CONSTRUCTIONS

A. Dans une variété de contact quelconque

On précise ici quelques termes et notions de base puis on montre qu'une variété de contact (K, η) est de type non-ordonnable si et seulement si le groupe $\mathscr{G}(K, \eta)$ contient des lacets strictement positifs contractiles et on explique comment utiliser ces lacets pour tasser des domaines dans le « produit de contact » de K par le cercle.

Soit (K, η) une variété de contact, c'est-à-dire une variété munie d'une structure de contact coorientée. L'opération appelée *symplectisation* associe à K une variété symplectique SK munie d'une action conformément symplectique de $\mathbf{R}_+^*$ et à toute transformation de contact $\phi\colon K \to K$ une transformation symplectique $\mathbf{R}_+^*$-équivariante $S\phi\colon SK \to SK$. Par définition, $SK \subset T^*K$ est l'ensemble des covecteurs non nuls $\beta \in T_q^*K$, $q \in K$, ayant pour noyau coorienté η_q. C'est donc un fibré $\mathbf{R}_+^*$-principal au-dessus de K dont les sections sont les équations (de Pfaff) globales de η. Une telle équation β fournit une trivialisation $SK \simeq \mathbf{R}_+^* \times K$. Dans ces coordonnées, la forme λ

induite sur SK par la forme (de Liouville) canonique de T^*K s'écrit $\lambda = r\beta$, $r \in \mathbf{R}_+^*$, d'où il ressort que $d\lambda$ est une forme symplectique. De plus, le champ de vecteurs $d\lambda$-dual de λ est $r\partial_r$ et engendre donc l'action de $\mathbf{R}_+^*$. Enfin, la projection $SK \to K$ envoie le noyau de λ sur la structure de contact η.

Exemple 2.1. — La structure de contact standard η_0 sur $\mathbf{S}^{2n-1}$ est le champ des hyperplans complexes tangents à la sphère unité de $\mathbf{C}^n = \mathbf{R}^{2n}$. La symplectisation de $(\mathbf{S}^{2n-1}, \eta_0)$ s'identifie à $\mathbf{C}^n \setminus \{0\}$ muni de la 1-forme

$$\lambda_0 = \tfrac{1}{2}\sum_{j=1}^{n}(x_j dy_j - y_j dx_j), \qquad x + iy = z \in \mathbf{C}^n,$$

dont le champ de vecteurs dual est le champ radial

$$\tfrac{1}{2}\sum_{j=1}^{n}(x_j \partial_{x_j} + y_j \partial_{y_j}).$$

Dans ce modèle, l'action de $\mathbf{R}_+^*$ est donnée par $r \cdot z = \sqrt{r}z$.

La tautologie suivante sera utilisée constamment :

LEMME 2.2. — *Pour toute variété de contact (K, η), la projection $SK \to K$ induit une bijection entre les transformations de contact de K et les transformations symplectiques $\mathbf{R}_+^*$-équivariantes de SK.*

Démonstration. — Étant donné une transformation de contact ϕ de K, son application cotangente est une transformation symplectique de T^*K linéaire sur les fibres qui préserve SK. Sa restriction à SK est donc une transformation symplectique $\mathbf{R}_+^*$-équivariante. Inversement, toute transformation $\mathbf{R}_+^*$-équivariante $\widetilde{\phi}$ de SK préserve le champ de vecteurs qui engendre l'action. Si elle est symplectique, elle préserve aussi la 1-forme duale, c'est-à-dire λ. Par suite, $\widetilde{\phi}$ descend sur $K = SK/\mathbf{R}_+^*$ en une transformation qui préserve la projection du noyau de λ, à savoir la structure de contact η. □

Soit ϕ_t, $t \in [0,1]$, une isotopie de contact de (K, η). Le champ de vecteurs dépendant du temps X_t qui l'engendre, donné par

$$X_t\big(\phi_t(q)\big) = \dot{\phi}_t(q) = \frac{d}{dt}\phi_t(q)$$

définit une fonction sur SK :

$$H\phi_t \colon SK \times [0,1] \longrightarrow \mathbf{R}, \quad (\beta_q, t) \longmapsto \beta_q\big(X_t(q)\big).$$

Cette fonction $H\phi_t$ est homogène au sens où

$$H\phi_t(r\beta, t) = rH\phi_t(\beta, t).$$

C'est le *hamiltonien de contact* de l'isotopie ϕ_t — et aussi le hamiltonien de l'isotopie symplectique $S\phi_t$. Elle sera parfois notée $H_t\phi_t$ quand on voudra expliciter la variable par rapport à laquelle on prend le hamiltonien. La règle de composition suivante sera très utile :

$$H(\psi_t \circ \phi_t) = H\psi_t + H\phi_t \circ (S\psi_t^{-1} \times \iota)$$
$$\text{donc} \qquad H(\phi_t^{-1}) = -H\phi_t \circ (S\phi_t \times \iota),$$

où ψ_t et ϕ_t, $t \in [0,1]$, sont deux isotopies de contact de (K,η).

PROPOSITION 2.3 ([**13**]). — *Une variété de contact close (K,η) est de type non-ordonnable si et seulement si $\mathcal{G}(K,\eta)$ contient un lacet contractile strictement positif.*

Démonstration. — Les définitions stipulent que (K,η) est de type non-ordonnable si et seulement si $\mathcal{G}(K,\eta)$ contient un lacet contractile positif (ou nul) non constant. Il reste à voir que l'existence d'un tel lacet γ_t, $t \in \mathbf{T} = \mathbf{R}/\mathbf{Z}$, implique celle d'un lacet contractile strictement positif. Comme γ_t n'est pas constant, son hamiltonien $H\gamma_t \colon SK \times \mathbf{T} \to \mathbf{R}$ (qui est partout positif ou nul) est strictement positif en un point (p_0,t_0). Quitte à remplacer γ_t par le lacet $\gamma_{t+t_0} \circ \gamma_{t_0}^{-1}$, on peut supposer que $t_0 = 0$. On note U un ouvert de M tel que $H\gamma_t$ soit strictement positif sur $SU \times \{0\}$.

Soit $\phi \in \mathcal{G}(K,\eta)$. Le lacet $\phi \circ \gamma_t \circ \phi^{-1}$ est contractile et son hamiltonien vaut $H\gamma_t \circ (S\phi^{-1} \times \iota)$ donc le lacet composé $\gamma_t \circ \phi \circ \gamma_t \circ \phi^{-1}$ a pour hamiltonien

$$H\gamma_t + H\gamma_t \circ (S\phi^{-1} \circ S\gamma_t^{-1} \times \iota)$$

qui, comme $\gamma_0 = \iota$, est strictement positif sur $S(U \cup \phi(U)) \times \{0\}$. Comme K est close et que $\mathcal{G}(K,\eta)$ opère transitivement, une itération de ce procédé donne un lacet contractile noté de nouveau γ_t dont le hamiltonien $H\gamma_t$ est strictement positif sur $SK \times \{0\}$, donc aussi sur $SK \times [-1/m, 1/m]$ pour un certain entier m. Le lacet

$$\gamma_t \circ (\gamma_{t+\frac{1}{m}} \circ \gamma_{\frac{1}{m}}^{-1}) \circ \cdots \circ (\gamma_{t+\frac{m-1}{m}} \circ \gamma_{\frac{m-1}{m}}^{-1})$$

est alors contractile et strictement positif. □

Le lemme suivant a déjà été invoqué pour prouver la proposition 1.11 :

LEMME 2.4 ([**13**]). — *Toute isotopie strictement positive $\phi_t \in \mathcal{G}(K,\eta)$, $t \in [0,1]$, est homotope, à extrémités fixes et parmi les isotopies strictement positives, à une isotopie dont le hamiltonien est 1-périodique en temps.*

Démonstration. — L'idée est de composer une isotopie $\phi_t \circ \psi_t$ qui a les mêmes extrémités que ϕ_t et vérifie

$$H(\phi_t \circ \psi_t)(z,0) = H\phi_t(z,0) + H\phi_t(z,1) = H(\phi_t \circ \psi_t)(z,1).$$

On commence par prolonger l'isotopie ϕ_t au-delà de ses extrémités en une isotopie toujours strictement positive paramétrée par $t \in [-a, 1+a]$, $a > 0$, et on pose $f_t(z) = H\phi_t(z,t)$, $(z,t) \in SK \times [-a, 1+a]$. On considère ensuite, pour $t \in [-a,a] \cup [1-a, 1+a]$, les fonctions $g_t \colon SK \to \mathbf{R}$ définies par

$$g_t = \begin{cases} f_{1+t} \circ S\phi_{1+t} & \text{si } t \in [-a,a], \\ f_{t-1} \circ S\phi_{t-1} & \text{si } t \in [1-a, 1+a]. \end{cases}$$

On désigne par ψ_t^0, $t \in [-a,a]$, et ψ_t^1, $t \in [1-a, 1+a]$, les isotopies de contact engendrées par les hamiltoniens g_t dans les intervalles respectifs et vérifiant $\psi_0^0 = \psi_1^1 = \iota$. On note enfin $\psi_{s,t}$, l'homotopie d'isotopies de contact donnée, pour tout $s \in [0,1]$, par

$$\psi_{s,t} = \begin{cases} \psi^0_{su(t)} & \text{si } t \in [-a,a], \\ \psi^1_{1+su(t-1)} & \text{si } t \in [1-a, 1+a], \\ \iota & \text{si } t \in [a, 1-a], \end{cases}$$

où $u \colon [-a,a] \to [-a,a]$ coïncide avec l'identité près de 0, est nulle près du bord et a une dérivée partout supérieure à $-\varepsilon$. Les hamiltoniens correspondants sont

$$H_t\psi_{s,t}(z,t) = \begin{cases} su'(t)\, g(z, su(t)) & \text{si } t \in [-a,a], \\ su'(t-1)\, g(z, 1+su(t-1)) & \text{si } t \in [1-a, 1+a], \\ 0 & \text{si } t \in [a, 1-a]. \end{cases}$$

On pose maintenant $\phi_{s,t} = \phi_t \circ \psi_{s,t}$, $s \in [0,1]$. Comme $\psi_{s,0} = \psi_{s,1} = \iota$ pour tout $s \in [0,1]$, c'est une homotopie à extrémités fixes qui part de $\phi_{0,t} = \phi_t$. Enfin, les hamiltoniens sont

$$H_t\phi_{s,t}(z,t) = f_t + H_t\psi_{s,t} \circ S\phi_t^{-1}.$$

À partir des expressions des $H_t\psi_{s,t}$, on vérifie sans peine que toutes ces fonctions sont positives si on prend ε assez petit et que

$$H_t\phi_{1,t}(z,t) = \begin{cases} f_t(z) + f_{1+t}(z) & \text{pour } t \text{ proche de } 0, \\ f_t(z) + f_{t-1}(z) & \text{pour } t \text{ proche de } 1 \end{cases}$$

s'étend bien en une fonction 1-périodique en temps. □

On explique maintenant comment utiliser les lacets[1] contractiles strictement positifs dans $\mathscr{G}(K,\eta)$ pour tasser des domaines dans $V = SK \times \mathbf{T}$ muni de la structure de contact d'équation $dt - \lambda = 0$ — de manière plus symétrique, V est le « produit de contact » de K et $\mathbf{T}$, c'est-à-dire le quotient de $SK \times S\mathbf{T}$ par l'action diagonale de $\mathbf{R}_+^*$. On commence par un lemme qui résulte d'un simple calcul :

[1] Tous les lacets considérés dans la suite sont basés en l'identité.

LEMME 2.5 ([**12**]). — *Soit γ_t, $t \in \mathbf{T}$, un lacet dans $\mathscr{G}(K,\eta)$ et soit*

$$W = \big\{(z,t) \in V \mid H\gamma_t(S\gamma_t(z),t) > -1\big\}.$$

L'application $\Gamma\colon W \to V$ définie par

$$\Gamma(z,t) = \big(S\gamma_t(u_t(z)\cdot z),t\big), \quad \text{où} \quad u_t(z) = \frac{1}{1+H\gamma_t\big(S\gamma_t(z),t\big)},$$

est un plongement de contact.

Soit ϕ_s, $s \in [0,1]$, une homotopie de chemins strictement positifs $\phi_{s,t} \in \mathscr{G}(K,\eta)$, $t \in [0,1]$. On suppose que les chemins partent tous de $\phi_{s,0} = \iota$, aboutissent tous en un même point $\phi_{s,1} = \phi$ et que les hamiltoniens $H_t\phi_{s,t}$ sont 1-périodiques. On pose

$$W_s = \big\{(z,t) \in V \mid H_t\phi_{s,t}(z,t) < 1\big\}$$
$$\text{et} \qquad \gamma_{s,t} = \phi_{s,t}\circ\phi_{0,t}^{-1}, \qquad s,t \in [0,1].$$

Le lecteur pourra vérifier que les plongements Γ_s, $s \in [0,1]$, associés par le lemme ci-dessus aux lacets γ_s sont définis sur W_0 et l'envoient sur W_s.

Pour toute équation β de η, on note K_β son graphe dans SK et $SK_{<\beta}$ la moitié inférieure de SK délimitée par K_β :

$$SK_{<\beta} = \big\{(z,t) \in SK \mid z = r\beta(q),\ q \in K,\ 0 < r < 1\big\}.$$

On dira qu'un domaine U de SK est étoilé s'il est du type $SK_{<\beta}$ pour une certaine équation β. De même, un domaine W dans V est étoilé s'il est du type

$$V_{<\beta_t} = \big\{(z,t) \in V \mid z \in SK_{<\beta_t}\big\}$$

où β_t, $t \in \mathbf{T}$, est un lacet d'équations de η.

THÉORÈME 2.6 ([**12**]). — *Soit γ_s, $s \in [0,1]$, une homotopie de lacets $\gamma_{s,t} \in \mathscr{G}(K,\eta)$, $t \in \mathbf{T}$, entre le lacet constant $\gamma_{0,t} = \iota$ et un lacet strictement positif $\gamma_{1,t} = \gamma_t$. Si β est une équation de η telle que*

$$(H_t\gamma_{s,t})\,|_{K_\beta\times\mathbf{T}} > -1 \quad \textit{pour tout} \quad s \in [0,1],$$

alors, pour tout $r < 1$:

(a) *on peut tasser $V_{<r\beta}$ dans $V_{<r'\beta}$ où $r' = r/(1+cr)$, le nombre $c > 0$ ne dépendant pas de r ;*

(b) *on peut tasser $V_{<r\beta}$ dans lui-même au sein de $V_{<r_0\beta}$ si on prend $r_0 > r/(1-r)$.*

De plus, dans les deux cas, le tassement se fait parmi les domaines étoilés.

Autrement dit, l'homotopie de lacets γ_s permet de tasser les domaines étoilés dont le bord est en dessous du niveau -1 de tous les hamiltoniens $H_t\gamma_{s,t}$.

Démonstration. — Soit $r < 1$ et ψ_t le flot de Reeb de β, de sorte que $K_\beta \times \mathbf{T}$ est le niveau 1 de $H\psi_t$. Le hamiltonien du chemin $\phi_t = \psi_{t/r}$ vaut alors $H\phi_t = r^{-1}H\psi_t$. On déforme, à extrémités fixes, le chemin ϕ_t, $t \in [0,1]$, par l'homotopie

$$\phi_{s,t} = \phi_t \circ \gamma_{s,t}, \quad s,t \in [0,1].$$

D'après la règle de composition des hamiltoniens,

$$H_t\phi_{s,t}(z,t) = H\phi_t(z) + H_t\gamma_{s,t}\big(S\phi_{-t}(z),t\big).$$

Par hypothèse, $H_t\gamma_{s,t} > -H\psi_t$ donc

$$H_t\phi_{s,t}(z,t) > r^{-1}H\psi_t(z) - H\psi_t\big(\psi_{-t/r}(z)\big).$$

Comme ψ_t est un flot, $H\psi_t$ est constant sur les orbites et

$$H_t\phi_{s,t}(z,t) > (r^{-1}-1)H\psi_t(z) > r_0^{-1}H\psi_t(z).$$

Grâce à un petit reparamétrage des chemins $\phi_{s,t}$ et à une version paramétrique du lemme 2.4, on peut faire en sorte que les hamiltoniens $H_t\phi_{s,t}$ soient en outre 1-périodiques en temps. Si on pose

$$W_s = \big\{(z,t) \in V \mid H_t\phi_{s,t}(z,t) < 1\big\},$$

on a donc $W_s \subset V_{<r_0\beta}$. De plus, le lemme 2.5 et le commentaire qui le suit donnent une isotopie de contact de V qui envoie $W_0 = V_{<r\beta}$ sur W_s, donc l'adhérence de W_0 dans $W_1 \subset W_0$ car le lacet γ_1 est strictement positif. On a ainsi tassé $V_{<r\beta}$ dans lui-même au sein de $V_{<r_0\beta}$, ce qui démontre le point (b). Soit maintenant $c > 0$ tel que $H\gamma_{1,t} > cH\phi_t$. Les estimations ci-dessus disent que

$$H\phi_{1,t}(z,t) > r^{-1}H\psi_t(z) + cH\psi_t(z) = \frac{1+cr}{r}H\psi_t(z)$$

et donc que $W_1 \subset V_{<r'\beta}$ où $r' = r/(1+cr)$. Ceci démontre le point (a). □

B. Dans la sphère de contact standard

On commence par une incursion dans l'espace hyperbolique complexe. On considère sur $\mathbf{C}^{n+1}$ la forme pseudohermitienne de signature $(n,1)$ donnée par

$$\langle z', z''\rangle = -\overline{z}'_0 z''_0 + \sum_{j=1}^{n} \overline{z}'_j z''_j, \qquad z', z'' \in \mathbf{C}^{n+1}.$$

Par définition, l'espace hyperbolique complexe $\mathbf{CH}^n$ est l'image dans l'espace projectif $\mathbf{CP}^n$ de l'ouvert $\{z \in \mathbf{C}^{n+1} \mid \langle z,z\rangle < 0\}$. Dans la carte affine $\{z_0 = 1\}$, il apparaît ainsi comme la boule unité ouverte de $\mathbf{C}^n$,

$$\mathbf{CH}^n = \big\{(z_1,\dots,z_n) \in \mathbf{C}^n \mid -1 + \sum_{j=1}^{n} |z_j|^2 < 0\big\},$$

et a pour bord naturel la sphère $\mathbf{S}^{2n-1}$, projection dans $\mathbf{CP}^n$ du cône isotrope $Q_0 = \{z \in \mathbf{C}^{n+1} \mid \langle z, z\rangle = 0,\ z \neq 0\}$.

De manière équivalente, $\mathbf{CH}^n$ est le quotient de la quadrique affine (au sens réel) $Q = \{z \in \mathbf{C}^{n+1} \mid \langle z, z\rangle = -1\}$ par l'action du cercle $\mathbf{S}^1 \subset \mathbf{C}^*$, action qui préserve la forme $\langle .,. \rangle$. Comme cette forme est définie positive sur l'orthogonal des orbites de $\mathbf{S}^1$, elle induit sur $\mathbf{CH}^n$ une métrique hermitienne, qui est même kählérienne puisque la partie imaginaire de la forme $\langle .,. \rangle$ est la forme symplectique

$$\omega = -dx_0 \wedge dy_0 + \sum_{j=1}^{n} dx_j \wedge dy_j, \qquad x + iy \in \mathbf{C}^{n+1}.$$

Les isométries de $\mathbf{CH}^n$ sont les transformations projectives de $\mathbf{CP}^n$ issues du groupe $\mathrm{U}(n,1)$ des applications linéaires de $\mathbf{C}^{n+1}$ qui préservent la forme $\langle .,. \rangle$. Elles composent le groupe $\mathrm{PU}(n,1) = \mathrm{U}(n,1)/\mathbf{S}^1$ où le cercle $\mathbf{S}^1$ est le sous-groupe des homothéties. Comme ces isométries sont des transformations holomorphes de $\mathbf{CP}^n$ et préservent la sphère $\mathbf{S}^{2n-1} = Q_0/\mathbf{C}^*$, elles préservent aussi sur $\mathbf{S}^{2n-1}$ le champ des hyperplans complexes tangents, c'est-à-dire la structure de contact standard η_0. Autrement dit, $\mathrm{PU}(n,1)$ se plonge naturellement dans le groupe $\mathscr{D}(\mathbf{S}^{2n-1}, \eta_0)$ des transformations de contact de $\mathbf{S}^{2n-1}$.

Pour comprendre les transformations de contact, il est utile de connaître leur action sur la symplectisation. Dans le cas des éléments de $\mathrm{U}(n)$ (qui est un sous-groupe de $\mathrm{PU}(n,1)$ via l'inclusion $\mathrm{U}(n) \to \mathrm{U}(n,1)$, $\phi \mapsto \iota \times \phi$), cette action est celle de la transformation linéaire en dehors de l'origine. Dans le cas des éléments de $\mathrm{PU}(n,1)$, il convient d'identifier la symplectisation de $\mathbf{S}^{2n-1}$ comme le saturé de $\mathbf{S}^{2n-1} = Q_0 \cap \{z_0 = 1\}$ par le flot de Liouville radial, ou encore comme le quotient de Q_0 par l'action de $\mathbf{S}^1$.

Exemple 2.7. — Pour $t \in \mathbf{R}$, soit $\widetilde{\tau}_t \in \mathrm{U}(n,1)$ la transformation de $\mathbf{C}^{n+1}$ associée à la matrice

$$\begin{pmatrix} \operatorname{ch} t & \operatorname{sh} t & 0 \\ \operatorname{sh} t & \operatorname{ch} t & 0 \\ 0 & 0 & 1_{n-1} \end{pmatrix}.$$

L'élément de $\mathrm{PU}(n,1)$ correspondant est la transformation projective τ_t qui, dans la carte $\{z_0 = 1\}$, s'écrit

$$\tau_t(z_1, z_2, \ldots, z_n) = \frac{1}{\operatorname{ch} t + \operatorname{sh} t\, z_1}(\operatorname{sh} t + \operatorname{ch} t\, z_1, z_2, \ldots, z_n).$$

Géométriquement, τ_t, $t \in \mathbf{R}$, est le flot des translations de $\mathbf{CH}^n$ le long de la géodésique entre les points $(-1, 0, \ldots, 0)$ et $(1, 0, \ldots, 0)$.

Quant à la transformation $S\tau_t$ sur la symplectisation de $\mathbf{S}^{2n-1}$ vue comme $\mathbf{C}^n\setminus\{0\}$, elle est donnée par

$$S\tau_t(z_1,z_2,\dots,z_n)=\frac{\big|\operatorname{ch} t\,|z|+\operatorname{sh} t\,z_1\big|}{\operatorname{ch} t\,|z|+\operatorname{sh} t\,z_1}\big(\operatorname{sh} t\,|z|+\operatorname{ch} t\,z_1,z_2,\dots,z_n\big).$$

Pour compléter le tableau succinct, on décrit le préordre sur $\mathrm{PU}(n,1)$ induit par $\mathscr{D}(\mathbf{S}^{2n-1},\eta_0)$ et, pour commencer, on explique comment calculer le hamiltonien de contact d'un chemin dans $\mathrm{PU}(n,1)$.

Toute transformation linéaire $\phi\in\mathrm{U}(n,1)$ préserve la forme symplectique ω et, mieux, sa primitive obtenue par produit intérieur avec le champ radial. Du coup, tout chemin $\widetilde{\phi}_t$ dans $\mathrm{U}(n,1)$, $t\in[0,1]$, définit une isotopie hamiltonienne de $(\mathbf{C}^{n+1},\omega)$. Celle-ci est engendrée par une unique fonction hamiltonienne homogène de degré 2

$$\widetilde{f}=\widetilde{H}\widetilde{\phi}_t\colon \mathbf{C}^{n+1}\times[0,1]\longrightarrow\mathbf{R}$$

où la condition d'homogénéité s'écrit

$$\widetilde{f}(uz,t)=|u|^2\widetilde{f}(z,t),\qquad u\in\mathbf{C},\ z\in\mathbf{C}^{n+1},\ t\in[0,1].$$

Par exemple, si $\widetilde{\phi}_t(z)=e^{2i\pi t}z$, alors $\widetilde{H}\widetilde{\phi}_t(z,t)=\pi\langle z,z\rangle$ est une fonction qui s'annule sur le cône isotrope Q_0.

Vu la règle de composition pour les hamiltoniens (qui s'adapte immédiatement), cet exemple montre que, si ϕ_t est un chemin dans $\mathrm{PU}(n,1)$ et $\widetilde{\phi}_t$ un relèvement de ce chemin dans $\mathrm{U}(n,1)$, la restriction au cône Q_0 de la fonction $\widetilde{H}\widetilde{\phi}_t$ ne dépend pas du relèvement choisi. D'autre part, la condition d'homogénéité assure que cette restriction est invariante par l'action du cercle et descend donc en une fonction $H\phi_t$ sur $Q_0/\mathbf{S}^1$ qui, comme on l'a dit plus haut, est la symplectisation de $\mathbf{S}^{2n-1}$. La fonction $H\phi_t$ ainsi obtenue est bien sûr le hamiltonien de contact du chemin ϕ_t.

Dorénavant, les groupes $\mathrm{SU}(n)\subset\mathrm{U}(n)\subset\mathrm{PU}(n,1)$ seront regardés comme des sous-groupes du groupe $\mathscr{D}(\mathbf{S}^{2n-1},\eta_0)$ des transformations de contact de la sphère. Pour tout chemin ϕ_t dans l'un d'eux, $S\phi_t$ désigne ainsi son relèvement à la symplectisation de $\mathbf{S}^{2n-1}$, identifiée à $\mathbf{C}^n\setminus\{0\}$, et $H\phi_t$ une fonction sur $\mathbf{C}^n\setminus\{0\}$ qui dépend *a priori* du temps t.

Le théorème suivant peut être déduit des résultats généraux de G. I. Olshanskii [**22**] sur les ordres invariants dans les groupes de Lie (voir aussi [**18**]). On en donne ci-après une preuve totalement explicite et élémentaire en exhibant un lacet contractile strictement positif.

THÉORÈME 2.8 ([**12, 22**]). — *Pour $a,t\in\mathbf{R}$, soit*

$$\gamma_t=\phi_{2t}\circ[\rho_{-t},\tau_{-a}]=\phi_{2t}\circ\rho_{-t}\circ\tau_{-a}\circ\rho_t\circ\tau_a\in\mathrm{PU}(n,1),$$

où $\phi_t \in \mathrm{SU}(n)$, $\rho_t \in \mathrm{U}(n)$ *et* $\tau_a \in \mathrm{PU}(n,1)$ *sont les transformations de* $\mathbf{S}^{2n-1}$ *définies par*

$$\begin{aligned}
\phi_t(z_1, z_2, \dots, z_n) &= \big(e^{2i\pi(n-1)t} z_1, e^{-2i\pi t} z_2, \dots, e^{-2i\pi t} z_n\big),\\
\rho_t(z_1, z_2, \dots, z_n) &= e^{2i\pi t}(z_1, z_2, \dots, z_n),\\
\tau_a(z_1, z_2, \dots, z_n) &= \frac{1}{\operatorname{ch} a + \operatorname{sh} a\, z_1}(\operatorname{sh} a + \operatorname{ch} a\, z_1, z_2, \dots, z_n).
\end{aligned}$$

Le lacet γ_t, $t \in \mathbf{T}$, *est contractile et positif pour peu que* a *soit assez grand.*

Corollaire 2.9 ([**12**]). — *La sphère de contact standard de dimension* $2n-1$ *est de type non-ordonnable.*

La démonstration qui suit recourt, pour $z = (z_1, z_2, \dots, z_n) \in \mathbf{C}^n$, aux notations suivantes :

$$r_j(z) = \pi |z_j|^2,\ 1 \le j \le n, \quad r(z) = \sum_{j=1}^{n} r_j(z), \quad r'(z) = \sum_{j=2}^{n} r_j(z).$$

Démonstration. — Le lacet γ_t est contractile parce que les lacets $[\rho_{-t}, \tau_{-a}]$ et ϕ_t le sont. Comme le hamiltonien du flot ρ_t est $H\rho_t = r$ et ne dépend pas de t, celui du lacet $[\rho_{-t}, \tau_{-a}]$ vaut

$$H[\rho_{-t}, \tau_{-a}] = H\rho_{-t} + H\rho_t \circ S\tau_a \circ S\rho_t = -H\rho_t \circ S\rho_t + H\rho_t \circ S\tau_a \circ S\rho_t = (-r + r \circ S\tau_a) \circ S\rho_t.$$

Comme on l'a vu, pour $z = (z_1, \dots, z_n)$,

$$S\tau_a(z_1, z_2, \dots, z_n) = \frac{\big|\operatorname{ch} a\, |z| + \operatorname{sh} a\, z_1\big|}{\operatorname{ch} a\, |z| + \operatorname{sh} a\, z_1}\big(\operatorname{sh} a\, |z| + \operatorname{ch} a\, z_1, z_2, \dots, z_n\big),$$

donc $r' \circ S\tau_a = r'$ (mieux, $r_j \circ S\tau_a = r_j$ pour $j \ge 2$) et

$$\begin{aligned}
r_1 \circ S\tau_a(z) &= \pi \big|\operatorname{sh} a\, |z| + \operatorname{ch} a\, z_1\big|^2\\
&= \operatorname{ch} a\, r(z) \big|\operatorname{th} a + z_1/|z|\big|^2 = \operatorname{ch} a\, r(z) \big|(\operatorname{th} a - 1) + (1 + z_1/|z|)\big|^2.
\end{aligned}$$

Soit $W_\varepsilon = \{r' \ge \varepsilon r_1\} \subset \mathbf{C}^n \setminus \{0\}$, avec $\varepsilon > 0$: c'est un cône invariant par $\mathbf{C}^*$ et évitant un voisinage de la demi-droite réelle engendrée par $(-1, 0, \dots, 0)$. Dès que $a > 0$ est assez grand (voir la figure 1), ε étant donné,

$$r_1 \circ S\tau_a(z) \ge 3r(z) \quad \text{pour tout} \quad z \in W_\varepsilon$$

et, comme W_ε est invariant par $S\rho_t$,

$$H[\rho_{-t}, \tau_{-a}] = (-r + r \circ S\tau_a) \circ S\rho_t = (-r_1 + r_1 \circ S\tau_a) \circ S\rho_t \ge (-r_1 + 3r) \circ S\rho_t \ge r' + 2r > 0.$$

Ainsi, le hamiltonien du lacet $[\rho_{-t}, \tau_{-a}]$ est positif partout sauf peut-être dans la région $r' < \varepsilon r_1$ complémentaire de W_ε, région où le hamiltonien du flot ϕ_t, à savoir

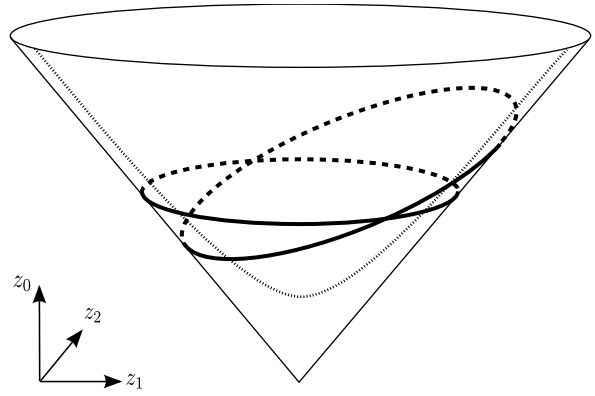


FIGURE 1. Image de la sphère par la translation hyperbolique.

$H\phi_t = (n-1)r_1 - r'$, est positif. Pour voir que la composition par ϕ_{2t} apporte la compensation nécessaire, on écrit

$$\begin{aligned} H\gamma_t(z,t) &= H\phi_{2t}(z) + H[\rho_{-t}, \tau_{-a}](S\phi_{-2t}(z), t) \\ &= 2(n-1)r_1(z) - 2r'(z) + (-r_1 + r_1 \circ S\tau_a) \circ S\rho_t \circ S\phi_{-2t}(z). \end{aligned}$$

Comme les fonctions r_1 et r' sont invariantes par les flots $S\phi_t$ et $S\rho_t$, il reste à vérifier en tout point que

$$2(n-1)r_1 - 2r' - r_1 + r_1 \circ S\tau_a = (2n-3)r_1 - 2r' + r_1 \circ S\tau_a > 0.$$

Or, dans W_ε, l'estimation $r_1 \circ S\tau_a \geq 3r$ donne

$$(2n-3)r_1 - 2r' + r_1 \circ S\tau_a \geq (2n-3)r_1 - 2r' + 3r = r + (2n-1)r_1 > 0$$

et, dans le complémentaire, $r_1 > \varepsilon r'$ donc $r_1 > 0$ et

$$(2n-3)r_1 - 2r' + r_1 \circ S\tau_a \geq (2n-3-2\varepsilon)r_1 > 0,$$

ce qui conclut. □

Remarque 2.10. — On peut se demander s'il n'y a pas, dans $\mathscr{D}(\mathbf{S}^{2n-1}, \eta_0)$, un lacet positif contractile plus simple que dans $\mathrm{PU}(n,1)$ et par exemple si le lacet ρ_t, $t \in \mathbf{T}$, n'est pas contractile. C'est peu probable mais je n'en ai pas la preuve. Ce lacet n'est pas contractile dans $\mathrm{U}(n)$ mais, pour n pair, il l'est dans $\mathrm{SO}(2n)$ et donc dans $\mathscr{D}(\mathbf{S}^{2n-1})$. Pour $n = 2$ cependant, les résultats de Y. Eliashberg **[10]** sur l'espace des structures de contact tendues de $\mathbf{S}^3$ et le théorème de A. Hatcher **[17]** (voir aussi **[20]**) qui confirme la conjecture de Smale sur les difféomorphismes de $\mathbf{S}^3$ s'articulent pour montrer que l'inclusion de $\mathrm{U}(2)$ dans $\mathscr{D}(\mathbf{S}^3, \eta_0)$ est une équivalence d'homotopie faible.

Les deux énoncés qui suivent serviront pour démontrer le théorème 1.2-(a).

PROPOSITION 2.11 ([**12**]). — *On peut relier dans* $\mathrm{PU}(n,1)$ *le lacet constant* $\gamma_{0,t} = \iota$ *au lacet* $\gamma_{3,t} = \gamma_t$, $t \in \mathbf{T}$, *par une homotopie* $\gamma_{s,t}$, $s \in [0,3]$, *de lacets dont les hamiltoniens*

$$g_s = H_t\gamma_{s,t} \colon \big(\mathbf{C}^n \setminus \{0\}\big) \times \mathbf{T} \to \mathbf{R}, \quad (z,t) \mapsto g_s(z,t) = g_{s,t}(z),$$

vérifient

$$g_{s,t} \geq -r \quad \textit{pour tous} \quad t \in \mathbf{T},\ s \in [0,3]\,.$$

LEMME 2.12 ([**12**]). — *On peut relier dans* $\mathrm{SU}(n)$ *le lacet constant* $\phi_{0,t} = \iota$ *au lacet* $\phi_{1,t} = \phi_t$, $t \in \mathbf{T}$, *par une homotopie* $\phi_{s,t}$, $s \in [0,1]$, *de lacets dont les hamiltoniens*

$$f_s = H_t\phi_{s,t} \colon \big(\mathbf{C}^n \setminus \{0\}\big) \times \mathbf{T} \to \mathbf{R}, \quad (z,t) \mapsto f_s(z,t) = f_{s,t}(z),$$

vérifient

$$f_{s,t} \circ S\phi_{s,t} \geq -r' \quad \textit{pour tous} \quad t \in \mathbf{T},\ s \in [0,1].$$

Démonstration du lemme. — On commence par le cas $n = 2$. On considère les transformations unitaires

$$\delta_t(z_1,z_2) = \big(z_1, e^{2i\pi t}z_2\big), \qquad t \in \mathbf{T},$$
$$\chi_s(z_1,z_2) = \big(\cos(\pi s/2)\, z_1 - \sin(\pi s/2)\, z_2, \sin(\pi s/2)\, z_1 + \cos(\pi s/2)\, z_2\big), \qquad s \in [0,1]$$

et on pose

$$\phi_{s,t} = [\chi_s, \delta_t] = \chi_s \circ \delta_t \circ \chi_{-s} \circ \delta_{-t}.$$

Ceci est une homotopie entre le lacet constant $\phi_{0,t} = \iota$ et le lacet $\phi_{1,t} = \phi_t$. Pour estimer $f_{s,t} \circ S\phi_{s,t}$, on note d'abord que cette fonction n'est autre que $-H_t(\phi_{s,t}^{-1})$. Comme

$$\phi_{s,t}^{-1} = [\delta_t, \chi_s] = \delta_t \circ \chi_s \circ \delta_{-t} \circ \chi_{-s}$$

et que le hamiltonien de δ_t est $H\delta_t = r_2 = r'$, la règle de composition pour les hamiltoniens donne

$$f_{s,t}\circ S\phi_{s,t} = -H\delta_t - H\delta_{-t}\circ S\chi_{-s}\circ S\delta_{-t} = -H\delta_t + H\delta_t\circ S\delta_t\circ S\chi_{-s}\circ S\delta_{-t} \geq -H\delta_t = -r',$$

comme voulu.

Dans le cas $n > 2$, on écrit

$$\phi_t = \phi_t^{(2)} \circ \cdots \circ \phi_t^{(n)}$$

où chaque $\phi_t^{(j)} \in \mathrm{SU}(n)$ opère diagonalement en multipliant les coordonnées z_1 et z_j par $e^{2i\pi t}$ et $e^{-2i\pi t}$, respectivement, et en fixant les autres. La solution du cas $n = 2$ fournit, pour tout j, une homotopie $\phi_{s,t}^{(j)}$, $s \in [0,1]$, entre le lacet constant et le lacet $\phi_t^{(j)}$. La règle de composition pour les hamiltoniens montre que le produit

$$\phi_{s,t} = \phi_{s,t}^{(2)} \circ \cdots \circ \phi_{s,t}^{(n)}, \qquad s \in [0,1],$$

est une homotopie qui convient. □

Démonstration de la proposition. — Avec les notations du lemme 2.12, on pose

$$\gamma_{s,t} = \begin{cases} [\rho_{-t}, \tau_{-sa}] & s \in [0,1], \\ \phi_{s-1,t} \circ [\rho_{-t}, \tau_{-a}] & s \in [1,2], \\ \phi_{s-2,t} \circ \phi_t \circ [\rho_{-t}, \tau_{-a}] & s \in [2,3]. \end{cases}$$

Comme $\phi_{2t} = \phi_t \circ \phi_t$, c'est bien une homotopie entre le lacet constant $\gamma_{0,t} = \iota$ et le lacet $\gamma_{3,t} = \gamma_t$.

Pour $s \in [0,1]$, moyennant les calculs de la démonstration du théorème 2.8,

$$g_{s,t} = (-r_1 + r_1 \circ \tau_{-sa}) \circ S\rho_t \geq -r_1 \circ S\rho_t = -r_1 \geq -r.$$

Pour $s \in [1,2]$, en utilisant l'estimation du lemme 2.12,

$$g_{s,t} \circ S\phi_{s-1,t} = f_{s-1,t} \circ S\phi_{s-1,t} + (-r_1 + r_1 \circ S\tau_{-a}) \circ S\rho_t \geq -r' - r_1 \circ S\rho_t = -r$$

et, comme $S\phi_{s-1,t}$ préserve r (toute l'homotopie se fait dans $\mathrm{U}(n)$), on a bien $g_{s,t} \geq -r$.

Pour $s \in [2,3]$,

$$\begin{aligned} g_{s,t} \circ S\phi_{s-2,t} &= f_{s-2,t} \circ S\phi_{s-2,t} + (n-1)r_1 - r' + (-r_1 + r_1 \circ S\tau_{-a}) \circ S\rho_t \circ S\phi_{-t} \\ &\geq (n-1)r_1 - 2r' + (-r_1 + r_1 \circ S\tau_{-a}) \circ S\rho_t \circ S\phi_{-t} \\ &= (n-2)r_1 - 2r' + r_1 \circ S\tau_{-a} \circ S\rho_t \circ S\phi_{-t}. \end{aligned}$$

On observe alors que le domaine $W_\varepsilon = \{r' \geq \varepsilon r_1\}$ est invariant non seulement par $S\rho_t$ mais aussi par $S\phi_t$ (voir la démonstration du théorème 2.8). Or, dans W_ε, l'estimation $r_1 \circ S\tau_a \geq 3r$ donne

$$(n-2)r_1 - 2r' + r_1 \circ S\tau_{-a} \circ S\rho_t \circ S\phi_{-t} \geq (n-2)r_1 - 2r' + 3r = r + nr_1 > 0$$

et, dans le complémentaire, $r_1 > \varepsilon r'$ donc $r_1 > 0$ et

$$(n-2)r_1 - 2r' + r_1 \circ S\tau_{-a} \circ S\rho_t \circ S\phi_{-t} \geq (n-2-2\varepsilon)r_1 \geq -r_1 \geq -r,$$

ce qui conclut puisque $S\phi_{s-2,t}$ préserve r. □

On termine cette partie par la :

Démonstration du théorème 1.2-(a). — On rappelle qu'il faut prouver que, pour tout compact $A \subset C(1) \times \mathbf{T}$ et tout nombre $d > 0$, il existe une isotopie de contact de V qui envoie A dans $B(d) \times \mathbf{T}$. Avec les mêmes notations que précédemment, $C = C(1)$ désigne ici le cylindre $\{r_2 < 1\} \subset \mathbf{C}^n$. On pose d'autre part

$$D = \{r' < 1/n\}, \qquad E = \{r_2 - r - 1 + nr' < 1\}$$

et on considère les lacets dans $\mathrm{SU}(n)$ donnés par

$$\phi_t(z_1, z_2, \dots, z_n) = \big(e^{2i\pi(n-1)t} z_1, e^{-2i\pi t} z_2, \dots, e^{-2i\pi t} z_n\big),$$
$$\psi_t(z_1, z_2, \dots, z_n) = \big(e^{2i\pi t} z_1, e^{-2i\pi t} z_2, z_3, \dots, z_n\big).$$

Ces lacets ont pour hamiltoniens $H\phi_t = (n-1)r_1 - r'$ et $H\psi_t = r_1 - r_2$. En outre, le lemme 2.5 leur associe les plongements de contact (d'un ouvert) de $\mathbf{C}^n \times \mathbf{T}$ dans $\mathbf{C}^n \times \mathbf{T}$ définis par

$$\Phi(z,t) = \frac{S\phi_t(z)}{\sqrt{1 + H\phi_t(z)}}$$
$$\text{et} \quad \Psi(z,t) = \frac{S\psi_t(z)}{\sqrt{1 + H\psi_t(z)}}.$$

(On rappelle que l'action de $\mathbf{R}_+^*$ sur $\mathbf{C}^n \setminus \{0\}$ vu comme symplectisation de la sphère est donnée par $c \cdot z = \sqrt{c} z$.)

D'après le lemme 2.12, le lacet ϕ_t est contractile parmi les lacets de hamiltoniens minorés par $-r'$. Le plongement Φ est donc bien défini sur $D \times \mathbf{T}$ et isotope à l'inclusion parmi les plongements de contact.

De même, le lacet ψ_t est contractile parmi les lacets de hamiltoniens minorés par $-r_2$ de sorte que Ψ est bien défini sur $C \times \mathbf{T}$ et isotope à l'inclusion parmi les plongements de contact.

Soit $A \subset C$ un compact. Pour $a > 0$ assez grand, la translation hyperbolique τ_a envoie A dans $E \cap C$. Or, pour $z \in E \cap C$ et $(w,t) = \Psi(z,t)$,

$$r'(w) = \frac{r'(z)}{1 + r_1(z) - r_2(z)} < \frac{1}{n}$$

donc $\Psi \circ (\tau_a \times \iota)(A) \subset D$. D'autre part, pour $z \in D$ et $(w,t) = \Phi(z,t)$,

$$r(w) = \frac{r_1(z) + r'(z)}{1 + (n-1)r_1(z) - r'(z)} \leq \frac{r_1(z) + 1/n}{1 + (n-1)r_1(z) - 1/n} = \frac{1}{n-1}$$

donc

$$\Phi \circ \Psi \circ (\tau_a \times \iota)(A) \subset B\big(1/(n-1)\big) \times \mathbf{T}.$$

On utilise maintenant le lacet contractile strictement positif dans $\mathrm{PU}(n,1)$. Grâce à la proposition 2.11, le théorème 2.6-(a) permet, pour tout $s < 1$, de tasser $B(s) \times \mathbf{T}$ dans $B\big(s/(1+cs)\big)$ où $c > 0$ est indépendant de s. En itérant cette opération, on tasse $B(s) \times \mathbf{T}$ dans $B(d) \times \mathbf{T}$ avec d arbitrairement petit. □

3. OBSTRUCTIONS

A. Fonctions génératrices

Soit M une variété et $f\colon M \to \mathbf{R}$ une fonction. Le graphe de $j^1 f$ est une sous-variété legendrienne M_f de J^1M et sa projection dans T^*M, à savoir le graphe M_{df} de df, une sous-variété lagrangienne exacte (c'est-à-dire sur laquelle la 1-forme canonique de Liouville induit une forme exacte). La fonction f renseigne beaucoup sur la géométrie de ces sous-variétés. Typiquement, les points critiques de f sont les points d'intersection de M_{df} avec la section nulle M_0 et ses valeurs critiques sont les altitudes des relèvements de ces points dans $M_f \subset J^1M$, la fonction hauteur étant la projection $J^1M \to \mathbf{R}$. Malheureusement, rares sont les sous-variétés du type M_f ou M_{df} : ce sont uniquement les sous-variétés legendriennes de J^1M ou lagrangiennes exactes de T^*M qui se projettent difféomorphiquement sur la base M.

La théorie des fonctions génératrices permet de décrire beaucoup plus de sous-variétés legendriennes dans J^1M et lagrangiennes exactes dans T^*M par des fonctions. Cependant, ces fonctions sont définies non plus sur M mais sur un fibré au-dessus de M. Ce mode de représentation a été inventé par L. Hörmander [**19**].

Une *fonction génératrice* est une fonction

$$S\colon E \to \mathbf{R}, \quad (x,u) \mapsto S(x,u)$$

définie sur un fibré vectoriel E au-dessus de M (dont les éléments sont notés (x,u) avec u dans la fibre de $x \in M$) et dont la dérivée partielle $\partial_u S$ s'annule transversalement. Une telle fonction engendre deux sous-variétés (*a priori* seulement immergées), l'une legendrienne dans J^1M et l'autre lagrangienne exacte dans T^*M (la projection de la première), à savoir

$$L_S = \big\{(x, \partial_x S(x,u), S(x,u)) \in J^1M \mid \partial_u S(x,u) = 0\big\}$$
$$\text{et} \quad L_{dS} = \big\{(x, \partial_x S(x,u)) \in T^*M \mid \partial_u S(x,u) = 0\big\}.$$

Une sous-variété L, legendrienne dans J^1M ou lagrangienne exacte dans T^*M, est dite représentée par la fonction génératrice S si $L = L_S$ ou $L = L_{dS}$. Comme plus haut, les points critiques de S sont alors les points d'intersection de L_{dS} avec la section nulle M_0 et ses valeurs critiques sont les altitudes des relèvements de ces points dans $L_S \subset J^1M$. Cela dit, même si M est une variété close, S est définie sur un espace E non compact et peut très bien n'avoir aucun point critique. On coupe court à cette éventualité en se restreignant à des fonctions qui coïncident à l'infini avec une forme quadratique non-dégénérée et sont ainsi passibles d'un avatar de la théorie de Morse.

On appellera *fonction quadratrice* de L toute fonction génératrice S de L qui, en dehors d'un compact de E, est de la forme $S(x,u) = Q(x,u) + c$ où $c \in \mathbf{R}$ est une

constante et $Q\colon E \to \mathbf{R}$ une (fonction dont la restriction à chaque fibre est une) forme quadratique non dégénérée.

Pour énoncer le théorème fondamental des fonctions génératrices, on a encore besoin d'une notion d'équivalence. Étant donné des fonctions $S\colon E \to \mathbf{R}$ et $T\colon F \to \mathbf{R}$ sur des fibrés vectoriels de même base M, on note $S \oplus T\colon E \oplus F \to \mathbf{R}$ la fonction $(x,u,v) \mapsto S(x,u) + T(x,v)$. On dit que deux fonctions quadratrices $S_1\colon E_1 \to \mathbf{R}$ et $S_2\colon E_2 \to \mathbf{R}$ sont équivalentes s'il existe des formes quadratiques non dégénérées $Q_i\colon F_i \to \mathbf{R}$, $i \in \{1,2\}$, et un difféomorphisme fibré $\Phi\colon E_2 \oplus F_2 \to E_1 \oplus F_1$ tels que $S_2 \oplus Q_2 = (S_1 \oplus Q_1) \circ \Phi$.

Remarque 3.1. — L'opération de somme directe employée ici sur les fonctions quadratrices est liée à une opération géométrique sur les sous-variétés [**30**]. Pour toutes sous-variétés legendriennes L_1 et L_2 de J^1M, on pose

$$L_1 + L_2 = \big\{(x, y_1+y_2, z_1+z_2) \in J^1M \mid (x,y_1,z_1) \in L_1,\ (x,y_2,z_2) \in L_2\big\}.$$

Sous une hypothèse de transversalité *ad hoc*, $L_1 + L_2$ est une sous-variété automatiquement legendrienne. De plus, si L_1 et L_2 ont des fonctions quadratrices respectives S_1 et S_2, alors $S_1 \oplus S_2$ est une fonction quadratrice de $L_1 + L_2$.

Dans le prolongement, pour toute sous-variété legendrienne L de J^1M, on note $-L$ la sous-variété

$$-L = \big\{(x,-y,-z) \in J^1M \mid (x,y,z) \in L\big\},$$

dont les fonctions quadratrices sont opposées à celles de L. Noter que $L - L = L + (-L)$ n'est la section nulle M_0 que si L se projette sur M par difféomorphisme.

Le théorème ci-dessous est principalement dû à J.-C. Sikorav et à C. Viterbo qui ont établi respectivement l'existence de fonctions quadratrices et leur unicité à équivalence près dans le cas lagrangien. Les extensions au cas legendrien sont dues respectivement à Y. Chekanov et D. Théret.

THÉORÈME 3.2 ([**5, 26, 28, 30**]). — *Soit M une variété close.*

(a) *Dans T^*M, toute isotopie L_t de sous-variétés lagrangiennes exactes partant de la section nulle est représentée par un chemin de fonctions quadratrices $S_t\colon E \to \mathbf{R}$. De plus, toute fonction quadratrice de L_t est équivalente à $S_t + c$ pour une certaine constante c.*

(b) *Dans J^1M, toute isotopie L_t de sous-variétés legendriennes partant de la section nulle est représentée par un chemin de fonctions quadratrices $S_t\colon E \to \mathbf{R}$. De plus, toute fonction quadratrice de L_t est équivalente à S_t.*

Quelques mots sur la construction des fonctions quadratrices seront utiles par la suite. On commence par une observation sur le cas lagrangien due à A. Weinstein. Soit E l'espace des chemins $u\colon [0,1] \to T^*M$ dont l'origine $u(0)$ est sur la section nulle M_0,

espace fibré au-dessus de M (supposée close) par l'application $u \mapsto u(1)$. Soit d'autre part H_t, $t \in [0,1]$, un hamiltonien dépendant du temps sur T^*M et M_t l'image de M_0 à l'instant t par l'isotopie hamiltonienne engendrée par H_t. L'observation est que la fonctionnelle d'action $S\colon E \to \mathbf{R}$ donnée par

$$S(u) = \int_0^1 u^*\lambda - H_t\big(u(t)\big)\,dt$$

est formellement une fonction génératrice de M_1 définie sur un fibré non vectoriel et de dimension infinie.

C'est M. Chaperon **[3]** qui a eu l'idée d'approximer E par un espace de dimension finie formé de « géodésiques brisées », idée développée par la suite dans **[21, 25]** avant d'aboutir dans **[26]** à l'existence de fonctions quadratrices.

En suivant **[4]** à travers **[24]**, on décrit maintenant une construction de fonctions quadratrices pour une isotopie legendrienne L_t à support compact dans $J^1\mathbf{R}^n$ (le schéma général de toutes les constructions connues est le même).

Soit $f\colon J^1\mathbf{R}^n \to \mathbf{R}$ la fonction $(x,y,z) \mapsto xy+z$ qu'on regarde comme une famille $f_{y,z}$ de fonctions sur $\mathbf{R}^n$. Les sections

$$j^1 f_{y,z}\colon \mathbf{R}^n \to J^1\mathbf{R}^n, \quad x \mapsto (x,y,xy+z), \qquad (y,z) \in \mathbf{R}^n \times \mathbf{R},$$

sont des plongements legendriens dont les images $F_{y,z}$ feuillettent $J^1\mathbf{R}^n$.

Si $\psi \in \mathscr{G}(J^1\mathbf{R}^n)$ est une transformation de contact $\mathscr{C}^1$-proche de l'identité, chaque sous-variété $\psi(F_{y,z})$ reste le graphe d'une section $j^1 g_{y,z}\colon \mathbf{R}^n \to J^1\mathbf{R}^n$, où $g_{y,z} = g(.,y,z)$ est une fonction sur $\mathbf{R}^n$. On appellera ci-dessous fonction de transition associée à ψ la fonction

$$G\colon J^1\mathbf{R}^n \to \mathbf{R} \quad G(x,y,z) = g(x,y,z) - f(x,y,z),$$

qui vérifie notamment $G = 0$ si et seulement si $\psi = \iota$.

LEMME 3.3. — *Soit L une sous-variété legendrienne de $J^1\mathbf{R}^n$ ayant une fonction quadratrice $S\colon \mathbf{R}^n \times \mathbf{R}^d \to \mathbf{R}$. Soit d'autre part $\psi_t \in \mathscr{G}(J^1\mathbf{R}^n)$, $t \in [0,1]$, une isotopie de contact $\mathscr{C}^1$-petite, G_t les fonctions de transition associées et X_t le champ de vecteurs qui l'engendre.*

(a) *Chaque sous-variété $\psi_t(L)$ a pour fonction quadratrice la fonction*

$$S_t\colon \mathbf{R}^n \times \mathbf{R}^d \times \mathbf{R}^n \times \mathbf{R}^n \longrightarrow \mathbf{R}, \quad S_t(x,u,v,w) = G_t\big(x,w,S(v,u) - vw\big).$$

(b) *Pour tout $(x,y,z) \in J^1\mathbf{R}^n$,*

$$\frac{d}{dt} G_t(x_t,y,z) = h_t(x_t,y_t,z_t')$$

où $(x_t,y_t,z_t') = \psi_t(x,y,z+xy)$ et $h_t = \alpha(X_t)$ est le hamiltonien de contact vu comme fonction sur $J^1\mathbf{R}^n$ grâce à la forme de contact canonique $\alpha = dz - y\,dx$.

Noter que la fonction quadratrice S_0 fournie par le (a) du lemme ne coïncide pas avec S. Elle s'écrit $S_0(x,u,v,w) = S(v,u) + w(x-v)$ et doit être vue comme une stabilisation de S.

On renvoie à [**4**] pour la preuve de ce lemme et on explique comment en déduire des fonctions quadratrices pour les images L_t de la section nulle par une isotopie de contact $\psi_t \in \mathscr{G}(J^1\mathbf{R}^n)$, $t \in [0,1]$. On prend un entier k assez grand pour que toutes les transformations

$$\psi_t \circ \psi_{i/k}^{-1}, \qquad t \in [i/k, (i+1)/k],\ 0 \leq i \leq k-1,$$

soient suffisamment $\mathscr{C}^1$-proches de l'identité pour définir des fonctions de transition G_t. Pour $t \in [0,1/k]$, le lemme s'applique avec $S=0$ et fournit une fonction quadratrice S_t^1 de L_t variant continûment avec t. Pour $t \in [1/k, 2/k]$, le lemme s'applique avec $S = S_{1/k}^1$ et donne une fonction quadratrice S_t^2 de L_t. De plus, le chemin S_t^2, $t \in [1/k, 2/k]$, se raccorde au chemin S_t^2, $t \in [0,1/k]$, des stabilisations des fonctions S_t^1. En répétant le procédé, on obtient un chemin $S_t = S_t^k$ de fonctions quadratrices pour toutes les sous-variétés L_t, $t \in [0,1]$.

La proposition suivante servira plus loin :

PROPOSITION 3.4 ([**24, 29**]). — *Soit $L_0 \subset J^1\mathbf{R}^n$ une sous-variété legendrienne ayant une fonction quadratrice $S\colon \mathbf{R}^n \times \mathbf{R}^d \to \mathbf{R}$ et $\psi_t \in \mathscr{G}(J^1\mathbf{R}^n)$, $t \in [0,1]$, une isotopie positive. Alors L_0 et $L_1 = \psi_1(L_0)$ ont des fonctions quadratrices $S_0, S_1\colon E \to \mathbf{R}$ telles que $S_0 \leq S_1$.*

Démonstration. — On suppose d'abord que l'isotopie ψ_t est $\mathscr{C}^1$-petite et définit donc des fonctions de transition G_t. D'après le lemme 3.3-(b), la fonction dG_t/dt est partout positive. D'autre part, le lemme 3.3-(a) dit que $L_t = \psi_t(L_0)$ a pour fonction quadratrice $S_t(x,u,v,w) = G_t\big(x,w,S(v,u)-vw\big)$. Par suite $dS_t/dt \geq 0$ donc $S_1 \geq S_0$. Le cas d'une grande isotopie se traite par l'argument de subdivision et de recomposition décrit plus haut. □

B. Invariants des sous-variétés

L'application principale du théorème de Sikorav sur l'existence de fonctions quadratrices est une minoration par la théorie de Morse du nombre de points d'intersection entre la section nulle de T^*M et son image par toute isotopie hamiltonienne, minoration qui confirme une conjecture d'Arnold. La motivation du théorème de Viterbo sur l'unicité de ces fonctions est leur utilisation systématique pour fabriquer des invariants capables de détecter la rigidité symplectique. L'extension des constructions de Viterbo au cadre de la géométrie de contact est due à Bhupal [**2**].

Soit M une variété close orientable et $\mathscr{L} = \mathscr{G}M_0$ l'ensemble des sous-variétés legendriennes de J^1M images de la section nulle M_0 par une transformation de contact de

$\mathscr{G} = \mathscr{G}(J^1M)$. D'après le théorème 3.2, toute sous-variété $L \in \mathscr{L}$ admet une fonction quadratrice $S\colon E \to \mathbf{R}$ et l'ensemble des valeurs critiques de S ne dépend que de L : c'est le spectre de L. On explique ci-dessous comment Viterbo associe un point de ce spectre à toute classe de cohomologie $\nu \in \mathrm{H}^*(M)$.

Soit $L \in \mathscr{L}$ et $S\colon E \to \mathbf{R}$ une fonction quadratrice de L. On note Q la forme quadratique sur E telle que $S - Q$ soit constante à l'infini et E_- un sous-fibré vectoriel maximal sur lequel Q est définie négative, donc de rang égal à l'indice i de Q. On note aussi E^a, $a \in \mathbf{R}$, le sous-niveau $\{S \leq a\}$ dans E et $E^{-\infty}$ un sous-niveau E^a pour $a < 0$ très grand (en dessous d'un certain seuil, le type topologique ne change plus). La cohomologie en vigueur ci-après est à coefficients dans un corps. Par excision,

$$\mathrm{H}^*\big(E, E^{-\infty}\big) \cong \mathrm{H}^*\big(\mathrm{B}(E_-), \partial\mathrm{B}(E_-)\big)$$

où $\mathrm{B}(E_-)$ est le fibré en boules de E_-. D'autre part, le rappel par la projection $\pi\colon E_- \to M$ et le *cup*-produit avec la classe de Thom $\tau \in \mathrm{H}^i\big(\mathrm{B}(E_-), \partial\mathrm{B}(E_-)\big)$ (classe Poincaré duale de la section nulle de E_-) donnent l'isomorphisme de Thom

$$\mathrm{H}^*(M) \longrightarrow \mathrm{H}^{i+*}\big(\mathrm{B}(E_-), \partial\mathrm{B}(E_-)\big), \quad \nu \longmapsto \tau \smile \pi^*\nu.$$

Par composition, l'inclusion $E^a \to E$ induit, pour tout $a \in \mathbf{R}$, un homomorphisme

$$\theta_a\colon \mathrm{H}^*(M) \longrightarrow \mathrm{H}^*\big(\mathrm{B}(E_-), \partial\mathrm{B}(E_-)\big) \longrightarrow \mathrm{H}^*(E, E^{-\infty}) \longrightarrow \mathrm{H}^*(E^a, E^{-\infty}).$$

Pour tout $\nu \in \mathrm{H}^*(M)$, on pose alors

$$\mathrm{c}(\nu, L) = \inf\big\{a \in \mathbf{R} \mid \theta_a(\nu) \neq 0\big\}.$$

Le nombre $\mathrm{c}(\nu, L)$ est clairement un point du spectre de L si $\nu \neq 0$. En voici quelques propriétés (les arguments passent mot pour mot du cas lagrangien traité dans **[30]** au cas legendrien) :

PROPOSITION 3.5 (**[30]**). — (a) *Si $L_1, L_2 \in \mathscr{L}$ ont des fonctions quadratrices $S_1, S_2\colon E \to \mathbf{R}$ vérifiant $\|S_1 - S_2\|_{\mathscr{C}^0} \leq \varepsilon$, alors*

$$|\mathrm{c}(\nu, L_1) - \mathrm{c}(\nu, L_2)| \leq \varepsilon \quad \textit{pour tout} \quad \nu \in \mathrm{H}^*(M).$$

(b) *Pour tous $L_1, L_2 \in \mathscr{L}$ et tous $\nu_1, \nu_2 \in \mathrm{H}^*(M)$,*

$$\mathrm{c}(\nu_1 \smile \nu_2, L_1 + L_2) \geq \mathrm{c}(\nu_1, L_1) + \mathrm{c}(\nu_2, L_2).$$

(c) *Soit $1, \mu$ des générateurs respectifs de $\mathrm{H}^0(M)$ et $\mathrm{H}^n(M)$. Pour tout $L \in \mathscr{L}$,*

$$\mathrm{c}(\mu, -L) = -\mathrm{c}(1, L).$$

De plus, si L rencontre la section nulle M_0,

$$\mathrm{c}(\mu, L) - \mathrm{c}(1, L) = 0 \quad \textit{si et seulement si} \quad L = M_0.$$

Démonstration. — On renvoie à **[30]** pour (b) et (c) mais on explique le point facile (a) qui servira ci-dessous. Pour $j \in \{1,2\}$ et $a \in \mathbf{R}$, soit E_j^a le sous-niveau $\{S_j \leq a\}$ et $\theta_{j,a}$ l'homomorphisme $\mathrm{H}^*(M) \to \mathrm{H}^{i_j+*}(E_j^a, E_j^{-\infty})$ associé à S_j. L'hypothèse $\|S_1 - S_2\|_{\mathscr{C}^0} \leq \varepsilon$ garantit que

$$E_2^{a-\varepsilon} \subset E_1^a \subset E_2^{a+\varepsilon}$$

et ces sous-niveaux ont le même type topologique si $|a|$ est assez grand. Pour $a < \mathrm{c}(\nu, L_1)$, la classe $\theta_{1,a}(\nu)$ est nulle donc aussi la classe $\theta_{2,a-\varepsilon}(\nu)$ à cause de l'inclusion ci-dessus. Par suite, $\mathrm{c}(\nu, L_1) - \mathrm{c}(\nu, L_2) \leq \varepsilon$. De même, $\mathrm{c}(\nu, L_2) - \mathrm{c}(\nu, L_1) \leq \varepsilon$, d'où le résultat. □

Dans la proposition ci-dessous, $\mathbf{R}$ — donc aussi $\mathbf{Z} \subset \mathbf{R}$ — opère sur J^1M par les translations (de contact) $(x, y, z) \mapsto (x, y, z + c)$, $c \in \mathbf{R}$. On considère ainsi, en plus du groupe $\mathscr{G} = \mathscr{G}(J^1M)$, les deux groupes suivants :

- $\mathscr{G}' = \mathscr{G}'(J^1M)$ est la composante neutre du groupe des transformations de contact $\mathbf{Z}$-équivariantes dont le support se projette sur un compact de T^*M ;
- $\mathscr{H} = \mathscr{H}(J^1M)$ est la composante neutre du groupe des transformations de contact $\mathbf{R}$-équivariantes dont le support se projette sur un compact de T^*M — ces transformations sont les relèvements des transformations hamiltoniennes de T^*M à support compact.

Proposition 3.6 (**[2, 24, 30]**). — *Soit $\nu \in \mathrm{H}^*(M)$ une classe non nulle, $L \in \mathscr{L}$ une sous-variété legendrienne et ψ une transformation de contact. Alors*

$$\mathrm{c}\big(\nu, \psi(L)\big) = \mathrm{c}\big(\nu, L - \psi^{-1}(M_0)\big)$$

dans les trois cas suivants :

1. *$\psi \in \mathscr{G}$ et $\mathrm{c}(\nu, \psi(L)) = 0$;*
2. *$\psi \in \mathscr{G}'$ et $\mathrm{c}(\nu, \psi(L)) \in \mathbf{Z}$;*
3. *$\psi \in \mathscr{H}$.*

Démonstration. — On relie $\psi = \psi_1$ à l'identité $\iota = \psi_0$ par un chemin ψ_t dans $\mathscr{G}$, $\mathscr{G}'$ ou $\mathscr{H}$ selon le cas. On pose

$$c_t = \mathrm{c}(\nu, L_t) \quad \text{où} \quad L_t = \psi_t^{-1}\big(\psi(L)\big) - \psi_t^{-1}(M_0)$$

et on note $S_t \colon E \to \mathbf{R}$ un chemin de fonctions quadratrices pour les L_t, $t \in [0,1]$. La proposition 3.5-(a) assure que la valeur critique c_t de S_t varie continûment. Par ailleurs,

$$L_0 = \psi(L) - M_0 = \psi(L) \quad \text{et} \quad L_1 = L - \psi^{-1}(M_0)$$

donc il suffit de montrer que le chemin c_t est forcément constant dans les trois cas envisagés.

L'affirmation clé est que, dans ces trois cas, ψ_t^{-1} induit une bijection entre les points critiques de S_0 de valeur c_0 et ceux de S_t de même valeur c_0. Pour le vérifier, on observe d'abord que les points critiques de S_t de valeur $c \in \mathbf{R}$ paramètrent $L_t \cap M_c$ où $M_c = \tau_c(M_0)$ et $\tau_c \colon J^1M \to J^1M$ est la translation $(x, y, z) \mapsto (x, y, z + c)$. On distingue ensuite les trois cas.

Cas 1. L'ensemble $L_t \cap M_0$ est en bijection évidente avec

$$\psi_t^{-1}\big(\psi(L)\big) \cap \psi_t^{-1}(M_0) = \psi_t^{-1}\big(\psi(L) \cap M_0\big) = \psi_t^{-1}(L_0 \cap M_0).$$

Cas 2. Comme τ_k commute avec ψ_t, l'ensemble $L_t \cap M_k$ est en bijection avec

$$\psi_t^{-1}\big(\psi(L)\big) \cap \tau_k\big(\psi_t^{-1}(M_0)\big) = \psi_t^{-1}\big(\psi(L) \cap \tau_k(M_0)\big) = \psi_t^{-1}(L_0 \cap M_k).$$

Cas 3. Comme τ_c commute avec ψ_t, l'ensemble $L_t \cap M_c$ est en bijection avec

$$\psi_t^{-1}\big(\psi(L)\big) \cap \tau_c\big(\psi_t^{-1}(M_0)\big) = \psi_t^{-1}\big(\psi(L) \cap \tau_c(M_0)\big) = \psi_t^{-1}(L_0 \cap M_c).$$

Pour conclure, on suppose d'abord que les points critiques de S_0 sont non dégénérés, c'est-à-dire que $\psi(L) = L_0$ est transversale à $M_0 \times \mathbf{R} \subset J^1M$. Il est facile de voir que les sous-variétés L_t (legendriennes) sont alors toutes transversales à $M_0 \times \mathbf{R}$ de sorte que les points critiques des fonctions S_t sont eux aussi non dégénérés. Ils forment donc des arcs paramétrés par $t \in [0, 1]$ et seuls ceux qui restent constamment à l'altitude c_0 passent par l'altitude c_0. Dans le cas général, la proposition 3.5-(a) permet de se ramener au cas particulier par perturbation. □

COROLLAIRE 3.7. — *Soit $\nu \in \mathrm{H}^*(M)$ une classe non nulle, $L \in \mathscr{L}$ une sous-variété legendrienne et ψ une transformation de contact. Les nombres*

$$\mathrm{c}\big(\nu, \psi(L)\big) \quad \textit{et} \quad \mathrm{c}\big(\nu, L - \psi^{-1}(M_0)\big)$$

ont le même signe si $\psi \in \mathscr{G}$ et la même partie entière si $\psi \in \mathscr{G}'$.

C. Invariants des transformations

On décrit ici les invariants associés aux transformations hamiltoniennes de $\mathbf{R}^{2n}$ par Viterbo [**30**] et étendus aux transformations de contact de $\mathbf{R}^{2n+1}$ et $\mathbf{R}^{2n} \times \mathbf{T}$ respectivement par Bhupal [**2**] et Sandon [**24**]. Pour se placer dans un cadre unique, on tient compte des actions de $\mathbf{R}$ et $\mathbf{Z}$ sur $\mathbf{R}^{2n+1} = J^1\mathbf{R}^n$ par les translations (entières dans le cas de $\mathbf{Z}$) sur le dernier facteur. On travaille alors avec les trois groupes $\mathscr{G}$, $\mathscr{G}'$ et $\mathscr{H}$ intervenant déjà dans la proposition 3.6. On a d'autre part besoin de quelques constructions qu'on présente tout de suite.

Étant donné des variétés F et K munies respectivement d'une structure symplectique ω et d'une structure de contact η, on note $\overline{F}$ et $\overline{K}$ les mêmes variétés munies des

structures $-\omega$ et $-\eta$ (le même champ d'hyperplans avec la coorientation opposée). Toute transformation symplectique $\phi\colon F \to F$ définit un plongement lagrangien

$$G\phi\colon F \longrightarrow \overline{F} \times F, \quad p \longmapsto \big(p, \phi(p)\big),$$

donc toute transformation de contact $\psi\colon K \to K$ détermine le plongement lagrangien

$$GS\psi\colon SK \longrightarrow \overline{S}K \times SK, \quad p \longmapsto \big(p, S\psi(p)\big).$$

Ce plongement est $\mathbf{R}_+^*$-équivariant pour les actions normale à gauche et diagonale à droite et le quotient de $\overline{S}K \times SK$ par cette dernière action est une variété de contact (à regarder comme le produit de contact de $\overline{K}$ et K) qu'on notera $\overline{K} \boxtimes K$. Par passage au quotient, ψ définit un plongement legendrien de K dans $\overline{K} \boxtimes K$ qu'on désigne par $G\psi$.

D'autre part, il existe un difféomorphisme symplectique $\delta\colon \overline{\mathbf{R}}^{2n} \times \mathbf{R}^{2n} \to T^*\mathbf{R}^{2n}$ qui envoie la diagonale sur la section nulle. On peut prendre par exemple

$$\delta(x', y', x'', y'') = ((x' + x'')/2, (y' + y'')/2, y' - y'', x'' - x').$$

Noter que δ conjugue l'involution $(x', y', x'', y'') \mapsto (x'', y'', x', y')$ à la symétrie $(x, y) \mapsto (x, -y)$.

On laisse au lecteur le soin de construire l'analogue de δ pour les 1-jets, à savoir un plongement de contact $\widehat{\delta}\colon \overline{\mathbf{R}}^{2n+1} \boxtimes \mathbf{R}^{2n+1} \to J^1\mathbf{R}^{2n+1}$ ayant les propriétés suivantes :

- $\widehat{\delta}$ envoie la « diagonale » sur la section nulle ;
- $\widehat{\delta}$ est équivariant pour les actions de $\mathbf{Z}$ respectivement diagonale à gauche et horizontale à droite (autrement dit relevée de la base) ;
- $\widehat{\delta}$ est équivariant pour les actions de $\mathbf{R}$ respectivement verticale à gauche (c'est-à-dire induite par l'action sur le second facteur) et normale à droite.

Pour $n = 0$ — et $J^1\mathbf{R}^0 = \mathbf{R}$ —, un exemple de tel plongement (sous sa forme symplectisée) est

$$S\widehat{\delta}(z', s', z'', s'') = (z', s'/s'' - 1, z'' - z', s'').$$

Les transformations symplectiques $\phi\colon \mathbf{R}^{2n} \to \mathbf{R}^{2n}$ et de contact $\psi\colon \mathbf{R}^{2n+1} \to \mathbf{R}^{2n+1}$ définissent ainsi des plongements respectivement lagrangiens et legendriens

$$L\phi = \delta \circ G\phi\colon \mathbf{R}^{2n} \longrightarrow T^*\mathbf{R}^{2n}$$
$$\text{et} \quad L\psi = \widehat{\delta} \circ G\psi\colon \mathbf{R}^{2n+1} \longrightarrow J^1\mathbf{R}^{2n+1}.$$

Les conditions de support permettent en outre de compactifier ces plongements :

- en un plongement lagrangien exact $L\phi\colon \mathbf{S}^{2n} \to T^*\mathbf{S}^{2n}$ si $\phi \in \mathscr{H}$, plongement qu'on peut aussi voir comme projection d'un plongement legendrien $L\phi\colon \mathbf{S}^{2n} \to J^1\mathbf{S}^{2n}$;
- en un plongement legendrien $L\psi\colon \mathbf{S}^{2n+1} \to J^1\mathbf{S}^{2n+1}$ si $\psi \in \mathscr{G}$;

– en un plongement legendrien $L\psi\colon \mathbf{S}^{2n}\times\mathbf{T}\to J^1(\mathbf{S}^{2n}\times\mathbf{T})$ si $\psi\in\mathscr{G}'$.

Dans la suite, $L\phi$ et $L\psi$ désignent aussi bien ces plongements que leurs images. On observe que $L(\phi^{-1})=-L\phi$ et $L(\psi^{-1})=-L\psi$ (au sens des opérations dans $T^*\mathbf{R}^{2n}$ et $J^1\mathbf{R}^{2n+1}$).

Remarque 3.8. — Pour $\phi\in\mathscr{H}$, on dispose en fait de deux plongements legendriens $L\phi$ différents. D'une part, la transformation hamiltonienne de $\mathbf{R}^{2n}$ fournit un plongement lagrangien exact dans $T^*\mathbf{R}^{2n}$, sa compactification dans $T^*\mathbf{S}^{2n}$ et son relèvement legendrien dans $J^1\mathbf{S}^{2n}$. D'autre part, la transformation de contact équivariante de $\mathbf{R}^{2n+1}$ donne un élément de $\mathscr{G}'$, d'où un plongement legendrien dans $J^1(\mathbf{S}^{2n}\times\mathbf{T})$. Ces deux plongements conduisent aux mêmes invariants pour la raison suivante : si $S\colon E\to\mathbf{R}$ est une fonction quadratrice du premier, une fonction quadratrice du second est simplement la fonction $E\times\mathbf{T}\to\mathbf{R}$, $(x,u,z)\mapsto S(x,u)$. Dans la suite, c'est le second point de vue qu'on adopte parce qu'il permet de traiter ensemble les trois groupes $\mathscr{G}$, $\mathscr{G}'$ et $\mathscr{H}$.

Pour toute transformation ϕ dans $\mathscr{G}$, $\mathscr{G}'$ ou $\mathscr{H}$, on pose

$$\mathrm{c}(\phi)=\mathrm{c}(\mu,L\phi)$$

où μ est la classe fondamentale de $\mathbf{S}^{2n+1}$, $\mathbf{S}^{2n}\times\mathbf{T}$ (ou même éventuellement $\mathbf{S}^{2n}$), selon le cas.

Ce nombre est donc une valeur critique d'une quelconque fonction quadratrice S de $L\phi$. Il est utile d'avoir une interprétation géométrique des points et des valeurs critiques de S. Vu l'équivariance de $\widehat{\delta}$ par rapport aux actions verticale et normale de $\mathbf{R}$, les points critiques paramètrent les points q de $\mathbf{R}^{2n+1}$ que ϕ translate parallèlement à l'axe des z avec une dérivée $\partial_z\phi(q)$ égale à 1. La quantité dont q est translaté donne la valeur critique correspondante. On appellera ces points les points translatés par ϕ. Noter que, si $\phi\in\mathscr{H}$, la dérivée $\partial_z\phi$ vaut partout 1.

La proposition qui suit résume les propriétés principales de $\mathrm{c}(\phi)$.

PROPOSITION 3.9 ([**2, 24, 30**]). — *Pour tous ϕ et ψ dans $\mathscr{G}$, dans $\mathscr{G}'$ ou dans $\mathscr{H}$:*

(a) $\mathrm{c}(\phi)\geq 0$.

(b) $\mathrm{c}(\phi)=\mathrm{c}(\phi^{-1})=0$ *si et seulement si* $\phi=\iota$.

(c) $\mathrm{c}(\psi\circ\phi)\leq\mathrm{c}(\phi)+\mathrm{c}(\psi)$ *dans les cas suivants :*

- $\phi,\psi\in\mathscr{G}$ *et* $\mathrm{c}(\phi)=0$ *;*
- $\phi,\psi\in\mathscr{G}'$ *et* $\mathrm{c}(\phi)\in\mathbf{Z}$ *;*
- $\phi,\psi\in\mathscr{H}$.

(d) $\mathrm{c}(\psi\circ\phi\circ\psi^{-1}) = \mathrm{c}(\phi)$ *dans les trois mêmes cas que ci-dessus.*

(e) $\mathrm{c}(\psi) \leq \mathrm{c}(\phi)$ *si on peut aller de* ψ *à* ϕ *par un chemin positif — dans* $\mathscr{G}$, $\mathscr{G}'$ *ou* $\mathscr{H}$ *selon le cas.*

Démonstration. — (a) On traite le cas de $\mathscr{G}$ mais les autres sont identiques aux notations près. D'après la proposition 3.5-(c),

$$\mathrm{c}(\phi) = \mathrm{c}(\mu, L\phi) = -\mathrm{c}(1, -L\phi)$$

donc il revient au même de montrer que $\mathrm{c}(1, -L\phi) \leq 0$.

Soit $S\colon E \to \mathbf{R}$ une fonction quadratrice de $-L\phi$, où E est un fibré sur $\mathbf{S}^{2n+1} = \mathbf{R}^{2n+1} \cup \{q\}$. On note E_q la fibre au-dessus de q et on contemple le diagramme commutatif

$$\begin{array}{ccc} \mathrm{H}^*(\mathbf{S}^{2n}) & \longrightarrow & \mathrm{H}^*(\{q\}) \\ {\scriptstyle\theta_0}\big\downarrow & & \big\downarrow{\scriptstyle\cong} \\ \mathrm{H}^*(E^0, E^{-\infty}) & \longrightarrow & \mathrm{H}^*(E_q^0, E_o^{-\infty}) \end{array}$$

dans lequel les flèches horizontales proviennent des inclusions. Comme q est hors du support de ϕ, la sous-variété $-L\phi$ coïncide avec la section nulle au-dessus d'un voisinage de q et $S\,|_{E_q}$ est une forme quadratique non dégénérée. Par suite, la flèche verticale de droite est un isomorphisme. Comme la flèche horizontale du haut envoie 1 sur 1, la classe $\theta_0(1)$ est non nulle donc $\mathrm{c}(1, -L\phi) \leq 0$.

(b) Comme $L(\phi^{-1}) = -L\phi$, la propriété résulte directement de la proposition 3.5-(c).

(c) Dans les trois cas, la proposition 3.6 dit que, pour toute classe ν non nulle,

$$\mathrm{c}(\nu, L\phi) = \mathrm{c}(\nu, L(\psi\circ\phi) - L\psi).$$

Par suite,

$$\begin{aligned} \mathrm{c}(\phi) = \mathrm{c}(\mu, L\phi) &= \mathrm{c}(\mu \smile 1, L(\psi\circ\phi) - L\psi) \geq \mathrm{c}(\mu, L(\psi\circ\phi)) + \mathrm{c}(1, -L\psi) \\ &= \mathrm{c}(\mu, L(\psi\circ\phi)) - \mathrm{c}(\mu, L\psi) = \mathrm{c}(\psi\circ\phi) - \mathrm{c}(\phi) \end{aligned}$$

d'où le résultat.

(d) L'argument est le même que pour démontrer la proposition 3.6.

(e) Cette propriété découle directement de la proposition 3.4. □

Une conséquence des points (c) et (d) de cette proposition (voir le corollaire 3.7) est :

Corollaire 3.10 ([**24**]). — *Pour tous* $\phi, \psi \in \mathscr{G}'$,

$$\lfloor \mathrm{c}(\psi\circ\phi\circ\psi^{-1}) \rfloor = \lfloor \mathrm{c}(\phi) \rfloor$$

$$\textit{et} \qquad \lfloor \mathrm{c}(\psi\circ\phi) \rfloor \leq \lfloor \mathrm{c}(\phi) \rfloor + \lfloor \mathrm{c}(\psi) \rfloor + 1.$$

Une autre conséquence immédiate de la proposition ci-dessus est :

COROLLAIRE 3.11 ([**2, 30**]). — *Les groupes $\mathscr{G}$, $\mathscr{G}'$ et $\mathscr{H}$ ne contiennent aucun lacet positif non constant.*

Démonstration. — Sur un lacet positif (basé en ι), toute transformation ϕ est, de même que son inverse ϕ^{-1}, accessible depuis l'identité par un chemin négatif. Par suite, $\mathrm{c}(\phi) \leq 0$ et $\mathrm{c}(\phi^{-1}) \leq 0$. Comme ce sont des nombres positifs, $\mathrm{c}(\phi) = \mathrm{c}(\phi^{-1}) = 0$ et $\phi = \iota$. □

Ce corollaire assure que la relation $\preceq$ induit des ordres partiels sur $\mathscr{G}$, $\mathscr{G}'$ et $\mathscr{H}$. En fait, comme Viterbo et Bhupal l'ont vu, la proposition 3.9 permet de vérifier que la relation

$$\phi_0 \leqslant \phi_1 \quad \text{si et seulement si} \quad \mathrm{c}(\phi_0 \circ \phi_1^{-1}) = 0$$

définit des ordres partiels invariants (au sens de 1.C) sur $\mathscr{G}$, $\mathscr{G}'$ et $\mathscr{H}$.

On définit maintenant les groupes d'homologie associés par L. Traynor [**29**] aux transformations de $\mathscr{H}$ et par Sandon aux transformations de $\mathscr{G}'$.

Pour toute transformation ϕ dans $\mathscr{H}$ ou dans $\mathscr{G}'$, on note $\Sigma\phi$ l'ensemble des valeurs critiques des fonctions quadratrices de $L\phi$. On choisit une telle fonction quadratrice $S\colon E \to \mathbf{R}$, on note i l'indice de sa partie quadratique (à l'infini) et, pour $a < b \in]-\infty, \infty] \setminus \Sigma\phi$, on pose

$$\mathrm{SGH}^k(\phi; a, b) = \mathrm{H}_{k+i}(E^b, E^a) \qquad \text{si} \quad \phi \in \mathscr{H}$$
$$\text{et} \quad \mathrm{CGH}^k(\phi; a, b) = \mathrm{H}_{k+i}(E^b, E^a) \qquad \text{si} \quad \phi \in \mathscr{G}'.$$

Il résulte du théorème fondamental des fonctions quadratrices que ces groupes sont définis et ne dépendent pas du choix de S.

Le lecteur attentif pourra se demander pourquoi on introduit deux notations alors que $\mathscr{H}$ est un sous-groupe de $\mathscr{G}'$ et que, d'après la remarque 3.8, les deux interprétations qu'on peut faire de la définition ci-dessus pour $\phi \in \mathscr{H}$ donnent le même résultat. La raison est dans la proposition suivante :

PROPOSITION 3.12 ([**24, 29**]). — *Soit $a < b \in]-\infty, \infty] \setminus \Sigma\phi$.*

(a) *Si $\phi \in \mathscr{H}$, toute transformation $\psi \in \mathscr{H}$ induit un isomorphisme*

$$\psi^*\colon \mathrm{SGH}^*(\psi \circ \phi \circ \psi^{-1}; a, b) \longrightarrow \mathrm{SGH}^*(\phi; a, b).$$

(b) *Si $\phi \in \mathscr{G}'$ et si a et b sont entiers, toute transformation $\psi \in \mathscr{G}'$ induit un isomorphisme*

$$\psi^*\colon \mathrm{CGH}^*(\psi \circ \phi \circ \psi^{-1}; a, b) \longrightarrow \mathrm{CGH}^*(\phi; a, b).$$

Démonstration. — Soit ψ_t, $t \in [0,1]$, un chemin dans $\mathscr{H}$ ou $\mathscr{G}'$ selon le cas entre $\psi_0 = \iota$ et $\psi = \psi_1$. On pose $\phi_t = \psi_t \circ \phi \circ \psi_t^{-1}$ et on se donne un chemin de fonctions quadratrices $S_t \colon E \to \mathbf{R}$ pour les plongements legendriens $L\phi_t$.

(a) Comme ψ_t est $\mathbf{R}$-équivariante, elle envoie les points translatés par ϕ sur les points translatés de la même quantité par ϕ_t donc le spectre $\Sigma\phi_t$ ne varie pas. D'autre part, comme les niveaux $\{S_t = a\}$ et $\{S_t = b\}$ sont réguliers à tout instant et ne changent pas hors d'un compact, ils se déforment par isotopie et toute isotopie ambiante de E qui prolonge celle de ces niveaux induit le même isomorphisme

$$\mathrm{SGH}^*(\phi; a, b) = \mathrm{H}_{*+i}(E_0^b, E_0^a) \longrightarrow \mathrm{H}_{*+i}(E_1^b, E_1^a) = \mathrm{SGH}^*(\psi \circ \phi \circ \psi^{-1}; a, b),$$

à savoir l'inverse de l'isomorphisme ψ^* cherché.

(b) L'observation clé est la suivante : si $a \in \mathbf{Z}$ est une valeur régulière entière de $S = S_0$, c'est une valeur régulière de S_t pour tout t.

On suppose que a est valeur critique de S_{t_0} pour un certain $t_0 > 0$. Comme ϕ et ψ_t sont $\mathbf{Z}$-équivariants, ψ_t envoie tout point translaté d'un entier k par ϕ sur un point translaté de k par ϕ_t. La valeur critique a de $S - t_0$ détecte un point $\psi_{t_0}(q)$ translaté de a par ϕ_{t_0} et $\psi_t(q)$ est donc, pour tout t, un point translaté de a par ϕ_t. En particulier, q est translaté de a par ϕ donc a est valeur critique de S_0. □

D. Invariants des domaines

Le premier invariant qu'on décrit est la capacité de Viterbo **[30]** pour les domaines de $\mathbf{R}^{2n}$ et sa généralisation due à Sandon **[24]** pour les domaines de $\mathbf{R}^{2n+1} \times \mathbf{T}$.

Pour tout ouvert $U \subset \mathbf{R}^{2n}$, on note $\mathscr{H}(U)$ la composante neutre du sous-groupe de $\mathscr{H}$ formé des transformations dont le support se projette dans U. Les chemins dans $\mathscr{H}(U)$ relèvent ainsi les isotopies de $\mathbf{R}^{2n}$ engendrées par les hamiltoniens à support compact dans U. On pose alors

$$\mathrm{c}_v(U) = \sup\bigl\{\mathrm{c}(\phi),\ \phi \in \mathscr{H}(U)\bigr\}.$$

De même, pour tout ouvert $U \subset \mathbf{R}^{2n} \times \mathbf{T}$, on note $\mathscr{G}'(U)$ la composante neutre du sous-groupe de $\mathscr{G}'$ formé des transformations dont le support se projette dans U et on pose

$$\mathrm{c}_s(U) = \sup\bigl\{\mathrm{c}(\phi),\ \phi \in \mathscr{G}'(U)\bigr\}.$$

Le lemme suivant montre que ces quantités sont finies sur les ouverts bornés :

Lemme 3.13. — (a) *Soit $U \subset \mathbf{R}^{2n}$ un ouvert borné, $\phi \in \mathscr{H}(U)$ et $\psi \in \mathscr{H}$. Si $\psi(U)$ est disjoint de U,*

$$\mathrm{c}(\phi) \leq \mathrm{c}(\psi) + \mathrm{c}(\psi^{-1}).$$

(b) *Soit $U \subset \mathbf{R}^{2n} \times \mathbf{T}$ un ouvert borné, $\phi \in \mathscr{G}'(U)$ et $\psi \in \mathscr{G}'$. Si $\psi(U)$ est disjoint de U,*

$$\lfloor \mathrm{c}(\phi) \rfloor \leq \lfloor \mathrm{c}(\psi) \rfloor + \lfloor \mathrm{c}(\psi) \rfloor + 1.$$

Dans ce qui suit, on identifie parfois subrepticement U à sa préimage dans $\mathbf{R}^{2n+1}$.

Démonstration. — (a) Comme, d'après la proposition 3.9-(c),

$$\mathrm{c}(\phi) \leq \mathrm{c}(\psi \circ \phi) + \mathrm{c}(\psi^{-1}),$$

il suffit de montrer que $\mathrm{c}(\psi \circ \phi) = \mathrm{c}(\psi)$.

Soit $\phi_t \in \mathscr{H}(U)$ un chemin de $\phi_0 = \iota$ à $\phi_1 = \phi$ et q_t un point translaté par $\psi \circ \phi_t$ d'une quantité égale à la valeur critique $c_t = \mathrm{c}(\psi \circ \phi_t)$. Comme $\psi \circ \phi_t(U) \subset \psi(U)$ est disjoint de U, le point q_t n'est pas dans U. Par suite, q_t est un point fixe de ϕ_t (hors support) donc un point translaté par ψ. Il en résulte que c_t est, pour tout t, valeur critique des fonctions quadratrices de $L\psi$. Comme cet ensemble est totalement discontinu et que c_t varie continûment, $\mathrm{c}(\psi \circ \phi) = c_1 = c_0 = \mathrm{c}(\psi)$.

(b) Comme, d'après le corollaire 3.10,

$$\lfloor \mathrm{c}(\phi) \rfloor \leq \lfloor \mathrm{c}(\psi \circ \phi) \rfloor + \lfloor \mathrm{c}(\psi^{-1}) \rfloor,$$

il suffit de montrer encore que $\mathrm{c}(\psi \circ \phi) = \mathrm{c}(\psi)$ et l'argument est le même que plus haut. □

On admettra le résultat suivant (dont les deux premiers points sont évidents) qui dit que la fonction c_v définit une capacité au sens de Ekeland et Hofer pour les ouverts de $\mathbf{R}^{2n}$:

THÉORÈME 3.14 ([**30**]). — *La fonction c_v sur les ouverts de $\mathbf{R}^{2n}$ a les propriétés suivantes :*

- $\mathrm{c}_v(U) \leq \mathrm{c}_v(V)$ *si* $U \subset V$;
- $\mathrm{c}_v\big(\psi(U)\big) = \lambda \mathrm{c}_v(U)$ *pour toute transformation ψ de $\mathbf{R}^{2n}$ qui dilate la forme symplectique par λ ;*
- $\mathrm{c}_v\big(B^{2n}(1)\big) = \mathrm{c}_v\big(C^{2n}(1)\big) = 1$.

Le pendant de ce théorème pour les ouverts de $\mathbf{R}^{2n} \times \mathbf{T}$ est :

THÉORÈME 3.15 ([**24**]). — *La fonction c_s sur les ouverts de $\mathbf{R}^{2n} \times \mathbf{T}$ a les propriétés suivantes :*

- $\mathrm{c}_s(U) \leq \mathrm{c}_s(V)$ *si* $U \subset V$;
- $\mathrm{c}_s\big(\psi(U)\big) = \mathrm{c}_s(U)$ *pour toute transformation de contact ψ de $\mathbf{R}^{2n} \times \mathbf{T}$;*
- $\mathrm{c}_s(U \times \mathbf{T}) = \lfloor \mathrm{c}_v(U) \rfloor$ *pour tout ouvert U de $\mathbf{R}^{2n}$.*

Démonstration. — La première propriété est évidente et la deuxième découle du corollaire 3.10. Quant à la troisième, elle résulte d'une part de l'inclusion de $\mathscr{H}(U)$ dans $\mathscr{G}'(U \times \mathbf{T})$ et d'autre part du point (e) de la proposition 3.9. □

Le théorème de Sandon ci-dessus implique immédiatement le théorème 1.2-(b) :

COROLLAIRE 3.16 ([**12, 24**]). — *Si $r_2 \leq k \leq r_1$ pour un certain entier k, aucune transformation de contact de $\mathbf{R}^{2n} \times \mathbf{T}$ n'envoie* $\operatorname{Adh} B(r_1) \times \mathbf{T}$ *dans $B(r_2) \times \mathbf{T}$.*

Pour finir, on explique comment Traynor [**29**] et Sandon [**24**] associent des groupes de cohomologie pour les domaines respectivement de $\mathbf{R}^{2n}$ et $\mathbf{R}^{2n} \times \mathbf{T}$.

Pour tout domaine $U \subset \mathbf{R}^{2n}$ et tous $a < b \in \mathbf{R} \cup \{\infty\}$, soit $\mathscr{H}(U; a, b)$ le sous-ensemble de $\mathscr{H}(U)$ composé des transformations dont les fonctions quadratrices ont a et b comme valeurs régulières. Pour $\phi_0, \phi_1 \in \mathscr{H}(U; a, b)$ tels que $\phi_0 \preceq \phi_1$, la proposition 3.4 montre que les plongements legendriens $L\phi_0, L\phi_1$ ont des fonctions quadratrices respectives $S_0, S_1 \colon E \to \mathbf{R}$ vérifiant $S_0 \leq S_1$. Il en résulte des inclusions entre les sous-niveaux — à savoir que $E_1^a \subset E_0^a$ et $E_1^b \subset E_0^b$ — et un homomorphisme

$$\mathrm{SGH}^*(\phi_1; a, b) \longrightarrow \mathrm{SGH}^*(\phi_0; a, b).$$

Les homomorphismes de ce type ont une fonctorialité manifeste qui fait des groupes $\mathrm{SGH}^*(\phi; a, b)$, $\phi \in \big(\mathscr{H}(U; a, b), \preceq\big)$ un système projectif. On définit alors $\mathrm{SGH}^*(U; a, b)$ comme la limite projective de ce système :

$$\mathrm{SGH}^*(U; a, b) = \varprojlim_{\phi} \mathrm{SGH}^*(\phi; a, b).$$

Clairement, toute transformation hamiltonienne ψ de $\mathbf{R}^{2n}$ induit un isomorphisme

$$\psi^* \colon \mathrm{SGH}^*(\psi(U); a, b) \longrightarrow \mathrm{SGH}^*(U; a, b),$$

tout sous-domaine $U' \subset U$ hérite d'un homomorphisme

$$\mathrm{SGH}^*(U; a, b) \longrightarrow \mathrm{SGH}^*(U'; a, b)$$

et tous ces morphismes se comportent fonctoriellement.

Pour tout domaine $U \subset \mathbf{R}^{2n} \times \mathbf{T}$ et tous $a < b \in \mathbf{Z} \cup \{\infty\}$, le même discours fournit un système projectif de groupes $\mathrm{CGH}^*(\phi; a, b)$, $\phi \in \big(\mathscr{G}(U; a, b), \preceq\big)$, et un groupe

$$\mathrm{CGH}^*(U; a, b) = \varprojlim_{\phi} \mathrm{CGH}^*(\phi; a, b).$$

Toute transformation ψ de $\mathbf{R}^{2n} \times \mathbf{T}$ qui est le temps 1 d'une isotopie de contact induit un isomorphisme

$$\psi^* \colon \mathrm{CGH}^*(\psi(U); a, b) \longrightarrow \mathrm{CGH}^*(U; a, b),$$

tout sous-domaine $U' \subset U$ hérite d'un homomorphisme

$$\mathrm{CGH}^*(U; a, b) \longrightarrow \mathrm{CGH}^*(U'; a, b)$$

et ces morphismes se comportent à nouveau fonctoriellement. En outre :

PROPOSITION 3.17 ([**24**]). — *Soit $a < b \in \mathbf{Z} \cup \{\infty\}$. Pour tout domaine U de $\mathbf{R}^{2n}$, il y a un isomorphisme*

$$\mathrm{CGH}(U \times \mathbf{T}; a, b) \longrightarrow \mathrm{SGH}^*(U; a, b)$$

qui se comporte fonctoriellement vis-à-vis des morphismes induits par les inclusions et les transformations hamiltoniennes.

Exemple 3.18 ([**29**]). — Traynor a calculé les groupes SGH^* à coefficients dans $\mathbf{Z}_2$ pour les ellipsoïdes de $\mathbf{R}^{2n}$. Pour $a > 0$ et $j = 2mn$, $m \geq 1$, elle obtient par exemple :

$$\mathrm{SGH}^j(B(r); a, \infty) = \begin{cases} \mathbf{Z}_2 & \text{si } a/m < r < a/(m-1), \\ 0 & \text{sinon.} \end{cases}$$

En particulier,

$$\mathrm{SGH}^{2n}(B(r); a, \infty) = \begin{cases} \mathbf{Z}_2 & \text{si } r > a, \\ 0 & \text{sinon.} \end{cases}$$

Pour toutes les autres valeurs de j, le groupe $\mathrm{SGH}^j(B(r); a, \infty)$ est nul. En outre, pour $a/m < r_2 < r_1 \leq a/(m-1)$, l'homomorphisme

$$\mathrm{SGH}^j(B(r_1); a, \infty) \longrightarrow \mathrm{SGH}^j(B(r_2); a, \infty)$$

induit par l'inclusion $B(r_2) \subset B(r_1)$ est un isomorphisme.

Pour terminer, on démontre le théorème 1.4 en suivant [**24**].

Démonstration. — Soit $r_0, r_1, r_2 > 0$ tels que $r_2 \leq k/m < r_1 < r_0 < k/(m-1)$. On suppose qu'il existe une transformation de contact ϕ de $\mathbf{R}^{2n} \times \mathbf{T}$ qui envoie $\widehat{B}(r_1)$ dans $\widehat{B}(r_2)$ et qui préserve $\widehat{B}(r_0)$. On regarde alors le diagramme commutatif

$$\begin{array}{ccccc} \mathrm{CGH}^j(\widehat{B}(r_0); k, \infty) & \longrightarrow & \mathrm{CGH}^j(\widehat{B}(r_1); k, \infty) & & \\ \uparrow{\scriptstyle \phi^*} & & & \nwarrow{\scriptstyle \phi^*} & \\ \mathrm{CGH}^j(\widehat{B}(r_0); k, \infty) & \longrightarrow & \mathrm{CGH}^j(\widehat{B}(r_2); k, \infty) & \longrightarrow & \mathrm{CGH}^j(\phi(\widehat{B}(r_1)); k, \infty). \end{array}$$

Comme dans l'exemple ci-dessus, on regarde la cohomologie à coefficients dans $\mathbf{Z}_2$ en degré $j = 2mn$.

$$\mathrm{CGH}^j(\widehat{B}(r_2); k, \infty) = 0, \quad \mathrm{CGH}^j(\widehat{B}(r_1); k, \infty) = \mathrm{CGH}^j(\widehat{B}(r_0); k, \infty) = \mathbf{Z}_2$$

et la flèche horizontale du haut est un isomorphisme. Le diagramme donne alors la contradiction voulue. □

Remerciements

Je remercie Hélène Eynard-Bontemps, Patrick Massot et Jean-Yves Welschinger pour leur relecture attentive de ce texte. Je remercie le second aussi pour avoir réalisé la figure et m'avoir signalé promptement la parution sur arXiv du joli article de Sheila Sandon, à qui j'exprime d'ailleurs ma gratitude dans la mesure où elle a facilité ma tâche.

RÉFÉRENCES

[1] G. Ben Simon & T. Hartnick – Reconstructing quasimorphisms from the associated partial order and a question of Polterovich, prépublication arXiv:0811.2608, 2008.

[2] M. Bhupal – A partial order on the group of contactomorphisms of $\mathbb{R}^{2n+1}$ via generating functions, *Turkish J. Math.* **25** (2001), p. 125–135.

[3] M. Chaperon – Une idée du type « géodésiques brisées » pour les systèmes hamiltoniens, *C. R. Acad. Sci. Paris Sér. I Math.* **298** (1984), p. 293–296.

[4] ———, On generating families, in *The Floer memorial volume*, Progr. Math., vol. 133, Birkhäuser, 1995, p. 283–296.

[5] Y. V. Chekanov – Critical points of quasifunctions, and generating families of Legendrian manifolds, *Funktsional. Anal. i Prilozhen.* **30** (1996), p. 56–69 ; traduction anglaise : *Funct. Anal. Appl.* **30** (1996), p. 118–128.

[6] V. Chernov & S. Nemirovski – Non-negative Legendrian isotopy in ST^*M, prépublication arXiv:0905.0983.

[7] K. Cieliebak, A. Floer & H. Hofer – Symplectic homology. II. A general construction, *Math. Z.* **218** (1995), p. 103–122.

[8] V. Colin, E. Ferrand & P. Pushkar – Positive isotopies of Legendrian submanifolds, texte non publié.

[9] Y. Eliashberg – New invariants of open symplectic and contact manifolds, *J. Amer. Math. Soc.* **4** (1991), p. 513–520.

[10] ———, Contact 3-manifolds twenty years since J. Martinet's work, *Ann. Inst. Fourier (Grenoble)* **42** (1992), p. 165–192.

[11] Y. Eliashberg, A. Givental & H. Hofer – Introduction to symplectic field theory, *Geom. Funct. Anal.* Special Volume, Part II (2000), p. 560–673.

[12] Y. Eliashberg, S. S. Kim & L. Polterovich – Geometry of contact transformations and domains : orderability versus squeezing, *Geom. Topol.* **10** (2006), p. 1635–1747.

[13] Y. Eliashberg & L. Polterovich – Partially ordered groups and geometry of contact transformations, *Geom. Funct. Anal.* **10** (2000), p. 1448–1476.

[14] A. Floer & H. Hofer – Symplectic homology. I. Open sets in $\mathbf{C}^n$, *Math. Z.* **215** (1994), p. 37–88.

[15] E. Giroux – Topologie de contact en dimension 3 (autour des travaux de Yakov Eliashberg), Séminaire Bourbaki, vol. 1992/93, exposé n° 760, *Astérisque* **216** (1993), p. 7–33.

[16] M. Gromov – Pseudoholomorphic curves in symplectic manifolds, *Invent. Math.* **82** (1985), p. 307–347.

[17] A. E. Hatcher – A proof of the Smale conjecture, $\mathrm{Diff}(S^3) \simeq \mathrm{O}(4)$, *Ann. of Math.* **117** (1983), p. 553–607.

[18] J. Hilgert & K.-H. Neeb – *Lie semigroups and their applications*, Lecture Notes in Math., vol. 1552, Springer, 1993.

[19] L. Hörmander – Fourier integral operators. I, *Acta Math.* **127** (1971), p. 79–183.

[20] F. Laudenbach – Les 2-sphères de $\mathbb{R}^3$ vues par A. Hatcher et la conjecture de Smale $\mathrm{Diff}(\mathbb{S}^3) \sim \mathrm{O}(4)$, Séminaire Bourbaki, vol. 1983/84, exposé n° 629, *Astérisque* **121-122** (1985), p. 279–293.

[21] F. Laudenbach & J.-C. Sikorav – Persistance d'intersection avec la section nulle au cours d'une isotopie hamiltonienne dans un fibré cotangent, *Invent. Math.* **82** (1985), p. 349–357.

[22] G. I. Ol′shanskiĭ – Invariant orderings in simple Lie groups. Solution of a problem of È. B. Vinberg, *Funktsional. Anal. i Prilozhen.* **16** (1982), p. 80–81 ; traduction anglaise : *Funct. Anal. Appl.* **16** (1982), p. 311–315.

[23] L. Polterovich – Geometry on the group of Hamiltonian diffeomorphisms, in *Proceedings of the International Congress of Mathematicians (Berlin 1998)*, vol. II, 1998, p. 401–410.

[24] S. Sandon – Contact homology, capacity and non-squeezing in $\mathbb{R}^{2n} \times \mathbb{S}^1$ via generating functions, prépublication arXiv:0901.3112, 2009.

[25] J.-C. Sikorav – Sur les immersions lagrangiennes dans un fibré cotangent admettant une phase génératrice globale, *C. R. Acad. Sci. Paris Sér. I Math.* **302** (1986), p. 119–122.

[26] ———, Problèmes d'intersections et de points fixes en géométrie hamiltonienne, *Comment. Math. Helv.* **62** (1987), p. 62–73.

[27] ———, Rigidité symplectique dans le cotangent de T^n, *Duke Math. J.* **59** (1989), p. 759–763.

[28] D. Théret – A complete proof of Viterbo's uniqueness theorem on generating functions, *Topology Appl.* **96** (1999), p. 249–266.

[29] L. Traynor – Symplectic homology via generating functions, *Geom. Funct. Anal.* **4** (1994), p. 718–748.

[30] C. Viterbo – Symplectic topology as the geometry of generating functions, *Math. Ann.* **292** (1992), p. 685–710.

[31] ———, Generating functions, symplectic geometry, and applications, in *Proceedings of the International Congress of Mathematicians (Zürich, 1994)*, Birkhäuser, 1995, p. 537–547.

[32] ———, Functors and computations in Floer homology with applications. II, prépublication, 1996.

[33] ———, Functors and computations in Floer homology with applications. I, *Geom. Funct. Anal.* **9** (1999), p. 985–1033.

Emmanuel GIROUX

École normale supérieure de Lyon
UMPA – UMR CNRS 5669
46, allée d'Italie
F–69364 LYON Cedex 07
E-mail : Emmanuel.Giroux@umpa.ens-lyon.fr

332

ASTÉRISQUE

2010

SÉMINAIRE BOURBAKI
VOLUME 2008/2009
EXPOSÉS 997-1011

Topological modular forms

Paul G. GOERSS

SOCIÉTÉ MATHÉMATIQUE DE FRANCE

Publié avec le concours du CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

Séminaire BOURBAKI Mars 2009
61e année, 2008-2009, no 1005, p. 221 à 255

TOPOLOGICAL MODULAR FORMS
[after Hopkins, Miller, and Lurie]

by **Paul G. GOERSS**

INTRODUCTION

In the early 1970s, Quillen [**55**] noticed a strong connection between 1-parameter formal Lie groups and cohomology theories with a natural theory of Chern classes. The algebraic geometry of these formal Lie groups allowed Morava, Ravenel, and others to make predictions about large scale phenomena in stable homotopy theory, and the resulting theorems completely changed the field. For example, the solution of Ravenel's nilpotence conjectures by Devinatz, Hopkins, and Smith ([**18**] and [**29**]) was one of the great advances of the 1980s.

An example of a 1-parameter formal Lie group can be obtained by taking the formal neighborhood of the identity in a smooth algebraic group of dimension one. The additive group and the multiplicative group correspond to ordinary cohomology and complex K-theory respectively, and the only other algebraic groups of dimension 1 are elliptic curves. This class is different because elliptic curves can come in families over a base scheme S and the geometry of the fibers can vary significantly as we move through S. Thus there are many elliptic cohomology theories and it should be possible to produce them in families over schemes. In retrospect, the realization, by Hopkins and Miller, of a good theory of elliptic cohomology theories provided a centerpiece for the emerging field of derived algebraic geometry.

Derived algebraic geometry has origins in a number of diverse sources. In geometry, there is the work of Serre on multiplicities in intersection theory [**61**] and the work of Illusie [**33**] on the cotangent complex. For an overview of the roots in stable homotopy theory, the article [**27**] is very useful. In algebraic K-theory, again originating with Quillen, it was important very early to consider algebraic varieties and schemes with sheaves of generalized ring objects and, indeed, it was mathematicians in this field who first wrote down a systematic theory [**34**]. Thanks to work of Toën, Vezzosi [**66**],

and Lurie [**47**], we now have a fairly mature theory. The purpose here is to concentrate on the example of elliptic cohomology theories; that is, I would like to make precise the following statement of a theorem of Mike Hopkins and Haynes Miller, as refined by Jacob Lurie: the compactified Deligne-Mumford moduli stack of elliptic curves is canonically and essentially uniquely an object in derived algebraic geometry. The homotopy global sections of this derived stack form the ring spectrum of topological modular forms.

1. AN OVERVIEW

1.1. The moduli stack of elliptic curves

In the late 1960s, Deligne and Mumford [**15**] defined a moduli object $\mathcal{M}_g$ for algebraic curves of genus g. Thus, morphisms $X \to \mathcal{M}_g$ from a scheme X to $\mathcal{M}_g$ are in one-to-one correspondence with smooth proper morphisms

$$q : C \longrightarrow X$$

of relative dimension 1 such that each fiber is a curve of genus g. It was known that $\mathcal{M}_g$ could not be a scheme; one way to see this is to note that automorphisms of the fibers of q do not vary nicely with the fiber. However, Deligne and Mumford noticed that $\mathcal{M}_g$ exists if we enlarge the category of schemes slightly to include what we now call Deligne-Mumford stacks. From this example, the whole theory of algebraic stacks emerged. There is an extremely brief exposition on algebraic stacks at the beginning of Section 3.1.

From this collection, we single out the moduli stack $\mathcal{M}_{1,1}$ of elliptic curves. These are curves of genus 1 with 1 marked point; that is diagrams of the form

$$C \underset{e}{\overset{q}{\rightleftarrows}} X$$

where q is a family of smooth curves of genus 1 over X and e is a section identifying a distinguished point in each fiber. A classical, but still remarkable, property of these curves is that C becomes an abelian group over X: there is a canonical commutative group multiplication on C with e as the unit. Furthermore, any morphism of elliptic curves is a group homomorphism. There are a number of reasons for singling out this stack: elliptic curves are central to algebraic number theory, for example. For algebraic topologists, the formal neighborhood C_e of e in C gives a family of 1-parameter formal Lie groups ("formal groups") which, in turn, gives rise to a rich family of elliptic cohomology theories.

The stack $\mathcal{M}_{1,1}$ is not compact – more precisely, it is not proper over $\mathbb{Z}$. Indeed, the morphism to the affine line

$$j : \mathcal{M}_{1,1} \longrightarrow \mathbb{A}^1$$

which assigns to each elliptic curve C the j-invariant $j(C)$ is almost (but not quite) a covering map. Deligne and Mumford found a canonical compactification $\bar{\mathcal{M}}_{ell}$ of $\mathcal{M}_{1,1}$ which now classifies generalized elliptic curves with, possibly, nodal singularities. In an extensive study of this stack **[16]** Deligne and Rapoport showed that the j-invariant extends to a morphism

$$j : \bar{\mathcal{M}}_{ell} \longrightarrow \mathbb{P}^1.$$

1.2. Derived schemes

The basic objects of algebraic geometry are schemes, which are locally ringed spaces $(X, \mathcal{O}_X)$ which are locally isomorphic to the prime spectrum of some ring. Thus, among other things, X is a space and $\mathcal{O}_X$ is a sheaf of rings on X. The basic idea of derived algebraic geometry is to replace rings by some more generalized ring object. For example, Serre **[61]** considered schemes with a sheaf of commutative differential graded algebras. This has the advantage that it is relatively simple to define and, indeed, over the rational numbers it is equivalent to the more general theory. However, commutative DGAs only have good homotopy theory when we work over the rational numbers; over the integers or in characteristic p a more rigid theory is needed. In his work on the cotangent complex, Illusie **[33]** worked with schemes with sheaves of simplicial commutative algebras. This was also the point of view of Lurie in his thesis **[45]**. However, there are basic examples arising from homotopy theory which cannot come from simplicial algebras – complex K-theory is an important example. Thus, a derived scheme (or stack) will be a scheme equipped with a sheaf of commutative ring spectra. I immediately remark that "commutative" is a difficult notion to define in ring spectra: here I mean "E_∞-ring spectra". The foundations of commutative ring spectra are forbidding, but I'll make some attempt at an exposition below in Section 2.1.

A spectrum X has homotopy groups $\pi_k X$, for $k \in \mathbb{Z}$. If X is a commutative ring spectrum, $\pi_0 X$ is a commutative ring and the graded abelian group $\pi_* X$ is a graded skew-commutative $\pi_0 X$-algbera. In particular, $\pi_k X$ is a $\pi_0 X$-module.

DEFINITION 1.1. — *A* derived scheme $(X, \mathcal{O})$ *is a pair with* X *a topological space and* $\mathcal{O}$ *a sheaf of commutative ring spectra on* X *so that*

(1) *the pair* $(X, \pi_0 \mathcal{O})$ *is a scheme; and*
(2) *the sheaf* $\pi_k \mathcal{O}$ *is a quasi-coherent sheaf of* $\pi_0 \mathcal{O}$*-modules.*

It is somewhat subtle to define the notion of a sheaf of spectra; I will come back to that point below in Remark 2.7. One definition of "quasi-coherent" is to require the sheaf to be locally the cokernel of a morphism between locally free sheaves. On the spectrum of a ring R, these are the module sheaves which arise from R-modules. There is a completely analogous definition of a derived Deligne-Mumford stack, except that now we must be careful about the topology we use to define sheaves: for these objects we must use the étale topology.

There is a rich structure inherent in Definition 1.1. The homotopy groups of a commutative ring spectrum support far more structure than simply that of a graded skew-commutative ring; in particular, it is a ring with "power operations". See Remark 2.6 below. Thus, if $(X, \mathcal{O})$ is a derived scheme, the graded sheaf $\pi_*\mathcal{O}$ is a sheaf of graded rings with all of this higher order structure.

1.3. Topological modular forms

On the compactified Deligne-Mumford stack $\bar{\mathcal{M}}_{ell}$, there is a canonical quasi-coherent sheaf ω. If C is an elliptic curve over X, then C is an abelian variety over X of relative dimension 1 and we can construct the sheaf of invariant 1-forms ω_C for C. This is a locally free sheaf of rank 1 on X and the assignment

$$\omega(C : X \to \bar{\mathcal{M}}_{ell}) = \omega_C$$

defines a quasi-coherent sheaf on $\bar{\mathcal{M}}_{ell}$. The sheaf ω is locally free of rank 1, hence invertible, and the tensor powers $\omega^{\otimes k}$, $k \in \mathbb{Z}$, are all quasi-coherent. Here is the main result; see **[26]** and **[46]**.

THEOREM 1.2 (Hopkins-Miller-Lurie). — *There exists a derived Deligne-Mumford stack $(\bar{\mathcal{M}}_{ell}, \mathcal{O})$ so that*

(1) *the underlying algebraic stack $(\bar{\mathcal{M}}_{ell}, \pi_0\mathcal{O})$ is equivalent to the compactified Deligne-Mumford moduli stack $(\bar{\mathcal{M}}_{ell}, \mathcal{O}_{ell})$ of generalized elliptic curves; and*
(2) *there are isomorphisms of quasi-coherent sheaves $\pi_{2k}\mathcal{O} \cong \omega^{\otimes k}$ and $\pi_{2k+1}\mathcal{O} = 0$.*

Furthemore, the derived stack $(\bar{\mathcal{M}}_{ell}, \mathcal{O})$ is determined up to equivalence by conditions (1) *and* (2).

The doubling of degrees in (2) is quite typical: when the homotopy groups of a commutative ring spectrum X are concentrated in even degrees, then π_*X is a graded commutative ring, not just commutative up to sign.

We can rephrase the uniqueness statement of Theorem 1.2 as follows: there is a space of all derived stacks which satisfy points (1) and (2) and this space is path-connected. In fact, Lurie's construction gives a canonical base-point; that is, a canonical model for $(\bar{\mathcal{M}}_{ell}, \mathcal{O})$.

Here is a definition of the object in the title of this manuscript.

DEFINITION 1.3. — *The spectrum* **tmf** *of* topological modular forms *is the derived global sections of the sheaf $\mathcal{O}$ of commutative ring spectra of the derived Deligne-Mumford stack $(\bar{\mathcal{M}}_{ell}, \mathcal{O})$.*

The spectrum **tmf** is itself a commutative ring spectrum and formal considerations now give a descent spectral sequence

$$H^s(\bar{\mathcal{M}}_{ell}, \omega^{\otimes t}) \Longrightarrow \pi_{2t-s}\mathbf{tmf}. \tag{1}$$

This spectral sequence has been completely calculated in [**7**], [**28**] and [**59**].

The spectrum **tmf** is called "topological modular forms" for the following reason. One definition of modular forms of weight k (and level 1) is as the global sections of $\omega^{\otimes k}$ over $\bar{\mathcal{M}}_{ell}$; that is, the graded ring

$$M_* = H^0(\bar{\mathcal{M}}_{ell}, \omega^{\otimes *})$$

is the ring of modular forms for generalized elliptic curves. This ring is well understood [**14**]: there are modular forms c_4, c_6, and Δ of weights 4, 6, and 12 respectively and an isomorphism of graded rings

$$\mathbb{Z}[c_4, c_6, \Delta]/(c_4^3 - c_6^2 = 1728\Delta) \cong M_*.$$

Note $1728 = (12)^3$, indicating that the primes 2 and 3 are special in this subject. The modular form Δ is the discriminant and is the test for smoothness: a generalized elliptic curve C is smooth if and only if $\Delta(C)$ is invertible; in fact, $\mathcal{M}_{1,1} \subseteq \bar{\mathcal{M}}_{ell}$ is the open substack obtained by inverting Δ.

Modular forms then form the zero line of the spectral sequence of (1), at least up to degree doubling, and we can now ask which modular forms give homotopy classes in **tmf**. The higher cohomology groups of (1) are all 2 and 3-torsion and, as mentioned above, the differentials have also been calculated. Thus, for example, we know that Δ is not a homotopy class, but 24Δ is; similarly c_6 is not a homotopy class, but $2c_6$ is. The class c_4 is a homotopy class. In the last section of this note, I will uncover some of the details of this calculation.

1.4. Impact

Let me write down three areas of algebraic topology where **tmf** has had significant impact. I will go in more-or-less chronological order.

Remark 1.4 (The Witten genus). — In his work in string theory, Witten noticed that one could define a genus for certain spin manifolds that takes values in modular forms. A compact differentiable manifold M has a spin structure if the first and second Stiefel-Whitney classes vanish and a spin manifold M is a manifold with a choice of spin structure. For spin manifolds there is a new characteristic class λ which has the property that twice this class is the first Pontrjagin class. If λ also vanishes, the

manifold M has a *string* structure. There is a cobordism ring $\mathrm{MString}_*$ of string manifolds, and Witten wrote down a rich ring homomorphism

$$\sigma_* : \mathrm{MString}_* \longrightarrow \mathbb{Q} \otimes M_*,$$

which we now call the Witten genus. Let me say a little about his methods.

Every modular form can be written as a power series over the integers in q; this is the q-expansion. One way to do this is to evaluate the modular form on the Tate curve, which is a generalized elliptic curve over $\mathbb{Z}[[q]]$. This defines a monomorphism $M_* \to \mathbb{Z}[[q]]$ but it is definitely not onto. For each string manifold, Witten [**67**] wrote down a power series in q over the rationals and then used physics to argue it must be a modular form.

Almost immediately, homotopy theorists began searching for a spectrum-level construction of this map. By the Thom-Pontrjagin construction, $\mathrm{MString}_*$ is isomorphic to the homotopy groups of a commutative ring spectrum MString and the Witten genus was posited to be given by a morphism of ring spectra from MString to some appropriate ring spectrum or family of ring spectra; this is what happens with the Atiyah-Bott-Shapiro realization of the $\hat{A}$-genus on spin manifolds. Motivated by this problem, among others, there were an early important work on elliptic cohomology theories and orientations and a major conference on elliptic cohomology theories in the late 1980s; [**39**] is a highlight of this period. There were also very influential papers in the following years. For example, the Bourbaki exposé by Segal [**60**] has been especially important in the search for elliptic cohomologies arising from differential geometry and mathematical physics and the paper by Franke [**20**] gave a profusion of examples of cohomology theories arising from elliptic curves.

The Witten genus was addressed specifically in the work of Ando, Hopkins, Rezk, and Strickland. Earlier results are in [**5**] and [**25**]; the definitive results are in [**4**] and [**6**].

THEOREM 1.5. — *The Witten genus can be realized as a morphism of commutative ring spectra*

$$\sigma : \mathrm{MString} \longrightarrow \mathbf{tmf}.$$

The map σ is surjective on homotopy in non-negative degrees.

Besides giving a rigid construction of the Witten genus, this has other consequences. Not every modular form is in the image, for example, as not every modular form is a homotopy class. On the other hand, there is a plenty of 2 and 3 torsion in both source and target and the map σ detects whole families of this torsion.

Remark 1.6 (Homotopy groups of spheres). — One of the fundamental problems of stable homotopy theory is to compute $\pi_* S$, where S is the stable sphere; thus

$\pi_k S = \operatorname{colim} \pi_{n+k} S^n$. This is impossible, at least at this point, but we can measure progress against this problem. Early successful calculation focused on the image of the J-homomorphism $\pi_* SO(n) \to \pi_* S^n$; see [**1**] and [**48**]. There were many calculations with higher order phenomena, most notably in [**53**] and the work of Shimomura and his coauthors (see [**63**] and [**62**] among many papers), but the very richness of these results hindered comprehension. Using **tmf**, related spectra, and the algebra and geometry of elliptic curves, it is now possible to reorganize the calculations in a way that better reveals the larger structure. See for example, [**21**], [**23**], and [**8**]. The latter, in particular, makes the connections with elliptic curves explicit.

Remark 1.7 (Congruences among modular forms). — There is a remarkable interplay between homotopy theory and the theory of modular forms. For example, by thinking about the descent spectral sequence of Equation (1) above, Hopkins [**26**] discovered a new proof of a congruence of Borcherds [**11**]. We can also vary the moduli problem to consider elliptic curves with appropriate homomorphisms $(\mathbb{Z}/N\mathbb{Z})^2 \to C$ ("level-structures"). This new moduli problem is étale over $\mathcal{M}_{1,1}[1/N]$ and Theorem 1.2 immediately produces a new spectrum $\mathbf{tmf}[N]$ in $\mathbb{Z}[1/N]$-local homotopy theory, which might be called topological modular forms of level N. There is a descent spectral sequence analogous to (1) beginning with modular forms of level N. This has been used to effect in [**50**] and [**8**].

More subtle connections and congruences emerge when we work p-adically. For example, the ring of divided congruences of Katz [**36**] appears in topology as the p-complete K-theory of **tmf**. This observation, due to Hopkins, appears in [**41**] for the prime 2. More recently, Behrens [**9**] was able to explicitly compare certain congruences among modular forms that hold near supersingular curves with the intricate and beautiful patterns in the stable homotopy groups of spheres that first appeared in [**53**].

2. BASICS FROM HOMOTOPY THEORY

2.1. Spectra and commutative ring spectra

The need for a good theory of spectra arose in the 1950s while trying to find a framework to encode a variety of examples that displayed similar phenomena. In each case there was a *spectrum*; that is, a sequence of based (or pointed) spaces X_n and suspension maps

$$\Sigma X = S^1 \wedge X_n \to X_{n+1}.$$

Here $A \wedge B$ is the smash product of two pointed spaces characterized by the adjunction formula in spaces of pointed maps

$$\operatorname{map}_*(A \wedge B, Y) \cong \operatorname{map}_*(A, \operatorname{map}_*(B, Y)).$$

Example 2.1. — 1) The space $S^1 \wedge A$ is the suspension of A and a basic example is the suspension spectrum $\Sigma^\infty A$ of a pointed finite CW complex A; thus, $(\Sigma^\infty A)_n = \Sigma^n A$ is the iterated suspension. In particular, $S = \Sigma^\infty S^0$.

2) A generalized cohomology theory is a contravariant functor $E^*(-)$ from the homotopy category of spaces to graded abelian groups which takes disjoint unions to products and has a Mayer-Vietoris sequence. By Brown's representability theorem there are spaces E_n so the reduced cohomology $\tilde{E}^n(A)$ is naturally isomorphic to the pointed homotopy classes of maps $A \to E_n$. The suspension isomorphism $\tilde{E}^n A \cong \tilde{E}^{n+1}\Sigma A$ yields the map $\Sigma E_n \to E_{n+1}$.

3) The Thom spectra MO and its variants such as MString. The n-th space of MO is the Thom space $\mathrm{MO}(n) = T(\gamma_n)$ of the universal real n-plane bundle γ_n over the infinite Grassmannian. The map $\mathrm{Gr}_n(\mathbb{R}^\infty) \to \mathrm{Gr}_{n+1}(\mathbb{R}^\infty)$ classifying the Whitney sum $\epsilon \oplus \gamma_n$, where ϵ is a trivial 1-plane bundle, defines the map

$$S^1 \wedge \mathrm{MO}(n) \cong T(\epsilon \oplus \gamma_n) \to T(\gamma_{n+1}) = \mathrm{MO}(n+1).$$

A basic invariant of a spectrum X is the *stable homotopy groups*

$$\pi_k X = \operatorname{colim} \pi_{n+k} X_n, \qquad k \in \mathbb{Z}.$$

For example, the Thom-Pontrjagin construction shows that $\pi_k\mathrm{MO}$ is the group of cobordism classes of closed differentiable k-manifolds; it was an early triumph of homotopy theory that Thom was able to compute these groups.

Remark 2.2 (The stable homotopy category). — A basic difficulty was to decide how to build a homotopy theory of spectra. Seemingly anomalous examples led to some very ingenious ideas, most notably the "cells now maps later" construction of Boardman and Adams [**2**]. The language of model categories gives a simple and elegant definition. We define a morphism $f : X \to Y$ of spectra to be a set of pointed maps $f_n : X_n \to Y_n$ which commute with the suspension maps. Next we stipulate that such a morphism is a weak equivalence if $\pi_* f : \pi_* X \cong \pi_* Y$. Bousfield and Friedlander [**12**] showed that there is a model category structure on spectra with these weak equivalences, and we obtain the homotopy category by formally inverting the weak equivalences.

There was a much more serious difficulty in the theory, however. Suppose that we have a cohomology theory $E^*(-)$ with natural, associative, and graded commutative cup products

$$E^m X \otimes E^n X \longrightarrow E^{m+n} X.$$

Then, by considering the universal examples, we get a map

$$E_m \wedge E_n \longrightarrow E_{n+m}.$$

From this we ought to get a map $E \wedge E \to E$ of spectra making E into a homotopy associative and commutative ring object. However, given two spectra X and Y the object $\{X_m \wedge Y_n\}$ is not a spectrum, but a "bispectrum". We can extract a spectrum Z by taking Z_k to be any of the spaces $X_m \wedge Y_n$ with $m+n=k$ with m and n non-decreasing; if, in addition, we ask that n and m go to ∞ with k, we get a well-defined homotopy type – and hence a symmetric monoidal smash product on the homotopy category. However, this does not descend from such a structure on the category of spectra.

Worse, there were important examples that indicated that there should be such a structure on spectra. The Eilenberg-MacLane spectra HR, R a commutative ring, the cobordism spectrum MO and the complex analog MU, spectra arising from algebraic and topological K-theory, and the sphere spectrum S itself all had more structure than simply giving ring objects in the homotopy category.

Remark 2.3. — The first, and still a very elegant, solution to this problem was due to Lewis, May and Steinberger [**43**]. The idea was to expand the notion of a spectrum X to be a collection of spaces $\{X_V\}$ indexed on the finite-dimensional subspaces of the infinite inner product space $\mathbb{R}^\infty$; the suspension maps then went $S^V \wedge X_W \to X_{V\oplus W}$, where S^V was the one-point compactification of V. If X_i, $1 \le i \le n$ are spectra, then $X_1 \wedge \cdots \wedge X_n$ naturally yields a spectrum object over $(\mathbb{R}^\infty)^n$ and any linear isometry $f : \mathbb{R}^\infty \to (\mathbb{R}^\infty)^n$ then returns a spectrum $f^*(X_1 \wedge \cdots \wedge X_n)$ over $\mathbb{R}^\infty$. The crucial observation is that the space $\mathscr{L}(n)$ of all such choices of isometries is contractible and the construction of $f^*(X_1 \wedge \cdots \wedge X_n)$ could be extended to a construction of a functor

$$(X_1, \cdots, X_n) \mapsto \mathscr{L}(n)_+ \wedge X_1 \wedge \cdots \wedge X_n.$$

This functor, in the case $n=2$, descended to the smash product on the homotopy category. More importantly, the collection of spaces $\mathscr{L} = \{\mathscr{L}(n)\}$ form an E_∞-operad. This implies that there is a coherent way to compose the above functors and that the action of the symmetric group on $\mathscr{L}(n)$ is free.

We still did not have a spectrum level symmetric monoidal structure, but we did now know what a ring object should be.

Definition 2.4. — *A* commutative ring spectrum *is an algebra over the operad $\mathscr{L}$; that is, an algebra for the monad*

$$X \mapsto \vee_n \ \mathscr{L}(n)_+ \wedge_{\Sigma_n} X^{\wedge n}.$$

Here $\vee$ is the coproduct or wedge in spectra; and $\wedge_{\Sigma_n}$ means divide out by the diagonal action of the symmetric group. We can also call a commutative ring spectrum an E_∞-ring spectrum.

This definition has the distinct advantage of being a machine with an input slot; for example, this theory is ideal for showing that the Thom spectra MO and MU are commutative ring spectra; similarly, the Eilenberg-MacLane spectra HR, with R a commutative ring, and S are easily seen to be commutative ring spectra.

Example 2.5. — If R is a commutative ring, we could consider algebras over an E_∞-operad in R-chain complexes; that is, we could work with differential graded E_∞-algebras. However, in a very strong sense (the technical notion is *Quillen equivalence* or equivalence of ∞-categories) the category of differential graded E_∞-algebras over R has the same homotopy theory as the category of commutative HR-algebras in spectra, where HR is the associated Eilenberg-MacLane spectra. This is actually a fairly restrictive example of commutative ring spectra; many important examples, such as MO, MU, and K-theory are not of this type. I note that if R is a $\mathbb{Q}$-algebra, then E_∞-algebras over R and commutative dgas over R also form Quillen equivalent categories.

Remark 2.6. — Since the work of Lewis, May and Steinberger, many authors (including May himself) have built models for the stable category which indeed have a symmetric monoidal structure. See [**19**], [**31**], and [**51**]. Then a commutative ring spectrum is simply a commutative monoid for that structure. However, it is only a slight exaggeration to say that all such models build an E_∞-operad into the spectrum-level smash product in some way or another; thus, the theory is elegant, but the computations remain the same. For example, the homotopy of an E_∞-ring spectrum is a commutative ring, but there is much more structure as well: the map in homotopy

$$\pi_*\mathcal{L}(n)_+ \wedge_{\Sigma_n} X^{\wedge n} \longrightarrow \pi_* X$$

adds *power operations* (such as Steenrod or Dyer-Lashof operations) to $\pi_* X$. The spectral sequence

$$H_p(\Sigma_n, \pi_q(X^{\wedge n})) \Longrightarrow \pi_{p+q}\mathcal{L}(n)_+ \wedge_{\Sigma_n} X^{\wedge n}$$

makes the role of the homology of the symmetric groups explicit in the construction of these operations.

Remark 2.7 (Sheaves of spectra). — Once we have ring spectra, we must confront what we mean by a sheaf of ring spectra. There is an issue here as well. Suppose we have a presheaf $\mathcal{F}$ on X and $\{V_i\}$ is a cover of $U \subseteq X$. Then if $\mathcal{F}$ is a sheaf we have an equalizer diagram

$$\mathcal{F}(U) \longrightarrow \prod \mathcal{F}(V_i) \rightrightarrows \prod \mathcal{F}(V_i \times_U V_j).$$

The problem is that equalizers are not homotopy invariant. Again model categories help. We define a morphism of presheaves $\mathcal{E} \to \mathcal{F}$ to be a weak equivalence if it induces

an isomorphism $\pi_*\mathscr{E} \to \pi_*\mathscr{T}$ of associated homotopy sheaves. Thus, for example, a presheaf is weakly equivalent to its associated sheaf. The theorem, due to Jardine **[34]**, **[35]**, is that there is a model category structure on presheaves with these weak equivalences. A *sheaf of (ring) spectra* is then a fibrant/cofibrant object in this model category. This has the effect of building in the usual homological algebra for sheaves; that is, if $\mathscr{T}$ is a module sheaf and $\mathscr{T} \to \mathscr{I}^\bullet$ is an injective resolution; then the associated presheaf of generalized Eilenberg-MacLane spectra $K\mathscr{I}^\bullet$ is a fibrant model for $K\mathscr{T}$. In this setting, global sections $\Gamma(-)$ are inherently derived so I may write $R\Gamma(-)$ instead. In good cases there is a descent spectral sequence

$$H^s(X, \pi_t\mathscr{T}) \Longrightarrow \pi_{t-s}\Gamma(\mathscr{T}).$$

2.2. Cohomology theories and formal groups

If E is a spectrum associated to a cohomology theory E^*, we get a homology theory by setting $E_*X = \pi_*E \wedge X_+$, where X_+ is X with a disjoint basepoint.

DEFINITION 2.8. — *Let $E^*(-)$ be a cohomology theory. Then E^* is* 2-periodic *if*

(1) *the functor $X \mapsto E^*(X)$ is a functor to graded commutative rings;*
(2) *for all integers k, $E^{2k+1} = E^{2k+1}(*) = 0$;*
(3) *E^2 is a projective module of rank* 1 *over E^0; and*
(4) *for all integers k, the cup product map $(E^2)^{\otimes k} \to E^{2k}$ is an isomorphism.*

Note that E^2 is an invertible module over E^0 and E^{-2} is the dual module. If E^2 is actually free, then so is $E_2 = E^{-2}$ and a choice of generator $u \in E_2$ defines an isomorphism $E_0[u^{\pm 1}] \cong E_*$. This often happens; for example, in complex K-theory. However, there are elliptic cohomology theories for which E_2 does not have a global generator.

From 2-periodic cohomology theories we automatically get a formal group. Let $\mathbb{C}\mathrm{P}^\infty = \mathrm{Gr}_1(\mathbb{C}^\infty)$ be the infinite Grassmannian classifying complex line bundles. Then $\mathbb{C}\mathrm{P}^\infty$ is a topological monoid where the multiplication $\mathbb{C}\mathrm{P}^\infty \times \mathbb{C}\mathrm{P}^\infty \to \mathbb{C}\mathrm{P}^\infty$ classifies the tensor product of line bundles. If E^* is a 2-periodic homology theory, then $E^0\mathbb{C}\mathrm{P}^\infty$ is complete with respect to the augmentation ideal

$$I(e) \stackrel{\mathrm{def}}{=} \tilde{E}^0\mathbb{C}\mathrm{P}^\infty = \mathrm{Ker}\{\, E^0\mathbb{C}\mathrm{P}^\infty \to E^0(*)\,\}$$

and, using the monoid structure on $\mathbb{C}\mathrm{P}^\infty$, we get a commutative group object in formal schemes

$$G_E = \mathrm{Spf}(E^0\mathbb{C}\mathrm{P}^\infty).$$

This formal group is smooth and one-dimensional in the following sense. Define the E_0-module ω_G by

$$\omega_G = I(e)/I(e)^2 \cong \tilde{E}^0 S^2 \cong E_2. \tag{2}$$

This module is locally free of rank 1, hence projective, and any choice of splitting of $I(e) \to \omega_G$ defines a homomorphism out of the symmetric algebra

$$S_{E_0}(\omega_G) \longrightarrow E^0\mathbb{C}\mathrm{P}^\infty$$

which becomes an isomorphism after completion. For example, if E_2 is actually free we get a non-canonical isomorphism

$$E^0\mathbb{C}\mathrm{P}^\infty \cong E^0[[x]].$$

Such an x is called a *coordinate*. Whether G_E has a coordinate or not, the ring E_0 and the formal group determine the graded coefficient ring E_*; indeed, $E_{2t+1} = 0$ and for all $t \in \mathbb{Z}$,

$$\omega_G^{\otimes t} \cong \tilde{E}^0 S^{2t} \cong E_{2t}. \tag{3}$$

Remark 2.9 (Formal group laws). — The standard literature on chromatic homotopy theory, such as **[2]** and **[56]**, emphasizes formal group laws. If $E^*(-)$ is a two-periodic theory with a coordinate, then the group multiplication

$$G_E \times G_E = \mathrm{Spf}(E^0(\mathbb{C}\mathrm{P}^\infty \times \mathbb{C}\mathrm{P}^\infty)) \to \mathrm{Spf}(E^0\mathbb{C}\mathrm{P}^\infty) = G_E$$

determines and is determined by a power series

$$x +_F y = F(x, y) \in E^0[[x, y]] \cong E^0(\mathbb{C}\mathrm{P}^\infty \times \mathbb{C}\mathrm{P}^\infty).$$

This power series is a 1-dimensional formal group law.

Homomorphisms can also be described by power series, and a homomorphism ϕ is an isomorphism if $\phi'(0)$ is a unit.

Remark 2.10 (Invariant differentials). — The module ω_{G_E} is defined as the conormal module of the embedding

$$e : \mathrm{Spec}(E^0) \to \mathrm{Spf}(E^0\mathbb{C}\mathrm{P}^\infty) = G_E$$

given by the basepoint. This definition extends to any formal group G over any base scheme X and we get a sheaf ω_G on X. It is isomorphic to the sheaf of invariant differentials of G via an evident inclusion $\omega_G \to q_*\Omega_{G/X}$. Here $\Omega_{G/X}$ is the sheaf of continuous differentials and $q : G \to X$ is the structure map.

If $X = \mathrm{Spec}(R)$ for some ring and G has a coordinate x, then the invariant differentials form the free R-module generated by the *canonical* invariant differential

$$\eta_G = \frac{dx}{F_y(x, 0)}$$

where $F_y(x,y)$ is the partial derivative of the associated formal group law. It is an exercise to calculate that if $\phi : G_1 \to G_2$ is a homomorphism of formal groups with coordinate, then $d\phi : \omega_{G_2} \to \omega_{G_1}$ is determined by

$$d\phi(\eta_{G_2}) = \phi'(0)\eta_{G_1}. \tag{4}$$

3. FORMAL GROUPS AND STABLE HOMOTOPY THEORY

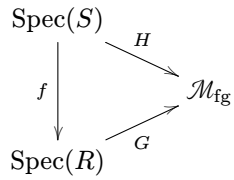
3.1. The moduli stack of formal groups

Let $\mathcal{M}_{\mathrm{fg}}$ be the moduli stack of formal groups: this is the algebro-geometric object which classifies all smooth 1-parameter formal Lie groups and their isomorphisms. Thus, if R is a commutative ring, the morphisms

$$G : \mathrm{Spec}(R) \longrightarrow \mathcal{M}_{\mathrm{fg}}$$

are in one-to-one correspondence with formal groups G over R. Furthermore, the 2-commutative diagrams

(5)



correspond to pairs $(f : R \to S,\ \phi : H \xrightarrow{\cong} f^*G\ )$.

Remark 3.1. — Schemes are defined as locally ringed spaces $(X, \mathcal{O}_X)$ which have an open cover, as locally ringed spaces, by affine schemes. Equivalently, schemes can be defined as functors from rings to sets which are sheaves in the Zariski topology and have an open cover, as functors, by functors of the form

$$A \mapsto \mathbf{Rings}(R, A).$$

Stacks are a generalization of the second definition. A stack is a sheaf of groupoids on commutative rings satisfying an effective descent condition [**40**], §3. For example, $\mathcal{M}_{\mathrm{fg}}$ assigns to each ring R the groupoid of formal groups over $\mathrm{Spec}(R)$ [(1)].

Algebraic stacks have a suitable cover by schemes. A morphism $\mathcal{M} \to \mathcal{N}$ of stacks is *representable* if for all morphisms $X \to \mathcal{N}$ with X a scheme, the 2-category pull-back (or homotopy pull-back) $X \times_{\mathcal{N}} \mathcal{M}$ is equivalent to a scheme. A representable

(1) As in [**40**], §2, we should really speak of categories fibered in groupoids, rather than sheaves of groupoids – for f^*g^*G is only isomorphic to $(gf)^*G$. However, there are standard ways to pass between the two notions.

morphism then has algebraic property P (flat, smooth, surjective, étale, etc.) if all the resulting morphisms

$$X \times_{\mathcal{N}} \mathcal{M} \to X$$

have that property.

A stack $\mathcal{M}$ is then called *algebraic*[(2)] if

(1) every morphism $Y \to \mathcal{M}$ with Y a scheme is representable; and
(2) there is a smooth surjective map $q : X \to \mathcal{M}$ with X a scheme.

The morphism q is called a presentation. Note that an algebraic stack may have many presentations. If a presentation can be chosen to be étale, we have a *Deligne-Mumford stack*.

Remark 3.2. — The stack $\mathcal{M}_{\mathrm{fg}}$ is not algebraic, in this sense, as it only has a flat presentation, not a smooth presentation. If we define **fgl** to be the functor which assigns to each ring R the set of formal group laws over R, then Lazard's theorem **[42]** says that $\mathbf{fgl} = \mathrm{Spec}(L)$ where L is (non-canonically) isomorphic to $\mathbb{Z}[t_1, t_2, \ldots]$. The map

$$\mathbf{fgl} \longrightarrow \mathcal{M}_{\mathrm{fg}}$$

which assigns a formal group law to its underlying formal group is flat and surjective, but not smooth since it is not finitely presented. It is *pro-algebraic* however; that is, it can be written as the 2-category inverse limit of a tower of the algebraic stacks of "buds" of formal groups. This is inherent in **[42]** and explicit in **[65]**.

Remark 3.3. — A sheaf in the *fpqc*-topology on an algebraic stack $\mathcal{M}$ is a functor $\mathcal{F}$ on the category of affine schemes over $\mathcal{M}$ which satisfies faithfully flat descent. For example, define the structure sheaf $\mathcal{O}_{\mathrm{fg}}$ to be the functor on affine schemes over $\mathcal{M}_{\mathrm{fg}}$ with

$$\mathcal{O}_{\mathrm{fg}}(R, G) = \mathcal{O}_{\mathrm{fg}}(G : \mathrm{Spec}(R) \to \mathcal{M}) = R.$$

A module sheaf $\mathcal{F}$ over $\mathcal{O}_{\mathcal{M}}$ is *quasi-coherent* if, for each 2-commutative diagram over $\mathcal{M}$, the restriction map $\mathcal{F}(R, G) \to \mathcal{F}(S, H)$ extends to an isomorphism

$$S \otimes_R \mathcal{F}(R, G) \cong \mathcal{F}(S, H).$$

This isomorphism can be very non-trivial, as it depends on the choice of isomorphism ϕ which makes the diagram 2-commute.

A fundamental example of a quasi-coherent sheaf is the sheaf of invariant differentials ω on $\mathcal{M}_{\mathrm{fg}}$ with

$$\omega(R, G) = \omega_G$$

(2) The notion defined here is stronger than what is usually called an algebraic (or Artin) stack, which requires a cover only by an algebraic space. Algebraic spaces are sheaves which themselves have an appropriate cover by a scheme. Details are in **[40]**.

the invariant differentials on G. This is locally free of rank 1 and hence all powers $\omega^{\otimes n}$, $n \in \mathbb{Z}$, are also quasi-coherent sheaves. The effect of the choice of isomorphism in the 2-commuting diagram on the transition maps for $\omega^{\otimes n}$ is displayed in Equation 4.

3.2. The height filtration

Consider a formal group G over a scheme X over $\mathbb{F}_p$. If we let $f : X \to X$ be the Frobenius, we get a new formal group $G^{(p)} = f^*G$. The Frobenius $f : G \to G$ factors through the *relative Frobenius* $F : G \to G^{(p)}$. We know that if $\phi : G \to H$ is a homomorphism of formal groups over X for which $d\phi = 0 : \omega_H \to \omega_G$, there is then a factoring $\phi = \psi F : G \to H$. Then we can test $d\psi$ to see if we can factor further.

For example, let $\phi = p : G \to G$ be the p-th power map. Then we obtain a factoring

$$G \xrightarrow[F]{} G^{(p)} \xrightarrow[V_1]{} G, \qquad G \xrightarrow{\ p\ } G.$$

This yields an element

$$dV_1 \in \mathrm{Hom}(\omega_G, \omega_{G^{(p)}})$$

and we can factor further if $dV_1 = 0$. Since G is of dimension 1, $\omega_{G^{(p)}} = \omega_G^{\otimes p}$; since ω_G is invertible,

$$\mathrm{Hom}(\omega_G, \omega_G^{\otimes p}) = \mathrm{Hom}(\mathcal{O}_X, \omega_G^{\otimes p-1}).$$

Thus dV_1 defines a global section $v_1(G)$ of $\omega_G^{\otimes p-1}$. If $v_1(G) = 0$, then we obtain a further factorization and a global section $v_2(G) \in \omega_G^{\otimes p^2-1}$. This can be continued to define sections $v_n(G) \in \omega_G^{\otimes p^n-1}$ and G has *height* at least n if

$$v_1(G) = \cdots = v_{n-1}(G) = 0.$$

We say G has height *exactly* n if $v_n(G) : \mathcal{O}_X \to \omega_G^{\otimes p^n-1}$ is an isomorphism. Note that a formal group may have infinite height.

The assignment $G \mapsto v_1(G)$ defines a global section v_1 of the sheaf $\omega^{\otimes p-1}$ on the closed substack

$$\mathbb{F}_p \otimes \mathcal{M}_{\mathrm{fg}} \stackrel{\mathrm{def}}{=} \mathcal{M}(1) \subseteq \mathcal{M}_{\mathrm{fg}}.$$

Indeed, we obtain a sequence of closed substacks

$$\cdots \subseteq \mathcal{M}(n+1) \subseteq \mathcal{M}(n) \subseteq \cdots \subseteq \mathcal{M}(1) \subseteq \mathcal{M}_{\mathrm{fg}}$$

where $\mathcal{M}(n+1) \subseteq \mathcal{M}(n)$ is defined by the vanishing of the global section v_n of $\omega^{\otimes p^n-1}$. Thus $\mathcal{M}(n)$ classifies formal groups of height at least n. The relative open

$$\mathcal{H}(n) = \mathcal{M}(n) - \mathcal{M}(n+1)$$

classifies formal groups of height exactly n. One of Lazard's theorems [**42**], rephrased, says that $\mathcal{H}(n)$ has a single geometric point given by a formal group G of height n over any algebraically closed field $\mathbb{F}$ of characteristic p. The pair $(\mathbb{F}, G)$ has plenty of

automorphisms, however, so $\mathscr{H}(n)$ is not a scheme; indeed, it is a *neutral gerbe*. See [**65**].

Remark 3.4 (Landweber's criterion for flatness). — If $G : \mathrm{Spec}(R) \to \mathscr{M}_{\mathrm{fg}}$ is any flat map, then there is a 2-periodic homology theory $E(R, G)$ with $E(R, G)_0 \cong R$ and with G as the associated formal group. Landweber's Exact Functor Theorem gives an easily checked criterion to decide when a representable morphism $\mathscr{N} \to \mathscr{M}_{\mathrm{fg}}$ is flat.

Let $\mathscr{O}_n$ be the structure sheaf of the substack $\mathscr{M}(n)$ of $\mathscr{M}_{\mathrm{fg}}$. Then the global section $v_n \in H^0(\mathscr{M}(n), \omega^{\otimes p^n-1})$ defines an *injection* of sheaves

$$0 \to \mathscr{O}_n \xrightarrow{v_n} \omega^{\otimes p^n-1} .$$

This yields a short exact sequence

$$0 \to \omega^{\otimes -(p^n-1)} \xrightarrow{v_n} \mathscr{O}_n \longrightarrow j_*\mathscr{O}_{n+1} \to 0.$$

This identifies $\omega^{\otimes -(p^n-1)}$ with the ideal defining the closed immersion of $\mathscr{M}(n+1)$ in $\mathscr{M}(n)$.

Now let $f : \mathscr{N} \to \mathscr{M}_{\mathrm{fg}}$ be a representable morphism of stacks and let

$$\mathscr{N}(n) = \mathscr{M}(n) \times_{\mathscr{M}_{\mathrm{fg}}} \mathscr{N} \subseteq \mathscr{N}.$$

Then $\mathscr{N}(n+1) \subseteq \mathscr{N}(n)$ remains a closed immersion and, if f is flat, then

$$\mathscr{O}_{\mathscr{N}(n)} \xrightarrow{v_n} \omega^{\otimes p^n-1} \tag{6}$$

remains an injection. Landweber's theorem now says that this is sufficient; that is, if for all primes p and all n the morphism of Equation (6) is an injection, then the representable morphism $\mathscr{N} \to \mathscr{M}_{\mathrm{fg}}$ is flat. The original source is [**38**]; in the form presented here, it appears in [**24**] and [**54**].

Remark 3.5 (Chromatic stable homotopy theory). — The moduli stack $\mathscr{M}_{\mathrm{fg}}$ of formal groups has not been shown to be a derived stack and it may not be. One technical difficulty is that we must use the *fpqc*-topology on $\mathscr{M}_{\mathrm{fg}}$ and there are many flat maps. Nonetheless, the geometry of the stack $\mathscr{M}_{\mathrm{fg}}$ has been successfully used to predict theorems about large scale phenomena in stable homotopy theory. Here are some of the basic results.

If $\mathscr{M}_{\mathrm{fg}}$ could be lifted to a derived stack, there would be a descent spectral sequence

$$H^s(\mathscr{M}_{\mathrm{fg}}, \omega^{\otimes t}) \Longrightarrow \pi_{2t-s}S.$$

This spectral sequence exists: it is the Adams-Novikov Spectral Sequence. It remains our most sensitive algebraic approximation to π_*S. The paper [**53**] initiated the modern era of calculations in stable homotopy theory; much of the algebra there is driven by Morava's meditations on the geometry of formal groups.

Another example is the Hopkins-Ravenel chromatic convergence theorem. If we define

$$\mathcal{U}(n) = \mathcal{M}_{\mathrm{fg}} - \mathcal{M}(n+1)$$

to be the open substack classifying formal groups over schemes over $\mathbb{Z}_{(p)}$ of height at most n, then we get an ascending chain of open substacks

$$\mathbb{Q} \otimes \mathcal{M}_{\mathrm{fg}} \simeq \mathcal{U}(0) \subseteq \mathcal{U}(1) \subseteq \mathcal{U}(2) \subseteq \cdots \subseteq \mathbb{Z}_{(p)} \otimes \mathcal{M}_{\mathrm{fg}}.$$

This sequence is not exhaustive: the additive formal group of $\mathbb{F}_p$ has infinite height and does not give a point in any $\mathcal{U}(n)$.

Now let $G_n : \mathrm{Spec}(R_n) \to \mathcal{M}_{\mathrm{fg}}$ be a flat map classifying a formal group of exact height n; this gives a 2-periodic homology theory $E(R_n, G_n)$. Let $L_n(-)$ be the localization with respect to this homology theory. For example, L_0X is the rational localization and L_1X is the localization with respect to p-local K-theory. Then chromatic convergence [**57**] says that for a spectrum X there is a tower

$$\cdots \to L_2X \to L_1X \to L_0X$$

under X; furthermore, if $\pi_k X = 0$ for k sufficiently negative and $H_*(X, \mathbb{Z}_{(p)})$ is finitely generated as a graded $\mathbb{Z}_{(p)}$-module, then

$$X \longrightarrow \mathrm{holim} L_nX$$

is the localization with respect to $H_*(-, \mathbb{Z}_{(p)})$.

Next we might like to decompose the L_nX. For this we use the open inclusion $i : \mathcal{U}(n-1) \to \mathcal{U}(n)$ and its closed complement $\mathcal{H}(n) \subseteq \mathcal{U}(n)$. Recall $\mathcal{H}(n)$ classifies formal groups of exact height n and has a single geometric point given by any formal group Γ_n over a field $\mathbb{F}$ of characteristic p of exact height n. The classifying morphism $\Gamma_n : \mathrm{Spec}(\mathbb{F}) \to \mathcal{M}_{\mathrm{fg}}$ is not flat, but the homology theory $K(\mathbb{F}, \Gamma_n)$ exists nonetheless and is remarkably computable. These are the *Morava K-theories*. If we write $L_{K(n)}(-)$ to denote localization with respect to any of these theories (all the localizations are equivalent), then there is a homotopy pull-back square

$$\begin{array}{ccc} L_nX & \longrightarrow & L_{K(n)}X \\ \downarrow & & \downarrow \\ L_{n-1}X & \longrightarrow & L_{n-1}L_{K(n)}X. \end{array}$$

The square can be deduced from [**32**]; the paper [**30**] contains a detailed analysis of a conjecture on how these squares behave.

We are thus left with calculating the pieces $L_{K(n)}X$. Here the theory is actually fully realized. This is the subject of the next subsection.

3.3. Deformations and the local Hopkins-Miller Theorem

We will need the language of deformation theory at several points; here it is used to explain how one might compute $L_{K(n)}X$. Let $\mathcal{M}$ be a stack and $A_0/\mathbb{F}$ be an $\mathcal{M}$-object over a field $\mathbb{F}$. Recall that an Artin local ring $(R, \mathfrak{m})$ is a local ring with nilpotent maximal ideal $\mathfrak{m}$. If $q : R \to \mathbb{F}$ is a surjective morphism of rings, then a *deformation* of A_0 to R is an $\mathcal{M}$-object A and a pull-back diagram

$$\begin{array}{ccc} A_0 & \longrightarrow & A \\ \downarrow & & \downarrow \\ \mathrm{Spec}(\mathbb{F}) & \longrightarrow & \mathrm{Spec}(R). \end{array}$$

Deformations form a groupoid-valued functor $\mathrm{Def}_{\mathcal{M}}(\mathbb{F}, A_0)$ on an appropriate category of Artin local rings.

If Γ is a formal group of finite height n over a perfect field $\mathbb{F}$, then Lubin-Tate theory [**44**] says that the groupoid-valued functor $\mathrm{Def}_{\mathcal{M}_{\mathrm{fg}}}(\mathbb{F}, \Gamma)$ is discrete; that is, the natural map

$$\mathrm{Def}_{\mathcal{M}_{\mathrm{fg}}}(\mathbb{F}, \Gamma) \to \pi_0\mathrm{Def}_{\mathcal{M}_{\mathrm{fg}}}(\mathbb{F}, \Gamma) \stackrel{\mathrm{def}}{=} \text{Isomorphism classes in } \mathrm{Def}_{\mathcal{M}_{\mathrm{fg}}}(\mathbb{F}, \Gamma)$$

is an equivalence. Furthermore, $\pi_0\mathrm{Def}_{\mathcal{M}_{\mathrm{fg}}}(\mathbb{F}, \Gamma)$ is pro-represented by a complete local ring $R(\mathbb{F}, \Gamma)$; that is, there is a natural isomorphism

$$\pi_0\mathrm{Def}_{\mathcal{M}_{\mathrm{fg}}}(\mathbb{F}, \Gamma) \cong \mathrm{Spf}(R(\mathbb{F}, \Gamma)).$$

An appropriate choice of coordinate for the universal deformation of Γ over $R(\mathbb{F}, \Gamma)$ defines an isomorphism

$$W(\mathbb{F})[[u_1, \ldots, u_{n-1}]] \cong R(\mathbb{F}, \Gamma)$$

where $W(-)$ is the Witt vector functor. If $\mathrm{Aut}(\mathbb{F}, \Gamma)$ is the group of automorphisms of the pair $(\mathbb{F}, \Gamma)$, then $\mathrm{Aut}(\mathbb{F}, \Gamma)$ acts on $\mathrm{Def}(\mathbb{F}, \Gamma)$ and hence on $R(\mathbb{F}, \Gamma)$. The formal spectrum $\mathrm{Spf}(R(\mathbb{F}, \Gamma))$ with the action of $\mathrm{Aut}(\mathbb{F}, \Gamma)$ is called Lubin-Tate space; if we insist that $\mathbb{F}$ is algebraically closed it is independent of the choice of the pair $(\mathbb{F}, \Gamma)$, by Lazard's classification theorem. See [**22**] for the following result.

THEOREM 3.6 (Local Hopkins-Miller). — *There is a 2-periodic commutative ring spectrum $E(\mathbb{F}, \Gamma)$ with $E(\mathbb{F}, \Gamma)_0 \cong R(\mathbb{F}, \Gamma)$ and an associated formal group isomorphic to a universal deformation of Γ. Furthermore,*

1. *the space of all such commutative ring spectrum realizations of the universal deformation of $(\mathbb{F}, \Gamma)$ is contractible; and*
2. *the group $\mathrm{Aut}(\mathbb{F}, \Gamma)$ acts on $E(\mathbb{F}, \Gamma)$ through maps of commutative ring spectra.*

The assignment $(\mathbb{F},\Gamma)\mapsto E(\mathbb{F},\Gamma)$ is actually a functor from a category of height n formal groups to commutative ring spectra.

In [**17**] Devinatz and Hopkins show that when $\mathbb{F}$ contains enough roots of unity, there is a weak equivalence

$$L_{K(n)}X \simeq \operatorname{holim}_{\mathrm{Aut}(\mathbb{F},\Gamma)} E(\mathbb{F},\Gamma)\wedge X$$

for all finite CW spectra and, in particular, for $X=S$ [(3)]. This gives, for example, a spectral sequence

$$H^s(\mathrm{Aut}(\mathbb{F},\Gamma),\omega_G^{\otimes t}) \Longrightarrow \pi_{2t-s}L_{K(n)}S$$

where G is a universal deformation of Γ. If p is large with respect to n, this spectral sequence collapses and there are no extensions and the problem becomes purely algebraic – if not easy. For $n=1$ this happens if $p>2$ and, if $n=2$, we need $p>3$ and in these cases all the calculation have been done. See [**63**]. The case $p=2$ and $n=1$ is not hard; there has also been extensive calculation at $p=3$ and $n=2$. In both cases p-torsion in $\mathrm{Aut}(\mathbb{F},\Gamma)$ creates differentials. As a sample of the sort of large-scale periodic phenomena we see, I offer the following classical result of Adams [**1**]. Let $p>2$. Then $\pi_0 L_{K(1)}S=\pi_{-1}L_{K(1)}S=\mathbb{Z}_p$ and $\pi_k L_{K(1)}S$ is zero for all other k unless

$$k=2p^t s(p-1)-1, \qquad (s,p)=1$$

and then

$$\pi_k L_{K(1)}S=\mathbb{Z}/p^{t+1}.$$

All the elements in positive degree come from π_*S itself; in fact, the natural map $\pi_*S\to\pi_*L_{K(1)}S$ splits off the image of the classical J-homomorphism.

4. CONSTRUCTION AND DECONSTRUCTION ON $\bar{\mathcal{M}}_{ell}$.

4.1. A realization problem

If $X\to\bar{\mathcal{M}}_{ell}$ classifies a generalized elliptic curve $C\to X$, let $\mathscr{I}(e)\subseteq\mathcal{O}_C$ be the ideal sheaf defining the identity section $e:X\to C$. Then $\mathscr{I}(e)/\mathscr{I}(e)^2$ is an $e_*\mathcal{O}_X$-module sheaf and thus it determines a unique $\mathcal{O}_X$-module sheaf ω_C on X. This sheaf is locally free of rank 1. Even if C is singular, it is never singular at e and the smooth locus of C has the structure of an abelian group scheme of dimension 1 over X; hence, ω_C is isomorphic to the invariant differentials of C. The assignment

$$\omega_{ell}(C:X\longrightarrow\bar{\mathcal{M}}_{ell})=\omega_C$$

(3) There are technical issues arising from the fact that $\mathrm{Aut}(\mathbb{F},\Gamma)$ is a profinite group and that we must use continuous cohomology. Again, see [**17**].

defines an invertible sheaf on $\bar{\mathcal{M}}_{e\ell\ell}$.

More is true. If C_e denotes the formal completion of C at e, then C_e is a smooth 1-parameter formal Lie group and $\omega_C \cong \omega_{C_e}$. This data defines a morphism

$$q : \bar{\mathcal{M}}_{e\ell\ell} \longrightarrow \mathcal{M}_{\mathrm{fg}}$$

with the property that $q^*\omega = \omega_{e\ell\ell}$; for this reason I will simply write $\omega_{e\ell\ell}$ as ω, leaving the base stack to determine which sheaf I mean.

If C is a generalized elliptic curve over an affine scheme X and ω_C has a trivializing global section $\eta : \mathcal{O}_X \to \omega_C$, then there is a choice of closed immersion $C \to \mathbb{P}^2$ over X with e sent to $[0,1,0]$. This closed immersion is defined by a Weierstrass equation

$$Y^2Z + a_1XYZ + a_3YZ^2 = X^3 + a_2X^2Z + a_4XZ^2 + a_6Z^3$$

which is normally written as

$$y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6$$

since there is only one point where $Z = 0$. This closed immersion is not unique, but is adapted to η in the sense that $z = -x/y$ reduces to η in ω_C. Other choices of η and the associated adapted immersion produce a different immersion; all such immersions differ by a projective transformation. Given η we see that the invariant differentials for the associated formal group C_e form the trivial sheaf; hence C_e can be given a coordinate. Once we have chosen a coordinate t and an immersion adapted to η we have an equation

$$z = t + e_1t^2 + e_2t^3 + \cdots$$

since z is also a coordinate. Then we can uniquely specify the adapted immersion by requiring $e_i = 0$ for $i \leq 3$. From this we can conclude that the morphism $\bar{\mathcal{M}}_{e\ell\ell} \to \mathcal{M}_{\mathrm{fg}}$ is representable. This argument is due to Hopkins.

This morphism is also flat, by Landweber's criterion 3.4. Indeed, in this setting the global section

$$v_1 \in H^0(\mathbb{F}_p \otimes \bar{\mathcal{M}}_{e\ell\ell}, \omega^{\otimes p-1})$$

is the *Hasse invariant*. A generalized elliptic curve C over X over $\mathbb{F}_p$ is *ordinary* if $v_1(C) : \mathcal{O}_X \to \omega_C$ is an isomorphism; on the other hand, if $v_1(C) = 0$, then C is *supersingular* and automatically smooth. We define $\mathcal{M}_{\mathrm{ss}} \subseteq \mathbb{F}_p \otimes \bar{\mathcal{M}}_{e\ell\ell}$ to be the closed substack defined by the vanishing of v_1 and then, over $\mathcal{M}_{\mathrm{ss}}$, the global section

$$v_2 : \mathcal{O}_{\mathrm{ss}} \longrightarrow \omega^{\otimes p^2-1}$$

is an isomorphism. This is a rephrasing of the statement that formal group of a supersingular curve has exact height 2.

Remark 4.1 (The realization problem). — We can now ask the following question. Suppose $\mathscr{N} \to \mathscr{M}_{\mathrm{fg}}$ is a representable and flat morphism from a Deligne-Mumford stack to the moduli stack of formal groups. A realization of $\mathscr{N}$ is a derived Deligne-Mumford stack $(\mathscr{N}, \mathcal{O})$ with

$$\pi_k \mathcal{O} \cong \begin{cases} \omega_{\mathscr{N}}^{\otimes t}, & k = 2t; \\ \\ 0, & k = 2t+1. \end{cases}$$

Does $\mathscr{N}$ have a realization? If so, how many are there? Better, what is the homotopy type of the space of all realizations?

This question is naïve for a variety of reasons. A simple one is that if $\mathcal{O}$ exists, then $\pi_* \mathcal{O}$ will be a sheaf of graded commutative rings with power operations and, in general, there is not enough data in formal groups to specify those operations. In the case of the local Hopkins-Miller theorem 3.6 these operations are determined by the subgroup structure of the formal groups. See [**3**] and [**58**]. In the case of $\bar{\mathscr{M}}_{e\ell\ell}$, they will be determined by the Serre-Tate theorem and the subgroup structure of elliptic curves.

The work of Mark Behrens and Tyler Lawson [**10**] solves the realization problem for certain Shimura varieties, which are moduli stacks of highly structured abelian varieties. The extra structure is needed to get formal groups of higher heights.

The realization problem is essentially a p-adic question. If X is a scheme or a stack, let $j_n : X(p^n) \to X$ be the closed immersion defined by the vanishing of p^n. The formal completion of X at p is the colimit sheaf $X^\wedge_p = \mathrm{colim} X(p^n)$; as a functor, $X^\wedge_p$ is the restriction of X to rings in which p is nilpotent; thus $X^\wedge_p$ is a formal scheme over $\mathrm{Spec}(\mathbb{Z})^\wedge_p = \mathrm{Spf}(\mathbb{Z}_p)$. If $\mathcal{F}$ is a derived module sheaf on X, the derived completion of $\mathcal{F}$ is

$$\mathcal{F}_p = \mathrm{holim} (j_n)_* j_n^* \mathcal{F}.$$

Then if $\mathcal{F}$ is a derived sheaf of modules on $\mathscr{N}$, there is a homotopy pull-back square

$$\begin{array}{ccc} \mathcal{F} & \longrightarrow & \prod_p \mathcal{F}_p \\ \downarrow & & \downarrow \\ \mathbb{Q} \otimes \mathcal{F} & \longrightarrow & \mathbb{Q} \otimes (\prod_p \mathcal{F}_p) \end{array} \tag{7}$$

where $\mathbb{Q} \otimes (-)$ is the rational localization. Here I am regarding $\mathcal{F}_p$ as a sheaf on X; however, $\mathcal{F}_p$ is determined by its restriction to X_p.

The category of commutative ring spectra over the rational numbers is Quillen equivalent to the category of differential graded algebras over $\mathbb{Q}$; hence, many questions in derived algebraic geometry over $\mathbb{Q}$ become classical.

Example 4.2 ($\bar{\mathcal{M}}_{e\ell\ell}$ over $\mathbb{Q}$). — If S_* is a graded ring with $S_0 = R$, then the grading defines an action of the multiplicative group $\mathbb{G}_m$ on $Y = \mathrm{Spec}(S_*)$. Define $\mathrm{Proj}(S_*)$ to be the quotient stack over R of this action on $Y - \{0\}$ where $\{0\} \subseteq Y$ is the closed subscheme defined by the ideal of elements of positive degree. Then

$$\mathbb{Q} \otimes \bar{\mathcal{M}}_{e\ell\ell} \cong \mathrm{Proj}(\mathbb{Q}[c_4, c_6])$$

and the graded sheaf of rings $\mathbb{Q}[\omega^{\pm 1}]$ is a dga with trivial differential. This defines the derived scheme $\mathbb{Q}\otimes\bar{\mathcal{M}}_{e\ell\ell}$ and the sheaf $\mathbb{Q}\otimes\mathcal{O}$. Similarly $\mathbb{Q}\otimes(\bar{\mathcal{M}}_{e\ell\ell})_p = \mathrm{Proj}(\mathbb{Q}_p[c_4, c_6])$ and the map across the bottom in Diagram (7) will be the obvious one. I note that we can define $\mathbb{Z}[1/6]\otimes\bar{\mathcal{M}}_{e\ell\ell}$ as $\mathrm{Proj}(\mathbb{Z}[1/6][c_4, c_6])$; however, we cannot define the derived scheme this way as the homotopy theory of commutative ring spectra over $\mathbb{Z}[1/6]$ is not equivalent to a category of dgas.

4.2. Lurie's theorem and p-divisible groups

In this section, we pick a prime p and work over $\mathrm{Spf}(\mathbb{Z}_p)$; that is, p is implicitly nilpotent in all our rings. I will leave this out of the notation.

Definition 4.3. — *Let R be a ring and G a sheaf of abelian groups on R-algebras. Then G is a* p-divisible group *of* height n *if*

(1) *$p^k : G \to G$ is surjective for all k;*
(2) *$G(p^k) = \mathrm{Ker}(p^k : G \to G)$ is a finite and flat group scheme over R of rank p^{kn};*
(3) *$\mathrm{colim} G(p^k) \cong G$.*

Remark 4.4. — 1) If G is a p-divisible group, then completion at $e \in G$ gives an abelian formal group $G_{\mathrm{for}} \subseteq G$, not necessarily of dimension 1. The quotient G/G_{for} is étale over R; thus we get a natural short exact sequence

$$0 \to G_{\mathrm{for}} \to G \to G_{\mathrm{et}} \to 0.$$

This is split over fields, but not in general.

2) If C is a smooth elliptic curve, then $C(p^\infty) = \mathrm{colim}\ C(p^n)$ is p-divisible of height 2 with formal part of dimension 1.

3) If G is a p-divisible group over a scheme X, the function which assigns to each geometric point x of X the height of the fiber G_x of G at x is constant. This is not true of formal groups, as the example of elliptic curves shows. Indeed, if G is p-divisible of height n with G_{for} of dimension 1, then the height of G_{for} can be any integer between 1 and n.

Definition 4.5. — *Let $\mathcal{M}_p(n)$ be the moduli stack of p-divisible groups of height n and with* $\dim G_{\mathrm{for}} = 1$.

Remark 4.6. — The stack $\mathcal{M}_p(n)$ is not an algebraic stack, but rather pro-algebraic in the same way $\mathcal{M}_{\mathrm{fg}}$ is pro-algebraic. This can be deduced from the material in the first chapter of **[52]**.

There is a morphism of stacks $\mathcal{M}_p(n) \to \mathcal{M}_{\mathrm{fg}}$ sending G to G_{for}. This morphism is *not* representable. By definition, there is a factoring of this map as

$$\mathcal{M}_p(n) \longrightarrow \mathcal{U}(n) \longrightarrow \mathcal{M}_{\mathrm{fg}}$$

through the open substack of formal groups of height at most n.

We now can state Lurie's realization result **[46]**. Since we are working over $\mathbb{Z}_p$, one must take care with the hypotheses: the notions of algebraic stack and étale must be the appropriate notions over $\mathrm{Spf}(\mathbb{Z}_p)$.

THEOREM 4.7 (Lurie). — *Let $\mathcal{M}$ be a Deligne-Mumford stack equipped with a formally étale morphism*

$$\mathcal{M} \longrightarrow \mathcal{M}_p(n).$$

Then the realization problem for the composition

$$\mathcal{M} \longrightarrow \mathcal{M}_p(n) \longrightarrow \mathcal{M}_{\mathrm{fg}}$$

has a canonical solution; that is, the space of all solutions has a preferred basepoint.

It is worth emphasizing that this theorem uses the local Hopkins-Miller theorem 3.3 in an essential way.

Example 4.8 (Serre-Tate theory). — As an addendum to this theorem, Lurie shows that the morphism $\epsilon : \mathcal{M} \to \mathcal{M}_p(n)$ is formally étale if it satisfies the Serre-Tate theorem. This applies to the open substack $\mathcal{M}_{1,1}$ of $\bar{\mathcal{M}}_{ell}$ and we recover the main theorem 1.2, at least for smooth elliptic curves. Then, in **[46]**, Lurie gives an argument extending the result to the compactification $\bar{\mathcal{M}}_{ell}$.

The Serre-Tate theorem asserts an equivalence of deformation groupoids; compare the Lubin-Tate result of Section 3.3. Let $\mathcal{M}$ be a stack over $\mathcal{M}_p(n)$ and $A_0/\mathbb{F}$ be an $\mathcal{M}$-object over a field $\mathbb{F}$, necessarily of characteristic p since we are working over $\mathrm{Spf}(\mathbb{Z}_p)$. Then the deformations form a groupoid functor $\mathrm{Def}_{\mathcal{M}}(\mathbb{F}, A_0)$ on Artin local rings. The Serre-Tate theorem holds if the evident morphism

$$\mathrm{Def}_{\mathcal{M}}(\mathbb{F}, A_0) \longrightarrow \mathrm{Def}_{\mathcal{M}_p(n)}(\mathbb{F}, \epsilon A_0)$$

is an equivalence. This result holds for elliptic curves, but actually in much wider generality. See **[52]**.

Remark 4.9 (Deformations of p-divisible groups). — We discussed the deformation theory of formal groups and Lubin-Tate theory in Section 3.3; the theory is very similar for p-divisible groups. Let G be a p-divisible group over an algebraically closed field $\mathbb{F}$. Then we have the split short exact sequence

$$0 \to G_{\text{for}} \to G \to G_{\text{et}} \to 0.$$

Since G_{et} has a unique deformation up to isomorphism, by the definition of étale, the deformations of G are determined by the deformations of G_{for} and an extension class. From this it follows that the groupoid-valued functor $\text{Def}_{\mathcal{M}_p(n)}(\mathbb{F}, G)$ is discrete and $\pi_0\text{Def}_{\mathcal{M}_p(n)}(\mathbb{F}, G)$ is pro-represented by

$$R(\mathbb{F}, G_{\text{for}})[[t_1, \cdots, t_{n-h}]] \cong W(\mathbb{F})[[u_1, \cdots, u_{h-1}, t_1, \cdots, t_{n-h}]].$$

Note that this is always a power series in $n-1$ variables. Using this remark it is possible to give a local criterion for when a morphism of stacks $\mathcal{M} \to \mathcal{M}_p(n)$ is étale. It is in this guise that Lurie's theorem appears in **[10]**.

4.3. Decomposing the structure sheaf

Both the original Hopkins-Miller argument and Lurie's derived algebraic geometry construction of the derived $\bar{\mathcal{M}}_{ell}$ rely on a decomposition of the moduli stack of elliptic curve into its ordinary and supersingular components. Specifically, let $\mathcal{M}_{\text{ss}} \subseteq \bar{\mathcal{M}}_{ell}$ be the closed substack of supersingular curves defined by the vanishing of both a fixed prime p and the Hasse invariant. There are inclusion of closed substacks

$$\mathcal{M}_{\text{ss}} \subseteq \mathbb{F}_p \otimes \bar{\mathcal{M}}_{ell} \subseteq \bar{\mathcal{M}}_{ell};$$

from this we get an inclusion of formal substacks $\hat{\mathcal{M}}_{\text{ss}} \subseteq (\bar{\mathcal{M}}_{ell})^\wedge_p$. Let $\hat{\mathcal{M}}_{\text{ord}} \subseteq (\bar{\mathcal{M}}_{ell})^\wedge_p$ be the open complement of $\hat{\mathcal{M}}_{\text{ss}}$; thus, $\hat{\mathcal{M}}_{\text{ord}}$ classifies ordinary elliptic curves over rings R in which p is nilpotent. If X is a sheaf on $\bar{\mathcal{M}}_{ell}$, we obtain completions of X from each of these formal stacks, which we write L_pX, $L_{\text{ss}}X$, and $L_{\text{ord}}X$, respectively.

THEOREM 4.10. — *There is a homotopy pull-back diagram of sheaves of E_∞-ring spectra on $\bar{\mathcal{M}}_{ell}$*

$$\begin{array}{ccc} L_p\mathcal{O} & \longrightarrow & L_{\text{ss}}\mathcal{O} \\ \downarrow & & \downarrow \\ L_{\text{ord}}\mathcal{O} & \longrightarrow & L_{\text{ord}}L_{\text{ss}}\mathcal{O}. \end{array}$$

This statement belies history: the original Hopkins-Miller proof used obstruction theory to build the sheaves $L_{\text{ord}}\mathcal{O}$ and $L_{\text{ss}}\mathcal{O}$ and the map across the bottom of this diagram, thus building $L_p\mathcal{O}$. Lurie, on the other hand, produces a candidate for $L_p\mathcal{O}$, and then must show it has the right homotopy type. This is done by analyzing the pieces in this square separately. In either case, the pieces $L_{\text{ord}}\mathcal{O}$ and $L_{\text{ss}}\mathcal{O}$ have intrinsic

interest and the calculations of their homotopy types call on classical calculations with modular forms.

4.4. The supersingular locus

The description of $L_{\mathrm{ss}}\mathcal{O}$ begins with an alternative description of the stack $\mathcal{M}_{\mathrm{ss}}$ of supersingular curves. This is standard theory for elliptic curves; see **[37]**, Chapter 12. Since the vanishing of the Hasse invariant of C depends only on the isomorphism class of the curve C, it depends only on the j-invariant of C as well. Write

$$j : \mathbb{F}_p \otimes \bar{\mathcal{M}}_{ell} \longrightarrow \mathbb{F}_p \otimes \mathbb{P}^1 = \mathbb{P}^1$$

with $j(C) = [c_4^3(C), \Delta(C)]$ and call the point $[1,0] \in \mathbb{P}^1$ the point at infinity. If $\Delta(C)$ is invertible – that is, C is an elliptic curve – we also write

$$j(C) = c_4^3(C)/\Delta(C) \in \mathbb{A}^1_j = \mathbb{P}^1 - \{\infty\}.$$

A point in $\mathbb{P}^1$ will be ordinary if it is the image of an ordinary curve; supersingular otherwise.

THEOREM 4.11. — *Fix a prime p. The point at ∞ in $\mathbb{P}^1$ is ordinary and there is a separable polynomial $\Phi(j) \in \mathbb{F}_p[j]$ so that C is supersingular if and only if $j(C)$ is a root of Φ. All the roots of Φ lie in $\mathbb{F}_{p^2}$.*

For example, if $p = 2$ or 3, then $\Phi(j) = j$. The degree of Φ is $[(p-1)/12] + \epsilon_p$ where ϵ_p is 0, 1, or 2 depending on the prime.

PROPOSITION 4.12. — *Let $R_{\mathrm{ss}} = \mathbb{F}_p[j]/(\Phi(j))$. There is a 2-category pull-back*

$$\begin{array}{ccc} \mathcal{M}_{\mathrm{ss}} & \longrightarrow & \mathbb{F}_p \otimes \bar{\mathcal{M}}_{ell} \\ \downarrow & & \downarrow j \\ \mathrm{Spec}(R_{\mathrm{ss}}) & \longrightarrow & \mathbb{P}^1. \end{array}$$

Furthermore, there is a supersingular curve C_{ss} over $\mathbb{F}_{p^2} \otimes R_{\mathrm{ss}}$ which gives an equivalence of stacks

$$B\mathrm{Aut}(C_{\mathrm{ss}}) \simeq \mathbb{F}_{p^2} \otimes \mathcal{M}_{\mathrm{ss}}.$$

Here I have written $\mathrm{Aut}(C_{\mathrm{ss}})$ for $\mathrm{Aut}(C_{\mathrm{ss}}/(\mathbb{F}_{p^2} \otimes R_{\mathrm{ss}}))$. If G is a group scheme, BG is the moduli stack of G-torsors.

To be concrete, write $C_{\mathrm{ss}} = \coprod C_a$ where C_a is a representative for the isomorphism class of supersingular curves over $\mathbb{F}_{p^2}$ with $j(C_a) = a$. Then

$$\mathbb{F}_{p^2} \otimes \mathcal{M}_{\mathrm{ss}} \simeq \coprod B\mathrm{Aut}(C_a/\mathbb{F}_{p^2}). \tag{8}$$

The algebraic groups $\mathrm{Aut}(C_a/\mathbb{F}_{p^2})$ are all known. See **[64]**, among many other sources. The difficulties and the interest lie at the primes 2 and 3, where the $\mathbb{F}_{p^2}$ points of $\mathrm{Aut}(C_0/\mathbb{F}_p)$ have elements of order p. For example, if $p = 3$,

$$\mathrm{Aut}(C_0/\mathbb{F}_9) \cong \mu_4 \rtimes \mathbb{Z}/3,$$

where μ_4 is the 4-th roots of unity.

Let $C : \mathrm{Spec}(B) \to \bar{\mathcal{M}}_{e\ell\ell}$ be étale and $I = I(B, C) \subseteq B$ the ideal generated by p and the Hasse invariant. Then B/I is a separable $\mathbb{F}_p$-algebra and q^*C_e is a formal group of exact height 2. The evident extension of the local Hopkins-Miller Theorem 3.6 to separable algebras gives a sheaf $\mathscr{E}_{\mathrm{ss}}$ of commutative ring spectra with

$$\mathscr{E}_{\mathrm{ss}}(C : \mathrm{Spec}(B) \to \bar{\mathcal{M}}_{e\ell\ell}) \simeq E(B/I, q^*C_e)$$

and we have $L_{\mathrm{ss}}\mathscr{O} \simeq \mathscr{E}_{\mathrm{ss}}$.

It is straightforward from here to understand the homotopy type of the global sections of $L_{\mathrm{ss}}\mathscr{O}$; indeed, we have

$$R\Gamma L_{\mathrm{ss}}\mathscr{O} \simeq \mathrm{holim}_G E(\mathbb{F}_{p^2} \otimes R_{\mathrm{ss}}, (C_{\mathrm{ss}})_e)$$

where $G = \mathrm{Gal}(\mathbb{F}_{p^2}/\mathbb{F}_p) \rtimes \mathrm{Aut}(C_{\mathrm{ss}})$.

The homotopy groups of this spectrum have been computed. If $p > 5$ this is fairly easy. If $p = 3$ the most explicit source is **[21]**, and if $p = 2$ it is implicit in **[28]** although that source needs to be combined with **[7]** to get complete answers.

4.5. The ordinary locus

For the sheaf $L_{\mathrm{ord}}\mathscr{O}$ we use the map $\bar{\mathcal{M}}_{e\ell\ell} \to \mathcal{M}_{\mathrm{fg}}$ classifying the associated formal group to make preliminary computations. By construction, this morphism restricts to a morphism $\hat{\mathcal{M}}_{\mathrm{ord}} \to \mathcal{U}(1)$ to the open substack of formal groups of height 1. The map $\mathcal{U}(1) \to \mathrm{Spf}(\mathbb{Z}_p)$ is formally étale and has section $g : \mathrm{Spf}(\mathbb{Z}_p) \to \mathcal{U}(1)$ classifying the multiplicative formal group $\hat{\mathbb{G}}_m$. The map g is pro-Galois with Galois group $\mathrm{Aut}(\hat{\mathbb{G}}_m) = \mathbb{Z}_p^\times$, the units in the p-adics. We use this cover not to define $L_{\mathrm{ord}}\mathscr{O}$ directly, but to first specify the resulting homology sheaf $K_*\mathscr{O}$, where K_* is p-complete complex K-theory[(4)]. From this we then can construct $L_{\mathrm{ord}}\mathscr{O}$.

Let R be a p-complete ring and let $C : \mathrm{Spec}(R) \to \bar{\mathcal{M}}_{e\ell\ell}$ be an étale morphism. Consider the following diagram, where both squares are 2-category pull-backs

$$\begin{array}{ccccc} \mathrm{Spf}K_0\mathscr{O}(R, C) & \longrightarrow & \mathrm{Spf}(V) & \longrightarrow & \mathrm{Spf}(\mathbb{Z}_p) \\ \downarrow & & \downarrow & & \downarrow{\scriptstyle \hat{\mathbb{G}}_m} \\ \mathrm{Spec}(R) & \longrightarrow & \bar{\mathcal{M}}_{e\ell\ell} & \longrightarrow & \mathcal{M}_{\mathrm{fg}}. \end{array}$$

(4) Because we are working p-adically, $K_*X = \pi_* L_{K(1)} K \wedge X$, a completion of the usual K-theory. The issues here are very technical, but carefully worked out in **[32]**.

This defines $K_0\mathcal{O}(R,C)$ and the ring V. The latter ring is Katz's ring of *divided congruences* **[36]** – and, as it turns out, $K_0\mathbf{tmf}$. Note that $\mathrm{Spf}(V)$ solves the moduli problem which assigns to each ring A with p nilpotent the set of pairs (C,ϕ) where C is an elliptic curve over A and $\phi : \hat{\mathbb{G}}_m \to C_e$ is an isomorphism. By construction $V \to K_0\mathcal{O}(R,C)$ is étale. We extend the resulting sheaf of rings $K_0\mathcal{O}$ to a graded sheaf $K_*\mathcal{O}$ by twisting by the various powers of the sheaf of invariant differentials, as in Equation (3).

The sheaf of rings $K_0\mathcal{O}$ has a great deal of structure. First, it comes equipped with an action of $\mathbb{Z}_p^\times$ from the automorphisms of $\hat{\mathbb{G}}_m$. This gives the action of the Adams operations in K-theory. Second, and more subtly, there is an extra ring operation ψ which is a lift of the Frobenius and commutes with the action of $\mathrm{Aut}(\mathbb{G}_m)$. To construct ψ, we use that if C is an ordinary elliptic curve, then the kernel of $p : C_e \to C_e$ defines a canonical subgroup $C(p) \to C$ of order p. Then $\psi : V \to V$ is defined by specifying the natural transformation on the functor V:

$$\psi(C, \phi : \hat{\mathbb{G}}_m \to C) = (C/C(p), \hat{\mathbb{G}}_m \cong \hat{\mathbb{G}}_m/\hat{\mathbb{G}}_m(p) \xrightarrow{\bar{\phi}} (C/C(p))_e).$$

If C is an ordinary curve over an $\mathbb{F}_p$-algebra A, then $C/C(p) = C^{(p)}$; this explains why $\psi : V \to V$ is a lift of the Frobenius. To extend ψ to all of $K_0\mathcal{O}$, note that since $V \to K_0\mathcal{O}(R,C)$ is étale there is a unique morphism $\psi : K_0\mathcal{O}(R,C) \to K_0\mathcal{O}(R,C)$ lifting the Frobenius and extending ψ on V.

Since $K_0\mathcal{O}$ is torsion-free, $\psi(x) = x^p + p\theta(x)$ for some unique operator θ. All this structure extends in a unique way to $K_*\mathcal{O}$, making this a sheaf of *theta-algebras*. By McClure's work **[13]**, this is exactly the algebraic structure supported by K_*X when X is a commutative ring spectrum. The existence of $L_{\mathrm{ord}}\mathcal{O}$ is now guaranteed by the following result.

THEOREM 4.13. — *The space of all sheaves of $L_{K(1)}$-local ring spectra X with $K_*X \cong K_*\mathcal{O}$ as sheaves of theta-algebras is non-empty and connected.*

Any of these sheaves – they are all homotopy equivalent – gives a model for $L_{\mathrm{ord}}\mathcal{O}$. Lurie's work does better: it gives a canonical model.

Originally, Theorem 4.13 was proved by an obstruction theory argument; the obstruction groups are computed using an appropriate cotangent complex for the sheaf $K_0\mathcal{O}$. This argument uses in an essential way that $\bar{\mathcal{M}}_{ell}$ is smooth of dimension 1 over $\mathbb{Z}$. At the crucial primes 2 and 3 there is a very elegant construction originating with Hopkins of $L_{K(1)}\mathbf{tmf}$ itself which short-circuits the obstruction theory. See the paper by Laures **[41]**.

Remark 4.14. — The homotopy groups of the derived global sections $R\Gamma(L_{\mathrm{ord}}\mathcal{O}) \simeq L_{K(1)}\mathbf{tmf}$ have been computed. For example, at the prime 2, we have

$$\pi_* R\Gamma(L_{\mathrm{ord}}\mathcal{O}) \cong (\mathbb{Z}_2[1/j])^\wedge_2[\eta, v, b^{\pm 1}]/I$$

where the completed polynomial ring on the inverse of the j-invariant is in degree 0 and η, v, and b have degrees 1, 4, and 8 respectively. The ideal I is generated by the relations

$$2\eta = \eta^3 = v\eta = 0$$
$$v^2 = 2b.$$

There is a map $R\Gamma(L_{\mathrm{ord}}\mathcal{O}) \to KO$ to the spectrum of 2-completed real K-theory which, in homotopy, is the quotient by the ideal generated by $1/j$.

Remark 4.15. — To complete the construction of $L_p\mathcal{O}$, we must produce the map $L_{\mathrm{ord}}\mathcal{O} \to L_{\mathrm{ord}}L_{\mathrm{ss}}\mathcal{O}$. To do this, we calculate $K_*L_{\mathrm{ss}}\mathcal{O}$ as a theta-algebra and again use obstruction theory. The complication is that the lift ψ of the Frobenius on $K_*L_{\mathrm{ss}}\mathcal{O}$ is determined by the E_∞-ring structure on $L_{\mathrm{ss}}\mathcal{O}$. For the algebra to work, we need to check that this is the same lift as that determined by the subgroup structure of an appropriate elliptic curve. To obtain the elliptic curve, we use the Serre-Tate theorem to show that the universal deformation of the formal group of a supersingular curve is actually the p-divisible group of an elliptic curve. Then we can apply the general theory of power operations in complex orientable homology theories as developed by Ando [**3**] and Rezk [**58**]. I have to thank Charles Rezk for explaining this to me. Lurie's construction avoids this question, because the map already exists.

4.6. Derived modular forms and duality

The descent spectral sequence

$$H^s(\bar{\mathcal{M}}_{e\ell\ell}, \omega^{\otimes t}) \Longrightarrow \pi_{2t-s}\mathbf{tmf} \tag{9}$$

has been completely calculated and the homotopy groups of **tmf** exhibit a very strong form of duality not present in the cohomology groups of $\bar{\mathcal{M}}_{e\ell\ell}$. Let me try to give some flavor of the results. The first observation is that, by [**14**], we have an isomorphism

$$M_* = \mathbb{Z}[c_4, c_6, \Delta] \cong H^0(\bar{\mathcal{M}}_{e\ell\ell}, \omega^{\otimes *})$$

from the ring of modular forms of level 1 to the global sections.

Next, the stack $\bar{\mathcal{M}}_{e\ell\ell}$ is smooth of dimension 1 over $\mathbb{Z}$ and the cotangent sheaf is identified by the isomorphism

$$\Omega_{\bar{\mathcal{M}}_{e\ell\ell}/\mathbb{Z}} \cong \omega^{\otimes -10}.$$

I learned this from Hopkins, and it can be deduced from [**37**], Chapter 10. There is an isomorphism $H^1(\bar{\mathcal{M}}_{e\ell\ell}, \omega^{\otimes -10}) \cong \mathbb{Z}$ and a multiplication homomorphism

$$H^1(\bar{\mathcal{M}}_{e\ell\ell}, \omega^{\otimes -k-10}) \otimes H^0(\bar{\mathcal{M}}_{e\ell\ell}, \omega^{\otimes k}) \to H^0(\bar{\mathcal{M}}_{e\ell\ell}, \omega^{\otimes -10}) = \mathbb{Z}. \tag{10}$$

If we invert the primes 2 and 3, $H^s(\bar{\mathcal{M}}_{e\ell\ell}, \omega^{\otimes *}) = 0$ for $s > 1$ and the multiplication of Equation (10) becomes a perfect pairing. Thus, if we invert 6 the spectral sequence of (9) collapses, there can be no possible extensions, and both the coherent cohomology of $\bar{\mathcal{M}}_{e\ell\ell}$ and the homotopy groups of **tmf** exhibit a strong form of Serre-type duality. This is not very surprising as, when 6 is inverted, $\bar{\mathcal{M}}_{e\ell\ell}$ is isomorphic to the projective stack obtained from the graded ring M_*. See Remark 4.2.

Over the integers, however, the behavior is more subtle. The bookkeeping at the prime 2 is difficult, so let me say what happens when 2 is inverted. The presence of an element of order 3 in the automorphisms of the supersingular curve at the prime 3 forces higher cohomology. There is an injection

$$M_*^* \stackrel{\text{def}}{=} M_*[\alpha, \beta]/R \longrightarrow H^*(\bar{\mathcal{M}}_{e\ell\ell}, \omega^{\otimes *}) \tag{11}$$

where $\alpha \in H^1(\bar{\mathcal{M}}_{e\ell\ell}, \omega^{\otimes 2})$ and $\beta \in H^2(\bar{\mathcal{M}}_{e\ell\ell}, \omega^{\otimes 6})$ and R is the ideal of relations

$$3\alpha = 3\beta = \alpha^2 = c_i\alpha = c_i\beta = 0.$$

The element α is the image of the Hasse invariant under the boundary map

$$H^0(\mathbb{F}_3 \otimes \bar{\mathcal{M}}_{e\ell\ell}, \omega^{\otimes 2}) \to H^1(\bar{\mathcal{M}}_{e\ell\ell}, \omega^{\otimes 2})$$

and β is the Massey product $\langle \alpha, \alpha, \alpha \rangle$. Both α and β arise from the homotopy groups of spheres. The ring M_*^* is actually the coherent cohomology of a moduli stack of Weierstrass curves. See [**7**].

We now define a bigraded M_*^*-module C_*^* by the short exact sequence of M_*^*-modules

$$0 \to M_*^* \to H^*(\bar{\mathcal{M}}_{e\ell\ell}, \omega^{\otimes *}) \to C_*^* \to 0. \tag{12}$$

We then need to write down the M_*^*-module C_*^*. It will turn out that in any given bidegree one of (and possibly both) M_*^* and C_*^* is zero, so we get an unambiguous splitting of bigraded groups

$$H^*(\bar{\mathcal{M}}_{e\ell\ell}, \omega^{\otimes *}) = M_*^* \oplus C_*^*.$$

It is not quite split as M_*^*-modules as multiplication by Δ will link the two pieces.

To describe the modules C_*^*, we first note that the pairing of (10) is no longer a perfect pairing with only 2 inverted: it only induces an injection. First one computes $C_t^1 = H^1(\bar{\mathcal{M}}_{e\ell\ell}, \omega^{\otimes t})$ if $t < 0$ and $C_t^1 = 0$ otherwise. Now let $K_*^1 \subseteq C_*^1$ be the kernel of

multiplication by β. Then, the pairing induces a morphism of short exact sequences of graded M_*-modules

$$\begin{array}{ccccccccc} 0 & \longrightarrow & K_*^1 & \longrightarrow & C_*^1 & \longrightarrow & \mathrm{Hom}_{\mathbb{F}_3}(\mathbb{F}_3[\Delta], \mathbb{F}_3) & \longrightarrow & 0 \\ & & \cong\downarrow & & \downarrow & & \downarrow & & \\ 0 & \longrightarrow & \mathrm{Hom}_{\mathbb{Z}}(M_*, 3\mathbb{Z}[1/2]) & \longrightarrow & \mathrm{Hom}_{\mathbb{Z}}(M_*, \mathbb{Z}[1/2]) & \longrightarrow & \mathrm{Hom}_{\mathbb{F}_3}(M_*/3, \mathbb{F}_3) & \longrightarrow & 0 \end{array}$$

where the right vertical inclusion is induced by the quotient by the ideal (c_4, c_6) of M_*.

To complete the description of C_*^* we extend the top short exact sequence of our diagram of M^*-modules to a short exact sequence of M_*^*-modules by

$$0 \to K_*^1 \to C_*^* \to M_*^* \otimes_{M_*} \mathrm{Hom}_{\mathbb{F}_3}(\mathbb{F}_3[\Delta], \mathbb{F}_3) \to 0.$$

The extension of M_*^*-modules in the exact sequence (12) is now determined by the requirement that multiplication by Δ is surjective in positive cohomological degrees. In these degrees multiplication by c_4 and c_6 is necessarily zero, for degree reasons.

There must be differentials in the descent spectral sequence (9). A classical relation, due to Toda, states that $\alpha\beta^2 = 0$ in homotopy; this forces $d_5(\Delta) = \pm\alpha\beta^2$ in the spectral sequence. This differential and elementary considerations dictate the entire spectral sequence. Let $DM_* \subset M_*$ be the subring generated by $c_4, c_6, 3\Delta, 3\Delta^2$, and Δ^3. The inclusion $DM_* \subseteq \mathbf{tmf}_*$ extends to a split inclusion

$$DM_*^* = DM_*[\alpha, \beta, x]/DR \subseteq \mathbf{tmf}_*$$

where DR is the ideal of relations

$$\begin{gathered} 3\alpha = 3\beta = 3x = c_i\alpha = c_i\beta = c_ix = 0 \\ \alpha^2 = \beta^5 = \alpha\beta^2 = x\beta^2 = 0 \\ \alpha x = \beta^3. \end{gathered}$$

The class x is the Toda bracket $\langle \alpha, \alpha, \beta^2 \rangle$ and it is detected by $\alpha\Delta$ in the spectral sequence. The inclusion $DM_*^* \subseteq \mathbf{tmf}_*$ of DM_*^*-modules has cokernel DC_*^*. There is a short exact sequence of $DM^{*,*}$-modules

$$0 \to K_*^1 \to DC_*^* \to DM_*^* \otimes_{DM_*} \mathrm{Hom}_{\mathbb{F}_3}(\mathbb{F}_3[\Delta^3], \mathbb{F}_3) \to 0.$$

Note that there is a quotient $DM_* \to \mathbb{F}_3[\Delta^3]$; this defines the module structure needed for the tensor product.

Remark 4.16 (Duality for **tmf**). — At this point a new duality emerges – one not apparent before completing these homotopy theory calculations. A first remarkable

feature is that DC^*_* is in degrees less than -20 and DM^*_* is concentrated in non-negative degrees. In particular DM^*_* is the homotopy groups of the connected cover of $\mathbf{tmf}$.

A second feature is that DM^*_* and DC^*_* are almost dual as DM^*_*-modules. There are a number of ways to make this precise; a simple one is to say that for all primes p there is a homomorphism $\mathbf{tmf}_{-21} \to \mathbb{F}_p$ so the induced map given by the ring multiplication

$$\pi_k(\mathbf{tmf}/p) \to \mathrm{Hom}(\pi_{-k-21}\mathbf{tmf}/p, \mathbb{F}_p)$$

is an isomorphism. Here $\mathbf{tmf}/p$ is the derived global sections of the topological structure sheaf on $\mathbb{F}_p \otimes \bar{\mathcal{M}}_{ell}$.

This duality has an elegant homotopy theoretic explanation, given by Mahowald and Rezk [**49**]; in that source, there is also a simple explanation of the number -21. In derived geometry, this number surely appears because the dualizing sheaf is $\omega^{\otimes -10}$. However, I know of no explanation for the duality from the point of view of derived algebraic geometry.

REFERENCES

[1] J. F. Adams – "On the groups $J(X)$. IV", *Topology* **5** (1966), p. 21–71.

[2] ———, *Stable homotopy and generalised homology*, University of Chicago Press, 1974.

[3] M. Ando – "Isogenies of formal group laws and power operations in the cohomology theories E_n", *Duke Math. J.* **79** (1995), p. 423–485.

[4] M. Ando, M. J. Hopkins & C. Rezk – "Multiplicative orientations of KO-theory and of the spectrum of topological modular forms", preprint `http://www.math.uiuc.edu/~mando/papers/koandtmf.pdf`.

[5] M. Ando, M. J. Hopkins & N. P. Strickland – "Elliptic spectra, the Witten genus and the theorem of the cube", *Invent. Math.* **146** (2001), p. 595–687.

[6] ———, "The sigma orientation is an H_∞ map", *Amer. J. Math.* **126** (2004), p. 247–334.

[7] T. Bauer – "Computation of the homotopy of the spectrum `tmf`", in *Groups, homotopy and configuration spaces*, Geom. Topol. Monogr., vol. 13, Geom. Topol. Publ., Coventry, 2008, p. 11–40.

[8] M. Behrens – "A modular description of the $K(2)$-local sphere at the prime 3", *Topology* **45** (2006), p. 343–402.

[9] ———, "Congruences between modular forms given by the divided β family in homotopy theory", *Geom. Topol.* **13** (2009), p. 319–357.

[10] M. Behrens & T. Lawson – "Topological automorphic forms", to appear in *Mem. A.M.S.*

[11] R. E. Borcherds – "Automorphic forms on $O_{s+2,2}(\mathbf{R})$ and infinite products", *Invent. Math.* **120** (1995), p. 161–213.

[12] A. K. Bousfield & E. M. Friedlander – "Homotopy theory of Γ-spaces, spectra, and bisimplicial sets", in *Geometric applications of homotopy theory (Proc. Conf., Evanston, Ill., 1977), II*, Lecture Notes in Math., vol. 658, Springer, 1978, p. 80–130.

[13] R. R. Bruner, J. P. May, J. E. McClure & M. Steinberger – *H_∞ ring spectra and their applications*, Lecture Notes in Math., vol. 1176, Springer, 1986.

[14] P. Deligne – "Courbes elliptiques : formulaire d'après J. Tate", in *Modular functions of one variable, IV (Proc. Internat. Summer School, Univ. Antwerp, Antwerp, 1972)*, Lecture Notes in Math., vol. 476, Springer, 1975, p. 53–73.

[15] P. Deligne & D. Mumford – "The irreducibility of the space of curves of given genus", *Publ. Math. I.H.É.S.* **36** (1969), p. 75–109.

[16] P. Deligne & M. Rapoport – "Les schémas de modules de courbes elliptiques", in *Modular functions of one variable, II (Proc. Internat. Summer School, Univ. Antwerp, Antwerp, 1972)*, Lecture Notes in Math., vol. 349, Springer, 1973, p. 143–316.

[17] E. S. Devinatz & M. J. Hopkins – "Homotopy fixed point spectra for closed subgroups of the Morava stabilizer groups", *Topology* **43** (2004), p. 1–47.

[18] E. S. Devinatz, M. J. Hopkins & J. H. Smith – "Nilpotence and stable homotopy theory. I", *Ann. of Math.* **128** (1988), p. 207–241.

[19] A. D. Elmendorf, I. Kriz, M. A. Mandell & J. P. May – *Rings, modules, and algebras in stable homotopy theory*, Mathematical Surveys and Monographs, vol. 47, Amer. Math. Soc., 1997.

[20] J. Franke – "On the construction of elliptic cohomology", *Math. Nachr.* **158** (1992), p. 43–65.

[21] P. G. Goerss, H. W. Henn, M. E. Mahowald & C. Rezk – "A resolution of the $K(2)$-local sphere at the prime 3", *Ann. of Math.* **162** (2005), p. 777–822.

[22] P. G. Goerss & M. J. Hopkins – "Moduli spaces of commutative ring spectra", in *Structured ring spectra*, London Math. Soc. Lecture Note Ser., vol. 315, Cambridge Univ. Press, 2004, p. 151–200.

[23] H. W. Henn, N. Karamanov & M. E. Mahowald – "The homotopy of the $K(2)$-local Moore spectrum at the prime 3 revisited", preprint arXiv:0811.0235.

[24] S. Hollander – "Geometric criteria for Landweber exactness", *Proc. London Math. Soc.* **99** (2009), p. 697–724.

[25] M. J. Hopkins – "Topological modular forms, the Witten genus, and the theorem of the cube", in *Proceedings of the International Congress of Mathematicians, Vol. 1, 2 (Zürich, 1994)*, Birkhäuser, 1995, p. 554–565.

[26] ———, "Algebraic topology and modular forms", in *Proceedings of the International Congress of Mathematicians, Vol. I (Beijing, 2002)*, Higher Ed. Press, 2002, p. 291–317.

[27] M. J. Hopkins & B. H. Gross – "The rigid analytic period mapping, Lubin-Tate space, and stable homotopy theory", *Bull. Amer. Math. Soc. (N.S.)* **30** (1994), p. 76–86.

[28] M. J. Hopkins & M. E. Mahowald – "From elliptic curves to homotopy theory", preprint `http://hopf.math.purdue.edu//Hopkins-Mahowald/eo2homotopy.pdf`, 2005.

[29] M. J. Hopkins & J. H. Smith – "Nilpotence and stable homotopy theory. II", *Ann. of Math.* **148** (1998), p. 1–49.

[30] M. Hovey – "Bousfield localization functors and Hopkins' chromatic splitting conjecture", in *The Čech centennial (Boston, MA, 1993)*, Contemp. Math., vol. 181, Amer. Math. Soc., 1995, p. 225–250.

[31] M. Hovey, B. Shipley & J. H. Smith – "Symmetric spectra", *J. Amer. Math. Soc.* **13** (2000), p. 149–208.

[32] M. Hovey & N. P. Strickland – "Morava K-theories and localisation", *Mem. Amer. Math. Soc.* **139** (1999).

[33] L. Illusie – *Complexe cotangent et déformations. I*, Lecture Notes in Math., vol. 239, Springer, 1971.

[34] J. F. Jardine – "Stable homotopy theory of simplicial presheaves", *Canad. J. Math.* **39** (1987), p. 733–747.

[35] ———, "Presheaves of symmetric spectra", *J. Pure Appl. Algebra* **150** (2000), p. 137–154.

[36] N. M. Katz – "Higher congruences between modular forms", *Ann. of Math.* **101** (1975), p. 332–367.

[37] N. M. Katz & B. Mazur – *Arithmetic moduli of elliptic curves*, Annals of Math. Studies, vol. 108, Princeton Univ. Press, 1985.

[38] P. S. Landweber – "Homological properties of comodules over $MU_*(MU)$ and $BP_*(BP)$", *Amer. J. Math.* **98** (1976), p. 591–610.

[39] P. S. Landweber, D. C. Ravenel & R. E. Stong – "Periodic cohomology theories defined by elliptic curves", in *The Čech centennial (Boston, MA, 1993)*, Contemp. Math., vol. 181, Amer. Math. Soc., 1995, p. 317–337.

[40] G. Laumon & L. Moret-Bailly – *Champs algébriques*, Ergebnisse Math. Grenzg., vol. 39, Springer, 2000.

[41] G. Laures – "$K(1)$-local topological modular forms", *Invent. Math.* **157** (2004), p. 371–403.

[42] M. Lazard – "Sur les groupes de Lie formels à un paramètre", *Bull. Soc. Math. France* **83** (1955), p. 251–274.

[43] L. G. J. Lewis, J. P. May, M. Steinberger & J. E. McClure – *Equivariant stable homotopy theory*, Lecture Notes in Math., vol. 1213, Springer, 1986.

[44] J. Lubin & J. Tate – "Formal moduli for one-parameter formal Lie groups", *Bull. Soc. Math. France* **94** (1966), p. 49–59.

[45] J. Lurie – "Derived algebraic geometry", Ph.D. Thesis, MIT, 2004.

[46] ———, "Survey article on elliptic cohomology", preprint `http://www.math.harvard.edu/~lurie/papers/survey.pdf`, 2007.

[47] ———, *Higher topos theory*, Annals of Math. Studies, vol. 170, Princeton Univ. Press, 2009.

[48] M. E. Mahowald – "On the order of the image of J", *Topology* **6** (1967), p. 371–378.

[49] M. E. Mahowald & C. Rezk – "Brown-Comenetz duality and the Adams spectral sequence", *Amer. J. Math.* **121** (1999), p. 1153–1177.

[50] ———, "Topological modular forms of level 3", *Pure Appl. Math. Q.* **5** (2009), p. 853–872.

[51] M. A. Mandell, J. P. May, S. Schwede & B. Shipley – "Model categories of diagram spectra", *Proc. London Math. Soc.* **82** (2001), p. 441–512.

[52] W. Messing – *The crystals associated to Barsotti-Tate groups: with applications to abelian schemes*, Lecture Notes in Math., vol. 264, Springer, 1972.

[53] H. R. Miller, D. C. Ravenel & W. S. Wilson – "Periodic phenomena in the Adams-Novikov spectral sequence", *Ann. Math.* **106** (1977), p. 469–516.

[54] N. Naumann – "The stack of formal groups in stable homotopy theory", *Adv. in Math.* **215** (2007), p. 569–600.

[55] D. Quillen – "Elementary proofs of some results of cobordism theory using Steenrod operations", *Adv. in Math.* **7** (1971), p. 29–56.

[56] D. C. Ravenel – *Complex cobordism and stable homotopy groups of spheres*, Pure and Applied Mathematics, vol. 121, Academic Press Inc., 1986.

[57] ———, *Nilpotence and periodicity in stable homotopy theory*, Annals of Math. Studies, vol. 128, Princeton Univ. Press, 1992.

[58] C. Rezk – "Lectures on power operations", notes of a course given at MIT in 2006, `http://www.math.uiuc.edu/~rezk/power-operation-lectures.dvi`.

[59] ———, "Supplementary notes for Math 512", notes of a course on topological modular forms given at Northwestern University in 2001, `http://www.math.uiuc.edu/~rezk/512-spr2001-notes.pdf`.

[60] G. Segal – "Elliptic cohomology (after Landweber-Stong, Ochanine, Witten, and others)", Séminaire Bourbaki, vol. 1987/88, exposé n° 695, *Astérisque* **161-162** (1988), p. 187–201.

[61] J-P. Serre – *Algèbre locale. Multiplicités*, Cours au Collège de France, 1957–1958, rédigé par Pierre Gabriel. Seconde édition, 1965. Lecture Notes in Math., vol. 11, Springer, 1965.

[62] K. Shimomura & X. Wang – "The homotopy groups $\pi_*(L_2S^0)$ at the prime 3", *Topology* **41** (2002), p. 1183–1198.

[63] K. Shimomura & A. Yabe – "The homotopy groups $\pi_*(L_2S^0)$", *Topology* **34** (1995), p. 261–289.

[64] J. H. Silverman – *The arithmetic of elliptic curves*, Graduate Texts in Math., vol. 106, Springer, 1986.

[65] B. Smithling – "On the moduli stack of commutative, 1-parameter formal Lie groups", Ph.D. Thesis, University of Chicago, 2007.

[66] B. Toën & G. Vezzosi – "Homotopical algebraic geometry. II. Geometric stacks and applications", *Mem. Amer. Math. Soc.* **193** (2008).

[67] E. Witten – "Elliptic genera and quantum field theory", *Comm. Math. Phys.* **109** (1987), p. 525–536.

Paul G. GOERSS

Northwestern University
Department of Mathematics
2033 Sheridan Road
EVANSTON, Il 60208-2730 – USA
E-mail : pgoerss@math.northwestern.edu

332

ASTÉRISQUE

2010

SÉMINAIRE BOURBAKI
VOLUME 2008/2009
EXPOSÉS 997-1011

(1006) *Corps de classes des schémas arithmétiques*

Tamás SZAMUELY

SOCIÉTÉ MATHÉMATIQUE DE FRANCE
Publié avec le concours du CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

Séminaire BOURBAKI Mars 2009
61e année, 2008-2009, no 1006, p. 257 à 286

CORPS DE CLASSES DES SCHÉMAS ARITHMÉTIQUES

par **Tamás SZAMUELY**

INTRODUCTION

La théorie des corps de classes représente l'acmé de l'étude classique des corps de nombres. Son objet est la description des extensions abéliennes d'un corps global K au moyen d'invariants arithmétiques attachés à K. C'est le point culminant d'une longue série de travaux partant des recherches de Gauss sur les lois de réciprocité, et le fruit des efforts de mathématiciens aussi éminents que Weber, Hilbert, Takagi, Artin, Hasse ou Chevalley, pour ne nommer que quelques-uns. C'est aussi le point de départ de nombre de sujets centraux en arithmétique moderne, tels que le programme de Langlands, la théorie d'Iwasawa, ou les principes locaux-globaux sur les points rationnels. Et c'est également un outil fondamental dans l'étude de cet objet plein de mystère qui est le groupe de Galois absolu de $\mathbf{Q}$.

Les premiers pas vers une généralisation en dimension supérieure ont été faits par Lang [**28**], [**27**], pour les variétés sur les corps finis. C'est sauf erreur la dernière fois où le sujet a été évoqué dans ce séminaire [**45**]. Pourtant, des développements très significatifs ont émergé vers la fin des années 1970, grâce aux travaux de Parshin [**33**] dans le cas local et Bloch [**3**] dans le cas global. Leur idée était d'utiliser des K-groupes de Milnor dans la construction des invariants arithmétiques attachés à des schémas de dimension supérieure, le cas classique étant celui du groupe $K_1 = \mathbf{G}_m$. Ce programme a été mené à bien dans une série de travaux ([**17**], [**21**], [**20**], [**22**], [**36**]) de K. Kato et S. Saito, qui donnent une description des revêtements étales abéliens d'un schéma régulier connexe de type fini sur $\mathbf{Z}$. Un point faible de cette théorie est que les invariants construits via la K-théorie sont très compliqués (sauf dans le cas propre), ce qui les rend peu propices aux applications. Mais l'utilisation de la K-théorie a également un avantage : elle fournit un lien avec la théorie des régulateurs sur les

K-groupes et les groupes de cycles des schémas arithmétiques, objets d'un faisceau de conjectures « motiviques » justement célèbres.

Récemment, une nouvelle approche plus élémentaire a été trouvée par le regretté G. Wiesend. Elle permet de retrouver la plupart des résultats principaux de Kato et Saito sous une forme simplifiée, notamment le cas de caractéristique 0, ainsi que le cas où X est propre sur $\mathbf{Z}$. En particulier, elle donne une nouvelle démonstration de la finitude du groupe de Chow $CH_0(X)$ des zéro-cycles d'un schéma régulier propre et plat sur $\mathbf{Z}$, et de la finitude de la partie de degré zéro $CH_0(X)^0 \subset CH_0(X)$ pour une variété projective et lisse sur un corps fini ; dans ce dernier cas le résultat de Wiesend est même un peu plus général, et ne requiert que l'hypothèse de propreté au lieu de la projectivité. Sa méthode n'utilise pas la K-théorie et est purement globale, évitant les localisations élaborées utilisées précédemment. Par contre, le lien avec les conjectures motiviques n'apparaît pas explicitement. Nous avons donc décidé de discuter en annexe ce qui est peut-être le plus joli exemple de l'approche cohomologique : la théorie non ramifiée pour les variétés sur les corps finis.

Les arguments de Wiesend, publiés dans les notes [**51**], [**52**] et [**53**], contenaient des lacunes et des imprécisions. Ils ont été mis au propre dans les publications [**25**] et [**26**] de Kerz et Schmidt, qui ont également trouvé des améliorations. L'article de Kerz [**24**] y apporte des compléments et des arguments alternatifs. Nous avons suivi ces sources lors de la rédaction de la majeure partie de cet exposé.

Bien entendu, beaucoup reste encore à faire dans le domaine. Un des problèmes ouverts majeurs est la description fine des revêtements abéliens ramifiés via une théorie généralisée de conducteurs ; des premiers pas sont faits dans [**19**]. Une autre tâche, non sans lien avec la précédente, est le développement d'une théorie analytique satisfaisante qui jouerait le rôle de la thèse de Tate en dimension supérieure.

Mais il existe également des aspects déjà bien étudiés dont nous ne pouvons rendre compte dans le cadre du présent exposé. Mentionnons-en deux. Le premier est la théorie des corps de classes pour les corps locaux supérieurs. Ce sont les corps qui apparaissent quand on complète successivement les anneaux locaux de schémas arithmétiques réguliers le long d'une chaîne maximale de sous-schémas fermés. Cette théorie, qui a joué un rôle clef dans l'approche de Kato et Saito à la théorie globale, a été développée par Kato dans la série d'articles [**17**], et aussi d'une autre façon par Fesenko et ses collaborateurs. Nous renvoyons le lecteur intéressé au livre [**6**] et aux références citées dedans. Un deuxième sujet fascinant est l'ensemble de conjectures proposées par Kato dans son article fondamental [**18**]. Elles donnent une généralisation en dimension supérieure de la suite exacte d'Albert–Brauer–Hasse–Noether décrivant le groupe de Brauer d'un corps global, elle-même l'un des résultats centraux de la théorie des corps de classes classique. Une grande partie de ces conjectures a été démontrée au cours de la dernière quinzaine d'années par Jannsen, Saito et leurs

collaborateurs. Ces résultats très importants n'ont été rédigés que partiellement à ce jour (voir [**14**], [**15**]), et mériteront certainement un exposé à part entière.

Le rédacteur tient à exprimer sa gratitude à Jean-Louis Colliot-Thélène et à Alexander Schmidt pour des corrections apportées *in extremis.*

1. RAPPELS SUR LA THÉORIE CLASSIQUE

Soit K un corps global, i.e. une extension finie de $\mathbf{Q}$ ou de $\mathbf{F}(t)$ pour un corps fini $\mathbf{F}$. Pour une place v de K notons K_v le complété de K par rapport à v et, pour v fini, notons $\mathcal{O}_v \subset K_v$ son anneau des entiers. Suivant Chevalley, l'anneau $\mathbf{A}_K$ des adèles de K est défini comme le sous-anneau du produit direct de tous les K_v constitué des suites (a_v) avec $a_v \in \mathcal{O}_v$ pour tout v sauf pour un nombre fini de places finies. Le groupe $\mathbf{I}_K \subset \mathbf{A}_K$ des unités de $\mathbf{A}_K$ est appelé le groupe des idèles de K. Il est muni de la topologie dans laquelle une base de sous-groupes ouverts de 1 est donné par les sous-groupes

$$\prod_{v\in S} W_v \times \prod_{v\notin S} \mathcal{O}_v^\times,$$

où S est un ensemble fini de places contenant les places infinies, et W_v est un sous-groupe ouvert de $K_v^\times$. Pour toute place v le morphisme

$$i_v : K_v^\times \to \mathbf{I}_K, \qquad a_v \mapsto (1, \dots, 1, a_v, 1, \dots, 1)$$

est un plongement topologique. On considère également l'application diagonale

$$i : K^\times \to \mathbf{I}_K, \qquad a \mapsto (a, a, \dots, a),$$

dont on note C_K le conoyau, muni de la topologie quotient. C'est *le groupe des classes d'idèles* de K.

Pour v fini on dispose de *l'application de réciprocité locale*

$$\rho_v : K_v^\times \to \mathrm{Gal}\,(\overline{K}_v|K_v)^{\mathrm{ab}},$$

où $\overline{K}_v$ est une clôture séparable de K_v et G^{ab} désigne l'abélianisé du groupe G. Elle est définie par exemple dans [**44**], chap. XIII, §4. On définit ρ_v aux places infinies comme suit : si $K_v = \mathbf{C}$, on pose $\rho_v = 0$; si $K_v = \mathbf{R}$, on envoie $\mathbf{R}_+^\times$ sur 0, et -1 sur la conjugaison complexe engendrant $\mathrm{Gal}\,(\mathbf{C}|\mathbf{R})$.

Avec ces notations on peut résumer les énoncés principaux de la théorie globale dans la formulation de Chevalley comme suit.

THÉORÈME 1.1. — *Soit K un corps global, et soit $\overline{K}$ une clôture séparable de K.*

(1) *Il existe un unique homomorphisme* $\widetilde{\rho} : \mathbf{I}_K \to \mathrm{Gal}\,(\overline{K}|K)^{\mathrm{ab}}$ *faisant commuter les diagrammes*

$$\begin{array}{ccc} K_v^\times & \xrightarrow{\rho_v} & \mathrm{Gal}\,(\overline{K}_v|K_v)^{\mathrm{ab}} \\ {\scriptstyle i_v}\downarrow & & \downarrow \\ \mathbf{I}_K & \xrightarrow{\widetilde{\rho}} & \mathrm{Gal}\,(\overline{K}|K)^{\mathrm{ab}} \end{array}$$

pour toute place v. *Ici le morphisme vertical de droite est induit par l'envoi de* $\mathrm{Gal}\,(\overline{K}_v|K_v)$ *sur un sous-groupe de décomposition de* v *dans* $\mathrm{Gal}\,(\overline{K}|K)$.

(2) (Loi de réciprocité globale) *On a*

$$(\widetilde{\rho} \circ i)(K^\times) = 0,$$

d'où un morphisme induit $\rho : C_K \to \mathrm{Gal}\,(\overline{K}|K)^{\mathrm{ab}}$, *appelé application de réciprocité globale.*

(3) *Si* $\mathrm{car}(K) = 0$, *le morphisme* ρ *est surjectif, et son noyau est la composante connexe de l'identité du groupe abélien topologique* C_K. *C'est aussi le sous-groupe divisible maximal de* C_K.

Si $\mathrm{car}(K) > 0$, *le groupe* $\mathrm{Gal}\,(\overline{K}|K)^{\mathrm{ab}}$ *admet comme quotient canonique le groupe* $\mathrm{Gal}\,(\overline{\mathbf{F}}|\mathbf{F}) \cong \widehat{\mathbf{Z}}$, *engendré topologiquement par l'automorphisme de Frobenius* F. *Le morphisme* ρ *est injectif, et son image dans* $\mathrm{Gal}\,(\overline{K}|K)^{\mathrm{ab}}$ *est le sous-groupe des éléments qui s'envoient sur une puissance de* F *dans* $\mathrm{Gal}\,(\overline{\mathbf{F}}|\mathbf{F})$.

(4) (Isomorphisme de réciprocité global) *Pour tout quotient fini de la forme* $\mathrm{Gal}\,(L|K)$ *de* $\mathrm{Gal}\,(\overline{K}|K)^{\mathrm{ab}}$ *l'application de réciprocité* ρ *induit un isomorphisme*

$$C_K/N_{L|K}C_L \xrightarrow{\sim} \mathrm{Gal}\,(L|K).$$

Ici l'application norme $N_{L|K} : C_L \to C_K$ *est induite par les applications norme usuelles* $N_{L_w|K_v} : L_w^\times \to K_v^\times$, *où* w *est une place de l'extension finie* L *de* K *au-dessus de* v.

(5) (Théorème d'existence) *L'application de réciprocité* ρ *induit une bijection entre les sous-groupes ouverts d'indice fini* $U \subset C_K$ *et les sous-groupes ouverts de* $\rho(C_K) \subset \mathrm{Gal}\,(\overline{K}|K)^{\mathrm{ab}}$. *En outre, tout* U *comme ci-dessus est de la forme* $N_{L|K}(C_L)$ *pour une extension finie abélienne* $L|K$ *convenable. Si* $\mathrm{car}(K) = 0$, *tout sous-groupe ouvert de* C_K *est d'indice fini.*

Pour les démonstrations, voir par exemple [**1**], chapitres 7 et 8.

Une théorie plus fine est obtenue en considérant des *modules*. Par définition, un module est une série d'entiers $\mathfrak{m} = (n_v)$, avec v parcourant les places de K. Les n_v doivent être nuls sauf pour un nombre fini de places finies ou réelles, et égaux à 0 ou à 1 pour une place réelle. L'ordre usuel de $\mathbf{Z}$ induit un ordre partiel naturel sur l'ensemble des modules.

Étant donné un module $\mathfrak{m} = (n_v)$, on lui associe le sous-groupe

$$\mathbf{I}_K^{\mathfrak{m}} := \prod_v U_v^{(n_v)} \subset \mathbf{I}_K,$$

où $U_v^{(n_v)}$ est le groupe des unités principales $\{x \in K_v : v(x-1) \geq n_v\}$ dans K_v pour v finie, et le groupe multiplicatif $K_v^\times$ pour v infinie, sauf pour $K_v = \mathbf{R}$ et $n_v = 1$, auquel cas c'est le groupe multiplicatif de $\mathbf{R}_+$.

L'image de $\mathbf{I}_K^{\mathfrak{m}}$ dans C_K est noté $C_K^{\mathfrak{m}}$; c'est un sous-groupe ouvert. Par le théorème d'existence ci-dessus, il lui correspond une extension finie abélienne $K^{\mathfrak{m}}|K$, appelée le *corps de classes de rayon (Strahlklassenkörper)* associé au module $\mathfrak{m}$. Également par le théorème d'existence (et la définition de la topologie de C_K) toute extension finie abélienne $L|K$ est contenue dans un $K^{\mathfrak{m}}$. La borne inférieure $\mathfrak{f}$ des $\mathfrak{m}$ tels que $K^{\mathfrak{m}} \supset L$ est le *conducteur (Führer)* de l'extension $L|K$. Il jouit de la propriété fondamentale suivante (démontrée dans [**1**], Chapter 8 et [**42**], chapitre VI, §6, par exemple) :

Complément 1.2. — *Une place finie v est ramifiée dans $L|K$ si et seulement si $f_v > 0$ dans $\mathfrak{f} = (f_v)$.*

Pour $\mathfrak{m} = (0, 0, \ldots, 0)$ on a $C_K/C_K^{\mathfrak{m}} = Cl_K$, le groupe de classes de K. Si $\mathrm{car}(K) = 0$, il est fini, et l'extension abélienne correspondante de K est *le corps de classes de Hilbert,* traditionnellement noté H. Si $\mathrm{car}(K) > 0$, le groupe Cl_K n'est plus fini, mais il existe un morphisme degré $Cl_K \to \mathbf{Z}$ dont le noyau Cl_K^0 est fini.

Le théorème et son complément impliquent alors :

Corollaire 1.3 (Corps de classes non ramifié)

- *Si $\mathrm{car}(K) = 0$, le corps de classes de Hilbert H est l'extension abélienne non ramifiée maximale de K dans laquelle les places réelles de K sont totalement décomposées. L'application de réciprocité induit un isomorphisme de groupes finis*
$$Cl_K \xrightarrow{\sim} \mathrm{Gal}\,(H|K).$$
- *Si $\mathrm{car}(K) > 0$, l'application de réciprocité induit un isomorphisme de groupes finis*
$$Cl_K^0 \xrightarrow{\sim} \ker(\mathrm{Gal}\,(H|K) \to \mathrm{Gal}\,(\overline{\mathbf{F}}|\mathbf{F})),$$
où H est l'extension abélienne non ramifiée maximale de K.

Soit maintenant S un ensemble fini de places de K contenant les places infinies, et soit M_S l'ensemble des modules $\mathfrak{m} = (n_v)$ avec $n_v = 0$ si $v \notin S$. Les groupes $\mathrm{Gal}\,(K^{\mathfrak{m}}|K)$ pour $\mathfrak{m} \in M_S$ forment un système projectif dont la limite est le groupe $\mathrm{Gal}\,(K_S^{\mathrm{ab}}|K)$, où K_S^{ab} est l'extension abélienne maximale de K non ramifiée en dehors de S. En prenant la limite projective des quotients correspondants de C_K, on obtient :

COROLLAIRE 1.4. — *L'application de réciprocité induit un morphisme de groupes topologiques*

$$\rho_S : \operatorname{coker}(K^\times \to \bigoplus_{v \notin S} \mathbf{Z} \oplus \bigoplus_{v \in S} K_v^\times) \to \operatorname{Gal}(K_S^{\mathrm{ab}}|K).$$

Il jouit de propriétés analogues à celles décrites dans les énoncés (3)–(5) du théorème 1.1.

En langage géométrique, le corollaire décrit les revêtements finis étales abéliens des ouverts de Spec $\mathcal{O}_K$, où $\mathcal{O}_K$ est l'anneau des entiers de K. Ce sont les deux corollaires ci-dessus que l'on se propose de généraliser dans les sections suivantes.

Enfin, rappelons que la théorie des corps de classes fournit des renseignements non seulement sur la ramification des idéaux premiers dans les extensions abéliennes, mais aussi sur la décomposition des idéaux. Citons ici le résultat qui est peut-être le plus célèbre, et qui sera également généralisé dans la suite :

THÉORÈME 1.5 (Hauptidealsatz). — *Soit K un corps de nombres, et soit $\mathcal{O}_K$ son anneau d'entiers. Tout idéal premier de $\mathcal{O}_K$ engendre un idéal principal dans l'anneau des entiers $\mathcal{O}_H$ du corps de classes de Hilbert de K.*

Pour la démonstration, voir par exemple **[31]**, chapitre VI, théorème 7.5 ou **[44]**, chapitre VII, §8.

2. L'APPLICATION DE RÉCIPROCITÉ DE WIESEND

Soit désormais X un schéma régulier connexe, séparé et de type fini sur $\mathbf{Z}$. Par *courbe sur X* on entend un sous-schéma fermé intègre $C \subset X$ de dimension de Krull un. Notons $\widetilde{C}$ le normalisé de C et $k(C)$ son corps de fonctions ; c'est un corps global classique. Soient Ω_C l'ensemble des places de $k(C)$, et $\Omega_C^\infty \subset \Omega_C$ le sous-ensemble des places ne provenant pas d'un point fermé de C ; elle contient les places infinies si $k(C)$ est de caractéristique 0, mais aussi des places finies quand C n'est pas propre. Pour $v \in \Omega_C$ on note $k(C)_v$ le complété de $k(C)$ en v.

DÉFINITION 2.1. — Le groupe d'idèles *(selon Wiesend) de X est la somme directe*

$$I_X := Z_0(X) \oplus \bigoplus_{C \subset X} \bigoplus_{v \in \Omega_C^\infty} k(C)_v^\times.$$

Ici $Z_0(X)$ est le groupe des zéro-cycles de X (i.e. le groupe abélien libre engendré par les points fermés de X), et la deuxième somme est indexée par les courbes sur X au sens précédent.

Pour toute courbe C sur X il existe un morphisme naturel

$$i_C : k(C)^\times \to I_X$$

défini comme suit : sur la composante $Z_0(X)$ c'est le composé $k(C)^\times \overset{\mathrm{div}}{\to} Z_0(\widetilde{C}) \to Z_0(X)$, où la première flèche est la flèche diviseur et la deuxième la poussette des zéro-cycles ; sur une composante indexée par $v \in \Omega_C^\infty$, c'est l'inclusion $k(C)^\times \to k(C)_v^\times$; sur les autres composantes, c'est 0.

Définition 2.2. — *Le quotient*

$$C_X := \mathrm{coker}\left(\bigoplus_{C\subset X} k(C)^\times \overset{\sum i_C}{\longrightarrow} I_X\right)$$

est le groupe des classes d'idèles de X.

Notons que le groupe de Chow $CH_0(X)$ des zéro-cycles apparaît naturellement comme quotient de C_X : il suffit de projeter I_X sur sa composante $Z_0(X)$, et prendre l'image dans C_X. En dimension 1, on retrouve le groupe du corollaire 1.4.

Munissons $Z_0(X)$ de la topologie discrète, et les groupes $k(C)_v$ de la topologie v-adique. On peut alors équiper I_X de la topologie somme directe, et C_X de la topologie quotient. Cette topologie n'est ni séparée, ni localement compacte en général (*cf.* **[25]**, Example 7.1).

Lemme 2.3. — *L'image de $Z_0(X)$ dans C_X est un sous-groupe dense pour la topologie de C_X.*

Démonstration. — Pour X de dimension 1, ceci résulte du lemme d'approximation classique (**[44]**, chap. I, §3). Le cas général s'y ramène aussitôt. □

Nous allons maintenant définir un morphisme $\rho_X : C_X \to \pi_1^{\mathrm{ab}}(X)$, où $\pi_1^{\mathrm{ab}}(X)$ est l'abélianisé du groupe fondamental du schéma régulier X. Ce groupe classifie les revêtements finis étales galoisiens de X de groupe de Galois abélien. Pour $X = \mathrm{Spec}\,(F)$ c'est l'abélianisé du groupe de Galois absolu du corps F. Si l'on prend pour X le spectre de l'anneau des S-entiers de K, où S est un ensemble fini de places contenant les places infinies, c'est le groupe noté $\mathrm{Gal}\,(K_S^{\mathrm{ab}}|K)$ dans le corollaire 1.4.

L'association $X \to \pi_1^{\mathrm{ab}}(X)$ étant un foncteur covariant en X, il suffit de définir des morphismes $i_x : \mathbf{Z} \to \mathrm{Gal}\,(\overline{\kappa(x)}|\kappa(x))^{\mathrm{ab}}$ pour chaque point fermé $x \in X_0$, et des morphismes $i_C^v : k(C)_v^\times \to \mathrm{Gal}\,(\overline{k(C)_v}|k(C)_v)^{\mathrm{ab}}$ pour toute courbe C sur X. Le premier envoie $1 \in \mathbf{Z}$ sur le générateur canonique du groupe $\mathrm{Gal}\,(\overline{\kappa(x)}|\kappa(x))^{\mathrm{ab}} \cong \widehat{\mathbf{Z}}$, à savoir le morphisme de Frobenius ; le second n'est autre que l'application de réciprocité locale classique ρ_v mentionnée dans la section précédente. En composant la somme

directe de ces morphismes avec la somme des morphismes $\pi_1^{\mathrm{ab}}(\operatorname{Spec}\kappa(x)) \to \pi_1^{\mathrm{ab}}(X)$ et $\pi_1^{\mathrm{ab}}(\operatorname{Spec}k(C)_v) \to \pi_1^{\mathrm{ab}}(X)$ donnés par fonctorialité, on obtient un morphisme

$$\widetilde{\rho}_X : I_X \to \pi_1^{\mathrm{ab}}(X).$$

LEMME 2.4. — *Pour toute courbe C sur X le morphisme composé*

$$k(C)^\times \xrightarrow{i_C} I_X \xrightarrow{\widetilde{\rho}_X} \pi_1^{\mathrm{ab}}(X)$$

est trivial.

Démonstration. — Ceci n'est autre que la loi de réciprocité globale classique (théorème 1.1 (2)). □

Vu le lemme, le morphisme $\widetilde{\rho}_X$ induit un morphisme

$$\rho_X : C_X \to \pi_1^{\mathrm{ab}}(X),$$

appelé *application de réciprocité*.

Remarque 2.5. — On voit donc que dans l'approche de Wiesend il n'y a pas de nouvelle loi de réciprocité qui apparaît. Ce n'était pas ainsi dans la construction plus ancienne de Kato et Saito **[22]** : pour définir leur application de réciprocité, ils avaient besoin à la fois de la loi de réciprocité classique et d'une loi de réciprocité « supérieure », pour les anneaux locaux de dimension 2 de X.

L'application de réciprocité admet une fonctorialité covariante pour des morphismes *quelconques* qui sera très utile dans la suite. Pour l'expliquer, considérons un morphisme $\phi : Y \to X$ de schémas réguliers connexes de type fini sur $\mathbf{Z}$, et définissons une poussette $\widetilde{\phi}_* : I_Y \to I_X$ comme suit. Sur la composante $Z_0(Y)$ c'est la poussette $Z_0(Y) \to Z_0(X)$ des zéro-cycles. Définissons maintenant ϕ_* sur la composante $k(C)_v$, avec C une courbe sur Y et $v \in \Omega_C^\infty$. Si $D := \overline{\phi(C)} \subset X$ est une courbe sur X, alors ou bien $v|_{k(D)} \in \Omega_D^\infty$, auquel cas on prend la norme $k(C)_v^\times \to k(D)_v^\times$, ou bien $v|_{k(D)}$ provient d'un point fermé $x \in D$, et on prend le morphisme $k(C)_v^\times \to \mathbf{Z}.x \subset Z_0(X)$ induit par v. Enfin si C est contractée par ϕ en un point fermé $x \in X$ (ceci n'est possible que si Y et X sont de caractéristique $p > 0$), on envoie $k(C)_v^\times$ dans la composante $\mathbf{Z}.x \subset Z_0(X)$ par la valuation v (qui est forcément discrète).

LEMME 2.6. — *La poussette $\widetilde{\phi}_*$ induit un morphisme $\phi_* : C_Y \to C_X$ faisant commuter le diagramme*

$$\begin{array}{ccc} C_Y & \xrightarrow{\rho_Y} & \pi_1^{\mathrm{ab}}(Y) \\ {\scriptstyle \phi_*}\downarrow & & \downarrow \\ C_X & \xrightarrow{\rho_X} & \pi_1^{\mathrm{ab}}(X), \end{array}$$

où le morphisme vertical de droite est donné par la fonctorialité de π_1^{ab}.

Démonstration. — Que $\widetilde{\phi}_*$ induise un morphisme $C_Y \to C_X$ résulte du cas classique de dimension 1. Que ce morphisme soit compatible avec la poussette $\pi_1^{\mathrm{ab}}(Y) \to \pi_1^{\mathrm{ab}}(X)$ est immédiat sur la composante $CH_0(X)$, et résulte de la fonctorialité covariante de l'application de réciprocité locale ([**44**], chapitre XIII, proposition 10 (a)) sur les composantes locales. □

3. LES THÉORÈMES PRINCIPAUX

Énonçons maintenant les résultats principaux concernant l'application de réciprocité.

THÉORÈME 3.1

(a) *Supposons X plat sur $\mathbf{Z}$. Alors le morphisme ρ_X est surjectif, et son noyau est la composante connexe de l'identité dans C_X.*
(b) *Supposons que X est lisse et géométriquement connexe sur un corps fini $\mathbf{F}$, et qu'il admet un morphisme propre surjectif vers une courbe lisse Z, à fibre générique lisse et géométriquement intègre. Alors l'image de ρ_X est le sous-groupe de $\pi_1^{\mathrm{ab}}(X)$ formé des éléments qui s'envoient sur une puissance de Frobenius dans le quotient* $\mathrm{Gal}\,(\overline{\mathbf{F}}|\mathbf{F})$*, et son noyau est la composante connexe de l'identité dans C_X.*

Remarques 3.2

1. Dans le cas (b) la fibration ne sert que dans la démonstration du théorème pour la p-partie de $\mathrm{Im}\,(\rho_X)$ pour $p = \mathrm{car}(\mathbf{F})$; on espère pouvoir s'en débarrasser un jour. Mais si l'on ne s'intéresse qu'à la description des revêtements abéliens de X dont le degré est premier à p ou, plus généralement, aux revêtements modérément ramifiés, on n'a pas besoin de la condition de fibration. Voir ([**25**], Theorem 8.3) pour cette variante.
2. Kerz a démontré dans ([**24**], §5) que la composante connexe de l'identité dans C_X est égale au sous-groupe divisible maximal dans le cas où X est propre et plat sur un ouvert du spectre de l'anneau des entiers d'un corps de nombres.

Les démonstrations seront expliquées dans les paragraphes suivants. Déduisons maintenant des conséquences plus ou moins immédiates. Commençons par le cas propre, démontré avant par d'autres méthodes par Bloch, Kato et Saito.

Corollaire 3.3 (Corps de classes non ramifié)

(1) *Supposons X propre et plat sur $\mathbf{Z}$, et notons $\widetilde{\pi}_1^{\mathrm{ab}}(X)$ le quotient de $\pi_1^{\mathrm{ab}}(X)$ classifiant les revêtements finis étales abéliens de X pour lesquels tout point réel de X est complètement décomposé. Alors le morphisme composé $C_X \overset{\rho_X}{\to} \pi_1^{\mathrm{ab}}(X) \to \widetilde{\pi}_1^{\mathrm{ab}}(X)$ se factorise à travers un isomorphisme de groupes finis*

$$CH_0(X) \overset{\sim}{\to} \widetilde{\pi}_1^{\mathrm{ab}}(X).$$

(2) *Supposons X projectif et lisse sur un corps fini $\mathbf{F}$. Notons $\pi_1^{\mathrm{ab}}(X)^0$ le noyau du morphisme naturel $\pi_1^{\mathrm{ab}}(X) \to \mathrm{Gal}\,(\overline{\mathbf{F}}|\mathbf{F})$, et $CH^0(X)^0 \subset CH^0(X)$ le noyau du morphisme degré. L'application de réciprocité ρ_X se factorise à travers le quotient $CH_0(X)$ de C_X et induit un isomorphisme de groupes finis*

$$CH_0(X)^0 \overset{\sim}{\to} \pi_1^{\mathrm{ab}}(X)^0.$$

Notons qu'à ce jour le seul argument connu pour démontrer la finitude de $CH_0(X)$ (resp. de $CH_0(X)^0$) sous les hypothèses ci-dessus est en utilisant les isomorphismes du corollaire et la finitude de $\pi_1^{\mathrm{ab}}(X)$ (resp. $\pi_1^{\mathrm{ab}}(X)^0$), ce dernier fait résultant d'un théorème célèbre de Katz et Lang (théorème 4.3 ci-dessous).

Démonstration. — Comme X est supposé propre, dans le cas (a) le groupe C_X est extension de $CH_0(X)$ par un sous-groupe topologique A_X provenant de places archimédiennes. Le sous-groupe A_X contient la composante connexe de l'identité (puisque $CH_0(X)$ est discret) et s'envoie sur 0 dans $\widetilde{\pi}_1^{\mathrm{ab}}(X)$. Ceci étant remarqué, il suffit d'appliquer le théorème.

Dans le cas (b) on a $C_X = CH_0(X)$ sous l'hypothèse de projectivité. En faisant des éclatements convenables, on obtient un morphisme propre birationnel $Y \to X$, avec Y projective et lisse satisfaisant à l'hypothèse (b) du théorème 3.1 (utiliser des projections, par exemple). Puisque les groupes $CH_0(X)^0$ et $\pi_1^{\mathrm{ab}}(X)^0$ sont des invariants birationnels pour des variétés propres et lisses, on se ramène au cas $Y = X$. L'assertion résulte alors du cas (b) du théorème, ainsi que du diagramme commutatif

$$\begin{array}{ccc} CH_0(X) & \xrightarrow{\rho_X} & \pi_1^{\mathrm{ab}}(X) \\ {\scriptstyle \deg}\downarrow & & \downarrow \\ \mathbf{Z} & \longrightarrow & \mathrm{Gal}\,(\overline{\mathbf{F}}|\mathbf{F}) \end{array}$$

où le morphisme du bas est l'inclusion naturelle $\mathbf{Z} \subset \widehat{\mathbf{Z}}$. □

Remarques 3.4

(1) Ces énoncés ont été démontrés pour la première fois par Bloch [**3**] en dimension 2 pour le cas (a), et par Kato et Saito [**21**], [**36**] en général (avec l'aide de Colliot-Thélène dans le cas (b)). Voir aussi le chapitre 5 du rapport de Raskind [**35**] pour des démonstrations utilisant des techniques proches des preuves originales. Toutes ces preuves utilisaient la K-théorie algébrique. Cette approche, bien que plus chère, est toujours la plus éclairante dans le cas géométrique selon l'avis personnel du rédacteur, et sera discutée en annexe.

(2) Une variante « modérée » a été étudiée par Schmidt et Spiess [**41**] dans le cas géométrique et par Schmidt [**38**] dans le cas arithmétique. Ici X n'est plus supposé propre, et l'on remplace $\pi_1^{\mathrm{ab}}(X)$ par son quotient classifiant les revêtements abéliens modérément ramifiés à l'infini (voir [**26**], [**37**] pour cette notion). L'objet correspondant à $CH_0(X)$ est un groupe $H_0^{\mathrm{sing}}(X, \mathbf{Z})$ défini par Schmidt [**39**]. Sur un corps il coïncide avec le groupe $h_0(X)$ de Suslin (*cf.* [**46**]). Comme expliqué dans [**40**], ces résultats peuvent également être retrouvés par la méthode initiée par Wiesend.

(3) Comme noté dans ([**25**], Theorem 9.2), le cas (b) du corollaire vaut plus généralement pour X propre et lisse sur un corps fini. En effet, par un petit argument expliqué aux p. 2576–2577 de [**25**] on peut généraliser le lemme 4.4 ci-dessous, et ne demander l'existence de la fibration $X \to Z$ dans le cas (b) qu'après avoir fait une altération à la de Jong [**16**]. Le théorème de de Jong permet alors de supposer X projective et lisse admettant une fibration $X \to Z$ avec les propriétés requises, et c'est le seul endroit dans la démonstration du théorème 3.1 (b) où l'hypothèse de fibration sert.

La démonstration de ([**25**], Theorem 9.2) est un peu différente, et utilise la variante modérée de la théorie mentionnée ci-dessus.

(4) Il est amusant de noter que l'on a déduit la finitude de $CH_0(X)$ de la finitude de $\pi_1^{\mathrm{ab}}(X)$, tandis qu'en théorie des corps de classes classique on procède en sens inverse : on déduit la finitude du degré du corps de classes de Hilbert de la finitude du groupe des classes.

(5) Comme expliqué dans ([**22**], §6) ou ([**25**], §9), des arguments de dévissage faciles permettent de déduire des résultats ci-dessus qu'en toute généralité le groupe de Chow $CH_0(S)$ des zéro-cycles d'un schéma S connexe et séparé de type fini sur $\mathbf{Z}$ est fini, sauf si S_{red} est propre sur un corps fini, auquel cas $CH_0(S) \cong \mathbf{Z} \oplus F$, avec F fini.

Terminons cette section par une conséquence frappante de la théorie non ramifiée. Plaçons-nous dans la situation du cas (a) du corollaire 3.3, i.e. supposons X propre et plat sur $\mathbf{Z}$. Par la théorie du groupe fondamental il existe alors à isomorphisme près

un seul revêtement fini étale galoisien $X' \to X$ dont le groupe de Galois est le groupe fini abélien $\widetilde{\pi}_1^{\mathrm{ab}}(X)$. Il mérite le nom de *revêtement de Hilbert.*

COROLLAIRE 3.5 (Hauptidealsatz généralisé). — *Avec les notations ci-dessus, le morphisme $CH_0(X) \to CH_0(X')$ induit par la tirette des zéro-cycles est trivial.*

Démonstration. — L'argument est le même que celui d'Artin dans le cas classique. Soit $X'' \to X'$ le revêtement de Hilbert de X'. Son groupe de Galois $\pi_1^{\mathrm{ab}}(X')$ s'identifie naturellement à un sous-groupe du groupe de Galois G du revêtement composé $X'' \to X$. De plus, le quotient $G/\pi_1^{\mathrm{ab}}(X') \cong \widetilde{\pi}_1^{\mathrm{ab}}(X)$ est le quotient abélien maximal de G, donc le groupe abélien $\pi_1^{\mathrm{ab}}(X')$ est en fait le sous-groupe de commutateurs G' de G. Le transfert $\mathrm{Ver} : G/G' \to G'/G''$ s'identifie donc à un morphisme $\widetilde{\pi}_1^{\mathrm{ab}}(X) \to \widetilde{\pi}_1^{\mathrm{ab}}(X')$. On vérifie comme dans **[44]**, chapitre VI, §8, que le diagramme

$$\begin{array}{ccc} CH_0(X) & \xrightarrow{\rho_X} & \widetilde{\pi}_1^{\mathrm{ab}}(X) \\ \downarrow & & \downarrow \mathrm{Ver} \\ CH_0(X') & \xrightarrow{\rho_{X'}} & \widetilde{\pi}_1^{\mathrm{ab}}(X') \end{array}$$

commute. Mais d'après Furtwängler le transfert $\mathrm{Ver} : G/G' \to G'/G''$ est trivial pour tout groupe fini G (voir par exemple **[31]**, chapitre VI, théorème 7.6). Puisque ρ_X et $\rho_{X'}$ sont des isomorphismes, le corollaire est démontré. □

4. L'IMAGE DE L'APPLICATION DE RÉCIPROCITÉ

Le premier résultat concernant l'image de ρ_X remonte à Lang **[27]**.

PROPOSITION 4.1. — *L'image de ρ_X est dense pour la topologie profinie de $\pi_1^{\mathrm{ab}}(X)$.*

Suivant Lang, la proposition se démontre en utilisant la fonction zêta de X. Rappelons que c'est la fonction complexe définie par le produit eulérien

$$\zeta_X(s) = \prod_{P \in X_0} \frac{1}{(1 - \mathbf{N}P)^{-s}},$$

où X_0 est l'ensemble des points fermés de X, et $\mathbf{N}P$ le cardinal du corps résiduel de P. On sait (*cf.* **[43]**) que le produit converge pour $\mathrm{Re}\,(s) > \dim(X)$, et s'étend méromorphiquement au demi-plan $\mathrm{Re}\,(s) > \dim(X) - 1/2$, avec un pôle simple en $s = \dim(X)$.

Démonstration. — Nous allons montrer que déjà le sous-groupe $H \subset \pi_1^{\mathrm{ab}}(X)$ engendré par les éléments de Frobenius associés aux points fermés de X est dense dans $\pi_1^{\mathrm{ab}}(X)$. Supposons le contraire, et notons $\overline{H}$ l'adhérence de H. Par hypothèse on trouve un sous-groupe ouvert $U \subset \pi_1^{\mathrm{ab}}(X)$ contenant le sous-groupe fermé $\overline{H}$ qui n'est

pas $\pi_1^{\mathrm{ab}}(X)$ tout entier. Il correspond à un revêtement fini étale abélien $Y \to X$, de degré $d > 1$. Pour un point fermé x de X, la fibre Y_x en x doit être le spectre d'un produit fini de copies de $\kappa(x)$, car par construction de Y le Frobenius de x agit trivialement sur la fibre géométrique $Y_{\bar{x}}$. Par conséquent, il y a exactement d points fermés de Y au-dessus de x, et ceci pour tout x. Mais alors par définition $\zeta_Y = \zeta_X^d$, ce qui est impossible, car les deux fonctions zêta ont des pôles simples en $s = \dim(X) = \dim(Y)$. □

Nous traitons maintenant les énoncés du théorème 3.1 concernant l'image de ρ_X.

PROPOSITION 4.2. —

(a) *Supposons X plat sur $\mathbf{Z}$. Alors le morphisme ρ_X est surjectif.*
(b) *Supposons que X est lisse sur un corps fini $\mathbf{F}$, et qu'il admet un morphisme propre surjectif vers une courbe lisse, à fibre générique lisse et géométriquement intègre. Alors l'image de ρ_X est le sous-groupe de $\pi_1^{\mathrm{ab}}(X)$ formé des éléments qui s'envoient sur une puissance de Frobenius dans le quotient* $\mathrm{Gal}(\overline{\mathbf{F}}|\mathbf{F})$.

Nous avons besoin de deux résultats auxiliaires. Le premier est le théorème de finitude de Katz et Lang déjà mentionné, qui est un outil fondamental dans la théorie. Rappelons-le :

THÉORÈME 4.3 (Katz–Lang). — *Soit S un schéma normal intègre dont le corps de fonctions est de type fini sur le corps premier, de caractéristique $p \geq 0$. Soit $Y \to S$ un morphisme lisse surjectif, à fibre générique géométriquement intègre. Alors le noyau* $\ker(Y/S)$ *du morphisme induit $\pi_1^{\mathrm{ab}}(Y) \to \pi_1^{\mathrm{ab}}(S)$ est le produit d'un groupe fini et d'un pro-p-groupe. Si de plus $Y \to S$ est propre et toutes les fibres sont géométriquement intègres, alors* $\ker(Y/S)$ *est fini en toute caractéristique.*

La démonstration utilise la théorie des fibrations élémentaires (lemme 7.1 ci-dessous) pour se réduire au cas des courbes, ainsi que la théorie de la spécialisation du groupe fondamental et la finitude du groupe des points de la jacobienne d'une courbe sur un corps fini. Pour les détails, voir l'article original **[23]**, ainsi que **[35]**, §5 ou **[49]**, §5.8 (pour le cas propre, où l'on peut éviter l'argument de fibration en faisant appel à l'application d'Albanese).

Utilisant le théorème, on peut maintenant démontrer :

LEMME 4.4. — *Sous les hypothèses* (a) *ou* (b) *de la proposition il existe un morphisme $C \to X$ de schémas de type fini sur $\mathbf{Z}$ tel que C est régulier connexe de dimension 1 et l'image du morphisme induit $\pi_1^{\mathrm{ab}}(C) \to \pi_1^{\mathrm{ab}}(X)$ est ouverte. Dans le cas* (a) *on peut de plus prendre C plat sur $\mathbf{Z}$.*

Démonstration. — Pour démontrer le lemme on peut remplacer X par un ouvert étale $\phi : X' \to X$ convenable. En effet, dans le cas où ϕ est même fini étale, on peut le supposer galoisien de groupe G, et on conclut par la suite exacte de groupes profinis

$$\pi_1^{\mathrm{ab}}(X') \to \pi_1^{\mathrm{ab}}(X) \to G^{\mathrm{ab}} \to 0.$$

Dans le cas où ϕ est une immersion ouverte, le morphisme $\pi_1^{\mathrm{ab}}(X') \to \pi_1^{\mathrm{ab}}(X)$ est surjectif car la restriction de tout revêtement connexe de X reste connexe. Le cas général résulte de ces deux cas particuliers puisque ϕ est fini sur un ouvert dense $U \subset X$ (en effet, d'après le « main theorem » de Zariski on trouve une factorisation $\phi = \mu \circ \lambda$ avec $\lambda : X' \hookrightarrow \overline{X}$ une immersion ouverte et $\mu : \overline{X} \to X$ fini, et on peut prendre $U = X \setminus \mu(\overline{X} \setminus X')$).

Ceci étant dit, les hypothèses (a) ou (b) impliquent que l'on peut trouver un morphisme $X \to Z$ lisse et surjectif, à fibres géométriquement connexes, de schémas séparés de type fini sur $\mathbf{Z}$, avec Z régulier de dimension 1 ; dans le cas (a) on peut en outre supposer Z plat sur $\mathbf{Z}$, et dans le cas (b) X propre sur Z. D'après le théorème de Katz–Lang, le noyau $\pi_1^{\mathrm{ab}}(X/Z)$ du morphisme $\pi_1^{\mathrm{ab}}(X) \to \pi_1^{\mathrm{ab}}(Z)$ est fini. Mais le morphisme $X \to Z$ étant lisse, il admet une section s après passage à un ouvert étale $Z' \to Z$. En remplaçant Z par Z' et X par $X \times_Z Z'$, la courbe $s(Z) \subset X$ convient, puisque l'image de $\pi_1^{\mathrm{ab}}(s(Z)) \to \pi_1^{\mathrm{ab}}(X)$ est compacte, donc fermée, et d'indice fini par ce qui précède. □

Démonstration de la proposition 4.2. – Prenons C comme dans le lemme. Dans le cas (a) on sait par le théorème 1.1 (3) que ρ_C est surjectif. Le lemme 2.6 appliqué à $Y = C$ implique donc que $\mathrm{Im}\,(\rho_X)$ contient un sous-groupe ouvert de $\pi_1^{\mathrm{ab}}(X)$. Il est alors lui-même ouvert, donc fermé dans $\pi_1^{\mathrm{ab}}(X)$. Comme par ailleurs il est dense par la proposition 4.1, il est égal à $\pi_1^{\mathrm{ab}}(X)$. La démonstration du cas (b) est similaire, en considérant l'intersection de $\mathrm{Im}\,(\rho_X)$ avec $\pi_1^{\mathrm{ab}}(X)^0 = \ker(\pi_1^{\mathrm{ab}}(X) \to \mathrm{Gal}\,(\overline{\mathbf{F}}|\mathbf{F}))$.

5. L'ISOMORPHISME DE RÉCIPROCITÉ

Gardant les hypothèses précédentes sur X, nous allons maintenant démontrer l'analogue du théorème 1.1 (4) en dimension supérieure.

THÉORÈME 5.1. — *Soit $\phi : Y \to X$ un revêtement fini étale galoisien, de groupe G. L'application de réciprocité ρ_X induit un isomorphisme*

$$C_X/\phi_* C_Y \xrightarrow{\sim} G^{\mathrm{ab}}.$$

Comme on va le voir, ce résultat se ramène sans grande difficulté au cas de dimension 1. En caractéristique positive on va utiliser un argument de type Bertini. En caractéristique 0 le lemme clef est le suivant.

LEMME 5.2 (lemme d'approximation de Bloch–Raskind–Kerz)

Supposons de plus que X est quasi-projectif et lisse sur Spec $\mathcal{O}_k$, *avec $\mathcal{O}_k$ l'anneau des entiers d'un corps de nombres k. Étant donné un ensemble fini $x_1, \dots, x_n$ de points fermés de X, il existe une courbe C sur X contenant les x_i comme points réguliers.*

Cette formulation du lemme se trouve dans ([**35**], Lemma 6.21), sous l'hypothèse supplémentaire que les x_i ont des images distinctes dans Spec $\mathcal{O}_k$; elle repose sur un énoncé un peu plus faible démontré dans [**3**]. La démonstration exploite le théorème d'irréductibilité de Hilbert. Kerz ([**24**], Proposition 2.2) a donné une nouvelle preuve du lemme, sans hypothèse supplémentaire sur les x_i. Il utilise les théorèmes de Bertini sur les corps finis trouvés récemment par Gabber [**7**] et Poonen [**34**].

Comme observé par Wiesend dans [**51**], le lemme peut être renforcé comme suit.

COMPLÉMENT 5.3. — *Dans la situation du lemme, considérons de plus un revêtement étale galoisien connexe $\phi : Y \to X$. Alors on peut choisir C de sorte que le produit fibré $Y \times_X C$ soit connexe.*

La preuve ci-après est celle de [**25**].

Démonstration. — Soient $\Gamma_1, \dots, \Gamma_m$ les classes de conjugaison de G. D'après le théorème de Tchébotarev généralisé ([**43**], Theorem 7) on peut trouver pour chaque i un point fermé x_i' de X pour lequel Γ_i contient des générateurs des groupes de décomposition (cycliques) H_{ij} des points de la fibre $Y_{x_i'}$. Appliquant le lemme à l'ensemble fini $S = \{x_1, \dots, x_n, x_1', \dots x_m'\}$, on trouve une courbe C sur X dont le lieu régulier C^{reg} contient S. Si D est une composante de $Y \times_X C^{\mathrm{reg}}$, le groupe de Galois H du revêtement étale galoisien $D \to C^{\mathrm{reg}}$ est un sous-groupe de G ; il contient au moins un H_{ij} pour tout i, car D contient un point de Y au-dessus de chaque x_i'. Mais alors G est la réunion des conjugués de H, ce qui n'est possible que si $H = G$ par un lemme classique (*cf.* par exemple [**50**], §112, Hilfssatz). Ceci montre que $D = Y \times_X C^{\mathrm{reg}}$, et alors l'adhérence $Y \times_X C$ est également connexe. □

Démonstration du théorème 5.1. – Remarquons que pour $\dim X = 1$ l'énoncé n'est autre que le th. 1.1 (4) ; on va y ramener le cas général.

Le groupe G^{ab} est le quotient de $\pi_1^{\mathrm{ab}}(X)$ par l'image de $\pi_1^{\mathrm{ab}}(Y)$, donc la fonctorialité covariante de ρ_X induit bien un morphisme $\rho_{XY} : C_X/\phi_* C_Y \to G^{\mathrm{ab}}$. L'image de ρ_{XY} est dense dans le groupe fini discret G^{ab} par la proposition 4.1, donc ρ_{XY} est surjectif.

Remarquons maintenant que l'application composée $Z_0(X) \to C_X \to C_X/\phi_* C_Y$ est surjective. En effet, la norme est une application ouverte pour les extensions finies galoisiennes de corps locaux (ceci résulte des calculs du chap. V de [**44**]), donc $\phi_* : C_Y \to C_X$ est également une application ouverte, et l'assertion résulte du lemme 2.3. On peut donc relever tout élément $\alpha \in \ker(\rho_{XY})$ en un zéro-cycle $z_\alpha \in Z_0(X)$, et

choisir une courbe C passant par le support de z satisfaisant aux propriétés dans 5.2 et 5.3 (en caractéristique positive on remplace l'application de 5.2 par un argument de Bertini–Gabber–Poonen). Notant $\widetilde{C}$ le normalisé de C, le revêtement $\phi_{\widetilde{C}} : Y \times_X \widetilde{C} \to \widetilde{C}$ est galoisien de groupe G par construction. L'élément α provient donc d'un élément $\widetilde{\alpha}$ de $C_{\widetilde{C}}$ qui s'annule dans G^{ab} par fonctorialité de C_X pour $\widetilde{C} \to X$. Par le cas de dimension 1 on a $\widetilde{\alpha} = 0$ modulo $\phi_{\widetilde{C}*} C_{Y \times_X \widetilde{C}}$, et a fortiori $\alpha = 0$ dans $C_X/\phi_* C_Y$. □

6. LE THÉORÈME D'EXISTENCE

Du théorème 3.1 seuls les énoncés concernant le noyau de ρ_X restent à démontrer. Or des considérations générales sur les groupes topologiques détaillées dans la section 6 de [**25**] montrent que dans le groupe C_X la composante connexe de l'identité est l'intersection des sous-groupes ouverts. Ainsi les énoncés requis résultent de la première assertion du théorème suivant.

THÉORÈME 6.1 (Théorème d'existence). — *Supposons que X satisfait aux hypothèses* (a) *ou* (b) *du théorème 3.1. Alors l'application de réciprocité ρ_X induit une bijection entre les sous-groupes ouverts d'indice fini $U \subset C_X$ et les sous-groupes ouverts de $\rho_X(C_X) \subset \pi_1^{\mathrm{ab}}(X)$. En outre, tout U comme ci-dessus est de la forme $\phi_*(C_{X^1})$ pour un revêtement fini étale abélien $\phi : X^1 \to X$ convenable. Dans le cas (a) tout sous-groupe ouvert de C_X est d'indice fini.*

La démonstration de ce théorème comprend la majeure partie des innovations de Wiesend, mais elle utilise également des idées remontant aux articles de Bloch [**3**] et Katz–Lang [**23**]. *On se place dans la situation du cas* (a), *i.e. on suppose X plat sur* $\mathbf{Z}$. Pour les modifications à faire dans le cas (b), voir la remarque 7.5 ci-dessous.

Le cœur technique de la démonstration du théorème 6.1 est la proposition ci-dessous, dont la démonstration est reportée à la section suivante.

PROPOSITION 6.2. — *Soit $U \subset C_X$ un sous-groupe ouvert d'indice fini. Il existe un morphisme étale $\psi : Z \to X$ avec Z connexe et $\psi_*^{-1}(U) = C_Z$.*

Pour démontrer le théorème 6.1, nous avons encore besoin du lemme non trivial suivant.

LEMME 6.3. — *Soient R un anneau local normal intègre excellent, $S' \subset S = \mathrm{Spec}\,(R)$ un ouvert dense, et $T' \to S'$ un revêtement étale galoisien de degré premier. Considérons le normalisé T de S dans le corps de fonctions de T'. Si le morphisme $T \to S$ n'est pas étale en codimension 1, il existe un point p de dimension 1 de S', d'adhérence C sur S, tel que le revêtement $T \times_S \widetilde{C} \to \widetilde{C}$ du normalisé $\widetilde{C}$ soit ramifié.*

Le lemme est démontré par une analyse détaillée de la ramification dans [**26**], Lemma 2.4. Certaines idées de la preuve proviennent de [**53**].

COROLLAIRE 6.4. — *Soient S un schéma régulier connexe excellent, $S' \subset S$ un ouvert dense, et $T' \to S'$ un revêtement étale connexe. Notons T le normalisé de S dans le corps de fonctions de T'. Le morphisme $T \to S$ est étale si et seulement si pour tout point p de dimension 1 de S' le morphisme $T' \times_S \widetilde{C} \to S' \times_S \widetilde{C}$ s'étend en un revêtement étale du normalisé $\widetilde{C}$ de l'adhérence de p sur S.*

Démonstration. — La nécessité est évidente. Pour la suffisance on peut supposer $T \to S$ galoisien de groupe G en passant à la clôture galoisienne. Si $T \to S$ n'est pas étale, le théorème de pureté de Zariski–Nagata permet de trouver un point $t \in T$ de codimension 1 dont le sous-groupe d'inertie $I_t \subset G$ est non trivial. Prenons un sous-groupe $H_t \subset I_t$ de degré premier, et un point s du quotient T/H_t en dessous de t. Appliquant le lemme à l'anneau local R de T/H_t en s, on trouve un point p de dimension 1 de l'image réciproque de S' dans $\mathrm{Spec}\,(R)$ tel que T induit un revêtement ramifié du normalisé de l'adhérence de p. L'image de p dans S contredit l'hypothèse. □

Démonstration du théorème 6.1 (dans le cas où X est plat sur $\mathbf{Z}$). – Traitons d'abord le cas où il existe un morphisme $\psi : Z \to X$ comme dans la proposition 6.2 qui est en plus *fini.* En remplaçant Z par sa clôture galoisienne on peut le supposer galoisien de groupe G. Par hypothèse on a $\psi_*(C_Z) \subset U$, et par le théorème 5.1 on dispose d'un isomorphisme $\rho_{XZ} : C_X/\psi_*(C_Z) \xrightarrow{\sim} G^{\mathrm{ab}}$. Ceci montre que $\rho_X^{-1}(\rho_X(U)) = U$, il ne reste donc qu'à se rappeler que $\rho_X(U)$ est ouvert dans $\pi_1^{\mathrm{ab}}(X)$ (puisqu'il contient l'ouvert $\rho_Z(C_Z) = \pi_1^{\mathrm{ab}}(Z)$).

En général, la proposition 6.2 nous donne un $\psi : Z \to X$ non nécessairement fini. Toutefois, comme on l'a vu lors de la démonstration du lemme 4.4, on trouve un ouvert dense $X' \subset X$ au-dessus duquel ψ est fini. Par le paragraphe précédent l'image réciproque U' de U dans $C_{X'}$ correspond à un sous-groupe ouvert $\rho_{X'}(U') \subset \pi_1^{\mathrm{ab}}(X')$, d'où un revêtement étale abélien $Y' \to X'$. Il suffit alors de montrer que le normalisé Y de X dans le corps de fonctions de Y' est étale sur X ; en effet, on vérifie alors par restriction aux courbes sur X que le sous-groupe ouvert correspondant à Y ne peut être autre que $\rho_X(U)$, et l'on conclut comme ci-dessus. Maintenant pour que $Y \to X$ soit étale, il suffit de montrer d'après le corollaire précédent que pour toute courbe C sur X rencontrant X' le revêtement étale $Y' \times_X \widetilde{C} \to Y' \times_X \widetilde{C}$ s'étend en un revêtement étale du normalisé $\widetilde{C}$ de C. Mais une telle extension est possible par le théorème d'existence classique sur $\widetilde{C}$ (théorème 1.1 (5)) : elle est définie par le sous-groupe ouvert $\rho_{\widetilde{C}}(U_{\widetilde{C}}) \subset \pi_1^{\mathrm{ab}}(\widetilde{C})$, où $U_{\widetilde{C}}$ est l'image réciproque de $U \subset C_X$ par la poussette $C_{\widetilde{C}} \to C_X$.

Enfin, que U soit d'indice fini dans C_X résulte de la surjectivité de ρ_X et du fait que $\rho_X(U)$ est ouvert, donc d'indice fini dans $\pi_1^{\mathrm{ab}}(X)$. Soit A le quotient, et soit $\phi : X^1 \to X$ le revêtement fini étale abélien correspondant à la projection $\pi_1^{\mathrm{ab}}(X) \twoheadrightarrow A$. D'après le théorème 5.1 le noyau du morphisme composé $C_X \to \pi_1^{\mathrm{ab}}(X) \to A$ contient $\phi_*(C_{X^1})$, et les quotients modulo ces deux sous-groupes d'indice fini sont isomorphes à A. Ceci donne bien $U = \phi_*(C_{X^1})$. □

7. DÉMONSTRATION DE LA PROPOSITION 6.2

Dans cette section nous démontrons la proposition 6.2. Par la nature même de l'énoncé, on est autorisé à remplacer X par des ouverts étales tout au long de la preuve. On va procéder par récurrence sur la dimension, le cas de dimension 1 résultant de la théorie classique (on peut prendre pour ψ le revêtement étale abélien de X associé à U). Pour traiter le cas général on se sert de la théorie des fibrations élémentaires de M. Artin, que l'on peut résumer comme suit.

LEMME 7.1. — *Après avoir remplacé X par un ouvert étale, il existe des morphismes de schémas réguliers $\iota : X \hookrightarrow \mathcal{X}$ et $\pi : \mathcal{X} \to W$ satisfaisant aux propriétés suivantes :*

- *ι est une immersion ouverte dont l'image est dense dans chaque fibre de π ;*
- *π est projectif et lisse, à fibres géométriquement intègres de dimension 1 ;*
- *π induit un morphisme fini étale surjectif $\mathcal{X} \setminus \iota(X) \twoheadrightarrow W$;*
- *il existe une section $\sigma : W \to X$ de $\pi \circ \iota$.*

Démonstration. — Pour l'existence de ι et de π satisfaisant aux trois premières propriétés, voir **[2]**, exp. XI, §3. Comme $\pi \circ \iota$ est lisse, quitte à remplacer W par un ouvert étale $W' \to W$ et $\mathcal{X}$ par $X \times_W W'$ on trouve également une section σ de $\pi \circ \iota$. □

Par l'hypothèse de récurrence on peut supposer que la proposition vaut pour W. Ainsi, en le remplaçant par un ouvert étale *on peut supposer dans toute la suite que* $\sigma_*^{-1}(U) = C_W$.

Soit maintenant w un point fermé de W, et soit X_w la fibre de $\pi \circ \iota$ en w. D'après le théorème d'existence classique (théorème 1.1 (5)), l'image inverse $U_w \subset C_{X_w}$ de U par la poussette associée au morphisme $X_w \to X$ correspond à un sous-groupe ouvert $V_w \subset \pi_1^{\mathrm{ab}}(X_w)$. Soit $\bar{w}$ un point géométrique au-dessus de w. La courbe $X_{\bar{w}}$ est connexe, donc on dispose d'une suite exacte « d'homotopie »

$$1 \to \pi_1(X_{\bar{w}}, \bar{x}) \to \pi_1(X_w, \bar{x}) \to \mathrm{Gal}\,(\overline{\kappa(w)}|\kappa(w)) \to 1$$

pour un point géométrique $\bar{x}$ convenable. La section σ induit un point $\kappa(w)$-rationnel de X_w, d'où un scindage de la suite exacte. Passant aux abélianisés, on obtient alors

une suite exacte scindée

$$0 \to \pi_1^{\mathrm{ab}}(X_{\bar{w}})_{\mathrm{Gal}\,(\overline{\kappa(w)}|\kappa(w))} \to \pi_1^{\mathrm{ab}}(X_w) \to \mathrm{Gal}\,(\overline{\kappa(w)}|\kappa(w)) \to 0$$

où $\pi_1^{\mathrm{ab}}(X_{\bar{w}})_{\mathrm{Gal}\,(\overline{\kappa(w)}|\kappa(w))}$ est le groupe des coinvariants de $\mathrm{Gal}\,(\overline{\kappa(w)}|\kappa(w))$, i.e. le quotient maximal de $\pi_1^{\mathrm{ab}}(X_{\bar{w}})$ avec action triviale de $\mathrm{Gal}\,(\overline{\kappa(w)}|\kappa(w))$ (voir la section 5.8 de **[49]** pour plus de détails).

Remarque 7.2. — Notre hypothèse $\sigma_*^{-1}(U) = C_W$ implique que le revêtement $Y_w \to X_w$ associé à V_w est géométriquement intègre. En effet, pour toute courbe $C \subset \sigma(W)$ contenant $\sigma(w)$ dans son lieu régulier (une telle C existe d'après le lemme 5.2) le revêtement étale du normalisé $\widetilde{C}$ correspondant à l'image inverse de U dans $C_{\widetilde{C}}$ est trivial. En particulier, la fibre en $\sigma(w)$ est complètement décomposée. Alors il en est de même pour la fibre en $\sigma(w)$ de $Y_w \to X_w$ (en effet, dans les deux cas le sous-groupe de décomposition d'un point au-dessus de $\sigma(w)$ est un groupe cyclique dont l'ordre est égal à l'ordre de l'image de $\sigma(w)$ dans C_X/U, par la théorie des corps de classes classique – donc la trivialité de l'un implique la trivialité de l'autre). Ceci n'est possible que si Y_w est géométriquement intègre, puisque $\sigma(w)$ est un point $\kappa(w)$-rationnel de X_w.

Par la remarque qu'on vient de faire, le sous-groupe V_w se surjecte sur $\mathrm{Gal}\,(\overline{\kappa(w)}|\kappa(w))$, d'où un isomorphisme

$$\pi_1^{\mathrm{ab}}(X_w)/V_w \cong \pi_1^{\mathrm{ab}}(X_{\bar{w}})_{\mathrm{Gal}\,(\overline{\kappa(w)}|\kappa(w))}/(V_w \cap \pi_1^{\mathrm{ab}}(X_{\bar{w}})_{\mathrm{Gal}\,(\overline{\kappa(w)}|\kappa(w))}).$$

Notons n_w l'exposant de ces groupes abéliens finis. Cet exposant est majoré par l'ordre de $\pi_1^{\mathrm{ab}}(X_{\bar{w}})_{\mathrm{Gal}\,(\overline{\kappa(w)}|\kappa(w))}$ qui est un groupe fini d'après Katz–Lang. On veut maintenant majorer les n_w *indépendamment de* w. Pour des raisons techniques on arrive à le faire sous l'hypothèse supplémentaire qu'ils sont tous des puissances d'un nombre premier ℓ.

LEMME 7.3. — *Supposons que les n_w soient des puissances d'un nombre premier ℓ fixé. Alors quitte à remplacer W (et a fortiori X) par un ouvert étale, ils admettent une borne supérieure finie.*

Démonstration. — Pour un point géométrique $\bar{x}$ de W écrivons $\Pi_{\bar{x}}^{\ell}$ pour la pro-ℓ-composante de $\pi_1^{\mathrm{ab}}(X_{\bar{x}})$. Quitte à remplacer W par un ouvert de Zariski, on peut supposer ℓ inversible sur W. Il résulte du théorème de changement de base lisse en cohomologie étale (**[30]**, Chapter VI, §4) que les $\Pi_{\bar{x}}^{\ell}$ sont les fibres géométriques d'un système local ℓ-adique, à savoir le dual du système local $\mathbf{R}^1(\pi \circ \iota)_*\mathbf{Q}_\ell/\mathbf{Z}_\ell$. Pour η le point générique et w un point fermé de W, on peut identifier $\mathrm{Gal}\,(\overline{\kappa(w)}|\kappa(w))$ à un groupe de décomposition D_w dans $\Gamma := \mathrm{Gal}\,(\overline{\kappa(\eta)}|\kappa(\eta))$. L'isomorphisme de spécialisation $\Pi_{\bar{\eta}}^{\ell} \cong \Pi_{\bar{w}}^{\ell}$ donné par le théorème de changement de base lisse est compatible

avec l'action de D_w, d'où une surjection $(\Pi^\ell_{\bar{w}})_{D_w} \twoheadrightarrow (\Pi^\ell_{\bar{\eta}})_\Gamma$. Puisque ces deux groupes sont finis par Katz–Lang, on trouve un sous-groupe ouvert $H \subset \Gamma$ contenant D_w tel que $(\Pi^\ell_{\bar{w}})_{D_w}$ soit envoyé isomorphiquement sur le groupe des coinvariants de H sur $\Pi^\ell_{\bar{\eta}}$. Soit W_H le normalisé de W dans $\overline{\kappa(\eta)}^H$; puisque $H \supset D_w$, il induit un revêtement étale d'un ouvert de W. Remplaçant W par l'ouvert étale ainsi obtenu et w par une image réciproque, on peut supposer que la surjection $(\Pi^\ell_{\bar{w}})_{D_w} \to (\Pi^\ell_{\bar{\eta}})_\Gamma$ est un isomorphisme. Tout point fermé w' ayant un groupe de décomposition $D_{w'} \subset \Gamma$ contenant D_w a la même propriété.

L'ordre ℓ^N de $(\Pi^\ell_{\bar{\eta}})_\Gamma$ fournit donc une majoration des exposants $n_{w'}$ pour tout w' comme ci-dessus. Suivant une astuce remarquable de Wiesend [**51**], on en déduit une majoration en tous les points fermés de W comme suit. Soit $W' \to W$ un revêtement fini étale trivialisant le système local des groupes abéliens finis $\Pi^\ell_{\bar{x}}/\ell^N\Pi^\ell_{\bar{x}}$, et soit ℓ^d son degré. On va montrer que la borne ℓ^{N+d} convient.

Supposons le contraire : il existe un point fermé w' avec $n_{w'} > \ell^{N+d}$. Rappelons que $n_{w'}$ est l'exposant du groupe abélien $\pi_1^{\mathrm{ab}}(X_{w'})/V_{w'}$. Le théorème de densité de Tchébotarev implique alors qu'il existe un point fermé $x' \in X_{w'}$ dont le groupe de décomposition dans $\pi_1^{\mathrm{ab}}(X_{w'})/V_{w'}$ est d'ordre $> \ell^{N+d}$. Par le lemme 5.2 on peut trouver une courbe C sur X, de normalisé $\widetilde{C}$, contenant x' et $x := \sigma(w)$ comme points réguliers. L'image réciproque de U dans $C_{\widetilde{C}}$ correspond à un revêtement étale abélien $\widetilde{D} \to \widetilde{C}$; notons G son groupe de Galois. L'exposant de G est $> \ell^{N+d}$, puisqu'il contient le groupe de décomposition de x' (qui a le même ordre dans G que dans $\pi_1^{\mathrm{ab}}(X_{w'})/V_{w'}$, à savoir l'ordre de x' dans C_X/U). Ainsi l'indice du sous-groupe $G_N \subset G$ des éléments d'ordre au plus ℓ^N est $> \ell^d$, donc par Tchébotarev la densité de Dirichlet des points fermés $x \in \widetilde{C}$ dont le groupe de décomposition est contenu dans G_N est $< 1/\ell^d$.

Or, en comparant cette fois avec $\pi_1^{\mathrm{ab}}(X_w)/V_w$, on voit que le groupe de décomposition dans G de x est contenu dans G_N. Par les considérations ci-dessus, la même chose vaut pour les points fermés de C contenus dans des fibres au-dessus de $w' \in W$ avec $D_{w'} \supset D_w$. Minorons la densité de ces points. D'après le complément 5.3 on peut supposer que le revêtement $C' := W' \times_W \widetilde{C} \cong (W' \times_W X) \times_X \widetilde{C}$ de $\widetilde{C}$ est intègre. Le groupe de Galois du revêtement $W' \to W$, et a fortiori de $C' \to \widetilde{C}$ est un quotient de Γ, donc on peut y considérer les images des $D_{w'}$. Puisque le degré commun de $W' \to W$ et $C' \to \widetilde{C}$ est ℓ^d, Tchébotarev nous dit encore que la densité des points fermés qui nous intéressent est au moins $1/\ell^d$. Contradiction. □

La réduction suivante montre que pour démontrer la proposition 6.2 on peut effectivement se placer dans la situation du lemme ci-dessus.

LEMME 7.4. — *Il suffit de démontrer la proposition 6.2 dans le cas où C_X/U est de torsion ℓ-primaire pour un nombre premier ℓ.*

Démonstration. — Le groupe C_X/U est en tout cas un groupe de torsion, puisqu'il est quotient de la somme des C_{X_w}/U_w qui le sont (utiliser le lemme 2.3). Pour ℓ fixé soit $U_\ell \supset U$ le sous-groupe ouvert de C_X pour lequel C_X/U_ℓ est la composante ℓ-primaire de C_X/U. Si la proposition vaut pour U_ℓ, on trouve un morphisme étale $\psi_\ell : Z_\ell \to X$ trivialisant U_ℓ, de degré générique une puissance de ℓ. Il se restreint à un revêtement étale d'un ouvert de X, trivial sur la trace de $\sigma(W)$. Ainsi, la fibre du revêtement au point générique η de W est géométriquement intègre ; elle correspond donc à un quotient de $\pi_1^{\mathrm{ab}}(X_{\bar\eta})_{\mathrm{Gal}\,(\overline{\kappa(\eta)}|\kappa(\eta))}$. Or ce dernier groupe est fini par le théorème de Katz–Lang (théorème 4.3). A fortiori, quand ℓ est premier à son ordre, le revêtement est un isomorphisme, et il en est de même pour le morphisme ψ_ℓ. On conclut alors en prenant pour ψ le produit fibré des ψ_ℓ. □

Fin de la démonstration de la proposition 6.2.– Par le lemme 7.4 on peut supposer que C_X/U est de torsion ℓ-primaire pour un nombre premier ℓ.

Soient maintenant η le point générique de W, et $\bar\eta$ un point géométrique au-dessus de η. La fibre $X_{\bar\eta}$ est une courbe lisse sur un corps algébriquement clos de caractéristique 0, donc son groupe fondamental est topologiquement de type fini. *A fortiori*, pour tout $n > 0$ l'intersection des sous-groupes ouverts $V \subset \pi_1^{\mathrm{ab}}(X_{\bar\eta})^{(\ell)}$ tels que l'exposant de $\pi_1^{\mathrm{ab}}(X_{\bar\eta})^{(\ell)}/V$ soit au plus ℓ^n est un sous-groupe ouvert, correspondant à un revêtement étale $Z_{\bar\eta} \to X_{\bar\eta}$. Quitte à remplacer $k(W)$ par une extension finie et W, X par des ouverts étales, on peut supposer que $Z_{\bar\eta}$ provient par changement de base d'un revêtement étale $Z_\eta \to X_\eta$. Quitte à remplacer encore X par un ouvert de Zariski et W par l'image de cet ouvert, on peut étendre Z_η en un revêtement étale $Z \to X$; pour un point fermé $w \in W$ il induit un revêtement $Z_w \to X_w$. Si $\bar w$ est un point géométrique au-dessus de w, l'isomorphisme de spécialisation $\pi_1^{\mathrm{ab}}(X_{\bar\eta})^{(\ell)} \xrightarrow{\sim} \pi_1^{\mathrm{ab}}(X_{\bar w})^{(\ell)}$ montre que le sous-groupe ouvert de $\pi_1^{\mathrm{ab}}(X_{\bar w})^{(\ell)}$ correspondant à $Z_{\bar w}$ est encore contenu dans l'intersection des sous-groupes ouverts tels que le quotient soit d'indice au plus ℓ^n.

Prenons maintenant pour ℓ^n la borne donnée par le lemme 7.3, et considérons pour chaque point fermé w de W le revêtement étale abélien $Y_w \to X_w$ correspondant au sous-groupe ouvert $V_w \subset \pi_1^{\mathrm{ab}}(X_w)$ défini ci-dessus. Nous allons montrer que, quitte à remplacer X et à fortiori Z par des ouverts étales, les revêtements étales

$$Y_w \times_X Z = Y_w \times_{X_w} (Z \times_X X_w) \to Z \times_X X_w$$

sont triviaux pour tout w, ce qui impliquera l'égalité $\psi_*^{-1}(U) = C_Z$. En effet, chaque composante D d'un $Z \times_X X_w$ est une courbe lisse sur Z, et le revêtement étale $Y_w \times_{X_w} D \to D$ correspond à l'image réciproque U_D de $U \subset C_X$ dans C_D par la

poussette associée au morphisme composé $D \to Z \to X$. La trivialité du revêtement veut dire que $U_D = C_D$. On peut alors conclure par le lemme 2.3, puisque la réunion des courbes D comme ci-dessus couvre Z, donc induit une surjection $\bigoplus_D Z_0(D) \twoheadrightarrow Z_0(Z)$.

Pour démontrer la trivialité requise, notons d'abord que le choix de n implique que le revêtement étale $Y_w \times_X Z \to Z \times_X X_w$ induit un revêtement trivial de $Z \times_X X_{\bar{w}}$ pour tout point géométrique $\bar{w}$ au-dessus de w. Considérons le normalisé $\widetilde{W}$ de W dans le corps de fonctions de Z. Le composé $Z \to X \to W$ se factorise à travers un morphisme $Z \to \widetilde{W}$ à fibres lisses et géométriquement connexes. Remplaçons X par Z, W par $\widetilde{W}$ et U par $\psi_*^{-1}(U)$. Pour tout point fermé $w \in W$ le revêtement $Y_w \to X_w$ associé à U_w induit un revêtement trivial de $X_{\bar{w}}$ pour tout point géométrique $\bar{w}$ au-dessus de w. Quitte à remplacer W par un ouvert étale convenable on peut supposer que le morphisme lisse $X \to W$ admet une section σ, et que $\sigma_*^{-1}(U) = C_W$ (ce dernier fait par l'hypothèse de récurrence). Par le même raisonnement que dans la remarque 7.2, chaque revêtement $Y_w \to X_w$ est alors trivial au-dessus du point rationnel $\sigma(w)$, ce qui n'est possible que si le revêtement lui-même est trivial. □

Remarque 7.5. — Dans le cas (b) du théorème 3.1 les arguments ci-dessus doivent être modifiés. L'observation principale est que les revêtements $Y_w \to X_w$ sont tous *modérément ramifiés* aux points à l'infini de la compactification lisse $\mathcal{X}_w$ de X_w. En fait, par la théorie des corps de classes classique pour la courbe X_w on obtient même une extension en un revêtement étale de $\mathcal{X}_w$ en prenant le revêtement associé à l'image de U_w par la surjection $C_{X_w} \twoheadrightarrow CH_0(\mathcal{X}_w)$.

Ainsi, dans la démonstration de la proposition 6.2 on peut travailler avec le groupe fondamental modéré $\pi_1^{\mathrm{mod}}(X_{\bar{x}})$ de $X_{\bar{x}}$; la surjectivité du morphisme de spécialisation $\pi_1^{\mathrm{mod,ab}}(X_{\bar{\eta}}) \to \pi_1^{\mathrm{mod,ab}}(X_{\bar{w}})$ vaut alors également pour la p-partie (voir [**32**] pour une excellente exposition). Cette surjectivité suffit pour faire marcher l'argument.

Par contre, dans la démonstration du lemme 7.3 on n'obtient plus de système local pour $\ell = p$. Toutefois, comme expliqué dans ([**25**], §5), les pro-p-composantes $\Pi_{\bar{x}}^p$ des groupes $\pi_1^{\mathrm{ab,mod}}(X_{\bar{x}})$ (qui sont les mêmes que celles des $\pi_1^{\mathrm{ab}}(\mathcal{X}_{\bar{x}})$) sont les fibres d'un faisceau constructible, donc localement constant sur un ouvert de Zariski $W' \subset W$. Alternativement, on peut appliquer un résultat de Kato et Saito ([**21**], Proposition 6) : ils démontrent en utilisant la théorie des groupes p-divisibles qu'il existe un ouvert de Zariski de W au-dessus duquel on a des isomorphismes $(\Pi_{\bar{w}}^p)_{D_w} \cong (\Pi_{\bar{\eta}}^p)_\Gamma$. Après restriction à cet ouvert la preuve se conclut de la même façon.

Enfin, dans la démonstration du théorème 6.1 on a utilisé la surjectivité de ρ_X, ce qui ne vaut plus dans le cas géométrique ; il faut alors modifier les arguments en travaillant avec la partie de degré 0 des groupes en question.

8. ANNEXE : L'APPROCHE COHOMOLOGIQUE POUR UNE VARIÉTÉ PROJECTIVE ET LISSE SUR UN CORPS FINI

Cette section annexe est consacrée à une version « remastérisée » d'une démonstration plus ancienne du corollaire 3.3, dans le cas où X est projectif, lisse et géométriquement connexe sur un corps fini $\mathbf{F}$ de caractéristique p. Rappelons l'énoncé :

THÉORÈME. — *Sous les hypothèses ci-dessus, l'application de réciprocité induit un isomorphisme de groupes finis*

$$CH_0(X)^0 \xrightarrow{\sim} \pi_1^{\mathrm{ab}}(X)^0.$$

Notons qu'ici l'application de réciprocité ρ_X n'est autre que celle de Lang [**28**] : on envoie un point fermé sur l'élément de Frobenius associé. La surjectivité résulte, comme précédemment, de l'énoncé de densité de Lang (Proposition 2.3) et de la finitude du groupe $\pi_1^{\mathrm{ab}}(X)^0$ (théorème 4.3).

Nous démontrons maintenant l'injectivité par un argument de Colliot-Thélène–Sansuc–Soulé [**5**] (complété par Gros [**11**] pour la partie p-primaire) que nous présentons dans l'habit neuf de la cohomologie motivique.

Soit d la dimension de X. Pour un nombre premier ℓ et des entiers $n > 0, i \geq 0$, considérons les objets de la catégorie dérivée des faisceaux étales sur X donnés par

$$\mathbf{Z}/\ell^n\mathbf{Z}(i) := \begin{cases} \mu_{\ell^n}^{\otimes i} & \ell \neq p\,; \\ \nu_n(i)[-i] & \ell = p\,, \end{cases}$$

où $\nu_n(i)$ est le sous-faisceau du faisceau de Hodge–Witt $W_n\Omega_X^i$ d'Illusie [**13**] engendré étale localement par des différentielles logarithmiques $(dx_1/x_1) \wedge \cdots \wedge (dx_i/x_i)$. Le décalage par $-i$ veut dire, comme d'habitude, qu'il est placé en degré i.

LEMME 8.1. — *Il existe des isomorphismes canoniques*

$$\pi_1^{\mathrm{ab}}(X)/\ell^n \cong H^{2d}_{\text{ét}}(X, \mathbf{Z}/\ell^n\mathbf{Z}(d))$$

pour tout ℓ^n.

Démonstration. — Le groupe abélien dual de $\pi_1^{\mathrm{ab}}(X)/\ell^n$ est le groupe fini $H^1_{\text{ét}}(X, \mathbf{Z}/\ell^n\mathbf{Z})$. Ce dernier est dual de $H^{2d}_{\text{ét}}(X, \mathbf{Z}/\ell^n\mathbf{Z}(d))$ par une combinaison des dualités de Tate et de Poincaré pour $\ell \neq p$, et par une dualité de Milne en cohomologie plate pour $\ell = p$ (cf. [**35**], théorèmes 1.12 et 1.13). □

Rappelons maintenant que l'on dispose pour tout $i \geq 0$ d'applications cycle

$$c^i_{\ell^n} : CH^i(X) \to H^{2i}_{\text{ét}}(X, \mathbf{Z}/\ell^n\mathbf{Z}(i)),$$

le cas $\ell = p$ ayant été défini par Gros [**10**].

LEMME 8.2. — *En composant $c^d_{\ell^n}$ avec l'isomorphisme du lemme précédent on obtient l'application de réciprocité modulo ℓ^n.*

Démonstration. — Appliquant la fonctorialité covariante des deux morphismes aux inclusions $\operatorname{Spec}\kappa(x) \to X$ des points fermés de X, on se réduit au cas de dimension 0, où l'assertion est évidente. □

Observons que le groupe $CH_0(X)^0$ est de torsion ; cela se vérifie en considérant la surjection $\bigoplus_C CH_0(\widetilde{C})^0 \twoheadrightarrow CH_0(X)^0$, où C parcourt les courbes sur X. Notons $CH_0(X)^0\{\ell\}$ la composante ℓ-primaire. Par le lemme ci-dessus, l'énoncé d'injectivité du théorème est équivalent à l'injectivité des applications cycle

$$c^i_\ell : CH_0(X)^0\{\ell\} \to H^{2d}_{\text{ét}}(X, \mathbf{Z}_\ell(d))$$

pour tout ℓ, obtenues à partir des $c^i_{\ell^n}$ par passage à la limite projective. Pour démontrer cette injectivité, on va utiliser la théorie des complexes motiviques.

Étant donnés une variété lisse V sur un corps parfait et un entier $i \geq 0$, on définit le complexe motivique $\mathbf{Z}(i)$ sur le site zariskien V_{Zar} de V par $\mathbf{Z}(i) := z^i(-,*)[-2i]$, où $z^i(U,*)$ est le complexe définissant les groupes de Chow supérieurs $CH^i(U,*)$ de Bloch [**4**] ; Suslin et Voevodsky ont donné dans [**47**] une autre construction donnant lieu aux mêmes groupes de cohomologie. Le complexe $\mathbf{Z}(i)$ est acyclique en degrés $> i$, et ses termes sont des faisceaux abéliens sans torsion. On va utiliser deux de ses propriétés. Le premier est l'isomorphisme

$$\mathbf{H}^{2i}_{\text{Zar}}(V, \mathbf{Z}(i)) \cong CH^i(V) \tag{1}$$

valable pour tout $i \geq 0$. Le deuxième est le quasi-isomorphisme de complexes de faiscaux étales

$$\pi^*\mathbf{Z}(i) \otimes \mathbf{Z}/\ell^n\mathbf{Z} \xrightarrow{\sim} \mathbf{Z}/\ell^n\mathbf{Z}(i)$$

où $\pi : V_{\text{ét}} \to V_{\text{Zar}}$ est le morphisme de changement de site. Ici on emploie la même convention que ci-dessus ; en particulier, pour $\ell = \operatorname{car}(F)$ on a $\mathbf{Z}/\ell^n\mathbf{Z}(i) = \nu_n(i)[-i]$. Cet isomorphisme a été démontré par Suslin–Voevodsky [**47**] en caractéristique 0, et par Geisser–Levine [**8**], [**9**] en général. Joint à l'acyclicité de $\mathbf{Z}(i)$ en degrés $> i$, il fournit un morphisme

$$s_{i,\ell^n} : \mathbf{Z}(i) \otimes \mathbf{Z}/\ell^n\mathbf{Z} \to \tau_{\leq i}\mathbf{R}\pi_*\mathbf{Z}/\ell^n\mathbf{Z}(i)$$

dans la catégorie dérivée des faisceaux sur V_{Zar}, où $\tau_{\leq i}$ est la troncation sophistiquée. C'est un isomorphisme pour $\ell = \operatorname{car}(F)$ d'après ([**8**], Theorem 8.5). Pour $\ell \neq \operatorname{car}(\mathrm{F})$ un célèbre théorème de Suslin–Voevodsky [**46**], étendu en caractéristique quelconque par Geisser–Levine [**9**], montre que s_{i,ℓ^n} est un isomorphisme si et seulement si la célèbre conjecture de Bloch–Kato–Milnor sur la K-théorie de Milnor vaut en degré i à coefficients $\mu_{\ell^n}^{\otimes i}$. La conjecture est maintenant un théorème grâce aux efforts de

Rost, Voevodsky, Weibel et leurs collaborateurs. Dans l'argument ci-dessous seul le cas $i = 2$, démontré par Merkurjev–Suslin [**29**], sera utilisé.

Le triangle distingué

$$\mathbf{Z}(i) \xrightarrow{\ell^n} \mathbf{Z}(i) \to \mathbf{Z}(i) \otimes \mathbf{Z}/\ell^n\mathbf{Z} \to \mathbf{Z}(i)[1]$$

et l'isomorphisme (1) induisent une surjection

$$\mathbf{H}^{2i-1}_{\mathrm{Zar}}(V, \mathbf{Z}(i) \otimes \mathbf{Z}/\ell^n\mathbf{Z}) \twoheadrightarrow {}_{\ell^n}CH^i(V),$$

où ${}_{\ell^n}CH^i(V)$ est le sous-groupe de ℓ^n-torsion de $CH^i(V)$. Elle s'insère dans le diagramme commutatif

$$\begin{array}{ccc} \mathbf{H}^{2i-1}_{\mathrm{Zar}}(V, \mathbf{Z}(i) \otimes \mathbf{Z}/\ell^n\mathbf{Z}) & \longrightarrow & {}_{\ell^n}CH^i(V) \\ \downarrow & & \downarrow \\ H^{2i-1}_{\text{ét}}(V, \mathbf{Z}/\ell^n\mathbf{Z}(i)) & \longrightarrow & H^{2i}_{\text{ét}}(V, \mathbf{Z}/\ell^m\mathbf{Z}(i)) \end{array}$$

où le morphisme vertical de gauche est le morphisme

$$\mathbf{H}^{2i-1}_{\mathrm{Zar}}(V, \mathbf{Z}(i) \otimes \mathbf{Z}/\ell^n\mathbf{Z}) \to \mathbf{H}^{2i-1}_{\mathrm{Zar}}(V, \tau_{\leq i}\mathbf{R}\pi_*\mathbf{Z}/\ell^n\mathbf{Z}(i))$$

induit par $s^i_{\ell^n}$ suivi du morphisme naturel

$$\mathbf{H}^{2i-1}_{\mathrm{Zar}}(V, \tau_{\leq i}\mathbf{R}\pi_*\mathbf{Z}/\ell^n\mathbf{Z}(i)) \to \mathbf{H}^{2i-1}_{\mathrm{Zar}}(V, \mathbf{R}\pi_*\mathbf{Z}/\ell^n\mathbf{Z}(i)) \cong H^{2i-1}_{\text{ét}}(V, \mathbf{Z}/\ell^n\mathbf{Z}(i)).$$

Le morphisme de droite dans le diagramme est l'application cycle, et celui du bas est le morphisme de Bockstein provenant de la suite exacte

$$0 \to \mathbf{Z}/\ell^n\mathbf{Z}(i) \to \mathbf{Z}/\ell^{n+m}\mathbf{Z}(i) \to \mathbf{Z}/\ell^m\mathbf{Z}(i) \to 0 \tag{2}$$

sur $V_{\text{ét}}$ ([**5**], proposition 1 et [**11**], (1.1.4)). La commutativité du diagramme résulte de la compatibilité des morphismes $s^i_{\ell^n}$ avec les suites exactes de type (2) ; pour la vérification détaillée d'une compatibilité similaire, voir le §3 de [**48**].

Plaçons-nous maintenant au-dessus du corps fini $\mathbf{F}$, et substituons la variété propre et lisse X du début à la place de V. Prenant $i = d = \dim X$, par passage à la limite inductive suivant n et à la limite projective suivant m dans le diagramme ci-dessus, on obtient le diagramme commutatif

$$\begin{array}{ccc} \mathbf{H}^{2d-1}_{\mathrm{Zar}}(X, \mathbf{Z}(d) \otimes \mathbf{Q}_\ell/\mathbf{Z}_\ell) & \longrightarrow & CH^d(X)\{\ell\} \\ \downarrow & & \downarrow \\ H^{2d-1}_{\text{ét}}(X, \mathbf{Q}_\ell/\mathbf{Z}_\ell(d)) & \longrightarrow & H^{2d}_{\text{ét}}(X, \mathbf{Z}_\ell(d)). \end{array}$$

Ici on obtient les groupes de cohomologie étale habituels pour $\ell \neq p$ (ceci utilise la finitude des groupes à coefficients finis) ; pour $\ell = p$ on les *définit* comme les limites des groupes à coefficients finis.

Remarquons que le morphisme horizontal du bas est injectif. En effet, pour $\ell \neq p$ son noyau est couvert par $H^{2d-1}(X, \mathbf{Q}_\ell(d))$; or ce groupe est trivial, ce qu'on

voit par un argument standard utilisant la suite spectrale de Hochschild–Serre et le théorème de Deligne sur les poids de Frobenius (cf. par exemple [**5**], théorème 2 pour les détails). Pour $\ell = p$ l'injectivité se démontre de manière analogue, utilisant les résultats de Katz–Messing sur les poids de Frobenius sur la cohomologie cristalline (cf. [**11**], lemme 2.1.16).

On voit donc que pour démontrer l'injectivité du morphisme vertical de droite dans le diagramme ci-dessus, il suffit de démontrer l'injectivité de celui de gauche. Or la considération de la suite spectrale d'hypercohomologie et la nullité des groupes $H^i_{\text{ét}}(U, \mathbf{Q}/\mathbf{Z}_\ell(d)$ pour $U \subset X$ affine et $i > d+1$ montrent que pour $\ell \neq p$ le noyau du morphisme $\mathbf{H}^{2d-1}_{\text{Zar}}(X, \tau_{\leq d}\mathbf{R}\pi_*\mathbf{Q}_\ell/\mathbf{Z}_\ell(d)) \to H^{2d-1}_{\text{ét}}(X, \mathbf{Q}_\ell/\mathbf{Z}_\ell(d))$ est couvert par le groupe $\mathbf{H}^{d-3}_{\text{Zar}}(X, \mathbf{R}^{d+1}\pi_*\mathbf{Q}_\ell/\mathbf{Z}_\ell(d))$. Ce groupe s'annule trivialement pour $d \leq 2$, et on a la bijectivité des morphismes $c^d_{\ell^n}$ grâce à Merkurjev–Suslin pour $d = 2$. Ceci montre l'injectivité du morphisme de gauche pour $d \leq 2$ et $\ell \neq p$; pour $\ell = p$ son injectivité résulte de la construction ([**11**], (2.1.4)). Le théorème est donc démontré pour $d \leq 2$.

Pour $d > 2$ la nullité du groupe $\mathbf{H}^{d-3}_{\text{Zar}}(X, \mathbf{R}^{d+1}\pi_*\mathbf{Q}_\ell/\mathbf{Z}_\ell(d))$ résulterait d'une des conjectures de Kato [**18**] mentionnées dans l'introduction, sous la seule hypothèse que X est propre et lisse sur un corps fini. Cette conjecture n'est que partiellement démontrée à ce jour. Mais on peut démontrer le cas général du théorème par une astuce de Colliot-Thélène qui permet de le réduire au cas de dimension 2 ; c'est ici que l'hypothèse de projectivité sert. Supposons en effet le résultat connu en dimension $d-1$. Étant donné un zéro-cycle α dans le noyau du morphisme $\rho_X : CH_0(X)^0 \to \pi_1^{\text{ab}}(X)^0$, le théorème de Bertini–Gabber–Poonen permet de trouver une section d'hypersurface lisse $Y \subset X$ contenant le support de α. La fonctorialité du groupe de Chow et celle du groupe fondamental fournissent un diagramme commutatif

$$\begin{array}{ccc} CH_0(Y)^0 & \xrightarrow{\rho_Y} & \pi_1^{\text{ab}}(Y)^0 \\ \downarrow & & \downarrow \\ CH_0(X)^0 & \xrightarrow{\rho_X} & \pi_1^{\text{ab}}(X)^0 \end{array}$$

où le morphisme vertical de droite est un isomorphisme par le théorème de Lefschetz (applicable puisque $d \geq 3$). Ainsi, $\alpha = 0$ dans $CH_0(Y)^0$ par l'hypothèse de récurrence, ce qui conclut la démonstration.

RÉFÉRENCES

[1] E. Artin & J. Tate – *Class field theory*, W. A. Benjamin, Inc., New York-Amsterdam, 1968.

[2] M. Artin, A. Grothendieck & J.-L. Verdier (éds.) – *Théorie des topos et cohomologie étale des schémas. Tome 3 (SGA 4)*, Lecture Notes in Math., vol. 305, Springer, 1973.

[3] S. Bloch – Algebraic K-theory and classfield theory for arithmetic surfaces, *Ann. of Math.* **114** (1981), p. 229–265.

[4] ———, Algebraic cycles and higher K-theory, *Adv. in Math.* **61** (1986), p. 267–304.

[5] J.-L. Colliot-Thélène, J.-J. Sansuc & C. Soulé – Torsion dans le groupe de Chow de codimension deux, *Duke Math. J.* **50** (1983), p. 763–801.

[6] I. Fesenko & M. Kurihara (éds.) – *Invitation to higher local fields*, Geometry & Topology Monographs, vol. 3, Geometry & Topology Publications, Coventry, 2000.

[7] O. Gabber – On space filling curves and Albanese varieties, *Geom. Funct. Anal.* **11** (2001), p. 1192–1200.

[8] T. Geisser & M. Levine – The K-theory of fields in characteristic p, *Invent. Math.* **139** (2000), p. 459–493.

[9] ———, The Bloch-Kato conjecture and a theorem of Suslin-Voevodsky, *J. reine angew. Math.* **530** (2001), p. 55–103.

[10] M. Gros – Classes de Chern et classes de cycles en cohomologie de Hodge-Witt logarithmique, *Mém. Soc. Math. France* **21** (1985).

[11] ———, Sur la partie p-primaire du groupe de Chow de codimension deux, *Comm. Algebra* **13** (1985), p. 2407–2420.

[12] A. Grothendieck (éd.) – *Revêtements étales et groupe fondamental (SGA 1)*, Lecture Notes in Math., vol. 224, Springer, 1971. Nouvelle édition annotée : Documents Mathématiques, vol. 3, Soc. Math. France, 2003.

[13] L. Illusie – Complexe de de Rham-Witt et cohomologie cristalline, *Ann. Sci. École Norm. Sup.* **12** (1979), p. 501–661.

[14] U. Jannsen – Hasse principles for higher-dimensional fields, prépublication, 2005.

[15] U. Jannsen & S. Saito – Kato homology of arithmetic schemes and higher class field theory over local fields, *Doc. Math.* (2003), p. 479–538.

[16] A. J. de Jong – Smoothness, semi-stability and alterations, *Publ. Math. I.H.É.S.* **83** (1996), p. 51–93.

[17] K. Kato – A generalization of local class field theory by using K-groups. I, II, III, *J. Fac. Sci. Univ. Tokyo Sect. IA Math.* **26** (1979), p. 303–376, **27** (1980), p. 603–683, **29** (1982), p. 31–43.

[18] ———, A Hasse principle for two-dimensional global fields, *J. reine angew. Math.* **366** (1986), p. 142–183.

[19] ———, Class field theory, $\mathscr{D}$-modules, and ramification on higher-dimensional schemes. I, *Amer. J. Math.* **116** (1994), p. 757–784.

[20] K. Kato & S. Saito – Two-dimensional class field theory, in *Galois groups and their representations (Nagoya, 1981)*, Adv. Stud. Pure Math., vol. 2, North-Holland, 1983, p. 103–152.

[21] ———, Unramified class field theory of arithmetical surfaces, *Ann. of Math.* **118** (1983), p. 241–275.

[22] ———, Global class field theory of arithmetic schemes, in *Applications of algebraic K-theory to algebraic geometry and number theory, Part I, II (Boulder, Colo., 1983)*, Contemp. Math., vol. 55, Amer. Math. Soc., 1986, p. 255–331.

[23] N. M. Katz & S. Lang – Finiteness theorems in geometric classfield theory, *Enseign. Math.* **27** (1981), p. 285–319.

[24] M. Kerz – Higher class field theory and the connected component, prépublication arXiv:0711.4485.

[25] M. Kerz & A. Schmidt – Covering data and higher dimensional global class field theory, *J. Number Theory* **129** (2009), p. 2569–2599.

[26] ———, On different notions of tameness in arithmetic geometry, *Math. Ann.* **346** (2010), p. 641–668.

[27] S. Lang – Sur les séries L d'une variété algébrique, *Bull. Soc. Math. France* **84** (1956), p. 385–407.

[28] ———, Unramified class field theory over function fields in several variables, *Ann. of Math.* **64** (1956), p. 285–325.

[29] A. Merkurjev & A. Suslin – K-cohomologie des variétés de Severi-Brauer et l'homomorphisme de résidu normique (en russe), *Izv. Akad. Nauk SSSR* **46** (1982) ; traduction anglaise : *Math. USSR Izvestiya* **21** (1983), p. 307–340.

[30] J. S. Milne – *Étale cohomology*, Princeton Mathematical Series, vol. 33, Princeton Univ. Press, 1980.

[31] J. Neukirch – *Algebraic number theory*, Grund. Math. Wiss., vol. 322, Springer, 1999.

[32] F. Orgogozo & I. Vidal – Le théorème de spécialisation du groupe fondamental, in *Courbes semi-stables et groupe fondamental en géométrie algébrique (Luminy, 1998)*, Progr. Math., vol. 187, Birkhäuser, 2000, p. 169–184.

[33] A. N. Parshin – Revêtements abéliens de schémas arithmétiques (en russe), *Dokl. Akad. Nauk SSSR* **243** (1978), p. 855–858 ; traduction anglaise : *Math. Dokl.* **19** (1978), p. 1438–1442.

[34] B. Poonen – Bertini theorems over finite fields, *Ann. of Math.* **160** (2004), p. 1099–1127.

[35] W. Raskind – Abelian class field theory of arithmetic schemes, in *K-theory and algebraic geometry : connections with quadratic forms and division algebras (Santa Barbara, CA, 1992)*, Proc. Sympos. Pure Math., vol. 58, Amer. Math. Soc., 1995, p. 85–187.

[36] S. Saito – Unramified class field theory of arithmetical schemes, *Ann. of Math.* **121** (1985), p. 251–281.

[37] A. Schmidt – Tame coverings of arithmetic schemes, *Math. Ann.* **322** (2002), p. 1–18.

[38] ———, Tame class field theory for arithmetic schemes, *Invent. Math.* **160** (2005), p. 527–565.

[39] ———, Singular homology of arithmetic schemes, *Algebra Number Theory* **1** (2007), p. 183–222.

[40] ———, Some consequences of Wiesend's higher dimensional class field theory, *Math. Z.* **256** (2007), p. 731–736.

[41] A. Schmidt & M. Spiess – Singular homology and class field theory of varieties over finite fields, *J. reine angew. Math.* **527** (2000), p. 13–36.

[42] J-P. Serre – *Groupes algébriques et corps de classes*, Publications de l'institut de mathématique de l'université de Nancago, VII. Hermann, Paris, 1959.

[43] ———, Zeta and L functions, in *Arithmetical Algebraic Geometry (Proc. Conf. Purdue Univ., 1963)*, Harper & Row, 1965, p. 82–92.

[44] ———, *Corps locaux*, 2e éd., Hermann, 1968.

[45] ———, Théorie du corps de classes pour les revêtements non ramifiés de variétés algébriques (d'après S. Lang), Séminaire Bourbaki, vol. 1955/56, exposé n° 133, Collection hors série, vol. 3, Soc. Math. France, 1995, p. 347–355.

[46] A. Suslin & V. Voevodsky – Singular homology of abstract algebraic varieties, *Invent. Math.* **123** (1996), p. 61–94.

[47] ———, Bloch-Kato conjecture and motivic cohomology with finite coefficients, in *The arithmetic and geometry of algebraic cycles (Banff, AB, 1998)*, NATO Sci. Ser. C Math. Phys. Sci., vol. 548, Kluwer Acad. Publ., 2000, p. 117–189.

[48] T. Szamuely – Sur la théorie des corps de classes pour les variétés sur les corps p-adiques, *J. reine angew. Math.* **525** (2000), p. 183–212.

[49] ———, *Galois groups and fundamental groups*, Cambridge Studies in Advanced Math., vol. 117, Cambridge Univ. Press, 2009.

[50] B. L. van der Waerden – *Algebra II*, 5e éd., Springer, 1967.

[51] G. Wiesend – A construction of covers of arithmetic schemes, *J. Number Theory* **121** (2006), p. 118–131.

[52] ———, Class field theory for arithmetic schemes, *Math. Z.* **256** (2007), p. 717–729.

[53] ———, Tamely ramified covers of varieties and arithmetic schemes, *Forum Math.* **20** (2008), p. 515–522.

Tamás SZAMUELY

Alfréd Rényi Institute of Mathematics
Hungarian Academy of Sciences
Reáltanoda utca 13–15.
H-1053 BUDAPEST – Hongrie
E-mail : szamuely@renyi.hu

332

ASTÉRISQUE

2010

SÉMINAIRE BOURBAKI
VOLUME 2008/2009
EXPOSÉS 997-1011

(1007) *Un théorème de la limite centrale pour les ensembles convexes*

Franck BARTHE

SOCIÉTÉ MATHÉMATIQUE DE FRANCE
Publié avec le concours du CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

Séminaire BOURBAKI Juin 2009
61e année, 2008-2009, no 1007, p. 287 à 304

UN THÉORÈME DE LA LIMITE CENTRALE POUR LES ENSEMBLES CONVEXES [d'après Klartag et Fleury-Guédon-Paouris]

par **Franck BARTHE**

INTRODUCTION

À l'interface de la théorie locale des espaces de Banach et de la théorie de Brunn-Minkowski-Lusternik, la géométrie asymptotique des convexes étudie les propriétés métriques ou volumiques des corps convexes (i.e. les sous-ensembles convexes, compacts de $\mathbb{R}^n$ dont l'intérieur n'est pas vide) en grande dimension n en cherchant les bonnes dépendances dimensionnelles. Le résultat fondateur de ce domaine est certainement le théorème de Dvoretzky, revisité par Milman **[32, 33]** qui assure que, pour tout corps convexe $K \subset \mathbb{R}^n$, symétrique par rapport à l'origine, il existe un sous-espace vectoriel F de dimension au moins $c(\varepsilon)\log n$ tel que l'intersection $K \cap F$ soit euclidienne à ε près, au sens suivant : il existe un ellipsoïde $\mathcal{E} \subset F$ tel que

$$(1-\varepsilon)\mathcal{E} \subset K \cap F \subset (1+\varepsilon)\mathcal{E}.$$

Par dualité il existe aussi des projections de K de dimension au moins $c(\varepsilon)\log n$ qui sont presque euclidiennes. Ce résultat structurel est un exemple des phénomènes de grande dimension, qui vont souvent à l'encontre de notre intuition spatiale.

Formulé il y a plus de dix ans, le problème de la limite centrale pour les corps convexes demande essentiellement si la mesure uniforme sur un corps convexe de grande dimension possède une marginale presque gaussienne. Il s'agit donc d'établir un analogue du théorème de Dvoretzky où l'on considère les projections des mesures sur des sous-espaces plutôt que les projections des ensembles. Il est à noter que le problème de la limite centrale est non-trivial pour des projections de rang 1, ce qui est bien différent du contexte ensembliste. Après de multiples contributions qui ont apporté des réponses partielles, la conjecture a été résolue indépendamment par Klartag et Fleury-Guédon-Paouris en 2006. Par la suite, Klartag a grandement amélioré la précision des estimations quantitatives.

Le cadre de cet exposé est l'espace $\mathbb{R}^n$ muni de la structure euclidienne induite par le produit scalaire $\langle x, y\rangle = \sum_{i=1}^n x_i y_i$ et la norme associée $|x| = \langle x, x\rangle^{1/2}$. On note $B_2^n = \{x \in \mathbb{R}^n;\ |x| \leq 1\}$ la boule unité et S^{n-1} la sphère unité, que l'on munit de la probabilité uniforme σ_{n-1}. Nous utiliserons aussi les probabilités invariantes μ_n sur le groupe spécial orthogonal $SO(n)$ et $\sigma_{n,k}$ sur la grassmannienne $G_{n,k}$ formée des sous-espaces vectoriels de dimension k de $\mathbb{R}^n$. La projection orthogonale sur un sous-espace vectoriel E sera notée P_E. Dans tout l'exposé, $c, C, c', C', \ldots$ sont des constantes numériques qui peuvent changer d'une ligne à l'autre, mais que nous notons de la même manière pour ne pas alourdir les notations.

1. LE PROBLÈME DE LA LIMITE CENTRALE

1.1. Deux exemples significatifs

Si un vecteur aléatoire $\Theta^{(N)} = (\Theta_1^{(N)}, \ldots, \Theta_N^{(N)})$ est uniformément distribué sur la sphère de rayon $\sqrt{N}$, il est bien connu depuis Maxwell que $\Theta_1^{(N)}$ converge en loi vers une variable gaussienne standard G (de loi $\exp(-t^2/2)\, dt/\sqrt{2\pi}$, $t \in \mathbb{R}$). Par ailleurs le principe d'Archimède nous assure que $(\Theta_1^{(N)}, \ldots, \Theta_{N-2}^{(N)})$ est uniformément distribué sur $\sqrt{N} B_2^{N-2}$. On en déduit aisément que si X est un vecteur uniforme sur $\sqrt{n+2}\, B_2^n$, alors $\mathbb{E}X_i = 0$, $\mathbb{E}(X_i X_j) = \delta_{i,j}$, et que si n est grand la loi de X_1 est presque gaussienne. L'invariance par rotation permet de dire pour tout $\theta \in S^{n-1}$ que la loi de $\langle X, \theta\rangle$ est presque gaussienne.

Considérons maintenant $X^{(n)} = (X_1, \ldots, X_n)$ uniformément distribué sur le cube $[-\sqrt{3}, \sqrt{3}]^n$. Comme les variables X_i sont indépendantes et vérifient $\mathbb{E}X_i = 0$, $\mathbb{E}(X_i^2) = 1$ le théorème de la limite centrale nous assure de la convergence en loi

$$\frac{X_1 + \cdots + X_n}{\sqrt{n}} \underset{n \to +\infty}{\longrightarrow} G.$$

Il est cependant naturel de considérer d'autres directions que $(1/\sqrt{n}, \ldots, 1/\sqrt{n}) \in S^{n-1}$. En notant $F_T(t) = \mathbb{P}(T \leq t)$ la fonction de répartition d'une variable aléatoire réelle T, le théorème de Berry-Esseen nous donne que, pour tout $\theta \in S^{n-1}$ et $t \in \mathbb{R}$,

$$|F_{\langle X^{(n)}, \theta\rangle}(t) - F_G(t)| \leq 6 \frac{\sum_{i=1}^n \mathbb{E}\big(|\theta_i X_i|^3\big)}{\Big(\sum_{i=1}^n \mathbb{E}\big(|\theta_i X_i|^2\big)\Big)^3} \leq 20 \max_{1 \leq i \leq n} |\theta_i|.$$

Il n'est pas difficile de vérifier que cette dernière quantité est petite en moyenne

$$\int_{S^{n-1}} \max_{1 \leq i \leq n} |\theta_i|\, d\sigma_{n-1}(\theta) \approx \sqrt{\frac{\log n}{n}}.$$

En utilisant la concentration de la mesure sphérique (voir plus loin pour un énoncé précis), on en déduit aisément l'existence de constantes numériques c, C telles que

$$\sigma_{n-1}\left(\left\{\theta \in S^{n-1};\ \max_{1\leq i\leq n} |\theta_i| \leq C\sqrt{\frac{\log n}{n}}\right\}\right) \geq 1 - \frac{1}{n^c}.$$

Donc pour la plupart des directions $\theta \in S^{n-1}$, la marginale de la mesure uniforme sur le cube dans la direction de θ est presque gaussienne et les estimées s'améliorent avec la dimension.

1.2. Formulation précise

Le problème de la limite centrale demande si le phénomène observé sur ces deux exemples est universel. Ceci revient à postuler une variante inédite du théorème de la limite centrale où une hypothèse géométrique de convexité remplace la notion d'indépendance. Afin de formuler plus explicitement le problème, il convient de préciser la normalisation.

Définition 1.1. — *On dira qu'un corps convexe $K \subset \mathbb{R}^n$ est (en position) isotrope si* $\mathrm{vol}_n(K) = 1$, $\int_K x\,dx = 0$ *et s'il existe un nombre* $L_K > 0$ *tel que pour tout* $\theta \in S^{n-1}$

$$\int_K \langle x, \theta\rangle^2 dx = L_K^2.$$

Pour tout corps convexe, il existe une transformation affine qui permet de le mettre en position isotrope.

Conjecture 1.2. — *Existe-t-il des suites (ε_n) et (η_n) décroissantes et de limite nulle telles que, pour tout $n \geq 1$ et pour tout corps convexe isotrope $K \subset \mathbb{R}^n$, on ait*

$$\sigma_{n-1}\left(\left\{\theta \in S^{n-1};\ d\Big(\big\langle \frac{X}{L_K}, \theta\big\rangle, G\Big) \geq \varepsilon_n\right\}\right) \leq \eta_n,$$

où X est un vecteur aléatoire uniforme sur K, G est une variable gaussienne standard et d est une distance sur les lois des variables (norme uniforme entre fonctions de répartition par exemple) ?

Cette question, posée dans les articles d'Antilla-Ball-Perissinaki **[2]** (paru en 2003 mais disponible dès 1998) et Brehm-Voigt **[12]**, a suscité beaucoup d'intérêt (voir notamment **[4, 6, 10, 11, 25, 28, 30, 36, 37, 41, 44]**). Elle a été résolue par Klartag **[21]** et Fleury-Guédon-Paouris **[16]** (indépendamment, Klartag légèrement plus tôt) avec des suites $\varepsilon_n \leq (\log n)^{-\kappa}$. Par la suite, Klartag **[22]** a obtenu une forte amélioration avec $\varepsilon_n \leq n^{-\kappa}$, pour un $\kappa < 1$. Par ailleurs ses travaux montrent l'existence de beaucoup de marginales presque gaussiennes de dimension de l'ordre de n^κ et se placent dans le cadre fonctionnel naturel des mesures de probabilités log-concaves.

DÉFINITION 1.3. — *Une mesure de probabilité borélienne μ sur $\mathbb{R}^n$ est log-concave si elle admet une densité $f : \mathbb{R}^n \to \mathbb{R}^+$ vérifiant pour tous $\lambda \in [0,1]$, $x, y \in \mathbb{R}^n$*

$$f\big(\lambda x + (1-\lambda)y\big) \geq f(x)^\lambda f(y)^{1-\lambda}.$$

On dira qu'un vecteur aléatoire Y est centré réduit si son espérance est nulle et sa matrice de covariance est égale à l'identité ($\mathbb{E}Y = 0$, $\mathrm{Cov}(Y) = \mathrm{Id}$). Il est clair que dans la formulation du problème central limite, le vecteur aléatoire $Y = X/L_K$ est log-concave et centré réduit.

1.3. Réduction à l'hypothèse de concentration

Il est apparu dès le début que le problème se réduit à la question suivante, d'apparence bien plus simple : existe-t-il une suite (ε_n) de limite nulle telle que pour tout vecteur aléatoire Y à valeurs dans $\mathbb{R}^n$ centré réduit et uniforme sur un convexe (ou plus généralement log-concave)

$$\mathbb{P}\left(\left|\frac{|Y|}{\sqrt{n}} - 1\right| \geq \varepsilon_n\right) \leq \varepsilon_n ? \tag{1}$$

Notons que les hypothèses de normalisation assurent que $\sqrt{n} = \big(\mathbb{E}(Y^2)\big)^{1/2}$. Dans le cas d'un vecteur uniforme sur un convexe, il s'agit de montrer que la plus grande partie du volume du convexe est contenue dans une couronne dont l'épaisseur est négligeable devant le rayon.

L'énoncé qui justifie cette nouvelle formulation du problème se trouve dans **[2]** pour le cas des mesures uniformes sur les convexes (voir aussi **[4]** pour l'extension au cadre log-concave et des raffinements) :

PROPOSITION 1.4. — *Soit (ε_n) une suite qui vérifie l'hypothèse* (1)*. Alors, pour tous $n \geq 1$, $\delta > 0$ et tout vecteur aléatoire $Y \in \mathbb{R}^n$ centré réduit et log-concave, on a*

$$\sigma_{n-1}\left(\left\{\theta \in S^{n-1};\ \sup_{t\in\mathbb{R}}\Big|\mathbb{P}\big(\langle Y,\theta\rangle \leq t\big) - \mathbb{P}\big(G \leq t\big)\Big| \geq \frac{c}{\sqrt{n}} + \varepsilon_n + \delta\right\}\right) \leq 2e^{-cn\delta^2}.$$

Nous allons expliquer les idées qui permettent de démontrer cette proposition. Elles remontent à Sudakov **[43]**, qui a remarqué que lorsque Y est centré réduit, mais pas forcément log-concave, la plupart des variables $(\langle Y,\theta\rangle)_{\theta\in S^{n-1}}$ ont à peu près la même loi. Le point essentiel dans son argument est l'utilisation du phénomène de concentration de la mesure sphérique, qui est une conséquence de l'inégalité isopérimétrique de Lévy (voir par exemple **[35, 39]**). Nous en rappelons l'énoncé.

THÉORÈME 1.5. — *Soit $f : S^{n-1} \to \mathbb{R}$ une fonction L-lipschitzienne pour la distance induite par $\mathbb{R}^n$, alors pour tout $u > 0$,*

$$\sigma_{n-1}\left(\left\{\theta \in S^{n-1};\ \left|f(\theta) - \int f\, d\sigma_{n-1}\right| \geq u\right\}\right) \leq 2e^{-\frac{nu^2}{2L^2}}.$$

L'hypothèse de log-concavité sur Y n'est pas essentielle, mais elle nous simplifie la tâche puisqu'elle assure que les fonctions $\theta \mapsto \mathbb{P}(\langle Y, \theta\rangle \leq t)$ sont C-lipschitziennes pour une constante universelle C et ce, bien que $\theta \mapsto \mathbf{1}_{\langle Y,\theta\rangle \leq t}$ ne le soit pas. L'inégalité de concentration garantit que, pour t fixé et pour la plupart des valeurs de θ, la fonction $F_\theta(t) = \mathbb{P}(\langle Y, \theta\rangle \leq t)$ est proche de

$$F(t) := \int_{S^{n-1}} F_\theta(t)\, d\sigma_{n-1}(\theta).$$

Pour obtenir un résultat uniforme en t on utilise un argument classique de réseau, qui tire parti de la forte décroissance dans l'inégalité de déviation. En effet, par une borne d'union, on obtient que pour tout entier k et tous réels $t_1, \ldots, t_k$,

$$\begin{aligned}
&\sigma_{n-1}\left(\left\{\theta \in S^{n-1};\ \sup_{i\leq k} |F_\theta(t_i) - F(t_i)| \geq u\right\}\right) \\
&\quad\leq \sum_{i=1}^{k} \sigma_{n-1}\Big(\big\{\theta \in S^{n-1};\ |F_\theta(t_i) - F(t_i)| \geq u\big\}\Big) \leq 2ke^{-\frac{nu^2}{2C^2}}.
\end{aligned}$$

Pour $u > 0$ donné, on peut choisir k le plus petit possible et $t_1, \ldots, t_k$ de telle manière que la condition $\sup_{t\in\mathbb{R}} |F_\theta(t) - F(t)| \geq 2u$ implique $\sup_{1\leq i\leq k} |F_\theta(t_i) - F(t_i)| \geq u$ (on utilise ici le fait que $t \mapsto F_\theta(t) - F(t)$ est lipschitzienne et tend vers 0 à l'infini avec une vitesse explicite). Ceci donne une inégalité de déviation qui montre que, pour la plupart des directions, la fonction de répartition de $\langle Y, \theta\rangle$ est proche de F.

Il reste à identifier la loi dont la fonction de répartition est F. En utilisant le théorème de Fubini et l'invariance par rotation, il vient

$$\begin{aligned}
F(\theta) &= \int_{S^{n-1}} \mathbb{E}\mathbf{1}_{\langle Y,\theta\rangle\leq t}\, d\sigma_{n-1} \\
&= \mathbb{E}\sigma_{n-1}(\{\theta;\ \langle \theta, Y\rangle \leq t\}) = \mathbb{E}\sigma_{n-1}(\{\theta;\ \theta_1|Y| \leq t\}).
\end{aligned}$$

Soit $\Theta^{(n)}$ un vecteur de loi uniforme sur S^{n-1} et indépendant de Y. Le calcul précédent montre que F est la fonction de répartition de

$$\Theta_1^{(n)}|Y| = \sqrt{n}\Theta_1^{(n)} \times \frac{|Y|}{\sqrt{n}}.$$

Par le principe de Maxwell, $\sqrt{n}\Theta_1^{(n)}$ converge en loi vers une gaussienne standard lorsque n tend vers l'infini. Il est alors clair que, pour que la loi « presque » commune des marginales devienne gaussienne quand la dimension augmente, il suffit que $|Y|/\sqrt{n}$ soit très proche d'une constante.

1.4. Vers l'estimation de la concentration de la norme

Dans cette section, nous présentons un bref historique des résultats connus sur la distribution de $|X|$ quand X est uniforme sur un convexe et centré réduit. Nous essayons de mettre en avant ceux qui ont joué un rôle dans la démonstration de la concentration dans une fine couronne.

Le résultat fondamental est certainement l'inégalité de Prékopa-Leindler (voir e.g. **[39]**) :

THÉORÈME 1.6. — *Soit $\lambda \in (0,1)$ et soient f, g, h des fonctions mesurables de $\mathbb{R}^n$ dans $\mathbb{R}^+$ telles que pour tous $x, y \in \mathbb{R}^n$*

$$h(\lambda x + (1-\lambda)y) \geq f(x)^\lambda g(y)^{1-\lambda}.$$

Alors

$$\int_{\mathbb{R}^n} h \geq \left(\int_{\mathbb{R}^n} f\right)^\lambda \left(\int_{\mathbb{R}^n} g\right)^{1-\lambda}.$$

Il en découle des propriétés qui nous seront utiles par la suite : si μ et ν sont des mesures log-concaves sur $\mathbb{R}^n$ alors $\mu * \nu$ l'est aussi. De plus si E est un sous-espace vectoriel de $\mathbb{R}^n$ la marginale de μ dans la direction de E, notée $P_E\mu$ et définie pour tout ensemble borélien de E par $P_E\mu(B) = \mu(P_E^{-1}(B))$, est encore log-concave. Enfin μ vérifie une inégalité de Brunn-Minkowski : si $\lambda \in [0,1]$ et $A, B \in \mathbb{R}^n$, on a

$$\mu(\lambda A + (1-\lambda)B) \geq \mu(A)^\lambda \mu(B)^{1-\lambda}. \tag{2}$$

Si $D \subset \mathbb{R}^n$ est un convexe symétrique par rapport à l'origine, on vérifie facilement pour $t \geq 1$ l'inclusion

$$\frac{2}{t+1}(\mathbb{R}^n \setminus tD) + \frac{t-1}{t+1}D \subset \mathbb{R}^n \setminus D.$$

En appliquant l'inégalité (2) on obtient le résultat suivant, dû à Borell **[9]**, qui met en avant la décroissance sous-exponentielle des mesures log-concaves :

LEMME 1.7. — *Soit μ une mesure de probabilité log-concave sur $\mathbb{R}^n$ et soit $D \subset \mathbb{R}^n$ un convexe symétrique par rapport à l'origine. Si $\mu(D) > 1/2$, alors pour tout $t \geq 1$*

$$\mu\big(\mathbb{R}^n \setminus tD\big) \leq \mu(D)\left(\frac{1-\mu(D)}{\mu(D)}\right)^{\frac{t+1}{2}}.$$

On en déduit aisément une inégalité de grande déviation pour $|X|$. En effet par l'inégalité de Markov

$$\mathbb{P}(|X| > \sqrt{3n}) \leq \frac{\mathbb{E}\big(|X|^2\big)}{3n} = \frac{1}{3}.$$

Puisque la loi de X est log-concave, le lemme de Borell pour $D = \sqrt{3n}B_2^n$ donne pour $t \geq 1$

$$\mathbb{P}(|X| > t\sqrt{3n}) \leq \mathbb{P}(|X| \leq \sqrt{3n}) \left(\frac{\mathbb{P}(|X| > \sqrt{3n})}{\mathbb{P}(|X| \leq \sqrt{3n})} \right)^{\frac{t+1}{2}} \leq 2^{-t/2}.$$

On retiendra que pour $t \geq \sqrt{3}$ on a $\mathbb{P}(|X|/\sqrt{n} \geq t) \leq \exp(-ct)$. Un raisonnement similaire montre que, pour tout $\theta \in S^{n-1}$ et tout $t \geq \sqrt{3}$,

$$\mathbb{P}(|\langle X, \theta \rangle| \geq t) \leq e^{-ct}.$$

Ce dernier résultat est optimal aux constantes près. On peut le voir lorsque X est uniforme sur un multiple de $B_1^n = \{x \in \mathbb{R}^n;\ \sum_i |x_i| \leq 1\}$ et $\theta = (1, 0, \ldots, 0)$; dans ce cas la mesure marginale est presque exponentielle. Cependant X ne peut pas être aussi étalé dans toutes les directions. Dans cet esprit Alesker **[1]** a montré qu'en fait

$$\mathbb{P}\left(\frac{|X|}{\sqrt{n}} \geq t \right) \leq e^{-ct^2}, \qquad t \geq t_0.$$

Une concentration plus forte avait été observée dans le cas concret où X est uniforme sur un multiple de B_1^n et centré réduit **[40]** :

$$\mathbb{P}\left(\frac{|X|}{\sqrt{n}} \geq t \right) \leq e^{-ct\sqrt{n}}, \qquad t \geq t_0 \tag{3}$$

et la borne est optimale aux constantes près. Ce résultat a été démontré pour X uniforme sur un convexe inconditionnel (c'est-à-dire stable par les changements de signe des coordonnées) par Bobkov et Nazarov **[7, 8]** grâce à un principe de comparaison avec B_1^n. Il est à noter que cette inégalité est toujours plus forte que la précédente en $\exp(-t^2)$ car, pour X centré réduit et uniforme sur un convexe de $\mathbb{R}^n$, on a $|X| \leq cn$ presque sûrement (voir **[34]**). Ainsi seules les déviations $t \leq \sqrt{n}$ sont significatives.

Une avancée remarquable a été accomplie par Paouris **[38]** en 2006. Il a réussi à établir l'inégalité (3) pour X uniforme sur un convexe supposé seulement symétrique par rapport à l'origine. Cette percée a permis de débloquer plusieurs problèmes importants en géométrie asymptotique. L'argument de Paouris combine de manière novatrice et sophistiquée des outils pourtant bien connus. Nous en donnons maintenant un bref aperçu.

Les inégalités de concentration exponentielles peuvent s'obtenir en analysant la transformée de Laplace ou la croissance des moments. Paouris adopte ce dernier point de vue. L'inégalité de concentration gaussienne d'Alesker se traduit comme suit sur les moments :

$$\forall p \geq 2, \quad (\mathbb{E}|X|^p)^{\frac{1}{p}} \leq c\sqrt{p} \left(\mathbb{E}|X|^2 \right)^{\frac{1}{2}}.$$

L'amélioration que démontre Paouris revient à l'inégalité inverse-Hölder suivante

$$\forall p \in [2, c\sqrt{n}], \quad (\mathbb{E}|X|^p)^{\frac{1}{p}} \leq c \left(\mathbb{E}|X|^2 \right)^{\frac{1}{2}}.$$

Il est commode ici de revenir aux notations classiques de convexité : écrivons $X = X_K/L_K$ où K est un corps convexe isotrope, L_K sa constante d'isotropie et X_K un vecteur uniformément distribué sur K. Par ailleurs nous aurons besoin des notions suivantes.

DÉFINITION 1.8. — *Soit $K \subset \mathbb{R}^n$ un corps convexe. On définit*

- *sa fonction d'appui $h_K : \mathbb{R}^n \to \mathbb{R}$ telle que pour tout $x \in \mathbb{R}^n$,*

$$h_K(x) = \sup_{y \in K} \langle x, y \rangle,$$

- *sa demi-épaisseur moyenne $W(K) = \displaystyle\int_{S^{n-1}} h_K(\theta)\, d\sigma_{n-1}(\theta)$*
- *et pour $p \geq 1$ le corps p-centroïde $Z_p(K)$ associé à K par la formule*

$$h_{Z_p(K)}(x) = \left(\int_K |\langle x, y \rangle|^p dy \right)^{\frac{1}{p}}.$$

Notons que les corps $Z_p(K)$ ont été introduits par Lutwak et Zhang [**27**] dans le cadre de la théorie de Brunn-Minkowski L_p. Paouris observe que, pour $p \leq \sqrt{n}$,

$$(\mathbb{E}|X_K|^p)^{\frac{1}{p}} \approx \sqrt{\frac{n}{p}} W(Z_p(K)).$$

L'idée essentielle est ensuite d'interpréter cette épaisseur moyenne comme le rayon des projections de $Z_p(K)$ sur des sous-espaces aléatoires. En effet, une application précise du théorème de Dvoretzky montre que si $k \leq c\sqrt{n}$, pour la plupart des sous-espaces vectoriels $E \subset \mathbb{R}^n$ de dimension k (au sens de la probabilité uniforme sur la grassmannienne), on a

$$\frac{1}{2} W(Z_p(K))\, P_E(B_2^n) \subset P_E(Z_p(K)) \subset 2W(Z_p(K))\, P_E(B_2^n).$$

Notons que $P_E(B_2^n) = B_2^n \cap E$ est la boule unité de E ; par ailleurs le fait que le résultat soit vrai jusqu'en dimension $c\sqrt{n}$ utilise déjà des informations sur $Z_p(K)$ (en effet dans le cas général on ne peut aller au-delà de $\log n$). Il reste à majorer le volume de $P_E(Z_p(K))$. Pour ce faire, Paouris note que cet ensemble est de la forme $Z_p(B_K)$ où B_K est un autre corps convexe associé à K, introduit par Ball. Cette structure supplémentaire et le fait que $Z_p(B_K)$ soit presque une boule euclidienne permettent à Paouris d'en estimer le volume et d'obtenir la majoration voulue de $\mathbb{E}|X_K|^p$.

Le théorème de Paouris donne une estimation optimale des grandes déviations de la norme sur un convexe isotrope. Il ne dit rien sur les petites déviations autour de la moyenne. Cependant, les techniques de preuves et surtout l'idée de passer par des sous-espaces aléatoires dans lesquels la situation est plus « ronde » ont inspiré les auteurs qui ont résolu le problème de la limite centrale pour les convexes.

2. ESQUISSE DE PREUVE

Nous commençons par évoquer la solution proposée par Fleury, Guédon et Paouris **[16]**. Elle consiste en une version presque isométrique (à ε près pour ε petit) de l'approche « isomorphe » de Paouris (à une grande constante C près). Elle contient en particulier un résultat quantitatif de stabilité du corps p-centroïde qui assure que si K a le même volume qu'une boule D et si $Z_p(K)$ et $Z_p(D)$ sont très proches, alors Z et D sont eux-mêmes très proches.

THÉORÈME 2.1 (**[16]**). — *Soit X un vecteur centré réduit, uniforme sur un corps convexe de $\mathbb{R}^n$ qui est symétrique par rapport à 0. Alors pour une constante universelle c, on a*

$$\forall p \in [2, (\log n)^{1/3}], \quad (\mathbb{E}|X|^p)^{\frac{1}{p}} \leq \left(1 + \frac{cp}{(\log n)^{1/3}}\right) \left(\mathbb{E}|X|^2\right)^{\frac{1}{2}}.$$

Pour illustrer ce résultat, nous donnons un argument simple (mais grossier) qui permet d'en déduire la propriété de concentration dans une fine couronne. Il passe par l'estimation de la variance de $|X|^2$, dont nous reparlerons plus loin. Notons que l'estimation triviale de la variance de $|X|^2$ donne,

$$\mathrm{var}(|X|^2) \leq n \sum_{i=1}^n \mathrm{var}(|X_i|^2) \leq n \sum_{i=1}^n \mathbb{E}(X_i^4) \leq cn \sum_{i=1}^n \mathbb{E}(X_i^2)^2 = cn^2$$

en utilisant le fait que la norme L_4 de X_i est comparable à la norme L_2, ce qui est une conséquence du caractère sous-exponentiel des coordonnées.

Le théorème précédent donne, pour $p = 4$ et n grand,

$$\mathrm{var}(|X|^2) = \mathbb{E}(|X|^4) - (\mathbb{E}|X|^2)^2 \leq \frac{c}{(\log n)^{1/3}} (\mathbb{E}|X|^2)^2 = \frac{cn^2}{(\log n)^{1/3}}.$$

Cette faible amélioration suffit pour conclure. En effet puisque $||a| - 1| \leq ||a| - 1|(|a| + 1)$, l'inégalité de Markov donne pour $t > 0$

$$\mathbb{P}\left(\left|\frac{|X|}{\sqrt{n}} - 1\right| \geq t\right) \leq \mathbb{P}\left(\left|\frac{|X|^2}{n} - 1\right| \geq t\right) \leq \frac{1}{t^2} \mathrm{var}\left(\frac{|X|^2}{n}\right) \leq \frac{c}{t^2 (\log n)^{1/3}}.$$

On obtient ainsi (1) avec $\varepsilon_n = c(\log n)^{-1/9}$.

Nous allons donner un peu plus de détails sur l'argument du second article de Klartag, qui nous paraît plus direct et qui a l'avantage de donner de meilleures estimations dans un cadre plus général (notons cependant que la méthode de Paouris a le mérite de donner des estimations optimales pour les grandes déviations).

THÉORÈME 2.2 ([**22**]). — *Il existe $c > 0$ et $\alpha \in (0,1)$ tels que, pour tout $n \geq 1$ et tout vecteur aléatoire Y centré, réduit et log-concave à valeurs dans $\mathbb{R}^n$, on ait*

$$\mathbb{P}\left(\left|\frac{|Y|}{\sqrt{n}} - 1\right| \geq \frac{1}{n^\alpha}\right) \leq ce^{-n^\alpha}.$$

De plus, pour tout entier $k \leq cn^\alpha$,

$$\sigma_{n,k}\left(\left\{E \in G_{n,k};\ d_{TV}\Big(P_E(Y), P_E(G)\Big) \geq \frac{1}{n^\alpha}\right\}\right) \leq e^{-cn^{0.99}},$$

où G est un vecteur gaussien centré réduit sur $\mathbb{R}^n$. Ici d_{TV} désigne la distance en variation totale.

Ce théorème constitue donc un analogue pour les mesures du théorème de Dvoretzky. Il est à noter que l'on obtient des marginales presque gaussiennes de dimension puissance n^α, alors que dans le cas général le théorème de Dvoretzky ne donne que des projections presque euclidiennes de dimension $\log n$. La preuve de Klartag est dans le même esprit que la preuve de Milman pour le théorème de Dvoretzky : on montre qu'un sous-espace aléatoire a la propriété recherchée avec une grande probabilité en combinant des arguments de concentration de la mesure et d'approximation par un réseau. Les détails techniques (contrôle des normes Lipschitz, approximations, réglage des constantes) sont souvent très délicats et demandent de manier avec dextérité les nombreux résultats connus sur les mesures log-concaves (inégalités inverses Hölder, estimation des quantiles, comparaison de la densité maximale et de la densité au centre de gravité, ...). Voici les grandes étapes.

Si $f : \mathbb{R}^n \to \mathbb{R}^+$ est la densité de la loi de Y et si E est un sous-espace vectoriel de $\mathbb{R}^n$, la densité de la loi de $P_E(Y)$ par rapport à la mesure de Lebesgue sur E est donnée par

$$\pi_E(f)(x) = \int_{x+E^\perp} f(z)\, dz.$$

C'est encore une fonction log-concave. Pour $U \in SO(n)$ une isométrie directe et $x \in E$, on considère

$$M_{f,x,E}(U) = \log \pi_E(f \circ U)(x) = \log \pi_{U(E)}(f)(Ux).$$

Klartag commence par régulariser la densité f par convolution avec une gaussienne. Cette petite perturbation ne modifiera pas la nature des résultats.

LEMME 2.3 ([**22**]). — *Soient E_0 un sous-espace de dimension k de $\mathbb{R}^n$ et $x \in E_0$. Soit $\beta > 0$. Si G est un vecteur gaussien centré réduit à valeurs dans $\mathbb{R}^n$ et indépendant de Y, on note g la densité de $Y + k^{-\beta/2}G$. Alors l'application*

$$(x, U) \mapsto M_{g,E,x}(U)$$

restreinte à $\{x \in E_0;\ |x| \leq 10\sqrt{k}\} \times SO(n)$ est $Ck^{2\beta+2}$-lipschitzienne par rapport à U et $Ck^{\beta+1/2}$-lipschitzienne en la variable x.

Ici $SO(n)$ est considéré comme une sous-variété riemannienne de $\mathbb{R}^{n^2}$. La distance géodésique est alors comparable à la distance de Hilbert-Schmidt. Il est alors possible d'appliquer l'inégalité de concentration des fonctions lipschitziennes sur $SO(n)$ pour la probabilité bi-invariante μ_n dont l'énoncé est analogue à celui que nous avons donné pour la sphère. Notons que l'invariance par rotation assure que

$$\int_{SO(n)} M_{g,E_0,x_0}(U)\, d\mu_n(U)$$

ne dépend que de $|x_0|$. On note cette quantité $M(|x_0|)$. On obtient ainsi

$$\mu_n\left(\left\{U \in SO(n);\ \left|M_{g,E_0,x_0}(U) - M(|x_0|)\right| > \delta\right\}\right) \leq Ce^{-\frac{c\delta^2 b}{k^{4\beta+4}}}.$$

On considère alors un ε-réseau $\mathcal{N}$ de $10\sqrt{k}B_2^n \cap E_0$ de cardinal au plus $(C'\sqrt{k}/\varepsilon)^k$, ce qui signifie que pour tout $y \in 10\sqrt{k}B_2^n \cap E_0$ il existe $z \in \mathcal{N}$ tel que $|y-z| \leq \varepsilon$. Nous renvoyons par exemple à **[39]** pour l'existence d'un tel réseau. Par une borne d'union

$$\mu_n\left(\left\{U \in SO(n);\ \exists x \in \mathcal{N}, \left|M_{g,E_0,x}(U) - M(|x|)\right| > \delta\right\}\right) \leq (C'\sqrt{k}/\varepsilon)^k Ce^{-\frac{c\delta^2 b}{k^{4\beta+4}}}.$$

Comme $x \mapsto M_{g,E_0,x}(U)$ est aussi lipschitzienne, on en déduit que

$$\mu_n\left(\left\{U \in SO(n);\ \forall x \in 10\sqrt{k}B_2^n \cap E_0, \left|M_{g,E_0,x}(U) - M(|x|)\right| \leq \delta + 2\varepsilon C k^{\beta+1/2}\right\}\right)$$

vaut au moins $1 - (C'\sqrt{k}/\varepsilon)^k Ce^{-\frac{c\delta^2 b}{k^{4\beta+4}}}$. Pour un bon choix des paramètres, qui nécessite $k \leq n^\alpha$, l'événement ci-dessus est de grande mesure. On en déduit donc que les marginales de g dans la direction $E = U(E_0)$ sont presque radiales (jusqu'à un rayon de $10\sqrt{k}$) pour un choix typique de U, donc pour un choix typique de E dans $G_{n,k}$. Rappelons qu'elles sont log-concaves, de par les conséquences de l'inégalité de Prékopa-Leindler.

Ensuite Klartag étudie plus précisément la concentration de la norme pour les mesures presque radiales. Par intégration polaire, le cas des mesures exactement radiales se ramène à une question en dimension 1, que Klartag traite par des méthodes élémentaires mais fines :

LEMME 2.4 (**[22]**). — *Soient $d \geq 2$ et $f : \mathbb{R}^+ \to \mathbb{R}^+$ une fonction log-concave continue, C^2 sur $(0,+\infty)$ et intégrable. Alors, pour tout $\varepsilon \in [0,1]$,*

$$\int_{(1-\varepsilon)t_d(f)}^{(1+\varepsilon)t_d(f)} t^{d-1} f(t)\, dt \geq \left(1 - Ce^{-c\varepsilon^2 d}\right) \int_0^\infty t^{d-1} f(t)\, dt,$$

où $t_d(f)$ est le point où $t \mapsto t^{d-1}f(t)$ atteint son maximum.

Etant donnée une marginale presque radiale (obtenue pour la plupart des sous-espaces k-dimensionnels comme ci-dessus), il est possible de montrer qu'à l'instar des mesures exactement radiales, elle est concentrée sur une fine couronne. Ainsi, cette

mesure est très proche de la mesure uniforme sur une sphère de rayon de l'ordre de $\sqrt{k}$ en dimension k. Une version quantitative de l'observation de Maxwell, due à Diaconis et Friedman **[13]**, montre alors que ses marginales de dimension $\sqrt{k}$ sont presque gaussiennes. Ceci termine la preuve du fait que la plupart des marginales de la mesure initiale sont presque gaussiennes.

On peut remarquer que l'inégalité de concentration dans une fine couronne n'est intervenue que pour les marginales k-dimensionnelles typiques et non pour la mesure elle-même. Comme cette propriété est intéressante en soi, nous présentons de manière informelle l'argument de **[21]** qui permet de l'obtenir pour la mesure initiale sur $\mathbb{R}^n$. Il est basé sur le lemme maintenant classique de Johnson-Lindenstrauss **[18]**, qui peut être déduit de la concentration de la mesure sphérique :

LEMME 2.5. — *Soient $1 \leq \ell \leq n$ et $x \in \mathbb{R}^n$ non nul; alors pour tout $\varepsilon \in [0,1]$*

$$\sigma_{n,\ell}\left(\left\{E \in G_{n,\ell};\ \left|\frac{\sqrt{n}\,|P_E(x)|}{\sqrt{\ell}\,|x|} - 1\right| \geq \varepsilon\right\}\right) \leq Ce^{-c\varepsilon^2\ell}.$$

Si l'on considère un sous-espace vectoriel aléatoire F de loi uniforme sur $G_{n,\ell}$ et indépendant de Y, on obtient

$$\mathbb{P}\left((1-\varepsilon)\sqrt{\frac{\ell}{n}}|Y| \leq |P_F(Y)| \leq (1+\varepsilon)\sqrt{\frac{\ell}{n}}|Y|\right) \geq 1 - Ce^{-c\varepsilon^2\ell}. \tag{4}$$

Supposons que la plupart des marginales sont concentrées dans une fine couronne, ou plus précisément qu'il existe $\mathcal{E} \subset G_{n,\ell}$ avec $\sigma_{n,\ell}(\mathcal{E}) \geq 1-\eta$ tel que, pour tout $E \in \mathcal{E}$,

$$\mathbb{P}\left(\left|\frac{|P_E(Y)|}{\sqrt{\ell}} - 1\right| \geq \varepsilon\right) \leq Ce^{-c\varepsilon^2\ell}.$$

En termes du sous-espace aléatoire, on obtient donc

$$\mathbb{P}\left((1-\varepsilon)\sqrt{\ell} \leq |P_F(Y)| \leq (1+\varepsilon)\sqrt{\ell}\right) \geq 1 - \eta - Ce^{-c\varepsilon^2\ell}. \tag{5}$$

L'intersection des événements qui apparaissent dans (4) et (5) est encore de probabilité très proche de 1 pour η assez petit. Mais dans cette intersection on a à la fois

$$|Y| \approx \sqrt{\frac{n}{\ell}}\,|P_F(Y)| \quad \text{et} \quad |P_F(Y)| \approx \sqrt{\ell}.$$

Il s'ensuit que $|Y| \approx \sqrt{n}$ avec grande probabilité.

3. QUESTIONS CONNEXES

Commençons par évoquer le problème des directions sous-gaussiennes (aussi appelées directions ψ_2), posé par V. Milman. Il demande s'il existe une constante universelle c telle que, pour toute dimension et pour tout vecteur aléatoire X centré réduit

et uniforme sur un corps convexe de $\mathbb{R}^n$, il existe au moins une direction $\theta \in S^{n-1}$ telle que, pour tout $t \geq 1$,

$$\mathbb{P}(\langle X, \theta \rangle \geq t) \leq e^{-ct^2}.$$

Cette question est encore ouverte en général, même si elle a beaucoup progressé avec les travaux de Paouris et Klartag. Nous renvoyons aux articles **[23]** et **[17]** pour les meilleurs résultats disponibles et plus de références bibliographiques.

Passons maintenant à des questions qui prolongent directement l'étude de la concentration dans une fine couronne. Nous avons vu que, pour Y vecteur aléatoire à valeurs dans $\mathbb{R}^n$ centré réduit et log-concave, toute amélioration de l'estimation triviale $\mathrm{var}(|Y|^2) \leq cn^2$ permettait de résoudre le problème des marginales presque gaussiennes. Dans le cas où les coordonnées de Y sont indépendantes

$$\mathrm{var}(|Y|^2) = \sum_{i=1}^{n} \mathrm{var}(Y_i^2) \approx cn.$$

Il est naturel de se demander si l'on a toujours $\mathrm{var}(|Y|^2) \leq cn$ pour une constante universelle c (le théorème de Klartag donne en général majoration par une puissance de n un peu inférieure à 2). Cette conjecture a été confirmée par Antilla-Ball-Perissinaki **[2]** lorsque Y est uniforme sur un multiple de $B_p^n = \{x \in \mathbb{R}^n;\ \sum_i |x_i|^p \leq 1\}$, grâce à une propriété de sous-indépendance des coordonnées qui assure que, pour $i \neq j$, $\mathrm{cov}(Y_i^2, Y_j^2) \leq 0$ et donne la même majoration que dans le cas indépendant. Wojtaszczyk **[44]** a étendu la sous-indépendance, et donc la borne précise de la variance, aux boules d'Orlicz généralisées

$$\Big\{x \in \mathbb{R}^n;\ \sum_{i=1}^{n} f_i(|x_i|) \leq 1\Big\},$$

où les fonctions $f_i : \mathbb{R}^+ \to \mathbb{R}^+$ sont convexes et croissantes. La propriété de sous-indépendance peut être fausse pour certains corps convexes inconditionnels. Cependant Klartag **[24]** a récemment établi la borne $\mathrm{var}(|X|^2) \leq cn$ lorsque X est uniforme sur un tel convexe et centré réduit. Sa preuve, plus classique, repose sur des méthodes L_2 à la Hörmander. Il obtient une inégalité du type $\mathrm{var}(f(X)) \leq \mathscr{E}(f, X)$ pour toute fonction inconditionnelle suffisamment régulière, où $\mathscr{E}(f, X)$ est un terme d'énergie inhabituel. La borne de variance en cn permet, par une amélioration du raisonnement que nous avons présenté plus haut, de montrer que la plupart des marginales de dimension 1 sont à distance $1/\sqrt{n}$ d'une gaussienne. Pour les convexes inconditionnels, l'existence de symétries permet de trouver des directions presque gaussiennes explicites :

THÉORÈME 3.1 ([**24**]). — *Soient X un vecteur aléatoire centré réduit, uniformément distribué sur un convexe inconditionnel de $\mathbb{R}^n$ et G une variable gaussienne centrée*

réduite. Alors

$$\sup_{\alpha<\beta}\left|\mathbb{P}\left(\frac{\sum_{i=1}^n X_i}{\sqrt{n}}\in[\alpha,\beta]\right)-\mathbb{P}\left(G\in[\alpha,\beta]\right)\right|\leq\frac{C}{\sqrt{n}}.$$

En utilisant la méthode de Stein, Meckes et Meckes **[28, 29]** ont étudié des convexes possédant d'autres types de symétrie.

La conjecture de Kannan-Lovász-Simonovits **[19]** postule l'existence d'une constante universelle C telle que, pour tout $n\geq 1$ et tout vecteur aléatoire à valeurs dans $\mathbb{R}^n$, centré réduit et uniforme sur un convexe (ou plus généralement log-concave), l'inégalité de Poincaré suivante est vérifiée : pour toute fonction $f:\mathbb{R}^n\to\mathbb{R}$ localement lipschitzienne

$$\operatorname{var}(f(X))\leq C\,\mathbb{E}\Big(|\nabla f(X)|^2\Big). \tag{6}$$

Ceci revient à dire que les fonctions linéaires sont extrémales, à constante près, dans cette inégalité. Il s'agit d'une conjecture très forte, qui admet des formulations équivalentes en termes d'isopérimétrie (voir **[31]** pour des progrès récents et une présentation des travaux antérieurs). Appliquée à $f(x)=|x|^2$, elle prédit la borne $\operatorname{var}(|X|^2)\leq 4Cn$. Elle a été vérifiée pour des vecteurs uniformes sur des multiples B_p^n **[26, 42]** et des simplexes réguliers **[3]**, ainsi que pour les mesure log-concaves invariantes par rotation **[4]**. L'inégalité de Poincaré (6) a été démontrée avec des constantes C dépendant de la dimension en n^κ pour un $\kappa<1$ (en combinant les meilleures estimations pour le théorème de la limite centrale pour les convexes et une inégalité isopérimétrique de Bobkov qui assure que l'on peut prendre $C=\operatorname{var}(|X|^2)^{1/2}$ **[5]**). Klartag **[24]** a obtenu une borne en $(\log n)^2$ pour les mesures uniformes sur un convexe inconditionnel.

Comme l'a remarqué Fleury **[14]**, la conjecture KLS prédit des inégalités de concentration pour $|X|$ qui unifieraient les inégalités de grandes déviations de Paouris et amélioreraient les inégalités de petites déviations de Klartag : pour tout $t\geq 0$,

$$\mathbb{P}\left(\left|\frac{|X|}{\sqrt{n}}-1\right|\geq t\right)\leq 2e^{-ct\sqrt{n}}.$$

Fleury a réussi à démontrer cette inégalité pour X uniforme sur une boule d'Orlicz généralisée. Il reste cependant beaucoup à faire dans cette direction.

Pour finir, nous mentionnons la conjecture de l'hyperplan qui demande si les constantes d'isotropie des corps convexes sont uniformément bornées

$$\sup_{n\geq 1}\ \sup_{K\subset\mathbb{R}^n\ \text{convexe}} L_K<+\infty\,?$$

Voir **[20]** pour la meilleure borne sur L_K dépendant de la dimension n et pour plus de références. Ce problème récurrent en géométrie asymptotique équivaut à demander

s'il existe une constante universelle $c > 0$ telle que pour tout convexe isotrope $K \subset \mathbb{R}^n$ et tout $\theta \in S^{n-1}$

$$\mathrm{vol}_{n-1}(K \cap \theta^{\perp}) \geq c.$$

Ceci revient donc à une minoration de la valeur en 0 des marginales de dimension 1 de la mesure uniforme sur un convexe. L'engouement pour le problème central limite pour les convexes s'explique en partie par le fait qu'il s'agit d'une question liée (sur l'aspect gaussien des marginales) mais indépendante de la conjecture de l'hyperplan, puisque L_K n'y apparaît que comme un facteur de normalisation qu'il n'est pas besoin de borner. Notons pour finir que Ball a montré (sans le publier) que la conjecture KLS entraîne la conjecture de l'hyperplan, ce qui a conduit certains spécialistes à penser que cette conjecture est trop forte pour être vraie...

Remarque de dernière minute : le résultat suivant a été annoncé par Fleury **[15]**. Pour tout vecteur aléatoire centré réduit et log-concave Y à valeurs dans $\mathbb{R}^n$ et pour tout $t \in [0, n^{\kappa}]$, on a

$$\mathbb{P}\left(\left|\frac{|Y|}{\sqrt{n}} - 1\right| \geq \frac{t}{n^{\kappa}}\right) \leq Ce^{-ct},$$

où $\kappa = 1/8$. Ceci améliore la valeur de l'exposant κ obtenue par Klartag (légèrement inférieure à $0,1$). Rappelons que la conjecture KLS prévoit que l'énoncé est valable pour $\kappa = 1/2$.

RÉFÉRENCES

[1] S. Alesker – ψ_2-estimate for the Euclidean norm on a convex body in isotropic position, in *Geometric aspects of functional analysis (Israel, 1992–1994)*, Oper. Theory Adv. Appl., vol. 77, Birkhäuser, 1995, p. 1–4.

[2] M. Anttila, K. Ball & I. Perissinaki – The central limit problem for convex bodies, *Trans. Amer. Math. Soc.* **355** (2003), p. 4723–4735.

[3] F. Barthe & P. Wolff – Remarks on non-interacting conservative spin systems : the case of gamma distributions, *Stochastic Process. Appl.* **119** (2009), p. 2711–2723.

[4] S. G. Bobkov – On concentration of distributions of random weighted sums, *Ann. Probab.* **31** (2003), p. 195–215.

[5] ———, On isoperimetric constants for log-concave probability distributions, in *Geometric aspects of functional analysis*, Lecture Notes in Math., vol. 1910, Springer, 2007, p. 81–88.

[6] S. G. Bobkov & A. Koldobsky – On the central limit property of convex bodies, in *Geometric aspects of functional analysis*, Lecture Notes in Math., vol. 1807, Springer, 2003, p. 44–52.

[7] S. G. Bobkov & F. L. Nazarov – Large deviations of typical linear functionals on a convex body with unconditional basis, in *Stochastic inequalities and applications*, Progr. Probab., vol. 56, Birkhäuser, 2003, p. 3–13.

[8] ———, On convex bodies and log-concave probability measures with unconditional basis, in *Geometric aspects of functional analysis*, Lecture Notes in Math., vol. 1807, Springer, 2003, p. 53–69.

[9] C. Borell – Convex measures on locally convex spaces, *Ark. Mat.* **12** (1974), p. 239–252.

[10] U. Brehm, P. Hinow, H. Vogt & J. Voigt – Moment inequalities and central limit properties of isotropic convex bodies, *Math. Z.* **240** (2002), p. 37–51.

[11] U. Brehm, H. Vogt & J. Voigt – Permanence of moment estimates for p-products of convex bodies, *Studia Math.* **150** (2002), p. 243–260.

[12] U. Brehm & J. Voigt – Asymptotics of cross sections for convex bodies, *Beiträge Algebra Geom.* **41** (2000), p. 437–454.

[13] P. Diaconis & D. Freedman – A dozen de Finetti-style results in search of a theory, *Ann. Inst. H. Poincaré Probab. Statist.* **23** (1987), p. 397–423.

[14] B. Fleury – Between Paouris concentration inequality and variance conjecture, à paraître dans *Ann. Inst. H. Poincaré Probab. Statist*, 2009.

[15] ———, Concentration in a thin Euclidean shell for log-concave measures, prépublication, 2009.

[16] B. Fleury, O. Guédon & G. Paouris – A stability result for mean width of L_p-centroid bodies, *Adv. Math.* **214** (2007), p. 865–877.

[17] A. Giannopoulos, A. Pajor & G. Paouris – A note on subgaussian estimates for linear functionals on convex bodies, *Proc. Amer. Math. Soc.* **135** (2007), p. 2599–2606.

[18] W. B. Johnson & J. Lindenstrauss – Extensions of Lipschitz mappings into a Hilbert space, in *Conference in modern analysis and probability (New Haven, Conn., 1982)*, Contemp. Math., vol. 26, Amer. Math. Soc., 1984, p. 189–206.

[19] R. Kannan, L. Lovász & M. Simonovits – Isoperimetric problems for convex bodies and a localization lemma, *Discrete Comput. Geom.* **13** (1995), p. 541–559.

[20] B. Klartag – On convex perturbations with a bounded isotropic constant, *Geom. Funct. Anal.* **16** (2006), p. 1274–1290.

[21] ———, A central limit theorem for convex sets, *Invent. Math.* **168** (2007), p. 91–131.

[22] ———, Power-law estimates for the central limit theorem for convex sets, *J. Funct. Anal.* **245** (2007), p. 284–310.

[23] ———, Uniform almost sub-Gaussian estimates for linear functionals on convex sets, *Algebra i Analiz* **19** (2007), p. 109–148.

[24] ———, A Berry-Esseen type inequality for convex bodies with an unconditional basis, *Probab. Theory Related Fields* **45** (2009), p. 1–33.

[25] A. Koldobsky & M. Lifshits – Average volume of sections of star bodies, in *Geometric aspects of functional analysis*, Lecture Notes in Math., vol. 1745, Springer, 2000, p. 119–146.

[26] R. Latała & J. O. Wojtaszczyk – On the infimum convolution inequality, *Studia Math.* **189** (2008), p. 147–187.

[27] E. Lutwak & G. Zhang – Blaschke-Santaló inequalities, *J. Differential Geom.* **47** (1997), p. 1–16.

[28] E. S. Meckes & M. W. Meckes – The central limit problem for random vectors with symmetries, *J. Theoret. Probab.* **20** (2007), p. 697–720.

[29] M. W. Meckes – Gaussian marginals of convex bodies with symmetries, *Beiträge Algebra Geom.* **50** (2009), p. 101–118.

[30] E. Milman – On Gaussian marginals of uniformly convex bodies, *J. Theoret. Probab.* **22** (2009), p. 256–278.

[31] ———, On the role of convexity in isoperimetry, spectral gap and concentration, *Invent. Math.* **177** (2009), p. 1–43.

[32] V. D. Milman – A new proof of A. Dvoretzky's theorem on cross-sections of convex bodies, *Funkcional. Anal. i Priložen.* **5** (1971), p. 28–37.

[33] ———, Dvoretzky's theorem—thirty years later, *Geom. Funct. Anal.* **2** (1992), p. 455–479.

[34] V. D. Milman & A. Pajor – Isotropic position and inertia ellipsoids and zonoids of the unit ball of a normed n-dimensional space, in *Geometric aspects of functional analysis (1987–88)*, Lecture Notes in Math., vol. 1376, Springer, 1989, p. 64–104.

[35] V. D. Milman & G. Schechtman – *Asymptotic theory of finite-dimensional normed spaces*, Lecture Notes in Math., vol. 1200, Springer, 1986.

[36] A. Naor & D. Romik – Projecting the surface measure of the sphere of ℓ_p^n, *Ann. Inst. H. Poincaré Probab. Statist.* **39** (2003), p. 241–261.

[37] G. Paouris – Concentration of mass and central limit properties of isotropic convex bodies, *Proc. Amer. Math. Soc.* **133** (2005), p. 565–575.

[38] ———, Concentration of mass on convex bodies, *Geom. Funct. Anal.* **16** (2006), p. 1021–1049.

[39] G. Pisier – *The volume of convex bodies and Banach space geometry*, Cambridge Tracts in Mathematics, vol. 94, Cambridge Univ. Press, 1989.

[40] G. Schechtman & J. Zinn – On the volume of the intersection of two L_p^n balls, *Proc. Amer. Math. Soc.* **110** (1990), p. 217–224.

[41] S. Sodin – Tail-sensitive Gaussian asymptotics for marginals of concentrated measures in high dimension, in *Geometric aspects of functional analysis*, Lecture Notes in Math., vol. 1910, Springer, 2007, p. 271–295.

[42] ———, An isoperimetric inequality on the l_p balls, *Ann. Inst. H. Poincaré Probab. Stat.* **44** (2008), p. 362–373.

[43] V. N. Sudakov – Typical distributions of linear functionals in finite-dimensional spaces of high dimension, *Dokl. Akad. Nauk SSSR* **243** (1978), p. 1402–1405.

[44] J. O. Wojtaszczyk – The square negative correlation property for generalized Orlicz balls, in *Geometric aspects of functional analysis*, Lecture Notes in Math., vol. 1910, Springer, 2007, p. 305–313.

Franck BARTHE

Équipe de Statistique et Probabilités
Institut de Mathématiques de Toulouse (IMT)
CNRS UMR 5219
Université Paul-Sabatier
F–31062 Toulouse cedex 9
E-mail : barthe@math.univ-toulouse.fr

332

ASTÉRISQUE

2010

SÉMINAIRE BOURBAKI
VOLUME 2008/2009
EXPOSÉS 997-1011

(1008) *Équidistribution des orbites toriques sur les espaces homogènes*

Emmanuel BREUILLARD

SOCIÉTÉ MATHÉMATIQUE DE FRANCE
Publié avec le concours du CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

Séminaire BOURBAKI Juin 2009
61e année, 2008-2009, no 1008, p. 305 à 339

ÉQUIDISTRIBUTION DES ORBITES TORIQUES SUR LES ESPACES HOMOGÈNES
[d'après M. Einsiedler, E. Lindenstrauss, Ph. Michel, A. Venkatesh]

par **Emmanuel BREUILLARD**

1. INTRODUCTION

Soit $G = SL(d,\mathbb{R})$, $\Gamma = SL(d,\mathbb{Z})$, A le sous-groupe des matrices diagonales et $\Omega = \Gamma\backslash G$, l'espace des réseaux unimodulaires de $\mathbb{R}^d$. Soit H un sous-groupe fermé de G et dg une mesure de Haar sur G ou sur Ω. Depuis que Raghunathan a mis en évidence à la fin des années 1970 tout le profit qu'il y aurait à tirer de l'étude des H-orbites $x \cdot H$ dans Ω, $x \in \Omega$, pour diverses questions classiques de théorie des nombres, beaucoup de chemin a été parcouru. Margulis (1986) a démontré la conjecture d'Oppenheim en suivant cette stratégie et Ratner (1990) a répondu aux conjectures de Raghunathan en donnant une description précise des propriétés topologiques et statistiques du système dynamique (Ω, H, dg) *lorsque H est un sous-groupe fermé engendré par des unipotents de G*. Pour de tels sous-groupes H (par exemple si H est semi-simple sans facteurs compacts) l'adhérence des orbites est « algébrique » : cela veut dire que cette adhérence est elle-même l'orbite d'un sous-groupe fermé L contenant H et possédant une mesure L-invariante finie. Un des aspects surprenants de la preuve de Ratner est qu'elle déduit ce théorème topologique d'un théorème métrique : elle classifie d'abord les mesures ergodiques H-invariantes sur Ω. Forts de ce résultat, de ses généralisations naturelles dans le contexte S-arithmétique, et des nouvelles techniques dites de linéarisation développées par Dani, Margulis et Shah, de nombreux auteurs (voir par exemple [**29**], [**28**], [**26**], [**32**]) ont fait intervenir ces théorèmes dans diverses situations comme le comptage des points entiers sur une variété homogène, l'équidistribution des translatées de H-orbites, et même récemment la conjecture d'André-Oort ([**72**],[**38**]).

Dans cet exposé, nous nous intéressons au cas opposé où le sous-groupe H n'est pas engendré par des unipotents. Plus précisément nous supposons que $H = A$ est le tore diagonal. Là aussi de nombreuses études ont été poursuivies, mais certaines questions

importantes demeurent. Le cas $d = 2$ est dans une large mesure bien compris : il correspond au flot géodésique sur une surface hyperbolique, ici la surface modulaire. On sait alors décrire les A-orbites via le codage markovien (voir par exemple **[67]**). Dans ce cas, la plupart des A-orbites ont un comportement chaotique. Par exemple on construit facilement des A-orbites dont l'adhérence a une dimension de Hausdorff qui peut prendre une infinité de valeurs entre $1 = \dim A$ et $3 = \dim G$. De même ces compacts invariants peuvent être les supports de mesures de probabilité A-invariantes et ces mesures peuvent avoir de l'entropie pour l'action de A. Lorsque $d \geq 3$ la situation est très différente et un phénomène de rigidité lié au rang supérieur (i.e. l'existence d'au moins deux actions linéairement indépendantes qui commutent) apparaît : on s'attend au contraire – voir plus bas §5 et l'article de Margulis **[46]** pour une formulation précise de ces conjectures – à ce que les adhérences de A-orbites soient beaucoup mieux contrôlées et que les mesures de probabilité A-invariantes et ergodiques soient elles aussi « algébriques », c'est-à-dire des mesures de Haar sur des sous-espaces homogènes de Ω. Après le travail initial de Katok et Spatzier **[37]** deux méthodes (dites de haute entropie et basse entropie) permettant de préciser la nature algébrique des mesures ergodiques A-invariantes ont vu le jour pour culminer avec le théorème de Einsiedler-Katok-Lindenstrauss **[20]** affirmant

THÉORÈME 1.1 (Einsiedler-Katok-Lindenstrauss **[20]**). — *Toute mesure de probabilité A-invariante ergodique sur Ω ayant de l'entropie positive pour au moins un sous-groupe à un paramètre de A est algébrique.*

Comme pour le théorème de Ratner, la classification des mesures A-invariantes sur Ω, et donc le théorème d'Einsiedler-Katok-Lindenstrauss, possède potentiellement des applications arithmétiques. Par exemple, Margulis a observé qu'une réponse positive à sa conjecture (sans condition supplémentaire d'entropie positive) impliquerait la fameuse conjecture de Littlewood en approximation diophantienne (voir plus bas §4). Malheureusement, vérifier la condition d'entropie positive demande souvent une hypothèse supplémentaire : par exemple on montre dans **[20]** que l'ensemble des exceptions possibles à Littlewood est de dimension de Hausdorff zéro en raisonnant par l'absurde et en construisant une mesure d'entropie positive. Il en est de même pour l'ensemble des contre-exemples éventuels (A-orbites irrégulières) à la conjecture de Margulis.

Dans **[22]** Einsiedler, Lindenstrauss, Michel et Venkatesh s'intéressent à la distribution des A-orbites compactes dans Ω. Ils posent entre autres les questions suivantes : est-ce que seul un nombre fini de A-orbites compactes peuvent rester confinées dans un compact donné de Ω ? Une suite d'orbites compactes dont le volume tend vers l'infini devient-elle toujours équidistribuée pour la mesure de Haar dans Ω ?

Alors qu'on s'attend (d'après ces mêmes conjectures de Margulis) à une réponse positive à la première question, les auteurs de [**22**] obtiennent la réponse approchée suivante (on définit au §2 la notion de discriminant d'une orbite compacte) :

THÉORÈME 1.2 ([**22**]). — *On fixe $d \geq 3$. Soit C un compact de Ω. Le nombre de A-orbites compactes contenues dans C et de discriminant au plus D est $\ll_\varepsilon D^\varepsilon$.*

C'est donc très peu d'orbites car le nombre total d'orbites de discriminant au plus D est polynomial en D (cf. proposition 2.3). La réponse à la seconde question est non (voir plus bas §3). Cependant, les auteurs de [**22**] conjecturent :

CONJECTURE 1.3. — *Soit $\rho > 0$ et $(x_n \cdot A)_n$ une suite de A-orbites compactes dans Ω de discriminant D_n telle que $\mathrm{vol}(x_n \cdot A) \geq (D_n)^\rho$ pour tout entier n. Alors si μ_n est la mesure de probabilité A-invariante sur $x_n \cdot A$, la suite μ_n converge vers la mesure de Haar normalisée sur Ω.*

Dans les deux cas il s'agit de vérifier que l'hypothèse faite (sur le nombre d'orbites de discriminant au plus D dans le cas du théorème 1.2 et sur le volume de l'orbite dans le cas de la conjecture 1.3) se traduit par une condition d'entropie positive pour toute limite faible des μ_n. Cette idée repose sur ce que ses auteurs appellent le « principe de Linnik » en hommage à Y. Linnik qui, dans son livre [**44**], a été un des premiers à étudier le problème de la distribution des A-orbites compactes d'un point de vue ergodique.

Soit Y_n une A-orbite ou un ensemble fini de A-orbites de discriminant au plus D_n. On note $\mathrm{vol}(Y_n)$ le volume total de Y_n et μ_n la mesure de Haar normalisée sur Y_n (i.e. la mesure de probabilité assignant un poids relatif à chaque orbite égal à son volume A-invariant). Si μ est une mesure A-invariante sur Ω et $a \in A$, on note $h_\mu(a)$ l'entropie de μ pour la translation à droite par a.

PROPOSITION 1.4 ([**22**] « Principe de Linnik »). — *Supposons qu'il existe un $\rho > 0$ tel que $\mathrm{vol}(Y_n) \geq (D_n)^\rho$ et soit μ_∞ une valeur d'adhérence de la suite μ_n qui est de masse totale 1 sur Ω. Alors $h_{\mu_\infty}(a) \geq \rho \cdot h(a)$ pour tout $a \in A$, où $h(a)$ est une fonction positive sur A et strictement positive sur les éléments réguliers de A.*

Moyennant ce principe on obtient facilement la

Preuve du théorème 1.2. — Soit μ_∞ une valeur d'adhérence de μ_D la mesure A-invariante normalisée sur Y_D la réunion des A-orbites compactes de discriminant au plus D incluses dans C. Comme C est compact, il n'y a pas de fuite de masse et $\mu_\infty(\Omega) = 1$. Le volume d'une A-orbite compacte est clairement minoré par une constante strictement positive car la plus grande valeur propre d'un élément semi-simple $\mathbb{R}$-déployé de Γ est elle-même minorée. Ainsi si le cardinal de Y_D est au moins D^ε, c'est que le volume total de Y_D est aussi au moins D^ε. Par le principe de

Linnik, μ_∞ a de l'entropie positive, donc possède une composante ergodique d'entropie positive. Celle-ci est algébrique d'après le théorème 1.1. Mais son support est compact : c'est donc la probabilité A-invariante sur une A-orbite compacte. Mais une telle mesure est d'entropie nulle (translation sur un tore). D'où la contradiction. □

L'application du principe à la seconde question, i.e. à la conjecture 1.3 permet de montrer que toute limite μ_∞ possède de l'entropie. Le problème est que, bien que clairement A-invariante, μ_∞ n'est pas nécessairement ergodique. Ce type de problème est classique dans les applications arithmétiques du théorème de Ratner et les techniques de linéarisation de Dani-Margulis et Shah ([**13**] [**68**]) ont été développées en partie pour y remédier. Il peut aussi y avoir fuite de masse (voir §3). Dans cette situation il faut donc se contenter de :

COROLLAIRE 1.5 ([**22**]). — *On fixe $d = 3$ et on se place dans la situation de la conjecture 1.3. Pour toute mesure limite μ_∞ des μ_n telle que $\mu_\infty(\Omega) > 0$ on note $\nu_\infty = \frac{\mu_\infty}{\mu_\infty(\Omega)}$ la mesure renormalisée. Alors $\nu_\infty \geq \rho \cdot c \cdot dg$ où dg est la mesure de Haar normalisée sur Ω et $c > 0$ est une constante.*

Démonstration. — C'est la conséquence de la combinaison du principe de Linnik et du théorème 1.1 d'Einsiedler-Katok-Lindenstrauss. Plus précisément, fixons $a \in A$ un élément régulier (i.e. à valeurs propres de modules distincts). Soit $\mu_\infty = \int_X \nu_\xi dm(\xi)$ la décomposition ergodique de μ_∞ pour l'action de a. Écrivons $\mu_\infty = t\nu_1 + (1-t)\nu_2$ où ν_1 (resp. ν_2) est la partie de la décomposition de μ_∞ dont les composantes ergodiques ν_ξ sont d'entropie strictement positive (resp. nulle) pour a. D'après le théorème 1.1 chaque ν_ξ est algébrique, i.e. L-invariante pour un certain sous-groupe fermé L contenant A. Notons que $A \subsetneq L$ car la mesure de Haar sur une A-orbite compacte est d'entropie nulle pour tout $a \in A$. Mais comme $d = 3$, la seule possibilité est que $L = G$ (voir la discussion après le théorème 4.3 plus bas) et donc $\nu_\xi = dg$ et $h_{\nu_1}(a) = h_{dg}(a)$. On sait par ailleurs calculer $h_{dg}(a)$, c'est la somme des $\log^+$ des valeurs propres de a, il est donc strictement positif. Mais $h_{\mu_\infty}(a) = \int h_{\nu_\xi}(a) dm(\xi) = t h_{\nu_1}(a)$ et $h_{\mu_\infty} \geq c(a) \cdot \rho$ d'après le principe de Linnik 1.4. D'où $t \geq \frac{c(a)}{h_{dg}(a)} \cdot \rho$. □

Lorsque d est premier impair la conclusion subsiste avec la même preuve. Pour d quelconque en revanche des mesures algébriques intermédiaires peuvent en principe apparaître et la conclusion qu'on tire par cette méthode est donc plus faible. De toute façon cet argument n'interdit pas une fuite de masse éventuelle. Néanmoins lorsque l'on considère $\Omega = \Gamma' \backslash G$ où Γ' est un réseau arithmétique co-compact associé à une algèbre à division sur $\mathbb{Q}$, alors le principe de Linnik reste valide et la co-compacité fait qu'il n'y a pas de problème de fuite de masse : les auteurs de [**22**] en déduisent alors semblablement que la réunion d'une *suite stricte* de A-orbites compactes (i.e. telle que seul un nombre fini d'entre elles soient contenues dans une L-orbite périodique

donnée associée à un sous-groupe $L \subsetneq G$) est dense dans Ω pourvu qu'il existe un $\rho > 0$ tel que la réunion Y_D des orbites de la suite dont le discriminant est D vérifie $\mathrm{vol}(Y_D) \geq D^\rho$.

En sus de ce corollaire c'est le résultat principal de Einsiedler, Lindenstrauss, Michel et Venkatesh dans [**24**] qui vient apporter le plus de créance à la conjecture 1.3 : ce résultat confirme la conjecture pour $d = 3$ dans le cas où Y_n est une réunion de « paquets » d'orbites :

THÉORÈME 1.6 (équidistribution des paquets [**24**]). — *On fixe $d = 3$. Soit Y_n une suite infinie de « paquets » de A-orbites. Alors μ_n converge vers la mesure de Haar sur Ω. En particulier, si Y_n est la réunion des toutes les A-orbites compactes de même volume $V = V_n \uparrow +\infty$, alors μ_n converge vers Haar.*

Ce théorème s'inscrit dans le cadre d'une conjecture plus générale due à Clozel et Ullmo concernant l'équidistribution des translatés adéliques de mesures algébriques (voir §6 plus bas). La seconde assertion est aussi l'analogue du résultat fameux de Bowen [**6**] et Margulis [**47**] sur l'équidistribution des géodésiques fermées de longueur au plus L dans une variété compacte à courbure strictement négative et des travaux de Zelditch [**77**] sur la surface modulaire. Nous définissons les paquets d'orbites au paragraphe 6. Pour l'instant contentons-nous d'indiquer que ce sont des sous-ensembles finis d'orbites formant une partition de l'ensemble des A-orbites compactes. Dans un même paquet toutes les orbites ont le même volume (ce qui explique pourquoi la deuxième partie de l'énoncé découle de la première) et le même discriminant. Pour obtenir la convergence vers Haar et non plus seulement un résultat partiel comme le corollaire précédent qui donnait que toute limite n'est pas étrangère à Haar, il faut bien sûr montrer que toutes les composantes ergodiques d'une limite sont d'entropie positive. Les auteurs de [**24**] établissent cette propriété en montrant la convergence des μ_n sur une partie seulement des fonctions sur Ω à savoir sur les séries de Siegel-Eisenstein. Nous verrons au paragraphe 7 comment l'analyse harmonique permet de déduire cette convergence de la sous-convexité de certaines fonctions L. C'est là le second ingrédient essentiel de la preuve du théorème 1.6.

Signalons que l'on peut envisager d'autres groupements naturels (et moins fins) de A-orbites, comme le groupement par discriminant fixé, par ordre fixé, ou bien encore par l'action des opérateurs de Hecke. Cette dernière option est étudiée en détail par Benoist et Oh dans [**1**] où est établie l'équidistribution des orbites sous Hecke d'une A_0-orbite compacte quelconque pour un sous-tore A_0 de A non nécessairement maximal.

Quand $d = 2$, bien qu'il n'y ait plus rigidité des mesures d'entropie positive, les auteurs de [**24**] conjecturent que l'énoncé de la conjecture 1.3 reste vrai. En fait l'analogue pour $d = 2$ du théorème 1.6 lui aussi est vrai : c'est le théorème de Duke

[**14**] sur l'équidistribution des points de Heegner et des géodésiques fermées de la surface modulaire. La preuve de Duke reposait sur la traduction classique du problème via la série théta associée : l'équidistribution revient alors à montrer une borne non triviale sur les coefficients de Fourier de cette forme modulaire (qui est de poids demi-entier), ce que Duke établit en généralisant une idée due à Iwaniec. Cependant, grâce à la formule de Waldspurger [**75**] (voir aussi [**57**]) qui relie l'intégrale d'une forme automorphe sur un paquet aux valeurs spéciales de certaines fonctions L (voir §7) il est aussi possible de déduire le théorème de Duke d'estimations de sous-convexité relatives à ces fonctions L – voir en particulier l'article de Clozel et Ullmo [**10**] où ces techniques sont mises en œuvre. Ces estimations, reposant sur les travaux initiaux de Burgess [**7**], puis de Duke-Friedlander-Iwaniec [**15**], permettent de montrer (pour $d = 2$) l'équidistribution des μ_n évaluées contre n'importe quelle forme automorphe, ce qui suffit bien sûr à montrer la convergence de μ_n vers Haar.

Pour $d = 3$ cette méthode est difficile à mettre en œuvre, surtout pour la partie cuspidale du spectre car on ne dispose pas de formule liant l'intégrale sur un paquet aux fonctions L associées. Cependant pour les séries de Siegel-Eisenstein, on dispose d'une telle formule, c'est la formule de Hecke qui exprime l'intégrale sur un paquet comme un produit de la fonction de Dedekind ζ_K du corps cubique relatif à l'orbite et d'un terme local associé à l'ordre correspondant. La borne de sous-convexité nécessaire pour ζ_K est connue par les travaux de Duke-Friedlander-Iwaniec [**16**] et Blomer-Harcos-Michel [**4**]. Une partie importante de l'article [**24**] consiste à montrer le contrôle du terme local. Ceci entraîne finalement l'équidistribution sur les séries de Siegel-Eisenstein. Mais il se trouve que cette information est suffisante pour établir l'entropie positive de toute mesure limite. On peut alors conclure en utilisant le théorème 1.1. L'hypothèse d'entropie positive nécessaire à la preuve ergodique est donc fournie par la sous-convexité. Comme Einsiedler, Lindenstrauss, Michel et Venkatesh l'ont observé ce principe était déjà présent en filigrane dans les travaux de Linnik ([**44**]) qui, moyennant l'hypothèse de Riemann (ou bien une « condition de Linnik » imposant une condition de congruence sur le discriminant du paquet), obtenait le cas $d = 2$ par une méthode purement ergodique.

Signalons que tout récemment Michel et Venkatesh [**51**] ont grandement généralisé ces bornes de sous-convexité en obtenant un énoncé général et uniforme valable pour toute forme automorphe sur GL_2. Aussi, dans un troisième volet toujours en préparation (voir [**24**] Theorem 4.6), Einsiedler, Lindenstrauss, Michel et Venkatesh obtiennent une vaste généralisation du théorème de Duke au cadre adélique sur GL_2, i.e. l'équidistribution des paquets dans n'importe quelle algèbre de quaternions sur un corps de nombres et sans restriction sur l'espace des paramètres (tore, places). Enfin ils obtiennent aussi une preuve purement ergodique du théorème de Duke pour les géodésiques en montrant une version forte du principe de Linnik énoncé ci-dessus

consistant à établir que l'entropie de toute mesure limite est maximale, donc Haar. Cette preuve est le produit de la relecture moderne, via l'entropie, que font Einsiedler, Lindenstrauss, Michel et Venkatesh du livre de Linnik [**44**] dans lequel un tel argument dynamique est mis en œuvre pour établir l'équidistribution sous une condition de congruence (condition de Linnik).

Dans la suite de l'exposé nous décrivons plus précisément les travaux de Einsiedler, Lindenstrauss, Michel et Venkatesh et donnons les ingrédients de preuve des théorèmes ci-dessus. Ce texte n'est pas un survol de fond du sujet, au plus c'est une invitation à la lecture des articles [**22**] et [**24**] où le lecteur pourra trouver tous les détails des preuves ainsi que plusieurs extensions des énoncés donnés dans cette introduction.

2. ORBITES TORIQUES COMPACTES : DISCRIMINANT ET VOLUME

Lorsque $d = 2$, les A-orbites compactes de Ω peuvent être décrites géométriquement : ce sont les géodésiques orientées fermées de la surface modulaire. Tout relevé $\widetilde{\xi}$ d'une telle géodésique au revêtement universel, i.e. au demi-plan de Poincaré, est un arc de cercle orienté d'extrémités ξ_- et ξ_+ situées sur la droite réelle. Il se trouve que ξ_- et ξ_+ sont les racines d'un trinôme $ax^2 + bx + c$ à coefficients entiers. Réciproquement tout tel trinôme avec $D > 0$ donne lieu à une géodésique fermée. Si on choisit $pgcd(a, b, c) = 1$, alors $D = b^2 - 4ac$ est un entier qui est indépendant du choix du relevé. C'est le discriminant de l'orbite.

En dimension quelconque, les A-orbites compactes de Ω sont mieux décrites via l'arithmétique. Soit K un corps de nombres totalement réel de degré d sur $\mathbb{Q}$ et L un réseau de K, c'est-à-dire un $\mathbb{Z}$-module libre de rang d inclus dans K qui engendre K comme $\mathbb{Q}$-espace vectoriel. Soit $\sigma : K \to \mathbb{R}^d$, $x \mapsto (\sigma_1(x), \ldots, \sigma_d(x))$ un plongement canonique de K via les d différents $\mathbb{Q}$-isomorphismes $\sigma_1, \ldots, \sigma_d : K \hookrightarrow \mathbb{R}$. L'image de L par σ est un réseau de $\mathbb{R}^d$ qu'on peut renormaliser pour le rendre unimodulaire (noter que $\Omega = SL(d, \mathbb{Z})\backslash SL(d, \mathbb{R}) = PGL(d, \mathbb{Z})\backslash PGL(d, \mathbb{R})$). Sa A-orbite est compacte dans Ω. En effet $a = \mathrm{diag}(a_1, \ldots, a_d) \in A$ est dans le stabilisateur de $\sigma(L)$ si et seulement si $a_i = \sigma_i(x)$ pour un certain $x \in \mathcal{O}_L^\times$ le groupe des éléments inversibles de l'ordre $\mathcal{O}_L = \{k \in K, kL \subseteq L\}$. Clairement $\mathcal{O}_L$ (resp. $\mathcal{O}_L^\times$) est d'indice fini l'ordre maximal $\mathcal{O}_K$ (resp. $\mathcal{O}_K^\times$) et d'après le théorème des unités de Dirichlet, $\{\mathrm{diag}(\sigma_1(x), \ldots, \sigma_d(x)), x \in \mathcal{O}_K^\times\}$ est discret et co-compact dans A. Notons que deux choix L et L' donnent lieu à la même orbite si et seulement si $L' = kL$ pour un certain $k \in K^\times$. Comme on le vérifie aisément, cette construction produit en fait toutes les A-orbites compactes :

PROPOSITION 2.1. — *L'ensemble des A-orbites compactes de Ω est naturellement paramétré par les triplets $(K, [L], \sigma)$, où K est un corps de nombres totalement réel, $[L]$ une classe d'homothétie de réseaux de K et σ le choix (parmi $d!$) d'un plongement canonique.*

Le corps K et la classe $[L]$ sont uniquement déterminés par la A-orbite ; il se peut en revanche que deux σ distincts donnent la même orbite. Le corps K est tout simplement la sous $\mathbb{Q}$-algèbre engendrée par le stabilisateur dans A d'un point de l'orbite, lequel est un sous-groupe d'indice fini du groupe des unités de K.

On peut préciser cette correspondance en étudiant le tore associé. La A-orbite de $\sigma(L)$ s'écrit $\Gamma g A$ où g est la matrice dont les colonnes sont les vecteurs $(\sigma_1(v_i), \dots, \sigma_d(v_i))$ où $v_1, \dots, v_d$ est une $\mathbb{Z}$-base de L. Le tore $T = gAg^{-1}$ est un tore maximal $\mathbb{R}$-déployé de $SL(d, \mathbb{R})$ qui est défini sur $\mathbb{Q}$ car il est clairement stable sous $Gal(\overline{\mathbb{Q}}|\mathbb{Q})$. C'est un tore $\mathbb{Q}$-anisotrope car $T \cap \Gamma$ est co-compact dans T. En tant que $\mathbb{Q}$-tore, il est isomorphe aux éléments de norme 1 dans la restriction des scalaires $Res^K_{\mathbb{Q}} \mathbb{G}_m$ du groupe multiplicatif $\mathbb{G}_m$ de K à $\mathbb{Q}$. L'application $x \to g \cdot \mathrm{diag}(\sigma_1(x), \dots, \sigma_d(x)) \cdot g^{-1}$ permet aussi d'identifier K à une sous-algèbre $\mathfrak{t}_{\mathbb{Q}}$ de $M_d(\mathbb{Q})$. Sous cette identification, les éléments de norme 1 dans K s'identifient à $T(\mathbb{Q})$. Réciproquement tout tore maximal $\mathbb{R}$-déployé et $\mathbb{Q}$-anisotrope donne lieu à une orbite compacte, ou plutôt $d!$ orbites compactes. Notons en effet que deux orbites $\Gamma g_1 A$ et $\Gamma g_2 A$ produisent le même tore si et seulement si $g_2 \in N_G(A)$ le normalisateur de A dans G. Il y a donc au plus $d!$ orbites distinctes correspondant à un même tore : en projection dans l'espace localement symétrique $X = \Gamma \backslash G / K$ ces $d!$ orbites s'envoient sur un même tore immergé et le choix d'une chambre de Weyl les distingue (modulo une éventuelle auto-intersection) dans $\Gamma \backslash G$.

Prendre un conjugué sur $\mathbb{Q}$ de T revient à faire varier le réseau L associé à l'orbite sur $\mathbb{Q}$. Enfin, par Skolem-Noether, deux tores sont associés au même corps de nombres K (i.e. sont isomorphes sur $\mathbb{Q}$) si et seulement si ils sont conjugués dans $GL_n(\mathbb{Q})$.

On retiendra surtout de cette classification qu'à toute A-orbite compacte on peut associer deux données essentielles : *son volume V et son discriminant D*. Le volume est par définition le volume A-invariant naturel pour un choix de mesure de Haar sur A fixé une fois pour toutes. Dans l'identification ci-dessus, le volume est le régulateur $reg(\mathcal{O}_L)$ de l'ordre $\mathcal{O}_L$, c'est-à-dire le co-volume de $\sigma(\mathcal{O}_L^\times)$ après le plongement habituel dans l'hyperplan de $\mathbb{R}^d$ formé des points de coordonnées à somme nulle via l'application $\log|\cdot|$. Le discriminant de l'orbite est par définition le discriminant $\mathrm{disc}(\mathcal{O}_L)$ de l'ordre associé. C'est une quantité arithmétique associée à l'orbite, en particulier c'est un entier. Dans l'identification ci-dessus le discriminant correspond, à un facteur multiplicatif fixe près, au co-volume de $\mathfrak{t}_{\mathbb{Q}} \cap M_d(\mathbb{Z})$ dans $\mathfrak{t}_{\mathbb{R}} = \mathfrak{t}_{\mathbb{Q}} \otimes_{\mathbb{Q}} \mathbb{R}$

pour la mesure de Lebesgue sur $\mathfrak{t}_{\mathbb{R}}$ associée à la structure euclidienne standard sur $M_d(\mathbb{R})$.

Discriminant et volume sont liés par la relation fondamentale suivante :

PROPOSITION 2.2 ([**22**]). — *On fixe $d \geq 2$. Il existe une constante $C = C(d) > 0$ telle que pour toute union Y de A-orbites compactes de même discriminant D*

$$\log D \ll \mathrm{vol}(Y) \ll D^C$$

où $\ll$ signifie comme d'habitude une inégalité à des constantes multiplicatives près (dépendant de d seulement).

Comme le volume d'une A-orbite est uniformément minoré (rayon d'injectivité minoré), on en déduit que le nombre de A-orbites compactes de volume (ou discriminant) borné est fini (un fait établi précédemment dans [**54**]). Plus précisément on montre :

PROPOSITION 2.3 ([**22**]). — 1) *Le nombre $N(D)$ de A-orbites compactes de discriminant au plus D vérifie $D^\alpha \ll N(D) \ll D^\beta$ pour des constantes $\alpha, \beta > 0$.*

2) *Si d est premier, le nombre $N(V)$ de A-orbites compactes de volume au plus V vérifie $V^{1/(d-1)} \ll \log N(V) \ll V^{1/(d-1)}$.*

3) *Le volume d'une A-orbite compacte de discriminant D est $\ll_\varepsilon D^{\frac{1}{2}+\varepsilon}$ et le volume total des A-orbites compactes de même déterminant D est $\gg_\varepsilon D^{\frac{1}{2}-\varepsilon}$.*

Disons quelques mots sur ces estimées. La borne supérieure dans 1) résulte de la proposition 2.2 et de la remarque qui la suit. Pour la borne inférieure, il suffit d'exhiber suffisamment d'orbites compactes de petit discriminant ; l'exemple de Cassels donné à la fin de la section 3 est à cet égard suffisant car il exhibe n orbites distinctes de discriminant au plus $n^{d(d-1)}$. Pour la borne supérieure de 2) remarquons que, si d est premier, il n'y a pas de corps intermédiaire entre $\mathbb{Q}$ et K et on peut alors améliorer la borne inférieure de la proposition 2.2 comme suit $\mathrm{vol}(Y) \gg (\log D)^{d-1}$. Cette dernière minoration du volume résulte par le théorème de Minkowski sur les corps convexes de ce que $\log ||\sigma(x)|| \geq O(\log D)$ pour toute unité x du corps K associé à Y. Et cette dernière borne est la conséquence du fait que le $\mathbb{Z}$-module engendré par $1, x, \ldots, x^{d-1}$ est d'indice fini dans $\mathcal{O}_K$ car $\mathbb{Q}(x) = K$. La borne inférieure du 2) résulte aussi de l'exemple de Cassels. Le 3) en revanche est une conséquence de l'estimation du volume d'un paquet d'orbite (voir §6 ci-dessous) par le théorème de Brauer-Siegel.

Le point le plus délicat est certainement la borne supérieure dans la proposition 2.2. Intuitivement elle résulte du fait que deux orbites distinctes de discriminant au plus D ne peuvent s'approcher trop près l'une de l'autre. Plus précisément on a le :

LEMME 2.4 (Isolation des orbites compactes [**22**]). — *Il existe une constante $c = c(d) > 0$ telle que si g_1 et g_2 sont deux éléments de G vérifiant $g_1 A \neq g_2 A$, $||g_i|| \leq R$, $i = 1, 2$, et si les A-orbites $\Gamma\backslash\Gamma g_i A$ sont toutes deux compactes et de discriminant au plus D, alors $d(g_1, g_2) \geq (RD)^{-c}$.*

Ici d est une métrique riemannienne invariante à gauche sur G. Comme une A-orbite compacte de discriminant au plus D ne peut s'aventurer trop loin vers l'infini (cf. proposition 3.1) on conclut de l'énoncé précédent que les A-orbites compactes de discriminant au plus D sont $D^{-O_d(1)}$-isolées les unes des autres, i.e. les $D^{-O_d(1)}$-voisinages tubulaires autour d'elles sont disjoints, ce qui fournit immédiatement la borne supérieure dans la proposition 2.2 par la finitude du volume de $\Gamma\backslash G$. Le lemme 2.4 est aussi le point clé pour montrer le principe de Linnik (cf. proposition 1.4 ci-dessus et 5.2 ci-dessous).

3. FUITE DE MASSE

Pour une A-orbite compacte, le fait d'aller plus ou moins loin à l'infini se traduit aisément en termes de l'ordre associé. Appelons Ω_ε l'ensemble des réseaux Λ de Ω dont la systole $\delta(\Lambda) = \inf\{||v||, v \in \Lambda\backslash\{0\}\}$ est $\leq \varepsilon^{\frac{1}{d}}$. Rappelons que d'après le critère de Mahler (voir [**5**]) les Ω_ε, $\varepsilon > 0$, forment une base de voisinages de l'infini dans Ω.

PROPOSITION 3.1. — *On considère la A-orbite compacte associée à $(K, [L], \sigma)$ et D son discriminant (voir §2).*

a) *Elle rencontre Ω_ε si et seulement si $\exists x \in L$ tel que $N_{K|\mathbb{Q}}(x) \leq \varepsilon \cdot covol_{\mathbb{R}^d}(\sigma(L))$.*

b) *Si $\varepsilon < D^{-\frac{1}{2}}$ alors elle ne rencontre pas Ω_ε.*

Démonstration. — En effet dire que $N_{K|\mathbb{Q}}(x) \leq \varepsilon \cdot covol_{\mathbb{R}^d}(\sigma(L))$ revient à dire que le réseau renormalisé de covolume 1 associé possède un vecteur non-nul dont le produit des coordonnées est au plus ε, si bien que quitte à appliquer un élément de A on a bien un vecteur de norme au plus $\varepsilon^{\frac{1}{d}}$. Pour le b) noter que $x\mathcal{O}_L \subseteq L$ pour tout $x \in L$ et donc que $N_{K|\mathbb{Q}}(x) D^{\frac{1}{2}} \geq covol_{\mathbb{R}^d}(\sigma(L))$. □

Le théorème suivant montre qu'il est nécessaire de grouper les A-orbites compactes si on souhaite avoir de l'équidistribution. Une orbite individuelle peut très bien avoir un comportement pathologique du point de vue de l'équidistribution : il peut y avoir fuite de masse à l'infini dans l'espace des réseaux.

THÉORÈME 3.2 (Fuite de masse individuelle). — *On fixe $d \geq 2$. Il existe une suite de A-orbites compactes Y_n telles que pour tout compact C de Ω on ait $\mu_{Y_n}(C) \to 0$ pour n assez grand.*

Lorsque $d = 2$ ce phénomène est facile à mettre en évidence grâce au codage markovien : il existe des géodésiques fermées dans la surface modulaire qui passent une proportion négligeable de leur vie dans toute partie compacte fixée. En effet, le codage symbolique du flot géodésique permet aisément de construire de telles géodésiques : il suffit de considérer une géodésique du demi-plan de Poincaré dont le codage dans la triangulation de Farey est périodique de la forme $\ldots DG^nDG^nDG^nD\ldots$, c'est-à-dire que cette géodésique laisse la pointe n fois sur sa gauche (G) avant de la laisser une fois sur sa droite (D) puis de recommencer ad libitum. Les points limites de cette géodésique sont les nombres quadratiques ξ_- et ξ_+ dont le développement en fractions continues est donné par $\xi_+ = [n, 1, n, 1, \ldots]$ et $-1/\xi_- = [1, n, 1, n, \ldots]$. Quand n devient très grand, on constate aisément qu'une telle géodésique fermée passe une fraction $1 - \varepsilon$ de son temps à distance au moins $\varepsilon \log n$ d'une origine fixe dans la surface. En particulier toute la masse s'échappe à l'infini : la mesure de probabilité uniforme sur la géodésique converge vers la masse de Dirac en l'infini. Pour une explication de ce codage via la triangulation de Farey et du lien avec les fractions continues, on renvoie le lecteur à l'article **[67]**.

On remarquera que, dans cet exemple de géodésique non équirépartie ci-dessus en dimension 2, le discriminant de l'orbite compacte est égal à $D = n(n+4)$, tandis que la longueur de l'orbite est $2\log\lambda_D$ où $\lambda_D = \frac{n+2+\sqrt{D}}{2}$, c'est-à-dire que le volume (ici la longueur) de l'orbite est comparable au logarithme du discriminant. Cela veut dire que le volume de l'orbite est petit : il est proche de la borne inférieure de l'encadrement (2.2). Le contenu de la conjecture 1.3 est que ce phénomène est le seul obstacle à l'équidistribution. Selon ce principe, les orbites mal équiréparties auraient un petit volume par rapport au discriminant et plus le volume de l'orbite est gros, plus l'orbite aurait de chance d'être bien équirépartie.

3.1. Exemple de Cassels

Montrons le théorème 3.2. Il est en fait dû à Cassels[(1)] qui construit dans **[8]** une suite de corps totalement réels dont le régulateur est petit par rapport au discriminant. Décrivons son exemple. On fixe $a_1 < a_2 < \cdots < a_d$ des entiers positifs non nuls. Pour tout entier $n > 0$, on définit $P_n \in \mathbb{Z}[t]$ par $P_n(t) - 1 = \Pi_{1\leq i\leq d}(t - na_i)$. On vérifie aisément que P_n est irréductible sur $\mathbb{Q}$ et que le corps $K_n = \mathbb{Q}(\theta)$ engendré par une racine est totalement réel. Aussi on voit que, lorsque n est grand, les conjugués de Galois de θ sont très proches des na_i. Il en résulte que son discriminant est de l'ordre de $n^{d(d-1)}$. Mais l'équation $\Pi_{1\leq i\leq d}(\theta - na_i) = 1$ montre que chaque $\theta - na_i$ est une unité de K_n. On vérifie aisément qu'elle engendre un sous-groupe $\mathcal{U}$ de rang $d-1$ dans le groupe des unités de K_n. Ainsi le régulateur de K_n est un $O((\log n)^{d-1})$.

(1) Je remercie Uri Shapira et Elon Lindenstrauss pour m'avoir signalé cet exemple et l'article **[8]**.

On considère maintenant l'orbite compacte associée au réseau $L = \mathbb{Z}[\theta]$ de K_n. On a $\mathcal{U} \subseteq \mathcal{O}_L^\times$ et $\mathrm{disc}(L) \simeq n^{d(d-1)}$. D'après le critère de Mahler (voir [**5**]) montrer qu'il y a fuite de masse ou plus précisément que toute la masse part à l'infini revient à vérifier que pour un ensemble de $a \in A$ mod $\mathcal{U}$ de mesure presque totale le réseau aL possède un petit vecteur, i.e. de norme $o(\mathrm{disc}(L)^{1/2})$. C'est bien le cas car si $a = \mathrm{diag}(\Pi_{1\leq i\leq d}(\sigma_j(\theta) - na_i)^{r_i})_j$ avec $r_i \in \mathbb{R}$ avec $\sum r_i = 1$ et si v est le vecteur de L correspondant à $1 \in K_n$ alors la norme de av est un $o(n^{(d-1)/2})$ dès que les r_i vérifient le système d'inégalités $\left|\sum_{j\neq i} r_j - (d-1)r_i\right| < \frac{d-1}{2}$. Mais cette région de l'hyperplan $\sum r_i = 1$ est de volume 1 modulo $\mathbb{Z}^d$, ce qui termine la preuve. Dans [**22**] les auteurs utilisent une construction similaire due à Duke.

Une question intéressante – pour $d \geq 3$ – est de déterminer si la fuite de masse peut aussi s'opérer vers d'autres orbites compactes, au lieu de l'infini, ou bien les deux à la fois. En particulier, une valeur d'adhérence de μ_{Y_n} peut-elle être non ergodique ?

4. DYNAMIQUE DES FLOTS DIAGONAUX DANS Ω : ADHÉRENCES D'ORBITES ET CLASSIFICATION TOPOLOGIQUE

Peut-être la plus frappante des conjectures concernant les orbites du flot diagonal est la conjecture de Margulis ([**45**]) :

Conjecture 4.1 (Margulis). — *On fixe $d \geq 3$. Tout compact A-invariant de Ω est une réunion de A-orbites compactes.*

Notons que cet énoncé équivaut à demander que toute A-orbite bornée soit compacte. Remarquons d'autre part qu'il ne conclut pas qu'il s'agit d'une réunion *finie* d'orbites, néanmoins nous verrons plus bas (théorème d'isolation) que ce n'est qu'une apparence : la conjecture 4.1 entraîne bien que tout compact A-invariant est une réunion finie de A-orbites compactes.

Dans [**45**], Margulis explique aussi comment une réponse positive à cette conjecture entraînerait une réponse positive à la célèbre conjecture de Littlewood :

Conjecture 4.2 (Littlewood). — *Pour tous réels a et b on a*

$$\lim_{n\to+\infty} \inf n \cdot ||na|| \cdot ||nb|| = 0,$$

où $||x||$ désigne la distance à l'entier le plus proche.

Expliquons brièvement quel est le raisonnement de Margulis (voir [**45**]) pour passer de la conjecture (4.1) à la conjecture (4.2). En fait, comme Margulis l'a découvert a posteriori, ce raisonnement était déjà présent, bien que formulé différemment, dans un article des années 50 de Cassels et Swinnerton-Dyer [**9**]. Bien que ses auteurs

n'aient pas employé ce langage dynamique, cet article est l'un des tout premiers à aborder la question des orbites du sous-groupe diagonal dans l'espace des réseaux Ω. On pose $F(x,y,z) = x(ax+y)(bx+z)$. C'est une forme cubique ; $F = F_0 \circ u_{a,b}$ où $F_0(x,y,z) = xyz$ et $(x,y,z)\cdot u_{a,b} = (x, y+ax, z+bx)$ est une matrice unipotente de $SL_3(\mathbb{R})$. Le stabilisateur $\{g \in SL_3(\mathbb{R}), F_0 \circ g = F_0\}$ est le tore diagonal A. Ainsi $F(\mathbb{Z}^3) = F_0(\Delta \cdot A)$ où Δ est le réseau $\mathbb{Z}^3 \cdot u_{a,b}$.

Notons qu'on a toujours $F(0) = 0$ et que par conséquent (critère de Mahler) la A-orbite $\Delta \cdot A$ est non bornée. Toutefois, si le couple (a,b) est une exception à la conjecture (4.2), alors $F(\mathbb{Z}^3)$ ne contient pas 0 comme point d'accumulation et donc $\Delta \cdot A$ n'est pas dense dans Ω, mieux l'orbite *du cône positif* $A^+ = \{\mathrm{diag}(e^{-(t+s)}, e^t, e^s), t, s > 0\}$, i.e. $\Delta \cdot A^+$ est bornée. Ceci permet de construire une vraie A-orbite bornée : il suffit en effet de considérer un point d'accumulation disons Δ_∞ de $\Delta \cdot A^+$ pour avoir que $\Delta_\infty \cdot A$ est bornée. Si l'on croit à la conjecture 4.1, alors cette orbite doit être compacte. On va en déduire que l'orbite de départ, à savoir $\Delta \cdot A$ est aussi compacte ; ce sera une contradiction car par la remarque faite ci-dessus (F représente 0), cette orbite n'est pas compacte. Cette déduction est aujourd'hui appelée *théorème d'isolation* et peut se formuler comme suit :

THÉORÈME 4.3 (Théorème d'isolation). — *On fixe $d = 3$. Soit $\Delta \cdot A$ une A-orbite dans Ω. Supposons que l'adhérence $\overline{\Delta \cdot A}$ contienne une A-orbite compacte. Alors soit $\overline{\Delta \cdot A} = \Omega$ soit $\overline{\Delta \cdot A} = \Delta \cdot A$.*

La preuve moderne de ce théorème consiste à montrer que si $\Delta \cdot A$ possède un point d'accumulation situé sur une A-orbite compacte, alors $\overline{\Delta \cdot A}$ est invariant par un sous-groupe strictement plus gros que A qui contient notamment un sous-groupe à un paramètre unipotent, et de conclure par le théorème de Ratner que $\overline{\Delta \cdot A} = \Omega$. Ce type d'argument de dérive qui permet d'obtenir une invariance supplémentaire dans une direction transverse au flot est omniprésent tant dans la preuve originale de la conjecture d'Oppenheim par Margulis que dans l'œuvre de Ratner[(2)].

Expliquons succinctement cette mécanique : si $\Delta \cdot a_n \to \Delta_0$ avec $\Delta_0 \cdot A$ compacte et distincte de $A \cdot \Delta$, alors on peut écrire pour chaque n grand $\Delta \cdot a_n = \Delta_0 \cdot g$ avec $g \notin A$ très proche de 1. On décompose g sous la forme $g^{-1} = au^+u^-$ avec $a \in A$ et u^+ et u^- des matrices unipotentes (respectivement triangulaires supérieure et inférieure) très proches de 1 avec disons u^- petit mais non trivial. Le groupe diagonal A normalise les groupes U^+ et U^- et les contracte (resp. dilate) : on peut ainsi trouver un grand élément $b \in A$ qui contracte u^+ (tout en dilatant u^-) de sorte que $u_b^+ := bu^+b^{-1}$ soit extrêmement proche de 1 tandis que $u_b^- := bu^-b^{-1}$ lui atteint

(2) Signalons à ce propos que Einsiedler a récemment donné dans [**17**] une preuve simplifiée du théorème de Ratner dans le cas particulier de l'action d'un sous-groupe semi-simple : la propriété de réductibilité complète d'un groupe semi-simple simplifie grandement l'argument de dérive !

une taille macroscopique. Finalement puisque $\Delta_0 \cdot bu_b^- u_b^+ \in \Delta \cdot A$ et que $\Delta_0 \cdot A$ est compact, on obtient à la limite (quand $n \uparrow +\infty$) un élément $u_\infty^- \in U^- \backslash \{1\}$ tel que $\Delta_0 \cdot u_\infty^- \in \overline{\Delta \cdot A}$ puis, en faisant agir A à nouveau, on constate que $\overline{\Delta \cdot A}$ contient en fait toute l'orbite du sous-groupe à un paramètre U qui contient u_∞^-. On peut donc appliquer Ratner !

Par Ratner $\overline{\Delta \cdot A}$ contient maintenant une orbite fermée $\Delta_0 \cdot H$ de H-volume invariant fini où H est un sous-groupe fermé normalisé par A. On conclut que $\Delta_0 \cdot AH$ est aussi fermée de AH-covolume fini. À ce stade on n'a plus que deux possibilités pour AH : soit c'est $G = SL_3(\mathbb{R})$ tout entier, soit c'est un sous-groupe localement isomorphe à $GL_2(\mathbb{R})$. Pour se débarrasser de ce dernier cas, il faut utiliser une propriété particulière de $SL_3(\mathbb{Z})$ qui n'est pas partagée par tous les réseaux de $SL_3(\mathbb{R})$: les éléments semi-simples d'ordre infini de $SL_3(\mathbb{Z})$ sont réguliers, or un réseau dans $GL_2(\mathbb{R})$ intersecte le centre non trivialement, donc contient un élément non régulier, i.e. avec deux valeurs propres identiques. On conclut que $AH = SL_3(\mathbb{R})$ et que $\overline{\Delta \cdot A} = \Omega$.

Cet argument, du moins la première partie, semble marcher aussi sur $SL_2(\mathbb{R})$: en fait on a triché au moment de passer de $\Delta_0 \cdot u_\infty^-$ à toute une orbite unipotente contenue dans $\overline{\Delta \cdot A}$. Pour cela il nous fallait faire agir à nouveau par un élément $a \in A$ *qui fixe* Δ_0 de sorte que $\{a^k u_\infty^- a^{-k}, k \in \mathbb{Z}\}$ soit dense dans U et ceci ne peut être réalisé sur $SL_2(\mathbb{R})$ car cette orbite est toujours discrète dans ce cas. Pour les détails de cette preuve ainsi qu'une généralisation de ce théorème d'isolation en dimension supérieure, on renvoie à l'article de Lindenstrauss et Weiss [**43**] : en général il est nécessaire de prendre en compte d'autres sous-groupes intermédiaires entre A et G (l'énoncé précédent reste vrai cependant si la dimension d est un nombre premier). Notons finalement que, dans l'argument ci-dessus, on peut remplacer $\overline{\Delta \cdot A}$ par n'importe quel fermé A-invariant F s'accumulant sur une A-orbite compacte : la conclusion est que soit $F = \Omega$ soit F est une réunion finie d'orbites compactes. En particulier on voit qu'étant donnée une A-orbite compacte $\Delta_0 \cdot A$ toute autre A-orbite compacte qui s'approche très près de $\Delta_0 \cdot A$ doit partir très loin à l'infini dans Ω : cela justifie le terme d'*isolation* pour désigner ce théorème. La conjecture 4.1 entraîne donc

CONJECTURE 4.4. — *On fixe $d \geq 3$. Etant donné un compact K de Ω, il n'existe qu'un nombre fini de A-orbites compactes contenues dans K.*

Cette propriété particulière de $SL_3(\mathbb{Z})$ de ne pas posséder d'éléments non réguliers est claire : les valeurs propres d'un élément de $SL_3(\mathbb{Z})$ sont des unités algébriques dont le produit vaut 1 et si deux d'entre elles sont égales alors le polynôme caractéristique n'est pas irréductible sur $\mathbb{Q}$ et possède ainsi une racine dans $\mathbb{Q}$ qui est donc ± 1, ce qui force ± 1 et $\pm i$ pour les autres racines et on a donc affaire à un élément d'ordre fini.

Comme Mary Rees l'a mis en évidence pour la première fois dans ce contexte dynamique, on peut néanmoins construire des exemples de réseaux Γ de $G = SL_3(\mathbb{R})$ pour lesquels cette propriété n'est pas vérifiée, c'est-à-dire qu'il existe un $\gamma_0 \in \Gamma$ non régulier semi-simple d'ordre infini. L'existence de ces exemples est problématique de plusieurs points de vue. Dans l'exemple de Rees le centralisateur H de γ_0 dans G est localement isomorphe à $GL_2(\mathbb{R})$ et $H \cap \Gamma$ est un réseau de H ; ainsi $\Gamma\backslash\Gamma H$, qui est une sous-variété fermée de $\Gamma\backslash G$, est aussi un fibré en cercles au-dessus de $\Gamma\backslash\Gamma H_0$, qui est un quotient de volume fini de $H_0 \simeq SL_2(\mathbb{R})$. L'action de A est le produit d'une action isométrique dans la fibre et du flot géodésique sur $\Gamma\backslash\Gamma H_0$. Ainsi toutes les pathologies du flot géodésique de $SL_2(\mathbb{R})$ apparaissent dans ce contexte. En particulier dans un tel espace $\Gamma\backslash G$ il y a pléthore de compacts A-invariants irréguliers : des fibrés en cercles de compacts irréguliers invariants par le flot géodésique sur la surface hyperbolique plongée ; ces compacts donnent aussi lieu à des mesures A-invariantes irrégulières et qui ont même entropie positive au moins pour des directions de A transverses aux fibres. On renvoie le lecteur à la section 9 de l'article de Einsiedler et Katok [**18**] pour une description détaillée de ces exemples.

En résumé, ces exemples démontrent que toute généralisation hâtive de la conjecture 4.1 ci-dessus est vouée à l'échec. Margulis avait donc proposé dans [**46**] un énoncé conjectural, décrivant l'adhérence dans $\Gamma\backslash G$ des orbites d'un sous-groupe fermé de G, énoncé qui englobait à la fois le théorème de Ratner et les orbites des flots diagonaux tout en tenant compte de la présence de ces exemples à la Rees. Malheureusement même cet énoncé s'est avéré insuffisant : Maucourant a en effet construit dans [**49**] un nouvel exemple non algébrique d'adhérence d'orbite. Pour $G = SL_3(\mathbb{R}) \times SL_3(\mathbb{R})$ et $\Gamma = SL_3(\mathbb{Z}) \times SL_3(\mathbb{Z})$ il considère l'orbite par le sous-groupe de codimension un A_1 de $A \times A = \mathbb{R}^{*2} \times \mathbb{R}^{*2}$ défini par $\lambda_1/\lambda_3 = \mu_3/\mu_1$ (où $\mathrm{diag}(\lambda_1, \lambda_2, \lambda_3)$ et $\mathrm{diag}(\mu_1, \mu_2, \mu_3)$ sont les deux composantes du tore diagonal) du point $p := (\Delta_0 \cdot u, \Delta_0 \cdot u)$ où $u = u(1)$ est un élément unipotent non trivial du sous-groupe racinaire $\{u(t)\}_{t\in\mathbb{R}}$ associé à la racine $\lambda_1 - \lambda_3$ (coin supérieur droit) et où Δ_0 est un réseau dont la A-orbite est compacte. Vu le choix de A_1, il est clair que $p \cdot A_1$ évite l'ensemble $C^c \times C^c$ où C^c est le complémentaire du voisinage compact $C = \Delta_0 \cdot A\{u(t)\}_{t\in[0,1]}$ de l'orbite compacte $\Delta_0 \cdot A$. Ainsi $p \cdot A_1$ n'est pas dense et Maucourant vérifie, en utilisant le théorème d'isolation de Lindenstrauss et Weiss, qu'elle n'est pas non plus algébrique.

Dans cet exemple de Maucourant, le groupe A_1 n'est pas le tore maximal. Tout récemment Shapira [**69**] a construit un autre exemple surprenant dans $\Omega = SL_3(\mathbb{Z})\backslash SL_3(\mathbb{R})$ cette fois-ci et avec le tore diagonal maximal A :

THÉORÈME 4.5 (Shapira [**69**]). — *Soit $d = 3$. Il existe un réseau unimodulaire $\Delta_0 \in \Omega$ dont la A-orbite n'est pas dense et n'est pas contenue dans un sous-espace homogène fermé propre de Ω.*

Le réseau dans l'exemple de Shapira peut être pris simplement de la forme $\Delta_0 = \mathbb{Z}^3 \cdot u_{a,a}$ (on rappelle la notation $u_{a,b} : (x,y,z) \mapsto (x, y+ax, z+bx)$). Sa méthode consiste à étudier le « minimum inhomogène » $\mu(\Delta) := \sup_{y\in\mathbb{R}^3} \inf_{v\in\Delta} F_0(v+y)$ où F_0 est la forme cubique $F_0(x,y,z) = xyz$ et à vérifier que pour ces exemples $\mu(\Delta_0) > 0$. Mais une adaptation de la preuve du théorème d'isolation lui permet d'établir que $\mu(\Delta) = 0$ dès que $\overline{\Delta \cdot A}$ contient une orbite compacte, en particulier dès que $\Delta \cdot A$ est dense. Cela prouve la non-densité ; pour obtenir la seconde assertion, il faut faire varier a et tirer parti du fait qu'il n'y a qu'un nombre dénombrable de sous-espaces homogènes fermés.

Dans l'exemple de Cassels décrit au paragraphe précédent, la suite de A-orbites compactes s'accumule sur la A-orbite (fermée mais non compacte) de $\mathbb{Z}^d$. Ainsi il peut exister des fermés A-invariants non algébriques de Ω qui contiennent une infinité de A-orbites compactes.

Pour une description des A-orbites fermées (compactes ou non) voir l'article de Tomanov et Weiss [**71**] où il est montré entre autres qu'une A-orbite est fermée si et seulement si c'est la translatée de l'orbite d'un $\mathbb{Q}$-tore et à l'article de Tomanov [**70**] pour le cas S-arithmétique et l'application concrète suivante : si $f = l_1 \cdot \cdots \cdot l_k$ est un produit de $k \leq d$ formes linéaires reélles linéairement indépendantes tel que $f(\mathbb{Z}^d)$ soit discret dans $\mathbb{R}$, alors f est proportionnel à un polynôme à coefficients entiers. Dans l'appendice de [**71**], on trouve aussi une preuve due à Margulis de la propriété structurelle importante suivante :

PROPOSITION 4.6. — *Il existe un compact C de Ω qui rencontre toutes les A-orbites.*

La preuve de ce résultat est basée sur le critère de Mahler et un très joli jeu d'élimination des petits vecteurs d'un réseau qui rappelle le fameux lemme de Kazhdan et Margulis ([**58**] Chapitre 8).

5. DYNAMIQUE DES FLOTS DIAGONAUX DANS Ω : CLASSIFICATION DES MESURES INVARIANTES, CONDITION D'ENTROPIE POSITIVE ET PRINCIPE DE LINNIK

Comme dans le théorème de Ratner, qui classifiait d'abord les mesures invariantes par un flot unipotent et ensuite les fermés invariants, on s'attend à ce que les mesures A-invariantes soient plus faciles à décrire que les fermés A-invariants. Par exemple, considérons dans $\Omega = \Gamma\backslash G$ pour $d = 3$ l'orbite $\Delta \cdot A$ d'un réseau Δ qui intersecte $\mathbb{Z}^2$ en un réseau $\Delta_{\mathbb{Z}^2}$ dont l'orbite par le flot diagonal (la partie de A qui préserve $\mathbb{R}^2$) est irrégulière dans $SL_2(\mathbb{Z})\backslash SL_2(\mathbb{R})$. Alors $\Delta \cdot A$ est irrégulière, mais en raison de la

récurrence de Poincaré (il y a un sous-groupe $\{a_t\}_t$ de A tel que $\Delta \cdot A \to \infty$), celle-ci ne supporte aucune mesure A-invariante finie.

La conjecture suivante est formulée par Margulis dans ([**46**]) et précisée par Lindenstrauss dans ([**41**]). Disons que l'orbite $\Delta \cdot A$ a « un facteur de rang 1 » s'il existe un sous-groupe fermé F de G contenant A et un morphisme surjectif de groupes de Lie $F \xrightarrow{\phi} L$ vers un groupe de Lie L tel que $\Delta \cdot F$ soit fermé, $\phi(F_\Delta)$ soit fermé – où $F_\Delta = \{f \in F, \Delta \cdot f = \Delta\}$ – et $\phi(A)$ soit un groupe à un paramètre non Ad-unipotent de L.

CONJECTURE 5.1. — *Soit μ une mesure de probabilité A-invariante et A-ergodique sur $\Omega = \Gamma\backslash G$. Supposons que pour μ-presque tout $\Delta \in \Omega$ l'orbite $\Delta \cdot A$ n'a pas de facteur de rang* 1. *Alors μ est algébrique – i.e. est la mesure de Haar normalisée d'un sous-espace homogène de Ω.*

Notons que lorsque $d = 3$ et $\Gamma = SL_3(\mathbb{Z})$ la condition d'absence d'orbite avec facteur de rang 1 est toujours satisfaite. Les exemples de Rees mentionnés ci-dessus démontrent que ce proviso est néanmoins nécessaire en dimension plus grande ou pour d'autres réseaux.

D'un certain point de vue, cette conjecture est l'analogue pour l'espace des réseaux de la célèbre conjecture de Furstenberg : toute mesure de probabilité non atomique sur le cercle invariante par la multiplication par 2 et par la multiplication par 3 est-elle égale à la mesure de Lebesgue ? Rudolph [**65**] a démontré la conjecture de Furstenberg sous l'hypothèse supplémentaire que la mesure a de l'entropie positive par rapport à la multiplication par 2, disons. De façon analogue, le théorème de Einsiedler-Katok-Lindenstrauss, théorème 1.1 ci-dessus, répond à la conjecture 5.1 sous l'hypothèse supplémentaire d'entropie positive par rapport à une direction du groupe diagonal A.

La preuve de Einsiedler-Katok-Lindenstrauss est basée sur deux méthodes de natures assez distinctes. La première, dite de haute entropie, tire ses origines dans les articles de Katok-Spatzier [**37**] et Einsiedler-Katok [**18**]. L'idée maîtresse consiste à tirer parti du fait que le système dynamique (Ω, A, μ) est partiellement hyperbolique et possède un feuilletage stable et instable correspondant aux orbites des groupes racinaires unipotents normalisés par A : on peut alors découper μ en tranches le long de ces feuilletages, c'est-à-dire étudier les mesures conditionnelles (d'une certaine façon c'est là que réside le principal avantage des mesures par rapport aux fermés : on peut mieux les découper). A chaque sous-groupe racinaire unipotent est associée une famille de mesures conditionnelles aussi appelées « mesures de feuille » – *leafwise measures* – et l'entropie totale de la mesure se scinde en contributions distinctes propres aux différentes feuilles (formule de Ledrappier-Young [**40**]). En particulier une condition plus ou moins forte d'entropie positive sur μ pour un élément $a \in A$ entraîne que les mesures de feuille associées à un nombre plus ou moins grand de racines sont

non atomiques. La méthode de [**18**] permet de montrer que, si les mesures de feuille associées à toutes les racines de G sont non atomiques, alors μ est invariante par rapport à tous les sous-groupes racinaires et donc est G-invariante. Dans son travail sur la conjecture d'unique ergodicité quantique Elon Lindenstrauss [**42**] a mis en œuvre une autre méthode, dite de basse entropie, dont l'analogue dans notre contexte permet d'obtenir la même conclusion en supposant seulement que les mesures de feuille sont non atomiques pour un certain nombre de racines (quand $d = 3$ une suffit). Sa méthode est inspirée de la preuve de Host [**36**] du théorème de Rudolph qui utilise la récurrence comme substitut à l'entropie positive, ainsi que de plusieurs arguments introduits par Ratner dans son étude des flots unipotents [**59**], [**60**], [**61**], [**63**]. Nous n'en dirons pas davantage sur ces travaux remarquables et renvoyons le lecteur aux articles originaux et aux nombreux survols écrits sur le sujet, notamment [**21**], [**3**], [**73**].

5.1. Rappels sur l'entropie[(3)]

Soit $(X, \mathcal{B}, \mu)$ un espace mesuré muni d'une mesure de probabilité μ. Soit T une transformation mesurable de X qui préserve la mesure μ. Soit $\mathcal{P}$ une partition finie de X en parties mesurables appelées atomes de $\mathcal{P}$. L'entropie de la partition $\mathcal{P}$ est par définition

$$H_\mu(\mathcal{P}) = \sum_{C \in \mathcal{P}} -\mu(C) \log \mu(C).$$

On note $\mathcal{P} \vee Q$ le joint de $\mathcal{P}$ et Q, i.e. la partition obtenue en prenant les intersections des parties de $\mathcal{P}$ avec celles de Q. On note $[x]_\mathcal{P}$ l'atome de $x \in X$ pour $\mathcal{P}$, i.e. l'élément C de la partition $\mathcal{P}$ qui contient x. Aussi $T^{-1}\mathcal{P}$ est la partition $\{T^{-1}C\}_{C \in \mathcal{P}}$. On définit l'entropie de T par rapport à $\mathcal{P}$ comme la limite

$$h_\mu(T, \mathcal{P}) = \lim_{n \to +\infty} \frac{1}{n} H_\mu(\vee_0^{n-1} T^{-k}\mathcal{P}),$$

et l'entropie de T comme le supremum $h_\mu(T)$ des $h_\mu(T, \mathcal{P})$ où $\mathcal{P}$ varie parmi toutes les partitions mesurables finies de X. Les propriétés de base de l'entropie dont nous aurons besoin sont les suivantes :

1) La sous-additivité : $H_\mu(\mathcal{P} \vee Q) \leq H_\mu(\mathcal{P}) + H_\mu(Q)$.

2) L'invariance : $H_\mu(T^{-1}\mathcal{P}) = H_\mu(\mathcal{P})$.

3) La génération : si $\left\{T^{-k}\mathcal{P}\right\}_{k \geq 0}$ engendre la tribu $\mathcal{B}$ alors $h_\mu(T) = h_\mu(T, \mathcal{P})$.

4) La décomposition ergodique : si $\mu = \int_\Xi \mu_\xi d\nu(\xi)$ est une décomposition de μ en composantes ergodiques pour T, alors $h_\mu(T) = \int_\Xi h_{\mu_\xi}(T) d\nu(\xi)$.

(3) Voir le livre de Einsiedler et Ward [**25**].

5.2. Principe de Linnik

Donnons maintenant l'idée de la démonstration de la proposition 1.4 de l'introduction. Il s'agit d'établir une borne inférieure sur l'entropie de toute mesure limite μ_∞. Modulo quelques difficultés purement techniques dues au fait que l'on travaille dans un espace non compact $\Gamma\backslash G$, il se trouve que moyennant la définition et les propriétés de base de l'entropie rappelées ci-dessus cette borne inférieure résulte presque directement de la séparation des A-orbites compactes de discriminant au plus D (proposition 2.4 ci-dessus).

En effet donnons l'argument en substance. Soit $a \in A$ un élément régulier. Si $\mathscr{P}$ est une partition finie de $X = \Omega = \Gamma\backslash G$, on va s'intéresser à $\mathscr{P}^m := \vee_{-m}^m T_a^k \mathscr{P}$ où T_a est la translation à droite par a. Le point clé est que notre système dynamique est partiellement hyperbolique : T_a contracte exponentiellement une partie de l'espace tangent $\mathfrak{g}^+$ (correspondant à la somme des espaces propres des racines positives $\sum_{\alpha>0} \mathfrak{g}_\alpha$), dilate $\mathfrak{g}^- = \sum_{\alpha<0} \mathfrak{g}_\alpha$ et laisse inchangée la direction du tore $Lie(A) = \mathfrak{g}_0$. Ainsi (et c'est là que quelques difficultés techniques apparaissent liées à la non-compacité de $\Gamma\backslash G$) lorsque m est grand les atomes de $\mathscr{P}^m$ sont inclus dans des $e^{-\kappa m}$-voisinages de morceaux de taille fixe (i.e. de A-volume borné) de A-orbites, où $\kappa > 0$ mesure le coefficient de contraction/dilatation de T_a, coefficient qui dépend de la distance de a aux murs de la chambre de Weyl, i.e. à la plus petite valeur de $|\alpha(a)|$ pour α une racine (ici, différence entre deux valeurs propres de a).

Rappelons maintenant le lemme 2.4 de séparation des orbites : les orbites dans le support de μ_n sont D_n^{-c}-séparées. Choisissons m de sorte que $e^{-\kappa m} \approx D_n^{-c}$. Il résulte que chaque atome $[x]_{\mathscr{P}^m}$ de $\mathscr{P}^m$ intersecte au plus une orbite compacte présente dans la réunion d'orbites Y_n et donc vérifie $\mu_n([x]_{\mathscr{P}^m}) = O(\mathrm{vol}(Y_n)^{-1}) = O(D_n^{-\rho})$. Par suite $H_{\mu_n}(\mathscr{P}^m) \geq \rho \log D_n \approx \rho \frac{\kappa}{c} m$. Maintenant, par sous-additivité et invariance de l'entropie, $H_{\mu_n}(\mathscr{P}^m) \leq (2m+1) \cdot H_{\mu_n}(\mathscr{P})$ et $H_{\mu_n}(\mathscr{P}) \geq 2\rho\frac{\kappa}{c}$. Faisant tendre n vers l'infini on obtient $H_{\mu_\infty}(\mathscr{P}) \geq 2\rho\frac{\kappa}{c}$. Mais ce raisonnement peut être répété verbatim avec $\mathscr{P}^p$ pour un entier $p \in \mathbb{N}$ à la place de $\mathscr{P}$. On obtient alors $H_{\mu_\infty}(\mathscr{P}^p) \geq 2\rho\frac{\kappa}{c}p$, et en faisant tendre p vers l'infini, $h_{\mu_\infty}(T_a) \geq 2\rho\frac{k}{c}$. $CQFD$.

6. PAQUETS ET FORMULATION ADÉLIQUE

Lorsque $d = 2$, l'étude des paquets de A-orbites compactes remonte implicitement à Gauss dans les Disquitiones lorsque celui-ci étudie les formes quadratiques à deux variables et à coefficients entiers à changement de variables à coefficients entiers près. À chaque telle classe de formes quadratiques $ax^2 + bxy + cy^2$ avec $pgcd(a,b,c) = 1$ on peut associer le discriminant $D = b^2 - 4ac$ et si $D > 0$ la géodésique du demi-plan supérieur $\mathbb{H}^2$ reliant les deux racines réelles du trinôme. Ces géodésiques sont

fermées dans la surface modulaire $SL_2(\mathbb{Z})\backslash\mathbb{H}^2$. Gauss montre que le nombre de classes d'équivalence de formes ayant même discriminant D est fini et que cet ensemble de classes s'identifie au groupe des classes d'idéaux du corps quadratique $\mathbb{Q}(\sqrt{D})$. Ce sont là les paquets d'orbites. On renvoie le lecteur à **[34]** pour un excellent résumé du cas $d = 2$ et du théorème d'équidistribution correspondant, i.e. le théorème de Duke (voir aussi les travaux antérieurs de Zelditch **[77]**).

Décrivons maintenant ces paquets en dimension supérieure. Soit K un corps de nombres totalement réel et de degré d sur $\mathbb{Q}$. La construction des orbites compactes décrite au §2 permet notamment d'associer une orbite compacte à chaque idéal I de l'anneau des entiers $\mathcal{O}_K$ de K, ou plus précisément à chaque classe d'idéal, car deux réseaux de K égaux à homothétie près par un scalaire non nul de K donnent lieu aux mêmes A-orbites compactes. Plus généralement si L est un réseau de K (i.e. un sous-$\mathbb{Z}$-module de rang d) alors on peut considérer les réseaux de la forme $I \cdot L$ où I varie parmi les idéaux inversibles de l'ordre $\mathcal{O}_L = \{x \in K, xL \subseteq L\}$. Ceci nous donne une action bien définie du groupe de Picard de l'ordre $\mathcal{O}_L$ sur l'ensemble des A-orbites compactes dont l'ordre associé est $\mathcal{O}_L$. Cette action est sans point fixe. Une classe d'équivalence pour cette relation est appelée un *paquet d'orbites*. On a (cf. **[53]** I.12.6) :

PROPOSITION 6.1. — *Les A-orbites compactes $(K, [L], \sigma)$ et $(K, [L'], \sigma)$ associées à deux réseaux L et L' de K sont dans le même paquet si et seulement si ils sont localement égaux à un scalaire près, i.e. si pour tout nombre premier p, il existe un élément inversible λ_p de $K \otimes_{\mathbb{Q}} \mathbb{Q}_p$ tel que $L \otimes_{\mathbb{Z}} \mathbb{Z}_p = \lambda_p(L' \otimes_{\mathbb{Z}} \mathbb{Z}_p)$.*

On peut aussi décrire un paquet en termes du $\mathbb{Q}$-tore associé au réseau L. Le bon cadre pour cela est le cadre adélique. Soit donc $\mathbb{A}$ l'anneau des adèles sur $\mathbb{Q}$ et K_f un sous-groupe compact ouvert des adèles finies $\mathbb{G}(\mathbb{A}_f)$ où $\mathbb{G} = PGL_d$ tel que $\Gamma = \mathbb{G}(\mathbb{Q}) \cap K_f = PGL_d(\mathbb{Z})$. On pose $X_{\mathbb{G}} = \mathbb{G}(\mathbb{Q})\backslash\mathbb{G}(\mathbb{A})/K_f$. D'après la finitude du nombre de classes de $\mathbb{G}$ (voir **[56]** ch. 5), l'espace $X_{\mathbb{G}}$ est une réunion finie de copies de $\Gamma\backslash\mathbb{G}(\mathbb{R})$. On note $X^0_{\mathbb{G}}$ la composante $\Gamma\backslash\mathbb{G}(\mathbb{R})$ correspondant à la double classe de l'identité dans $\mathbb{G}(\mathbb{A}_f)$ i.e. $\mathbb{G}(\mathbb{Q})K_f$.

Soit A le tore diagonal de $PGL_d(\mathbb{R})$ et $\mathbb{T}$ un $\mathbb{Q}$-tore $\mathbb{Q}$-anisotrope et $\mathbb{R}$-déployé de $\mathbb{G}$ et soit $g = (g_\infty, 1) \in \mathbb{G}(\mathbb{A})$, tel que $\mathbb{T}(\mathbb{R}) = g_\infty A g_\infty^{-1}$. On considère *l'orbite torique homogène* $\mathbb{T}(\mathbb{Q})\backslash\mathbb{T}(\mathbb{A})g$ vue comme partie de $\mathbb{G}(\mathbb{Q})\backslash\mathbb{G}(\mathbb{A})$ et on pose

$$S(\mathbb{T}, g) := \mathbb{T}(\mathbb{Q})\backslash\mathbb{T}(\mathbb{A})gK_f/K_f.$$

Cette partie est située dans la composante $X^0_{\mathbb{G}} = \Gamma\backslash\mathbb{G}(\mathbb{R}) = \Omega$. La finitude du nombre de classes de $\mathbb{T}$ et de $\mathbb{G}$ permet d'écrire $\mathbb{T}(\mathbb{A}_f)K_f = \cup_i \mathbb{T}(\mathbb{Q})\alpha_i^{-1}u_iK_f$ où $\alpha_i \in \mathbb{G}(\mathbb{Q}) \cap \mathbb{G}(\mathbb{A}_f)$ et $u_i \in \mathbb{G}(\mathbb{A}_f)$ est un représentant de $\mathbb{G}(\mathbb{Q})\backslash\mathbb{G}(\mathbb{A}_f)/K_f$. Et

puisque g commute à $\mathbb{G}(\mathbb{A}_f)$ on obtient
(1)
$$S(\mathbb{T},g)=\cup_i\mathbb{G}(\mathbb{Q})\backslash\mathbb{G}(\mathbb{Q})(\alpha_i\mathbb{T}(\mathbb{R})g,u_iK_f)/K_f=\cup_i\Gamma\cap\mathbb{T}_i(\mathbb{R})\backslash\mathbb{T}_i(\mathbb{R})h_i=\cup_i\Gamma\backslash\Gamma h_iA$$

où $h_i=\alpha_ig_\infty$ et $\mathbb{T}_i=\alpha_i\mathbb{T}\alpha_i^{-1}$ avec $\alpha_i\in\mathbb{G}(\mathbb{Q})\cap K_f\mathbb{T}(\mathbb{A}_f)$. On vérifie sans difficulté la proposition suivante (cf. **[24]**).

PROPOSITION 6.2. — *Les $S(\mathbb{T},g)$ où $\mathbb{T}$ et $g=(g_\infty,1)$ varient parmi les $\mathbb{Q}$-tores $\mathbb{Q}$-anisotropes de $\mathbb{G}$ tels que $\mathbb{T}(\mathbb{R})=g_\infty Ag_\infty^{-1}$ forment une partition de l'ensemble des A-orbites compactes. Ce sont exactement les paquets : deux A-orbites sont dans le même paquet si et seulement si elles appartiennent à un même $S(\mathbb{T},g)$. Le groupe des classes $\mathbb{T}(\mathbb{Q})\backslash\mathbb{T}(\mathbb{A}_f)/(\mathbb{T}(\mathbb{A}_f)\cap K_f)$ agit simplement transitivement sur $S(\mathbb{T},g)$. De plus deux A-orbites situées dans le même paquet ont même volume et même discriminant.*

Si $(\mathbb{T}_1,g_1)$ et $(\mathbb{T}_2,g_2)$ sont les données associées à deux orbites compactes $\Gamma\backslash\Gamma g_1A$ et $\Gamma\backslash\Gamma g_2A$ alors celles-ci appartiennent au même paquet si et seulement si $\mathbb{T}_2=\delta\mathbb{T}_1\delta^{-1}$ et $g_2=\delta tg_1$ où $\delta\in\mathbb{G}(\mathbb{Q})\cap K_f\mathbb{T}_1(\mathbb{A}_f)$ et $t\in\mathbb{T}_1(\mathbb{R})$. Ainsi les orbites compactes d'un même paquet sont paramétrées par l'ensemble des doubles classes $(\mathbb{G}(\mathbb{Q})\cap K_f)\backslash\mathbb{G}(\mathbb{Q})\cap K_f\mathbb{T}_1(\mathbb{A}_f)/\mathbb{T}_1(\mathbb{Q})$, qui lui-même s'identifie au groupe des classes $(\mathbb{T}_1(\mathbb{A}_f)\cap K_f)\backslash\mathbb{T}_1(\mathbb{A}_f)/\mathbb{T}_1(\mathbb{Q})$, i.e. le groupe de Picard de l'ordre correspondant.

L'estimation de volume suivante résulte de la borne inférieure de Brauer-Siegel sur le résidu au pôle de la fonction zeta de Dedekind d'un corps de nombres K (cf. **[39]**) :

PROPOSITION 6.3 (**[24]** Theorem 4.8.). — *Le volume total A-invariant $\mathrm{vol}(Y)$ d'un paquet Y de discriminant D vérifie $\log\mathrm{vol}(Y)=\frac{1}{2}\log D+o(\log D)$.*

Dans une série d'articles antérieurs (en particulier **[10]**) Clozel et Ullmo ont systématisé l'étude des paquets $S(\mathbb{H},g)$ et de leur répartition à d'autres sous-groupes $\mathbb{H}$. Soit $\mathbb{G}$ un $\mathbb{Q}$-groupe semi-simple, K_f un sous-groupe compact ouvert des adèles finies de $\mathbb{G}$ sur $\mathbb{Q}$. On pose $G=\mathbb{G}(\mathbb{R})^+$ la composante connexe (topologique) de l'identité, $\mathbb{G}(\mathbb{Q})^+=\mathbb{G}(\mathbb{Q})\cap\mathbb{G}(\mathbb{R})^+$, $\mathbb{G}(\mathbb{A})^+=\mathbb{G}(\mathbb{R})^+\mathbb{G}(\mathbb{A}_f)$ et $\Gamma=\mathbb{G}(\mathbb{Q})^+\cap\mathbb{K}_f$. L'ensemble $X_\mathbb{G}=\mathbb{G}(\mathbb{Q})^+\backslash\mathbb{G}(\mathbb{A})^+/K_f$ est une réunion finie de G orbites : on note $\Omega=\Gamma\backslash G$ la double classe de l'identité. Pour $g=(g_\infty,1)\in\mathbb{G}(\mathbb{A})$, comme précédemment on pose $S(\mathbb{H},g):=\mathbb{H}(\mathbb{Q})^+\backslash\mathbb{H}(\mathbb{A})^+g/\mathbb{H}(\mathbb{A})\cap K_f$ vu comme partie de $X_\mathbb{G}$ et $S^0(\mathbb{H},g)$ l'intersection avec Ω. On note α la donnée $\alpha=(\mathbb{H},g)$. Soit μ_α la mesure $g_\infty^{-1}\mathbb{H}(\mathbb{R})^+g_\infty$-invariante sur $S^0(\mathbb{H},g)$. Clozel et Ullmo conjecturent (voir **[10]** p. 1258 pour une forme plus faible sans les g_α) :

CONJECTURE 6.4. — *Soit $(\mathbb{H}_\alpha,g_\alpha)_\alpha$ une suite stricte de $\mathbb{Q}$-groupes sans $\mathbb{Q}$-caractères non triviaux avec $g_\alpha=(g_\infty,1)\in\mathbb{G}(\mathbb{A}_\mathbb{Q})$. Alors la suite de mesures de probabilité μ_α converge vers la mesure de Haar sur $\Gamma\backslash G$.*

La suite $(\mathbb{H}_\alpha, g_\alpha)_\alpha$ est dite *stricte* si pour tout $\mathbb{Q}$-sous-groupe $\mathbb{L}$ de $\mathbb{G}$ et tout $g \in \mathbb{G}(\mathbb{A})$ tel que $S^0(\mathbb{L}, g) \neq \Omega$ il n'y a qu'un nombre fini de α tels que $S^0(\mathbb{H}_\alpha, g_\alpha) \subseteq S^0(\mathbb{L}, g)$. C'est clairement une condition nécessaire.

Le théorème 1.6 de Einsiedler, Lindenstrauss, Michel et Venkatesh, résultat principal qui nous occupe dans ce Bourbaki, correspond à l'énoncé précédent dans le cas particulier où $\mathbb{G} = PGL_3$ et $\mathbb{H}_\alpha$ est une suite de tores maximaux anisotropes sur $\mathbb{Q}$ et déployés sur $\mathbb{R}$. Antérieurement, Clozel et Ullmo dans [**10**] ainsi que Cohen [**12**] et Zhang [**78**], ont traité le cas de PGL_2 conditionnellement à la sous-convexité, condition levée par Venkatesh dans [**74**]. On renvoie au troisième article [**23**] pour plus d'information sur le cas $G = PGL_2$.

Le cas où les H_α sont semi-simples a été étudié par différents auteurs. Dans [**10**] Clozel et Ullmo observent que, dans le cas particulier où les g_α sont tous égaux à l'identité et les $H_\alpha(\mathbb{R})$ sont sans facteurs compacts, la conjecture 6.4 est un corollaire des fameux théorèmes de Ratner et Mozes-Shah sur la classification des mesures invariantes par les flots unipotents. Dans un impressionnant travail récent [**33**] de Gorodnik et Oh, ce résultat est grandement étendu au cas où on permet aux g_α de varier et où les H_α sont semi-simples et possèdent une place où ils sont simultanément isotropes ; leur résultat décrit même les mesures limites possibles sans la condition de suite « stricte ». En fait, Gorodnik et Oh formulent et démontrent un résultat plus fort, à savoir l'équidistribution adélique, i.e. dans $\mathbb{G}(\mathbb{Q})\backslash\mathbb{G}(\mathbb{A})$. C'est aussi sous cette forme qu'est démontré le théorème 1.6 dans [**24**]. Cette généralisation est cruciale pour les applications arithmétiques tel le comptage des points S-entiers ou des points rationnels de hauteur bornée dans une variété homogène (i.e. $X = G/H$). On renvoie au survol de Oh [**55**] pour une introduction claire à ce travail ainsi que pour les applications arithmétiques au comptage de points (conjecture de Manin).

7. SÉRIES D'EISENSTEIN ET SOUS-CONVEXITÉ

Pour montrer l'équidistribution d'une suite de mesures (par exemple comme dans le théorème 1.6) deux méthodes s'offrent à nous a priori : ou bien utiliser l'analyse harmonique en vérifiant le critère de Weyl pour l'équidistribution (ici montrer que chaque composante de la décomposition spectrale de $L^2(\Gamma\backslash G)$ a une intégrale par rapport à μ_n qui tend vers zéro quand n tend vers l'infini, sauf pour les fonctions constantes) ou bien utiliser la théorie ergodique en montrant la rigidité des mesures valeurs d'adhérence de la suite μ_n. La première méthode est celle utilisée par Duke dans le cas $d = 2$. Quand $d \geq 3$ cette méthode n'est plus si aisée à mettre en œuvre parce qu'on n'a de lien direct avec les fonctions L que pour une petite partie du spectre, i.e. les séries de Siegel-Eisenstein, et on ne s'attend pas à ce qu'un lien semblable existe

pour le reste de la décomposition spectrale de $L^2(\Gamma\backslash G)$ en particulier pour sa partie cuspidale. La seconde méthode, ergodique, n'est pas non plus suffisante car elle ne permet pas par exemple de contrôler la fuite de masse ni d'exclure la possibilité de plusieurs composantes ergodiques à la limite. Le succès de l'approche de Einsiedler, Lindenstrauss, Michel et Venkatesh dans [**24**] *réside dans la combinaison des deux approches.* Le critère de Weyl est vérifié sur les séries de Siegel-Eisenstein. C'est le corollaire 7.4 ci-dessous. Comme nous l'expliquons plus bas il résulte, via la formule de Hecke, de la sous-convexité des fonctions zeta de Dedekind associées ainsi que d'un calcul local. Mais le point clé est que l'équidistribution des séries de Siegel-Eisenstein est suffisante à la fois pour exclure l'éventualité d'une fuite de masse et pour montrer que toute composante ergodique de toute mesure limite a de l'entropie positive. Il suffit alors d'utiliser le théorème 1.1 pour conclure. Voici maintenant plus de détails.

7.1. Séries de Siegel-Eisenstein

Soit f une fonction continue à support compact inclus dans $\mathbb{R}^d\backslash\{0\}$. La transformée de Siegel de f est la fonction définie sur l'espace des réseaux $\Omega = SL_d(\mathbb{Z})\backslash SL_d(\mathbb{R})$ par

$$\widetilde{f}(\Delta) = \sum_{v\in\Delta\backslash\{0\}} f(v).$$

Cette fonction est dans $L^{d-\varepsilon}(\Omega)$. L'identité suivante, dite formule de Siegel, relie l'intégrale de $\widetilde{f}$ sur Ω avec l'intégrale de f sur $\mathbb{R}^d$:

$$\int_{\mathbb{R}^d} f(x)dx = \int_\Omega \widetilde{f}d\Delta \tag{2}$$

où $d\Delta$ est la mesure de Haar normalisée sur Ω et dx la mesure de Lebesgue sur $\mathbb{R}^d$. Cette formule est facile à vérifier, en effet, $f \to \int_\Omega \widetilde{f}d\Delta$ définit une fonctionnelle continue sur $C_c(\mathbb{R}^d\backslash\{0\})$ invariante par $SL_d(\mathbb{R})$. Une telle mesure est nécessairement de la forme $\alpha dx + \beta\delta_0$. Le critère de Mahler implique que $\beta = 0$ et en prenant $f_t(x) := t^d f(tx)$ la transformée $\widetilde{f_t}$ devient une somme de Riemann pour $t \to 0$ qui tend point par point vers le membre de gauche de (2), donc $\alpha = 1$ par convergence dominée.

Afin de faire le lien avec les fonctions L on construit la *série de Siegel-Eisenstein de* f en appliquant la transformée de Mellin à f (i.e. la transformée de Fourier sur $\mathbb{R}_+^*$ multiplicatif c'est-à-dire $f_s(v) = \int_{\mathbb{R}^*} f(tv)t^s d^\times t$ où $d^\times t = \frac{dt}{|t|}$) puis en posant

$$E_f(s,\Delta) = \widetilde{f_s}(\Delta) = \sum_{v\in\Delta\backslash\{0\}} f_s(v).$$

Cette série converge si $\mathrm{Re}(s) > d$. C'est une forme automorphe qui admet un prolongement méromorphe à tout le plan complexe avec pour seuls pôles 0 et d. Elle vérifie

de plus une équation fonctionnelle reliant ses valeurs de part et d'autre de la droite critique $\mathrm{Re}(s) = \frac{d}{2}$.

Le résidu en $s = d$ se calcule aisément, il vaut $\int_{\mathbb{R}^d} f(x)dx$. Pour retrouver $\widetilde{f}$, la transformée de Siegel de f, on applique la transformée de Mellin inverse, i.e.

$$\widetilde{f}(\Delta) = \int_{\mathrm{Re}(s)=\sigma>d} E_f(s,\Delta)\frac{ds}{2\pi i}.$$

On peut alors déplacer le contour d'intégration vers la droite critique, i.e.

$$\widetilde{f}(\Delta) - \int_{\mathbb{R}^d} f(x)dx = \int_{\mathrm{Re}(s)=\frac{d}{2}} E_f(s,\Delta)\frac{ds}{2\pi i}. \tag{3}$$

Il va être utile et même nécessaire pour la suite de raisonner adéliquement et de réécrire $E_f(s,\Delta)$ après un calcul immédiat sous la forme plus générale

$$E_f(s,\Delta) = |\det g_\infty|^{s/d}\cdot \widetilde{E}_\psi(s,g) = |\det g_\infty|^{s/d}\cdot \int_{\mathbb{Q}^\times\backslash\mathbb{A}^\times} \sum_{v\in\mathbb{Q}^d\setminus\{0\}} \psi(vtg)|t|_{\mathbb{A}}^s d^\times t \tag{4}$$

où $d^\times t$ est la mesure de Haar normalisée sur $\mathbb{Q}^\times\backslash\mathbb{A}_{\mathbb{Q}}^\times$, $\psi(v) := f(v_\infty)1_{U_f}(v_f)$ si $v = (v_\infty, v_f) \in \mathbb{A}_{\mathbb{Q}}^d$ et $U_f = \prod_p' \mathbb{Z}_p^d$, $g = (g_\infty, 1) \in \mathbb{G}(\mathbb{A})$, $\mathbb{G} = PGL_d$, et $\mathbb{Z}^d\cdot g_\infty = \Delta$. Noter que $\widetilde{E}_\psi(s,g)$ est bien défini pour tout $g \in \mathbb{G}(\mathbb{A})$, $\mathbb{G} = PGL_d$, et $\mathrm{Re}(s) > d$. Noter aussi, vu la forme de ψ, que $\widetilde{E}_\psi(s,gk) = \widetilde{E}_\psi(s,g)$ pour tout $k \in K_f$.

7.2. Formule de Hecke pour les intégrales toroïdales

Dans l'optique du théorème 1.6 la quantité qui nous intéresse est $\int_\Omega \widetilde{f}d\mu_n$; c'est l'intégrale de $\widetilde{f}$ par rapport à un paquet d'orbites. À la lumière de (3), il nous suffit de calculer $\int_\Omega E_f(s,\Delta)d\mu_n$ sur la droite critique $\mathrm{Re}(s) = \frac{d}{2}$. La *formule de Hecke* (voir à ce titre l'exposé de Wielonsky [**75**]) relie cette intégrale à un produit eulérien qui fait intervenir à la fois la fonction zeta de Dedekind $\zeta_K(s) = \prod_{v\in V_K}\zeta_{K_v}(s)$ du corps de nombres K associé au paquet et un nombre fini de facteurs locaux I_p pour p divisant le discriminant du paquet. Explicitons-la.

Soit $\Delta\cdot A$ une A-orbite compacte et $(K,[L],\sigma)$ la donnée associée (cf. §2). Le corps K est identifié à une sous-$\mathbb{Q}$-algèbre $\phi(K)$ de $M_d(\mathbb{Q})$ égale à $\mathfrak{t}_{\mathbb{Q}}$ où $\mathfrak{t}$ est l'algèbre de Lie du $\mathbb{Q}$-tore $\mathbb{Q}$-anisotrope $\mathbb{T}$ associé à l'orbite. On fixe une identification de K avec $\mathbb{Q}^d$ en posant $\iota : K \to \mathbb{Q}^d$, $\iota(y) = e_1\cdot\phi(y)$ où e_1 est le premier vecteur de la base canonique de $\mathbb{Q}^d$ et $y \in K \overset{\phi}{\hookrightarrow} M_d(\mathbb{Q})$.

Proposition 7.1 (formule de Hecke). — *Soit Y_n un paquet de A-orbites compactes de discriminant D et μ_n la probabilité A-invariante associée. Alors*

$$\int_\Omega E_f(s,\Delta)d\mu_n(\Delta) = \frac{1}{\mathcal{V}}\int_{\mathbb{A}_K^\times} \psi_{K,g}(y)|y|_{\mathbb{A}_K}^{s/d}d^\times y = \frac{1}{\mathcal{V}}\cdot\zeta_K(\frac{s}{d})\cdot I_\infty(f)\cdot\prod_{p|D}\frac{I_p}{\zeta_{K_p}(\frac{s}{d})} \tag{5}$$

où on a noté $\Delta = \mathbb{Z}^d g_\infty$, $g = (g_\infty, 1)$, $\psi_{K,g}(y) = \psi(\iota(y)g)$ *avec* $\iota(y) = e_1 \cdot y$ *et* $\psi(v) := f(v_\infty)1_{U_f}(v_f)$, $d^\times y = \prod_{v\in V_K} d^\times y_v$ *est la mesure de Haar sur* $\mathbb{A}_K^\times$ *définie par la normalisation* $d^\times y_v(\mathcal{O}_{K_v}^\times) = \mathrm{disc}(K_v)^{-\frac{1}{2}}$, $\mathcal{V} = vol_{d^\times y}((\mathbb{A}_K^\times)^1/K^\times))$, $\zeta_{K_p}(s) = \prod_{v|p}\zeta_{K_v}(s)$ *et* $\zeta_{K_v}(s) = (1-q_v^{-s})^{-1}$ *où* q_v *est le cardinal du corps résiduel de* K_v, *et* $I_p = \int_{K_p^\times} 1_{\mathbb{Z}_p^d}(\iota(y))|y|_p^{s/d} d^\times y$ *où* $K_p = K\otimes_{\mathbb{Q}}\mathbb{Q}_p = \prod_{v|p} K_v$ *et* $|y|_p = \prod_{v|p}|y|_v$ *pour* p *premier ou* ∞.

Démonstration. — Notons que $g = (g_\infty, 1)$ et K_f commutent et $\widetilde{E}_\psi(s, kg) = \widetilde{E}_\psi(s,g)$ si $k \in K_f$. Mais par (1)

$$\begin{aligned}\int_\Omega E_f(s,\Delta)d\mu_n(\Delta) &= \frac{1}{\mathrm{vol}(Y_n)}\sum_i \int_{A\cap h_i^{-1}\Gamma h_i\backslash A} |\det h_i a|^{s/d}\cdot \widetilde{E}_\psi(s, h_i a)da\\ &= \frac{1}{\mathrm{vol}(Y_n)}\int_{S(\mathbb{T},g)} |\det tg|^{s/d}\cdot \widetilde{E}_\psi(s,tg)dt\\ &= \frac{1}{\mathcal{V}}\int_{\mathbb{T}(\mathbb{Q})\backslash\mathbb{T}(\mathbb{A})} \widetilde{E}_\psi(s,tg)|t|_{\mathbb{A}_K}^{s/d}dt\end{aligned}$$

où dt est la mesure de Haar sur $\mathbb{T}(\mathbb{Q})\backslash\mathbb{T}(\mathbb{A})$ image de la mesure $d^\times y$ par l'isomorphisme entre $(K^\times\backslash\mathbb{A}_K^\times)/(\mathbb{Q}^\times\backslash\mathbb{A}_\mathbb{Q}^\times)$ et $\mathbb{T}(\mathbb{Q})\backslash\mathbb{T}(\mathbb{A})$ induit par ϕ et $\mathcal{V}$ le volume total de $\mathbb{T}(\mathbb{Q})\backslash\mathbb{T}(\mathbb{A})$. Via cet isomorphisme on calcule facilement la dernière expression en injectant (4) et en dévissant :

$$\frac{1}{\mathcal{V}}\int_{\mathbb{T}(\mathbb{Q})\backslash\mathbb{T}(\mathbb{A})} \widetilde{E}_\psi(s,tg)|t|_{\mathbb{A}_K}^{s/d}dt = \frac{1}{\mathcal{V}}\int_{\mathbb{A}_K^\times}\psi_{K,g}(y)|y|_{\mathbb{A}_K}^{s/d}d^\times y. \tag{6}$$

Enfin la quantité $I = \int_{\mathbb{A}_K^\times}\psi_{K,g}(y)|y|_{\mathbb{A}_K}^{s/d}d^\times y$ se factorise en un produit de facteurs locaux $I = \prod_p I_p$ où p parcourt les nombres premiers et ∞. Si $p\nmid D$ alors $I_p = \prod_{v|p} I_v$ et comme $\psi_{K_v,g_v}(y) = 1_{\mathcal{O}_{K_v}}$, par un calcul facile $I_v = \zeta_{K_v}(\frac{s}{d})$, ce qui permet de mettre (6) sous la forme voulue. □

7.3. Sous-convexité

Il s'agit maintenant de majorer chacun des facteurs de la formule de Hecke sur la droite critique $\mathrm{Re}(s) = \frac{d}{2}$. Le facteur de volume $\frac{1}{\mathcal{V}}$ est un $O_\varepsilon(D^\varepsilon)$ par Brauer-Siegel. Comme conséquence triviale du théorème des nombres premiers, le nombre $n(D)$ de p qui divisent D est un $o(\log D)$, ainsi $\prod_{p|D}|\zeta_{K_p}(\frac{s}{d})|^{-1} \le 2^{n(D)}$ est aussi un $O_\varepsilon(D^\varepsilon)$ si $\mathrm{Re}(s) = \frac{d}{2}$. Reste donc $\zeta_K(\frac{s}{d})$ et les I_p. Lorsque $d = 3$, on dispose de l'estimation de sous-convexité pour la fonction zeta de Dedekind (qui suit des résultats de Duke-Friedlander-Iwaniec [**16**] et Blomer-Harcos-Michel [**4**]) :

THÉORÈME 7.2 (sous-convexité pour les corps cubiques [**16**], [**4**])

Il existe des constantes $N, \theta > 0$ telles que, pour tout corps de nombres K de degré 3 sur $\mathbb{Q}$, on ait sur la droite critique $Re(s) = \frac{1}{2}$

$$|\zeta_K(s)| \ll (1 + |\mathrm{Im}(s)|)^N \cdot \mathrm{disc}(K)^{\frac{1}{4}-\theta} \tag{7}$$

où la constante implicite dans $\ll$ est indépendante de K.

À la place infinie nous avions $I_\infty(f) = \int_{(K\otimes\mathbb{R})^\times} f(e_1 \cdot g_\infty \mathrm{diag}(y))|y|^{s/d} d^\times y = \int_{(\mathbb{R}^d)^\times} f(y)|y|^{\frac{s}{d}} \frac{dy}{|y|}$ où $|y| = \prod_{i=1}^d |y_i|$. Supposons f lisse : on remarque que sur toute droite verticale $\mathrm{Re}(s) = \sigma$, la quantité $I_\infty(f) = \int_{(\mathbb{R}^d)^\times} f(y)|y|^{\frac{s}{d}} d^\times y$ est la transformée de Fourier d'une fonction de $C_c^\infty(\mathbb{R}^d)$ et donc, pour tout $M > 0$,

$$I_\infty(f) \ll_{M,f} (1 + |\mathrm{Im}(s)|)^{-M} \tag{8}$$

sur la droite critique.

Aux places finies nous avions $I_p = \int_{K_p^\times} 1_{\mathbb{Z}_p^d}(\iota(y))|y|_p^{s/d} d^\times y$, avec $d^\times y = \zeta_{K_p}(1)\frac{dy}{|y|_p}$ et $vol_{dy}(\mathcal{O}_{K_p}) = (discK_p)^{-1/2}$ où $\mathcal{O}_{K_p} = \prod_{v|p} \mathcal{O}_{K_v}$. L'estimée clé de [**24**] est la borne suivante établie par Einsiedler, Lindenstrauss, Michel et Venkatesh en toute dimension :

PROPOSITION 7.3. — *Il existe $\alpha = \alpha(d) > 0$ tel que sur la droite $Re(s) = \frac{d}{2}$ et pour tout $\varepsilon > 0$ on ait*

$$\left|\prod_{p|D} I_p\right| \ll_\varepsilon D^\varepsilon \cdot \mathrm{disc}(K)^{-\frac{1}{4}} \cdot \left(\frac{D}{\mathrm{disc}(K)}\right)^{-\alpha}. \tag{9}$$

Finalement la combinaison de (5), (7), (8) avec $M \gg N$ et (9) donne immédiatement le résultat d'équidistribution escompté :

COROLLAIRE 7.4. — *Fixons $d = 3$. Soit f une fonction C^∞ à support compact dans $\mathbb{R}^d\backslash\{0\}$. Alors $\int_\Omega \widetilde{f} d\mu_n \to \int_{\mathbb{R}^d} f(x)dx$.*

Notons que quand n tend vers $+\infty$ le discriminant D_n du paquet Y_n lui aussi tend vers $+\infty$ mais que $\mathrm{disc}(K)$ peut ou non rester borné. S'il reste borné – par exemple si K est fixe – alors on peut se passer de (7) et (9) suffit à établir l'équidistribution, valable alors en toute dimension.

Disons maintenant quelques mots sur la proposition 7.3, qui occupe la majeure partie de [**24**]. L'isomorphisme $\iota : K_p = K \otimes \mathbb{Q}_p \xrightarrow{\iota} \mathbb{Q}_p^d$, qui code la façon dont le tore $\mathbb{T}$ associé au paquet est plongé dans PGL_d, identifie la norme naturelle de K_p, i.e. $N_{K_p}(y) = \max_{v|p} N_{K_v|\mathbb{Q}_p}(y)$, à une norme, disons $N_p(z) := N_{K_p}(\iota^{-1}z)$, sur $\mathbb{Q}_p^d$. D'autre part nous disposons de la norme standard sur $\mathbb{Q}_p^d$ associée à $\mathbb{Z}_p^d$, i.e. $N_0(z) = \max|z_i|$. À chacune de ces normes correspond un réseau de $\mathbb{Q}_p^d$ – i.e. un $\mathbb{Z}_p$-module libre de rang d – qui est l'ensemble des points de norme ≤ 1, soit $\iota(\mathcal{O}_{K_p})$

pour N_p et $\mathbb{Z}_p^d$ pour N_0. Mais $GL_d(\mathbb{Q}_p)$ agit transitivement sur les réseaux et il existe donc $h_p \in GL_d(\mathbb{Q}_p)$ tel que $N_p(z) = N_0(zh_p)$.

Pour avoir (9) il nous suffit d'établir

$$J_p := (vol_{dy}\{N_0(\iota y) \leq 1\})^{-\frac{1}{2}} \int_{K_p^\times} 1_{N_0(\iota y)\leq 1} |y|_p^{-1/2} dy \ll \left(\frac{D_p}{\mathrm{disc}(K_p)}\right)^{-\alpha} \tag{10}$$

pour chaque p divisant $D = \prod D_p$ (D_p est de discriminant local de l'ordre $\mathcal{O}$ associé au paquet). En effet $vol_{dy}\{N_0(\iota y) \leq 1\} = |\det h_p|(discK_p)^{-1/2}$ car dans notre normalisation $vol_{dy}(\mathcal{O}_{K_p}) = (discK_p)^{-1/2}$. Mais l'isomorphisme $\mathbb{A}_K \xrightarrow{\iota} \mathbb{A}_{\mathbb{Q}}^d$ préserve la mesure car il envoie K sur $\mathbb{Q}^d$ et nous avions $vol_{dy}(\mathbb{A}_K/K) = 1$. Ainsi $\prod_p |\det h_p|(discK_p)^{-1/2} = 1$. Comme $\zeta_{K_p}(1) \ll 1$, on peut écrire sur la droite critique $|I_p| \ll \frac{|\det h_p|^{1/2}}{(discK_p)^{1/2}} J_p$, d'où (9).

Le point est donc de borner J_p, i.e. montrer (10). C'est une très jolie méthode géométrique, inspirée de [**10**], qui est mise en œuvre dans [**24**] à cette fin. Cette quantité dépend de la norme $N_0(\iota y)$ sur K_p, et même seulement de la classe d'homothétie de la norme (à un scalaire de $\mathbb{Q}_p$ près) : on peut donc la voir comme une fonction sur l'immeuble de Bruhat-Tits de $PGL_d(\mathbb{Q}_p)$ qui par définition est l'espace des normes – ou des $\mathbb{Z}_p$-réseaux – de $\mathbb{Q}_p^d$ à homothétie près. Le point clé est de montrer que J_p décroît exponentiellement en fonction de la distance combinatoire dans l'immeuble entre $N_0(\iota y)$ – i.e. $\mathbb{Z}_p^d$ dans $\mathbb{Q}_p^d$ – et le sous-immeuble formé de l'orbite de N_{K_p} sous l'action de $K_p^\times$ – i.e. les réseaux $\iota(t\mathcal{O}_{K_p})$, $t \in K_p^\times$, dans $\mathbb{Q}_p^d$. Cette orbite correspond à un appartement de l'immeuble dans le cas où p est totalement décomposé dans K. De plus cette distance combinatoire (notée n) est minorée par $\log_p\left(\frac{D_p}{\mathrm{disc}(K_p)}\right)$: en effet si $p^n\mathbb{Z}_p^d \leq \iota(t\mathcal{O}_{K_p}) \leq \mathbb{Z}_p^d$ pour un certain $t \in K_p^\times$ alors $p^n\mathcal{O}_{K_p} \leq \mathcal{O}_p \leq \mathcal{O}_{K_p}$ car l'ordre local associé au paquet est par définition $\mathcal{O}_p = \{x \in K_p, \mathbb{Z}_p^d\phi(x) \subseteq \mathbb{Z}_p^d\}$. Mais le discriminant local est $D_p = [\mathcal{O}_p^* : \mathcal{O}_p]$ et $discK_p = [\mathcal{O}_{K_p}^* : \mathcal{O}_{K_p}]$ où $\mathcal{O}_p^*$ (resp. $\mathcal{O}_{K_p}^*$) est le dual de $\mathcal{O}_p$ via la forme trace ; par dualité $\mathcal{O}_{K_p}^* \leq \mathcal{O}_p^* \leq p^{-n}\mathcal{O}_{K_p}^*$ et donc $D_p \leq p^{2dn} discK_p$. Ceci donne donc (10) moyennant cette affirmation concernant la décroissance exponentielle de J_p en fonction de la distance de $N_0(\iota y)$ à la $K_p^\times$-orbite de N_{K_p}. Cette dernière assertion est obtenue en identifiant J_p à l'intégrale sur $K_p^\times/\mathbb{Q}_p$ d'un coefficient de matrice $\langle y \cdot F_1, F_2\rangle$ pour la représentation unitaire $L^2(\mathbb{P}^{d-1}(\mathbb{Q}_p))$ de $PGL_d(\mathbb{Q}_p)$ et en faisant usage des bornes exponentielles classiques sur ces coefficients via la fontion Ξ d'Harish-Chandra (cf. [**24**] §9.16).

7.4. Preuve du théorème 1.6

Soit μ_∞ une limite faible des mesures μ_n. Il nous suffit de montrer tout d'abord que μ_∞ est bien une probabilité, c'est-à-dire qu'il n'y a pas de fuite de masse, puis que μ_∞ a de l'entropie positive par rapport à au moins un sous-groupe à un paramètre de A. Le théorème sera alors une conséquence immédiate du théorème 1.1.

Ces deux propriétés vont résulter du corollaire 7.4 c'est-à-dire de l'équidistribution des transformées de Siegel. Soit f une approximation lisse à la fonction indicatrice $1_{B(0,\varepsilon)}$ de la boule de rayon ε centrée en 0 dans $\mathbb{R}^d$. Alors $\widetilde{f}(\Delta)$ domine la fonction indicatrice du voisinage Ω_{ε^d} de l'infini dans Ω défini au §3 et donné par le critère de Mahler, i.e. $\widetilde{f}(\Delta) \geq 1_{\Omega_{\varepsilon^d}}(\Delta)$. Le corollaire 7.4 implique donc que $\mu_\infty(\Omega_{\varepsilon^d}) = O(\varepsilon^d)$. En particulier μ_∞ est une probabilité, i.e. il n'y a pas de fuite de masse (ou en d'autres termes encore la famille $\{\mu_n\}_n$ est tendue). De façon similaire, si cette fois f est une approximation lisse à la fonction indicatrice $1_{B(v,\varepsilon)}$ de la boule euclidienne de rayon ε centrée en $v \neq 0$ dans $\mathbb{R}^d$ alors $\widetilde{f}(\Delta)$ domine la fonction indicatrice de la boule $B(\Delta, \varepsilon)$ (pour une métrique riemannienne quelconque fixée à l'avance) de rayon $O(\varepsilon)$ au voisinage de tout réseau Δ de Ω qui contient v. Ainsi $\mu_\infty(B(\Delta,\varepsilon)) = O(\varepsilon^d)$. La preuve du théorème est donc terminée à la lumière du fait classique suivant :

LEMME 7.5. — *Soit $\alpha > d-1$. Soit a un élément régulier de A. Soit μ une mesure A-invariante sur Ω telle que $\mu(B(\Delta,\varepsilon)) = O(\varepsilon^\alpha)$ uniformément pour Δ variant dans un compact de Ω. Alors $h_{\mu_\xi}(a) > 0$ pour presque toute composante ergodique μ_ξ de μ.*

Démonstration. — Vérifions d'abord que $h_\mu(a) > 0$. Ceci résulte de la définition de l'entropie (cf. §5.1 pour un rappel) et du fait que le système dynamique (Ω, T_a, μ) est partiellement hyperbolique (T_a est la translation par a) avec un sous-fibré tangent stable de dimension égale au rang de G soit $d-1$. En effet, si $\mathscr{P}$ est une partition fixe assez fine de Ω alors les atomes de $\mathscr{P}^m = \vee_{-m}^m T_a^k \mathscr{P}$ sont contenus dans des $e^{-\kappa m}$-voisinages de morceaux de taille fixe (i.e. de A-volume borné) de A-orbites, où $\kappa > 0$ mesure le coefficient de contraction/dilatation de T_a, coefficient qui dépend de la distance de a aux murs de la chambre de Weyl. Ces atomes $[x]_{\mathscr{P}^m}$ peuvent donc être recouverts par au plus $(e^{\kappa m})^{\dim A}$ boules de rayon $e^{-\kappa m}$. La condition $\mu(B(\Delta,\varepsilon)) = O(\varepsilon^\alpha)$ avec $\alpha > d-1$ entraîne alors $\mu([x]_{\mathscr{P}^m}) = O(e^{-(\alpha-(d-1))\kappa m})$ et donc $H_\mu(\mathscr{P}^m) \geq (\alpha-(d-1))\kappa m$, i.e. $h_\mu(T_a) \geq h_\mu(T_a, \mathscr{P}) \geq (\alpha-(d-1))\frac{\kappa}{2} > 0$.

Soit maintenant $\mu = \int_\Xi \mu_\xi d\nu(\xi)$ une décomposition ergodique de μ. Si $h_{\mu_\xi}(a) = 0$ pour tout $\xi \in \Xi_0$, un sous-ensemble de ν-mesure strictement positive, alors $\mu_0 := \frac{1}{\nu(\Xi_0)} \int_{\Xi_0} \mu_\xi d\nu(\xi)$ vérifie $h_{\mu_0}(a) = 0$ d'après le point 4) du §5.1. Mais μ domine $\nu(\Xi_0) \cdot \mu_0$, en particulier $\mu_0(B(\Delta,\varepsilon)) \leq \frac{1}{\nu(\Xi_0)}\mu(B(\Delta,\varepsilon)) = O_\Delta(\varepsilon^\alpha)$ avec $\alpha > d-1$. Comme précédemment ceci entraîne $h_{\mu_0}(a) > 0$, une contradiction. □

REMARQUE. — *Notons que l'équidistribution dans le théorème 1.6 ou dans la conjecture 1.3 pourrait être le reflet de deux phénomènes a priori distincts : ou bien Y_n est constitué de nombreuses orbites de volume assez petit qui sont individuellement mal équidistribuées mais qui néanmoins sont placées dans $\Omega = \Gamma\backslash G$ de façon homogène et bien répartie, ou bien il se pourrait qu'au contraire le nombres d'orbites dans Y_n est moindre et la plupart des orbites de Y_n sont déjà individuellement bien équidistribuées.*

Ce dernier phénomène est apparemment difficile à mettre en évidence : peut-on trouver par exemple une suite de A-orbites compactes dont la mesure A-invariante tend vers la mesure de Haar de Ω ? Pour $d = 2$ il faudrait trouver une infinité de discriminants $d > 0$ tels que le nombre de classes $h(d)$ soit borné. C'est un problème ouvert bien connu. Les résultats de Harcos et Michel **[35]** *vont cependant dans cette direction : ils obtiennent l'équidistribution – sparse equidistribution – pour des paquets maigres de géodésiques (i.e. correspondant à l'orbite d'un sous-groupe strict assez gros du groupe de classes) comme conséquence de leur borne de sous-convexité pour les fonctions L associées à des formes de Maass paraboliques tordues par des caractères du groupe de classes (fonctions L de Rankin-Selberg).*

8. DEUX APPLICATIONS

8.1. Équidistribution des matrices entières dans une classe de similitude

Soit $P \in \mathbb{Z}[T]$ unitaire de degré d, irréductible sur $\mathbb{Q}$ et avec d racines réelles. Soit $\mathcal{V}_P$ la variété des matrices carrées de taille d dont le polynôme caractéristique est égal à P. Clairement $\mathcal{V}_P(\mathbb{R})$ est une classe de similitude, i.e. $G = PGL_d(\mathbb{R})$ agit transitivement. Le stabilisateur d'un point est $N(A_P)$ où A_P est le centralisateur d'une matrice M_P de $\mathcal{V}_P$. Comme P a d racines réelles, A_P est un tore maximal et $\mathbb{R}$-déployé, il est donc conjugué à A, i.e. $A = h_P A_P h_P^{-1}$. On a donc une application naturelle G-équivariante et surjective de $X = G/A$ sur $\mathcal{V}_P(\mathbb{R})$ définie par $\iota_P : g \mapsto g h_P M_P h_P^{-1} g^{-1}$. Cette application est indépendante du choix de M_P.

On peut alors tirer en arrière les points entiers de $\mathcal{V}_P$ par ι_P et obtenir un ensemble de points entiers $X(P) = \iota_P^{-1}(M_d(\mathbb{Z}) \cap \mathcal{V}_P)$ sur X. Notons que X possède une mesure naturelle à normalisation près : la mesure $PGL_d(\mathbb{R})$-invariante, notée ν. On obtient :

THÉORÈME 8.1 (**[24]** Corollary 3.3). — *Fixons $d = 3$. Soit $(P_n)_{n\in\mathbb{N}}$ une suite infinie de polynômes comme ci-dessus. Alors la famille $\{X(P_n)\}_{n\in\mathbb{N}}$ devient équidistribuée dans X quand n tend vers l'infini, i.e. on a le théorème limite quotient suivant :*

$$\lim_{n\to+\infty} \frac{|\mathcal{U}_1 \cap X(P_n)|}{|\mathcal{U}_2 \cap X(P_n)|} = \frac{\nu(\mathcal{U}_1)}{\nu(\mathcal{U}_2)}$$

pour tous ouverts bornés $\mathcal{U}_1$ et $\mathcal{U}_2$ de frontière de ν-mesure négligeable.

Cet énoncé est à l'origine un cas particulier d'une question de Linnik (voir **[66]**). Le lien entre un tel énoncé et un énoncé d'équidistribution de périodes (tel le théorème 1.6) a été dégagé dans **[29]** où Eskin et Oh répondent à la question de Linnik sur l'asymptotique du nombre de points entiers sur une suite de variétés homogènes de la forme $V_m = \{f = m\}$, où f est un polynôme homogène G-invariant et où on compte les points dans un cône fixe de l'espace. Les variétés considérées sont isomorphes

à G/H avec H engendré par des unipotents pour que la théorie de Ratner puisse s'appliquer. Voir aussi l'article antérieur [**30**].

C'est un principe désormais classique en théorie ergodique des actions de groupes qu'à tout énoncé d'ergodicité ou d'équidistribution de H_1 sur G/H_2 correspond un énoncé analogue pour H_2 sur $H_1\backslash G$. Ici $\Gamma = PGL_d(\mathbb{Z})$ agit sur $X(P)$ par conjugaison, ce qui revient à la translation à gauche sur G/A via l'application ι_P. À chaque Γ-orbite d'une matrice M_0 dans $X(P)$ correspond ainsi une A-orbite de $\Gamma\backslash G$ qui est compacte car l'irréductibilité de P garantit la $\mathbb{Q}$-anisotropie du centralisateur A_0 de M_0. On vérifie enfin que l'ensemble $X(P)$ est une réunion finie de Γ-orbites à laquelle correspond par cette correspondance la réunion (finie) des paquets de A-orbites compactes Y_L associées aux réseaux L du corps de nombres totalement réel $K = K_P = \mathbb{Q}[T]/(P)$ dont l'ordre associé $\mathcal{O}_L$ contient l'anneau quotient $\mathcal{O}_P = \mathbb{Z}[T]/(P)$.

8.2. Amélioration du théorème de Minkowski

Pour démontrer la finitude du groupe des classes d'un corps de nombres K de degré d, on commence en général d'abord par établir la borne de Minkowski selon laquelle toute classe d'idéaux fractionnaires possède un représentant I dont la norme est un $O_d(\mathrm{disc}(K))$. Les auteurs de [**22**] conjecturent que ce $O_d(\cdot)$ peut être remplacé par un $o_d(\cdot)$. La proposition 3.1 (appliquée à l'idéal inverse I^{-1}) combinée au théorème 1.2 donne immédiatement l'approximation suivante à cette conjecture :

THÉORÈME 8.2 ([**22**] Theorem 6.3). — *Pour tous $\varepsilon, \delta > 0$, le nombre de corps de nombres K totalement réels de degré d et de discriminant $\mathrm{disc}(K) \leq X$ tels qu'au moins une classe d'idéaux ne puisse pas être représentée par un idéal fractionnaire I de norme au plus $\delta \cdot \mathrm{disc}(K)$ est un $O_{\delta,d,\varepsilon}(X^{\varepsilon})$.*

En revanche, comme le montre déjà l'exemple de Cassels au paragraphe 3.1, le nombre de corps de nombres totalement réels de degré d et de discriminant $\mathrm{disc}(K) \leq X$ est au moins polynomial en X.

Remerciements

Je remercie vivement Philippe Michel pour son soutien stratégique durant la préparation de cet exposé et pour avoir si patiemment répondu à mes questions. Uri Shapira et Elon Lindenstrauss m'ont aussi apporté une aide cruciale en me faisant part de leurs dernières réflexions sur le sujet, je leur suis particulièrement reconnaissant. Je remercie François Ledrappier, Hee Oh et Georges Tomanov pour leurs commentaires pertinents qui m'ont permis d'améliorer le texte. Je remercie aussi Emmanuel Ullmo et Laurent Clozel et les autres participants du groupe de travail organisé à Orsay en

2008 sur ce sujet pour les discussions que nous avons eues à propos des articles [**22**] et [**24**]. Merci aussi à Françoise Dal'bo pour ses remarques constructives.

RÉFÉRENCES

[1] Y. Benoist & H. Oh – Equidistribution of rational matrices in their conjugacy classes, *Geom. Funct. Anal.* **17** (2007), p. 1–32.

[2] ———, Effective equidistribution of S-integral points on symmetric varieties, prépublication arXiv:0706.1621.

[3] N. Bergeron – Spectre des surfaces hyperboliques, livre en préparation.

[4] V. Blomer, G. Harcos & P. Michel – Bounds for modular L-functions in the level aspect, *Ann. Sci. École Norm. Sup.* **40** (2007), p. 697–740.

[5] A. Borel – *Introduction aux groupes arithmétiques*, Publications de l'Institut de Mathématique de l'Université de Strasbourg, XV. Actualités Scientifiques et Industrielles, No. 1341, Hermann, 1969.

[6] R. Bowen – Periodic orbits for hyperbolic flows, *Amer. J. Math.* **94** (1972), p. 1–30.

[7] D. A. Burgess – On character sums and L-series. II, *Proc. London Math. Soc.* **13** (1963), p. 524–536.

[8] J. W. S. Cassels – The product of n inhomogeneous linear forms, *J. London Math. Soc.* (1952).

[9] J. W. S. Cassels & H. P. F. Swinnerton-Dyer – On the product of three homogeneous linear forms and the indefinite ternary quadratic forms, *Philos. Trans. Roy. Soc. London. Ser. A.* **248** (1955), p. 73–96.

[10] L. Clozel & E. Ullmo – Équidistribution de mesures algébriques, *Compos. Math.* **141** (2005), p. 1255–1309.

[11] ———, Équidistribution de sous-variétés spéciales, *Ann. of Math.* **161** (2005), p. 1571–1588.

[12] P. B. Cohen – Hyperbolic equidistribution problems on Siegel 3-folds and Hilbert modular varieties, *Duke Math. J.* **129** (2005), p. 87–127.

[13] S. G. Dani & G. A. Margulis – Limit distributions of orbits of unipotent flows and values of quadratic forms, in *I. M. Gel'fand Seminar*, Adv. Soviet Math., vol. 16, Amer. Math. Soc., 1993, p. 91–137.

[14] W. Duke – Hyperbolic distribution problems and half-integral weight Maass forms, *Invent. Math.* **92** (1988), p. 73–90.

[15] W. Duke, J. B. Friedlander & H. Iwaniec – Bounds for automorphic L-functions, *Invent. Math.* **112** (1993), p. 1–8.

[16] ———, The subconvexity problem for Artin L-functions, *Invent. Math.* **149** (2002), p. 489–577.

[17] M. Einsiedler – A simplified proof of Ratner's measure classification for the action of subgroups isomorphic to $SL(2,\mathbb{R})$, prépublication, 2008.

[18] M. Einsiedler & A. Katok – Invariant measures on G/Γ for split simple Lie groups G, *Comm. Pure Appl. Math.* **56** (2003), p. 1184–1221.

[19] ———, Rigidity of measures—the high entropy case and non-commuting foliations, *Israel J. Math.* **148** (2005), p. 169–238.

[20] M. Einsiedler, A. Katok & E. Lindenstrauss – Invariant measures and the set of exceptions to Littlewood's conjecture, *Ann. of Math.* **164** (2006), p. 513–560.

[21] M. Einsiedler & E. Lindenstrauss – Diagonalizable flows on locally homogeneous spaces and number theory, in *Proceedings of the International Congress of Mathematicians. Vol. II*, Eur. Math. Soc., Zürich, 2006, p. 1731–1759.

[22] M. Einsiedler, E. Lindenstrauss, P. Michel & A. Venkatesh – Distribution of periodic torus orbits on homogeneous spaces I, *Duke Math. J.* **148** (2009), p. 119–174.

[23] ———, Distribution of periodic torus orbits on homogeneous spaces II : Duke's theorem for quadratic fields, en préparation.

[24] ———, Distribution of periodic torus orbits on homogeneous spaces III : Duke's theorem for cubic fields, preprint, arXiv:math.0708.1113.

[25] M. Einsiedler & T. Ward – Ergodic theory with a view towards number theory, livre à paraître dans la collection *Springer Graduate Texts in Math.*

[26] A. Eskin – Counting problems and semisimple groups, in *Proceedings of the International Congress of Mathematicians, Vol. II (Berlin, 1998)*, 1998, p. 539–552.

[27] A. Eskin & C. McMullen – Mixing, counting, and equidistribution in Lie groups, *Duke Math. J.* **71** (1993), p. 181–209.

[28] A. Eskin, S. Mozes & N. A. Shah – Unipotent flows and counting lattice points on homogeneous varieties, *Ann. of Math.* **143** (1996), p. 253–299.

[29] A. Eskin & H. Oh – Representations of integers by an invariant polynomial and unipotent flows, *Duke Math. J.* **135** (2006), p. 481–506.

[30] W. T. Gan & H. Oh – Equidistribution of integer points on a family of homogeneous varieties : a problem of Linnik, *Compositio Math.* **136** (2003), p. 323–352.

[31] É. Ghys – Dynamique des flots unipotents sur les espaces homogènes, Séminaire Bourbaki, vol. 1991/92, exposé n° 747, *Astérisque* **206** (1992), p. 93–136.

[32] A. Gorodnik, F. Maucourant & H. Oh – Manin's and Peyre's conjectures on rational points and adelic mixing, *Ann. Sci. Éc. Norm. Supér.* **41** (2008), p. 383–435.

[33] A. Gorodnik & H. Oh – Rational points on homogeneous varieties and equidistribution of adelic periods, prépublication arXiv:0803.1996.

[34] G. Harcos – Equidistribution and L-functions, Lecture Notes Clay Math. School, Pisa 2007.

[35] G. Harcos & P. Michel – The subconvexity problem for Rankin-Selberg L-functions and equidistribution of Heegner points. II, *Invent. Math.* **163** (2006), p. 581–655.

[36] B. Host – Nombres normaux, entropie, translations, *Israel J. Math.* **91** (1995), p. 419–428.

[37] A. Katok & R. J. Spatzier – Invariant measures for higher-rank hyperbolic abelian actions, *Ergodic Theory Dynam. Systems* **16** (1996), p. 751–778.

[38] B. Klingler & A. Yafaev – On the André-Oort conjecture, prépublication, 2006.

[39] S. Lang – *Algebraic number theory*, 2e éd., Graduate Texts in Math., vol. 110, Springer, 1994.

[40] F. Ledrappier & L.-S. Young – The metric entropy of diffeomorphisms. I and II, *Ann. of Math.* **122** (1985), p. 509–574.

[41] E. Lindenstrauss – Rigidity of multiparameter actions, *Israel J. Math.* **149** (2005), p. 199–226.

[42] ———, Invariant measures and arithmetic quantum unique ergodicity, *Ann. of Math.* **163** (2006), p. 165–219.

[43] E. Lindenstrauss & B. Weiss – On sets invariant under the action of the diagonal group, *Ergodic Theory Dynam. Systems* **21** (2001), p. 1481–1500.

[44] Y. Linnik – *Ergodic properties of number fields*, Springer, 1970.

[45] G. A. Margulis – Oppenheim conjecture, in *Fields Medallists' lectures*, World Sci. Ser. 20th Century Math., vol. 5, World Sci. Publ., River Edge, NJ, 1997, p. 272–327.

[46] ———, Problems and conjectures in rigidity theory, in *Mathematics : frontiers and perspectives*, Amer. Math. Soc., 2000, p. 161–174.

[47] ———, *On some aspects of the theory of Anosov systems*, Springer Monographs in Math., Springer, 2004.

[48] G. A. Margulis & G. M. Tomanov – Invariant measures for actions of unipotent groups over local fields on homogeneous spaces, *Invent. Math.* **116** (1994), p. 347–392.

[49] F. Maucourant – A non-homogeneous orbit closure of a diagonal subgroup, prépublication arXiv:0707.2920, à paraître dans *Annals of Math.*

[50] P. Michel & A. Venkatesh – Equidistribution, L-functions and ergodic theory : on some problems of Yu. Linnik, in *Proceedings of the International Congress of Mathematicians. Vol. II*, Eur. Math. Soc., Zürich, 2006, p. 421–457.

[51] ———, Subconvexity for GL_2, prépublication, 2009.

[52] S. Mozes – Actions of Cartan subgroups, *Israel J. Math.* **90** (1995), p. 253–294.

[53] J. Neukirch – *Algebraic number theory*, Grund. Math. Wiss., vol. 322, Springer, 1999.

[54] H. Oh – Finiteness of compact maximal flats of bounded volume, *Ergodic Theory Dynam. Systems* **24** (2004), p. 217–225.

[55] ———, Orbital counting via mixing and unipotent flows, in *Scuola Normale Superiore Summer school, Pisa 2007*, Clay Math. Proceedings, vol. 8, 2008.

[56] V. Platonov & A. Rapinchuk – *Algebraic groups and number theory*, Pure and Applied Mathematics, vol. 139, Academic Press Inc., 1994.

[57] A. A. Popa – Central values of Rankin L-series over real quadratic fields, *Compos. Math.* **142** (2006), p. 811–866.

[58] M. S. Raghunathan – *Discrete subgroups of Lie groups*, Ergebnisse Math. Grenzg., vol. 68, Springer, 1972.

[59] M. Ratner – Factors of horocycle flows, *Ergodic Theory Dynam. Systems* **2** (1982), p. 465–489.

[60] ———, Rigidity of horocycle flows, *Ann. of Math.* **115** (1982), p. 597–614.

[61] ———, Horocycle flows, joinings and rigidity of products, *Ann. of Math.* **118** (1983), p. 277–313.

[62] ———, Distribution rigidity for unipotent actions on homogeneous spaces, *Bull. Amer. Math. Soc.* **24** (1991), p. 321–325.

[63] ———, On Raghunathan's measure conjecture, *Ann. of Math.* **134** (1991), p. 545–607.

[64] ———, Raghunathan's topological conjecture and distributions of unipotent flows, *Duke Math. J.* **63** (1991), p. 235–280.

[65] D. J. Rudolph – $\times 2$ and $\times 3$ invariant measures and entropy, *Ergodic Theory Dynam. Systems* **10** (1990), p. 395–406.

[66] P. C. Sarnak – Diophantine problems and linear groups, in *Proceedings of the International Congress of Mathematicians, Vol. I, II (Kyoto, 1990)*, Math. Soc. Japan, 1991, p. 459–471.

[67] C. Series – On coding geodesics with continued fractions, in *Ergodic theory (Sem., Les Plans-sur-Bex, 1980)*, Monograph. Enseign. Math., vol. 29, Univ. Genève, 1981, p. 67–76.

[68] N. A. Shah – Uniformly distributed orbits of certain flows on homogeneous spaces, *Math. Ann.* **289** (1991), p. 315–334.

[69] U. Shapira – A solution to a problem of Cassels and Diophantine properties of cubic fields, prépublication, 2008.

[70] G. M. Tomanov – Values of decomposable forms at S-integral points and orbits of tori on homogeneous spaces, *Duke Math. J.* **138** (2007), p. 533–562.

[71] G. M. Tomanov & B. Weiss – Closed orbits for actions of maximal tori on homogeneous spaces, *Duke Math. J.* **119** (2003), p. 367–392.

[72] E. Ullmo & A. Yafaev – Galois orbits and equidistribution of special subvarieties : towards the André-Oort conjecture, prépublication, 2006.

[73] A. Venkatesh – The work of Einsiedler, Katok and Lindenstrauss on the Littlewood conjecture, *Bulletin of the A.M.S.* **45** (2008), p. 117–134 (électronique).

[74] ———, Sparse equidistribution problems, period bounds, and subconvexity, prépublication arXiv:math.NT/0506224.

[75] J.-L. Waldspurger – Sur les valeurs de certaines fonctions L automorphes en leur centre de symétrie, *Compositio Math.* **54** (1985), p. 173–242.

[76] F. Wielonsky – Séries d'Eisenstein, intégrales toroïdales et une formule de Hecke, *Enseign. Math.* **31** (1985), p. 93–135.

[77] S. Zelditch – Selberg trace formulae and equidistribution theorems for closed geodesics and Laplace eigenfunctions : finite area surfaces, *Mem. Amer. Math. Soc.* **96** (1992).

[78] S.-W. Zhang – Equidistribution of CM-points on quaternion Shimura varieties, *Int. Math. Res. Not.* **2005** (2005), p. 3657–3689.

Emmanuel BREUILLARD

Laboratoire de Mathématiques
Université Paris XI
Bâtiment 425
F–91405 Orsay cedex
E-mail : breuillard@math.u-psud.fr

332

ASTÉRISQUE

2010

SÉMINAIRE BOURBAKI
VOLUME 2008/2009
EXPOSÉS 997-1011

(1009) *Regularity of optimal transport maps*

Alessio FIGALLI

SOCIÉTÉ MATHÉMATIQUE DE FRANCE
Publié avec le concours du CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

Séminaire BOURBAKI Juin 2009
61e année, 2008-2009, no 1009, p. 341 à 368

REGULARITY OF OPTIMAL TRANSPORT MAPS [after Ma–Trudinger–Wang and Loeper]

by **Alessio FIGALLI**

INTRODUCTION

In the field of optimal transportation, one important issue is the regularity of the optimal transport map. There are several motivations for the investigation of the smoothness of the optimal map:

- It is a typical PDE/analysis question.
- It is a step towards a qualitative understanding of the optimal transport map.
- If it is a general phenomenon, then non-smooth situations may be treated by regularization, instead of working directly on non-smooth objects.

In the special case "cost = squared distance" on $\mathbb{R}^n$, the problem was solved by Caffarelli [**4, 5, 6, 7**], who proved the smoothness of the map under suitable assumptions on the regularity of the densities and on the geometry of their support. However, a major open problem in the theory was the question of regularity for more general cost functions, or for the case "cost = squared distance" on a Riemannian manifold. A breakthrough in this problem has been achieved by Ma, Trudinger and Wang [**27**] and Loeper [**24**], who found a necessary and sufficient condition on the cost function in order to ensure regularity. This condition, now called MTW condition, involves a combination of derivatives of the cost, up to the fourth order. In the special case "cost = squared distance" on a Riemannian manifold, the MTW condition corresponds to the non-negativity of a new curvature tensor on the manifold (the so-called MTW tensor), which implies strong geometric consequences on the geometry of the manifold and on the structure of its cut-locus.

1. THE OPTIMAL TRANSPORTATION PROBLEM

The Monge transportation problem is more than 200 years old **[29]**, and it has generated a huge amount of work in the last years.

Originally Monge wanted to move, in the Euclidean space $\mathbb{R}^3$, a rubble (*déblais*) to build up a mound or fortification (*remblais*) minimizing the cost. To explain this in a simple case, suppose that the rubble consists of masses, say $m_1, \dots, m_n$, at locations $\{x_1, \dots x_n\}$, and one is interested in moving them into another set of positions $\{y_1, \dots, y_n\}$ by minimizing the weighted travelled distance. Then, one tries to minimize

$$\sum_{i=1}^{n} m_i |x_i - T(x_i)|,$$

over all bijections $T : \{x_1, \dots x_n\} \to \{y_1, \dots, y_n\}$.

Nowadays, influenced by physics and geometry, one would be more interested in minimizing the energy cost rather than the distance. Therefore, one wants to minimize

$$\sum_{i=1}^{n} m_i |x_i - T(x_i)|^2.$$

Of course, it is desirable to generalize this problem to continuous, rather than just discrete, distributions of matter. Hence, the optimal transport problem is now formulated in the following general form: given two probability measures μ and ν, defined on the measurable spaces X and Y, find a measurable map $T : X \to Y$ with $T_\sharp \mu = \nu$, i.e.

$$\nu(A) = \mu\left(T^{-1}(A)\right) \qquad \forall\, A \subset Y \text{ measurable},$$

in such a way that T minimizes the transportation cost. This means

$$\int_X c(x, T(x))\, d\mu(x) = \min_{S_\# \mu = \nu} \left\{ \int_X c(x, S(x))\, d\mu(x) \right\},$$

where $c : X \times Y \to \mathbb{R}$ is some given cost function, and the minimum is taken over all measurable maps $S : X \to Y$ such that $S_\# \mu = \nu$. When the transport condition $T_\# \mu = \nu$ is satisfied, we say that T is a *transport map*, and if T also minimizes the cost we call it an *optimal transport map*.

Even in Euclidean spaces, with the cost c equal to the Euclidean distance or its square, the problem of the existence of an optimal transport map is far from being trivial. Moreover, it is easy to build examples where the Monge problem is ill-posed simply because there is no transport map: this happens for instance when μ is a Dirac mass while ν is not. This means that one needs some restrictions on the measures μ and ν.

We further remark that, if $\mu(dx) = f(x)dx$ and $\nu(dy) = g(y)dy$, the condition $T_\# \mu = \nu$ formally gives the Jacobian equation $\left|\det(\nabla T)\right| = f/(g \circ T)$.

1.1. Existence and uniqueness of optimal maps on Riemannian manifolds

In [**1, 2**], Brenier considered the case $X = Y = \mathbb{R}^n$, $c(x,y) = |x-y|^2/2$, and he proved the following theorem (the same result was also proven independently by Cuesta-Albertos and Matrán [**10**] and by Rachev and Rüschendorf [**30**]):

THEOREM 1.1 ([**1, 2**]). — *Let μ and ν be two compactly supported probability measures on $\mathbb{R}^n$. If μ is absolutely continuous with respect to the Lebesgue measure, then:*

(i) *There exists a unique solution T to the Monge problem.*
(ii) *The optimal map T is characterized by the structure $T(x) = \nabla\phi(x)$, for some convex function $\phi : \mathbb{R}^n \to \mathbb{R}$.*

Furthermore, if $\mu(dx) = f(x)dx$ and $\nu(dy) = g(y)dy$,

$$\big|\det(\nabla T(x))\big| = \frac{f(x)}{g(T(x))} \qquad \text{for } \mu\text{-a.e. } x \in \mathbb{R}^n.$$

After this result, many researchers started to work on the problem, showing existence of optimal maps with more general costs, both in an Euclidean setting, in the case of compact (Riemannian and sub-Riemannian) manifolds, and in some particular classes on non-compact manifolds. In particular, exploiting some ideas introduced by Cabré in [**3**] for studying elliptic equations on manifolds, McCann was able to generalize Brenier's theorem to (compact) Riemannian manifolds [**28**].

REMARK. — *From now on, we will always implicitly assume that all manifolds have no boundary.*

To explain McCann's result, let us first introduce a few definitions.

We recall that a function $\varphi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is convex and lower semicontinuous if and only if

$$\varphi(x) = \sup_{y \in \mathbb{R}^n} \big[x \cdot y - \varphi^*(y)\big],$$

where

$$\varphi^*(x) := \sup_{x \in \mathbb{R}^n} \big[x \cdot y - \varphi(x)\big].$$

This fact is the basis for the notion of *c-convexity*, where $c : X \times Y \to \mathbb{R}$ is an arbitrary function:

DEFINITION 1.2. — *A function $\psi : X \to \mathbb{R} \cup \{+\infty\}$ is* c-convex *if*

$$\psi(x) = \sup_{y \in Y} \big[\psi^c(y) - c(x,y)\big] \qquad \forall\, x \in X,$$

where

$$\psi^c(y) := \inf_{x \in X} \big[\psi(x) + c(x,y)\big] \qquad \forall\, y \in Y.$$

Moreover, for a c-convex function ψ, we define its c-subdifferential *at x as*

$$\partial^c\psi(x) := \big\{y \in Y \,|\, \psi(x) = \psi^c(y) - c(x,y)\big\}.$$

With this general definition, when $X = Y = \mathbb{R}^n$ and $c(x,y) = -x \cdot y$, the usual convexity coincides with the c-convexity, and the usual subdifferential coincides with the c-subdifferential.

In particular, in the case $X = Y = \mathbb{R}^n$ and $c(x,y) = |x-y|^2/2$, a function ψ is c-convex if and only if $\psi(x) + \frac{|x|^2}{2}$ is convex. The following result is the generalization of Brenier's Theorem to Riemannian manifolds:

THEOREM 1.3 ([**28**]). — *Let (M,g) be a Riemannian manifold, take μ and ν two compactly supported probability measures on M, and consider the optimal transport problem from μ to ν with cost $c(x,y) = d(x,y)^2/2$, where $d(x,y)$ denotes the Riemannian distance on M. If μ is absolutely continuous with respect to the volume measure, then:*

(i) *There exists a unique solution T to the Monge problem.*
(ii) *T is characterized by the structure $T(x) = \exp_x\big(\nabla\psi(x)\big) \in \partial^c\psi(x)$ for some c-convex function $\psi : M \to \mathbb{R}$.*
(iii) *For μ_0-a.e. $x \in M$, there exists a unique minimizing geodesic from x to $T(x)$, which is given by $[0,1] \ni t \mapsto \exp_x\big(t\nabla\psi(x)\big)$.*

Furthermore, if $\mu(dx) = f(x)\mathrm{vol}(dx)$ and $\nu(dy) = g(y)\mathrm{vol}(dy)$,

$$\big|\det(\nabla T(x))\big| = \frac{f(x)}{g(T(x))} \qquad \text{for } \mu\text{-a.e. } x \in M.$$

The last formula in the above theorem needs a comment: given a function $T : M \to M$, the determinant of its Jacobian is not intrinsically defined. Indeed, in order to compute the determinant of $\nabla T(x) : T_xM \to T_{T(x)}M$, one needs to identify the tangent spaces. On the other hand, $\big|\det(\nabla T(x))\big|$ is intrinsically defined as

$$\big|\det(\nabla T(x))\big| = \lim_{r\to 0} \frac{\mathrm{vol}(T(B_r(x)))}{\mathrm{vol}(B_r(x))},$$

whenever the above limit exists.

2. THE REGULARITY ISSUE: THE EUCLIDEAN CASE

Let Ω and Ω' be two bounded smooth open sets in $\mathbb{R}^n$, and let $\mu(dx) = f(x)dx$, $\nu(y) = g(y)dy$ be two probability measures, with f and g such that $f = 0$ in $\mathbb{R}^2 \setminus \Omega$, $g = 0$ in $\mathbb{R}^2 \setminus \Omega'$. We assume that f and g are C^∞ and bounded away from zero and infinity on Ω and Ω', respectively. By Brenier's Theorem, the optimal transport

map T is given by the gradient of a convex function ϕ. Hence, at least formally, the Jacobian equation for T

$$\big|\det(\nabla T(x))\big| = \frac{f(x)}{g(T(x))},$$

gives a PDE for ϕ:

$$\det(D^2\phi(x)) = \frac{f(x)}{g(\nabla\phi(x))}. \tag{1}$$

This is a Monge-Ampère equation for ϕ, which is naturally coupled with the boundary condition

$$\nabla\phi(\Omega) = \Omega' \tag{2}$$

(which corresponds to the fact that T transports $f(x)dx$ onto $g(y)dy$).

As observed by Caffarelli [**6**], even for smooth densities, one cannot expect any general regularity result for ϕ without making some geometric assumptions on the support of the target measure. Indeed, suppose that $\Omega = B_1$ is the unit ball centered at the origin, and $\Omega' = \big(B_1^+ + e_n\big) \cup \big(B_1^- - e_n\big)$ is the union of two half-balls, where $(e_i)_{i=1,\dots,n}$ denote the canonical basis of $\mathbb{R}^n$, and

$$B_1^+ := \big(B_1 \cap \{x_n > 0\}\big), \qquad B_1^- := \big(B_1 \cap \{x_n < 0\}\big).$$

Then, if $f = g = \frac{1}{|B_1|}$ on Ω and Ω' respectively, it is easily seen that the optimal map T is given by

$$T(x) := \begin{cases} x + e_n & \text{if } x_n > 0, \\ x - e_n & \text{if } x_n < 0, \end{cases}$$

which corresponds to the gradient of the convex function $\phi(x) = |x_n| + |x|^2/2$.

Thus, as one could also show by an easy topological argument, in order to hope for a regularity result for ϕ, we need at least to assume the connectedness of Ω'. But, starting from the above construction and considering a sequence of domains Ω'_ε where one adds a small strip of width $\varepsilon > 0$ to glue together $\big(B_1^+ + e_n\big) \cup \big(B_1^- - e_n\big)$, one can also show that for $\varepsilon > 0$ small enough the optimal map will still be discontinuous (see [**6**]).

As proven by Caffarelli [**6**], the right geometric condition on Ω' which allows to prevent singularities of ϕ and to show the regularity of the optimal transport map is the convexity of the target: if Ω' is convex, and f and g are C^∞ and strictly positive on their respective support, then ϕ (and hence T) is C^∞ inside Ω [**4, 5, 6**]. Moreover, if one further assumes that both Ω and Ω' are smooth and uniformly convex, then $\phi \in C^\infty(\overline{\Omega})$, and $T : \overline{\Omega} \to \overline{\Omega'}$ is a smooth diffeomorphism [**7**] (the same result has been proven independently by Urbas [**33**]).

3. THE REGULARITY ISSUE: THE RIEMANNIAN CASE

The extension of Caffarelli's regularity theory to more general cost function or to the case of the squared distance function on Riemannian manifolds was for a long time a serious issue, not clear how to attack. To keep the exposition easier, we will focus on the case of the squared distance on Riemannian manifolds, although most of the arguments are exactly the same for a more general cost function. In what follows, we will use "smooth" as a synonymous of C^∞.

3.1. A PDE approach to the regularity issue

Let (M,g) be a (smooth) compact connected Riemannian manifold, let $\mu(dx) = f(x)\mathrm{vol}(dx)$ and $\nu(dy) = g(y)\mathrm{vol}(dy)$ be probability measures on M, and consider the cost $c(x,y) = d(x,y)^2/2$. Assume f and g to be C^∞ and strictly positive on M.

As before, we start from the Jacobian equation

$$\big|\det(\nabla T(x))\big| = \frac{f(x)}{g(T(x))}$$

to formally obtain an equation for ψ. It can be shown, by standard arguments of Riemannian geometry, that the relation $T(x) = \exp_x\big(\nabla\psi(x)\big)$ is equivalent to

$$\nabla\psi(x) + \nabla_x c(x, T(x)) = 0. \tag{3}$$

Writing everything in charts, we differentiate the above identity with respect to x, and by using the Jacobian equation we get

$$\begin{aligned}\det\big(D^2\psi(x) + D^2_x c\big(x, \exp_x\big(\nabla\psi(x)\big)\big)\big) &= \frac{f(x)\mathrm{vol}_x}{g(T(x))\mathrm{vol}_{T(x)}\big|\det(d_{\nabla\psi(x)}\exp_x)\big|}\\ &=: h(x, \nabla\psi(x)),\end{aligned} \tag{4}$$

where vol_z denotes the volume density at a point $z \in M$ computed with respect to the chart. (Because ψ is c-convex (cf. Theorem 1.3(ii)), the matrix $D^2\psi(x) + D^2_x c\big(x, \exp_x\big(\nabla\psi(x)\big)\big)$ is non-negative.) Hence ψ solves a Monge-Ampère type equation with a perturbation term $D^2_x c\big(x, \exp_x\big(\nabla\psi(x)\big)\big)$, which is of first order in ψ. Unfortunately, for Monge-Ampère type equations lower order terms do matter, and it turns out that it is exactly the term $D^2_x c\big(x, \exp_x\big(\nabla\psi(x)\big)\big)$ which can create obstructions to the smoothness.

The breakthrough in this problem came with the paper of Ma, Trudinger and Wang **[27]** (whose roots lie in an earlier work of Wang on the reflector antenna problem **[35]**), where the authors found a mysterious fourth-order condition on the cost functions, which turned out to be sufficient to prove the regularity of ψ. The idea was to differentiate twice Equation (4) in order to get a linear PDE for the second derivatives

of ψ, and then to try to show an a priori estimate on the second derivatives of ψ. In this computation, one ends up at a certain moment with a term which needs to have a sign in order to conclude the desired a priori estimate. This term is what is now called the Ma–Trudinger–Wang tensor (in short MTW tensor):

$$\mathfrak{S}_{(x,y)}(\xi,\eta) := \frac{3}{2}\sum_{ijklrs} \left(c_{ij,r}c^{r,s}c_{s,kl} - c_{ij,kl}\right)\xi^i\xi^j\eta^k\eta^l, \qquad \xi \in T_xM,\ \eta \in T_yM. \tag{5}$$

In the above formula the cost function is evaluated at (x,y), and we used the notation $c_j = \frac{\partial c}{\partial x^j}$, $c_{jk} = \frac{\partial^2 c}{\partial x^j \partial x^k}$, $c_{i,j} = \frac{\partial^2 c}{\partial x^i \partial y^j}$, $c^{i,j} = (c_{i,j})^{-1}$, and so on. Moreover, all the derivatives are computed by introducing a system of coordinates $(x^1,\dots,x^n)$ around x, and a system $(y^1,\dots,y^n)$ around y. (We will discuss later on the independence of this expression on the choice of the system of coordinates, see Paragraph 3.4.) The condition to impose on $\mathfrak{S}_{(x,y)}(\xi,\eta)$ is

$$\mathfrak{S}_{(x,y)}(\xi,\eta) \geq 0 \qquad \text{whenever } \sum_{ij} c_{i,j}\xi^i\eta^j = 0$$

(this is called the MTW condition). Under this hypothesis, and a geometric condition on the supports of the measures (which is the analogous of the convexity assumption of Caffarelli), Ma, Trudinger and Wang could prove the following result:

THEOREM 3.1 ([**27, 31, 32**]). — *Let (M,g) be a compact Riemannian manifold. Assume that the* MTW *condition holds, that f and g are smooth and bounded away from zero and infinity on their respective supports Ω and Ω', and that the cost function $c = d^2/2$ is smooth on the set $\overline{\Omega} \times \overline{\Omega'}$. Finally, suppose that:*

(a) *Ω and Ω' are smooth;*
(b) *$(\exp_x)^{-1}(\Omega') \subset T_xM$ is uniformly convex for all $x \in \Omega$;*
(c) *$(\exp_y)^{-1}(\Omega) \subset T_xM$ is uniformly convex for all $y \in \Omega'$.*

Then $\psi \in C^\infty(\overline{\Omega})$, and $T : \overline{\Omega} \to \overline{\Omega'}$ is a smooth diffeomorphism.

Sketch of the proof. — As we already pointed out before, the key point is to show an a priori estimate on second derivatives of smooth solutions of (4). Indeed, once such an estimate is proven, Equation (4) becomes uniformly elliptic, and standard PDE methods based on approximation allow to show the desired regularity of ψ inside Ω. (The regularity up to the boundary is more complicated, and needs a barrier argument.)

We will assume for simplicity that a stronger MTW condition holds: there exists a constant $K > 0$ such that

$$\mathfrak{S}_{(x,y)}(\xi,\eta) \geq K\,|\xi|_x^2 |\eta|_x^2 \qquad \text{whenever } \sum_{ij} c_{i,j}\xi^i\eta^j = 0.^{(1)} \tag{6}$$

Let us start from a smooth (say C^4) solution of (4), coupled with the boundary condition $T(\Omega) = \Omega'$, where $T(x) = \exp_x(\nabla\psi(x))$. The goal is to find a universal bound for the second derivatives of ψ.

We observe that, since $T(x) = \exp_x(\nabla\psi(x))$, we have

$$|\nabla\psi(x)| = d(x, T(x)) \leq \operatorname{diam}(M).$$

Hence ψ is globally Lipschitz, with a uniform Lipschitz bound. We define

$$w_{ij} := D^2_{x^i x^j}\psi + D^2_{x^i x^j} c\big(x, \exp_x(\nabla\psi(x))\big).$$

(Recall that by the c-convexity of ψ, (w_{ij}) is non-negative, and it is actually positive definite thanks to (4), as $h > 0$.) Then (4) can be written as

$$\det(w_{ij}) = h(x, \nabla\psi(x)), \tag{7}$$

or equivalently

$$\log\big(\det(w_{ij})\big) = \varphi,$$

with $\varphi(x) := \log\big(h(x, \nabla\psi(x))\big)$. By differentiating the above equation, and using the convention of summation over repeated indices, we get

$$w^{ij} w_{ij,k} = \varphi_k,$$

$$w^{ij} w_{ij,kk} = \varphi_{kk} + w^{is} w^{jt} w_{ij,k} w_{st,k} \geq \varphi_{kk},$$

where (w^{ij}) denotes the inverse of (w_{ij}). We use the notation $\psi_k = \frac{\partial}{\partial x^k}\psi$, $w_{ij,k} = \frac{\partial}{\partial x^k} w_{ij}$, $T_{s,k} = \frac{\partial}{\partial x^k} T_s$, and so on. Then the above equations become

$$w^{ij}\big[\psi_{ijk} + c_{ijk} + c_{ij,s} T_{s,k}\big] = \varphi_k, \tag{8}$$

$$w^{ij}\big[\psi_{ijkk} + c_{ijkk} + 2\,c_{ijk,s} T_{s,k} + c_{ij,s} T_{s,kk} + c_{ij,st} T_{s,k} T_{t,k}\big] \geq \varphi_{kk}. \tag{9}$$

(1) This stronger MTW condition is actually the one originally used in [**27, 31**]. The general case (i.e. $K = 0$) is treated in [**32**], where the authors relax the stronger assumption by applying a sort of barrier method, using a function $\tilde{u}$ which satisfies

$$\sum_{ij}\Big[D_{x^i x^j}\tilde{u} + \sum_k D_{p_k} A_{ij}(x, \nabla\psi(x))\, D_{x^k}\tilde{u}\Big]\xi^i\xi^j \geq \delta\,|\xi|^2, \qquad \delta > 0,$$

with $A_{ij}(x,p) := D^2_{x^i x^j} c\big(x, \exp_x(p)\big)$.

We fix now $\bar{x} \in \Omega$, we take η a cut-off function around $\bar{x}$, and define the function $G : \Omega \times \mathbb{S}^{n-1} \to \mathbb{R}$,

$$G(x,\xi) := \eta(x)^2\, w_{\xi\xi}, \qquad w_{\xi\xi} := \sum_{ij} w_{ij}\xi^i\xi^j.$$

We want to show that G is uniformly bounded by a universal constant C, depending only on $\mathrm{dist}(\bar{x}, \partial\Omega)$, n, the cost function, and the function $h(x,p)$. (Observe that $G \geq 0$, since (w_{ij}) is positive definite.) In fact, this will imply that

$$\eta(x)^2 \Big| D^2\psi(x) + D^2_x c\big(x, \exp_x\big(\nabla\psi(x)\big)\big)\Big| \leq C,$$

and since $\nabla\psi(x)$ is bounded and c is smooth, the above equation gives that $|D^2\psi|$ is locally uniformly bounded by a universal constant, which is the desired a priori estimate.

To prove the bound on G, the strategy is the following: let $x_0 \in \Omega$ and $\xi_0 \in \mathbb{S}^{n-1}$ be a point where G attains its maximum. By a rotation of coordinates, one can assume $\xi_0 = e_1$. Then at x_0 we have

$$0 = (\log G)_i = \frac{w_{11,i}}{w_{11}} + 2\,\frac{\eta_i}{\eta}, \tag{10}$$

$$(\log G)_{ij} = \frac{w_{11,ij}}{w_{11}} + 2\,\frac{\eta_{ij}}{\eta} - 6\,\frac{\eta_i\eta_j}{\eta^2}.$$

Since the above matrix is non-positive, we get

$$0 \geq w_{11}w^{ij}(\log G)_{ij} = w^{ij}w_{11,ij} + 2\,\frac{w_{11}}{\eta}w^{ij}\eta_{ij} - 6\,w_{11}w^{ij}\,\frac{\eta_i\eta_j}{\eta^2}. \tag{11}$$

We further observe that, differentiating (3), we obtain the relation

$$w_{ij} = c_{i,k}T_{k,j}. \tag{12}$$

This gives in particular $T_{k,j} = c^{k,i}w_{ij}$ (which implies $|\nabla T| \leq C\, w_{11}$), and allows to write derivatives of T in terms of that of w and c.

The idea is now to start from (11), and to combine the information coming from (8), (9), (10), (12), to end up with an inequality of the form

$$0 \geq w^{ij}\big[c^{k,\ell}c_{ij,k}c_{\ell,st} - c_{ij,st}\big]c^{s,p}c^{t,q}w_{p1}w_{q1} - C_0,$$

for some universal constant C_0. (When doing the computations, one has to remember that the derivatives of φ depend on derivatives of $\nabla\psi$, or equivalently on derivatives of T.) By a rotation of coordinates, one can further assume that (w_{ij}) is diagonal at x_0. We then obtain

$$w^{ii}\big[c^{k,\ell}c_{ii,k}c_{\ell,st} - c_{ii,st}\big]c^{s,1}c^{t,1}w_{11}w_{11} \leq C_0.$$

Up to now, the MTW condition has not been used. So, we now apply (6) to get

$$K\, w_{11}^2 \sum_i w^{ii} \leq C_0. \tag{13}$$

Observe that by the arithmetic-geometric inequality and by (7)

$$\sum_{i=1}^n w^{ii} \geq \sum_{i=2}^n w_{ii} \geq \Big(\prod_{i=2}^n w^{ii}\Big)^{1/(n-1)} \geq c_0\, w_{11}^{-1/(n-1)},$$

where $c_0 := \inf_{x\in\Omega} h(x, \nabla\psi(x))^{1/(n-1)} > 0$. Hence, combining the above estimate with (13) we finally obtain

$$c_0\, K \left[w_{11}(x_0)\right]^{2-1/(n-1)} \leq C_0,$$

which proves that $G(x,\xi) \leq G(x_0,\xi_0) \leq C_1$ for all $(x,\xi) \in \Omega \times \mathbb{S}^{n-1}$, as desired. □

3.2. A geometric interpretation of the MTW condition

Although the MTW condition seemed the right assumption to obtain regularity of optimal maps, it was only after Loeper's work **[24]** that people started to have a good understanding of this condition, and a more geometric insight. The idea of Loeper was the following: for the classical Monge-Ampère equation, a key property to prove regularity of convex solutions is that the subdifferential of a convex function is convex, and so in particular connected. Roughly speaking, this has the following consequence: whenever a convex function φ is not C^1 at a point x_0, there is at least a whole segment contained in the subdifferential of φ at x_0, and this fact combined with the Monge-Ampère equation provides a contradiction. (See also Theorem 3.6 below.) Hence, Loeper wanted to understand whether the c-subdifferential of a c-convex function is at least connected, believing that this fact had a link with the regularity. To explain all this in details, let us introduce some definitions.

Let $\varphi : \mathbb{R}^n \to \mathbb{R}$ be a convex function; its *subdifferential* $\partial\varphi(x)$ is given by

$$\begin{aligned}\partial\varphi(x) &= \big\{y \in \mathbb{R}^n \,|\, \varphi(x) + \varphi^*(y) = x\cdot y\big\}\\ &= \big\{y \in \mathbb{R}^n \,|\, \varphi(z) - z\cdot y \geq \varphi(x) - x\cdot y \quad \forall\, z \in \mathbb{R}^n\big\}.\end{aligned}$$

Then $\partial\varphi(x)$ is a convex set, a fortiori connected. More in general, given a semiconvex function $\phi : \mathbb{R}^n \to \mathbb{R}$ (i.e. ϕ can be locally written as the sum of a convex and a smooth function), its *subgradient* $\nabla^-\phi(\bar x)$ is defined as

$$\nabla^-\phi(\bar x) := \big\{p \,|\, \phi(\bar x + v) \geq \phi(\bar x) + \langle p, v\rangle + o(|v|) \quad \forall\, v\big\}.$$

We remark that, by working in charts, this definition makes sense also for functions ϕ defined on manifolds.

If we now consider $\psi : M \to \mathbb{R}$ a c-convex function, $c = d^2/2$, then

$$\begin{aligned}\partial^c\psi(x) &= \big\{y \in M \,|\, \psi(x) = \psi^c(y) - c(x,y)\big\}\\ &= \big\{y \in M \,|\, \psi(z) + c(z,y) \geq \psi(x) + c(x,y) \quad \forall\, z \in M\big\}\end{aligned}$$

(see Definition 1.2). In this generality there is no reason for $\partial^c\psi(x)$ to be connected, and in fact in general this is not the case!

Conditions for the connectedness of $\partial^c\psi$. — We now wish to find some simple enough conditions implying the connectedness of sets $\partial^c\psi$. In all the following arguments, we will assume for simplicity that points $(x,y) \in M \times M$ vary in a compact subset where the cost function $c = d^2/2$ is smooth. In particular it is well know that, under this assumption, for any pair (x,y) there exists a unique minimizing geodesic $\gamma_{x,y}$ joining them, which is given by $[0,1] \ni t \mapsto \exp_x(tv_{x,y})$, for some vector $v_{x,y} \in T_xM$. (See also Paragraph 3.5.1 below.) We will use the notation $(\exp_x)^{-1}(y) := v_{x,y}$.

First attempt to the connectedness. — Let us look first at the simplest c-convex functions:

$$\psi(x) := -c(x,y_0) + a_0.$$

Let $\bar{y} \in \partial^c\psi(\bar{x})$. Then the function $\psi(x) + c(x,\bar{y})$ achieves its minimum at $x = \bar{x}$, so that

$$-\nabla_x c(\bar{x},y_0) + \nabla_x c(\bar{x},\bar{y}) = 0.$$

This implies $(\exp_{\bar{x}})^{-1}(y_0) = (\exp_{\bar{x}})^{-1}(\bar{y})$, which gives $\bar{y} = y_0$. In conclusion $\partial^c\psi(\bar{x}) = \{y_0\}$ is a singleton, automatically connected, and so we do not get any information!

Second attempt to the connectedness. — The second simplest example of c-convex functions are

$$\psi(x) := \max\big\{-c(x,y_0) + a_0, -c(x,y_1) + a_1\big\}.$$

Take a point $\bar{x} \in \{x \,|\, -c(x,y_0) + a_0 = -c(x,y_1) + a_1\}$, and let $\bar{y} \in \partial^c\psi(\bar{x})$. Since $\psi(x) + c(x,\bar{y})$ attains its minimum at $x = \bar{x}$, we get

$$0 \in \nabla^-_{\bar{x}}\big(\psi + c(\cdot,\bar{y})\big),$$

or equivalently

$$-\nabla_x c(\bar{x},\bar{y}) \in \nabla^-\psi(\bar{x}).$$

From the above inclusion, one can easily deduce that $\bar{y} \in \exp_{\bar{x}}\big(\nabla^-\psi(\bar{x})\big)$. Moreover, it is not difficult to see that

$$\nabla^-\psi(\bar{x}) = \{(1-t)v_0 + tv_1 \,|\, t \in [0,1]\}, \qquad v_i := \nabla_x c(\bar{x},y_i) = (\exp_{\bar{x}})^{-1}(y_i), \quad i = 0,1.$$

Therefore, denoting by $[v_0,v_1]$ the segment joining v_0 and v_1, we obtain

$$\partial^c\psi(\bar{x}) \subset \exp_{\bar{x}}\big([v_0,v_1]\big).$$

The above formula suggests the following definition:

DEFINITION 3.2. — *Let $\bar{x} \in M$, $y_0, y_1 \notin \mathrm{cut}(\bar{x})$. Then we define the* c-segment *from y_0 to y_1 with base $\bar{x}$ as*

$$[y_0,y_1]_{\bar{x}} := \big\{y_t = \exp_{\bar{x}}\big((1-t)(\exp_{\bar{x}})^{-1}(y_0) + t(\exp_{\bar{x}})^{-1}(y_1)\big) \,|\, t \in [0,1]\big\}.$$

By slightly modifying some of the arguments in [**27**], Loeper showed that, under adequate assumptions, the connectedness of the c-subdifferential is a necessary condition for the smoothness of optimal transport (see also [**34**, Theorem 12.7]):

THEOREM 3.3 ([**24**]). — *Assume that there exist $\bar{x} \in M$ and $\psi : M \to \mathbb{R}$ c-convex such that $\partial^c\psi(\bar{x})$ is not (simply) connected. Then one can construct two probability densities f and g, C^∞ and strictly positive on M, such that the optimal map is discontinuous.*

While the above result was essentially contained in [**27**], Loeper's major contribution was to link the connectedness of the c-subdifferential to a differential condition on the cost function, which actually coincides with the MTW condition (see Paragraph 3.3). He proved (a slightly weaker version of) the following result, still assuming that the points (x, y) vary in a compact set where the cost function is smooth (see [**34**, Chapter 12] for a more general statement):

THEOREM 3.4 ([**24**]). — *The following conditions are equivalent:*

(i) *For any ψ c-convex, for all $\bar{x} \in M$, $\partial^c\psi(\bar{x})$ is connected.*

(ii) *For any ψ c-convex, for all $\bar{x} \in M$, $(\exp_{\bar{x}})^{-1}\big(\partial^c\psi(\bar{x})\big)$ is convex, and it coincides with $\nabla^-\psi(\bar{x})$.*

(iii) *For all $\bar{x} \in M$, for all y_0, y_1, if $[y_0, y_1]_{\bar{x}} = (y_t)_{t\in[0,1]}$, then*

$$d(x, y_t)^2 - d(\bar{x}, y_t)^2 \geq \min\big[d(x, y_0)^2 - d(\bar{x}, y_0)^2, d(x, y_1)^2 - d(\bar{x}, y_1)^2\big] \tag{14}$$

for all $x \in M$, $t \in [0, 1]$.

(iv) *For all $\bar{x}, y \in M$, for all $\eta, \xi \in T_{\bar{x}}M$ with $\xi \perp \eta$,*

$$\left.\frac{d^2}{ds^2}\right|_{s=0} \left.\frac{d^2}{dt^2}\right|_{t=0} d\big(\exp_{\bar{x}}(t\xi), \exp_{\bar{x}}(p + s\eta)\big)^2 \leq 0,$$

where $p = (\exp_{\bar{x}})^{-1}(y)$.

Moreover, if any of these conditions is not satisfied, C^1 c-convex functions are not dense in Lipschitz c-convex functions.

Sketch of the proof. — We give here only some elements of the proof.

(ii) $\Rightarrow$ (i): since $(\exp_{\bar{x}})^{-1}\big(\partial^c\psi(\bar{x})\big)$ is convex, it is connected, and so its image by $\exp_{\bar{x}}$ is connected too.

(i) $\Rightarrow$ (ii): for $\psi_{\bar{x},y_0,y_1} := \max\big\{-c(\cdot, y_0) + c(\bar{x}, y_0), -c(\cdot, y_1) + c(\bar{x}, y_1)\big\}$ we have $(\exp_{\bar{x}})^{-1}\big(\partial^c\psi_{\bar{x},y_0,y_1}(\bar{x})\big) \subset [(\exp_{\bar{x}})^{-1}(y_0), (\exp_{\bar{x}})^{-1}(y_1)]$, which is a segment. Since in this case connectedness is equivalent to convexity, if (i) holds we obtain $\partial^c\psi_{\bar{x},y_0,y_1}(\bar{x}) = [y_0, y_1]_{\bar{x}}$, and $\partial^c\psi_{\bar{x},y_0,y_1}(\bar{x}) = \exp_{\bar{x}}\big(\nabla^-\psi_{\bar{x},y_0,y_1}(\bar{x})\big)$

In the general case, we fix $y_0, y_1 \in \partial^c\psi(\bar{x})$. Then it is simple to see that

$$\partial^c\psi(\bar{x}) \supset \partial^c\psi_{\bar{x},y_0,y_1}(\bar{x}) = [y_0,y_1]_{\bar{x}},$$

and the result follows easily.

(ii) $\Leftrightarrow$ (iii): condition (14) is equivalent to $\partial^c\psi_{\bar{x},y_0,y_1} = [y_0,y_1]_{\bar{x}}$. Then the equivalence between (ii) and (iii) follows arguing as above.

(iii) $\Rightarrow$ (iv): fix $\bar{x} \in M$, and let $y := \exp_{\bar{x}}(p)$. Take ξ, η orthogonal and with unit norm, and define

$$y_0 := \exp_{\bar{x}}(p - \varepsilon\eta), \quad y_1 := \exp_{\bar{x}}(p + \varepsilon\eta) \qquad \text{for some } \varepsilon > 0 \text{ small.}$$

Moreover, let

$$h_0(x) := c(\bar{x},y_0) - c(x,y_0), \;\; h_1(x) := c(\bar{x},y_1) - c(x,y_1), \;\; \psi := \max\{h_0,h_1\} = \psi_{\bar{x},y_0,y_1}.$$

We now define $\gamma(t)$ as a curve contained in the set $\{h_0 = h_1\}$ such that $\gamma(0) = \bar{x}$, $\dot{\gamma}(0) = \xi$.

Since $y \in [y_0,y_1]_{\bar{x}}$, by (iii) we get $y \in \partial^c\psi(\bar{x})$, so that

$$\begin{aligned}\frac{1}{2}\big[h_0(\bar{x}) + h_1(\bar{x})\big] + c(\bar{x},y) &= \psi(\bar{x}) + c(\bar{x},y) \le \psi(\gamma(t)) + c(\gamma(t),y) \\ &= \frac{1}{2}\big[h_0(\gamma(t)) + h_1(\gamma(t))\big] + c(\gamma(t),y),\end{aligned}$$

where we used that $h_0 = h_1$ along γ. Recalling the definition of h_0 and h_1, we deduce

$$\frac{1}{2}\big[c(\gamma(t),y_0) + c(\gamma(t),y_1)\big] - c(\gamma(t),y) \le \frac{1}{2}\big[c(\bar{x},y_0) + c(\bar{x},y_1)\big] - c(\bar{x},y),$$

so the function $t \mapsto \frac{1}{2}\big[c(\gamma(t),y_0) + c(\gamma(t),y_1)\big] - c(\gamma(t),y)$ achieves its maximum at $t = 0$. This implies

$$\frac{d^2}{dt^2}\bigg|_{t=0}\Big[\frac{1}{2}\big(c(\gamma(t),y_0) + c(\gamma(t),y_1)\big) - c(\gamma(t),y)\Big] \le 0,$$

i.e.

$$\Big\langle\Big[\frac{1}{2}\big(D_x^2c(\bar{x},y_0) + D_x^2c(\bar{x},y_1)\big) - D_x^2c(\bar{x},y)\Big]\cdot\xi,\xi\Big\rangle \le 0$$

(here we used that $\nabla_x c(\bar{x},y) = \frac{1}{2}\big[\nabla_x c(\bar{x},y_0) + \nabla_x c(\bar{x},y_1)\big]$). Thus the function

$$\eta \mapsto \big\langle D_x^2c\big(\bar{x},\exp_{\bar{x}}(p+\eta)\big)\cdot\xi,\xi\big\rangle$$

is concave, and proves (iv). □

The above theorem leads to the definition of the *regularity* property:

DEFINITION 3.5. — *The cost function $c = d^2/2$ is said to be* regular *if the properties listed in Theorem 3.4 are satisfied.*

To understand why the above properties are related to smoothness, consider Theorem 3.4(iii). It says that, if we take the function

$$\psi_{\bar x,y_0,y_1} = \max\big\{-c(\cdot,y_0)+c(\bar x,y_0), -c(\cdot,y_1)+c(\bar x,y_1)\big\},$$

then we are able to touch the graph of this function from below at $\bar x$ with the family of functions $\{-c(\cdot,y_t)+c(\bar x,y_t)\}_{t\in[0,1]}$. This suggests that we could use this family to regularize the cusp of $\psi_{\bar x,y_0,y_1}$ at the point $\bar x$, by slightly moving above the graphs of the functions $-c(\cdot,y_t)+c(\bar x,y_t)$. On the other hand, if (14) does not hold, it is not clear how to regularize the cusp preserving the condition of being c-convex.

By what we said above, the regularity property seems mandatory to develop a theory of smoothness of optimal transport. Indeed, if it is not satisfied, we can construct C^∞ strictly positive densities f, g such that the optimal map is not continuous. Hence the natural question is when it is satisfied, and what is the link with the MTW condition.

3.3. A unified point of view

As we have seen in Theorem 3.4, the regularity of $c = d^2/2$ is equivalent to

$$\frac{d^2}{ds^2}\bigg|_{s=0}\ \frac{d^2}{dt^2}\bigg|_{t=0} c\big(\exp_x(t\xi), \exp_x(p+s\eta)\big) \le 0, \tag{15}$$

for all $p,\xi,\eta\in T_xM$, with ξ and η orthogonal, $p=(\exp_x)^{-1}(y)$.

By introducing a local system of coordinates $(x^1,\dots,x^n)$ around x, and a system $(y^1,\dots,y^n)$ around y, it is not difficult to check (by some standard but tedious computations) that the above expression *coincides* up to the sign with the MTW tensor. Hence the MTW condition is *equivalent* to the connectedness of the c-subdifferential of a c-convex function, and by Theorems 3.1 and 3.3, it is a necessary and sufficient condition for the smoothness of the optimal transport map. (At least, as long as the cost function is smooth on the supports of the two densities!)

By exploiting (a variant of) Theorem 3.4, Loeper proved the following regularity result:

THEOREM 3.6 ([**24**]). — *Let (M,g) be a compact Riemannian manifold, and let Ω and Ω' denote the support of f and g respectively. Assume that* (6) *holds for some $K>0$, f is bounded from above on Ω, g is bounded away from zero on Ω', and the cost function $c=d^2/2$ is smooth on the set $\overline{\Omega}\times\overline{\Omega'}$. Finally, suppose that $(\exp_x)^{-1}(\Omega')\subset T_xM$ is convex for any $x\in\Omega$. Then $\psi\in C^{1,\alpha}(\overline{\Omega})$, with $\alpha=1/(4n-1)$, so that $T\in C^{0,\alpha}(\overline{\Omega},\overline{\Omega'})$.*

A remarkable fact of the above result is that the Hölder exponent found by Loeper is explicit. (For instance, for the classical Monge-Ampère equation one can prove $C^{1,\alpha}$ regularity of solutions under weak assumptions on the densities [**4, 5, 6**], but there is

no explicit lower bound on the exponent.) As shown recently by Liu **[23]**, the optimal exponent in the above theorem is $\alpha = 1/(2n-1)$.

Sketch of the proof. — We will just prove C^1 regularity of ψ. A very similar argument, with a slightly more refined analysis, gives the $C^{1,\alpha}$-estimate.

Assume that ψ is not C^1 at x_0. Being a c-convex function, ψ is defined as a supremum of smooth functions of the form $x \mapsto -c(x,y) + \psi^c(y)$, and in particular is semiconvex. Hence it is not difficult to see that being not differentiable at x_0 means that there exist two points $y_0, y_1 \in \partial^c\psi(x_0) \cap \Omega'$. Let now $(y_t)_{t\in[0,1]} = [y_0,y_1]_{x_0} \subset \Omega'$ be the c-segment from y_0 to y_1 with base x_0. (Here we are using the assumption that $\exp_{x_0}^{-1}(\Omega')$ is convex.) Thanks to the (stronger) MTW condition (6), one can prove an improved version of (14) (see also (20) below): writing everything in charts,

$$\psi(x) + c(x,y_t) \geq \psi(x_0) + c(x_0,y_t) + \delta_0\,|y_1-y_0|^2\,|x-x_0|^2 + O(|x-x_0|^3)$$

for all $t \in [1/4,3/4]$. The idea is now the following: let y belong to a ε-neighborhood of the curve $(y_t)_{t\in[1/4,3/4]}$, and consider the function $f_a := -c(\cdot,y) + a$ with $a \in \mathbb{R}$. If a is sufficiently negative, then this function is below ψ in the closed ball $\overline{B}_{c_0\varepsilon}(x_0)$, where $c_0 > 0$ has to be chosen. Now, let a increase until f_a touches ψ from below inside $\overline{B}_{c_0\varepsilon}(x_0)$. Thanks to the above inequality, if c_0 is chosen sufficiently large (but fixed once for all, independently of ε), then for all $\varepsilon > 0$ sufficiently small the contact point will belong to the open ball $B_{c_0\varepsilon}(x_0)$. By this fact and Theorem 3.4(ii), we easily obtain that

$$\partial^c\psi\big(B_{c_0\varepsilon}(x_0)\big) \supset N_\varepsilon\big((y_t)_{t\in[1/4,3/4]}\big),$$

where N_ε denotes the ε-neighborhood. In terms of the optimal transport problem, this means that any point y belonging to $N_\varepsilon\big((y_t)_{t\in[1/4,3/4]}\big)$ is the image through the optimal transport map T of a point in the ball $B_{c_0\varepsilon}(x_0)$ (see Theorem 1.3(ii)). By the transport condition $T_\#(f\mathrm{vol}) = g\mathrm{vol}$, this implies

$$\int_{B_{c_0\varepsilon}} f\,d\mathrm{vol} \geq \int_{N_\varepsilon\left((y_t)_{t\in[1/4,3/4]}\right)} g\,d\mathrm{vol}.$$

However, by the assumptions on f and g we have

$$\int_{N_\varepsilon\left((y_t)_{t\in[1/4,3/4]}\right)} g\,d\mathrm{vol} \gtrsim \mathrm{vol}\big(N_\varepsilon\big((y_t)_{t\in[1/4,3/4]}\big)\big) \sim \varepsilon^{n-1},$$

$$\int_{B_{c_0\varepsilon}} f\,d\mathrm{vol} \lesssim \varepsilon^n,$$

and all these conditions are not compatible if $\varepsilon > 0$ is sufficiently small. This contradiction proves the C^1 regularity of ψ. □

Let us now consider the following geometric example.

Absence of regularity in presence of negative sectional curvature. — We want to show how negative sectional curvature is an obstruction to regularity (indeed even to continuity) of optimal maps. We refer to [**34**, Theorem 12.4] for more details on the construction given below.

Let $M = \mathbb{H}^2$ be the hyperbolic plane (or a compact quotient thereof). Fix a point O as the origin, and fix a local system of coordinates in a neighborhood of O such that the maps $(x_1, x_2) \mapsto (\pm x_1, \pm x_2)$ are local isometries (it suffices for instance to consider the model of the Poincaré disk, with O equal to the origin in $\mathbb{R}^2$). Then define the points

$$A^{\pm} = (0, \pm\varepsilon), \quad B^{\pm} = (\pm\varepsilon, 0) \qquad \text{for some } \varepsilon > 0.$$

Take a measure μ symmetric with respect to 0 and concentrated near $\{A^+\} \cup \{A^-\}$ (say 3/4 of the total mass belongs to a small neighborhood of $\{A^+\} \cup \{A^-\}$), and a measure ν symmetric with respect to 0 and concentrated near $\{B^+\} \cup \{B^-\}$. Moreover assume that μ and ν are absolutely continuous, and have strictly positive densities everywhere. We denote by T the unique optimal transport map, and we assume by contradiction that T is continuous. By symmetry, we deduce that $T(O) = O$. Then, by counting the total mass, there exists a point A' close to A^+ which is sent to a point B' near, say, B^+.

But, by negative curvature (if A' and B' are close enough to A and B respectively), Pythagoras Theorem becomes an inequality: $d(O, A')^2 + d(O, B')^2 < d(A', B')^2$, and this contradicts the optimality of the transport map, as transporting A' onto O and O onto B' would be more convenient than transporting A' onto B' and letting O stay at rest.

Now, the natural question is: how does the above example fit into Ma, Trudinger and Wang's and Loeper's results? The answer is actually pretty simple: in [**24**] Loeper noticed that the MTW tensor satisfies the following remarkable identity:

$$\mathfrak{S}_{(x,x)}(\xi, \eta) = -\frac{3}{4} \left.\frac{d^2}{ds^2}\right|_{s=0} \left.\frac{d^2}{dt^2}\right|_{t=0} d\big(\exp_x(t\xi), \exp_x(s\eta)\big)^2 = \mathrm{Sect}_x([\xi, \eta]), \tag{16}$$

where $\xi, \eta \in T_xM$ are two orthogonal unit vectors, and $\mathrm{Sect}_x([\xi, \eta])$ denotes the sectional curvature of the plane generated by ξ and η.

In fact, as shown by Kim and McCann [**20**], $\mathfrak{S}$ is the sectional curvature of the manifold $M \times M$, endowed with the pseudo-metric $-d^2_{xy}c$. Combining (16) with Theorems 3.3 and 3.4, we get the following important negative result:

THEOREM 3.7. — *Let (M, g) be a (compact) Riemannian manifold, and assume that there exist $x \in M$ and a plane $P \subset T_xM$ such that $\mathrm{Sect}_x(P) < 0$. Then there exist C^∞ strictly positive probability densities f and g such that the optimal map is discontinuous.*

After this negative result, one could still hope to develop a regularity theory on any manifold with non-negative sectional curvature. But this is not the case: as shown by Kim [**19**], the regularity condition is strictly stronger than the condition of non-negativity of sectional curvatures. In conclusion, except for some special cases (see Paragraphs 3.4 and 3.5.2 below), the optimal map is non-smooth!

3.4. More on the MTW condition

As shown above, the MTW tensor is a non-local version of the sectional curvature, and the MTW condition is a stronger condition than non-negative sectional curvature. We further remark that the MTW condition is intrinsic, and independent of the system of coordinates.

To see this, we first show that (5) can also be written as

$$\frac{\partial^2}{\partial p_\eta^2}\frac{\partial^2}{\partial x_\xi^2}\, c(x,y) \leq 0. \tag{17}$$

The meaning of the left-hand side in (17) is the following: first freeze y and differentiate $c(x,y)$ twice with respect to x in the direction $\xi \in T_xM$. Then, considering the result as a function of y, parameterize y by $p = -\nabla_x c(x,y)$, and differentiate twice with respect to p in the direction $\eta \in T_yM$. By the relation $p_i = -c_i(x,y)$ we get $\frac{\partial p_i}{\partial y^j} = -c_{i,j}$, which gives $\frac{\partial y^k}{\partial p_\ell} = -c^{k,\ell}$. Finally, using $-c_{i,j}$ and $-c^{i,j}$ to raise and lower indices ($\eta^k = -c^{k,l}\eta_i$, etc.), it is just a (tedious) exercise to show that the expression in (17) is equal to

$$\sum_{ijklrs} \left(c_{ij,kl} - c_{ij,r}c^{r,s}c_{s,kl}\right)\xi^i\xi^j\eta^k\eta^l,$$

where we used the formula $d(M^{-1})\cdot H = -M^{-1}HM^{-1}$.

Since the expression in (17) involves second derivatives (which are not intrinsic and depend on the choice of the coordinates), it is not a priori clear whether $\mathfrak{S}$ depends or not on the choice of coordinates. On the other hand, we can hope it does not, because of the (intrinsic) geometric interpretation of the regularity.

To see that $\mathfrak{S}$ is indeed independent of the choice of coordinates (so that one does not even need to use geodesic coordinates, as in (15)), we observe that, if we do a change of coordinates and compute first the second derivatives in x, we get some additional terms of the form

$$\Gamma_{ij}^k(x)c_k(x,y) = -\Gamma_{ij}^k(x)p_k(x,y) = \Gamma_{ij}^k(x)g_{k\ell}(x)p^\ell(x,y).$$

But when we differentiate twice with respect to p, this additional term disappears! This shows that the MTW tensor is *independent* of the system of coordinates.

Let us introduce the following:

DEFINITION 3.8. — *Given $K \geq 0$, we say that (M, g) satisfies the* MTW(K) *condition if, for all $(x, y) \in (M \times M)$ where $c = d^2/2$ is smooth, for all $\xi \in T_xM$, $\eta \in T_yM$,*

$$\mathfrak{S}_{(x,y)}(\xi, \eta) \geq K|\xi|_x^2|\tilde{\eta}|_x^2 \qquad \text{whenever } -c_{i,j}(x, y)\xi^i\eta^j = 0,$$

where $\tilde{\eta}^i = -g^{i,k}(x)c_{k,j}(x, y)\eta^j \in T_xM$.

Remark 3.9. — Observe that, thanks to (16), the MTW(K) condition implies in particular that all sectional curvatures are bounded from below by K. Therefore, if $K > 0$, by the Bonnet-Myers Theorem the diameter of the manifold is bounded, and the manifold is compact.

Some examples of manifolds satisfying the MTW condition are given in [**12, 20, 21, 24, 25**]:

- $\mathbb{R}^n$ and $\mathbb{T}^n$ satisfy MTW(0).
- $\mathbb{S}^n$, its quotients (like $\mathbb{RP}^n$), and its submersions (like $\mathbb{CP}^n$ or $\mathbb{HP}^n$), satisfy MTW(1).
- Products of any of the examples listed above (for instance, $\mathbb{S}^{n_1} \times \ldots \times \mathbb{S}^{n_k} \times \mathbb{R}^\ell$ or $\mathbb{S}^{n_1} \times \mathbb{CP}^{n_2} \times \mathbb{T}^{n_3}$) satisfy MTW(0).

We observe that the MTW condition is a non-standard curvature condition, as it is fourth order and nonlocal. Therefore an important open problem is whether this condition is stable under perturbation. More precisely, we ask for the following:

QUESTION. — *Assume that (M, g) satisfies the* MTW(K) *condition for $K > 0$, and let g_ε be a C^4-perturbation of g. Does (M, g_ε) satisfy the* MTW(K') *condition for some $K' > 0$?*

The answer is easily seen to be affirmative for manifolds with nonfocal cut-locus like the projective space $\mathbb{RP}^n$ (see [**12, 26**], and Theorem 3.12 below). Moreover, as proven by Figalli and Rifford [**12**], the answer is affirmative also for the 2-dimensional sphere $\mathbb{S}^2$ (see Theorem 3.14 below). The extension of this last result to arbitrary dimension has been recently achieved by Figalli, Rifford and Villani [**13**].

As a corollary of these facts, one can prove that regularity of optimal maps holds in all of these cases. In the next paragraph, we will explain the link between the MTW condition and the geometry of the cut-locus, and we will describe more in detail the aforementioned results.

3.5. Relation between the MTW condition and the geometry of the cut-locus

3.5.1. *The cut and focal-locus.* — We recall that, given a tangent vector $v \in T_xM$, the curve $(\exp_x(tv))_{t\geq 0}$ is a geodesic defined for all times, but in general is not minimizing for large times. On the other hand, it is possible to prove that $\exp_x(tv)$ is always minimizing between x and $\exp_x(\varepsilon v)$ for $\varepsilon > 0$ sufficiently small. We define the *cut-time* $t_C(x,v)$ as

$$t_C(x,v) := \inf\Big\{t > 0 \,|\, s \mapsto \exp_x(sv) \text{ is not minimizing between } x \text{ and } \exp_x(tv)\Big\}.$$

Given two points $x, y \in M$, whenever there exists a unique minimizing geodesic $(\exp_x(tv))_{0\leq t\leq 1}$ going from x to y in time 1, we write $(\exp_x)^{-1}(y) := v$.

Given $x \in M$, we define the *cut-locus* of x as

$$\mathrm{cut}(x) := \Big\{\exp_x\big(t_C(x,v)v\big) \,|\, v \in T_xM,\ |v|_x = 1\Big\}.$$

We further define

$$\mathrm{cut}(M) := \{(x,y) \in M \times M \ \mid\ y \in \mathrm{cut}(x)\}.$$

EXAMPLE. — *On the sphere $\mathbb{S}^n$, the geodesics starting from a point x with unit speed describe great circles passing through its antipodal point $-x$. These geodesics are minimizing exactly until they reach $-x$ after a time π. Thus $t_C(x,v) = \pi$ for any $v \in T_xM$ with unit norm, and $\mathrm{cut}(x) = \{-x\}$. By time-rescaling, we get $t_C(x,v) = \frac{\pi}{|v|_x}$ for any $x \in \mathbb{S}^n$, $v \in T_xM \setminus \{0\}$.*

It is possible to prove that, if $y \notin \mathrm{cut}(x)$, then x and y are joined by a unique minimizing geodesic. The converse is close to be true: $y \notin \mathrm{cut}(x)$ if and only if there are neighborhoods U of x and V of y such that any two points $x' \in U$, $y' \in V$ are joined by a unique minimizing geodesic. In particular $y \notin \mathrm{cut}(x)$ if and only if $x \notin \mathrm{cut}(y)$.

Given now $x \in M$ and $v \in T_xM$, we define the *focal-time* $t_F(x,v)$ as

$$t_F(x,v) := \inf\Big\{t > 0 \,|\, d_{tv}\exp_x : T_xM \to T_{\exp_x(tv)}M \text{ is not invertible}\Big\}.$$

We further introduce the *tangent cut-locus* of x

$$\mathrm{TCL}(x) = \big\{t_C(x,v)v \,|\, v \in T_xM,\ |v|_x = 1\big\},$$

the *tangent focal-locus* of x

$$\mathrm{TFL}(x) = \big\{t_F(x,v)v \,|\, v \in T_xM,\ |v|_x = 1\big\},$$

the *injectivity domain* of the exponential map at x

$$\mathrm{I}(x) = \big\{tv \,|\, 0 \leq t < t_C(x,v),\ v \in T_xM,\ |v|_x = 1\big\},$$

and the *nonfocal domain* of the exponential map at x

$$\mathrm{NF}(x) = \big\{tv \,|\, 0 \le t < t_F(x,v),\ v \in T_xM,\ |v|_x = 1\big\}.$$

With these definitions, we have

$$\mathrm{cut}(x) = \exp_x\big(\mathrm{TCL}(x)\big), \qquad \mathrm{TCL}(x) = \partial\big(\mathrm{I}(x)\big), \qquad \mathrm{TFL}(x) = \partial\big(\mathrm{NF}(x)\big).$$

We finally define the *focal cut-locus* of x as

$$\mathrm{fcut}(x) := \exp_x\big(\mathrm{TCL}(x) \cap \mathrm{TFL}(x)\big).$$

It is a well-known fact of Riemannian geometry that $t_C \le t_F$ (see for instance [**17**, Corollary 3.77]). In the case of the sphere, $t_C \equiv t_F$, and $\mathrm{cut}(x) = \mathrm{fcut}(x)$ for all $x \in \mathbb{S}^n$.

The fact that a point $y \in M$ belongs to $\mathrm{cut}(x)$ is a phenomenon which is captured by the regularity of the distance function. Indeed, it can be proven that the following hold (see for instance [**9**, Proposition 2.5]):

(a) The function $d(x,\cdot)^2$ is smooth (i.e. C^∞) in a neighborhood of y if and only if $y \notin \mathrm{cut}(x)$.

(b.1) The function $d(x,\cdot)^2$ has an upward cusp at y if and only if $y \in \mathrm{cut}(x)$ and there are at least two minimizing geodesics between x and y.

(b.2) The function $d(x,\cdot)^2$ is C^1 at y and its Hessian has an eigenvalue $-\infty$ if and only if $y \in \mathrm{cut}(x)$ and there is a unique minimizing geodesic between x and y. (In this case, y necessarily belongs to $\mathrm{fcut}(x)$.)

In the above statement, having an "upward cusp" means that there exist two vectors $p_1 \neq p_2$ both belonging to the supergradient of $f := d(x,\cdot)^2$ at y: writing everything in charts, we have

$$\{p_1, p_2\} \subset \nabla^+ f(y) := \big\{p \,|\, f(y+v) \le f(y) + \langle p, v\rangle + o(|v|) \quad \forall\, v\big\},$$

that is f is locally below the function $v \mapsto f(y) + \min\{\langle p_1, v\rangle, \langle p_2, v\rangle\} + o(|v|)$ near y. Hence (b.1) corresponds to roughly say that the second derivative (along the direction $p_2 - p_1$) of $d(x,\cdot)^2$ at y is $-\infty$. (The fact that there is an upward cusp, means that one of the second directional derivatives is a negative delta measure!)

Furthermore, saying that "Hessian has an eigenvalue $-\infty$" means that (always working in charts)

$$\liminf_{|v|\to 0} \frac{f(y+v) - 2f(y) + f(y-v)}{|v|^2} = -\infty.$$

Thus, all the above description of the cut-locus in terms of the squared distance can be informally summarized as follows:

$$y \in \mathrm{cut}(x) \quad \Leftrightarrow \quad \big\langle D_y^2 d^2(x,y)\cdot v, v\big\rangle = -\infty \quad \text{for some } v \in T_yM. \tag{18}$$

This observation will be of key importance in what follows.

3.5.2. *The* MTW *condition and the convexity of the tangent cut-locus.* — In **[26]**, Loeper and Villani noticed the existence of a deep connection between the MTW condition and the geometry of the cut-locus. The idea is the following: fix $x \in M$, and let $v_0, v_1 \in \mathrm{I}(x)$. Consider the segment $(v_t)_{t\in[0,1]}$, with $v_t := (1-t)v_0 + tv_1$. Set further $y_t := \exp_x(v_t)$. Since $v_0, v_1 \in \mathrm{I}(x)$, we have

$$y_0, y_1 \notin \mathrm{cut}(x).$$

In particular $c(x,\cdot) := d(x,\cdot)^2/2$ is smooth in a neighborhood of y_0 and y_1. Assume now that the MTW condition holds. Thanks to Theorem 3.4(iv), we know that the function

$$\eta \mapsto \big\langle D^2_x c\big(\bar{x}, \exp_{\bar{x}}(p+\eta)\big)\cdot\xi, \xi\big\rangle$$

is concave for all $\eta \perp \xi$. (This is just a formal argument, as the theorem applies a priori only if $\exp_{\bar{x}}(p+\eta) \notin \mathrm{cut}(\bar{x})$.) Applying this fact along the segment $(v_t)_{t\in[0,1]}$, and exploiting the smoothness of $d(x,\cdot)^2$ near y_0 and y_1, we obtain, for $\xi \perp (v_1 - v_0)$,

$$\inf_{t\in[0,1]} \big\langle D^2_x d^2(x,y_t)\cdot\xi,\xi\big\rangle \geq \min\big\{\big\langle D^2_x d^2(x,y_0)\cdot\xi,\xi\big\rangle,\, \big\langle D^2_x d^2(x,y_1)\cdot\xi,\xi\big\rangle\big\} \geq C_0,$$

for some constant $C_0 \in \mathbb{R}$. Hence, if we forget for a moment about the orthogonality assumption between $v_1 - v_0$ and ξ, we see that the above equation implies that $x \notin \mathrm{cut}(y_t)$ for all $t \in [0,1]$ (compare with (18)), which by symmetry gives

$$y_t \notin \mathrm{cut}(x) \qquad \forall\, t \in [0,1],$$

or equivalently

$$v_t \notin \mathrm{TCL}(x) \qquad \forall\, t \in [0,1].$$

Since $v_0, v_1 \in \mathrm{I}(x)$, we have obtained

$$v_t \in \mathrm{I}(x) \qquad \forall\, t \in [0,1],$$

that is $\mathrm{I}(x)$ is convex! In conclusion, this formal argument suggests that the MTW condition (or a variant of it) should imply that all tangent injectivity loci $\mathrm{I}(x)$ are convex, for every $x \in M$. This would be a remarkable property. Indeed, usually the only regularity results available for $\mathrm{I}(x)$ say that $\mathrm{TCL}(x)$ is just Lipschitz **[8, 18, 22]**. Moreover, such a result would be of a global nature, and not just local like a semi-convexity property.

Unfortunately, the argument described above is just formal, and up to now there is no complete result in that direction. However, one can actually prove some rigorous results. To do this, we will need to introduce some variant of the MTW condition.

Convexity of the cut-loci: the nonfocal case

DEFINITION 3.10 (uniform MTW condition). — *If $K, C \geq 0$ are given, it is said that M satisfies the* $\mathrm{MTW}(K,C)$ *condition if, for all $(x,y) \in (M \times M) \setminus \mathrm{cut}(M)$, for all $(\xi,\eta) \in T_xM \times T_yM$,*

$$\mathfrak{S}_{(x,y)}(\xi,\eta) \geq K\,|\xi|_x^2\,|\tilde{\eta}|_x^2 - C\,\langle \xi, \tilde{\eta}\rangle_x^2, \tag{19}$$

where $v = (\exp_x)^{-1}(y)$, $\tilde{\eta} = (d_v \exp_x)^{-1}(\eta)$.

DEFINITION 3.11. — *We say that Riemannian manifold (M,g) has nonfocal cut-locus if* $\mathrm{fcut}(x) = \varnothing$ *for all $x \in M$.*

As shown in **[26]** by a compactness argument, as long as $y \notin \mathrm{cut}(x)$ stays uniformly away from $\mathrm{fcut}(x)$, the $\mathrm{MTW}(K,C)$ condition is actually equivalent to the $\mathrm{MTW}(K)$ condition. In particular, if (M,g) is a compact manifold with nonfocal cut-locus, and the $\mathrm{MTW}(K)$ condition holds for some $K \geq 0$, then there exists a constant $C > 0$ such that the $\mathrm{MTW}(K,C)$ condition is true. Thanks to this fact, the authors can prove a variant of Theorem 3.4(iii), where they exploit the information coming from the fact that now the vectors ξ and η do not need to be orthogonal, in order to get an improved version of that result: with the same notation as in Theorem 3.4(iii), then there exists $\lambda = \lambda(K,C) > 0$ such that, for any $t \in (0,1)$,

$$\begin{aligned} d(x,y_t)^2 - d(\bar{x},y_t)^2 \geq \min\Big(d(x,y_0)^2 - d(\bar{x},y_0)^2,\, d(x,y_1)^2 - d(\bar{x},y_1)^2\Big) \\ + 2\lambda\, t(1-t)\, d(\bar{x},x)^2 |v_1 - v_0|_{\bar{x}}^2, \end{aligned} \tag{20}$$

where $v_0 = (\exp_{\bar{x}})^{-1}(y_0)$, $v_1 = (\exp_{\bar{x}})^{-1}(y_1)$. Moreover, they can even assume that y_t is not exactly a c-segment, but just a C^2-perturbation of it.

Thanks to this improved version of "regularity", Loeper and Villani showed the following result:

THEOREM 3.12 (**[26]**). — *Let (M,g) be a Riemannian manifold with nonfocal cut-locus, satisfying* $\mathrm{MTW}(K)$ *for some $K > 0$ (in particular, M is compact by Remark 3.9). Then there is $\kappa > 0$ such that all tangent injectivity domains* $\mathrm{I}(x)$ *are κ-uniformly convex.*

The (uniform) convexity of all injectivity loci is exactly what Ma, Trudinger and Wang needed as a geometric assumption in order to prove the regularity of the optimal map.

Hence, combining Theorem 3.1 with the strategy developed by Loeper in **[24]** (see Theorem 3.6), Loeper and Villani obtained the following theorem:

COROLLARY 3.13 ([**26**]). — *Let (M, g) be a Riemannian manifold with nonfocal cut-locus, satisfying* MTW(K) *for some $K > 0$. Assume that f and g are smooth probability densities, bounded away from zero and infinity on M. Then ψ (and hence T) is smooth.*

Sketch of the proof. — The first step of the proof consists in showing that ψ is C^1. This is done using the same strategy of Theorem 3.6, exploiting (20) and the convexity of all injectivity domains ensured by Theorem 3.12. We remark that the fact that (20) holds for C^2-perturbations of c-segments allows to simplify some technical parts of the original proof of Loeper, and to slightly relax some of his assumptions.

Then, one takes advantage of the nonfocality assumption to ensure the "stay-away property" $\operatorname{dist}(T(x), \operatorname{cut}(x)) \geq \sigma > 0$. To see how nonfocality plays a role in this estimate, we recall the description of the distance function given in Paragraph 3.5.1: roughly speaking

- $d(x, y)^2$ is smooth for $y \notin \operatorname{cut}(x)$.
- $d(x, y)^2$ is at most C^1 for $y \in \operatorname{fcut}(x)$.
- $d(x, y)^2$ is not C^1 for $y \in \operatorname{cut}(x) \setminus \operatorname{fcut}(x)$.

Hence, in presence of nonfocality, either $d(x, y)^2$ is smooth, or is not C^1, and in this last case there are at least two minimizing geodesics joining x to y. Now, when proving Theorem 1.3(iii), one actually shows that, whenever ψ is differentiable at x, there exists a unique minimizing geodesic from x to $T(x)$, given by $t \mapsto \exp_x(t\nabla\psi(x))$ [**28**]. Thus, if ψ is C^1, in the nonfocal case one immediately deduces that $T(x) \notin \operatorname{cut}(x)$ *for all* $x \in M$, and a simple compactness argument provides the existence of a positive $\sigma > 0$ such that $d(T(x), \operatorname{cut}(x)) \geq \sigma$.

Once the stay-away property is established, since all pairs $(x, T(x))$ belong to a set where d^2 is smooth, it is simple to localize the problem and apply the a priori estimates of Ma, Trudinger and Wang (see Theorem 3.1) to prove the smoothness of ψ. □

The above result applies for instance to the projective space $\mathbb{RP}^n$ and its perturbations. We also recall that the smoothness of optimal maps holds true in the case of the sphere $\mathbb{S}^n$, as shown by Loeper [**25**]. However, a non-trivial question is whether the regularity of optimal maps holds for perturbations of the sphere.

By imposing some uniform L^∞-bound on the logarithm of the densities (so that they are uniformly bounded away from zero and infinity), Delanoë and Ge showed that for small perturbations of the metric (the smallness depending on the L^∞-bound) the optimal map stays uniformly away from the cut-locus, in the sense that $\operatorname{dist}(T(x), \operatorname{cut}(x)) \geq \sigma$ for some $\sigma > 0$ [**11**], and in this case the regularity issue presents no real difficulties (see the last part of the proof of Corollary 3.13). However

this stay-away property does not necessarily hold for general smooth densities, and the problem becomes much more complicated. The case of perturbations of $\mathbb{S}^2$ has been solved by Figalli and Rifford [**12**], and their result has been recently extended to arbitrary dimension by Figalli, Rifford and Villani [**14**].

The extended MTW condition. — We observe that, from the point of view of the structure of the cut-locus, the perturbations of the sphere are in some sense the worst case to treat. Indeed, since for $\mathbb{S}^n$ one has $\mathrm{cut}(x) = \mathrm{fcut}(x)$ for all $x \in M$ (which is completely the opposite of nonfocality), when one slightly perturbs the metric the structure of the cut-locus can be very wild. (The idea is that the cut-locus behaves nicely under perturbations of the metric away from focalization, while it is very difficult to control its behavior near the focal-locus [**8**]).

To overcome these difficulties, Figalli and Rifford introduced the following strategy: first of all, we observe that the MTW condition is defined only for $(x, y) \in M \times M$ with $y \notin \mathrm{cut}(x)$. Hence, we can write it as a condition on the pairs (x, v) instead of (x, y), where $v := (\exp_x)^{-1}(y) \in \mathrm{I}(x)$.

We fix now $\bar{x} \in M$, and we observe that the MTW tensor at $(\bar{x}, v)$ (or equivalently at $(\bar{x}, \exp_{\bar{x}}(v))$) is expressed in terms of derivatives of $d^2/2$ at $(\bar{x}, \exp_{\bar{x}}(v))$. Now, assume that v approaches $\mathrm{TCL}(\bar{x})$ but it is still far from $\mathrm{TFL}(\bar{x})$. This means that the map $(x, w) \mapsto (x, \exp_x(w))$ is a local diffeomorphism near $(\bar{x}, v)$. Hence, we can define a new cost function for (x, y) near $(\bar{x}, \exp_{\bar{x}}(v))$ as

$$\hat{c}(x, y) := \frac{\|(\exp_x)^{-1}(y)\|_x^2}{2},$$

where now $(\exp_x)^{-1}$ denotes the local smooth inverse of $\exp_x$, as explained above. This new cost function coincides with $d(x, y)^2/2$ as long as $y = \exp_x(w)$ with $w \in \mathrm{I}(x)$, and it provides a smooth extension of it up to the first conjugate time. This allows to define an *extended* MTW condition, which makes sense for all pairs (x, v) with $v \in \mathrm{NF}(x)$ (and not only for $v \in \mathrm{I}(x)$). The advantage of having extended the MTW condition up to the focal-locus is twofold: on the one hand, the extended MTW condition is more "local", as one can easily show that it only concerns the geodesic flow, and not the global topology of the manifold. On the other hand, the fact of being allowed to cross the cut-locus away from the focal points makes this extended condition more flexible than the usual one, and this strongly helps when trying to prove the convexity of all tangent injectivity domains. Exploiting these facts, Figalli and Rifford proved the following result:

THEOREM 3.14 ([**12**]). — *Let (M, g) be a Riemannian manifold which satisfies the extended* $\mathrm{MTW}(K, C)$ *condition for some $K, C > 0$, and assume that* $\mathrm{NF}(x)$ *is (strictly) convex for all $x \in M$. Then* $\mathrm{I}(x)$ *is (strictly) convex for all $x \in M$.*

We observe that, in the above result, the authors replace the nonfocality assumption as in Theorem 3.13 with the convexity of all tangent nonfocal domains. This hypothesis is satisfied for instance by any perturbation of the sphere $\mathbb{S}^n$ (see for example **[8]**).

The above theorem allows also to prove a regularity result for optimal maps:

COROLLARY 3.15 (**[12]**). — *Let (M,g) be a Riemannian manifold which satisfies the extended* $\mathrm{MTW}(K,C)$ *condition for some $K,C>0$, and assume that* $\mathrm{NF}(x)$ *is (strictly) convex for all $x\in M$. Assume that f and g are two probability densities bounded away from zero and infinity on M. Then the optimal map is continuous.*

We remark that the statement of the above theorem does not say that if f and g are smooth, then T is smooth too. The difficulty to prove such a result comes again from focalization: if the cut-locus is nonfocal, as shown in the proof of Corollary 3.13 the continuity of the transport map implies the stay-away property $\mathrm{dist}(T(x),\mathrm{cut}(x))\geq\sigma>0$, and from this fact the higher regularity of T follows easily **[26]**. Unfortunately, without nonfocality (as in the above case), the continuity of T is not enough to ensure the stay-away property, and this is why the above statement is only about the continuity of the optimal map.

In **[12]** the authors show that the sphere $\mathbb{S}^n$ satisfies the (extended) $\mathrm{MTW}(K,C)$ condition for some $K=C>0$, and they prove that this condition survives for perturbations of the two-dimensional sphere. In particular, they obtain as a corollary the following result:

COROLLARY 3.16 (**[12]**). — *Let $(M,g)=(\mathbb{S}^2,g_\varepsilon)$, where g_ε is a C^4-perturbation of canonical metric on $\mathbb{S}^2$. Then, for ε small enough,* $\mathrm{I}(x)$ *is strictly convex for all $x\in M$. Moreover, if f and g are two probability densities bounded away from zero and infinity on M, then the optimal map is continuous.*

Conclusions. — An interesting remark to the above result is the following: the first part of the statement of Corollary 3.16 is a statement on perturbations of the 2-sphere, which has nothing to do with optimal transport! Moreover, the same is true for many of the results stated above, which are just statements on the structure of the cut-locus. So, what happened can be summarized as follows: to prove regularity of optimal maps, Ma, Trudinger and Wang discovered a new tensor by purely PDE methods, starting from a Monge-Ampère type equation. Then it was realized that this tensor is intrinsic and has a geometric meaning, and now the MTW tensor is used as a tool (like the Ricci or the Riemann tensor) to prove geometric statements on manifolds. (For a recent account on other possible links between optimal transport and geometry, see **[16]**.) This domain of research is new and extremely active, and there are still a lot of open problems. For instance, a complete understanding of the link between the MTW

condition and the convexity of the tangent cut-loci is still missing (although in **[15]** the authors have a quite complete answer in the case of 2-dimensional manifolds). Another formidable challenge is for example the description of positively curved Riemannian manifolds which satisfy $\mathrm{MTW}(K, C)$, for some $K, C > 0$.

REFERENCES

[1] Y. Brenier – "Décomposition polaire et réarrangement monotone des champs de vecteurs", *C. R. Acad. Sci. Paris Sér. I Math.* **305** (1987), p. 805–808.

[2] ———, "Polar factorization and monotone rearrangement of vector-valued functions", *Comm. Pure Appl. Math.* **44** (1991), p. 375–417.

[3] X. Cabré – "Nondivergent elliptic equations on manifolds with nonnegative curvature", *Comm. Pure Appl. Math.* **50** (1997), p. 623–665.

[4] L. A. Caffarelli – "A localization property of viscosity solutions to the Monge-Ampère equation and their strict convexity", *Ann. of Math.* **131** (1990), p. 129–134.

[5] ———, "Some regularity properties of solutions of Monge Ampère equation", *Comm. Pure Appl. Math.* **44** (1991), p. 965–969.

[6] ———, "The regularity of mappings with a convex potential", *J. Amer. Math. Soc.* **5** (1992), p. 99–104.

[7] ———, "Boundary regularity of maps with convex potentials. II", *Ann. of Math.* **144** (1996), p. 453–496.

[8] M. Castelpietra & L. Rifford – "Regularity properties of the distance function to conjugate and cut loci for viscosity solutions of Hamilton-Jacobi equations and applications in Riemannian geometry", to appear in *ESAIM Control Optim. Calc. Var.*.

[9] D. Cordero-Erausquin, R. J. McCann & M. Schmuckenschläger – "A Riemannian interpolation inequality à la Borell, Brascamp and Lieb", *Invent. Math.* **146** (2001), p. 219–257.

[10] J. A. Cuesta & C. Matrán – "Notes on the Wasserstein metric in Hilbert spaces", *Ann. Probab.* **17** (1989), p. 1264–1276.

[11] P. Delanoë & Y. Ge – "Regularity of optimal transportation maps on compact, locally nearly spherical, manifolds", to appear in *J. reine angew. Math.*

[12] A. Figalli & L. Rifford – "Continuity of optimal transport maps and convexity of injectivity domains on small deformations of $\mathbb{S}^2$", *Comm. Pure Appl. Math.* **62** (2009), p. 1670–1706.

[13] A. Figalli, L. Rifford & C. Villani – "Nearly round spheres look convex", preprint, 2009.

[14] ———, "Necessary and sufficient conditions for continuity of optimal transport maps on Riemannian manifolds", in preparation.

[15] ———, "On the Ma–Trudinger–Wang curvature tensor on surfaces", to appear in *Calc. Var. Partial Differential Equations*.

[16] A. Figalli & C. Villani – "Optimal transport and curvature", Lecture notes, `http://cvgmt.sns.it/cgi/get.cgi/papers/figvil08/OTcurv-6.pdf`.

[17] S. Gallot, D. Hulin & J. Lafontaine – *Riemannian geometry*, third ed., Universitext, Springer, 2004.

[18] J.-i. Itoh & M. Tanaka – "The Lipschitz continuity of the distance function to the cut locus", *Trans. Amer. Math. Soc.* **353** (2001), p. 21–40.

[19] Y.-H. Kim – "Counterexamples to continuity of optimal transport maps on positively curved Riemannian manifolds", *Int. Math. Res. Not.* **2008** (2008), art. ID rnn120, 15.

[20] Y.-H. Kim & R. J. McCann – "Continuity, curvature, and the general covariance of optimal transportation", to appear in *J. Eur. Math. Soc.*

[21] ———, "Towards the smoothness of optimal maps on Riemannian submersions and Riemannian products (of round spheres in particular)", preprint, to appear in *J. reine angew. Math.*

[22] Y. Li & L. Nirenberg – "The distance function to the boundary, Finsler geometry, and the singular set of viscosity solutions of some Hamilton-Jacobi equations", *Comm. Pure Appl. Math.* **58** (2005), p. 85–146.

[23] J. Liu – "Hölder regularity of optimal mappings in optimal transportation", *Calc. Var. Partial Differential Equations* **34** (2009), p. 435–451.

[24] G. Loeper – "On the regularity of solutions of optimal transportation problems", *Acta Math.* **202** (2009), p. 241–283.

[25] ———, "Regularity of optimal maps on the sphere: The quadratic cost and the reflector antenna", to appear in *Arch. Ration. Mech. Anal.*

[26] G. Loeper & C. Villani – "Regularity of optimal transport in curved geometry: the nonfocal case", preprint, to appear in *Duke Math. J.*

[27] X.-N. Ma, N. S. Trudinger & X.-J. Wang – "Regularity of potential functions of the optimal transportation problem", *Arch. Ration. Mech. Anal.* **177** (2005), p. 151–183.

[28] R. J. McCann – "Polar factorization of maps on Riemannian manifolds", *Geom. Funct. Anal.* **11** (2001), p. 589–608.

[29] G. Monge – "Mémoire sur la théorie des déblais et des remblais", *Hist. de l'Acad. des Sciences de Paris* (1781), p. 666–704.

[30] S. T. Rachev & L. Rüschendorf – *Mass transportation problems. Vol. I*, Probability and its Applications (New York), Springer, 1998.

[31] N. S. Trudinger & X.-J. Wang – "On strict convexity and continuous differentiability of potential functions in optimal transportation", *Arch. Ration. Mech. Anal.* **192** (2009), p. 403–418.

[32] ———, "On the second boundary value problem for Monge-Ampère type equations and optimal transportation", *Ann. Sc. Norm. Super. Pisa Cl. Sci.* **8** (2009), p. 143–174.

[33] J. Urbas – "On the second boundary value problem for equations of Monge-Ampère type", *J. reine angew. Math.* **487** (1997), p. 115–124.

[34] C. Villani – *Optimal transport, old and new*, Grundl. Math. Wiss., vol. 338, Springer, 2009.

[35] X.-J. Wang – "On the design of a reflector antenna", *Inverse Problems* **12** (1996), p. 351–375.

Alessio FIGALLI

Department of Mathematics
The University of Texas at Austin
1 University Station C1200
Austin, TX 78712 – USA
E-mail : figalli@math.utexas.edu

332

ASTÉRISQUE

2010

SÉMINAIRE BOURBAKI
VOLUME 2008/2009
EXPOSÉS 997-1011

(1010) *Gauge theory and Langlands duality*

Edward FRENKEL

SOCIÉTÉ MATHÉMATIQUE DE FRANCE
Publié avec le concours du CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

Séminaire BOURBAKI Juin 2009
61e année, 2008-2009, n° 1010, p. 369 à 403

GAUGE THEORY AND LANGLANDS DUALITY

by **Edward FRENKEL**

INTRODUCTION

In the late 1960s Robert Langlands launched what has become known as the Langlands Program with the ambitious goal of relating deep questions in Number Theory to Harmonic Analysis [**39**]. In particular, Langlands conjectured that Galois representations and motives can be described in terms of the more tangible data of automorphic representations. A striking application of this general principle is the celebrated Shimura–Taniyama–Weil conjecture (which implies Fermat's Last Theorem), proved by A. Wiles and others, which says that information about Galois representations associated to elliptic curves over $\mathbb{Q}$ is encoded in the Fourier expansion of certain modular forms on the upper-half plane.

One of the most fascinating and mysterious aspects of the Langlands Program is the appearance of the *Langlands dual group*. Given a reductive algebraic group G, one constructs its Langlands dual ${}^LG$ by applying an involution to its root data. Under the Langlands correspondence, automorphic representations of the group G correspond to Galois representations with values in ${}^LG$.

Surprisingly, the Langlands dual group also appears in Quantum Physics in what looks like an entirely different context; namely, the *electro-magnetic duality*. Looking at the Maxwell equations describing the classical electromagnetism, one quickly notices that they are invariant under the exchange of the electric and magnetic fields. It is natural to ask whether this duality exists at the quantum level. In quantum theory there is an important parameter, the electric charge e. Physicists have speculated that there is an electro-magnetic duality in the quantum theory under which $e \longleftrightarrow 1/e$.

(*) Supported by DARPA through the grant HR0011-09-1-0015 and by Fondation Sciences Mathématiques de Paris.

Under this duality the electrically charged particle should be exchanged with a magnetically charged particle, called magnetic monopole, first theorized by P. Dirac (so far, it has not been discovered experimentally).

In modern terms, Maxwell theory is an example of 4D *gauge theory* (or Yang–Mills theory) which is defined, classically, on the space of connections on various G_c-bundles on a four-manifold M, where G_c is a compact Lie group.[(1)] Electromagnetism corresponds to the simplest, abelian, compact Lie group $U(1)$. It is natural to ask whether there is a non-abelian analogue of the electro-magnetic duality for gauge theories with non-abelian gauge groups.

The answer was proposed in the late 1970s, by Montonen and Olive **[46]**, following Goddard, Nuyts and Olive **[25]** (see also **[12, 50]**). A gauge theory has a coupling constant g, which plays the role of the electric charge e. The conjectural non-abelian electro-magnetic duality, which has later become known as *S-duality*, has the form

$$(G_c, g) \longleftrightarrow ({}^L G_c, 1/g). \tag{0.1}$$

In other words, the duality states that the gauge theory with gauge group G_c (more precisely, its "$N = 4$ supersymmetric" version) and coupling constant g should be equivalent to the gauge theory with the Langlands dual gauge group ${}^L G_c$ and coupling constant $1/g$ (note that if $G_c = U(1)$, then ${}^L G_c$ is also $U(1)$). If true, this duality would have tremendous consequences for quantum gauge theory, because it would relate a theory at small values of the coupling constant (weak coupling) to a theory with large values of the coupling constant (strong coupling). Quantum gauge theory is usually defined as a power series expansion in g, which can only converge for small values of g. It is a very hard problem to show that these series make sense beyond perturbation theory. S-duality indicates that the theory does exist non-perturbatively and gives us a tool for understanding it at strong coupling. That is why it has become a holy grail of modern Quantum Field Theory.

Looking at (0.1), we see that the Langlands dual group shows up again. Could it be that the Langlands duality in Mathematics is somehow related to S-duality in Physics?

This question has remained a mystery until about five years ago. In March of 2004, DARPA sponsored a meeting of a small group of physicists and mathematicians at the Institute for Advanced Study in Princeton (which I co-organized) to tackle this question. At the end of this meeting Edward Witten gave a broad outline of a relation between the two topics. This was explained in more detail in his subsequent joint work **[34]** with Anton Kapustin. This paper, and the work that followed it, opened new

(1) We will use the notation G for a complex Lie group and G_c for its compact form. Note that physicists usually denote by G a compact Lie group and by $G_{\mathbb{C}}$ its complexification.

bridges between areas of great interest for both physicists and mathematicians, leading to new ideas, insights and directions of research.

The goal of these notes is to describe briefly some elements of the emerging picture. In Sections 1 and 2, we will discuss the Langlands Program and its three flavors, putting it in the context of André Weil's "big picture". This will eventually lead us to a formulation of the geometric Langlands correspondence as an equivalence of certain categories of sheaves in Section 3. In Section 4 we will turn to the S-duality in topological twisted $N = 4$ super-Yang–Mills theory. Its dimensional reduction gives rise to the Mirror Symmetry of two-dimensional sigma models associated to the Hitchin moduli spaces of Higgs bundles. In Section 5 we will describe a connection between the geometric Langlands correspondence and this Mirror Symmetry, following **[34]**, as well as its ramified analogue **[26]**. In Section 6 we will discuss subsequent work and open questions.

Acknowledgments. — I thank Sergei Gukov, Vincent Lafforgue, Robert Langlands, and Edward Witten for inspiring discussions. I also thank S. Gukov and V. Lafforgue for their comments on a draft of this paper.

I am grateful to DARPA (and especially Benjamin Mann) for generous support which has been instrumental not only for my research, but for the development of this whole area. I also thank Fondation Sciences Mathématiques de Paris for its support during my stay in Paris.

1. LANGLANDS PROGRAM

In 1940 André Weil was put in jail for his refusal to serve in the army. There, he wrote a letter to his sister Simone Weil (a noted philosopher) in response to her question as to what really interested him in his work **[36]**. This is a remarkable document, in which Weil tries to explain, in fairly elementary terms (presumably, accessible even to a philosopher), the "big picture" of mathematics, the way he saw it. I think this sets a great example to follow for all of us.

Weil writes about the role of *analogy* in mathematics, and he illustrates it by the analogy that interested him the most: between Number Theory and Geometry.

On one side we look at the field $\mathbb{Q}$ of rational numbers and its algebraic closure $\overline{\mathbb{Q}}$, obtained by adjoining all roots of all polynomial equations in one variable with rational coefficients (like $x^2 + 1 = 0$). The group of field automorphisms of $\overline{\mathbb{Q}}$ is the *Galois group* $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$. We are interested in the structure of this group and its finite-dimensional representations. We may also take a more general *number field*—that is, a finite extension F of $\mathbb{Q}$ (such as $\mathbb{Q}(i)$)—and study its Galois group and its representations.

On the other side we have Riemann surfaces: smooth compact orientable surfaces equipped with a complex structure, and various geometric objects associated to them: vector bundles, their endomorphisms, connections, etc.

At first glance, the two subjects are far apart. However, it turns out that there are many analogies between them. The key point is that there is another class of objects which are in-between the two. A Riemann surface may be viewed as the set of points of a projective algebraic curve over $\mathbb{C}$. In other words, Riemann surfaces may be described by algebraic equations, such as the equation

$$y^2 = x^3 + ax + b, \tag{1.1}$$

where $a, b \in \mathbb{C}$. The set of complex solutions of this equation (for generic a, b for which the polynomial on the right hand side has no multiple roots), compactified by a point at infinity, is a Riemann surface of genus 1. However, we may look at the equation (1.1) not only over $\mathbb{C}$, but also over other fields—for instance, over finite fields.

Recall that there is a unique, up to an isomorphism, finite field $\mathbb{F}_q$ of q elements for all q of the form p^n, where p is a prime. In particular, $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z} \simeq \{0, 1, \dots, p-1\}$, with the usual arithmetic modulo p. Let a, b be elements of $\mathbb{F}_q$. Then the equation (1.1) defines a curve over $\mathbb{F}_q$. These objects are clearly analogous to algebraic curves over $\mathbb{C}$ (that is, Riemann surfaces). But there is also a deep analogy with number fields!

Indeed, let X be a curve over $\mathbb{F}_q$ (such as an elliptic curve defined by (1.1)) and F the field of rational functions on X. This *function field* is very similar to a number field. For instance, if X is the projective line over $\mathbb{F}_q$, then F consists of all fractions $P(t)/Q(t)$, where P and Q are two relatively prime polynomials in one variable with coefficients in $\mathbb{F}_q$. The ring $\mathbb{F}_q[t]$ of polynomials in one variable over $\mathbb{F}_q$ is similar to the ring of integers and so the fractions $P(t)/Q(t)$ are similar to the fractions p/q, where $p, q \in \mathbb{Z}$.

Thus, we find a *bridge*, or a "turntable"—as Weil calls it—between Number Theory and Geometry, and that is the theory of algebraic curves over finite fields.

In other words, we can talk about three parallel tracks

Number Theory *Curves over* $\mathbb{F}_q$ *Riemann Surfaces*

Weil's idea is to exploit it in the following way: take a statement in one of the three columns and translate it into statements in the other columns [**36**]: "my work consists in deciphering a trilingual text; of each of the three columns I have only disparate fragments; I have some ideas about each of the three languages: but I know as well there are great differences in meaning from one column to another, for which nothing has prepared me in advance. In the several years I have worked at it, I have found little pieces of the dictionary." Weil went on to find one of the most spectacular

applications of this "Rosetta stone": what we now call the *Weil conjectures* describing analogues of the Riemann Hypothesis in Number Theory in the context of algebraic curves over finite fields.

It is instructive to look at the Langlands Program through the prism of Weil's big picture. Langlands' original formulation **[39]** concerned the two columns on the left. Part of the Langlands Program may be framed as the question of describing n-dimensional representations of the Galois group $\mathrm{Gal}(\overline{F}/F)$, where F is either a number field ($\mathbb{Q}$ or its finite extension) or the function field of a curve over $\mathbb{F}_q$.(2) Langlands proposed that such representations may be described in terms of *automorphic representations* of the group $GL_n(\mathbb{A}_F)$, where $\mathbb{A}_F$ is the ring of adèles of F. I will not attempt to explain this here referring the reader to the surveys **[13, 15, 23]**.

However, it is important for us to emphasize how the Langlands dual group appears in this story. Let us replace $GL_n(\mathbb{A}_F)$ by $G(\mathbb{A}_F)$, where G is a general reductive algebraic group (such as orthogonal or symplectic, or E_8). In the case when $G = GL_n$ its automorphic representations are related to the n-dimensional representations of $\mathrm{Gal}(\overline{F}/F)$, that is, homomorphisms $\mathrm{Gal}(\overline{F}/F) \to GL_n$. The general Langlands conjectures predict that automorphic representations of $G(\mathbb{A}_F)$ are related, in a similar way, to homomorphisms $\mathrm{Gal}(\overline{F}/F) \to {}^LG$, where ${}^LG$ is the Langlands dual group to G.(3)

It is easiest to define ${}^LG$ in the case when G, defined over a field k, is split over k, that is, contains a maximal split torus T (which is the product of copies of the multiplicative group GL_1 over k). We associate to T two lattices: the weight lattice $X^*(T)$ of homomorphisms $T \to GL_1$ and the coweight lattice $X_*(T)$ of homomorphisms $GL_1 \to T$. They contain the sets of roots $\Delta \subset X^*(T)$ and coroots $\Delta^\vee \subset X_*(T)$ of G, respectively. The quadruple $(X^*(T), X_*(T), \Delta, \Delta^\vee)$ is called the root data for G over k. The root data determines the split group G up to an isomorphism.

Let us now exchange the lattices of weights and coweights and the sets of simple roots and coroots. Then we obtain the root data

$$(X_*(T), X^*(T), \Delta^\vee, \Delta)$$

of another reductive algebraic group over $\mathbb{C}$ (or $\overline{\mathbb{Q}}_\ell$), which is denoted by ${}^LG$.(4) Here are some examples:

(2) Langlands' more general "functoriality principle" is beyond the scope of the present article.

(3) More precisely, ${}^LG$ should be defined over $\overline{\mathbb{Q}}_\ell$, where ℓ is relatively prime to q, and we should consider homomorphisms $\mathrm{Gal}(\overline{F}/F) \to {}^LG(\overline{\mathbb{Q}}_\ell)$ which are continuous with respect to natural topology (see, e.g., Section 2.2 of **[15]**).

(4) In Langlands' definition **[39]**, ${}^LG$ also includes the Galois group of a finite extension of F. This is needed for non-split groups, but since we focus here on the split case, this is not necessary.

G	${}^L G$
GL_n	GL_n
SL_n	PGL_n
Sp_{2n}	SO_{2n+1}
$Spin_{2n}$	$SO_{2n}/\mathbb{Z}_2$
E_8	E_8

In the function field case we expect to have a correspondence between homomorphisms[(5)] $\mathrm{Gal}(\overline{F}/F) \to {}^L G$ and automorphic representations of $G(\mathbb{A}_F)$, where $\mathbb{A}_F$ is the ring of adèles of F,

$$\mathbb{A}_F = \prod_{x \in X}{}' F_x,$$

$F_x \simeq \mathbb{F}_{q_x}((t_x))$ being the completion of the field of functions at a closed point x of X, and the prime means that we take the restricted product, in the sense that for all but finitely many x the element of F_x belongs to its ring of integers $\mathcal{O}_x \simeq \mathbb{F}_x[[t_x]]$. We have a natural diagonal inclusion $F \subset \mathbb{A}_F$ an hence $G(F) \subset G(\mathbb{A}_F)$. Roughly speaking, an irreducible representation of $G(\mathbb{A}_F)$ is called automorphic if it occurs in the decomposition of $L^2(G(F)\backslash G(\mathbb{A}_F))$ (with respect to the right action of $G(\mathbb{A}_F)$).

For $G = GL_n$, in the function field case, the Langlands correspondence is a bijection between equivalence classes of irreducible n-dimensional (ℓ-adic) representations of $\mathrm{Gal}(\overline{F}/F)$ (more precisely, the Weil group) and cuspidal automorphic representations of $GL_n(\mathbb{A}_F)$. It has been proved by V. Drinfeld **[8, 9, 10, 11]** for $n = 2$ and by L. Lafforgue **[37]** for $n > 2$. A lot of progress has also been made recently in proving the Langlands correspondence for GL_n in the number field case.

For other groups the correspondence is expected to be much more subtle; for instance, it is not one-to-one. Homomorphisms from the Weil group of F to ${}^L G$ (and more general parameters introduced by J. Arthur, see Section 6.2) should parametrize certain collections of automorphic representations called "L-packets." This has only been proved in a few cases so far.

2. GEOMETRIC LANGLANDS CORRESPONDENCE

The above discussion corresponds to the middle column in the Weil big picture. What should be its analogue in the right column—that is, for complex curves?

In order to explain this, we need a geometric reformulation of the Langlands correspondence which would make sense for curves defined both over a finite field and

(5) More precisely, the Galois group should be replaced by its subgroup called the Weil group.

over $\mathbb{C}$. Thus, we need to find geometric analogues of the notions of Galois representations and automorphic representations.

The former is fairly easy. Let X be a curve over a field k and $F = k(X)$ the field of rational functions on X. If $Y \to X$ is a covering of X, then the field $k(Y)$ of rational functions on Y is an extension of the field $F = k(X)$ of rational functions on X, and the Galois group $\mathrm{Gal}(k(Y)/k(X))$ may be viewed as the group of "deck transformations" of the cover. If our cover is *unramified*, then this group is a quotient of the (arithmetic) fundamental group of X. For a cover ramified at points $x_1, \ldots, x_n$, it is a quotient of the (arithmetic) fundamental group of $X \backslash \{x_1, \ldots, x_n\}$. From now on (with the exception of Section 5.4) we will focus on the unramified case. This means that we replace $\mathrm{Gal}(\overline{F}/F)$ by its maximal unramified quotient, which is nothing but the (arithmetic) fundamental group of X. Its geometric analogue, when X is defined over $\mathbb{C}$, is $\pi_1(X)$.

Thus, the geometric counterpart of a (unramified) homomorphism $\mathrm{Gal}(\overline{F}/F) \to {}^L G$ is a homomorphism $\pi_1(X) \to {}^L G$.

From now on, let X be a smooth projective connected algebraic curve defined over $\mathbb{C}$. Let G be a complex reductive algebraic group and ${}^L G$ its Langlands dual group. Then homomorphisms $\pi_1(X) \to {}^L G$ may be described in differential geometric terms as bundles with a flat connection (the monodromy of the flat connection gives rise to a homomorphism $\pi_1(X) \to {}^L G$). Let E be a smooth principal ${}^L G$-bundle on X. A flat connection on E has two components. The $(0,1)$ component, with respect to the complex structure on X, defines holomorphic structure on E, and the $(1,0)$ component defines a holomorphic connection ∇. Thus, an ${}^L G$-bundle with a flat connection on X is the same as a pair (E, ∇), where E is a holomorphic (equivalently, algebraic) ${}^L G$-bundle on X and ∇ is a holomorphic (equivalently, algebraic) connection on E.

Thus, for complex curves the objects on one side of the Langlands correspondence are equivalence classes of flat (holomorphic or algebraic) ${}^L G$-bundles (E, ∇).

What about the other side? Here the answer is not quite as obvious. I will sketch it briefly referring the reader to Section 3 of **[15]** for more details.

Recall that automorphic representations of $G(\mathbb{A}_F)$ (where F is a function field of a curve X defined over $\mathbb{F}_q$) are realized in functions on the quotient $G(F) \backslash G(\mathbb{A}_F)$. An unramified automorphic representation (which corresponds to an unramified homomorphism $\mathrm{Gal}(\overline{F}/F) \to {}^L G$) gives rise to a function on the double quotient $G(F) \backslash G(\mathbb{A}_F)/G(\mathcal{O}_F)$, where $\mathcal{O}_F = \prod_{x \in X} \mathcal{O}_x$. A key observation (which is due to Weil) is that this double quotient is precisely the set of isomorphism classes of principal G-bundles on our curve X.[6] This statement is also true if the curve X is

[6] From now on we will only consider algebraic bundles.

defined over $\mathbb{C}$. Thus, geometric analogues of unramified automorphic representations should be some geometric objects which "live" on a moduli space of G-bundles.

Unfortunately, for a non-abelian group G there is no algebraic variety whose set of $\mathbb{C}$-points is the set of isomorphism classes of G-bundles on X (for $G = GL_1$ we can take the Picard variety). However, there is an *algebraic moduli stack* denoted by Bun_G. It is not an algebraic variety, but it looks locally like the quotient of an algebraic variety by the action of an algebraic group (these actions are not free, and therefore the quotient is no longer an algebraic variety). It turns out that this is good enough for our purposes.

So which geometric objects on Bun_G will replace unramified automorphic representations? Here we need to recall that the function on the double quotient $G(F)\backslash G(\mathbb{A}_F)/G(\mathcal{O}_F)$ attached to an unramified automorphic representation has a special property: it is an eigenfunction of the so-called Hecke operators. Those are cousins of the classical Hecke operators one studies in the theory of modular forms (which is in the left column of Weil's big picture). The geometric objects we are looking for will be certain sheaves on Bun_G satisfying an analogue of the Hecke property. We will call them *Hecke eigensheaves*.

More precisely, these sheaves are *$\mathcal{D}$-modules* on Bun_G. Recall (see, e.g., **[24, 35]**) that a $\mathcal{D}$-module on a smooth algebraic variety Z is a sheaf of modules over the sheaf $\mathcal{D}_Z$ of differential operators on Z. An example of a $\mathcal{D}$-module is the sheaf of sections of a flat vector bundle on Z. The sheaf of functions on Z acts on sections by multiplication, so it is an $\mathcal{O}_Z$-module. But the flat connection also allows us to act on sections by vector fields on Z. This gives rise to an action of the sheaf $\mathcal{D}_Z$, because it is generated by vector fields and functions. Thus, we obtain the structure of a $\mathcal{D}$-module.

In our case, Bun_G is not a variety, but an algebraic stack, but the (derived) category of $\mathcal{D}$-modules on it has been defined in **[3]**. On this category act the so-called *Hecke functors*. These are labeled by pairs (x, V), where $x \in X$ and V is a finite-dimensional representation of the dual group ${}^L G$, and are defined using certain modifications of G-bundles.

Instead of giving a general definition (which may be found in **[3]** or **[15]**) we will consider two examples. First, consider the abelian case when $G = GL_1$ (thus, we have $G(\mathbb{C}) = \mathbb{C}^\times$). In this case Bun_G may be replaced by the Picard variety Pic which parametrizes line bundles on X. Given a point $x \in X$, consider the map $h_x : \mathrm{Pic} \to \mathrm{Pic}$ sending a line bundle $\mathcal{L}$ to $\mathcal{L}(x)$ (the line bundle whose sections are sections of $\mathcal{L}$ which are allowed to have a pole of order 1 at x). By definition, the Hecke functor $H_{1,x}$ corresponding to x and $1 \in \mathbb{Z}$ (which we identify with the set of one-dimensional

representations of ${}^L G = GL_1$), is given by the formula

$$H_{1,x}(\mathcal{F}) = h_x^*(\mathcal{F}).$$

Next, consider the case of $G = GL_n$ and $V = V_{\check{\omega}_1}$, the defining n-dimensional representation of ${}^L G = GL_n$. In this case Bun_{GL_n} is the moduli stack Bun_n of rank n bundles on X. There is an obvious analogue of the map h_x, sending a rank n bundle $\mathcal{M}$ to $\mathcal{M}(x)$. But then the degree of the bundle jumps by n. It is possible to increase it by 1, but we need to choose a line ℓ in the fiber of $\mathcal{M}$ at x. We then define a new rank n bundle $\mathcal{M}'$ by saying that its sections are the sections of $\mathcal{M}$ having a pole of order 1 at x, but the polar part has to belong to ℓ. Then $\deg \mathcal{M}' = \deg \mathcal{M} + 1$. However, we now have a $\mathbb{P}^{n-1}$ worth of modifications of $\mathcal{M}$ corresponding to different choices of the line ℓ. The Hecke functor $H_{V_{\check{\omega}_1,x}}$ is obtained by "integrating" over all of them.

More precisely, let $\mathcal{H}ecke_{\check{\omega}_1,x}$ be the moduli stack of pairs $(\mathcal{M}, \mathcal{M}')$ as above. It defines a correspondence over $\mathrm{Bun}_n \times \mathrm{Bun}_n$:

$$\begin{array}{ccc} & \mathcal{H}ecke_{\check{\omega}_1,x} & \\ {\scriptstyle h_x^{\leftarrow}} \swarrow & & \searrow {\scriptstyle h_x^{\rightarrow}} \\ \mathrm{Bun}_n & & \mathrm{Bun}_n \end{array} \tag{2.1}$$

By definition,

$$H_{\check{\omega}_1,x}(\mathcal{F}) = h^{\rightarrow}_{x*}\, h^{\leftarrow *}_x(\mathcal{F}). \tag{2.2}$$

For irreducible representations $V_{\check{\lambda}}$ of ${}^L G$ with general dominant integral highest weights $\check{\lambda}$ there is an analogous correspondence in which the role of the projective space $\mathbb{P}^{n-1}$ is played by the Schubert variety in the affine Grassmannian of G corresponding to $\check{\lambda}$ (see **[3, 45]**, and **[15]** for a brief outline).

Allowing the point x to vary, we obtain a correspondence between Bun_G and $X \times \mathrm{Bun}_G$ and Hecke functors acting from the category of $\mathcal{D}$-modules on Bun_G to the (derived) category of $\mathcal{D}$-modules on $X \times \mathrm{Bun}_G$, which we denote by $H_V, V \in \mathrm{Rep}\, {}^L G$.

Now let $\mathcal{E} = (E, \nabla)$ be a flat ${}^L G$-bundle on X. A $\mathcal{D}$-module $\mathcal{F}$ on Bun_G is called a *Hecke eigensheaf* with respect to $\mathcal{E}$ (or with "eigenvalue" $\mathcal{E}$) if we have a collection of isomorphisms

$$H_V(\mathcal{F}) \simeq V_{\mathcal{E}} \boxtimes \mathcal{F}, \tag{2.3}$$

compatible with the tensor product structures. Here

$$V_{\mathcal{E}} = \mathcal{E} \underset{{}^L G}{\times} V$$

is the flat vector bundle on X associated to $\mathcal{E}$ and V, viewed as a $\mathcal{D}$-module. Thus, in particular, we have a collection of isomorphisms

$$H_{V,x}(\mathcal{F}) \simeq V \otimes \mathcal{F}, \qquad x \in X.$$

When we vary the point x, the "eigenvalues", which are all isomorphic to the vector space underlying V, combine into the flat vector bundle $V_{\mathscr{E}}$ on X.

The *geometric Langlands conjecture* may be stated as follows: for any flat ${}^LG$-bundle $\mathscr{E}$ there exists a non-zero $\mathscr{D}$-module $\mathscr{F}_{\mathscr{E}}$ on Bun_G with eigenvalue $\mathscr{E}$.

Moreover, if $\mathscr{E}$ is irreducible, this $\mathscr{D}$-module is supposed to be irreducible (when restricted to each connected component of Bun_G) and unique up to an isomorphism (it should also be holonomic and have regular singularities). But if $\mathscr{E}$ is not irreducible, we might have a non-trivial (derived) category of Hecke eigensheaves, and the situation becomes more subtle.

Thus, at least for irreducible $\mathscr{E}$, we expect the following picture:

(2.4)

flat ${}^LG$-bundles on X	$\longrightarrow$	Hecke eigensheaves on Bun_G

$$\mathscr{E} \quad \longrightarrow \quad \mathscr{F}_{\mathscr{E}}.$$

The geometric Langlands correspondence has been constructed in many cases. For $G = GL_n$ the Hecke eigensheaves corresponding to irreducible $\mathscr{E}$ have been constructed in [**17, 22**], building on the work of P. Deligne for $n = 1$ (explained in [**40**] and [**15**]), V. Drinfeld [**9**] for $n = 2$, and G. Laumon [**40**] (this construction works for curves defined both over $\mathbb{F}_q$ or $\mathbb{C}$).

For all simple algebraic groups G the Hecke eigensheaves have been constructed in a different way (for curves over $\mathbb{C}$) by A. Beilinson and V. Drinfeld [**3**] in the case when $\mathscr{E}$ has an additional structure of an *oper* (this means that $\mathscr{E}$ belongs to a certain half-dimensional locus in $\mathrm{Loc}_{{}^LG}$). It is interesting that this construction is also closely related to quantum field theory, but in a seemingly different way. Namely, it uses methods of 2D Conformal Field Theory and representation theory of affine Kac–Moody algebras of critical level. For more on this, see Part III of [**15**].

3. CATEGORICAL VERSION

Looking at the correspondence (2.4), we notice that there is an essential asymmetry between the two sides. On the left we have flat ${}^LG$-bundles, which are points of a moduli stack $\mathrm{Loc}_{{}^LG}$ of flat ${}^LG$-bundles (or local systems) on X. But on the right we have Hecke eigensheaves, which are objects of a category; namely, the category of $\mathscr{D}$-modules on Bun_G. Beilinson and Drinfeld have suggested a natural way to formulate it in a more symmetrical way.

The idea is to replace a point $\mathcal{E} \in \mathrm{Loc}_{{}^L G}$ by an object of another category; namely, the skyscraper sheaf $\mathcal{O}_{\mathcal{E}}$ at $\mathcal{E}$ viewed as an object of the category of coherent $\mathcal{O}$-modules on $\mathrm{Loc}_{{}^L G}$. A much stronger, categorical, version of the geometric Langlands correspondence is then a conjectural equivalence of derived categories[(7)]

$$\boxed{\begin{array}{c}\text{derived category of}\\ \mathcal{O}\text{-modules on } \mathrm{Loc}_{{}^L G}\end{array}} \quad \longleftrightarrow \quad \boxed{\begin{array}{c}\text{derived category of}\\ \mathcal{D}\text{-modules on } \mathrm{Bun}_G\end{array}} \tag{3.1}$$

This equivalence should send the skyscraper sheaf $\mathcal{O}_{\mathcal{E}}$ on $\mathrm{Loc}_{{}^L G}$ supported at $\mathcal{E}$ to the Hecke eigensheaf $\mathcal{F}_E$. If this were true, it would mean that Hecke eigensheaves provide a good "basis" in the category of $\mathcal{D}$-modules on Bun_G, so we would obtain a kind of spectral decomposition of the derived category of $\mathcal{D}$-modules on Bun_G, like in the Fourier transform. (Recall that under the Fourier transform on the real line the delta-functions δ_x, analogues of $\mathcal{O}_{\mathcal{E}}$, go to the exponential functions e^{itx}, analogues of $\mathcal{F}_{\mathcal{E}}$.)

This equivalence has been proved by G. Laumon [**41**] and M. Rothstein [**51**] in the abelian case, when $G = GL_1$ (or a more general torus). They showed that in this case this is nothing but a version of the Fourier–Mukai transform. Thus, the categorical Langlands correspondence may be viewed as a kind of non-abelian Fourier–Mukai transform (see [**15**], Section 4.4).

Unfortunately, a precise formulation of such a correspondence, even as a conjecture, is not so clear because of various subtleties involved. One difficulty is the structure of $\mathrm{Loc}_{{}^L G}$. Unlike the case of ${}^L G = GL_1$, when all flat bundles have the same groups of automorphisms (namely, GL_1) and Loc_{GL_1} is smooth, for a general group ${}^L G$ the groups of automorphisms are different for different flat bundles, and so $\mathrm{Loc}_{{}^L G}$ is a complicated stack. For example, if ${}^L G$ is a simple Lie group of adjoint type, then a generic flat ${}^L G$-bundle has no automorphisms, while the group of automorphisms of the trivial flat bundle is isomorphic to ${}^L G$. In addition, unlike Bun_G, the stack $\mathrm{Loc}_{{}^L G}$ has singularities. All of this has to be reflected on the other side of the correspondence, in ways that have not yet been fully understood.

Nevertheless, the diagram (3.1) gives us a valuable guiding principle to the geometric Langlands correspondence. In particular, it gives us a natural explanation as to why the skyscraper sheaves on $\mathrm{Loc}_{{}^L G}$ should correspond to Hecke eigensheaves.

The point is that on the category of $\mathcal{O}$-modules on $\mathrm{Loc}_{{}^L G}$ we also have a collection of functors W_V, parametrized by the same data as the Hecke functors H_V. Following

(7) It is expected (see [**19**], Sect. 10) that there is in fact a $\mathbb{Z}_2$-gerbe of such equivalences. This gerbe is trivial, but not canonically trivialized. One gets a particular trivialization of this gerbe, and hence a particular equivalence, for each choice of the square root of the canonical line bundle K_X on X.

physics terminology, we will call them *Wilson functors*. These functors act from the category of $\mathcal{O}$-modules on $\mathrm{Loc}_{{}^L G}$ to the category of sheaves on $X \times \mathrm{Loc}_{{}^L G}$, which are $\mathcal{D}$-modules along X and $\mathcal{O}$-modules along $\mathrm{Loc}_{{}^L G}$.

To define them, observe that we have a tautological ${}^L G$-bundle $\mathcal{T}$ on $X \times \mathrm{Loc}_{{}^L G}$, whose restriction to $X \times \mathcal{E}$, where $\mathcal{E} = (E, \nabla)$, is E. Moreover, ∇ gives us a partial connection on $\mathcal{T}$ along X. For a representation V of ${}^L G$, let $V_{\mathcal{T}}$ be the associated vector bundle on $X \times \mathrm{Loc}_{{}^L G}$, with a connection along X.

Let $p : X \times \mathrm{Loc}_{{}^L G} \to \mathrm{Loc}_{{}^L G}$ be the projection onto the second factor. By definition,

$$W_V(\mathcal{F}) = V_{\mathcal{T}} \otimes p^*(\mathcal{F}) \tag{3.2}$$

(note that by construction $V_{\mathcal{T}}$ carries a connection along X and so the right hand side really is a $\mathcal{D}$-module along X).

Now, the conjectural equivalence (3.1) should be compatible with the Wilson/Hecke functors in the sense that

$$C(W_V(\mathcal{F})) \simeq H_V(C(\mathcal{F})), \qquad V \in \mathrm{Rep}\, {}^L G, \tag{3.3}$$

where C denotes this equivalence (from left to right).

In particular, observe that the skyscraper sheaf $\mathcal{O}_{\mathcal{E}}$ at $\mathcal{E} \in \mathrm{Loc}_{{}^L G}$ is obviously an eigensheaf of the Wilson functors:

$$W_V(\mathcal{O}_{\mathcal{E}}) = V_{\mathcal{E}} \boxtimes \mathcal{O}_{\mathcal{E}}.$$

Indeed, tensoring a skyscraper sheaf with a vector bundle is the same as tensoring it with the fiber of this vector bundle at the point of support of this skyscraper sheaf. Therefore (3.3) implies that $\mathcal{F}_{\mathcal{E}} = C(\mathcal{O}_{\mathcal{E}})$ must satisfy the Hecke property (2.3). In other words, $\mathcal{F}_{\mathcal{E}}$ should be a Hecke eigensheaf on Bun_G with eigenvalue $\mathcal{E}$. Thus, we obtain a natural explanation of the Hecke property of $\mathcal{F}_{\mathcal{E}}$: it follows from the compatibility of the categorical Langlands correspondence (3.1) with the Wilson/Hecke functors.

Let us summarize: the conjectural equivalence (3.1) gives us a natural and convenient framework for the geometric Langlands correspondence. It is this equivalence that Kapustin and Witten have related to the S-duality of 4D super-Yang–Mills.

4. ENTER PHYSICS

We will now add a fourth column to Weil's big picture, which we will call "Quantum Physics":

Number Theory *Curves over* $\mathbb{F}_q$ *Riemann Surfaces* *Quantum Physics*

In the context of the Langlands Program, the last column means S-duality and Mirror Symmetry of certain 4D and 2D quantum field theories, which we will now briefly describe following **[34]**.

We start with the pure 4D Yang–Mills (or gauge) theory on a Riemannian four-manifold M_4. Let G_c be a compact connected simple Lie group. The classical (Euclidean) action is a functional on the space of connections on arbitrary principal G_c-bundles $\mathscr{P}$ on M_4 given by the formula

$$I = \frac{1}{4g^2}\int_{M_4} \operatorname{Tr} F_A \wedge \star F_A + \frac{i\theta}{8\pi^2}\int_{M_4} \operatorname{Tr} F_A \wedge F_A.$$

Here F_A is the curvature of the connection A (a $\mathfrak{g}$-valued two-form on M_4), $\star$ is the Hodge star operator, and Tr is the invariant bilinear form on the Lie algebra $\mathfrak{g}$ normalized in such a way that the second term is equal to $i\theta k$, where k could be an arbitrary integer, if G_c is simply-connected. The second term is equal to $i\theta$ times the second Chern class $c_2(\mathscr{P})$ of the bundle $\mathscr{P}$ and hence is topological. Correlation functions are given by path integrals of the form $\int e^{-I}$ over the space of connections modulo gauge transformations. Hence they may be written as Fourier series in $e^{i\theta}$ (or its root if G_c is not simply-connected) such that the coefficient in front of $e^{i\theta n}$ is the sum of contributions from bundles $\mathscr{P}$ with $c_2(\mathscr{P}) = -n$.

It is customary to combine the two parameters, g and θ, into one complex coupling constant

$$\tau = \frac{\theta}{2\pi} + \frac{4\pi i}{g^2}.$$

Next, we consider $N = 4$ supersymmetric extension of this model. This means that we add fermionic and bosonic fields in such a way that the action of the Lorentz group (we will work in Euclidean signature, where this group becomes $SO(4)$) is extended to an action of an appropriate supergroup (see **[34]** or the books **[6]** for a background on supersymmetric quantum field theory).

The S-duality of this theory is the statement that the theory with gauge group G_c and complex coupling constant τ is equivalent to the theory with the Langlands dual gauge group ${}^L G_c$ and coupling constant ${}^L\tau = -1/n_{\mathfrak{g}}\tau$:

$$(G_c, \tau) \longleftrightarrow ({}^L G_c, -1/n_{\mathfrak{g}}\tau), \tag{4.1}$$

where $n_{\mathfrak{g}}$ is the lacing number of the Lie algebra $\mathfrak{g}$ (equal to 1 for simply-laced Lie algebras, 2 for B_n, C_n and F_4, and 3 for G_2). This is an extension of the duality (0.1) of **[46]** discussed in the introduction to non-zero values of θ (with g normalized in a slightly different way). In addition, for simply-laced G_c the path integral is a Fourier series in $e^{i\theta}$, so θ may be shifted by an integer multiple of 2π without changing the path integral. Thus, we also have the equivalence

$$(G_c, \tau) \longleftrightarrow (G_c, \tau + 1).$$

For general non-simply connected groups we have instead a symmetry $\tau \mapsto \tau + n\mathbb{Z}$, where n is a certain integer. Thus, we obtain an action of a subgroup of $SL_2(\mathbb{Z})$ on the super-Yang–Mills theories with gauge groups G_c and ${}^LG_c$.[8] As we discussed in the Introduction, this is a striking statement because it relates a theory at strong coupling to a theory at weak coupling.

We want to focus next on a "topological sector" of this theory. This means that we pick an element Q in the Lie superalgebra $\mathfrak{s}$ of the super-Lorentz group (the supergroup extension of $SO(4)$) such that $Q^2 = 0$, and such that the stress tensor (which is a field responsible for variation of the metric on M_4) is equal to the commutator of Q and another field. Let us restrict ourselves to those objects (fields, boundary conditions, etc.) in the theory which commute with this Q. This is a particular (and relatively small) sector of the full quantum field theory, in which all quantities (such as correlation functions) are topological, that is, metric-independent. This sector is what is usually referred to as Topological Field Theory (TFT).

There is a problem, however. For this $Q \in \mathfrak{s}$ to be well-defined on an arbitrary manifold M_4, it has to be invariant under the action of the Lorentz group $SO(4)$—more precisely, its double cover $Spin(4)$. Unfortunately, there are no such elements in our Lie superalgebra $\mathfrak{s}$. In order to obtain such an element, one uses a trick, called *twisting* (see, e.g., **[59]**). Our theory has an additional group of automorphisms commuting with the action of $Spin(4)$, called R-symmetry; namely, the group $Spin(6)$. We can use it to modify the action of $Spin(4)$ on the fields of the theory and on the Lie superalgebra $\mathfrak{s}$ as follows: define a new action of $Spin(4)$ equal to the old action together with the action coming from a homomorphism $Spin(4) \to Spin(6)$ and the action of $Spin(6)$ by R-symmetry. One might then be able to find a differential $Q \in \mathfrak{s}$ invariant under this new action of $Spin(4)$.

There are essentially three different choices for doing this, as explained in **[55]**. The first two are similar to the twists used in Witten's construction of a topological field theory that yields Donaldson invariants of four-manifolds (which is a topological twist of an $N = 2$ supersymmetric Yang–Mills theory) **[56]**. It is the third twist, studied in detail in **[34]**, that is relevant to the geometric Langlands. For this twist there are actually two linearly independent (and anti-commuting with each other) operators, Q_l and Q_r, which square to 0. We can therefore use any linear combination

$$Q = uQ_l + vQ_r$$

as the differential defining the topological field theory (for each of them the stress tensor will be a Q-commutator, so we will indeed obtain a topological field theory).

[8] In general, it is a proper subgroup of $SL_2(\mathbb{Z})$ for two reasons: first, we have the coefficient $n_{\mathfrak{g}}$ in formula (4.1) for non-simply laced G_c, and second, the dual of a simply-connected Lie group is not simply-connected in general, in which case the transformation $\tau \to \tau + 1$ is not a symmetry.

We obtain the same theory if we rescale both u and v by the same number. Hence we obtain a family of topological field theories parametrized by $\mathbb{P}^1$.

Let $t = v/u$ be a coordinate on this $\mathbb{P}^1$. We will refer to it as the "twisting parameter." The S-duality (4.1) should be accompanied by the change of the twisting parameter according to the rule

$$t \mapsto \frac{\tau}{|\tau|} t. \tag{4.2}$$

How can we test S-duality of these topological theories? Vafa and Witten have earlier tested the S-duality of a different (Donaldson type) topological twisting of $N = 4$ super-Yang–Mills by showing that the partition functions of these theories (depending on τ) are modular forms **[55]**. This proves the invariance of the partition functions under the action of a subgroup of $SL_2(\mathbb{Z})$ on τ. What turns out to be relevant to the geometric Langlands Program is the study of boundary conditions in these topological field theories.

Kapustin and Witten assume that the four-manifold M_4 has the form

$$M_4 = \Sigma \times X,$$

where X is a closed Riemann surface (this will be the algebraic curve of the geometric Langlands) and Σ is a Riemann surface with a boundary (which we may simply take to be a half-plane). They study the limit of the topological gauge theory on this manifold when X becomes very small (this is called "compactification of the theory on X"). In this limit the theory is described by an effective *two-dimensional* topological field theory on Σ. In earlier works **[4, 28]** the latter theory was identified with the (twisted) topological sigma model on Σ with the target manifold $\mathcal{M}_H(G)$, the *Hitchin moduli space* of Higgs G-bundles on X. Moreover, the S-duality of the supersymmetric gauge theories on $\Sigma \times X$ (for particular values of τ and t) becomes Mirror Symmetry between the topological sigma models with the targets $\mathcal{M}_H(G)$ and $\mathcal{M}_H({}^L G)$.

Next, we look at the boundary conditions in the gauge theories, which give rise to *branes* in these sigma models. S-duality yields an equivalence of the categories of branes for $\mathcal{M}_H(G)$ and $\mathcal{M}_H({}^L G)$ (also known, after M. Kontsevich, as Homological Mirror Symmetry). Kapustin and Witten have related this equivalence to the categorical geometric Langlands correspondence (3.1). Thus, they establish a link between S-duality and geometric Langlands duality. We describe this in more detail in the next section.

5. MIRROR SYMMETRY OF HITCHIN MODULI SPACES

In [**30**] N. Hitchin introduced a remarkable hyper-Kähler manifold $\mathcal{M}_H(G)$ for each smooth projective complex algebraic curve X and reductive Lie group G. It is easiest to describe it in its complex structure I, in which it is the moduli space of semi-stable *Higgs bundles* on X. Recall that a Higgs G-bundle on X is a pair (E, ϕ), where E is a (algebraic) G-bundle on X and ϕ is a Higgs field on it, that is,

$$\phi \in H^0(X, \mathfrak{g}_E \otimes K_X),$$

where $\mathfrak{g}_E = E \underset{G}{\times} \mathfrak{g}$ is the adjoint vector bundle.

In the complex structure J, however, $\mathcal{M}_H(G)$ is described as the moduli space of semi-stable *flat bundles*, that is, pairs (E, ∇), where E is again (algebraic) G-bundle on X and ∇ is (algebraic) connection on E. To distinguish between it as a complex algebraic variety from the moduli space of Higgs bundles we will denote it by $\mathcal{Y}(G)$. The two are isomorphic as real manifolds (this is the statement of non-abelian Hodge theory [**5, 30, 52**]), but not as complex (or algebraic) manifolds.

There are two types of twisted supersymmetric two-dimensional sigma models with Kähler target manifolds: A-model and B-model (see [**57**]). The former depends on the symplectic structure on the target manifold and the latter depends on the complex structure.

Kapustin and Witten start with two topological twisted super-Yang–Mills theories on $\Sigma \times X$. One has gauge group G_c, twisting parameter $t = 1$, and $\theta = 0$. The other, S-dual theory, has gauge group ${}^L G_c$, the twisting parameter ${}^L t = i$, and ${}^L \theta = 0$ (neither of these topological theories depends on g [**34**]).[(9)] They show that after compactification on X the first theory becomes the A-model with the target manifold $\mathcal{M}_H(G)$ and the symplectic structure ω_K, which is the Kähler form for the complex structure K on $\mathcal{M}_H(G)$. This symplectic structure has a nice geometric description. Note that the Higgs field ϕ is an element of $H^0(X, \mathfrak{g}_E \otimes K_X)$, which is isomorphic to the cotangent space to E, viewed as a point of Bun_G, the moduli stack of G-bundles on X. Thus, $\mathcal{M}_H(G)$ is almost the cotangent bundle to Bun_G; "almost" because we impose the semi-stability condition on the Higgs bundle. The symplectic form ω_K comes from the standard symplectic form on the cotangent bundle (which is the imaginary part of the holomorphic symplectic form).

The second gauge theory becomes, after compactification on X, the B-model with the target manifold $\mathcal{Y}({}^L G)$; that is, $\mathcal{M}_H({}^L G)$ with respect to the complex structure J.

After dimensional reduction from 4D to 2D, the S-duality of super-Yang–Mills theories becomes Mirror Symmetry between the A-model with the target manifold

(9) As explained in [**33, 34**], for some other values of parameters one obtains the so-called quantum geometric Langlands correspondence (see [**15**], Section 6.3).

$\mathcal{M}_H(G)$ (and symplectic structure ω_K) and the B-model with the target manifold $\mathcal{Y}({}^L G)$ (and complex structure J).

REMARK. — *As explained in* **[34]**, *there is also Mirror Symmetry between A- and B-models with respect to other symplectic and complex structures. For instance, there is Mirror Symmetry, studied in* **[1, 7, 32]**, *between the B-models on* $\mathcal{M}_H(G)$ *and* $\mathcal{M}_H({}^L G)$ *with respect to the complex structures I on both of them. In what follows we will not discuss these additional dualities.*

5.1. Dual Hitchin fibrations

In order to understand better this Mirror Symmetry, we recall the construction of the *Hitchin map*. For any Higgs bundle (E, ϕ) and an invariant polynomial P of degree d on the Lie algebra $\mathfrak{g}$, we can evaluate P on ϕ and obtain a well-defined section $P(\phi)$ of $K_X^{\otimes d}$. The algebra $(\mathrm{Fun}(\mathfrak{g}))^G$ of invariant polynomial functions on $\mathfrak{g}$ is a graded free polynomial algebra with $\ell = \mathrm{rank}(\mathfrak{g})$ generators of degrees $d_i, i = 1, \dots, \ell$, where the d_i are the exponents of $\mathfrak{g}$ plus 1. Let us choose a set of generators $P_i, i = 1, \dots, \ell$. Then we construct the Hitchin map **[30, 31]**

$$p : \mathcal{M}_H(G) \to \mathbf{B} = \bigoplus_{i=1}^{\ell} H^0(X, K_X^{\otimes d_i}),$$
$$(E, \phi) \mapsto (P_1(\phi), \dots, P_\ell(\phi)).$$

This is slightly non-canonical, because there is no canonical choice of generators P_i in general. More canonically, we have a map to

$$\mathbf{B} := H^0(X, (\mathfrak{g}//G)_{K_X}), \qquad (\mathfrak{g}//G)_{K_X} := K_X^\times \underset{\mathbb{C}^\times}{\times} \mathfrak{g}//G,$$

where $K_X^\times$ denotes the $\mathbb{C}^\times$-bundle associated to K_X, and $\mathfrak{g}//G := \mathrm{Spec}((\mathrm{Fun}(\mathfrak{g}))^G)$ is the graded vector space on which the P_i are coordinate functions (the $\mathbb{C}^\times$-action on it comes from the grading).

By Chevalley's theorem, $\mathfrak{g}//G = \mathfrak{h}//W := \mathrm{Spec}((\mathrm{Fun}(\mathfrak{h}))^W)$, where $\mathfrak{h}$ is a Cartan subalgebra. By definition, ${}^L\mathfrak{h} = \mathfrak{h}^*$. Hence ${}^L\mathfrak{g}//{}^L G = {}^L\mathfrak{h}//W = \mathfrak{h}^*//W$. Choosing a non-degenerate invariant bilinear form κ_0 on $\mathfrak{g}$, we identify $\mathfrak{h}$ and $\mathfrak{h}^*$, and hence $\mathbf{B}$ and ${}^L\mathbf{B}$. Any other invariant bilinear form is proportional to κ_0. Hence replacing κ_0 by another non-zero bilinear form would correspond to a $\mathbb{C}^\times$-action on the base. But this action can be lifted to a $\mathbb{C}^\times$-action on the total space $\mathcal{M}_H(G)$ (rescaling the Higgs field ϕ). Hence the ambiguity in the choice of κ_0 is not essential; it may be absorbed into an automorphism of one of the two Hitchin moduli spaces.

As the result, we obtain two fibrations over the same base $\mathbf{B}$:

$$\begin{array}{ccc} \mathcal{M}_H({}^L G) & & \mathcal{M}_H(G) \\ & \searrow \quad \swarrow & \\ & \mathbf{B} & \end{array} \tag{5.1}$$

For generic $b \in \mathbf{B}$ (the connected components of) the fibers ${}^L\mathbf{F}_b$ and $\mathbf{F}_b$ of these Hitchin fibrations are smooth tori, which are in fact isomorphic to abelian varieties in the complex structure I. For instance, for $G = SL_n$, $\mathbf{F}_b$ is the generalized Prym variety of the spectral curve associated to b, which is a smooth degree n cover of X if b is generic.

Moreover, the tori ${}^L\mathbf{F}_b$ and $\mathbf{F}_b$ (again, for generic b) are *dual* to each other. This can be expressed in the following way which will be convenient for our purposes: there is a bijection between points of ${}^L\mathbf{F}_b$ and flat unitary line bundles on $\mathbf{F}_b$.[10] The duality of tori is the simplest (abelian) example of Mirror Symmetry, also known as *T-duality*. Thus, we expect that the Mirror Symmetry of the Hitchin moduli spaces $\mathcal{M}_H({}^L G)$ and $\mathcal{M}_H(G)$ is realized via fiberwise T-duality (for generic fibers of the dual Hitchin fibrations). This is an example of a general proposal of Strominger, Yau and Zaslow [**54**] that Mirror Symmetry of two Calabi–Yau manifolds X and Y should be realized as T-duality of generic fibers in special Lagrangian fibrations of X and Y (what happens for the singular fibers is *a priori* far less clear; but see Section 6.1). It is in this sense that "S-duality reduces to T-duality" [**28**].

We are interested in the study of the B-model with the target $\mathcal{M}_H({}^L G)$, with respect to the complex structure J (that is, the moduli space $\mathcal{Y}({}^L G)$ of flat bundles), and the A-model with the target $\mathcal{M}_H(G)$, with respect to the symplectic structure ω_K. These two topological field theories are expected to be equivalent to each other. Therefore anything we can say about one of them should have a counterpart in the other. For instance, their cohomologies may be interpreted as the spaces of vacua in these field theories, and hence they should be isomorphic. This has indeed been verified by Hausel and Thaddeus [**29**] in the case when $G = SL_n, {}^L G = PGL_n$ (since the Hitchin moduli spaces are non-compact, special care has to be taken to properly define these cohomologies, see [**29**]).

To make contact with the geometric Langlands correspondence, Kapustin and Witten study in [**34**] the *categories of branes* in these two topological field theories.

[10] This bijection depends on the choice of base points in the fibers. Using the Hitchin section, one obtains such base points, but for this one may have to choose a square root of the canonical line bundle K_X. This is closely related to the $\mathbb{Z}_2$-gerbe ambiguity in the equivalence (3.1) discussed in the footnote on page 379.

5.2. Categories of branes

Branes in two-dimensional sigma models are certain generalizations of boundary conditions. When writing path integral for maps $\Phi : \Sigma \to M$, where Σ has a boundary, we need to specify boundary conditions for Φ on $\partial\Sigma$. We may also "couple" the sigma model to another quantum field theory on $\partial\Sigma$ (that is, modify the action by a boundary term) which may be interpreted as a decoration of the boundary condition. In topological field theory these conditions should preserve the supersymmetry, which leads to natural restrictions.

A typical example of a boundary condition is specifying that $\Phi(\partial\Sigma)$ belongs to a submanifold $M' \subset M$. In the B-model the target manifold M is a complex manifold, and in order to preserve the supersymmetry M' has to be a complex submanifold. In the A-model, M is a symplectic manifold and M' should be Lagrangian. Coupling to field theories on $\partial\Sigma$ allows us to introduce into the picture a holomorphic vector bundle on M' in the case of B-model, and a flat unitary vector bundle on M' in the case of A-model.

More generally, the category of branes in the B-model with a complex target manifold M (called B-branes) is the (derived) category of coherent sheaves on M, something that is fairly well understood mathematically. The category of branes in the A-model with a symplectic target manifold M (called A-branes) is less understood. It is believed to contain what mathematicians call the Fukaya category, typical objects of which are pairs (L, ∇), where $L \subset M$ is a Lagrangian submanifold and ∇ is a flat unitary vector bundle on L. However (and this turns out to be crucial for applications to the geometric Langlands), it also contains more general objects, such as coisotropic submanifolds of M equipped with vector bundles with unitary connection.

Under the Mirror Symmetry between the sigma models with the target manifolds $\mathcal{Y}({}^L G)$ and $\mathcal{M}_H(G)$ we therefore expect to have the following equivalence of (derived) categories of branes (often referred to, after Kontsevich, as Homological Mirror Symmetry):

$$\boxed{B\text{-branes on } \mathcal{Y}({}^L G)} \quad \longleftrightarrow \quad \boxed{A\text{-branes on } \mathcal{M}_H(G)} \tag{5.2}$$

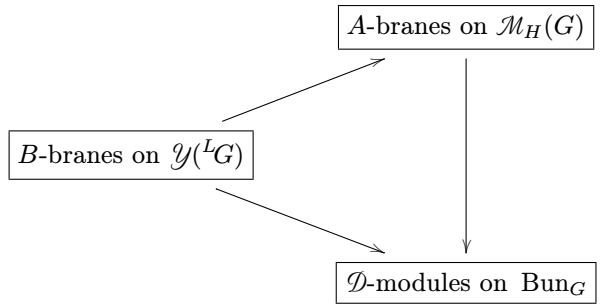
It is this equivalence that Kapustin and Witten have related to the categorical Langlands correspondence (3.1). The category on the left in (5.2) is the (derived) category of coherent sheaves on $\mathcal{Y}({}^L G)$, which is the moduli space of semi-stable flat ${}^L G$-bundles on X. It is closely related to the (derived) category of coherent sheaves (or, equivalently, $\mathcal{O}$-modules) on $\mathrm{Loc}_{{}^L G}$, which appears on the left of (3.1). The difference is that, first of all, $\mathrm{Loc}_{{}^L G}$ is the moduli *stack* of flat ${}^L G$-bundles on X, whereas $\mathcal{Y}({}^L G)$ is the moduli *space* of semi-stable ones. Second, from the physics perspective it is more natural to consider coherent sheaves on $\mathcal{Y}({}^L G)$ with respect to its complex analytic rather than algebraic structure, whereas in (3.1) we consider algebraic $\mathcal{O}$-modules on

$\mathrm{Loc}_{^LG}$. These differences aside, these two categories are very similar to each other. They certainly share many objects, such as skyscraper sheaves supported at points corresponding to stable flat ^LG-bundles which we will discuss momentarily.

5.3. Triangle of equivalences

The categories on the right in (3.1) and (5.2) appear at first glance to be quite different. But Kapustin and Witten have suggested that they should be equivalent to each other as well. Thus, we obtain the following triangle of derived categories:

(5.3)



The upper arrow represents Homological Mirror Symmetry (5.2) whereas the lower arrow represents the categorical Langlands correspondence (3.1).

According to **[34]**, Section 11, the vertical arrow is another equivalence that has nothing to do with either Mirror Symmetry or geometric Langlands. It should be a general statement linking the (derived) category of $\mathcal{D}$-modules on a variety M and the (derived) category of A-branes on its cotangent bundle T^*M (recall that $\mathcal{M}_H(G)$ is almost equal to $T^*\,\mathrm{Bun}_G$). Kapustin and Witten have proposed the following functor from the category of A-branes on T^*M with respect to the symplectic structure $\mathrm{Im}\,\Omega$ (where Ω is the holomorphic symplectic form on T^*M) to the category of $\mathcal{D}$-modules on M:

$$\mathcal{A} \mapsto \mathrm{Hom}(\mathcal{A}_{cc}, \mathcal{A}), \tag{5.4}$$

where $\mathcal{A}_{cc}$ is a "canonical coisotropic brane" on T^*M. This is T^*M itself (viewed as a coisotropic submanifold) equipped with a line bundle with connection satisfying special properties. They argued on physical grounds that the right hand side of (5.4) may be "sheafified" along M, and moreover that the corresponding sheaf of rings $\mathrm{Hom}(\mathcal{A}_{cc}, \mathcal{A}_{cc})$ is nothing but the sheaf of differential operators on $\mathcal{M}_H(G)$.[(11)] Hence the (sheafified) right hand side of (5.4) should be a $\mathcal{D}$-module. While this argument has not yet been made mathematically rigorous, it allows one to describe important characteristics of the $\mathcal{D}$-module associated to an A-brane, such as its reducibility, the

[(11)] More precisely, it is the sheaf of differential operators acting on a square root of the canonical line bundle on Bun_G, but we will ignore this subtlety here.

open subset of M where it is represented by a local system, the rank of this local system, and even its monodromy (see Section 4 of [**19**]).

An alternative (and mathematically rigorous) approach to establishing an equivalence between the categories of A-branes and $\mathcal{D}$-modules has also been proposed by D. Nadler and E. Zaslow [**47, 48**].

Though a lot of work still needs to be done to distill this connection and reconcile different approaches, this is clearly a very important and beautiful idea on its own right.

Thus, according to Kapustin and Witten, the (categorical) geometric Langlands correspondence (3.1) may be obtained in two steps. The first step is the Homological Mirror Symmetry (5.2) of the Hitchin moduli spaces for two dual groups, and the second step is the above link between the A-branes and $\mathcal{D}$-modules.

There is actually more structure in the triangle (5.3). On each of these three categories we have an action of certain functors, and all equivalences between them are supposed to commute with these functors. We have already described the functors on two of these categories in Section 3: these are the Wilson and Hecke functors. The functors acting on the categories of A-branes, were introduced in [**34**] as the two-dimensional shadows of the 't Hooft loop operators in 4D super-Yang–Mills theory. Like the Hecke functors, they are defined using modifications of G-bundles, but only those modifications which preserve the Higgs field. The S-duality of super-Yang–Mills theories is supposed to exchange the 't Hooft operators and the Wilson operators (whose two-dimensional shadows are the functors described in Section 3), and this is the reason why we expect the equivalence (5.2) to commute with the action of these functors.

As explained above, the central objects in the geometric Langlands correspondence are Hecke eigensheaves attached to flat ${}^L G$-bundles. Recall that the Hecke eigensheaf $\mathcal{F}_\mathcal{E}$ is the $\mathcal{D}$-module attached to the skyscraper sheaf $\mathcal{O}_\mathcal{E}$ supported at a point $\mathcal{E}$ of $\mathrm{Loc}_{{}^L G}$ under the conjectural equivalence (3.1). These $\mathcal{D}$-modules have very complicated structure. What can we learn about them from the point of view of Mirror Symmetry?

Let us assume first that $\mathcal{E}$ has no automorphisms other than those coming from the center of ${}^L G$. Then it is a smooth point of $\mathcal{Y}({}^L G)$. These skyscraper sheaves $\mathcal{O}_\mathcal{E}$ are the simplest examples of B-branes on $\mathcal{Y}({}^L G)$ (called 0-branes). What is the corresponding A-brane on $\mathcal{M}_H(G)$?

The answer is surprisingly simple. Let $b \in \mathbf{B}$ be the projection of $\mathcal{E}$ to the base of the Hitchin fibration. For $\mathcal{E}$ satisfying the above conditions the Hitchin fiber ${}^L \mathbf{F}_b$ is a smooth torus (it is actually an abelian variety in the complex structure I, but now we look at it from the point of view of complex structure J, so it is just a smooth torus). It is identified (possibly, up to a choice of the square root of the canonical line

bundle K_X, see the footnote on page 386) with the moduli space of flat unitary line bundles on the dual Hitchin fiber $\mathbf{F}_b$, which happens to be a Lagrangian submanifold of $\mathcal{M}_H(G)$. The Mirror Symmetry sends the B-brane $\mathcal{O}_{\mathcal{E}}$ to the A-brane which is the pair $(\mathbf{F}_b, \nabla_{\mathcal{E}})$, the Lagrangian submanifold $\mathbf{F}_b$ of $\mathcal{M}_H(G)$, together with the flat unitary line bundle on it corresponding to $\mathcal{E}$:

$$\mathcal{O}_{\mathcal{E}} \mapsto (\mathbf{F}_b, \nabla_{\mathcal{E}}).$$

Since $\mathcal{O}_{\mathcal{E}}$ is obviously an eigenbrane of the Wilson functors (as we discussed in Section 3), the A-brane $(\mathbf{F}_b, \nabla_{\mathcal{E}})$ should be an eigenbrane of the 't Hooft functors. This may in fact be made into a precise mathematical conjecture, and Kapustin and Witten have verified it explicitly in some cases.

Thus, the A-branes associated to the simplest B-branes turn out to be very nice and simple. This is in sharp contrast with the structure of the corresponding $\mathcal{D}$-modules, which is notoriously complicated in the non-abelian case. Therefore the formalism of A-branes developed in [**34**] has clear advantages. It replaces $\mathcal{D}$-modules with A-branes that are much easier to "observe experimentally" and to analyze explicitly. One can hope to use this new language in order to gain insights into the structure of the geometric Langlands correspondence. It has already been used in [**19**] for understanding what happens in the endoscopic case as explained in the next section.

5.4. Ramification

Up to now we have considered the *unramified* case of the geometric Langlands correspondence, in which the objects on the Galois side of the correspondence are holomorphic ${}^LG$-bundles on our curve X with a holomorphic connection. These flat bundles give rise to homomorphisms $\pi_1(X) \to {}^LG$. In the classical Langlands correspondence one looks at more general homomorphisms $\pi_1(X\backslash\{x_1, \dots, x_n\}) \to {}^LG$. Thus, we look at holomorphic ${}^LG$-bundles on X with *meromorphic* connections which have poles at finitely many points of X. The connections with poles of order one (regular singularities) correspond to *tame* ramification in the classical Langlands Program. Those with poles of orders higher than one (irregular singularities) correspond to *wild* ramification.

Mathematically, the ramified geometric Langlands correspondence has been studied in [**16**] and follow-up papers (see [**14**] for an exposition), using the affine Kac–Moody algebras of critical level and generalizing the Beilinson–Drinfeld approach [**3**] to allow ramification.

S. Gukov and E. Witten [**26**] have explained how to include tame ramification in the S-duality picture. Physicists have a general way of including into a quantum field theory on a manifold M objects supported on submanifolds of M. An example of this is the *surface operators* in 4D super-Yang–Mills theory, supported on two-dimensional

submanifolds of the four-manifold M_4. If we include such an operator, we obtain a certain modification of the theory. Let us again take $M_4 = \Sigma \times X$ and take this submanifold to be of the form $\Sigma \times x, x \in X$. Gukov and Witten show that for a particular class of surface operators the dimensional reduction of the resulting theory is the sigma model on Σ with a different target manifold $\mathcal{M}_H(G, x)$, the moduli space of semi-stable Higgs bundles with regular singularity at $x \in X$. It is again hyper-Kähler, and in the complex structure J it has a different incarnation as the moduli space of semi-stable bundles with a connection having regular singularity (see [**53**]).

The moduli space $\mathcal{M}_H(G, x)$ has parameters (α, β, γ), which lie in the (compact) Cartan subalgebra of $\mathfrak{g}$ (see [**53**]). For generic parameters, this moduli space parametrizes semi-stable triples $(E, \phi, \mathcal{L})$, where E is a (holomorphic) G-bundle, ϕ is a Higgs field which has a pole at x of order one whose residue belongs to the regular semi-simple conjugacy class of $\frac{1}{2}(\beta + i\gamma)$, and $\mathcal{L}$ is a flag in the fiber of E at x which is preserved by this residue (the remaining parameter α determines the flag). Various degenerations of parameters give rise to similar moduli spaces in which the residues of the Higgs fields could take arbitrary values. The moduli spaces $\mathcal{M}_H(G, x)$ and $\mathcal{M}_H({}^L G, x)$ (with matching parameters) are equipped with a pair of mirror dual Hitchin fibrations, and the Mirror Symmetry between them is again realized as fiberwise T-duality (for generic fibers which are again smooth dual tori).

The S-duality of the super-Yang–Mills theories, associated to the dual groups G_c and ${}^L G_c$, with surface operators gives rise to an equivalence of categories of A- and B-branes on $\mathcal{M}_H(G, x)$ and $\mathcal{M}_H({}^L G, x)$. The mirror dual for a generic 0-brane on $\mathcal{M}_H({}^L G, x)$ is the A-brane consisting of a Hitchin fiber and a flat unitary line bundle on it, as in the unramified case.

The analysis of [**26**] leads to many of the same conclusions as those obtained in [**16**] by using representations of affine Kac–Moody algebras and two-dimensional conformal field theory.

Gukov and Witten also considered [**27**] more general surface operators associated to coadjoint orbits in $\mathfrak{g}$ and ${}^L\mathfrak{g}$. The S-duality between these surface operators leads to some non-trivial and unexpected relations between these orbits. Gukov and Witten present many interesting examples of this in [**27**] drawing connections with earlier work done by mathematicians.

In [**58**], Witten has generalized the analysis of [**26, 27**] to the case of wild ramification.

6. MORE GENERAL BRANES

In the previous section we discussed applications of Mirror Symmetry of the dual Hitchin fibrations to the geometric Langlands correspondence. We saw that the A-branes associated to the B-branes supported at the generic flat ${}^LG$-bundles have very simple description: these are the Hitchin fibers equipped with flat unitary line bundles. But what about the B-branes supported at more general flat ${}^LG$-bundles? Can we describe explicitly the A-branes dual to them?

This question goes to the heart of the subtle interplay between physics and mathematics of Langlands duality. Trying to answer this question, we will see the limitations of the above analysis and a way for its generalization incorporating more general branes. This will lead us to surprising physical interpretation of deep mathematical concepts such as endoscopy and Arthur's SL_2.

The generic flat ${}^LG$-bundles, for which Mirror Symmetry works so nicely, are the ones that have no automorphisms (apart from those coming from the center of ${}^LG$). They correspond to smooth points of $\mathcal{Y}({}^LG)$ such that the corresponding Hitchin fiber is also smooth. We should consider next the singularities of $\mathcal{Y}({}^LG)$. The simplest of those are the orbifold singularities. We will discuss them, and their connection to endoscopy, in the next subsection, following [**19**]. We will then talk about more general singularities corresponding to flat ${}^LG$-bundles with continuous groups of automorphisms, and what we can learn about the corresponding categories from physics.

6.1. Geometric Endoscopy

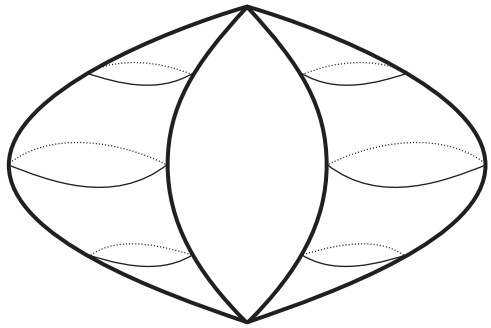
We start with the mildest possible singularities in $\mathcal{Y}({}^LG)$; namely, the orbifold singular points. The corresponding flat ${}^LG$-bundles are those having finite groups of automorphisms (modulo the center). In the classical Langlands correspondence the analogous Galois representations are called endoscopic. They, and the corresponding automorphic representations, play an important role in the stabilization of the trace formula.

The simplest example, analyzed in [**19**] and dubbed "geometric endoscopy", arises when ${}^LG = PGL_2$, which contains $O_2 = \mathbb{Z}_2 \ltimes \mathbb{C}^\times$ as a subgroup. Suppose that a flat PGL_2-bundle $\mathcal{E}$ on our curve X is reduced to this subgroup. Then generically it will have the group of automorphisms $\mathbb{Z}_2 = \{1, -1\} \subset \mathbb{C}^\times$, which is the center of O_2 (note that the center of PGL_2 itself is trivial). Therefore the corresponding points of $\mathcal{Y}({}^LG)$ are $\mathbb{Z}_2$-orbifold points. This means that the category of B-branes supported at such a point is equivalent to the category $\mathrm{Rep}(\mathbb{Z}_2)$ of representations of $\mathbb{Z}_2$.[(12)] Thus, it has two irreducible objects. Therefore we expect that the dual category of A-branes should

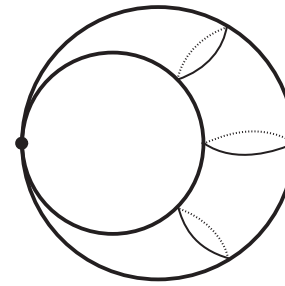
[(12)] The corresponding derived category has a more complicated structure, but we will not discuss it here.

also have two irreducible objects. In fact, it was shown in [**19**] that the dual Hitchin fiber has two irreducible components in this case, and the sought-after A-branes are *fractional branes* supported on these two components.

This was analyzed very explicitly in [**19**] in the case when X is an elliptic curve. Here we allow a single point of tame ramification (along the lines of Section 5.4)—this turns out to be better for our purposes than the unramified case. The corresponding Hitchin moduli spaces are two-dimensional. They fiber over the same one-dimensional vector space, and the fibers over all but three points in the base are smooth elliptic curves. The three pairs of dual singular fibers look as follows:



Singular Hitchin fiber in the A-model, $G = SL_2$.



Singular Hitchin fiber in the B-model, ${}^L G = SO_3$.

The fiber on the B-model side is a projective line with a double point, corresponding to a flat PGL_2-bundle that is reduced to the subgroup O_2. It is a $\mathbb{Z}_2$-orbifold point of the moduli space. The dual fiber on the A-model side is the union of two projective lines connected at two points. These two singular points of the fiber are actually smooth points of the ambient moduli space.

There are two irreducible B-branes supported at each of the $\mathbb{Z}_2$-orbifold points, corresponding to two irreducible representations of $\mathbb{Z}_2$. Let us denote them by $\mathcal{B}_+$ and $\mathcal{B}_-$. The corresponding fiber of the Hitchin fibration for SL_2 is the union of two components $\boldsymbol{F}_1$ and $\boldsymbol{F}_2$, and accordingly in the dual A-model there are two irreducible A-branes, $\mathcal{A}_1$ and $\mathcal{A}_2$ supported on these components (each component is a copy of $\mathbb{P}^1$, and therefore the only flat unitary line bundle on it is the trivial one). These A-branes are dual to the B-branes $\mathcal{B}_+$ and $\mathcal{B}_-$. Unlike $\mathcal{B}_+$ and $\mathcal{B}_-$, they are indistinguishable. An apparent contradiction is explained by the fact that in the equivalence (5.2) of the categories of A-branes and B-branes there is a twist by a $\mathbb{Z}_2$-gerbe which is not canonically trivialized (see Section 9 of [**19**] and the footnotes on pages 379 and 386). In order to set up an equivalence, we need to pick a trivialization of this gerbe, and

this breaks the symmetry between $\mathcal{A}_1$ and $\mathcal{A}_2$. We also have a similar picture when X has higher genus (see **[19]**).

What happens when we act on $\mathcal{B}_+$ or $\mathcal{B}_-$ by the Wilson operator $W_x, x \in X$, corresponding to the three-dimensional adjoint representation of ${}^LG = PGL_2$? Since $\mathcal{B}_+$ and $\mathcal{B}_-$ both have skyscraper support at the same point $\mathcal{E}$ of $\mathcal{Y}({}^LG)$, W_x acts on either of them by tensor product with the three-dimensional vector space $\mathcal{E}_x$, the fiber of $\mathcal{E}$ at x (in the adjoint representation). However, we should be more precise to keep track of the $\mathbb{Z}_2$-action. Recall that the structure group of our flat PGL_2-bundle $\mathcal{E}$ is reduced to the subgroup $O_2 = \mathbb{Z}_2 \ltimes \mathbb{C}^\times$. Denote by U the defining two-dimensional representation of O_2. Then $\det U$ is the one-dimensional sign representation induced by the homomorphism $O_2 \to \mathbb{Z}_2$. The adjoint representation of PGL_2 decomposes into the direct sum

$$(\det U \otimes I) \oplus (U \otimes S)$$

as a representation of $O_2 \times \mathbb{Z}_2$, where $\mathbb{Z}_2$ is the centralizer of O_2 in PGL_2 (the center of O_2), S is the sign representation of $\mathbb{Z}_2$, and I is the trivial representation of $\mathbb{Z}_2$. Therefore we have the following decomposition of the corresponding flat vector bundle:

$$(\det U_{\mathcal{E}} \otimes I) \oplus (U_{\mathcal{E}} \otimes S), \tag{6.1}$$

and there is a decomposition $U_{\mathcal{E}}|_x \oplus \det U_{\mathcal{E}}|_x$, where the non-trivial element of $\mathbb{Z}_2$ acts as -1 on the first summand and as $+1$ on the second summand. So we have

$$W_x \cdot \mathcal{B}_\pm = (\mathcal{B}_\mp \otimes U_{\mathcal{E}}|_x) \oplus (\mathcal{B}_\pm \otimes \det U_{\mathcal{E}}|_x)\,. \tag{6.2}$$

Thus, individual branes $\mathcal{B}_+$ and $\mathcal{B}_-$ are not eigenbranes of the Wilson operators; only their sum $\mathcal{B}_+ \oplus \mathcal{B}_-$ (corresponding to the regular representation of $\mathbb{Z}_2$) is.

The mirror dual statement, shown in **[19]**, is that we have a similar formula for the action of the corresponding 't Hooft operators on the $\mathcal{A}$-branes $\mathcal{A}_1$ and $\mathcal{A}_2$. Again, only their sum (or union), which gives the entire Hitchin fiber, is an eigenbrane of the 't Hoof operators.

Since an eigenbrane $\mathcal{A}$ decomposes into two irreducible branes $\mathcal{A}_1$ and $\mathcal{A}_2$, the corresponding Hecke eigensheaf $\mathcal{F}$ on Bun_G should also decompose as a direct sum of two $\mathcal{D}$-modules, $\mathcal{F}_1$ and $\mathcal{F}_2$, corresponding to $\mathcal{A}_1$ and $\mathcal{A}_2$, respectively. Furthermore, these two $\mathcal{D}$-modules should then separately satisfy an analogue of formula (6.2), which is a natural modification of the standard Hecke property. We called it in **[19]** the *fractional Hecke property*, and the $\mathcal{D}$-modules $\mathcal{F}_1$ and $\mathcal{F}_2$ *fractional Hecke eigensheaves*. We have also generalized this notion to other groups in **[19]**.

Thus, the Mirror Symmetry picture leads us to predict the existence of fractional Hecke eigensheaves for endoscopic flat ${}^LG$-bundles.[13] This has non-trivial consequences even for curves over $\mathbb{F}_q$, some of which have been verified in **[19]**. In addition, we obtain a relation between the group $\pi_0(\boldsymbol{P}_b)$ of components of the generalized Prym variety $\boldsymbol{P}_b$, which is an open dense part of the singular Hitchin fiber $\boldsymbol{F}_b$ arising in the A-model, and the group of automorphisms of the endoscopic flat ${}^LG$-bundles which are the singular points in the dual Hitchin fiber ${}^L\boldsymbol{F}_b$ of the B-model (this relation was independently observed by B.C. Ngô). Roughly speaking, elements of $\pi_0(\boldsymbol{P}_b)$ label the components of $\boldsymbol{F}_b$, and hence fractional A-branes. The B-branes dual to them correspond to characters of the group of automorphisms of an endoscopic flat ${}^LG$-bundle, viewed as a point in ${}^L\boldsymbol{F}_b$. Hence $\pi_0(\boldsymbol{P}_b)$ should be dual (as an abelian group) to this group of automorphisms.

The upshot of all this is that by analyzing the categories of A-branes supported on the singular Hitchin fibers, we learn many things about the geometric Langlands correspondence (and even the Langlands correspondence for curves over finite fields) which would have been very difficult to see directly using the conventional formalism of $\mathcal{D}$-modules. This is a good illustration of the power of this new method.

There is a link between our analysis and the classical theory of endoscopy, due to the fact that the geometry we use is similar to that exploited by B.C. Ngô in his recent proof of the fundamental lemma **[49]**. Ngô has discovered a striking connection between the orbital integrals appearing on the geometric side of the trace formula and the (ℓ-adic) cohomology of the Hitchin fibers in the moduli space $\mathcal{M}_H(G)$, but for curves over $\mathbb{F}_q$; more specifically, its decomposition under the action of the group $\pi_0(\boldsymbol{P}_b)$.[14]

However, there are important differences. First of all, we work over $\mathbb{C}$, whereas Ngô works over $\mathbb{F}_q$. In the latter setting there is no obvious analogue of the Homological Mirror Symmetry between $\mathcal{M}_H(G)$ and $\mathcal{M}_H({}^LG)$. Second, and more importantly, the objects we assign to the connected components of the singular Hitchin fiber $\boldsymbol{F}_b$—the A-branes—are objects of automorphic nature; we hope to relate them to Hecke eigensheaves and ultimately to the automorphic functions in the classical theory. Thus, these objects should live on the *spectral side* of the trace formula. On the other hand, in Ngô's work Hitchin fibers appear on the *geometric side* of the trace formula (more precisely, its Lie algebra version), through orbital integrals.

This raises the following question: could there be a more direct link between individual Hitchin fibers in the moduli space $\mathcal{M}_H(G)$ over $\mathbb{F}_q$ and automorphic

[13] In the case of SL_2 (as well as GSp_4) the existence of these $\mathcal{D}$-modules follows from the work of Lysenko **[42, 43, 44]**, but for other groups this is still a conjecture.

[14] More precisely, Ngô considers a generalization of $\mathcal{M}_H(G)$ parametrizing meromorphic Higgs fields with a sufficiently large divisor of poles.

representations? In other words, could it be that the passage from A-branes to Hecke eigensheaves discussed above has an analogue in the classical theory as a passage from orbital integrals to Hecke eigenfunctions? If so, then the Mirror Symmetry picture would give us valuable insights into the Langlands correspondence.

6.2. S-duality of more general boundary conditions

More general flat ${}^LG$-bundles have continuous groups of automorphisms. For instance, generic flat bundles reduced to a Cartan subalgebra ${}^LH$ have the group of automorphisms ${}^LH$. Or consider the trivial flat ${}^LG$-bundle, whose group of automorphisms is ${}^LG$ itself. What are the A-branes corresponding to these flat ${}^LG$-bundles?

The picture of two dual Hitchin fibrations discussed above is too naive to answer this question. The reason is that even if the flat bundles with continuous groups of automorphisms are semi-stable (which is not necessarily the case), they correspond to points of $\mathcal{Y}({}^LG)$ with singularity so severe that the category of B-branes corresponding to it cannot be described solely in terms of the moduli space $\mathcal{Y}({}^LG)$. In fact, the definition of the sigma model itself is problematic for singular target manifolds.

As an illustration, consider the quotient $\mathbb{C}^n/\mathbb{C}^\times$. The origin has the group of automorphisms $\mathbb{C}^\times$. What is the category of B-branes associated to this point? Because it is a singular point, there is no obvious answer (unlike the case of smooth points or orbifold points, discussed above). However, we can resolve the singularity by blowing it up. On general ground one can argue that this resolution will not change the category of B-branes. The category of B-branes after the resolution of singularities is the category of coherent sheaves on $\mathbb{P}^{n-1}$. So the singular point in $\mathbb{C}^n/\mathbb{C}^\times$ has "swallowed" an entire projective space! Likewise, singular points in $\mathcal{Y}({}^LG)$ also have complicated "inner structure" which needs to be uncovered to do justice to the corresponding categories of B-branes.

In order to understand better what is going on we should go back to the four-dimensional gauge theory and look more closely at the S-duality of boundary conditions there. From the physics perspective, this is the "master duality" and everything should follow from it. The Mirror Symmetry of the Hitchin fibrations is but the first approximation to the S-duality when we compactify the theory to two dimensions.

It is instructive to recall how one obtains the Hitchin moduli spaces in the first place: Each of the S-dual gauge theories has a differential Q such that $Q^2 = 0$, and we study the corresponding topological field theories. In the topological theory the path integral localizes on the moduli space of solutions to the "BPS equations", which read $Q \cdot \Psi = 0$, for all fermionic fields Ψ of our theory (since Q is fermionic and we want the equations on the bosonic degrees of freedom). After that we make dimensional reduction of these equations. This means that we assume that our four-manifold has

the form $\Sigma \times X$ and the fields on Σ vary "slowly" along Σ. The corresponding equations have been written in [**34**]:

$$F_A - \phi \wedge \phi = 0, \tag{6.3}$$
$$d_A \phi = d_A \star \phi = 0,$$

where d_A is the exterior derivative corresponding to the connection A, and $\star$ is the Hodge star operator. These are precisely the Hitchin equations [**30**] describing the moduli spaces $\mathcal{M}_H(G)$ or $\mathcal{M}_H({}^L G)$ (depending on which side of S-duality we are on). For example, points of $\mathcal{M}_H({}^L G)$ in the complex structure J are semi-stable flat ${}^L G$-bundles on X. The flat connection on this bundle is given by the formula $\nabla = A + i\phi$ (the flatness of ∇ is a corollary of (6.3)). This is how the (B-twisted) sigma model on Σ with values in $\mathcal{M}_H({}^L G)$ appears in this story. One obtains the (A-twisted) sigma model with target $\mathcal{M}_H(G)$ in the complex structure I in a similar way.

However, as Kapustin and Witten explain in [**34**], this derivation breaks down when we encounter singularities of the Hitchin moduli spaces. Thus, the sigma models with the targets $\mathcal{M}_H({}^L G)$ and $\mathcal{M}_H(G)$ are only approximations to the true physical theory. To understand what happens at the singularities we have to go back to the four-dimensional theory and analyze it more carefully (for more on this, see [**61**]).

There is also another problem: in the above derivation we have not taken into account all the fields of the super-Yang–Mills theory. In fact, there are additional scalar fields, denoted by σ and $\overline{\sigma}$ in [**34**], which we have ignored so far. The field σ is a section of the adjoint bundle $\mathfrak{g}_E$ on X, and $\overline{\sigma}$ is its complex conjugate. On the B-model side, which we have so far approximated by the sigma model with the target $\mathcal{M}_H({}^L G)$, we obtain from the BPS equations that σ is annihilated by the flat connection $\nabla = A + i\phi$, that is, $\nabla \cdot \sigma = 0$. In other words, σ belongs to the Lie algebra of infinitesimal automorphisms of the flat bundle (E, ∇).

Up to now we have considered generic flat ${}^L G$-bundles which have no non-trivial infinitesimal automorphisms. For such flat bundles we therefore have $\sigma \equiv 0$, and so we could safely ignore it. But for flat bundles with continuous automorphisms this field starts playing an important role.

The upshot of this discussion is that when we consider most general flat bundles there are new degrees of freedom that have to be taken into account. In order to find a physical interpretation of the geometric Langlands correspondence for such flat bundles we need to consider the S-duality of boundary conditions in the four-dimensional gauge theory with these degrees of freedom included.

A detailed study of S-duality of these boundary conditions has been undertaken by Gaiotto and Witten [**20, 21**]. We will only mention two important aspects of this work.

First of all, Gaiotto and Witten show that in the non-abelian gauge theory the S-dual of the Neumann boundary condition is not the usual Dirichlet boundary condition as one might naively hope, but a more complicated boundary condition in which the field σ has a pole at the boundary. This boundary condition corresponds to a solution of the Nahm equations, which is in turn determined by an embedding of the Lie algebra $\mathfrak{sl}_2$ into ${}^L\mathfrak{g}$ (see [**59**]).

This is parallel to the appearance of Arthur's SL_2 in the classical Langlands correspondence. Arthur has conjectured that the true parameters for (L-packets of) unitary automorphic representations of $G(\mathbb{A})$ are not homomorphisms $\mathrm{Gal}(\overline{F}/F) \to {}^LG$, but rather $\mathrm{Gal}(\overline{F}/F) \times SL_2 \to {}^LG$. The homomorphisms whose restriction to the SL_2 factor is trivial should correspond to the so-called tempered representations. (In the case of GL_n all cuspidal unitary representations are tempered, and that is why Arthur's SL_2 does not appear in the theorem of Drinfeld and Lafforgue quoted in Section 1.) An example of non-tempered unitary representation is the trivial representation of $G(\mathbb{A})$. According to [**2**], the corresponding parameter is the homomorphism $\mathrm{Gal}(\overline{F}/F) \times SL_2 \to {}^LG$, which is trivial on the first factor and is the principal embedding of the SL_2 factor.

In the geometric Langlands correspondence, Arthur's SL_2 may be observed in the following way. The analogue of the trivial representation is the constant sheaf $\mathbf{C}$ on Bun_G. It is a Hecke eigensheaf, but the "eigenvalues" are complexes of vector spaces with cohomological grading coming from the Cartan subalgebra of the principal SL_2 in ${}^LG$. For example, consider the case of $G = GL_n$ and let us apply the Hecke operators $H_{\check{\omega}_1,x}$ defined by formula (2.2) to the constant sheaf. It follows from the definition that

$$H_{\check{\omega}_1,x}(\mathbf{C}) \simeq H^\bullet(\mathbb{P}^{n-1},\mathbb{C}) \otimes \mathbf{C}.$$

Thus, the eigenvalue is a graded n-dimensional vector space with one-dimensional pieces in cohomological degrees $0, 2, 4, \ldots, 2(n-1)$. In the standard normalization of the Hecke operator the cohomological grading is shifted to $-(n-1), -(n-3), \ldots, (n-1)$—as in the grading on the n-dimensional representation of GL_n coming from the principal SL_2.

The A-brane corresponding to the constant sheaf on Bun_G is the Lagrangian submanifold of $\mathcal{M}_H(G)$ defined by the equation $\phi = 0$ (this is the zero section of the cotangent bundle to the moduli space of semi-stable G-bundles inside $\mathcal{M}_H(G)$). This A-brane corresponds to a Neumann boundary condition in the 4D gauge theory with gauge group G_c. According to Gaiotto and Witten, the dual boundary condition (in the theory with gauge group ${}^LG_c$) is a generalization of the Dirichlet boundary condition, in which the field σ has a pole at the boundary solving the Nahm equations

corresponding to the principal SL_2 embedding into ${}^L G$. Thus, we obtain a beautiful interpretation of Arthur's SL_2 from the point of view of S-duality of boundary conditions in gauge theory. For more on this, see **[18, 59]**.

Another important feature discovered in **[20, 21]** is that the S-duals of the general Dirichlet boundary conditions involve coupling the 4D super-Yang-Mills to 3D superconformal QFTs at the boundary. This means that there are some additional degrees of freedom that we have to include to describe the geometric Langlands correspondence.

What we learn from all this is that the true moduli spaces arising in the S-duality picture are not the Hitchin moduli spaces $\mathcal{M}_H(G)$ and $\mathcal{M}_H({}^L G)$, but some enhanced versions $\widetilde{\mathcal{M}}_H(G)$ and $\widetilde{\mathcal{M}}_H({}^L G)$ thereof, including, in addition to the Higgs bundle (E, ϕ), an element σ in the Lie algebra of its infinitesimal automorphisms as well as other data. (This will be discussed in more detail in the forthcoming paper **[18]**.) Physically, the field σ has non-zero "ghost number" 2. Mathematically, this means that these additional degrees of freedom have cohomological grading 2, and so $\widetilde{\mathcal{M}}_H(G)$ and $\widetilde{\mathcal{M}}_H({}^L G)$ are actually differential graded (DG) stacks. Similar DG stacks have been recently studied in the context of the categorical Langlands correspondence by V. Lafforgue **[38]**.

Thus, S-duality of super-Yang–Mills theory offers new insights into the Langlands correspondence and surprising new connections to geometry. Ultimately, we have to tackle the biggest question of all: what is the underlying reason for the Langlands duality? On the physics side the corresponding question is: why S-duality? In fact, physicists have the following elegant explanation (see **[60]**): there is a mysterious six-dimensional quantum field theory which gives rise to the four-dimensional super-Yang–Mills upon compactification on an elliptic curve. Roughly speaking (the argument should be modified slightly for non-simply laced groups), this elliptic curve is $E = \mathbb{C}/(\mathbb{Z} + \mathbb{Z}\tau)$, where τ is the coupling constant of this Yang–Mills theory. Since the action of $SL_2(\mathbb{Z})$ on τ does not change the elliptic curve E, we obtain that there are equivalences between the super-Yang–Mills theories whose coupling constants are related by the action of $SL_2(\mathbb{Z})$. This should explain the S-duality, which corresponds to the transformation $\tau \mapsto -1/\tau$, and hence the geometric Langlands correspondence. But that is a topic for a future Séminaire Bourbaki.

REFERENCES

[1] D. Arinkin – "Moduli of connections with a small parameter on a curve", preprint arXiv:math.AG/0409373.

[2] J. Arthur – "Unipotent automorphic representations: conjectures", *Astérisque* **171-172** (1989), p. 13–71.

[3] A. Beilinson & V. Drinfeld – "Quantization of Hitchin's integrable system and Hecke eigensheaves", preprint `http://www.math.uchicago.edu/~mitya/langlands/hitchin/BD-hitchin.pdf`.

[4] M. Bershadsky, A. Johansen, V. Sadov & C. Vafa – "Topological reduction of 4D SYM to 2D σ-models", *Nuclear Phys. B* **448** (1995), p. 166–186.

[5] K. Corlette – "Flat G-bundles with canonical metrics", *J. Differential Geom.* **28** (1988), p. 361–382.

[6] P. Deligne *et al.* – *Quantum fields and strings: A course for mathematicians, Vol. I and II*, Amer. Math. Soc., Institute for Advanced Study, 1999.

[7] R. Donagi & T. Pantev – "Langlands duality for Hitchin systems", preprint arXiv:math.AG/0604617.

[8] V. G. Drinfel′d – "Langlands' conjecture for GL(2) over functional fields", in *Proceedings of the International Congress of Mathematicians (Helsinki, 1978)*, Acad. Sci. Fennica, 1980, p. 565–574.

[9] ———, "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), p. 85–114.

[10] ———, "Moduli varieties of F-sheaves", *Funct. Anal. Appl.* **21** (1987), p. 107–122.

[11] ———, "The proof of Petersson's conjecture for $GL(2)$ over a global field of characteristic p", *Funct. Anal. Appl.* **22** (1988), p. 28–43.

[12] F. Englert & P. Windey – "Quantization condition for 't Hooft monopoles in compact simple Lie groups", *Phys. Rev D* **14** (1976), p. 2728–2731.

[13] E. Frenkel – "Recent advances in the Langlands program", *Bull. Amer. Math. Soc. (N.S.)* **41** (2004), p. 151–184.

[14] ———, *Langlands correspondence for loop groups*, Cambridge Studies in Advanced Math., vol. 103, Cambridge Univ. Press, 2007.

[15] ———, "Lectures on the Langlands program and conformal field theory", in *Frontiers in number theory, physics, and geometry. II*, Springer, 2007, p. 387–533.

[16] E. Frenkel & D. Gaitsgory – "Local geometric Langlands correspondence and affine Kac-Moody algebras", in *Algebraic geometry and number theory*, Progr. Math., vol. 253, Birkhäuser, 2006, p. 69–260.

[17] E. Frenkel, D. Gaitsgory & K. Vilonen – "On the geometric Langlands conjecture", *J. Amer. Math. Soc.* **15** (2002), p. 367–417.

[18] E. Frenkel & S. Gukov – "S-duality of branes and geometric Langlands correspondence", to appear.

[19] E. Frenkel & E. Witten – "Geometric endoscopy and mirror symmetry", *Commun. Number Theory Phys.* **2** (2008), p. 113–283.

[20] D. Gaiotto & E. Witten – "Supersymmetric boundary conditions in $\mathcal{N} = 4$ super Yang-Mills theory", *J. Stat. Phys.* **135** (2009), p. 789–855.

[21] ———, "S-duality of boundary conditions in $N = 4$ super Yang-Mills theory", preprint arXiv:0807.3720.

[22] D. Gaitsgory – "On a vanishing conjecture appearing in the geometric Langlands correspondence", *Ann. of Math.* **160** (2004), p. 617–682.

[23] S. Gelbart – "An elementary introduction to the Langlands program", *Bull. Amer. Math. Soc. (N.S.)* **10** (1984), p. 177–219.

[24] S. I. Gelfand & Y. I. Manin – *Homological algebra*, Encyclopaedia of Math. Sciences, vol. 38, Springer, 1994.

[25] P. Goddard, J. Nuyts & D. I. Olive – "Gauge theories and magnetic charge", *Nuclear Phys. B* **125** (1977), p. 1–28.

[26] S. Gukov & E. Witten – "Gauge theory, ramification, and the geometric Langlands program", in *Current developments in mathematics, 2006*, Int. Press, Somerville, MA, 2008, p. 35–180.

[27] ———, "Rigid surface operators", preprint arXiv:0804.1561.

[28] J. A. Harvey, G. Moore & A. Strominger – "Reducing S duality to T duality", *Phys. Rev. D* **52** (1995), p. 7161–7167.

[29] T. Hausel & M. Thaddeus – "Mirror symmetry, Langlands duality, and the Hitchin system", *Invent. Math.* **153** (2003), p. 197–229.

[30] N. J. Hitchin – "The self-duality equations on a Riemann surface", *Proc. London Math. Soc.* **55** (1987), p. 59–126.

[31] ———, "Stable bundles and integrable systems", *Duke Math. J.* **54** (1987), p. 91–114.

[32] ———, "Langlands duality and G_2 spectral curves", *Q. J. Math.* **58** (2007), p. 319–344.

[33] A. Kapustin – "A note on quantum geometric Langlands duality, gauge theory, and quantization of the moduli space of flat connections", preprint arXiv:0811.3264.

[34] A. Kapustin & E. Witten – "Electric-magnetic duality and the geometric Langlands program", *Commun. Number Theory Phys.* **1** (2007), p. 1–236.

[35] M. Kashiwara & P. Schapira – *Sheaves on manifolds*, Grund. Math. Wiss., vol. 292, Springer, 1990.

[36] M. H. Krieger – "A 1940 letter of André Weil on analogy in mathematics", *Notices Amer. Math. Soc.* **52** (2005), p. 334–341.

[37] L. Lafforgue – "Chtoucas de Drinfeld et correspondance de Langlands", *Invent. Math.* **147** (2002), p. 1–241.

[38] ———, "Quelques calculs reliés à la correspondance de Langlands géométrique pour $\mathbb{P}^1$", preprint `http://people.math.jussieu.fr/~vlafforg/geom.pdf`.

[39] R. P. Langlands – "Problems in the theory of automorphic forms", in *Lectures in modern analysis and applications, III*, Lecture Notes in Math., vol. 170, Springer, 1970, p. 18–61.

[40] G. Laumon – "Correspondance de Langlands géométrique pour les corps de fonctions", *Duke Math. J.* **54** (1987), p. 309–359.

[41] ———, "Transformation de Fourier généralisée", preprint arXiv:alg-geom/9603004.

[42] S. Lysenko – "Geometric Waldspurger periods", *Compos. Math.* **144** (2008), p. 377–438.

[43] ———, "Geometric theta-lifting for the dual pair $\mathrm{GSp}_{2n}, \mathrm{GO}_{2m}$", preprint arXiv:0802.0457.

[44] ———, "Geometric theta-lifting for the dual pair $\mathrm{SO}_{2m}, \mathrm{Sp}_{2n}$", preprint arXiv:math/0701170.

[45] I. Mirković & K. Vilonen – "Geometric Langlands duality and representations of algebraic groups over commutative rings", *Ann. of Math.* **166** (2007), p. 95–143.

[46] C. Montonen & D. I. Olive – "Magnetic monopoles as gauge particles?", *Phys. Lett. B* **72** (1977), p. 117–120.

[47] D. Nadler – "Microlocal branes are constructible sheaves", preprint arXiv:math/0612399.

[48] D. Nadler & E. Zaslow – "Constructible sheaves and the Fukaya category", *J. Amer. Math. Soc.* **22** (2009), p. 233–286.

[49] B. C. Ngô – "Le lemme fondamental pour les algèbres de Lie", preprint arXiv:0801.0446.

[50] H. Osborn – "Topological charges for $N = 4$ supersymmetric gauge theories and monopoles of Spin 1", *Phys. Lett. B* **83** (1979), 321–326.

[51] M. Rothstein – "Connections on the total Picard sheaf and the KP hierarchy", *Acta Appl. Math.* **42** (1996), p. 297–308.

[52] C. T. Simpson – "Constructing variations of Hodge structure using Yang-Mills theory and applications to uniformization", *J. Amer. Math. Soc.* **1** (1988), p. 867–918.

[53] ———, "Harmonic bundles on noncompact curves", *J. Amer. Math. Soc.* **3** (1990), p. 713–770.

[54] A. Strominger, S.-T. Yau & E. Zaslow – "Mirror symmetry is T-duality", *Nuclear Phys. B* **479** (1996), p. 243–259.

[55] C. Vafa & E. Witten – "A strong coupling test of S-duality", *Nuclear Phys. B* **431** (1994), p. 3–77.

[56] E. Witten – "Topological quantum field theory", *Comm. Math. Phys.* **117** (1988), p. 353–386.

[57] ———, "Mirror manifolds and topological field theory", in *Essays on mirror manifolds*, Int. Press, Hong Kong, 1992, p. 120–158.

[58] ———, "Gauge theory and wild ramification", *Anal. Appl. (Singap.)* **6** (2008), p. 429–501.

[59] ———, "Geometric Langlands and the equations of Nahm and Bogomolny", to appear.

[60] ———, "Geometric Langlands from six dimensions", preprint arXiv:0905.2720.

[61] ———, "Mirror symmetry, Hitchin's equations, and Langlands duality", preprint arXiv:0802.0999.

Edward FRENKEL

Department of Mathematics
University of California
Berkeley, CA 94720-3840 – U.S.A.
E-mail : frenkel@math.berkeley.edu

332

ASTÉRISQUE

2010

SÉMINAIRE BOURBAKI
VOLUME 2008/2009
EXPOSÉS 997-1011

(1011) *Variétés hyperboliques de petit volume*

Sylvain MAILLOT

SOCIÉTÉ MATHÉMATIQUE DE FRANCE
Publié avec le concours du CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

Séminaire BOURBAKI Juin 2009
61e année, 2008-2009, no 1011, p. 405 à 417

VARIÉTÉS HYPERBOLIQUES DE PETIT VOLUME
[d'après D. Gabai, R. Meyerhoff, P. Milley, ...]

par **Sylvain MAILLOT**

INTRODUCTION

Soit M une variété lisse. Une *métrique hyperbolique* sur M est une métrique riemannienne g_{hyp} complète de courbure sectionnelle constante égale à -1. Dans cet exposé, nous nous intéresserons principalement au cas où le volume de g_{hyp} est fini. Si la dimension de M est au moins 3, le théorème de rigidité de Mostow-Prasad implique qu'une variété donnée M ne peut admettre qu'une seule métrique hyperbolique de volume fini à isométrie près. Si c'est le cas, nous dirons que M est une *variété hyperbolique*. Les invariants métriques de g_{hyp}, en particulier le volume, ne dépendent donc que de M.

En dimension 2, le théorème de rigidité de Mostow ne s'applique pas, mais le volume est tout de même un invariant topologique. En effet, si M est une surface, disons fermée et orientable, la formule de Gauss-Bonnet donne $\mathrm{Vol}(g_{\mathrm{hyp}}) = 2\pi|\chi(M)|$ où $\chi(M)$ est la caractéristique d'Euler de M. Par conséquent, l'ensemble des volumes des surfaces hyperboliques est discret, et le volume croît avec la complexité topologique. En particulier, la surface de plus petit volume est celle de genre 2, qui est aussi la plus simple.

En dimension 3, la question est plus complexe. Il n'est pas évident *a priori* qu'il existe une variété hyperbolique de plus petit volume, ni même que l'infimum de l'ensemble des volumes soit non nul. Dans la section 1, on rappellera des résultats classiques de Kazhdan-Margulis, Thurston et Jørgensen qui impliquent que c'est le cas. On notera ici v_0 le plus petit élément de l'ensemble $\mathcal{V}$ des volumes des variétés hyperboliques orientables de dimension 3.

Se pose alors la question naturelle de déterminer la (ou les) variété(s) réalisant ce minimum. Un candidat a été proposé indépendamment par J. Weeks, d'une part, et A. Fomenko et S. Matveev [**36**], d'autre part. À l'aide du programme SnapPea,

J. Weeks et C. Hodgson [**29**] ont effectué un « recensement » des variétés hyperboliques et calculé des valeurs approchées de leurs volumes, donnant ainsi du poids à la conjecture selon laquelle la variété de Weeks-Fomenko-Matveev est l'unique variété de volume minimal, ce volume étant approximativement 0,9427.

La première minoration publiée de v_0 est due à R. Meyerhoff [**38**]. Elle est de 0,00064, qui comme on le voit est très loin du compte. Cette minoration a été ensuite améliorée dans une série de travaux dont certains seront traités plus en détail dans la section 2 [**2, 4, 20, 22, 23, 35, 37, 44, 45**].

Dans une autre direction, une série de travaux de M. Culler, P. Shalen et leurs collaborateurs (voir par exemple les articles [**3, 13**] et leurs références) ont montré qu'il existe une corrélation entre volume et complexité topologique.

Dans ce texte, on se limite par souci de simplicité aux variétés orientables de dimension 3. On peut bien entendu poser la même question en dimension supérieure, pour les variétés non-orientables, ou pour les orbifolds (voir par exemple les articles [**1, 14, 24, 27, 28, 31, 32**] et leurs références). Il ne sera ici que très peu question de ces généralisations. Nous n'entrerons pas non plus dans les détails de la preuve du théorème de Gabai-Meyerhoff-Milley ; le lecteur intéressé est renvoyé aux articles originaux, ainsi qu'au texte d'exposition [**18**].

1. STRUCTURE DE L'ENSEMBLE DES VOLUMES

1.1. Le lemme de Margulis

Soit (M, g_{hyp}) une variété hyperbolique de dimension n. Un résultat élémentaire de géométrie riemannienne affirme que le revêtement universel $\tilde{M}$ de M muni de la métrique induite $\tilde{g_{\mathrm{hyp}}}$ est isométrique à l'espace hyperbolique réel de dimension n, que nous notons $\mathbf{H}^n$. Une fois fixée une identification de $\tilde{M}$ avec $\mathbf{H}^n$, on peut considérer le groupe fondamental $\pi_1 M$ comme un sous-groupe du groupe $\mathrm{Isom}(\mathbf{H}^n)$ des isométries de $\mathbf{H}^n$. Si M est orientable, ce que l'on supposera toujours dans la suite, alors $\pi_1 M$ est inclus dans le sous-groupe $\mathrm{Isom}^+(\mathbf{H}^n)$ de $\mathrm{Isom}(\mathbf{H}^n)$ formé des éléments qui préservent l'orientation. De plus, ce sous-groupe est sans torsion et discret pour la topologie usuelle. Nous appellerons *groupe kleinéen* un sous-groupe de $\mathrm{Isom}^+(\mathbf{H}^n)$ ayant ces deux propriétés.

Pour $n = 3$, on peut identifier $\mathrm{Isom}^+(\mathbf{H}^3)$ avec $\mathrm{PSL}_2(\mathbf{C})$. Si l'on voit $\mathbf{H}^3$ comme la boule unité ouverte de $\mathbf{R}^3$ avec la métrique de Poincaré, alors l'action de $\mathrm{Isom}(\mathbf{H}^3)$ sur $\mathbf{H}^3$ se prolonge en une action sur la boule unité fermée $\overline{\mathbf{H}^3}$. Il est possible d'identifier la sphère $\overline{\mathbf{H}^3} - \mathbf{H}^3$ (la « sphère à l'infini ») avec $\mathbf{CP}^1$ de sorte que l'action induite de $\mathrm{PSL}_2(\mathbf{C})$ sur $\mathbf{CP}^1$ soit l'action standard.

On classifie les éléments de $\mathrm{PSL}_2(\mathbf{C})$ en trois types de la façon suivante. Soit γ un élément de $\mathrm{PSL}_2(\mathbf{C})$ différent de $\pm\mathrm{I}$. Si γ est diagonalisable, on dit qu'il est *semi-simple*. Il est alors conjugué à $\left(\begin{smallmatrix}\lambda & 0\\ 0 & \lambda^{-1}\end{smallmatrix}\right)$ pour un certain $\lambda \in \mathbf{C}^*$. Si $|\lambda| = 1$, on dit que γ est *elliptique*. Sinon, on dit que γ est *hyperbolique*.

Une isométrie semi-simple γ fixe exactement deux points de la sphère à l'infini $\mathbf{CP}^1$. L'unique géodésique de $\mathbf{H}^3$ reliant ces deux points est appelé l'*axe* de γ. Si l'on écrit la valeur propre λ comme $e^{(l+i\theta)/2}$ avec $l, \theta \in \mathbf{R}$, alors on peut interpréter géométriquement γ comme une sorte de vissage, dont l serait la longueur algébrique de translation et θ l'angle de rotation. Le cas elliptique correspond donc à une rotation pure.

Si γ n'est pas diagonalisable, alors γ est conjugué à $\left(\begin{smallmatrix}1 & 1\\ 0 & 1\end{smallmatrix}\right)$ et on dit que γ est *parabolique*. Dans ce cas γ fixe exactement un point de $\mathbf{CP}^1$.

Si $x \in \mathbf{CP}^1$, on appelle *horoboule* centrée en x une partie de $\mathbf{H}^3$ de la forme $B - \{x\}$, où B est une boule fermée de $\overline{\mathbf{H}^3}$ tangente à $\mathbf{CP}^1$ en x. Ces ensembles sont stabilisés par les isométries paraboliques ayant x comme point fixe.

Soit $\Gamma \subset \mathrm{PSL}_2(\mathbf{C})$ un groupe kleinéen. Il est facile de voir que Γ ne peut pas contenir d'élément elliptique. On dit que Γ est *élémentaire* s'il existe un point $x \in \mathbf{CP}^1$ qui est fixé par tous les éléments de Γ. Il existe trois types de sous-groupes élémentaires :

1. hyperbolique : un groupe infini cyclique constitué d'éléments hyperboliques de même axe ;
2. parabolique de rang 1 : un groupe infini cyclique constitué d'éléments paraboliques ayant même point fixe à l'infini ;
3. parabolique de rang 2 : un groupe abélien libre de rang 2 constitué d'éléments paraboliques ayant même point fixe à l'infini.

Nous nous intéressons à la partie de M formée des points où le rayon d'injectivité est petit. Comme nous sommes en courbure négative, ce sont exactement les points y par lesquels il passe un petit lacet non-homotope à zéro. Un tel lacet correspond à un élément du groupe fondamental de M dont l'action sur $\mathbf{H}^3$ a la propriété de ne pas beaucoup bouger un certain relevé de y.

Si Γ est un groupe kleinéen, x un point de $\mathbf{H}^3$ et ε un nombre réel strictement positif, on note $\Gamma_{x,\varepsilon}$ le sous-groupe de Γ engendré par les éléments γ tels que la distance (pour la métrique hyperbolique) entre x et $\gamma(x)$ est inférieure ou égale à ε.

THÉORÈME 1.1 (Lemme de Margulis [**30**]). — *Il existe une constante universelle $\mu_3 > 0$ telle que, pour tout $\varepsilon \in]0, \mu_3]$, pour tout groupe kleinéen Γ et tout $x \in \mathbf{H}^3$, le groupe $\Gamma_{x,\varepsilon}$ soit élémentaire.*

Dans la suite on fixe une telle constante μ_3 (appelée « constante de Margulis »).

DÉFINITION 1.2. — *Soient M une variété hyperbolique de dimension 3 et $\varepsilon > 0$. La* partie ε-mince *de M est l'ensemble des points $x \in M$ tel qu'il existe un élément $\gamma \in \pi_1 M - \pm \mathrm{I}$ et un relevé $\tilde{x} \in \mathbf{H}^3$ de x satisfaisant $d(\tilde{x}, \gamma\tilde{x}) \leq \varepsilon$. La* partie ε-épaisse *est le complémentaire de la partie ε-mince.*

Le lemme de Margulis nous permet d'obtenir une description précise de la partie ε-mince d'une variété hyperbolique pour ε inférieur ou égal à la constante de Margulis.

Soient $R > 0$, $\Gamma \subset \mathrm{PSL}_2(\mathbf{C})$ un groupe élémentaire hyperbolique et L l'axe commun des éléments de Γ. Rappelons que le *R-voisinage* de L est l'ensemble des points de $\mathbf{H}^3$ dont la distance à L est inférieure ou égale à R. Le quotient de cet ensemble par Γ est appelé *tube de Margulis*. On dit que R est le *rayon* de ce tube, et la géodésique fermée $L/\Gamma \subset M$ en est l'*âme*. Topologiquement, un tube de Margulis est un *tore solide*, c'est-à-dire qu'il est homéomorphe à $S^1 \times D^2$.

Si B est une horoboule centrée en un point $x \in \mathbf{CP}^1$ et Γ un groupe élémentaire parabolique dont le point fixe est x, alors on appelle *voisinage de cusp* le quotient de B par Γ. Cet ensemble est homéomorphe à $[0, +\infty[\times S^1 \times \mathbf{R}$ ou $[0, +\infty[\times S^1 \times S^1$ selon si le rang de Γ est 1 ou 2.

Le lemme de Margulis a la conséquence suivante :

COROLLAIRE 1.3. — *Soient M une variété hyperbolique de dimension 3 et $\varepsilon \in]0, \mu_3]$. Alors chaque composante connexe de la partie ε-mince de M est un tube de Margulis ou un voisinage de cusp.*

Cet énoncé est valable même si la métrique hyperbolique a un volume infini. Dans le cas où le volume est fini, on obtient des résultats plus précis. Premièrement, il n'y a pas de voisinage de cusp de rang 1, car ceux-ci ont un volume infini. Deuxièmement, il y a au plus un nombre fini de tubes de Margulis et de cusps de rang 2.

En particulier, si M est compacte, la partie mince est entièrement constituée de tubes de Margulis. Si M est non-compacte, chaque bout de M admet un voisinage homéomorphe à $T^2 \times [0, +\infty[$. Il existe donc une variété à bord $\bar{M}$ de dimension 3 compacte, dont les composantes de bord sont des tores, et telle que M soit homéomorphe à l'intérieur de $\bar{M}$. Les bouts de M sont appelés des *cusps*.

Du point de vue de la question qui nous intéresse, à savoir le problème du volume minimal, le lemme de Margulis implique que l'infimum des volumes des variétés hyperboliques de dimension 3 est strictement positif. En effet, il résulte du corollaire 1.3 que la partie μ_3-épaisse de M est non-vide, ce qui permet de minorer le volume de M par celui d'une boule de rayon μ_3 dans $\mathbf{H}^3$.

1.2. Remplissage de Dehn

À partir de maintenant, les variétés hyperboliques considérées sont toujours de dimension 3, orientables et de volume fini, et nous omettrons ces précisions.

On trouvera des discussions plus complètes et détaillées des résultats de cette sous-section dans l'exposé **[25]** et la monographie **[6]**, cette dernière traitant également de la généralisation de cette théorie aux orbifolds.

Nous commençons cette sous-section par une définition purement topologique. Soit X une variété à bord compacte de dimension 3 telle que chaque composante connexe de ∂X soit un tore. On appelle *remplissage de Dehn* sur X l'opération qui consiste à choisir un certain nombre de composantes de ∂X, recoller un tore solide à X le long de chacune de ces composantes, et retirer les autres.

Rappelons qu'on nomme *méridien* de $S^1 \times D^2$ la courbe $S^1 \times \partial D^2$. Si ∂X est connexe, le type topologique de la variété obtenue par remplissage de Dehn ne dépend que de la classe d'homologie de la courbe de ∂X le long de laquelle est recollé le méridien. Une fois choisie une base (m, l) de $H_1(\partial X)$, les remplissages de Dehn sur X sont donc paramétrés par les couples d'entiers premiers entre eux (à l'exception du « faux » remplissage qui consiste à simplement retirer ∂X.)

Dans le cas général où le nombre de composantes de ∂X est arbitraire, on choisit pour chaque composante T_i de ∂X une base (m_i, l_i) de $H_1(T_i)$. Les remplissages de Dehn sont paramétrés par les p-uples $(x_1, \dots, x_p) \in (\mathbf{Z}^2 \cup \{\infty\})^p$ où chaque x_i est soit un couple d'entiers premiers entre eux, soit ∞, ce dernier cas correspondant au fait de retirer T_i.

Soit M une variété hyperbolique. On a vu à la sous-section précédente que M admet une compactification $\bar{M}$. Par abus de langage, on appelle encore remplissage de Dehn sur M un remplissage de Dehn sur sa compactification $\bar{M}$. Ainsi, le remplissage $(\infty, \dots, \infty)$ correspond à la variété M elle-même. On dit qu'un remplissage de Dehn est hyperbolique si la variété obtenue est hyperbolique.

Le théorème suivant, dû à W. Thurston, fournit une foule d'exemples de variétés hyperboliques.

THÉORÈME 1.4 (Remplissage de Dehn hyperbolique). — *Soit M une variété hyperbolique ayant p cusps. Alors il existe un voisinage $\mathcal{U}$ de $(\infty, \dots, \infty)$ dans $(\mathbf{Z}^2 \cup \{\infty\})^p$ tel que, pour tout $x \in \mathcal{U}$, le remplissage de Dehn associé à x soit hyperbolique.*

Du point de vue des volumes, il faut noter deux faits : premièrement, si N est un remplissage de Dehn hyperbolique sur M, alors $\operatorname{Vol} N < \operatorname{Vol} M$; deuxièmement, si N_k est une suite de remplissages de Dehn sur M dont les paramètres tendent vers $(\infty, \dots, \infty)$, alors $\operatorname{Vol} N_k$ converge vers $\operatorname{Vol} M$ quand $k \to \infty$ [(1)]. Cette dernière

[(1)] Il existe des résultats plus précis concernant cette convergence, voir par exemple **[41]**.

propriété résulte du fait que la suite de variétés N_k « converge géométriquement » vers M.

Pour pouvoir utiliser le théorème 1.4, il faut bien entendu disposer d'exemples de variétés hyperboliques à cusps. La source la plus utile de tels exemples est le théorème d'hyperbolisation de Thurston.

Les exemples les plus connus sont le *complémentaire du nœud de huit* et le *complémentaire de l'entrelacs de Whitehead*. Le premier est $S^3 - k$, où k est une certaine courbe fermée simple dans S^3 appelée « nœud de huit », et le deuxième $S^3 - L$ où L est la réunion disjointe de deux courbes fermées simples formant « l'entrelacs de Whitehead ».

Nous pouvons à présent définir la variété W de Weeks-Fomenko-Matveev : c'est la variété obtenue par remplissage $((5,1),(5,2))$ sur le complémentaire de l'entrelacs de Whitehead. Notons au passage que la variété obtenue par remplissage de Dehn $((5,1),\infty)$ sur le complémentaire de l'entrelacs de Whitehead est hyperbolique et a le même volume que le complémentaire du nœud de huit (soit environ 2,02988.) Pour cette raison, on appelle cette variété la *sœur du complémentaire du nœud de huit*.

1.3. Théorème de Thurston-Jørgensen

Notons $\mathcal{V}$ l'ensemble des volumes des variétés hyperboliques (orientables de dimension 3.) C'est un sous-ensemble de la droite réelle, qu'on munit de l'ordre induit.

Il n'est pas difficile de construire pour chaque entier naturel p une variété hyperbolique ayant exactement p cusps. En calculant leurs volumes, on vérifie qu'on obtient ainsi un ensemble discret infini plongé dans $\mathcal{V}$. D'autre part, il résulte des considérations de la section précédente que le p-ième de ces points est un point d'accumulation de rang p, l'accumulation se faisant par la gauche.

Réciproquement, on peut montrer que tous les points d'accumulation de $\mathcal{V}$ sont obtenus de cette manière. On obtient le résultat suivant :

THÉORÈME 1.5. — *$\mathcal{V}$ est un ensemble bien ordonné de type ω^ω. De plus, chaque élément de $\mathcal{V}$ est réalisé par un nombre fini de variétés.*

Il existe donc un plus petit élément de $\mathcal{V}$, noté v_0. La (ou les) variété(s) correspondante(s) est (sont) compacte(s). Il y a un plus petit point d'accumulation, noté v_ω, qui est aussi le plus petit volume d'une variété hyperbolique non-compacte, et aussi le plus petit volume d'une variété hyperbolique ayant exactement un cusp. C. Cao et R. Meyerhoff **[10]** ont démontré que v_ω est réalisé par le complémentaire du nœud de huit et sa sœur.

Le théorème principal qui fait l'objet de cet exposé peut donc s'énoncer ainsi :

THÉORÈME 1.6 (Gabai-Meyerhoff-Milley [**17**]). — *La variété W est l'unique variété hyperbolique dont le volume est v_0.*

Remarque 1.7. — Pour des résultats concernant la structure des volumes des variétés modelées sur d'autres espaces symétriques, et notamment les aspects arithmétiques de cette question, voir par exemple l'exposé [**47**].

2. COMMENT MINORER LES VOLUMES HYPERBOLIQUES ?

2.1. Rayons des tubes de Margulis

Comme on l'a déjà remarqué, le lemme de Margulis implique que toute variété hyperbolique contient une boule métrique de $\mathbf{H}^3$ plongée de rayon une constante universelle μ_3. La première idée pour minorer v_0 est donc de déterminer une constante de Margulis explicite.

On obtient des résultats un peu meilleurs en minorant le rayon des tubes de Margulis, l'idée générale étant que plus l'âme du tube est courte, plus son rayon est grand. Ainsi, si M est une variété hyperbolique compacte, soit son rayon d'injectivité est assez grand, auquel cas on peut minorer son volume par celui d'une boule de $\mathbf{H}^3$ de rayon assez grand, soit elle possède une géodésique courte, qui est l'âme d'un tube de Margulis dont le rayon est relativement grand, et on minore le volume de M par celui de ce tube.

À titre d'exemple, R. Meyerhoff [**38**] en 1987 obtient une constante de Margulis explicite de 0,104 et prouve que $v_0 \geq 0{,}00064$, ce qui est bien sûr encore très loin du résultat optimal. La preuve repose sur l'*inégalité de Jørgensen* : si A, B sont deux matrices de $\mathrm{SL}_2(\mathbf{C})$ qui engendrent un sous-groupe discret non-élémentaire, alors

$$|\operatorname{tr}([A,B]) - 2| \geq 1 - |(\operatorname{tr} A)^2 - 4|.$$

Le lien provient du fait que la *distance de déplacement* d'une isométrie γ de $\mathbf{H}^3$, c'est-à-dire l'infimum des $d(x, \gamma x)$ pour tous les $x \in \mathbf{H}^3$, peut être obtenue à partir de la trace de la matrice correspondante.

Pour des travaux similaires, basés également sur l'inégalité de Jørgensen, voir [**8, 21, 48**]. Une autre méthode, qui s'applique également en dimension supérieure et pour des groupes discrets d'isométries d'autres espaces symétriques, consiste à minorer la distance de déplacement en fonction de la plus petite valeur propre du laplacien du groupe. Le lecteur intéressé par cette technique, ainsi que par les liens avec la constante isopérimétrique de Cheeger, pourra consulter [**9, 42**].

Remarquons que cette philosophie vaut aussi pour les variétés non-compactes, en remplaçant « tube de Margulis » par « cusp ». En fait, elle est même plus « facile » à mettre en œuvre dans ce cas-là, car il est plus aisé d'obtenir de bonnes minorations

des volumes de voisinages de cusps. C'est pourquoi historiquement le volume v_ω a été déterminé beaucoup plus tôt que v_0.

Dans la suite, il sera pratique d'utiliser la définition suivante : si γ est une géodésique fermée dans une variété hyperbolique M et qu'il existe un tube de Margulis $T \subset M$ dont l'âme est γ, alors on appelle *rayon* (en anglais « tuberadius ») de γ le rayon maximal d'un tel T.

Les méthodes « directes » de minoration du rayon des tubes de Margulis ont été poussées à un haut degré de sophistication et de complexité technique. L'avancée la plus importante est peut-être **[20]** : on y prouve en particulier que si M est une variété hyperbolique compacte, alors la géodésique la plus courte de M a un rayon d'au moins $\log 3/2$, sauf si M appartient à l'une de sept familles bien identifiées. Ces familles peuvent ensuite être étudiées par ordinateur. En particulier, un sous-produit de leur travail est le suivant :

THÉORÈME 2.1 (Gabai-Meyerhoff-N. Thurston **[20]**). — *Soit M une variété hyperbolique de volume v_0. Alors le rayon de la géodésique la plus courte dans M est supérieur ou égal à* $\log 3/2$.

Ce résultat, combiné avec des travaux de Gehring et Martin **[23]**, donne la minoration $v_0 \geq 0{,}16668$, qui est bien meilleure que les résultats précédemment connus. Le résultat de Gehring et Martin a été ensuite amélioré par Przeworski **[44]**, pour donner $v_0 \geq 0{,}276796$.

2.2. Le « drilling » ou forage

Soit N une variété hyperbolique compacte. Soit γ une géodésique fermée de N. Notons $N_\gamma := N - \gamma$. On dit que cette variété est obtenue à partir de N par *forage* (en anglais « drilling ».) Il résulte du théorème d'hyperbolisation de W. Thurston que N_γ est hyperbolique (cf. **[34]**.)

I. Agol **[2]** a découvert une inégalité permettant de majorer le volume de N_γ en fonction de celui de N et du rayon R de γ :

$$\mathrm{Vol}(N_\gamma) \leq (\coth R)^{5/2} (\coth 2R)^{1/2} \mathrm{Vol}(N). \tag{1}$$

La preuve de cette inégalité repose sur le résultat suivant, qui est une extension du théorème du volume minimal de G. Besson, G. Courtois et S. Gallot :

THÉORÈME 2.2 (Boland-Connell-Souto **[7]**). — *Soit M une variété hyperbolique. Alors la métrique hyperbolique réalise le plus petit volume parmi toutes les métriques riemanniennes complètes dont la courbure sectionnelle est bornée et majorée par -1, et la courbure de Ricci minorée par -2.*

L'idée de la preuve de l'inégalité (1) est la suivante : soit T un tube de Margulis d'âme γ et de rayon R . On construit une métrique riemannienne (de courbure variable) sur N_γ qui coïncide avec la métrique hyperbolique sur $N-T$ et dont on majore le volume tout en contrôlant la courbure sectionnelle et la courbure de Ricci. On voudrait appliquer le théorème ci-dessus à cette métrique. Il y a un point technique à surmonter : la métrique construite n'est que de classe $\mathcal{C}^1$. Ce problème est résolu par approximation.

L'inégalité (1), combinée avec le théorème 2.1 et la minoration de v_ω de Cao et Meyerhoff, donne $v_0 \geq 0{,}32$. Ce résultat est légèrement amélioré par Przeworski [**46**], qui obtient $v_0 \geq 0{,}33$.

Une amélioration plus substantielle de l'inégalité (1) a été obtenue par I. Agol, N. Dunfield, P. Storm et W. Thurston [**4**] :

$$\mathrm{Vol}(N_\gamma) \leq (\coth 2R)^3 \left(\mathrm{Vol}(N) + \frac{\pi}{2} l \tanh R \tanh 2R\right). \tag{2}$$

La méthode de preuve de l'inégalité (2) est similaire à celle de l'inégalité (1), et repose sur le théorème du volume minimal de Perelman, qui est une vaste généralisation du théorème de Besson-Courtois-Gallot :

THÉORÈME 2.3 (G. Perelman [**43**], voir aussi [**5, 11, 12, 33, 40**])

Soit M une variété hyperbolique compacte. Alors la métrique hyperbolique réalise le plus petit volume parmi les métriques riemanniennes sur M dont la courbure scalaire est supérieure ou égale à -6.

Pour démontrer l'inégalité (2), on construit sur N_γ une métrique $\mathcal{C}^0$ qui coïncide avec la métrique hyperbolique sur $N - C$ et dont on minore la courbure scalaire. On ne peut pas appliquer directement le théorème de Perelman pour deux raisons : la métrique n'est que $\mathcal{C}^0$, et la variété n'est pas compacte. La première difficulté est résolue par un procédé de lissage inspiré par les travaux de Bray et Miao sur la conjecture de Penrose (cf. l'exposé [**26**]) et repose sur un théorème de M. Simon permettant d'affirmer l'existence en temps petit du flot de Ricci pour une condition initiale $\mathcal{C}^0$. La deuxième l'est grâce à un argument d'approximation géométrique par remplissage de Dehn.

En combinant le théorème 2.1 avec l'inégalité (2), on obtient la minoration $v_0 \geq 0{,}67$.

2.3. Structures Mom

Les résultats de [**4**], [**20**] et [**46**] mis ensemble impliquent l'énoncé suivant :

THÉORÈME 2.4. — *Si N est une variété hyperbolique de volume inférieur ou égal à* $\mathrm{Vol}\, W$*, alors N est obtenue par remplissage de Dehn sur une variété à un cusp de volume au plus* 2,848.

Dans [**19**] et [**17**], D. Gabai, R. Meyerhoff et P. Milley ont obtenu une liste complète des dix variétés à un cusp de volume au plus 2,848. P. Milley [**39**] a analysé pour chacune de ces variétés l'espace des remplissages de Dehn et identifié W comme la variété de plus petit volume parmi elles. Cette analyse se base sur un théorème de D. Futer, E. Kalfagianni et J. Purcell [**15**] qui donne une minoration du volume dès que les coefficients du remplissage de Dehn sont assez grands, de sorte qu'il n'y a qu'un nombre fini de coefficients à regarder, ce qui se fait par ordinateur.

La liste des dix variétés à un cusp de volume au plus 2,848 s'obtient de la façon suivante : Gabai, Meyerhoff et Milley ont introduit une notion technique appelée structure $\mathrm{Mom} - n$. Sans rentrer dans les détails, si n est un entier naturel et M une variété hyperbolique, on dit que M est $\mathrm{Mom} - n$ s'il existe une décomposition en anses de $\bar{M}$ comprenant un voisinage collier d'une composante de $\partial \bar{M}$ à laquelle on attache n 1-anses et n 2-anses de sorte que certaines conditions combinatoires soient satisfaites.

La stratégie de [**17, 19**] consiste à procéder en deux temps : d'abord on classifie les variétés $\mathrm{Mom} - 2$ et $\mathrm{Mom} - 3$. Cela donne une liste de 21 variétés hyperboliques. On montre ensuite que toute variété hyperbolique à un cusp de volume au plus 2,848 est obtenue par remplissage de Dehn sur l'une de ces 21 variétés. Le travail est terminé dans [**39**] par une analyse rigoureuse des espaces de remplissages de Dehn des variétés en question.

Remerciements

Je voudrais remercier E. Breuillard, O. Debarre, D. Gabai, R. Kellerhals, R. Meyerhoff, P. Milley et B. Rémy pour leurs commentaires et suggestions.

RÉFÉRENCES

[1] C. C. ADAMS – The noncompact hyperbolic 3-manifold of minimal volume, *Proc. Amer. Math. Soc.* **100** (1987), p. 601–606.

[2] I. AGOL – Volume change under drilling, *Geom. Topol.* **6** (2002), p. 905–916 (électronique).

[3] I. AGOL, M. CULLER & P. B. SHALEN – Dehn surgery, homology and hyperbolic volume, *Algebr. Geom. Topol.* **6** (2006), p. 2297–2312.

[4] I. Agol, P. A. Storm & W. P. Thurston – Lower bounds on volumes of hyperbolic Haken 3-manifolds, avec un appendice de N. Dunfield, *J. Amer. Math. Soc.* **20** (2007), p. 1053–1077 (électronique).

[5] L. Bessières, G. Besson, M. Boileau, S. Maillot & J. Porti – Geometrisation of 3-manifolds, à paraître dans *Tracts of the E.M.S.*

[6] M. Boileau, S. Maillot & J. Porti – *Three-dimensional orbifolds and their geometric structures*, Panoramas et Synthèses, vol. 15, Soc. Math. France, 2003.

[7] J. Boland, C. Connell & J. Souto – Volume rigidity for finite volume manifolds, *Amer. J. Math.* **127** (2005), p. 535–550.

[8] R. Brooks & J. P. Matelski – Collars in Kleinian groups, *Duke Math. J.* **49** (1982), p. 163–182.

[9] P. Buser – On Cheeger's inequality $\lambda_1 \geq h^2/4$, in *Geometry of the Laplace operator (Proc. Sympos. Pure Math., Univ. Hawaii, Honolulu, Hawaii, 1979)*, Proc. Sympos. Pure Math., XXXVI, Amer. Math. Soc., 1980, p. 29–77.

[10] C. Cao & R. Meyerhoff – The orientable cusped hyperbolic 3-manifolds of minimum volume, *Invent. Math.* **146** (2001), p. 451–478.

[11] H.-D. Cao & X.-P. Zhu – A complete proof of the Poincaré and geometrization conjectures—application of the Hamilton-Perelman theory of the Ricci flow, *Asian J. Math.* **10** (2006), p. 165–492.

[12] ———, Erratum to [**11**], *Asian J. Math.* **10** (2006), p. 663–664.

[13] M. Culler, S. Hersonsky & P. B. Shalen – The first Betti number of the smallest closed hyperbolic 3-manifold, *Topology* **37** (1998), p. 805–849.

[14] B. Everitt, J. Ratcliffe & S. Tschantz – The smallest hyperbolic 6-manifolds, *Electron. Res. Announc. Amer. Math. Soc.* **11** (2005), p. 40–46.

[15] D. Futer, E. Kalfagianni & J. S. Purcell – Dehn filling, volume, and the Jones polynomial, *J. Differential Geom.* **78** (2008), p. 429–464.

[16] D. Gabai, R. Meyerhoff & P. Milley – Volumes of tubes in hyperbolic 3-manifolds, *J. Differential Geom.* **57** (2001), p. 23–46.

[17] ———, Minimum volume cusped hyperbolic three-manifolds, *J. Amer. Math. Soc.* **22** (2009), p. 1157–1215.

[18] ———, Mom technology and hyperbolic 3-manifolds, à paraître aux *Proceedings of the fourth Ahlfors-Bers Colloquium.*

[19] ———, Mom technology and volumes of hyperbolic 3-manifolds, prépublication arXiv:math.GT/0606072.

[20] D. Gabai, R. Meyerhoff & N. Thurston – Homotopy hyperbolic 3-manifolds are hyperbolic, *Ann. of Math.* **157** (2003), p. 335–431.

[21] D. Gallo – A 3-dimensional hyperbolic collar lemma, in *Kleinian groups and related topics (Oaxtepec, 1981)*, Lecture Notes in Math., vol. 971, Springer, 1983, p. 31–35.

[22] F. W. Gehring & G. J. Martin – Inequalities for Möbius transformations and discrete groups, *J. reine angew. Math.* **418** (1991), p. 31–76.

[23] ———, Precisely invariant collars and the volume of hyperbolic 3-folds, *J. Differential Geom.* **49** (1998), p. 411–435.

[24] ———, The volume of hyperbolic 3-folds with p-torsion, $p \geq 6$, *Quart. J. Math. Oxford Ser.* **50** (1999), p. 1–12.

[25] M. Gromov – Hyperbolic manifolds (according to Thurston and Jørgensen), Séminaire Bourbaki, vol. 1979/80, exposé n° 546, *Lecture Notes in Math.* **842** (1981), p. 40–53.

[26] M. Herzlich – L'inégalité de Penrose (d'après H. Bray, G. Huisken et T. Ilmanen...), Séminaire Bourbaki, vol. 2000/01, exposé n° 883, *Astérisque* **282** (2002), p. 85–111.

[27] T. Hild – The cusped hyperbolic orbifolds of minimal volume in dimensions less than ten, *J. Algebra* **313** (2007), p. 208–222.

[28] T. Hild & R. Kellerhals – The FCC lattice and the cusped hyperbolic 4-orbifold of minimal volume, *J. Lond. Math. Soc.* **75** (2007), p. 677–689.

[29] C. D. Hodgson & J. R. Weeks – Symmetries, isometries and length spectra of closed hyperbolic three-manifolds, *Experiment. Math.* **3** (1994), p. 261–274.

[30] D. A. Každan & G. A. Margulis – A proof of Selberg's hypothesis, *Mat. Sbornik* **75** (1998), p. 163–168.

[31] R. Kellerhals – Volumes of cusped hyperbolic manifolds, *Topology* **37** (1998), p. 719–734.

[32] R. Kellerhals & T. Zehrt – The Gauss-Bonnet formula for hyperbolic manifolds of finite volume, *Geom. Dedicata* **84** (2001), p. 49–62.

[33] B. Kleiner & J. Lott – Notes on Perelman's papers, *Geom. Topol.* **12** (2008), p. 2587–2855.

[34] S. Kojima – Deformations of hyperbolic 3-cone-manifolds, *J. Differential Geom.* **49** (1998), p. 469–516.

[35] T. H. Marshall & G. J. Martin – Volumes of hyperbolic 3-manifolds. Notes on a paper of D. Gabai, R. Meyerhoff, and P. Milley ([**16**]), *Conform. Geom. Dyn.* **7** (2003), p. 34–48 (électronique).

[36] S. V. Matveev & A. T. Fomenko – Isoenergetic surfaces of Hamiltonian systems, the enumeration of three-dimensional manifolds in order of growth of their complexity, and the calculation of the volumes of closed hyperbolic manifolds, *Russian Math. Surveys* **43** (1988), p. 3–24.

[37] R. Meyerhoff – Sphere-packing and volume in hyperbolic 3-space, *Comment. Math. Helv.* **61** (1986), p. 271–278.

[38] ———, A lower bound for the volume of hyperbolic 3-manifolds, *Canad. J. Math.* **39** (1987), p. 1038–1056.

[39] P. Milley – Minimum volume hyperbolic 3-manifolds, *J. Topol.* **2** (2009), p. 181–192.

[40] J. Morgan & G. Tian – *Ricci flow and the Poincaré conjecture*, Clay Mathematics Monographs, vol. 3, Amer. Math. Soc., 2007.

[41] W. D. Neumann & D. Zagier – Volumes of hyperbolic three-manifolds, *Topology* **24** (1985), p. 307–332.

[42] A. Nevo – The spectral theory of amenable actions and invariants of discrete groups, *Geom. Dedicata* **100** (2003), p. 187–218.

[43] G. Perelman – Ricci flow with surgery on three-manifolds, prépublication arXiv:math.DG/0303109.

[44] A. Przeworski – Tubes in hyperbolic 3-manifolds, *Topology Appl.* **128** (2003), p. 103–122.

[45] ———, Density of tube packings in hyperbolic space, *Pacific J. Math.* **214** (2004), p. 127–144.

[46] ———, A universal upper bound on density of tube packings in hyperbolic space, *J. Differential Geom.* **72** (2006), p. 113–127.

[47] B. Rémy – Covolume des groupes S-arithmétiques et faux plans projectifs (d'après Mumford, Prasad, Klingler, Yeung, Prasad–Yeung), Séminaire Bourbaki, vol. 2007/08, exposé n° 984, *Astérisque* **326** (2009), p. 83–130.

[48] P. L. Waterman – An inscribed ball for Kleinian groups, *Bull. London Math. Soc.* **16** (1984), p. 525–530.

Sylvain MAILLOT

Institut de Recherche Mathématique Avancée
Université Louis Pasteur et CNRS
7 rue René Descartes
F–67084 Strasbourg Cedex

et, à partir d'octobre 2009 :
Université de Montpellier II
Département de Mathématiques
Case 051
Place Eugène Bataillon
F–34095 Montpellier Cedex 5
E-mail : smaillot@math.univ-montp2.fr