

Astérisque

JEAN BELLISSARD

Le papillon de Hofstadter

Astérisque, tome 206 (1992), Séminaire Bourbaki, exp. n° 745, p. 7-39

http://www.numdam.org/item?id=SB_1991-1992__34__7_0

© Société mathématique de France, 1992, tous droits réservés.

L'accès aux archives de la collection « Astérisque » (<http://smf4.emath.fr/Publications/Asterisque/>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

LE PAPILLON DE HOFSTADTER
[d'après B. Helffer et J. Sjöstrand]

par Jean BELLISSARD

1. ÉLECTRONS DE BLOCH EN CHAMP MAGNÉTIQUE

Le papillon de Hofstadter auquel est consacrée cette contribution, est lié à la description du mouvement d'un électron cristallin plongé dans un champ magnétique uniforme. Ce problème est l'un de ceux qui a mobilisé le plus de physiciens du solide durant ce siècle, au point qu'une bibliographie devrait comporter un minimum de 200 articles essentiels, signés par les plus grands noms de ce domaine. Il n'est pas question d'en donner ici la liste, encore que l'histoire du sujet vaille sans doute la peine d'être écrite. Néanmoins, il est utile d'en retenir quelques points.

La première contribution à cette question est due à Landau [LA], qui, en 1930, donna pour la première fois le calcul du spectre d'énergie d'un électron libre dans un champ magnétique uniforme. Ce fut Peierls [PE] en 1933 qui jeta les bases de la théorie des électrons cristallins dans un métal soumis à un champ magnétique uniforme, et qui montra que la contribution des électrons de conduction à la susceptibilité magnétique était diamagnétique. Dans un cristal de dimension D , en l'absence de champ magnétique et en négligeant les interactions coulombiennes, les électrons de conduction sont décrits comme des particules indépendantes dont l'énergie est donnée par une ou plusieurs fonctions $E(\vec{k})$ du vecteur "quasi-moment" \vec{k} , périodique par rapport au groupe de symétrie du réseau réciproque du cristal considéré. Le quasi-moment représente, à une constante physique

près, la quantité de mouvement de la particule. Ces fonctions sont appelées “fonctions de bande”. Des considérations liées à la statistique de Fermi-Dirac montrent que seuls les porteurs de charges d’énergie proche du niveau de Fermi contribuent au transport de courant. Ainsi dans les cas les plus simples une seule fonction de bande suffit, tandis qu’en général, il en faudra un nombre fini, rarement supérieur à trois. Sachant que les champs magnétiques créés en laboratoire sont toujours beaucoup trop faibles pour produire des effets importants, Peierls proposa de décrire le mouvement de l’électron en présence d’un champ magnétique uniforme au moyen de l’opérateur Hamiltonien H obtenu en substituant au quasi-moment \vec{k} dans la fonction de bande, l’opérateur $\vec{K} = (\vec{P} - e\vec{A})/\hbar$, expression dans laquelle $\vec{P} = -\frac{i}{\hbar}\vec{\nabla}$ représente l’opérateur d’impulsion, tandis que \vec{A} représente le potentiel vecteur magnétique relié au champ magnétique \vec{B} par $\text{rot}(\vec{A}) = \vec{B}$.

La première justification de cette substitution fut l’objet d’un article de Luttinger [LU] en 1950, qui adapta à ce problème une méthode développée par Slater pour traiter les impuretés dans un cristal. Cet article fut le début d’une production abondante de travaux durant les années cinquante, consacrés à la construction de Hamiltoniens effectifs plus ou moins réalistes tenant compte de l’influence du champ magnétique. Parmi ces contributions, celle de Harper [HA], parue en 1955, propose un modèle simplifié susceptible de fournir une description qualitativement correcte du mouvement des porteurs de charge. Sachant que seul le mouvement perpendiculaire au champ magnétique est affecté par ce dernier, Harper se restreint à un problème bidimensionnel dans lequel $\vec{K} = (K_1, K_2)$. En outre, il ne retient dans la fonction de bande que les termes d’ordre les plus bas de son développement de Fourier. Pour un cristal carré, il obtient donc le Hamiltonien suivant:

$$(1) \quad H = 2t\{\cos(K_1) + \cos(K_2)\}$$

où t , appelé “transfert” ou “intégrale d’échange”, représente l’énergie nécessaire à l’électron pour son transfert d’un site à l’autre du cristal. Le point remarquable, dans cette expression, est que cet opérateur, ainsi d’ailleurs que tout autre modèle construit par cette méthode, ne dépend que des

“translations magnétiques” [ZA], à savoir les deux opérateurs unitaires $U_i = e^{iK_i}$ ($i=1,2$) qui satisfont aux relations de commutation:

$$(2) \quad U_1 U_2 = e^{2i\pi\alpha} U_2 U_1 \quad \alpha = \frac{\phi}{\phi_0}$$

où $\phi_0 = \frac{h}{e}$ est le quantum du flux magnétique, h étant la constante de Planck, et e la charge de l'électron. Ces relations sont les relations de commutation canoniques à condition de remplacer la constante de Planck \hbar par le flux normalisé $2\pi\alpha$. Or, pour un cristal réaliste comportant une maille élémentaire de quelques Å plongé dans un champ de quelques Teslas, ce flux normalisé est de l'ordre de 10^{-7} à 10^{-9} . Nous travaillerons dans ce cas dans le régime semi-classique avec une excellente approximation.

La justification mathématique de la substitution de Peierls a été apportée par [BE1, HS0] qui ont pu montrer que les électrons d'énergie proche du niveau de Fermi, sont décrits exactement par un Hamiltonien effectif construit comme une matrice de dimension finie dont les coefficients sont des séries convergentes de monômes dans les translations magnétiques. Ainsi, le modèle de Harper ne constitue-t-il que l'approximation à une bande la plus simple du Hamiltonien effectif décrivant les électrons de conduction.

Le calcul du spectre du modèle de Harper a lui-même une longue histoire. En effet, la présence du champ magnétique brise l'invariance par translation, interdisant le recours au théorème de Bloch pour son calcul. Les méthodes semi-classiques furent utilisées dès la fin des années cinquante pour décrire l'influence du réseau cristallin sur les niveaux de Landau. Cependant, les physiciens avaient aussi remarqué que lorsque le flux normalisé était un rationnel, le Hamiltonien redevenait périodique, si bien que l'on pouvait utiliser la théorie de Bloch dans ce cas. Alors en approximant α par un suite convenable de rationnels il devenait possible de calculer le spectre avec une précision croissante. Ce programme fut mis en oeuvre pour la première fois par Chambers [CH] en 1965, mais il ne calcula le spectre que pour un tout petit nombre de valeurs du flux estimant qu'il n'était pas nécessaire d'aller plus loin, ratant ainsi l'aspect le plus spectaculaire de ce problème, à savoir le caractère fractal du spectre comme fonction du flux. Il fallut attendre 1976, et la thèse de Hofstadter [HO] pour en apercevoir pour la première fois la structure (cf. Fig. 1)...en forme de papillon...fractal.

Il est remarquable de constater que le modèle de Harper ou ses dérivés furent redécouverts et employés dans des situations physiques très variées par plusieurs autres physiciens. En 1978, Aubry [AU] l'utilisa pour décrire la physique des chaînes unidimensionnelles spontanément modulées par le phénomène appelé "transition de Peierls". Dans les années quatre-vingts, à la suite de travaux dus à de Gennes [GE] et Alexander [AL] sur les réseaux de supraconducteurs, Rammal et plusieurs de ses collaborateurs montrèrent théoriquement et expérimentalement que le modèle de Harper permettait de décrire la courbe de transition métal normal-supraconducteur d'un réseau carré de fils supraconducteurs. En particulier la quantification du flux aux valeurs rationnelles du quantum de flux fut mise en évidence expérimentalement [PCR]. Ce modèle fut aussi utilisé pour comprendre la théorie de l'effet Hall quantique [TKN2] découvert en 1981 par von Klitzing et al. [KL]. Enfin tout récemment, ce modèle et ses dérivés sont devenus un passage obligé pour comprendre la théorie de la supraconduction à haute température dans les oxydes de cuivre supraconducteurs [RB1]. Le lecteur intéressé pourra trouver quelques informations dans [BE2].

2. PROPRIÉTÉS ÉLÉMENTAIRES DU MODÈLE DE HARPER

Dans ce paragraphe, nous considérons maintenant le modèle de Harper modifié comme suit:

$$(3) \quad H(\alpha, \mu) = U_1 + U_1^{-1} + \mu(U_2 + U_2^{-1})$$

où les opérateurs U_i sont unitaires et satisfont à l'équation (2) avec $\alpha = \frac{\phi}{\phi_0}$, et μ un nombre réel. Ce modèle sera désigné sous le nom de "Mathieu discret" dans la suite. Celui de Harper correspond à $\mu = 1$. Rappelons que le spectre d'un opérateur H est l'ensemble des nombres complexes z tels que $(z - H)$ n'admette pas d'inverse borné. Le spectre est un sous ensemble fermé du plan complexe. Pour un opérateur quelconque (resp. autoadjoint, unitaire), le spectre est contenu dans \mathbf{C} (resp. dans \mathbf{R} , dans le cercle unité) et nous conviendrons d'appeler "gap" les composantes connexes du complémentaire du spectre dans \mathbf{C} (resp. dans \mathbf{R} , dans le cercle unité).

Rappelons aussi qu'un ensemble de Cantor est un fermé nulle part dense sans point isolé.

Les résultats suivants résument les propriétés les plus simples de ce modèle.

THÉORÈME 2.1 (Dualité d'Aubry, cf. [AA, AU, BE3]). *Les spectres de $H(\alpha, \mu)$ de $\mu H(\alpha, \frac{1}{\mu})$, de $H(\alpha, -\mu)$ et de $H(1 \pm \alpha, \mu)$ sont égaux.*

THÉORÈME 2.2 (cf. [BES, vM, CEY]). *Si $\alpha = \frac{p}{q}$ est rationnel et si $\mu \neq 0$, le spectre de $H(\alpha, \mu)$ est formé de q bandes séparées par des gaps. Si q est impair, tous les gaps sont ouverts. Si q est pair, tous les gaps sauf le gap central correspondant à l'énergie nulle, sont ouverts.*

THÉORÈME 2.3 (cf. [BES, vM, CEY]). *Il existe dans $[0, 1]$ un G_δ dense Ω tel que si $\alpha \in \Omega$ et si $\mu \neq 0$, le spectre de $H(\alpha, \mu)$ est un ensemble de Cantor.*

THÉORÈME 2.4 (Formule d'Aubry, cf. [AA, TH, AMS]). *Si α est irrationnel, et si $\mu \neq 1$ la mesure de Lebesgue de $H(\alpha, \mu)$ est égale à $4|1 - \mu|$.*

THÉORÈME 2.5 (cf. [SI]). *Il existe $\mu_c > 0$ tel que si $\mu \geq \mu_c$, le spectre de $H(\alpha, \mu)$ est un ensemble de Cantor pour presque tout α .*

3. ASPECTS ALGÈBRIQUES

3.1. L'introduction des unitaires U_1 et U_2 dans le §1 permet de donner une description de l'opérateur de Harper et de ses dérivés en termes purement algébriques. En effet, un très grand nombre de résultats concernant le spectre comme ensemble ne dépendent que de la relation de commutation (2), et pas du tout du modèle considéré. Cette remarque a été exploitée dans [RB2, BE3] de façon systématique.

Désignons par \mathcal{P}_α l'algèbre complexe engendrée par deux unitaires U_1 et U_2 satisfaisant aux relations de commutation (2). Si nous posons

$$(4) \quad W(m) = U_1^{m_1} U_2^{m_2} e^{-i\pi\alpha m_1 m_2},$$

pour $m = (m_1, m_2) \in \mathbf{Z}^2$, cette algèbre est formée des polynômes de la forme

$$(5) \quad a = \sum_{m \in \mathbf{Z}^2} a_m W(m),$$

où $a_m \in \mathbf{C}$ seront appelés “coefficients de Fourier” de a et $a_m = 0$ sauf pour un nombre fini de termes. Les coefficients de Fourier du produit ab sont donnés par la formule

$$(6) \quad (ab)_m = \sum_{m \in \mathbf{Z}^2} a_{m'} b_{m-m'} e^{i\pi\alpha m' \wedge m},$$

où l'on a posé $m' \wedge m = m'_1 m_2 - m'_2 m_1$, tandis que ceux de l'adjoint a^* de a sont donnés par:

$$(7) \quad (a^*)_m = \overline{a_{-m}}.$$

L'opérateur identité est alors donné par la série de Fourier $\mathbf{1}_m = \delta_{m,0}$, tandis que les unitaires U_1 et U_2 sont donnés par les séries $\delta_{m,(1,0)}$ et $\delta_{m,(0,1)}$ respectivement.

La construction même de \mathcal{P}_α montre que cette algèbre est isomorphe à $\mathcal{P}_{\alpha+1}$, ce qui nous permettra de restreindre l'ensemble des valeurs de α à l'intervalle $[0, 1]$.

Nous pourrions aussi voir α comme une variable, en considérant, pour chaque compact I de \mathbf{R} , l'algèbre $\mathcal{P}(I)$ obtenue comme précédemment au moyen des séries de Fourier finies dont les coefficients sont maintenant des fonctions continues de la variable α sur I , et dont le produit et l'adjoint sont définis par les formules (6) et (7). Nous retrouvons l'algèbre \mathcal{P}_α si $I = \{\alpha\}$.

Il existe alors un homomorphisme naturel $\eta_\alpha : \mathcal{P}(I) \rightarrow \mathcal{P}_\alpha$ qui consiste à évaluer les coefficients de Fourier au point α .

Nous définissons une forme linéaire τ par la formule:

$$(8) \quad \tau(a) = a_0,$$

qui prend ses valeurs dans l'espace $\mathcal{C}(I)$ des fonctions continues sur I . Cette forme est une “trace”, c'est-à-dire qu'elle satisfait aux propriétés suivantes:

$$(9) \quad \tau(a^*a) \geq 0, \quad \tau(ab) = \tau(ba), \quad a, b \in \mathcal{P}(I).$$

3.2. Une représentation π de $\mathcal{P}(I)$ est la donnée d'un espace de Hilbert \mathcal{H}_π et d'un homomorphisme noté $\pi : \mathcal{P}(I) \rightarrow \mathcal{B}(\mathcal{H}_\pi)$ tel que $\pi(a^*) = \pi(a)^*$. Ici $\mathcal{B}(\mathcal{H}_\pi)$ désigne l'algèbre des opérateurs bornés sur \mathcal{H}_π . Nous ne distinguerons pas entre deux représentations unitairement équivalentes. Dans ce cas les classes d'équivalence unitaire forment un ensemble noté $\mathbf{Rep}(\mathcal{P}(I))$. Une norme de C*-algèbre est alors définie sur $\mathcal{P}(I)$ par:

$$(10) \quad \|a\| = \sup_{\pi \in \mathbf{Rep}(\mathcal{P}(I))} \|\pi(a)\|.$$

L'algèbre $\mathcal{A}(I)$ est obtenue comme la complétion de $\mathcal{P}(I)$ par rapport à cette norme. Il est alors facile de montrer que la trace, l'homomorphisme η_α ainsi que toutes les représentations de $\mathcal{P}(I)$ se prolongent par continuité de façon unique à $\mathcal{A}(I)$ [BE3].

Une famille de représentations π est dite "fidèle" si l'intersection des noyaux de chacun de ses éléments est réduit à zéro. Si la famille est réduite à un élément π , nous dirons que π est fidèle. L'intérêt de telles familles réside dans la propriété suivante:

LEMME 3.1 (cf. [PED]). *Soit $\{\pi_s\}_{s \in S}$, une famille fidèle de représentations de $\mathcal{A}(I)$ indexée par un ensemble S . Alors, pour tout $a \in \mathcal{A}(I)$, le spectre de a est la fermeture de la réunion sur S des spectres des opérateurs $\pi_s(a)$, à savoir :*

$$(11) \quad Sp(a) = \overline{\bigcup_{s \in S} Sp(\pi_s(a))}.$$

De plus, la norme de a est donnée par:

$$(12) \quad \|a\| = \sup_{s \in S} \|\pi_s(a)\|.$$

Les représentations suivantes de \mathcal{A}_α sont les plus utilisées.

(i)- La représentation de champ magnétique définie au §1, $\pi_{M,\alpha}$, agit sur l'espace $L^2(\mathbf{R}^2)$ par les relations :

$$(13) \quad \pi_{M,\alpha}(U_i) = e^{iK_i}, \quad K_i = \frac{\delta}{\hbar} \left(\frac{\hbar}{i} \frac{\partial}{\partial x_i} - eA_i \right),$$

où e est la charge de l'électron, $\mathbf{A} = (A_1, A_2)$ est le potentiel vecteur donné par le champ magnétique B au moyen de la relation $\partial_1 A_2 - \partial_2 A_1 = B = \text{const.}$, et où δ est une longueur représentant la période du réseau considéré. Dans ce cas, α est le rapport du flux dans la cellule carrée de côté δ et du quantum de flux h/e (cf. éq. (2)). Cette représentation est fidèle [BE3].

(ii)- π_{GNS} est définie sur l'espace $l^2(\mathbf{Z}^2)$ par les relations (où $\psi \in l^2(\mathbf{Z}^2)$) :

$$(14) \quad \pi_{GNS}(U_i)\psi(m) = e^{i\pi\alpha e_i \wedge m} \psi(m - r_i), \quad m \in \mathbf{Z}^2,$$

où $r_i (i = 1, 2)$ sont les deux vecteurs de la base canonique de \mathbf{Z}^2 . π_{GNS} aussi, est fidèle. Le nom de GNS se rapporte à Guelfand-Naïmark-Segal, qui associèrent à tout état sur une C^* -algèbre une représentation naturelle [PED]. Dans notre cas, l'état considéré est la trace. En outre cette représentation correspond à ce que les physiciens appellent "la représentation de liaison forte", dans laquelle les porteurs de charge sont considérés comme des particules sur le réseau \mathbf{Z}^2 au lieu de se mouvoir dans l'espace continu.

(iii)- La famille $\{\pi_{\alpha,x}; x \in \mathbf{R}/\mathbf{Z}\}$ agit sur l'espace $l^2(\mathbf{Z})$ par (où $\varphi \in l^2(\mathbf{Z})$) :

$$(15) \quad \pi_{\alpha,x}(U_1)\varphi(n) = \varphi(n - 1), \quad \pi_{\alpha,x}(U_2)\varphi(n) = e^{2i\pi(x-n\alpha)}\varphi(n), \quad n \in \mathbf{Z},$$

En particulier, l'opérateur de Mathieu discret $H(\mu)$ (cf. eq. (3)) agit dans cette représentation comme suit:

$$(16) \quad \pi_{\alpha,x}(H(\mu))\varphi(n) = \varphi(n + 1) + \varphi(n - 1) + 2\mu \cos 2\pi(x - n\alpha)\varphi(n).$$

C'est sous cette forme que Harper [HA] avait donné initialement son équation dans le cas $\mu = 1$. Cette famille est fidèle pour tout α . De plus, si α est irrationnel, chacune des représentations $\pi_{\alpha,x}$ est fidèle, ce qui n'est plus vrai par contre, si α est rationnel.

(iv)- La représentation de Weyl $\pi_{W,\alpha}$ est définie sur l'espace $L^2(\mathbf{R})$ au moyen de (où $f \in L^2(\mathbf{R})$) :

$$(17) \quad \pi_{W,\alpha}(U_1)f(x) = f(x + 2\pi\alpha), \quad \pi_{W,\alpha}(U_2)f(x) = e^{ix}f(x), \quad x \in \mathbf{R}.$$

Cette représentation est fidèle. C'est en fait celle qu'utilisent les experts du calcul pseudo-différentiel comme Helffer et Sjöstrand.

Les représentations précédentes fournissent aussi des représentations de $\mathcal{A}(I)$ en les composant avec l'homomorphisme η_α . Les familles obtenues en faisant varier α dans I sont, dans chacun des cas, des familles fidèles de représentations de $\mathcal{A}(I)$.

3.3. L'algèbre \mathcal{A}_α possède une structure assez simple, qui la rend particulièrement intéressante à utiliser. Dans les cas où α est irrationnel, nous obtenons la :

PROPOSITION 3.2 (cf. [BE3]). *Si α est un nombre irrationnel, l'algèbre \mathcal{A}_α est simple. En particulier, toute représentation de \mathcal{A}_α est fidèle.*

Dans le cas pour lequel $\alpha = p/q \in \mathbf{Q}$, où p et q sont des entiers premiers entre eux, nous obtenons la représentation suivante :

PROPOSITION 3.3 (cf. [BE3]). *Si $\alpha = p/q \in \mathbf{Q}$, où p et q sont des entiers premiers entre eux, l'algèbre \mathcal{A}_α est isomorphe à la sous-algèbre des fonctions continues définies sur le 2-tore $\mathbf{T}^2 = (\mathbf{R}/\mathbf{Z})^2$ à valeurs dans l'espace des matrices complexes $q \times q$, engendrée par les deux fonctions suivantes*

$$(18) \quad U_i(k) = e^{ik_i} u_i, \quad k = (k_1, k_2) \in \mathbf{T}^2, \quad i = 1, 2,$$

où les u_i sont des matrices unitaires $q \times q$ définies par

$$(19) \quad u_i^q = \mathbf{1}, \quad u_1 u_2 = e^{2i\pi \frac{p}{q}} u_2 u_1.$$

Nous pouvons choisir par exemple pour les u_i les matrices suivantes:

$$(20a) \quad u_1 = \begin{pmatrix} 0 & 1 & 0 \dots & 0 & 0 \\ 0 & 0 & 1 \dots & 0 & 0 \\ 0 & 0 & 0 \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 \dots & 0 & 1 \\ 1 & 0 & 0 \dots & 0 & 0 \end{pmatrix},$$

$$(20b) \quad u_2 = \begin{pmatrix} 1 & 0 & 0 \dots & 0 & 0 \\ 0 & e^{2i\pi \frac{p}{q}} & 0 \dots & 0 & 0 \\ 0 & 0 & e^{2i\pi \frac{2p}{q}} \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 \dots & e^{2i\pi \frac{(q-2)p}{q}} & 0 \\ 0 & 0 & 0 \dots & 0 & e^{2i\pi \frac{(q-1)p}{q}} \end{pmatrix}.$$

Cette représentation est particulièrement pratique pour le calcul numérique du spectre d'un élément H de $\mathcal{A}_{p/q}$. En effet, H devient une fonction de la variable k à valeurs matricielles. Pour chacune des valeurs de k , il suffit donc de diagonaliser une matrice de dimension finie, ce que l'on peut faire relativement aisément avec un ordinateur usuel. Chaque valeur propre définit ainsi une fonction $E_r(k)$ de k ($r \in [0, q - 1]$), et le spectre s'obtient en prenant la réunion de toutes ces valeurs possibles lorsque k varie dans le 2-tore. En fait, l'image de ce tore par E_r définit ce que l'on appelle une "bande", dont il suffit de connaître les extrémités. Dans la plupart des cas étudiés dans la littérature, cette remarque permet de réduire le calcul à la diagonalisation d'un nombre relativement petit de matrices. C'est par exemple le cas pour le modèle de Harper sur un triangle (cf. Fig. 2).

3.4. Il est remarquable de constater, au vu des figures 1 & 2 montrant des spectres obtenus pour des valeurs rationnelles de α par la méthode précédente, que lorsque α varie, ceux-ci prennent une forme fractale qui laisse à penser que dans le cas irrationnel, il s'agirait d'ensembles de Cantor. Cependant pour que cette conclusion soit correcte, il est nécessaire de s'assurer que le cas irrationnel est bien approximé par les cas rationnels. L'approche algébrique permet en effet de montrer que le spectre d'un élément de $\mathcal{A}(I)$ est continu par rapport à α :

PROPOSITION 3.4 (cf. [BE§]). *Si H est un élément normal de $\mathcal{A}(I)$ (i.e. $HH^* = H^*H$), les bords des gaps du spectre de $\eta_\alpha(H)$ sont des fonctions continues de α .*

Nous verrons dans les paragraphes suivants que l'étude du spectre pour des valeurs de α proches de zéro est possible par une analyse semi-classique.

Comme l'avait remarqué dès 1981 Sokoloff [SO], il est possible d'en déduire la forme du spectre au voisinage de n'importe quel rationnel. En effet, l'utilisation de la substitution de Peierls, d'une part et de la représentation de l'algèbre, dans le cas rationnel par des matrices dépendant de $k \in \mathbf{T}^2$ conduit à la représentation suivante de $\mathcal{A}_{p/q+\alpha}$:

PROPOSITION 3.5 (cf. [BE3]). *L'algèbre $\mathcal{A}_{p/q+\alpha}$ est isomorphe à la sous-algèbre de $\mathcal{A}_\alpha \otimes M_q$ engendrée par les éléments*

$$(21) \quad \hat{U}_i = U_i \otimes u_i, \quad i = 1, 2.$$

Ce dernier résultat permet donc de réaliser une analyse semi-classique au voisinage des rationnels, le prix à payer étant cependant d'avoir à travailler avec des opérateurs matriciels.

4. ANALYSE SEMI-CLASSIQUE

4.1. Le Groupe de Renormalisation, selon Wilkinson

Dans sa thèse [WI] Wilkinson a développé de façon systématique une méthode fondée sur l'analyse semi-classique, qui permet de comprendre pourquoi le spectre de Hofstadter se reproduit à toute échelle. Nous remarquons en effet que la relation de commutation (2) n'est autre qu'un cas particulier des relations de Weyl, à condition de remplacer la constante de Planck \hbar par $2\pi\alpha$. Ainsi la limite $\alpha \rightarrow 0$ n'est-elle autre que la limite semi-classique. Examinons tout d'abord la situation dans laquelle α est petit.

Suivant alors la méthode de Bohr & Sommerfeld, nous considérerons d'abord le problème classique associé. Pour cela, soit I un intervalle de la forme $[0, \varepsilon]$ et soit $H \in \mathcal{A}(I)$ un élément auto-adjoint. Pour simplifier l'étude nous supposerons que la série de Fourier de H décroît aussi rapidement qu'il sera nécessaire, et que les coefficients de Fourier sont des fonctions de α aussi régulières qu'il sera nécessaire. Nous associons à H la

fonction de Hamilton classique \mathcal{H} définie sur le 2-tore \mathbf{T}^2 par:

$$(22) \quad \mathcal{H}(k) = \sum_{m \in \mathbf{Z}^2} H_m e^{ik \cdot m},$$

où les composantes de k seront désignées par $k = (x, \xi)$ pour se conformer aux notations de Hörmander utilisées en analyse micro-locale, tandis que $k \cdot m = xm_1 + \xi m_2$ désigne le produit scalaire. Nous obtenons donc un problème de Mécanique Classique à un seul degré de liberté. L'espace des phases est ici \mathbf{R}^2 , mais la périodicité de \mathcal{H} par rapport aux deux variables (x, ξ) permet de ramener l'étude à une seule cellule de période. Par conséquent, les orbites classiques dans l'espace des phases sont les composantes connexes des courbes de niveau de \mathcal{H} qui forment des ensembles invariants par les translations multiples entières de 2π . Dans le cas du modèle de Harper (avec $\mu = 1$ par conséquent), $\mathcal{H}(x, \xi) = \cos x + \cos \xi$ et les courbes de niveau sont de deux sortes (cf. Fig.3) :

- (i)- Les orbites homotopes à un cercle centré sur un translaté de $(0, 0)$ ou de (π, π) . Ces orbites entourent donc un maximum ou un minimum respectivement.
- (ii)- Les séparatrices, à savoir les droites de la forme $x \pm \xi = \pi(\text{mod}.2\pi)$. Ces droites se croisent aux points de selle de \mathcal{H} , situés en $(0, \pi)$ et $(\pi, 0)$ ($\text{mod}.2\pi$). Les orbites sont alors les différents segments reliant deux points de selle voisins sur l'une de ces droites.

Remarquons que cette structure ne change pas à homotopie près si l'on considère un élément H_1 proche de H , à condition que les symétries de H soient préservées, à savoir par exemple, la symétrie de rotation de $\frac{\pi}{2}$ dans l'espace des phases. Pour un autre modèle la structure des orbites peut devenir plus compliquée. Nous ne nous en occuperons pas ici.

La quantification de Bohr-Sommerfeld consiste à ne retenir que les orbites de la forme Γ satisfaisant à la formule :

$$(23) \quad \oint_{\Gamma} \xi dx = (n + \frac{1}{2})2\pi\alpha.$$

La valeur de \mathcal{H} sur une telle orbite dépend de n et sera notée E_n qui sera désignée sous le nom d'énergie propre. Elle ne dépend pas de la composante

connexe considérée, en raison de l'invariance par translation. Ces valeurs sont alors les valeurs propres de l'opérateur H à l'ordre le plus bas en puissances de α . La méthode WKB usuelle permet alors de calculer un développement asymptotique de ces valeurs propres en puissances de α , ce qui permet de les connaître modulo $O(\alpha^\infty)$.

Un cas particulier intéressant concerne les orbites proches des maxima ou des minima (fonds de puits). Elles sont en effet presque des cercles, ce qui permet de calculer explicitement le développement des énergies propres sous la forme

$$(24) \quad E_n = \mathcal{H}(k_0) \mp 2\pi\alpha\left(n + \frac{1}{2}\right)\sqrt{\det \partial_i \partial_j \mathcal{H}(k_0)} + O(\alpha^2),$$

où le choix du signe \mp est $+$ pour un minimum et $-$ pour un maximum, tandis que k_0 désigne la position de l'extremum considéré, et n est un entier, le nombre quantique.

Dans la représentation de Weyl, chacune de ces orbites donne lieu à une fonction propre $\phi_{n,m}$ dans $L^2(\mathbf{R})$, où $m \in \mathbf{Z}^2$ numérote le centre de la cellule de période dans laquelle se trouve l'orbite considérée, si bien que chaque valeur propre E_n est infiniment dégénérée. Cette fonction propre approchée peut aussi être calculée par la méthode WKB. En raison de l'effet tunnel, l'intégrale de recouvrement $(\phi_{n,m} | \phi_{n,m'})$ décroît comme $e^{-const. \frac{|m-m'|}{\alpha}}$, si bien que la famille $\{\phi_{n,m}; m \in \mathbf{Z}^2\}$ est orthonormale modulo une erreur exponentiellement petite en α .

En raison même de l'effet tunnel entre les orbites de même énergie situées dans différentes cellules de période, la dégénérescence de la valeur propre E_n sera levée, ce qui va donner lieu à une bande \mathcal{B}_n dont la largeur est de l'ordre de la perturbation créée par cet effet tunnel, donc exponentiellement petite en α . Chacune de ces bandes permet de définir un projecteur spectral de H par la formule de Cauchy :

$$(25) \quad \Pi_n = \frac{1}{2i\pi} \oint_{\mathcal{C}} \frac{dz}{z - H},$$

expression dans laquelle \mathcal{C} désigne un cercle dans le plan complexe, centré sur E_n de rayon suffisamment grand pour entourer la bande \mathcal{B}_n et suffisamment petit pour exclure les autres bandes ainsi calculées. Ce projecteur

peut être calculé modulo une erreur exponentiellement petite en α comme le projecteur sur l'espace engendré par la famille $\{\phi_{n,m}; m \in \mathbf{Z}^2\}$. Il s'ensuit que l'on peut aussi calculer la restriction de H à cette bande en considérant la matrice de H dans la base $\{\phi_{n,m}; m \in \mathbf{Z}^2\}$. Supposant que nous sommes capables en outre, de contrôler les erreurs commises à chaque étape de ce raisonnement, nous obtenons une description de chaque bande ainsi définie par un opérateur effectif H_n donné par une matrice infinie indexée par les points du réseau \mathbf{Z}^2 numérotant les différentes cellules de période.

Une des remarques essentielles de Wilkinson consiste à montrer que cette matrice est de la forme :

$$(26) \quad H_{n;m,m'} = \tilde{E}_n(\alpha)\delta_{m,m'} + C(\alpha)e^{i\pi\alpha' m \wedge m'} \tilde{H}_{n;m'-m},$$

où $\tilde{E}_n(\alpha) = E_n + O(\alpha^2)$, α' est la partie fractionnaire de $\frac{1}{\alpha}$, $C(\alpha)$ est une constante exponentiellement petite en α , et \tilde{H}_m est une suite décroissant exponentiellement vite par rapport à $\frac{m}{\alpha}$. Nous reconnaissons là l'expression de la matrice d'un élément \tilde{H} de l'algèbre $\mathcal{A}_{\alpha'}$ dans la représentation GNS. De plus, le Hamiltonien de sous-bande hérite des mêmes symétries que H , et si α est petit, on peut approximer \tilde{H} , modulo une erreur exponentiellement petite en α , par les éléments de matrices correspondant aux plus proche voisins, si bien qu'à cette approximation, \tilde{H} est de nouveau égal au modèle de Harper. Ainsi le Hamiltonien de chacune des sous-bandes ainsi calculées est-il donné par:

$$(27) \quad \tilde{H}_n = H(\alpha', \mu = 1) + O(e^{-\frac{const.}{\alpha}}),$$

montrant que le spectre de H dans la sous-bande se calcule comme celui de H à condition de "renormaliser" les paramètres de transfert (ici la constante $C(\alpha)$) et le flux normalisé α , et de modifier cet opérateur par une perturbation exponentiellement petite en α .

Comme nous l'avons indiqué dans les précédents paragraphes, il est possible de généraliser cette stratégie à l'analyse du spectre près d'un rationnel p/q . Pour cela, la constante de Planck effective est maintenant égale à la différence $\delta = \alpha - p/q$, et le Hamiltonien classique associé est maintenant une fonction de $\mathcal{H}(x, \xi)$ à valeurs dans l'espace des matrices

$q \times q$. Dans le cas du modèle de Harper, cette matrice est tridiagonale, ce qui rend sa diagonalisation familière à n'importe quel physicien. Les valeurs propres $e_i(x, \xi)$ de cette matrice, à savoir les fonctions de bandes, vont alors jouer un rôle analogue à celui de la fonction de Hamilton classique précédente. En particulier, les “fonds de puits” pour ces fonctions, correspondent aux bords de la bande en question. Il faut cependant ajouter une correction dite de “courbure”, car les vecteurs propres de cette matrice définissent un fibré en droite périodique au dessus de l'espace des phases, ou, si l'on préfère, un fibré au dessus du 2-tore. Ce fibré est en général non trivial, et ses classes de Chern ont pu être reliées à la quantification de la conductivité Hall [TKN2]. Ce calcul conduit à un développement explicite des énergies propres près des bords de bandes, analogue à la formule (24) sous la forme [BE1, BE2, BE3, HS3, RB2]:

$$(28) \quad \begin{aligned} E_{i,n} = e_i(k_0) \mp 2\pi \left| \alpha - \frac{p}{q} \right| \left(n + \frac{1}{2} \right) \sqrt{\det \partial_i \partial_j e_i(k_0)} + \sigma 2\pi \left(\alpha - \frac{p}{q} \right) \\ + O\left(\left(\alpha - \frac{p}{q} \right)^2 \right), \end{aligned}$$

où le choix du signe \mp est $+$ pour un minimum et $-$ pour un maximum, tandis que k_0 désigne la position de l'extremum considéré, que n est un nombre quantique, et que σ désigne l'expression suivante

$$(29) \quad \sigma = \text{tr} \left(\mathcal{P}_i(k_0) \{ \partial_x \mathcal{P}_i(k_0) \partial_\xi \mathcal{H}(k_0) - \partial_\xi \mathcal{P}_i(k_0) \partial_x \mathcal{H}(k_0) \} \right),$$

où $\mathcal{P}_i(x, \xi)$ désigne le projecteur spectral associé à la valeur propre $e_i(x, \xi)$ de la matrice $\mathcal{H}(x, \xi)$. Cette formule, appelée “formule de Wilkinson-Rammal” a été démontrée et généralisée dans [BE1, BE2, HS2, RB2] au cas d'un Hamiltonien quelconque dans l'algèbre $\mathcal{A}(I)$.

4.2. La contribution d'Helfer et Sjöstrand

Les arguments de Wilkinson précédemment exposés forment l'armature de la stratégie de Helfer et Sjöstrand pour une preuve mathématiquement rigoureuse. Deux difficultés sont néanmoins passées sous silence dans cette argumentation: la première concerne le contrôle des erreurs en terme de

la constante de Planck effective, la seconde concerne le calcul du spectre au voisinage des séparatrices. La contribution d'Helffer et Sjöstrand est d'avoir comblé ces deux lacunes.

L'analyse microlocale est un outil mathématique qui permet de résoudre la première difficulté. Et c'est là un apport méthodologique essentiel de la contribution de Helffer et de Sjöstrand. Dans leur premier travail [HS1], ils examinent le cas α petit. Leur second article [HS2] concerne l'analyse près d'un rationnel.

La seconde difficulté est en fait rédibitoire. En effet divers travaux numériques ont permis de se rendre compte que l'essentiel du spectre est concentré au voisinage de l'énergie nulle, énergie qui correspond précisément aux orbites classiques proches de la séparatrice. Dans leurs deux premiers articles sur le papillon de Hofstadter [HS1, HS2], Helffer et Sjöstrand avaient évité la discussion de ce point délicat. C'est dans leur troisième article [HS3] qu'ils fournissent une solution à ce problème, s'inspirant en cela d'un travail ancien d'Azbel [AZ]. Ce résultat est sans doute la partie la plus originale, y compris par rapport aux travaux numériques ou théoriques des physiciens. En effet, l'analyse WKB près des séparatrices avait échappé jusqu'alors à la plupart des travaux antérieurs. Même les résultats d'Azbel ne sont pas exempts d'imprécisions sur ce point.

Pour résumer, donnons les résultats les plus importants auxquels on aboutit par cette méthode:

THÉORÈME 4.1 (cf. [HS3]). *Il existe un entier $C \geq 1$ tel que si α est un irrationnel admettant une décomposition en fraction continue $[a_1, a_2, \dots, a_n, \dots]$ pour laquelle $a_n \geq C, \forall n \geq 1$, le spectre de $H(\alpha, \mu = 1)$ est un ensemble de Cantor de mesure nulle.*

Ce théorème ne fournit pas la totalité des résultats. En effet, on peut aussi extraire de ce travail de nombreuses autres informations quant à la structure du spectre. C'est ainsi que la décomposition en fraction continue de α permet de définir une hiérarchie entre les bandes comme suit: étant donné une bande au voisinage d'une valeur propre approchée du type Bohr-Sommerfeld, on peut lui associer un opérateur renormalisé qui est soit du

type “Harper perturbé près de $\alpha = 0$ ”, soit du type “Harper perturbé près de $\alpha = 1/2$ ” et dans lequel il faut changer α en $\alpha' = \{1/\alpha\}$ (ici $\{x\}$ désigne la partie fractionnaire de x). Cet opérateur possède le même spectre que le Hamiltonien précédent sur cette bande. On peut donc décrire la hiérarchie complètement en itérant cette règle. Par ailleurs, chaque sous-bande de Bohr-Sommerfeld est séparée de ses voisines par une distance d’ordre $O(\alpha)$.

Notons cependant que ce problème est exemplaire en ceci qu’il permet de bien cerner les limites de la méthode. En effet, les erreurs ne sont jamais contrôlées que par des expressions implicites dont on ne connaît que le comportement asymptotique à petite constante de Planck. Pour pouvoir conclure, il est nécessaire de supposer que cette constante est “suffisamment petite”. Or la stratégie du groupe de renormalisation est récursive, en particulier, en passant d’une étape de la récurrence à la suivante, le flux renormalisé α change en $\alpha' = \{1/\alpha\}$, expression dans laquelle $\{x\}$ désigne la partie fractionnaire de x . Si l’on itère cette relation, on engendre le développement en fraction continue de α , à savoir:

$$(30) \quad \alpha = \frac{1}{a_1 + \frac{1}{a_2 + \dots \frac{1}{a_n + \dots}}}$$

Sachant qu’à chaque étape il est nécessaire de supposer que α est “suffisamment petit” pour pouvoir passer à l’étape suivante, ceci revient à prouver le résultat désiré pour des α dont la fraction continue satisfait à:

$$(31) \quad \exists C > 0, \quad a_n \geq C, \quad \forall n \geq 1.$$

Or, si $C > 1$, cette condition restreint l’ensemble des valeurs de α pour lesquelles le résultat s’applique à un ensemble de mesure de Lebesgue nulle! De plus, sont exclus de ce résultat la plupart des nombres auxquels se sont intéressés les physiciens, comme par exemple le nombre d’or. En d’autres termes, l’analyse micro-locale fournit certainement une approche qualitativement correcte et puissante, mais elle ne permet pas encore de fournir un résultat quantitativement optimal.

Parmi les conséquences de cette analyse, deux effets spectaculaires décrits soit dans les articles originaux [HS3], soit dans des articles ultérieurs [GH] permettent d'en illustrer la puissance prédictive. Le premier est la résolution d'un paradoxe issu d'une part du théorème 2.2, et d'autre part de la prédiction de Wilkinson selon laquelle le Hamiltonien effectif décrivant chaque sous bande est à nouveau celui de Harper. En effet, le théorème 2.2 prédit que si α est rationnel, tous les gaps sont ouverts sauf le gap central à énergie nulle. Soit donc p/q un rationnel de fraction continue $[a_1, a_2, \dots, a_n]$, et désignons par p'/q' un rationnel proche, de fraction continue $[a_1, a_2, \dots, a_n, 2b]$, où $b \geq 1$. Considérons une bande \mathcal{B} qui ne contienne pas l'énergie nulle dans le spectre du modèle de Harper correspondant à $\alpha = p/q$, et soit $\mathcal{B}(\alpha)$ la bande obtenue par déformation de \mathcal{B} lorsque α varie dans un petit voisinage de p/q . Le flux normalisé correspondant, apparaissant dans le Hamiltonien effectif, est égal au réel positif α' tel que $\alpha = [a_1, a_2, \dots, a_n + \alpha']$. Ainsi pour le nombre p'/q' nous obtenons $\alpha' = 1/2b$. $\mathcal{B}' = \mathcal{B}(p'/q')$ se décompose donc en $2b$ sous-bandes disjointes en vertu du théorème 2.2. Comme ce nombre est rationnel avec un dénominateur pair, le gap central existe et est ouvert. D'autre part si le Hamiltonien effectif était exactement égal à celui de Harper, en vertu du même théorème, le gap central de cette sous bande devrait être fermé. La solution de ce paradoxe apparent se trouve dans le fait que le Hamiltonien effectif sur \mathcal{B}' diffère du modèle de Harper par une perturbation exponentiellement petite en α' . Ainsi, bien que le gap central soit ouvert, sa largeur est d'ordre $O(e^{-const.2b})$! C'est bien en effet ce que révèle le calcul numérique illustré par la Fig.4 (cf.[GH]). Notons cependant que Elliott et al. [CEY] ont prouvé, par la méthode algébrique, que la largeur de chaque gap, dans le cas $\alpha = p/q$ est minorée par $C.8^{-q}$ où C est une constante numérique explicite.

L'autre effet concerne le gap central. Du fait du théorème 2.2, le gap central pour $\alpha = p/q$ lorsque q est pair, est fermé. Ceci signifie que deux bandes alors se touchent, et une analyse plus détaillée montre que ce contact se fait, dans le modèle de Harper, en un seul point de la cellule de période, et que ce point est conique. L'analyse semi-classique nécessite de tenir compte de cette dégénérescence accidentelle, en traitant un problème de matrices

2×2 . Il s'avère que la forme normale de cette matrice est un opérateur de Dirac dont les valeurs propres sont explicitement calculables (cf. [HS2] et aussi [RB2]). Les niveaux d'énergie correspondants sont alors de l'ordre de $\sqrt{|\alpha - p/q|}$ ce que l'on peut voir sur la Fig.1.

5. CONSTRUCTION DU HAMILTONIEN RENORMALISÉ

5.1. Problème à un puits

La première étape de l'analyse consiste à isoler chaque puits. Nous considérons par exemple le cas d'un maximum, celui d'un minimum se traitant de façon analogue.

Rappelons que si $f(x, \xi)$ est une fonction régulière bornée sur \mathbf{R}^2 , on définit un opérateur \hat{f} sur $L^2(\mathbf{R})$ par la quantification de Weyl comme suit:

$$(32) \quad \hat{f}\psi(x) = \int_{\mathbf{R}^2} \frac{dy d\xi}{4\pi^2\alpha} f\left(\frac{x+y}{2}, \xi\right) e^{\frac{i\xi(x-y)}{2\pi\alpha}} \psi(y), \quad \psi \in L^2(\mathbf{R}).$$

Par ailleurs nous définirons le quantifié gaussien \check{f} par la formule:

$$(33a) \quad (\check{f}\psi)(x) = \int_{\mathbf{R}} \check{f}(x, y)\psi(y)dy, \quad \psi \in L^2(\mathbf{R}),$$

où

$$(33b) \quad \check{f}(x, y) = \frac{1}{\sqrt{\pi}} \int_{\mathbf{R}^2} \frac{dx' d\xi}{4\pi^2\alpha} f(x', \xi) e^{-\frac{(x-x')^2 + (y-y')^2}{2}} e^{\frac{i\xi(x-y)}{2\pi\alpha}}.$$

Le quantifié gaussien d'une fonction positive f est un opérateur positif microlocalisé dans le support de f .

Soit alors $E > 0$ l'énergie au voisinage de laquelle nous cherchons à calculer le spectre, et soit $\epsilon > 0$ tel que $E - 2\epsilon > 0$. Soit aussi $m \in \mathbf{Z}^2$ l'index numérotant le puits considéré. On désigne par χ une fonction régulière de (x, ξ) s'annulant si $(x, \xi) \notin [-\pi, \pi]^2$ ou si $(x, \xi) \in [-\pi, \pi]^2$ et $\mathcal{H}(x, \xi) \leq E - 2\epsilon$, valant 1 sur l'ensemble des points $(x, \xi) \in [-\pi, \pi]^2$ pour lesquels $\mathcal{H}(x, \xi) \geq E - \epsilon$ et telle que $0 \leq \chi(x, \xi) \leq 1$. On pose alors $\chi_m(x, \xi) = \chi(x - 2\pi m_1, \xi - 2\pi m_2)$. L'opérateur $\check{\chi}_m$ désigne alors le quantifié

gaussien de χ_m . Remarquons que les opérateurs unitaires $T(m)$ ($m \in \mathbf{Z}^2$) donnés par:

$$(34) \quad T(m)\psi(x) = e^{im_2(x-2\pi\frac{m_1}{2})/\alpha} \psi(x - 2\pi m_1),$$

satisfont $T(m)\check{\chi}_0T(m)^{-1} = \check{\chi}_m$ et commutent aux opérateurs U_i , $i = 1, 2$ dans la représentation de Weyl. Ils quantifient par conséquent les translations correspondant aux multiples entiers de 2π .

Cette construction nous conduit à définir le Hamiltonien de puits H_m par la formule:

$$(35) \quad H_m = H - (E - \epsilon) \sum_{m'; m' \neq m} \check{\chi}_{m'}$$

Cet opérateur possède une symbole inférieur à $E - \epsilon$ sauf précisément dans le puits m . De plus, grâce à la quantification gaussienne de χ , $H_m \leq (E - \epsilon)\mathbf{1}$ modulo un opérateur compact. Ainsi, la partie du spectre de H_m contenue dans l'intervalle $[E - \frac{\epsilon}{2}, E + \frac{\epsilon}{2}]$ est-elle formée d'un nombre fini de valeurs propres de multiplicité finie. Une manière de calculer ces valeurs propres consiste à utiliser l'approximation WKB, mais pour montrer que l'on obtient ainsi tout le spectre dans la région considérée, Helffer et Robert [HR] ont été amenés à construire, en utilisant les techniques du calcul symbolique, une fonction $f(E, \alpha)$ telle que, dans l'intervalle $[E - \frac{\epsilon}{2}, E + \frac{\epsilon}{2}]$, et pour α suffisamment petit, l'opérateur $f(H_m, \alpha)$ ait même spectre que le Hamiltonien de l'oscillateur harmonique. Notons qu'un résultat analogue a été trouvé par [BV] dans le cadre du formalisme algébrique, où l'on construit perturbativement, mais de façon uniforme par rapport à α , un unitaire qui conjugue les deux opérateurs. Notons aussi que ceci est génériquement faux si le nombre de degrés de liberté n'est pas égal à un. L'avantage de cette méthode, est qu'elle fournit une construction explicite des quasi-modes correspondants, au moyen des vecteurs propres de l'oscillateur harmonique.

5.2. Traitement de l'effet tunnel entre puits

La technique de "comblement des puits" précédemment exposée, avait été auparavant développée pour traiter l'effet tunnel dans l'équation de

Schrödinger $H = P^2/2 + V$ où V est un potentiel à deux puits dégénérés que nous indexerons par \pm . Dans ce dernier cas, l'effet tunnel produit des niveaux d'énergie de la forme $E_{n,\pm} = E_n \pm \Delta E$, où E_n est l'énergie que nous trouverions si les deux puits étaient découplés, tandis que $\Delta E \approx e^{-S/\alpha}$ où S est l'action tunnel entre les deux orbites classiques d'énergie E_n . Les fonctions propres des deux Hamiltoniens H_{\pm} que l'on construit en comblant les puits, correspondant aux valeurs propres situées dans l'intervalle d'énergie considéré, sont égales dans chaque puits, à celles de H , modulo une erreur d'ordre $O(e^{-d/\alpha})$ où $d \approx S$ est "la distance d'Agmon" entre les orbites d'énergie $E + O(\epsilon)$ de deux puits voisins (la distance d'Agmon est définie par la métrique $ds^2 = \sqrt{(V(x) - E)_+} dx^2$). Le calcul de l'écart entre les deux sous-niveaux créés par effet tunnel, ne requiert la connaissance précise des fonctions propres qu'à mi-chemin entre les deux puits, c'est-à-dire en un point où la distance d'Agmon est $d/2$. Le calcul de cet écart au moyen des fonction propres des Hamiltoniens H_+ & H_- fournit par conséquent le résultat exact, modulo une erreur $O(e^{-2d/\alpha})$. Ainsi, par cette méthode, nous sommes sûr de trouver correctement le terme dominant de l'écart d'énergie dû à l'effet tunnel.

Il s'avère que dans le cas d'un Hamiltonien proche de celui de Harper, on peut généraliser la notion de distance d'Agmon dans l'espace des phases. C'est là un point technique à mettre à l'actif du travail d'Helfffer et Sjöstrand [HS1]. Nous savons donc, par cette méthode, que chaque puits m fournit des fonctions propres approchées $\varphi_{n,m}$, correspondant à des valeurs propres $E_n \in [E - \frac{\epsilon}{2}, E + \frac{\epsilon}{2}]$ indépendantes de m en raison de l'invariance de H par les translations multiples de 2π . De plus nous savons que la distance entre ces valeurs propres est $O(\alpha)$ et que le spectre de H ne diffère de cet ensemble de valeurs propres que par une erreur d'ordre $O(e^{-d/\alpha})$ due à l'effet tunnel entre les puits, où d est la distance d'Agmon entre deux puits voisins. En outre, les produits scalaires $(\varphi_{n,m}|\varphi_{n,m'})$ sont dominés par $O(e^{-d(m,m')/\alpha})$ si $d(m,m')$ est la distance d'Agmon entre les puits m et m' . Pour chaque n , la famille $\{\varphi_{n,m}; m \in \mathbf{Z}^2\}$ est donc presque orthonormale. Le procédé de Schmidt permet alors de former une base orthonormale $\{\hat{\varphi}_{n,m}; m \in \mathbf{Z}^2\}$ à partir de cette famille, qui ne diffère de la précédente que par une erreur exponentiellement petite en α .

On peut alors construire un nouvel opérateur $H_n(E)$, dépendant analytiquement de E dans un voisinage $O(\alpha)$ de E_n , agissant dans le sous-espace engendré par cette famille, et tel que le spectre de H dans ce voisinage soit donné par l'ensemble des E' telles que $E' \in Sp(H_n(E'))$. Pour cela, Helffer et Sjöstrand utilisent la "méthode de Grušin" qui consiste à considérer le nouveau problème suivant:

sur l'espace $L^2(\mathbf{R}) \oplus l^2(\mathbf{Z}^2)$, on considère l'opérateur \tilde{H} donné par l'expression :

$$(36a) \quad \tilde{H} = \begin{pmatrix} H - E & R_+ \\ R_- & 0 \end{pmatrix},$$

où

$$(36b) \quad R_- \psi_m = (\hat{\varphi}_{n,m} | \psi), \quad \psi \in L^2(\mathbf{R}),$$

et

$$(36c) \quad R_+ u = \sum_{m \in \mathbf{Z}^2} u_m \hat{\varphi}_{n,m}.$$

Si l'inverse de \tilde{H} est désigné par:

$$(37) \quad \tilde{H}^{-1} = \begin{pmatrix} Q & Q_+ \\ Q_- & Q_{-+} \end{pmatrix},$$

alors $H - E$ admet un inverse borné si et seulement si Q_{-+} possède la même propriété, et nous obtenons dans ce cas:

$$(38) \quad (H - E)^{-1} = Q - Q_+(Q_{-+})^{-1}Q_-, \quad (Q_{-+})^{-1} = -R_+(H - E)^{-1}R_-.$$

Le Hamiltonien effectif est donc donné par la matrice infinie Q_{-+} . Comme dans le raisonnement de Wilkinson, l'analyse des symétries montre que cette matrice peut être vue comme la matrice d'un élément $H_{eff,n}(E)$ de l'algèbre $\mathcal{A}_{\alpha'}$ dans la représentation GNS, auquel nous associerons son représentant $H_n(E)$ dans la représentation de Weyl, de façon à pouvoir réutiliser le calcul symbolique à l'étape suivante.

Comme nous l'avons déjà indiqué, la technique de "comblement des puits" nous permet de connaître avec précision le terme dominant de l'effet tunnel, qui consiste à ne retenir que les termes de plus proches voisins dans la matrice Q_{-+} , lesquels s'écrivent alors comme dans la formule (26) et conduisent donc à un nouvel Hamiltonien de Harper dont on aurait simplement renormalisé le paramètre de transfert t , donné par l'effet tunnel dominant. Ainsi $t = O(e^{-S/\alpha})$, où S est ici une action tunnel. Les autres termes sont des corrections exponentiellement petites en α relativement à ce terme dominant.

5.3. Traitement de la bande centrale

La construction précédente est valide tant que les orbites classiques ne sont pas trop proches de la séparatrice. Si nous nous contentions de n'examiner le spectre que dans les régions d'énergie éloignées de la bande centrale, le schéma de renormalisation nous obligerait à exclure à chaque étape un petit intervalle au centre de chaque sous-bande. Les sous-bandes dans cette partie du spectre sont de taille exponentiellement petite en α , ce qui, par récurrence implique que ce sous ensemble spectral est de dimension fractale nulle. Or les calculs numériques indiquent que la dimension fractale du spectre devrait être égale à $1/2$. C'est donc que l'essentiel de la densité spectrale est concentrée précisément dans la région ainsi exclue. Il est donc indispensable de comprendre ce qui s'y passe.

Pour y parvenir, Helffer et Sjöstrand construisent comme précédemment des quasi-modes micro-localisés près des séparatrices. Il faut pour cela conjuguer le Hamiltonien H (au moins au sens du calcul symbolique) à la forme normale dans un voisinage des points de selle. Cette forme normale peut être prise soit sous la forme $\xi^2 - x^2$ soit sous la forme $x\xi$ au niveau des symboles. C'est ce qu'indique le lemme ci-dessous, dans lequel h est un paramètre petit qui joue le rôle de la constante de Planck, et $D = -i\frac{\partial}{\partial x}$:

Lemme 5.1 (cf. [HS3]). Soit $H(x, hD, h)$ un opérateur pseudo-différentiel formellement analytique, formellement autoadjoint et d'ordre 0. Supposons que son symbole principal \mathcal{H} admette un point de selle non dégénéré en

$(0, 0)$ tel que $\mathcal{H}(0, 0) = 0$. Alors il existe un symbole analytique réel $F(t, h) \sim \sum_0^\infty f_j(t)h^j$ défini pour $|t|$ petit, ainsi qu'un opérateur Fourier intégral analytique et formellement unitaire U provenant d'une transformation canonique classique définie dans un voisinage de $(0, 0)$ et transformant ce point en lui même, tels que

$$(39) \quad U^*F(H, h)U = H_0 = \frac{1}{2}(xhD + hDx).$$

L'équation $(H_0 - \alpha)u = 0$ possède quatre solutions particulières données par $u_\pm(x) = \Theta(\pm x)|x|^{-\frac{1}{2}+i\alpha}$, où Θ est la fonction qui s'annule sur les réels négatifs et vaut 1 sur les réels positifs, ainsi que les transformées de Fourier v_\pm de u_\pm . Toute solution est alors combinaison linéaire soit des u_\pm soit des v_\pm et il existe par conséquent une matrice de "monodromie" permettant de passer d'une représentation à l'autre.

Revenant à l'opérateur H que nous étudions, et qui est égal au modèle de Harper modulo une erreur exponentiellement petite en α , l'invariance par rotation de $\frac{\pi}{2}$ de son symbole impose automatiquement que les points de selle soient situés en $(0, \pi)$ ou en $(\pi, 0)$ (modulo 2π). En chacun de ces points, arrivent donc deux morceaux de séparatrices correspondant à la variété stable du flot hamiltonien engendré par le symbole \mathcal{H} de H , et deux autres morceaux correspondant à la variété instable. Les deux premiers sont associés aux solutions u_\pm les autres à v_\pm de l'équation $(H - E)\psi = 0$ obtenues au moyen du Lemme 5.1 et du résultat précédent. Notons au passage que ces quatre solutions se déduisent les unes des autres par une rotation de $\frac{\pi}{2}$ dans l'espace des phases. Soit alors $\chi \in C_0^\infty$ égale à 1 dans un voisinage du point de selle $(0, \pi)$, et s'annulant dans le complémentaire du disque de rayon $3\sqrt{2}\pi/4$ centré sur ce point. On peut alors normaliser les solutions de sorte que $(i[H, \check{\chi}]u_+|u_+) = 1$, ce qui permet d'introduire la fonction $f = i[H, \check{\chi}]u_+$ ainsi que la famille $f_{m,j}$ obtenues en pivotant f de $j\pi/2$ (ce qui se fait en utilisant la transformation de Fourier) puis en la translatant de $2\pi m$ (au moyen des opérateurs $T(m)$ définis par la formule (34)) dans l'espace des phases.

On considère alors le problème de Grušin associé aux opérateurs R_\pm

avec $R_+ : L^2(\mathbf{R}) \rightarrow l^2(\mathbf{Z}^2 \times \{1, 3\})$ défini par:

$$(40a) \quad R_+ \psi_{m,j} = (f_{m,j} | \psi), \quad m \in \mathbf{Z}^2, j \in \{1, 3\}, \quad \psi \in L^2(\mathbf{R}),$$

et $R_- : l^2(\mathbf{Z}^2 \times \{2, 4\}) \rightarrow L^2(\mathbf{R})$ donné par:

$$(40b) \quad R_- u = \sum_{\mathbf{Z}^2 \times \{2,4\}} u_{m,j} f_{m,j}, \quad u \in l^2(\mathbf{Z}^2 \times \{2, 4\}),$$

qui fournira une matrice 2×2 d'opérateurs effectif $H_{j,j'}, j \in \{1, 3\}, j' \in \{2, 4\}$. Comme précédemment, nous pouvons alors interpréter cette matrice d'opérateurs comme représentant un élément de l'algèbre $\mathcal{A}_{\alpha'} \otimes M_2(\mathbf{C})$ dans la représentation GNS, permettant, en revenant à la représentation de Weyl, de définir l'opérateur renormalisé comme une matrice 2×2 dont les coefficients sont des opérateurs pseudo-différentiels provenant de notre algèbre.

De même qu'auparavant les termes qui ne correspondent pas aux plus proches voisins seront exponentiellement petit relativement aux autres, si bien que nous obtenons ici un Hamiltonien effectif sous forme d'une petite perturbation des termes de plus proches voisins.

Il nous reste donc à connaître la forme canonique prise par un tel opérateur. Notons que le symbole \mathcal{H}_{eff} est maintenant une fonction sur l'espace des phases à valeurs dans l'espace des matrices 2×2 . On se ramène donc à une situation analogue au cas de l'analyse semi-classique au voisinage d'un rationnel. Deux possibilités se présentent selon que les deux fonctions de bande de cette matrice ont des images disjointes ou non. Dans le premier, on se ramène à deux Hamiltoniens proches de celui de Harper, dans l'algèbre $\mathcal{A}_{\alpha'}$. Dans l'autre cas, les symétries du problème et le fait que le Hamiltonien initial soit proche du modèle de Harper, implique que les deux bandes se touchent selon un point conique. On peut alors se ramener au cas du modèle de Harper près de $\alpha = 1/2$.

Nous voyons donc que si l'on veut fermer la récurrence, il faut traiter aussi le cas pour lequel le Hamiltonien initial H est lui même une matrice 2×2 dont le symbole possède deux fonctions de bandes se rejoignant en un point conique. C'est ce que Helffer et Sjöstrand appellent "le cas

complètement dégénéré". L'étude dans ce cas impose un peu plus de combinatoire, mais son traitement n'apporte pas de difficultés nouvelles autres que des questions techniques que nous nous abstenons d'aborder ici.

Nous n'irons pas plus loin sur les détails de la stratégie, mais on peut se convaincre que le passage d'une étape du groupe de renormalisation à la suivante ne nécessite l'analyse que d'un nombre fini petit de cas typiques.

BIBLIOGRAPHIE

- [AL] S. ALEXANDER - *Superconductivity of networks. A percolation approach to the effect of disorder*, Phys. Rev., **B27**, (1983), 1541-1557.
- [AA] S. AUBRY, G. ANDRE - *Analyticity breaking and the Anderson localization in incommensurate lattices*, Ann. Israeli Phys. Soc., **3**, (1980), 133.
- [AU] S. AUBRY - *The new concept of transition by breaking of analyticity in crystallographic models*, Solid State Sci., **8**, (1978), 264.
- [AMS] J. AVRON, P.H.M. van MOUCHE, B. SIMON - *On the Measure of the Spectrum for the Almost Mathieu Operator*, Preprint CALTECH, 1990, Commun. Math. Phys..
- [AZ] Ya. AZBEL - *Energy spectrum of a conduction electron in a magnetic field*, Sov. Phys. JETP, **19(3)**, (1964), 634.
- [BE1] J. BELLISSARD - *C*-Algebras in Solid State Physics: 2D electrons in a uniform magnetic field*, in *Operator Algebras and Application*, Vol.2, D.E.Evans & M.Takesaki Eds., Cambridge University Press, (1988).
- [BE2] J. BELLISSARD - *Almost Periodicity in Solid State Physics and C*-Algebras*, in *The Harald Bohr Centenary*, C.Berg & B.Fuglede Eds., The Danish Royal Acad. Sci., **42:3**, (1989), 35-75.
- [BE3] J. BELLISSARD - *The Rotation Algebra*, Lecture donnée à l'école d'été de Nantes, Juin 1991, en préparation.
- [BKS] J. BELLISSARD, C. KREFT, R. SEILER - *Analysis of the spectrum of a particle on a triangular lattice with two magnetic fluxes by algebraic and numerical methods*, J. Phys., **A24**, (1991), 2329-2353.
- [BES] J. BELLISSARD, B. SIMON - *Cantor Spectrum for the Almost Mathieu Equation*, J. Funct. Anal., **48**, (1982), 408-419.

- [BV] J. BELLISSARD, M. VITTOT - *Heisenberg's picture and non commutative geometry in the semi-classical limit in quantum Mechanics*, Ann. Inst. Henri Poincaré, **52**, (1990), 175-235.
- [CH] W.G. CHAMBERS - Phys Rev., **A140**, (1965), 135-143.
- [CEY] M.D. CHOI, G. ELLIOTT, K. YUI - *Gauss polynomials and the rotation Algebra*, Inventiones Math., (1991).
- [GE] P.G. de GENNES - *Diamagnétisme de grains supraconducteurs près d'un seuil de percolation*, C.R. Acad. Sci., **B292**, (1981), 9-12;
Champ critique d'une boucle supraconductrice ramifiée, C.R. Acad. Sci., **B292**, (1981), 279-282.
- [GH] J.P. GUILLEMENT, B. HELFFER, P. TRETON - *Walk inside Hofstadter's butterfly*, J. Phys. France, **50**, (1989), 2019-2058.
- [HA] P.G. HARPER - *Single Band Motion of Conduction Electrons in a Uniform Magnetic Field*, Proc.Phys. Soc. London, **A68**, (1955), 874.
- [HR] B. HELFFER, D. ROBERT - *Calcul fonctionnel de la transformée de Mellin et applications*, J. Funct. Anal., **53**, (1983).
- [HS0] B. HELFFER, J. SJÖSTRAND - *Microlocal Analysis for the Periodic Magnetic Schrödinger Equation and related questions*, C.I.M.E Lectures (1989), in Lecture Notes in Math..
- [HS1] B. HELFFER, J. SJÖSTRAND - *Analyse semi-classique pour l'équation de Harper (application à l'étude de l'équation de Schrödinger avec champ magnétique)*, Mémoires de la S.M.F., (1989).
- [HS2] B. HELFFER, J. SJÖSTRAND - *Analyse semi-classique pour l'équation de Harper II. Comportement semi-classique près d'un rationnel*, Mémoires de la S.M.F n°40, **118**, (1990), 1-139.
- [HS3] B. HELFFER, J. SJÖSTRAND - *Semi-classical analysis for Harper's equation III*, Mémoires de la S.M.F., (1991).
- [HO] D.R. HOFSTADTER - *Energy levels and wave functions of Bloch electrons in a rational or irrational magnetic field*, Phys. Rev., **B14**, (1976), 2239.
- [KL] K. von KLITZING, G. DORDA, M. PEPPER - *Realization of a resistance standard based on fundamental constants*, Phys. Rev. Letters, **45**, (1980), 494-497.

- [LA] L. LANDAU - *Diamagnetismus der Metalle*, Z. für Phys., **64**, (1930), 629-637.
- [LU] J.M. LUTTINGER - *The Effect of a Magnetic Field on Electrons in a Periodic Potential*, Phys. Rev., **84**, (1951), 814.
- [MO] P. van MOUCHE - *The coexistence problem for the discrete Mathieu operator*, Comm. Math. Phys., **122**, (1989), 23-34.
- [PCR] B. PANNETIER, J. CHAUSSY, R. RAMMAL - *Experimental Determination of the (H, T) Phase Diagram of a Superconducting Network*, J. de Physique Lettres **44**, (1983), L853-L858.
- [PCRV] B. PANNETIER, J. CHAUSSY, R. RAMMAL, J.C. VILLEGIER - *Experimental Fine Tuning of the Frustration: 2D Superconducting Networks in a Magnetic Field*, Phys. Rev. Letters, **53**, (1984), 1845-1848.
- [PED] G. PEDERSEN - *C^* -Algebras and their automorphism groups*, Academic Press, London, New-York, (1979).
- [PE] R. PEIERLS - *Zur Theorie des Diamagnetismus von Leitungselektronen*, Z. für Phys., **80**, (1933), 763-791.
- [RB1] R. RAMMAL, J. BELLISSARD - *Ground State of the Fermi Gas in 2D Lattices with a Magnetic Field: New Exact Results*, Europhys. Letters, **13**, (1990), 205-210.
- [RB2] R. RAMMAL, J. BELLISSARD - *An algebraic semi-classical approach to Bloch electrons in a magnetic field*, J. Phys. France, **51**, (1990), 1803-1830.
- [SI] Ya.G. SINAI - *Anderson localization for the one dimensional difference Schrödinger operator with quasiperiodic potentials*, in Proc. Int. Congress Math. Phys. Marseille 1986, World Scientific Pub. Co., Singapore, (1987), pp. 870-903.
- [SO] J.B. SOKOLOFF - *Unusual band structure, wave functions and electrical conductance in crystals with incommensurable periodic potentials*, Phys. Reports, **126**, (1985), 189-244.
- [TH] D. THOULESS - *Bandwidths for a quasiperiodic tight-binding model*, Phys. Rev., **B28**, (1983), 4272-4276.
- [TKN2] D. THOULESS, M. KOHMOTO, M. NIGHTINGALE, M. den NIJS - *Quantized Hall Conductance in two dimensional periodic potential*, Phys. Rev. Letters, **49**, (1982), 405.

- [WI] M. WILKINSON - *Critical Properties of Electron Eigenstates in Incommensurate Systems*, Proc. Roy. Soc. London, **A391**, (1984), 305-350.
- [ZA] J. ZAK - *Magnetic Translation Group*, Phys. Rev., **A134**, (1964), 1602-1607; *Magnetic Translation Group II: Irreducible Representations*, Phys. Rev., **A134**, (1964), 1607-1611.

Jean BELLISSARD

Université Paul Sabatier

Laboratoire de Physique Quantique - IRSAMC

U.R.A. 505 du C.N.R.S.

118, route de Narbonne

F-31062 TOULOUSE CEDEX, FRANCE

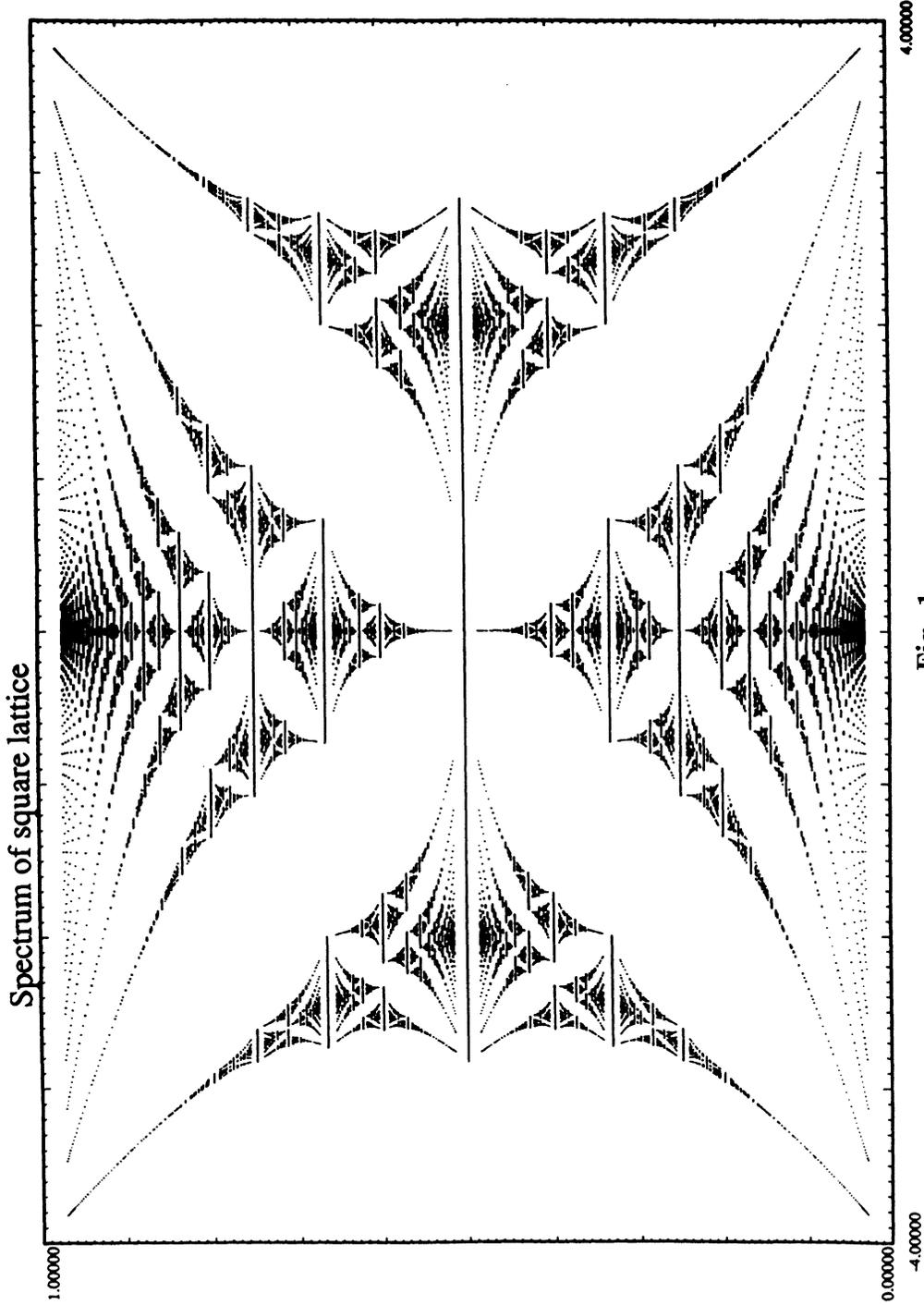


Fig. 1

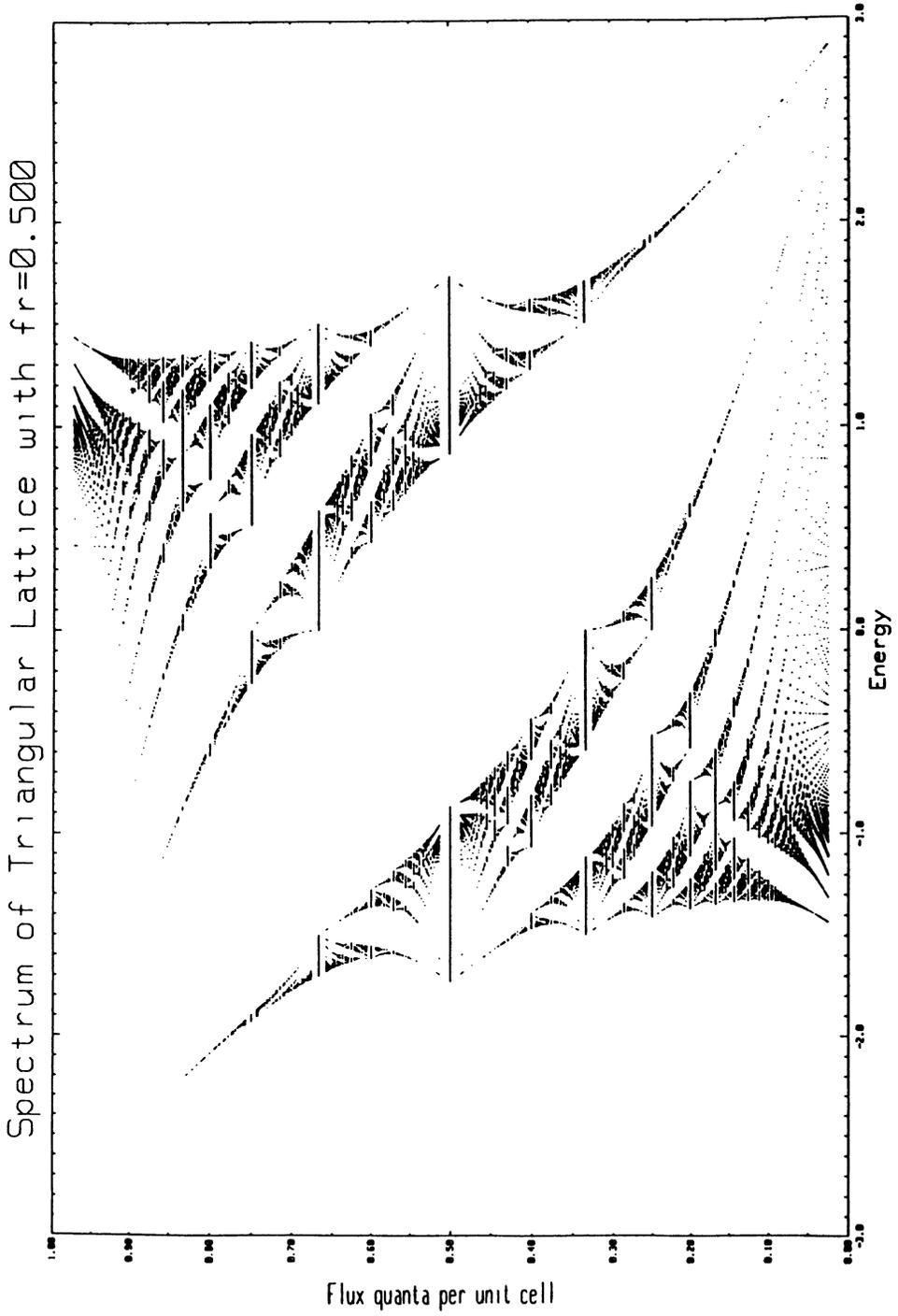
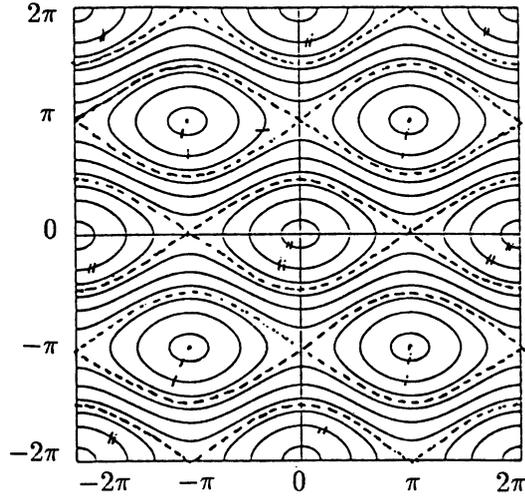


Fig. 2

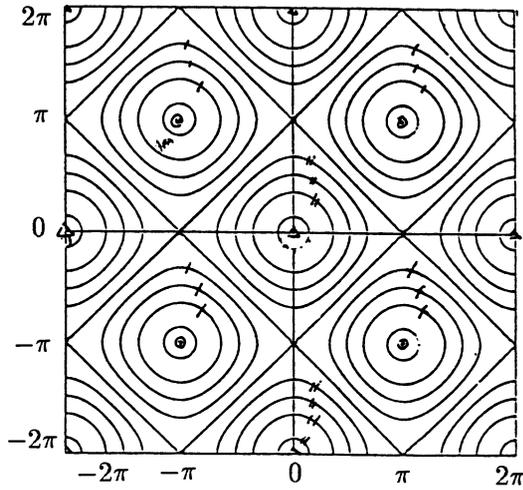


$- : -(1 - \mu) < E < (1 - \mu)$

$+ : -(1 + \mu) < E < -(1 - \mu)$

$\# : 1 - \mu \leq E \leq 1 + \mu$

Pour $\mu = 1$



$- : E = 0 \quad \text{O} : E = -2 \quad + : -2 < E < 0 \quad \Delta : E = 2 \quad \# : 0 < E < 2$

Fig. 3

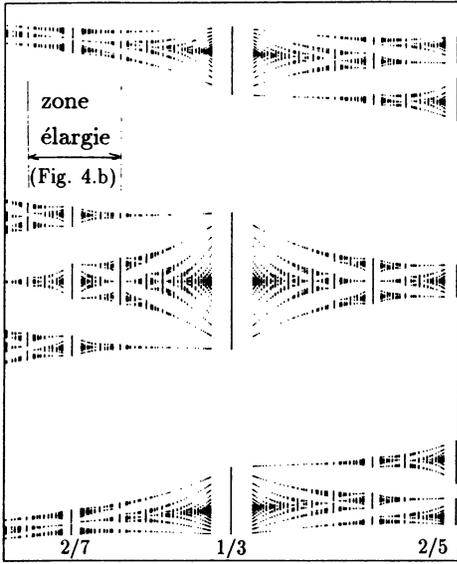


Fig. 4.a

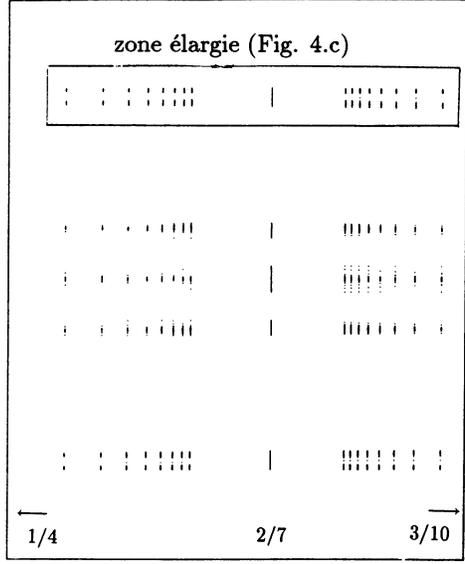


Fig. 4.b

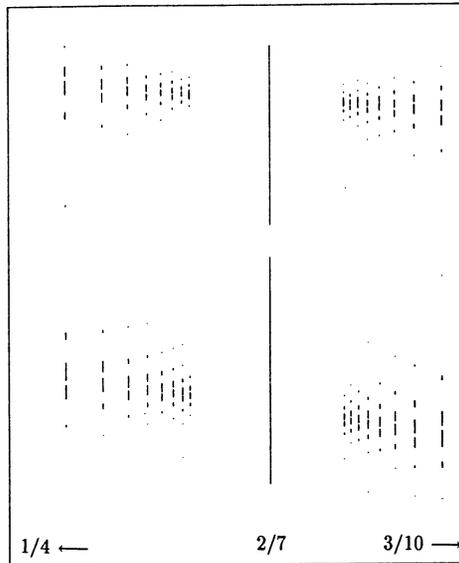


Fig. 4.c

- Spectrum for α near $1/(3 + 1/2)$, upper zoom, $\mu = 1$, partial sequence $\alpha = 1/(3 + 1/(2 + 1/q))$.

Astérisque

PIERRE CARTIER

**Démonstration « automatique » d'identités et
fonctions hypergéométriques**

Astérisque, tome 206 (1992), Séminaire Bourbaki, exp. n° 746, p. 41-91

http://www.numdam.org/item?id=SB_1991-1992__34__41_0

© Société mathématique de France, 1992, tous droits réservés.

L'accès aux archives de la collection « Astérisque » (<http://smf4.emath.fr/Publications/Asterisque/>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

**DÉMONSTRATION “AUTOMATIQUE”
D’IDENTITÉS ET FONCTIONS HYPERGÉOMETRIQUES
[d’après D. Zeilberger]**

par **Pierre Cartier**

INTRODUCTION

Considérons la sélection suivante d’identités entre coefficients binomiaux*

$$(1) \quad \sum_k \binom{n}{k} = 2^n \quad \sum_k (-1)^k \binom{n}{k} = 0$$

$$(2) \quad \sum_k \binom{n}{k}^2 = \binom{2n}{n} \quad \sum_k (-1)^k \binom{2n}{k}^2 = (-1)^n \binom{2n}{n}$$

$$(3) \quad \sum_k (-1)^k \binom{2n}{k}^3 = (-1)^n \binom{3n}{n} \binom{2n}{n} \quad (\text{Dixon 1890}).$$

Les deux premières sont banales, et résultent par spécialisation de la *formule du binôme*

$$(4) \quad (x + y)^n = \sum_{k=0}^n \binom{n}{k} x^k y^{n-k}.$$

Compte tenu de la loi de symétrie $\binom{n}{k} = \binom{n}{n-k}$, les deux formules (2) s’obtiennent en comparant les coefficients de monômes convenables dans les deux identités

$$(x + y)^n (x + y)^n = (x + y)^{2n}, \quad (x + y)^{2n} (x - y)^{2n} = (x^2 - y^2)^{2n}.$$

* Avec les conventions usuelles, on a $\binom{n}{k} = 0$ si $n < 0$ et si k n’est pas compris entre 0 et n . Toutes les séries écrites n’ont donc qu’un nombre fini de termes non nuls, et il est inutile d’écrire les bornes précises de sommation.

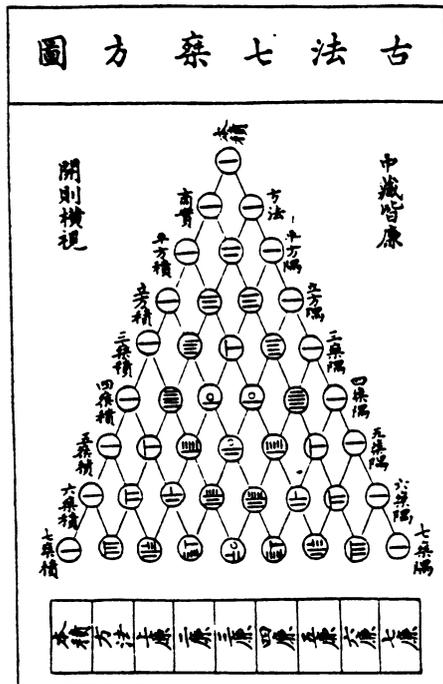
L'identité (3) de Dixon résiste à ce genre de méthodes ; il n'y a pas d'ailleurs de contrepartie portant sur $\sum_k \binom{2n}{k}^3$, ni sur les puissances d'ordre supérieur des coefficients binomiaux.

Pour beaucoup de raisonnements, on peut utiliser la formule du binôme (4) comme *définition* des coefficients binomiaux $\binom{n}{k}$. Comme les puissances se calculent de proche en proche selon la règle

$$(x + y)^n = (x + y)^{n-1}(x + y),$$

les coefficients binomiaux satisfont à la *formule fondamentale d'addition*

$$(5) \quad \binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}.$$



Le «triangle de Pascal» dans un ouvrage chinois de Chou Chi-kié (1303).

[Extrait de Jean-Paul Collette "Histoire des mathématiques", Editions Vuibert.]

Celle-ci permet un calcul rapide au moyen de la disposition connue sous le nom du “triangle de Pascal” (et attestée en Chine autour de 1300, voir figure)

La définition par la formule du binôme conduit à une *interprétation combinatoire* :

$\binom{n}{k}$ compte, parmi tous les mots de n lettres, ceux qui contiennent k fois x et $n - k$ fois y .

De manière équivalente, $\binom{n}{k}$ compte le nombre de parties à k éléments dans un ensemble donné à n éléments. Du coup, les formules entre coefficients binomiaux s’interprètent en termes de décompte de structures combinatoires (graphes, arbres, tableaux, chemins, mots, ...). Dans beaucoup de cas, on a pu prouver ces formules en construisant des *bijections* effectives entre divers types de structure. C’est de cette manière que Foata et moi-même avons établi dans [E 1] la formule de Dixon, et donné diverses généralisations. Cette direction est aujourd’hui vigoureusement poursuivie par les écoles lotharingienne (autour de Foata à Strasbourg) et québécoise (autour des frères Labelle).

Dans les identités (1) à (3), le résultat de la sommation est une expression “close” décrite comme un produit de puissances a^n et de divers coefficients binomiaux. On ne peut pas toujours exprimer aussi simplement le résultat : voici d’abord une formule élémentaire

$$(6) \quad \sum_k \binom{n-k}{k} = F_n,$$

où F_n est le n -ième nombre de Fibonacci, c’est-à-dire le solution de l’équation de récurrence

$$(7) \quad F_n = F_{n-1} + F_{n-2} \text{ pour } n \geq 2, F_0 = F_1 = 1.$$

Dans sa démonstration de l’irrationalité de $\zeta(3)$ Apéry [E 2] a dû sommer l’expression suivante

$$(8) \quad u_n := \sum_k \binom{n}{k}^2 \binom{n+k}{k}^2 ;$$

il a réussi à établir l'équation de récurrence

$$(9) \quad n^3 u_n - (34n^3 - 51n^2 + 27n - 5)u_{n-1} + (n-1)^3 u_{n-2} = 0 \quad (n \geq 2),$$

à joindre aux conditions initiales $u_0 = 1, u_1 = 5$. Apéry n'a pas indiqué sa démonstration de (9), qui fut fournie par Zagier et Cohen [E 3]. L'idée est

la suivante : posons $F(n, k) = \binom{n}{k}^2 \binom{n+k}{k}^2$; il existe alors une autre fonction $G(n, k)$ satisfaisant à l'identité

$$(10) \quad \begin{aligned} G(n, k) - G(n, k-1) &= n^3 F(n, k) - \\ &- (34n^3 - 51n^2 + 27n - 5)F(n-1, k) + (n-1)^3 F(n-2, k). \end{aligned}$$

On a évidemment

$$(11) \quad u_n = \sum_k F(n, k)$$

et de plus, pour n fixé, la "somme télescopique"

$$\sum_k [G(n, k) - G(n, k-1)]$$

s'annule ; la relation (9) se déduit donc de (10) par sommation sur k . La méthode s'appelle (en anglais) "creative telescoping".

On peut démontrer de manière analogue l'identité intégrale classique :

$$(12) \quad \int_{-\infty}^{+\infty} e^{2\pi i x y} e^{-\pi y^2} dy = e^{-\pi x^2}.$$

On note $F(x, y)$ la fonction à intégrer, puis l'on pose

$$(13) \quad u(x) := \int F(x, y) dy.$$

Il s'agit de prouver $u(x) = e^{-\pi x^2}$, ce qui se fait par la caractérisation suivante*

$$(14) \quad D_x u(x) + 2\pi x u(x) = 0, \quad u(0) = 1.$$

* On note D_x, D_y, \dots la dérivation par rapport à x, y, \dots

On détermine une autre fonction $G(x, y)$ telle que

$$(15) \quad D_x F(x, y) + 2\pi x F(x, y) = D_y G(x, y);$$

l'équation (14) se déduit de (15) en intégrant en y , compte tenu des deux règles fondamentales de *dérivation sous le signe intégral* et *d'intégration par parties*

$$(16) \quad D_x \int_{-\infty}^{+\infty} F(x, y) dy = \int_{-\infty}^{+\infty} D_x F(x, y) dy$$

$$(17) \quad \int_{-\infty}^{+\infty} D_y G(x, y) dy = 0 \quad \text{si} \quad G(x, \pm\infty) = 0.$$

Donc, au prix de déterminer la fonction $G(x, y)$, la *formule intégrale* (12) se ramène à une *relation différentielle* (15).

Quelle est la raison du succès dans ces deux cas ? Rappelons d'abord qu'on élimine les coefficients binomiaux au profit des factorielles par la formule

$$(18) \quad \binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

Puisque la factorielle $n!$ est définie par les conditions

$$(19) \quad n! = n.(n-1)! \text{ pour } n \geq 1, \quad 0! = 1,$$

la formule (18) résulte aussitôt de la découverte fondamentale de Pascal qui peut s'exprimer ainsi :

$$(20) \quad \binom{n-1}{k} = \frac{n-k}{n} \binom{n}{k}, \quad \binom{n-1}{k-1} = \frac{k}{n} \binom{n}{k}$$

(la formule d'addition est un corollaire évident !). Dans l'exemple d'Apéry, on a par conséquent

$$(21) \quad \frac{F(n+1, k)}{F(n, k)} = \left(\frac{n+k+1}{n-k+1} \right)^2$$

$$(22) \quad \frac{F(n, k+1)}{F(n, k)} = \frac{1}{(k+1)^4} \left(\frac{n+k+1}{n-k} \right)^2.$$

Il se trouve aussi que le quotient $G(n, k)/F(n, k)$ est une fonction rationnelle

$$(23) \quad \frac{G(n, k)}{F(n, k)} = 4 \left(\frac{n-k}{n+k} \right)^2 (2n-1) \{k(2k+1) - (2n-1)^2\}.$$

Pour démontrer l'identité (10) de Zagier et Cohen, on peut diviser les deux membres par $F(n, k)$ et *se ramener à un calcul trivial, mais incompréhensible, entre fonctions rationnelles de n et k .*

De manière analogue, dans l'exemple d'intégrale donné ci-dessus, les dérivées logarithmiques de la fonction $F(x, y) = e^{2\pi ixy} \cdot e^{-\pi y^2}$ sont des fonctions rationnelles

$$(24) \quad \frac{D_x F(x, y)}{F(x, y)} = 2\pi iy, \quad \frac{D_y F(x, y)}{F(x, y)} = 2\pi i(x + iy)$$

et la fonction $G(x, y)$ qui satisfait à l'équation différentielle (15) vérifie

$$(25) \quad \frac{G(x, y)}{F(x, y)} = -i.$$

La preuve de la relation différentielle (15) est donc ramenée (après division par $F(x, y)$) à celle d'une identité tout à fait élémentaire entre fonctions rationnelles de x et y . Pour achever de prouver la relation intégrale (12), il suffit de la vérifier dans le cas "initial" $x = 0$, c'est-à-dire d'établir la relation $\int_{-\infty}^{+\infty} e^{-\pi y^2} dy = 1$.

Voici pour la *phase de vérification*, mais qu'en est-il de la *phase inductive*? Au vu de la fonction $F(n, k) = \binom{n}{k}^2 \binom{n+k}{k}^2$, est-on certain qu'il existe une équation de récurrence du type (9) satisfaite par la somme $u_n = \sum_k F(n, k)$, et mieux, qu'il existe une fonction $G(n, k)$ satisfaisant à la relation (10). **Comment détermine-t-on explicitement l'équation de récurrence et la fonction $G(n, k)$?**

Dans une remarquable série d'articles, Zeilberger vient de donner deux réponses à ces problèmes :

- une réponse *théorique*, en utilisant la théorie des systèmes différentiels holonomes de Bernstein :

- une réponse *pratique*, en mettant au point des algorithmes qui fournissent l'équation de récurrence et $G(n, k)$.

Toutes les relations mentionnées ci-dessus, y compris l'extraordinaire récurrence d'Apéry, sont retrouvées de manière systématique et automatique, et l'on dispose d'un outil qui permet de *découvrir* et de *démontrer* des identités d'un certain type. Le jour est sans doute proche où les formulaires classiques sur les fonctions spéciales seront remplacés par un logiciel d'interrogation performant, une extension de MAPLE par exemple.

1. MODULES HOLONOMES [C 3]

1.1. On note \mathcal{A}_r l'anneau des *opérateurs différentiels à coefficients polynomiaux* agissant sur les fonctions de r variables : les constantes sont les nombres complexes. Tout opérateur s'écrit de manière unique sous la forme

$$(26) \quad P = \sum_{\alpha, \beta} c_{\alpha\beta} x^\alpha D^\beta.$$

Les notations sont classiques :

- on note $x = (x_1, \dots, x_r)$ le système des r variables et l'on note D_i la dérivation partielle par rapport à x_i :

- on note α, β, \dots des éléments de \mathbf{Z}_+^r , c'est-à-dire des vecteurs $\alpha = (\alpha_1, \dots, \alpha_r)$ à coordonnées entières positives :

- on pose

$$|\alpha| = \alpha_1 + \dots + \alpha_r, \quad x^\alpha = x_1^{\alpha_1} \dots x_r^{\alpha_r}, \quad D^\beta = D_1^{\beta_1} \dots D_r^{\beta_r}.$$

Dans la formule (26), il n'y a qu'un nombre fini de coefficients $c_{\alpha\beta}$ différents de 0.

L'ordre m de l'opérateur $P \neq 0$ est le maximum parmi les nombres $|\beta|$ pour les β tels que $c_{\alpha\beta} \neq 0$: le *symbole principal* $\sigma(P)$ de P est le polynôme $\sum_{|\beta|=m} c_{\alpha\beta} x^\alpha \xi^\beta$ en les variables $(x, \xi) = (x_1, \dots, x_r, \xi_1, \dots, \xi_r)$; il est homogène de degré m en ξ . Voici les règles de base :

- si P est d'ordre m et Q d'ordre n , alors PQ est d'ordre $m + n$ et le commutateur $[P, Q] = PQ - QP$ est d'ordre $\leq m + n - 1$;

- on a *

$$(27) \quad \sigma(PQ) = \sigma(P)\sigma(Q), \quad \sigma([P, Q]) = \{\sigma(P), \sigma(Q)\}.$$

Dans la dernière formule, on a utilisé le *crochet de Poisson* de deux polynômes en x, ξ défini par

$$(28) \quad \{F, G\} = \sum_{j=1}^r \left[\frac{\partial F}{\partial \xi_j} \cdot \frac{\partial G}{\partial x_j} - \frac{\partial G}{\partial \xi_j} \cdot \frac{\partial F}{\partial x_j} \right].$$

On peut présenter les choses de manière plus algébrique. Comme algèbre sur le corps \mathbf{C} des nombres complexes, \mathcal{A}_r est définie par les $2r$ générateurs x_i et D_j soumis aux relations

$$(29) \quad [x_i, x_j] = 0, \quad [D_i, D_j] = 0, \quad [D_i, x_j] = \delta_{ij}.$$

On note $F^m = F^m \mathcal{A}_r$ l'ensemble des opérateurs d'ordre $\leq m$. C'est un espace vectoriel sur \mathbf{C} , la suite (F^m) est croissante ($F^0 \subset F^1 \subset F^2 \subset \dots$), on a $F^0 = \mathbf{C}[x_1, \dots, x_r]$ et les règles données plus haut s'écrivent

$$(30) \quad F^m \cdot F^n \subset F^{m+n}, \quad [F^m, F^n] \subset F^{m+n-1}.$$

Grâce à ces propriétés, on peut définir une structure d'algèbre sur le gradué associé

$$gr^F \mathcal{A}_r = \bigoplus_{m \geq 0} F^m / F^{m-1};$$

on définit un isomorphisme de cette algèbre sur l'algèbre de polynômes $\mathbf{C}[x, \xi]$ qui associe $\sigma(P)$ à la classe de P modulo F^{m-1} (pour tout P dans F^m / F^{m-1}). L'anneau \mathcal{A}_r n'est pas commutatif, mais comme l'anneau $gr^F \mathcal{A}_r$ l'est, on peut considérer que \mathcal{A}_r est presque commutatif et beaucoup des propriétés des anneaux de polynômes s'étendent à \mathcal{A}_r .

* Dans la deuxième formule, on suppose que $[P, Q]$ est d'ordre $m + n - 1$, ce qui équivaut à $\{\sigma(P), \sigma(Q)\} \neq 0$.

1.2. Nous pouvons faire opérer les opérateurs différentiels de \mathcal{A}_r sur les polynômes en x , sur les fonctions de classe C^∞ définies sur \mathbf{R}^r (ou sur un ouvert de \mathbf{R}^r), sur les fonctions analytiques dans \mathbf{R}^r ou les fonctions holomorphes dans \mathbf{C}^r , sur divers espaces de distributions sur \mathbf{R}^r .

Soit par exemple f une fonction de classe C^∞ sur \mathbf{R}^r . Introduisons l'ensemble $I_f \subset \mathcal{A}_r$ formé des opérateurs différentiels $P \in \mathcal{A}_r$ tels que $Pf = 0$; c'est un idéal à gauche de \mathcal{A}_r . Les symboles principaux $\sigma(P)$ correspondant aux P dans I_f sont les éléments homogènes non nuls d'un idéal J_f de $\mathbf{C}[x, \xi]$. A cet idéal de polynômes on associe une variété algébrique dans \mathbf{C}^{2r} , définie par les équations

$$\sigma(P)(x, \xi) = 0 \quad \text{pour tout } P \neq 0 \text{ dans } I_f;$$

c'est la variété caractéristique V_f de f .

Un théorème profond de Kashiwara - Kawai - Sato affirme que la variété V_f est involutive; on en trouvera une démonstration dans l'exposé [C 11] de Malgrange à ce séminaire. Voici la signification: notons J_f^0 l'idéal des polynômes dans $\mathbf{C}[x, \xi]$ dont l'ensemble des zéros contient V_f ; d'après le théorème des zéros de Hilbert, on a $F \in J_f^0$ si et seulement si une puissance de F appartient à J_f . Alors J_f^0 est stable par crochet de Poisson (l'assertion correspondante pour J_f résulte aussitôt de la deuxième formule (27)).

D'après un résultat classique de géométrie symplectique, cela entraîne que la variété V_f est de dimension au moins égale à r : si $p = (x^{(0)}, \xi^{(0)})$ est un point lisse de V_f , l'espace tangent en p à V_f est l'orthogonal des différentielles $d_p F$ pour F parcourant J_f^0 . Mais il existe une forme bilinéaire symplectique J_p sur l'espace cotangent en p telle que

$$(31) \quad \{F, G\}(p) = J_p(d_p F, d_p G).$$

Si F et G sont dans J_f^0 , il en est de même de $\{F, G\}$ par l'involutivité, d'où $J_p(d_p F, d_p G) = 0$. Ceci entraîne que l'ensemble des $d_p F$ pour F dans J_f^0 est un espace vectoriel de dimension $\leq r$, et par dualité que l'espace tangent en p à V_f est de dimension $\geq 2r - r = r$.

1.3. On dira que la fonction (ou la distribution) f est *holonome** si la variété caractéristique V_f est de dimension r (le minimum possible).

En pratique, il est difficile de déterminer explicitement l'idéal à gauche I_f de \mathcal{A}_r , et encore moins les idéaux J_f et J_f^0 de $\mathbf{C}[x, \xi]$. Pour vérifier qu'une fonction f est holonome, on écrira un certain nombre d'équations différentielles

$$P_1 f = \dots = P_m f = 0;$$

soit σ_j le symbole principal de P_j . Sur la variété V_f s'annulent les polynômes σ_j , mais aussi les crochets de Poisson $\{\sigma_j, \sigma_k\}, \{\sigma_j, \{\sigma_k, \sigma_\ell\}\}$ etc... puisque l'idéal J_f est stable par crochet de Poisson. Si l'on peut trouver assez d'équations du type $\sigma_j = 0, \{\sigma_j, \sigma_k\} = 0, \dots$ pour que l'ensemble W des solutions dans \mathbf{C}^{2r} soit de dimension $\leq r$, on aura $\dim V_f \leq \dim W \leq r$ et f est holonome.

Exemples : 1) Si $r = 1$, une fonction holonome sur \mathbf{C} est une fonction qui satisfait à une équation différentielle

$$(32) \quad \sum_{j=0}^m p_j(z).D^j f(z) = 0 ,$$

où $p_0(z), \dots, p_m(z)$ sont des polynômes avec $p_m \neq 0$. Le symbole correspondant est $p_m(z)\zeta^m$ et la variété caractéristique est contenue dans

$$\{\zeta = 0\} \cup \{z = z_{(1)}\} \cup \dots \cup \{z = z_{(d)}\} ,$$

où $z_{(1)}, \dots, z_{(d)}$ sont les racines du polynôme $p_m(z)$. Par exemple, la *fonction hypergéométrique de Gauss* est solution de l'équation

$$z(1-z)\frac{d^2 F}{dz^2} + \{c - (a+b+1)z\}\frac{dF}{dz} - abF = 0;$$

elle est holonome avec la variété caractéristique

$$\{z = 0\} \cup \{z = 1\} \cup \{\zeta = 0\}$$

* Björk parle de la "classe de Bernstein" dans [C 3].

dans \mathbf{C}^2 .

2) Un polynôme $f(z)$ sur \mathbf{C}^r satisfait aux équations

$$D_1^m f = \cdots = D_r^m f = 0$$

pour m assez grand. La variété caractéristique est donc contenue dans la variété linéaire $\zeta_1 = \cdots = \zeta_r = 0$ de dimension r ; par suite, f est holonome.

3) La *distribution de Dirac* δ dans \mathbf{R}^r satisfait aux relations

$$x_1 \delta = \cdots = x_r \delta = 0.$$

Elle est donc holonome avec la variété caractéristique $x_1 = \cdots = x_r = 0$.

1.4. Si I est un idéal à gauche de \mathcal{A}_r , on appelle *variété caractéristique de I* la variété algébrique $V(I)$ dans \mathbf{C}^{2r} , intersection des zéros des polynômes $\sigma(P)$, où P parcourt I (et $P \neq 0$). C'est encore une variété involutive, donc de dimension $\geq r$. On dira que l'*idéal I est holonome* si $V(I)$ est de dimension r . Donc la fonction f est holonome si et seulement si l'idéal I_f est holonome.

Soit S un anneau de polynômes $\mathbf{C}[u_1, \dots, u_s]$. On sait que S est de *dimension homologique* s ; autrement dit, tout S -module M de type fini admet une résolution

$$0 \rightarrow L_s \rightarrow L_{s-1} \rightarrow \cdots \rightarrow L_0 \rightarrow M \rightarrow 0,$$

où les L_i sont des S -modules projectifs de type fini, et il existe un S -module M qui n'admet pas de résolution plus courte. Il revient encore au même de dire que l'on a $\text{Ext}_S^i(M, N) = 0$ pour $i > s$, quels que soient les S -modules de type fini M et N , et qu'il existe M et N tels que $\text{Ext}_S^s(M, N) \neq 0$.

Soit M un S -module de type fini ; introduisons l'idéal J de S formé des p tels que $pM = 0$, puis la variété algébrique W dans \mathbf{C}^s définie par les équations $p(u) = 0$, où p parcourt J (qu'on peut appeler la *variété caractéristique de M*). Notons d ou $d(M)$ la dimension de W . Un théorème classique affirme que l'on a

$$(33) \quad \begin{cases} \text{Ext}_S^j(M, S) = 0 & \text{pour } 0 \leq j < s - d \\ \text{Ext}_S^{s-d}(M, S) \neq 0. \end{cases}$$

Revenons à l'anneau \mathcal{A}_r , avec sa filtration $(F^m)_{m \geq 0}$; on convient qu'on a $F^m = 0$ pour $m < 0$. Soit alors M un \mathcal{A}_r -module à gauche de type fini, et soit (u_1, \dots, u_t) un système générateur ; nous choisirons aussi des entiers positifs d_1, \dots, d_t et nous poserons

$$(34) \quad \Gamma^m M = F^{m-d_1} \cdot u_1 + \dots + F^{m-d_t} \cdot u_t.$$

La suite croissante $\Gamma^0 M \subset \Gamma^1 M \subset \dots$ ainsi obtenue est appelée une *bonne filtration* de M ; comme on a $F^m \cdot \Gamma^n M \subset \Gamma^{n+m} M$, on peut définir le gradué associé $gr^\Gamma M = \bigoplus_{m \geq 0} \Gamma^m M / \Gamma^{m-1} M$, qui est un module sur l'anneau $gr^F \mathcal{A}_r$. Ce module est de type fini.* Comme l'anneau $gr^F \mathcal{A}_r$ est un anneau de polynômes $\mathbf{C}[x_1, \dots, x_r, \xi_1, \dots, \xi_r]$, ce qui précède s'applique. On introduit donc la variété caractéristique $W(gr^\Gamma M)$; compte tenu de la définition (34), on montre facilement que la variété caractéristique $W(gr^\Gamma M)$ est indépendante de Γ ; on la notera donc $W(M)$. Elle est involutive, donc de dimension $\geq r$.

En utilisant les résultats cités plus haut dans le cas où $S = \mathbf{C}[x_1, \dots, x_r, \xi_1, \dots, \xi_r]$, et un argument de suite spectrale lié à la filtration $(\Gamma^m M)_{m \geq 0}$, on montre qu'on a

$$(35) \quad \begin{cases} \text{Ext}_{\mathcal{A}_r}^j(M, \mathcal{A}_r) = 0 & \text{pour } 0 \leq j < 2r - d \\ \text{Ext}_{\mathcal{A}_r}^{2r-d}(M, \mathcal{A}_r) \neq 0, \end{cases}$$

où $d = d(M)$ est la dimension de la variété caractéristique $W(M) \subset \mathbf{C}^{2r}$. Ceci donne une caractérisation de $d(M)$ indépendante de toute filtration sur l'algèbre \mathcal{A}_r ou sur le \mathcal{A}_r -module M . On prouve aussi la relation

$$\text{Ext}_{\mathcal{A}_r}^j(M, \mathcal{A}_r) = 0 \quad \text{pour } j > r,$$

et l'on en déduit assez facilement que *la dimension homologique de \mathcal{A}_r est r (et non $2r$ comme celle de $gr^F \mathcal{A}_r$).*

* Cette propriété donne une caractérisation intrinsèque des "bonnes" filtrations.

On dira bien sûr que le \mathcal{A}_r -module M est holonome* si l'on a $d(M) = r$: il revient au même de supposer que l'on a

$$(36) \quad \text{Ext}_{\mathcal{A}_r}^j(M, \mathcal{A}_r) = 0 \quad \text{pour} \quad j \neq r .$$

Le \mathcal{A}_r -module à droite $\text{Ext}_{\mathcal{A}_r}^r(M, \mathcal{A}_r)$ joue alors le rôle de dual pour M ; il est lui aussi holonome.

1.5. La principale innovation de Bernstein a consisté en l'introduction d'une autre filtration : $B^m = B^m \mathcal{A}_r$ se compose des opérateurs différentiels de la forme

$$(37) \quad P = \sum_{|\alpha|+|\beta| \leq m} c_{\alpha\beta} x^\alpha D^\beta .$$

Autrement dit, on compte le degré total en les x_i et les D_i et non seulement le degré partiel en les D_i . On peut répéter presque mot pour mot ce qui a été énoncé pour la filtration (F^m) :

- anneau gradué associé $gr^B \mathcal{A}_r$, lui aussi isomorphe** à $\mathbf{C}[x_1, \dots, x_r, \xi_1, \dots, \xi_r]$: le B -symbole de P sera

$$(38) \quad \sigma_B(P) = \sum_{|\alpha|+|\beta|=m} c_{\alpha\beta} x^\alpha \xi^\beta$$

avec les notations de (37) :

- bonne B -filtration sur un \mathcal{A}_r -module de type fini M donnée par

$$(39) \quad \Delta^m M = B^{m-d_1} \cdot u_1 + \dots + B^{m-d_t} \cdot u_t ;$$

- gradué associé $gr^\Delta M$: c'est un module de type fini sur $gr^B \mathcal{A}_r$;

- variété B -caractéristique du module $gr^\Delta M$; elle ne dépend pas de la bonne B -filtration sur M ;

* L'idéal à gauche I de \mathcal{A}_r est holonome si et seulement si le \mathcal{A}_r -module à gauche \mathcal{A}_r/I est holonome. Si f est une fonction sur \mathbf{R}^r , le \mathcal{A}_r -module \mathcal{A}_r/I_f est isomorphe à $\mathcal{A}_r \cdot f$ (ensemble des $P \cdot f$ où P parcourt \mathcal{A}_r).

** La graduation utilisée dans $\mathbf{C}[x, \xi]$ sera le degré total en x, ξ .

- dimension $\delta = \delta(M)$ de la variété B -caractéristique du \mathcal{A}_r -module M ;
- caractérisation de $\delta(M)$: le plus petit entier j tel que $\text{Ext}_{\mathcal{A}_r}^j(M, \mathcal{A}_r)$ soit non nul est égal à $2r - \delta(M)$.

De ce dernier résultat, on déduit que la dimension $d(M)$ de la variété caractéristique de M est égale à la dimension $\delta(M)$ de la variété B -caractéristique (les variétés caractéristiques sont distinctes).

La dimension $\delta = \delta(M)$ peut se calculer ainsi : si un \mathcal{A}_r -module de type fini M est muni d'une bonne B -filtration $(\Delta^m M)_{m \geq 0}$, chacun des espaces vectoriels $\Delta^m M / \Delta^{m-1} M$ est de dimension finie sur \mathbf{C} ; on peut donc introduire la série de Poincaré

$$\begin{aligned}
 \chi_{\Delta}(t) &= \sum_{m \geq 0} t^m \cdot \dim \Delta^m M / \Delta^{m-1} M \\
 (40) \qquad &= (1-t) \sum_{m \geq 0} t^m \cdot \dim \Delta^m M.
 \end{aligned}$$

Comme $gr^{\Delta} M$ est un module gradué de type fini sur un anneau de polynômes, il est bien connu (*série de Hilbert - Samuel*) que $\chi_{\Delta}(t)$ est une fonction rationnelle de la forme $Q(t)/(1-t)^{\delta}$ où le polynôme $Q(t)$ satisfait à $Q(1) \neq 0$. De la formule (40) on déduit l'existence d'un entier m_0 et d'un polynôme $X(u)$ à coefficients rationnels tels que

$$\dim \Delta^m M = X(m) \quad \text{pour} \quad m \geq m_0.$$

De plus, δ est le degré du polynôme $X(u)$.

Tous les résultats des n° 1.4 et 1.5 se démontrent indépendamment du théorème d'involutivité de Kashiwara - Kawai - Sato. De plus, il existe une démonstration directe très simple de l'inégalité $\delta \geq r$ (voir [C 6], page 178), d'où une nouvelle preuve de l'inégalité $d \geq r$ (où d est la dimension de la variété caractéristique de M).

1.6. Concluons par un *guide pratique* des modules holonomes.

Un \mathcal{A}_r -module M est un module sur l'anneau de polynômes $\mathbf{C}[x_1, \dots, x_r]$ muni d'opérateurs \mathbf{C} -linéaires de dérivation D_1, \dots, D_r commutant deux à

deux et satisfaisant à la règle de Leibniz

$$(41) \quad D_i(F \cdot \Phi) = D_i F \cdot \Phi + F \cdot D_i \Phi$$

pour $1 \leq i \leq r$, F dans $\mathbf{C}[x_1, \dots, x_r]$ et Φ dans M . Par exemple, tout espace de fonctions ou de distributions à r variables qui est stable par multiplication par x_i et dérivation par rapport à x_i (pour $i = 1, \dots, r$) est un \mathcal{A}_r -module.

Un élément Φ d'un \mathcal{A}_r -module M est *holonome* s'il existe une constante C satisfaisant à la condition suivante :

(H) pour tout entier $m \geq 0$, le sous-espace vectoriel de M engendré par les éléments $x^\alpha D^\beta \Phi$, avec $|\alpha| + |\beta| \leq m$, est de dimension $\leq C \cdot m^r$.

Les éléments holonomes d'un \mathcal{A}_r -module forment un sous- \mathcal{A}_r -module de M . Un module holonome est un \mathcal{A}_r -module de type fini dont tout élément est holonome.

Un \mathcal{A}_r -module rationnel est un espace vectoriel V sur le corps $\mathbf{C}(x_1, \dots, x_r)$ des fonctions rationnelles muni d'opérateurs \mathbf{C} -linéaires D_1, \dots, D_r commutant deux à deux et satisfaisant à la règle de Leibniz. Soit Φ un élément holonome de V ; il existe alors des opérateurs différentiels non nuls P_1, \dots, P_r annihilant Φ et de la forme suivante

$$(42) \quad P_i = \sum_{j=0}^{s_i} p_{ij}(x_1, \dots, x_r) D_i^j$$

(les p_{ij} sont des polynômes). Le sous-espace vectoriel de V , engendré sur $\mathbf{C}(x_1, \dots, x_r)$ par les dérivées $D^\alpha \Phi$ pour $0 \leq \alpha_1 \leq s_1, \dots, 0 \leq \alpha_r \leq s_r$ en nombre fini, contient Φ et il est stable par les dérivations D_1, \dots, D_r .

On dira qu'un élément Φ de V est *rationnellement holonome** s'il est contenu dans un sous-espace vectoriel W de dimension finie sur $\mathbf{C}(x_1, \dots, x_r)$ stable par D_1, \dots, D_r (autrement dit, W est un sous- \mathcal{A}_r -module rationnel). Introduisons l'annulateur I_Φ de Φ ; c'est un idéal à gauche de \mathcal{A}_r auquel

* "D-finite" dans la terminologie de Zeilberger.

est associée une variété caractéristique $V(I_\Phi)$ dans \mathbf{C}^{2r} (cf n° 1.4). Alors Φ est rationnellement holonome si et seulement s'il existe un polynôme non nul $H(x_1, \dots, x_r)$ tel que $V(I_\Phi)$ soit contenue dans la réunion de l'hyper-surface $H = 0$ dans \mathbf{C}^{2r} et d'une variété de dimension r . D'après l'alinéa précédent, tout élément holonome est rationnellement holonome.

2. ÉTUDE DES SUITES ET DES FONCTIONS HOLONOMES

2.1. On appelle traditionnellement *série hypergéométrique* toute série de la forme $\sum_{n \geq 0} R(n)$ où $R(0) = 1$ et $R(n+1)/R(n)$ est une fonction rationnelle de n . En décomposant la fonction rationnelle en produit de facteurs linéaires, on écrit $R(n)$ sous la forme

$$(43) \quad R(n) = \frac{(a_1)_n \cdots (a_p)_n}{(b_1)_n \cdots (b_q)_n} \frac{z^n}{n!},$$

avec la définition suivante (classique !)

$$(44) \quad (a)_n = a(a+1) \cdots (a+n-1),$$

c'est-à-dire

$$(45) \quad (a)_0 = 1, \quad (a)_{n+1} = (a)_n(a+n).$$

La somme de la série $\sum_{n \geq 0} R(n)$ est alors ${}_pF_q \left(\begin{matrix} a_1 \cdots a_p \\ b_1 \cdots b_q \end{matrix} \middle| z \right)$ avec la définition classique de la *fonction hypergéométrique* [A 2] ; la fonction de Gauss est la fonction

$$(46) \quad {}_2F_1 \left(\begin{matrix} a & b \\ c \end{matrix} \middle| z \right) = \sum_{n \geq 0} \frac{(a)_n (b)_n}{(c)_n n!} z^n.$$

Sur l'espace des fonctions d'une variable entière n , on fait agir les opérateurs aux différences de la forme*

$$(47) \quad (Pu)(n) = \sum_{\alpha \geq 0} \sum_{\beta} c_{\alpha\beta} n^\alpha u(n+\beta).$$

* L'entier β est positif ou négatif.

En particulier, on introduit l'opérateur de translation N par

$$(48) \quad (Nu)(n) = u(n+1)$$

et l'opérateur précédent s'écrit $P(n, N) = \sum_{\alpha, \beta} c_{\alpha, \beta} n^\alpha N^\beta$. On a la relation de commutation

$$(49) \quad N.n = (n+1).N.$$

On dira qu'une suite est *holonome* s'il existe un opérateur $P(n, N)$ non nul tel que $P(n, N)u = 0$; c'est le cas de la suite hypergéométrique (43) (prendre

$$P(n, N) = (n+1)(n+b_1) \cdots (n+b_q)N - z(n+a_1) \cdots (n+a_p))$$

ou de la suite d'Apéry $u_n = \sum_{k=0}^n \binom{n}{k}^2 \binom{n+k}{k}$ (prendre

$$P(n, N) = n^3 - (34n^3 - 51n^2 + 27n - 5)N^{-1} + (n-1)^3 N^{-2}.$$

2.2. Tout ceci se généralise aux suites multiples, c'est-à-dire aux fonctions u à valeurs complexes définies dans \mathbf{Z}^s (ou dans une partie convenable de \mathbf{Z}^s). On introduit les opérateurs de translation K_1, \dots, K_s par

$$(50) \quad (K_i u)(k_1, \dots, k_s) = u(k_1, \dots, k_i + 1, \dots, k_s),$$

puis les opérateurs

$$(51) \quad P(\mathbf{k}, \mathbf{K}) = \sum_{\alpha \in \mathbf{Z}_+^s} \sum_{\beta \in \mathbf{Z}^s} c_{\alpha, \beta} k^\alpha K^\beta$$

(transformant la fonction $u(k_1, \dots, k_s)$ en la fonction

$$\sum_{\alpha} \sum_{\beta} c_{\alpha, \beta} k_1^{\alpha_1} \cdots k_s^{\alpha_s} u(k_1 + \beta_1, \dots, k_s + \beta_s)).$$

Les opérateurs de la forme (51) forment une algèbre \mathcal{B}_s qui est définie par les générateurs $k_1, \dots, k_s, K_1, \dots, K_s, K_1^{-1}, \dots, K_s^{-1}$ et les relations

$$(52) \quad k_i k_j = k_j k_i$$

$$(53) \quad K_i K_j = K_j K_i$$

$$(54) \quad k_i K_j = K_j k_i \quad (\text{pour } i \neq j)$$

$$(55) \quad K_i k_i = (k_i + 1) K_i$$

$$(56) \quad K_i K_i^{-1} = K_i^{-1} K_i = 1.$$

Cette algèbre n'est pas une inconnue : associons à la fonction $u : \mathbf{Z}^s \rightarrow \mathbf{C}$ la série de Laurent formelle

$$(57) \quad U(z_1, \dots, z_s) = \sum_{k_1, \dots, k_s} u(k_1, \dots, k_s) z_1^{k_1} \dots z_s^{k_s}.$$

A la fonction $k_i u$ correspond $z_i \frac{\partial U}{\partial z_i}$ et à $K_i u$ correspond $z_i^{-1} U$. On définit ainsi par

$$k_i \leftrightarrow z_i \frac{\partial}{\partial z_i}, \quad K_i \leftrightarrow z_i^{-1}$$

un isomorphisme entre \mathcal{B}_s et le localisé de \mathcal{A}_s obtenu en inversant z_1, \dots, z_s .

On est ainsi amené à définir le *symbole* de $P(\mathbf{k}, \mathbf{K})$ en gardant les termes de plus haut degré m par rapport à \mathbf{k} :

$$(58) \quad \sigma(P) = \sum_{|\alpha|=m} \sum_{\beta} c_{\alpha\beta} \zeta^\alpha z^{\alpha-\beta}$$

(polynôme en $z_1, \dots, z_s, z_1^{-1}, \dots, z_s^{-1}, \zeta_1, \dots, \zeta_s$) ; à une fonction $u(k_1, \dots, k_s)$ on associe ensuite la *variété caractéristique* V_u dans $(\mathbf{C}^\times)^s \times \mathbf{C}^s$ (coordonnées $z_1, \dots, z_s, \zeta_1, \dots, \zeta_s$) définie par les équations $\sigma(P)(z, \zeta) = 0$ où P parcourt l'ensemble des opérateurs tels que $P.u = 0$. Enfin, on dira que u est *holonome* si V_u est de dimension s .

Exemples : 1) On dit que u est *de type hypergéométrique* si $K_i u / u$ est une fonction rationnelle en k_1, \dots, k_s pour $i = 1, \dots, s$. Elle est (rationnellement) holonome.

2) On dit que u est *close* si c'est un produit de facteurs du type $(a)_{c_1 k_1 + \dots + c_s k_s}^{\pm 1}$ où c_1, \dots, c_s sont des entiers donnés, et d'une fonction rationnelle. C'est un cas particulier de suite de type hypergéométrique.

3) Considérons des fonctions $u(\mathbf{k})$ définies dans \mathbf{Z}^s , à valeurs vectorielles dans \mathbf{C}^N et satisfaisant aux équations

$$(59) \quad K_i u(\mathbf{k}) = A_i(\mathbf{k}) \cdot u(\mathbf{k}) \quad (1 \leq i \leq s).$$

Dans cette formule $A_1(\mathbf{k}), \dots, A_s(\mathbf{k})$ sont des matrices $N \times N$ dépendant rationnellement de \mathbf{k} , et l'on a les relations de compatibilité*

$$(60) \quad K_i A_j \cdot A_i = K_j A_i \cdot A_j.$$

Les composantes de $u(\mathbf{k})$ sont des suites rationnellement holonomes et le plus souvent holonomes.

2.3. On peut aussi considérer des *cas mixtes* de fonctions $u(x, \mathbf{k})$ sur $\mathbf{C}^r \times \mathbf{Z}^s$. On associe à cette fonction la série de Laurent formelle

$$(61) \quad U(x, z) = \sum_{k_1, \dots, k_s} u(x, \mathbf{k}) z_1^{k_1} \dots z_s^{k_s},$$

sur laquelle on fait agir les opérateurs différentiels à coefficients polynomiaux en $x_1, \dots, x_r, z_1, \dots, z_s, z_1^{-1}, \dots, z_s^{-1}$, d'où encore une théorie de fonctions holonomes.

Un exemple illustratif :

Posons $u(x, k) = P_k(x)$ (k -ième polynôme de Legendre). Ecrivant D pour $\frac{d}{dx}$ et K pour la translation en k , les relations classiques s'écrivent ainsi :

$$(62) \quad \begin{cases} Pu = 0 & \text{équation différentielle} \\ Qu = 0 & \text{récurrence} \end{cases}$$

* Ce système est l'analogie discret d'un système différentiel $D_i F = A_i F$, où les matrices A_i satisfont à la condition d'*intégrabilité complète*

$$D_i A_j - D_j A_i = [A_i, A_j].$$

avec

$$(63) \quad P = (1 - x^2)D^2 - 2xD + k(k + 1)$$

$$(64) \quad Q = (k + 2)K^2 - (2k + 3)xK + k + 1.$$

On a

$$(65) \quad \sigma(P) = (1 - x^2)\xi^2 + z^2\zeta^2$$

$$(66) \quad \sigma(Q) = \zeta(z + z^{-1} - 2x)$$

(recette : remplacer D par ξ , k par $z\zeta$, K par z^{-1} et garder les termes de plus haut degré en (ξ, ζ)). La variété caractéristique est définie par les équations $\sigma(P) = \sigma(Q) = \{\sigma(P), \sigma(Q)\} = 0$; elle a 4 composantes toutes de dimension 2 (dans l'espace $\mathbf{C}^3 \times \mathbf{C}^\times$ de coordonnées ξ, ζ, x, z)

$$V : 2x = z + z^{-1}, \quad 2\zeta = \xi(z^{-2} - 1)$$

$$V' : \xi = 0, \quad \zeta = 0$$

$$V_+'' : x = 1, \quad \zeta = 0$$

$$V_-'' : x = -1, \quad \zeta = 0.$$

On est dans le cas holonome. Il résulte d'un théorème général sur les systèmes holonomes que les solutions du système (62) forment un espace de dimension finie ν sur \mathbf{C} (ici $\nu = 4$ correspondant aux conditions initiales $u(0, 0), u(0, 1), Du(0, 0), Du(0, 1)$).

2.4. D'après la théorie générale des systèmes holonomes, les fonctions holonomes sur $\mathbf{C}^r \times \mathbf{Z}^s$ forment un espace vectoriel sur \mathbf{C} stable par les opérateurs $x_j, D_j, k_i, K_i^{\pm 1}$. Il est aussi facile de voir que si $u(x, \mathbf{k})$ et $v(y, \mathbf{m})$ sont holonomes, il en est de même du produit $u(x, \mathbf{k})v(y, \mathbf{m})$ à variables séparées (produit tensoriel). On peut aussi faire des changements de coordonnées linéaires sur les variables continues x_j , ou les variables discrètes k_i , ou encore des translations.

Soit $\Phi(x)$ une fonction holonome sur \mathbf{C}^r , dont la transformée de Laplace est donnée par

$$(67) \quad L(\xi) = \int_c e^{-x \cdot \xi} \Phi(x) d^r x;$$

l'intégrale porte sur une variété réelle convenable de dimension r dans \mathbf{C}^r , et l'on a posé $d^r x = dx_1 \wedge \cdots \wedge dx_r$, et $x \cdot \xi = x_1 \xi_1 + \cdots + x_r \xi_r$. On sait que si $P(x, D_x)$ est un opérateur à coefficients polynomiaux, la transformée de Laplace de $P(x, D_x)\Phi(x)$ est $Q(\xi, D_\xi)L(\xi)$ où l'opérateur $Q(\xi, D_\xi)$ s'obtient à partir de $P(x, D_x)$ en remplaçant x_j par $-\frac{\partial}{\partial \xi_j}$ et $\frac{\partial}{\partial x_j}$ par ξ_j . Si l'on se réfère au guide pratique (voir n° 1.6), on constate que la transformée de Laplace est une fonction holonome.*

Considérons maintenant la *spécialisation* $x_r = 0$ qui fait correspondre à une fonction $u(x_1, \dots, x_r)$ la fonction

$$(68) \quad S_r u(x_1, \dots, x_{r-1}) = u(x_1, \dots, x_{r-1}, 0).$$

Supposons u holonome et soit M le \mathcal{A}_r -module holonome formé des fonctions Pu où P parcourt \mathcal{A}_r . Identifiant \mathcal{A}_{r-1} à un sous-anneau de \mathcal{A}_r , il est clair que S_r est une application \mathcal{A}_{r-1} -linéaire et qu'on a $S_r(x_r v) = 0$, donc S_r définit une application \mathcal{A}_{r-1} -linéaire \bar{S}_r du \mathcal{A}_{r-1} -module $M/x_r M$ dans l'espace des fonctions de $r-1$ variables. Or le \mathcal{A}_{r-1} -module $M/x_r M$ est holonome (voir [C 6], page 193), et il est clair qu'un quotient d'un \mathcal{A}_{r-1} -module holonome est holonome : l'image de S_r est donc un \mathcal{A}_{r-1} -module holonome qui contient $S_r u$. Conclusion : la fonction $S_r u$ est holonome.

Combinant ce résultat avec des transformations linéaires, on voit par exemple que si $u(x, y, z)$ est holonome à 3 variables, la fonction $u(x, x, y)$ est holonome à 2 variables. Du résultat sur le produit tensoriel, on déduit, par spécialisation à la diagonale, que le produit ordinaire de deux fonctions holonomes à r variables est holonome.

* De manière plus algébrique, il existe un automorphisme σ de l'algèbre \mathcal{A}_r défini par $\sigma(x_j) = -D_j$, $\sigma(D_j) = x_j$ pour $1 \leq j \leq r$, et σ préserve la filtration de Bernstein dans \mathcal{A}_r .

Tous ces résultats s'étendent aux fonctions mixtes du type $u(x, \mathbf{k})$.

2.5. Disons quelques mots des q -analogues. On introduit une variable q , qu'on peut spécialiser en un nombre complexe $q \neq 0$. On opère sur des fonctions de s variables inversibles z_1, \dots, z_s au moyen des homothéties T_i :

$$(69) \quad T_i f(z_1, \dots, z_s) = f(z_1, \dots, qz_i, \dots, z_s).$$

On considère l'algèbre des opérateurs de la forme

$$(70) \quad P = \sum_{\alpha \in \mathbf{Z}^s} \sum_{\beta \in \mathbf{Z}^s} c_{\alpha\beta} z^\alpha T^\beta,$$

c'est-à-dire

$$(71) \quad Pf(z_1, \dots, z_s) = \sum_{\alpha, \beta} c_{\alpha\beta} z^\alpha f(q^{\beta_1} z_1, \dots, q^{\beta_s} z_s).$$

L'algèbre ${}_q\mathcal{C}_s$ de ces opérateurs est définie par les générateurs $z_1, \dots, z_s, T_1, \dots, T_s$ et leurs inverses soumis aux relations

$$(72) \quad z_i z_j = z_j z_i, T_i T_j = T_j T_i, z_i T_j = T_j z_i, \text{ pour } i \neq j$$

$$(73) \quad T_i z_i = q z_i T_i.$$

Sabbah [D 4] vient de donner les fondements de la théorie de cette algèbre, en imitant les résultats de Bernstein. On a ainsi une filtration par le degré total en les z_j et les T_j , on définit les bonnes filtrations sur les ${}_q\mathcal{C}_s$ -modules de type fini, on introduit une dimension δ dont on montre qu'elle satisfait à $\delta \geq s$ et l'on définit les modules holonomes sur ${}_q\mathcal{C}_s$ par la condition $\delta = s$.

On peut alors développer une théorie des *fonctions q -holonomes*. Par exemple, une fonction de type *q -hypergéométrique* est une fonction $F(z_1, \dots, z_s | q)$ telle que les quotients $T_i F / F$ soient des fonctions rationnelles de z_1, \dots, z_s et q (pour $1 \leq i \leq s$). Pour une variable, on a le produit infini

$$(74) \quad (z; q)_\infty = \prod_{i \geq 0} (1 - q^i z)$$

solution de l'équation

$$(75) \quad \frac{(qz; q)_\infty}{(z; q)_\infty} = \frac{1}{1-z}.$$

On pose

$$(76) \quad (z; q)_n = \frac{(z; q)_\infty}{(q^n z; q)_\infty} = \prod_{i=0}^{n-1} (1 - q^i z);$$

l'analogie des coefficients binomiaux

$$(77) \quad \left[\begin{matrix} n \\ k \end{matrix} \right]_q = \frac{(q; q)_n}{(q; q)_k (q; q)_{n-k}}$$

est un polynôme en q de degré $k(n-k)$ dont la valeur pour $q = 1$ est le coefficient binomial $\binom{n}{k}$. La formule d'addition devient

$$(78) \quad \left[\begin{matrix} n \\ k \end{matrix} \right]_q = \left[\begin{matrix} n-1 \\ k \end{matrix} \right]_q + q^{n-k} \left[\begin{matrix} n-1 \\ k-1 \end{matrix} \right]_q$$

et l'analogie de la fonction hypergéométrique de Gauss est la fonction "hypergéométrique basique" de Heine [A 20]

$$(79) \quad {}_2\Phi_1 \left(\begin{matrix} a & b \\ c \end{matrix} ; q ; z \right) = \sum_{n \geq 0} \frac{(a; q)_n (b; q)_n}{(c; q)_n (q; q)_n} z^n.$$

Si l'on substitue q^α à a , q^β à b et q^γ à c et qu'on fasse $q = 1$, on retrouve ${}_2F_1 \left(\begin{matrix} \alpha & \beta \\ \gamma \end{matrix} \middle| z \right)$.

3. INTÉGRATION ET SOMMATION DES FONCTIONS HOLONOMES

3.1. Posons $x = (x_1, \dots, x_r)$ et $x' = (x_1, \dots, x_{r-1})$, d'où $x = (x', x_r)$. A partir d'une fonction $\Phi(x)$, on définit par intégration une fonction

$$(80) \quad H(x') = \int \Phi(x', x_r) dx_r.$$

On peut supposer par exemple que x_r est une variable réelle et que $\Phi(x', x_r)$ pour x' fixé est nulle en dehors d'un intervalle compact. On peut aussi considérer une intégrale de Cauchy en la variable complexe x_r , ou des parties finies d'intégrales divergentes au sens de Hadamard. Dans tous ces exemples sont valables *les règles de dérivation sous le signe intégral et d'intégration par parties*.

Supposons que $\Phi(x)$ soit holonome, solution du système holonome

$$(81) \quad P_1(x, D)\Phi = \dots = P_N(x, D)\Phi = 0;$$

autrement dit, l'idéal à gauche I de \mathcal{A}_r engendré par P_1, \dots, P_N est holonome et le \mathcal{A}_r -module à gauche $M := \mathcal{A}_r/I$ est holonome. D'après [C 6], page 193, le \mathcal{A}_{r-1} -module $M/D_r M = \mathcal{A}_r/(I + D_r \mathcal{A}_r)$ est holonome. Si l'on pose

$$(82) \quad J = (I + D_r \mathcal{A}_r) \cap \mathcal{A}_{r-1}.$$

le \mathcal{A}_{r-1} -module \mathcal{A}_{r-1}/J est holonome, car isomorphe à un sous-module de $M/D_r M$, et l'idéal à gauche J de \mathcal{A}_{r-1} est holonome. Or J se compose des opérateurs différentiels de la forme

$$(83) \quad Q(x', D') = \sum_{j=1}^N A_j(x, D) P_j(x, D) + D_r B(x, D)$$

(avec $D' = (D_1, \dots, D_{r-1})$). D'après les équations (81), on a

$$(84) \quad Q(x', D')\Phi(x', x_r) = \partial \Phi_1(x', x_r) / \partial x_r$$

avec

$$(85) \quad \Phi_1(x) = B(x, D)\Phi(x).$$

On a alors

$$\begin{aligned} Q(x', D')H(x') &= Q(x', D') \int \Phi(x', x_r) dx_r \\ &= \int Q(x', D')\Phi(x', x_r) dx_r \\ &= \int \frac{\partial}{\partial x_r} \Phi_1(x', x_r) dx_r \\ &= 0 \end{aligned}$$

d'après les règles de calcul rappelées plus haut. Autrement dit, *l'intégrale* $H(x')$ *sera solution d'un système holonome*

$$(86) \quad Q_1(x', D')H = \dots = Q_M(x', D')H = 0,$$

où chacun des opérateurs $Q_j(x', D')$ est de la forme (83).

3.2. *L'analogie discret* est le suivant : on considère une solution d'un système holonome

$$(87) \quad A_1(\mathbf{k}, \mathbf{K})u(\mathbf{k}) = \dots = A_N(\mathbf{k}, \mathbf{K})u(\mathbf{k}) = 0$$

d'équations de récurrence. On pose

$$(88) \quad v(k_1, \dots, k_{r-1}) = \sum_{k_r} u(k_1, \dots, k_{r-1}, k_r),$$

où, par exemple, la sommation porte sur un nombre fini de valeurs non nulles. Alors $v(\mathbf{k}')$ satisfait à un système holonome

$$(89) \quad B_1(\mathbf{k}', \mathbf{K}')v(\mathbf{k}') = \dots = B_M(\mathbf{k}', \mathbf{K}')v(\mathbf{k}') = 0$$

où chaque opérateur $B_i(\mathbf{k}', \mathbf{K}')$ est de la forme

$$(90) \quad \sum_{j=1}^N C_j(\mathbf{k}, \mathbf{K}) \cdot A_j(\mathbf{k}, \mathbf{K}) + (L_r - 1)D(\mathbf{k}, \mathbf{K}).$$

Le problème est de faire les calculs explicitement.

Par exemple, la fonction $F(n, k) = \binom{n}{k}$ est solution du système holonome $PF = P'F = 0$ avec

$$(91) \quad P = (n - k + 1)N - (n + 1) , \quad P' = (k + 1)K - (n - k)$$

(on a posé $NF(n, k) = F(n + 1, k)$ et $KF(n, k) = F(n, k + 1)$).

Par un calcul facile, on trouve

$$(n + 1)(N - 2) = (K + 1)P + NP' - (K - 1)(Nn - n - 1).$$

Autrement dit, la fonction $v(n) = \sum_k \binom{n}{k}$ satisfait à l'équation de récurrence

$$(92) \quad (n + 1)(v(n + 1) - 2v(n)) = 0 .$$

Tenant compte de la condition initiale $v(0) = 1$, ceci entraîne la réponse souhaitée $v(n) = 2^n$ pour $n \geq 0$. La méthode précédente n'est pas la plus simple pour prouver l'identité $\sum_k \binom{n}{k} = 2^n$, mais elle a l'avantage de conduire à des algorithmes programmables.

3.3. Nous décrirons d'abord l'algorithme "lent" de Zeilberger.

a) Utilisant les relations de commutation

$$kn = nk , \quad KN = NK , \quad kN = Nk , \quad nK = Kn$$

et les relations

$$(93) \quad Nn = (n + 1)N , \quad Kk = (k + 1)K ,$$

on montre que tout opérateur $A(n, k, N, K)$ s'écrit de manière unique sous la forme

$$(94) \quad A(n, k, N, K) = \sum_{i=1}^b \sum_{j=0}^c a_{ij}(N, K) k^i n^j + R(n, N, K).$$

b) On considère deux opérateurs P et P' mis sous la forme normale précédente

$$(95) \quad P = \sum_{i=1}^b \sum_{j=0}^c a_{ij}(N, K) k^i n^j + R(n, N, K),$$

$$(96) \quad P' = \sum_{i=1}^{b'} \sum_{j=0}^c a'_{ij}(N, K) k^i n^j + R'(n, N, K).$$

c) On considère les opérateurs

$$k^{\alpha'} n^{\beta} P \quad \text{pour } 0 \leq \alpha' < b', \quad 0 \leq \beta \leq c(b + b' - 1)$$

$$k^{\alpha} n^{\beta} P' \quad \text{pour } 0 \leq \alpha < b, \quad 0 \leq \beta \leq c(b + b' - 1)$$

au nombre de $\mu = (b + b')[c(b + b' - 1) + 1]$. Ecrivons chacun de ces μ opérateurs sous la forme normale (94) : chacun d'eux est somme d'un opérateur dépendant uniquement de n, N et K et d'une combinaison linéaire de monômes $k^i n^j$ à coefficients dans l'anneau commutatif des polynômes en N, K . Les monômes $k^i n^j$ qui interviennent satisfont aux inégalités

$$1 \leq i \leq b + b' - 1, \quad 0 \leq j \leq c(b + b')$$

et leur nombre est $(b + b' - 1)[c(b + b') + 1]$, ce qui est égal à $\mu - 1$. Comme on a μ combinaisons linéaires à coefficients dans l'anneau commutatif $\mathbf{C}[N, K]$ de $\mu - 1$ monômes $k^i n^j$, il existe par l'algèbre linéaire usuelle une combinaison linéaire non nulle du type

$$\sum_{\alpha'=0}^{b'-1} \sum_{\beta=0}^{c(b+b'-1)} A_{\alpha',\beta}(N, K) k^{\alpha'} n^{\beta} P + \sum_{\alpha=0}^{b-1} \sum_{\beta=0}^{c(b+b'-1)} A'_{\alpha,\beta} k^{\alpha} n^{\beta} P'$$

qui ne fasse plus intervenir les monômes $k^i n^j$, donc soit un opérateur dépendant de n, N et k exclusivement.

d) On a donc trouvé deux opérateurs A et A' tels que $R = AP + A'P'$ ne dépende plus de k . Que cet opérateur ne soit pas nul se vérifie au moyen

de l'hypothèse que le système $PF = P'F = 0$ est holonome, c'est-à-dire que les symboles $\sigma(P)$ et $\sigma(P')$ sont "indépendants" en un sens convenable.

e) On écrit R sous la forme $(K - 1)^\lambda R'(K, N, n)$. Comme on a $PF = P'F = 0$, on a

$$(97) \quad (K - 1)^\lambda R'(K, N, n)F = 0 .$$

Cette relation signifie que la fonction $F' = R'(K, N, n)F$ est un polynôme en k . Supposons que, pour n fixé, on ait $F(n, k) = 0$ pour $|k|$ assez grand. Alors $F'(n, k)$ a la même propriété, et comme c'est un polynôme en k , on a $F'(n, k) = 0$.

f) Ecrivons l'opérateur R' sous la forme

$$(98) \quad R'(K, N, n) = \sum_{j=0}^d K^j R_j(N, n)$$

et posons

$$(99) \quad S(N, n) = \sum_{j=0}^d R_j(N, n);$$

il en découle une relation

$$(100) \quad R'(K, N, n) = S(N, n) + (K - 1)U(K, N, n).$$

En conclusion, de l'équation $R'F = 0$, il résulte que la fonction $a(n) = \sum_k F(n, k)$ satisfait à l'équation de récurrence

$$(101) \quad S(N, n)a = 0.$$

3.4. L'algorithme précédent est très exigeant en temps de calcul et surtout en mémoire. Un algorithme beaucoup plus performant utilise le procédé de sommation de Gosper [C 8] que nous décrivons maintenant.

Soit $(S(k))_{k \geq 0}$ une suite de type hypergéométrique, de sorte que l'on ait

$$(102) \quad \frac{S(k)}{S(k-1)} = s(k) \quad \text{pour } k \geq 1.$$

où $s(k)$ est fonction rationnelle de k . Si l'on pose

$$(103) \quad A(k) = S(k) - S(k-1),$$

la suite $(A(k))_{k \geq 0}$ est de type hypergéométrique : en effet, le quotient $a(k) = A(k)/A(k-1)$ est donné par

$$(104) \quad a(k) = s(k-1) \frac{s(k) - 1}{s(k-1) - 1}.$$

De même, le quotient $S(k)/A(k)$ est égal à la fonction rationnelle $s(k)/(s(k) - 1)$.

Le problème de la *sommation hypergéométrique indéfinie* est le suivant : étant donnée la suite de type hypergéométrique $(A(k))_{k \geq 0}$, décider si la suite $(S(k))_{k \geq 0}$ donnée par

$$(105) \quad S(k) = A(0) + \cdots + A(k)$$

est aussi de type hypergéométrique. Comme on a évidemment la relation (104), le problème devient le suivant :

Étant donnée la fonction rationnelle $a(k)$, décider si l'équation (104) admet une solution qui soit une fonction rationnelle $s(k)$, et la calculer si possible.

On commence par écrire la fonction rationnelle $a(k) = A(k)/A(k-1)$ sous la forme

$$(106) \quad a(k) = \frac{p(k)}{p(k-1)} \frac{q(k)}{r(k)},$$

où les polynômes $p(k), q(k), r(k)$ sont tels que $q(k)$ n'ait de diviseur commun non constant avec aucun des polynômes $r(k), r(k+1), r(k+2), \dots$ [si $q(k)$ et $r(k+j)$ avaient un facteur commun non constant $u(k)$, faire la substitution

$$\begin{aligned} p(k) &\leftarrow p(k)u(k)u(k-1) \cdots u(k-j+1) \\ q(k) &\leftarrow q(k)/u(k) \\ r(k) &\leftarrow r(k)/u(k-j), \end{aligned}$$

puis recommencer si nécessaire].

Introduisons maintenant une nouvelle suite $(\varphi(k))_{k \geq 0}$ par

$$(107) \quad S(k) = \frac{q(k+1)}{p(k)} \varphi(k) A(k) ;$$

la relation $A(k) = S(k) - S(k-1)$ se traduit alors par l'équation

$$(108) \quad q(k+1)\varphi(k) - r(k)\varphi(k-1) = p(k).$$

Si la suite $(S(k))$ est de type hypergéométrique, on a vu que $S(k)/A(k)$, donc aussi $\varphi(k)$, est fonction rationnelle de k . Le miracle est le suivant : *si une fonction rationnelle $\varphi(k)$ est solution de l'équation (108), c'est un polynôme dont on peut majorer explicitement le degré par un entier J calculable à partir de $p(k)$, $q(k)$ et $r(k)$. Si l'on a $J < 0$, il n'y a pas de solution et la suite $(S(k))$ n'est pas de type hypergéométrique ; sinon on pose*

$$(109) \quad \varphi(k) = f_0 + f_1 k + \dots + f_J k^J$$

et l'équation (108) se traduit en un système d'équations linéaires pour les coefficients f_0, \dots, f_J .

3.5. Posons le problème de la *sommation hypergéométrique définie*. On dispose d'une fonction $F(n, k)$ de type hypergéométrique, satisfaisant aux deux relations

$$(110) \quad \frac{F(n+1, k)}{F(n, k)} = A(n, k), \quad \frac{F(n, k+1)}{F(n, k)} = B(n, k),$$

où $A(n, k)$ et $B(n, k)$ sont deux fonctions rationnelles. On cherche à déterminer des polynômes $s_0(n), \dots, s_I(n)$ et une fonction rationnelle $R(n, k)$ satisfaisant à la relation

$$(111) \quad s(n, N)F(n, k) = G(n, k) - G(n, k-1),$$

où l'on a posé

$$(112) \quad s(n, N) = \sum_{i=0}^I s_i(n) N^i$$

$$(113) \quad G(n, k) = R(n, k)F(n, k) .$$

Compte tenu des relations (110) et après division par $F(n, k)$, la relation (111) devient une identité entre fonctions rationnelles

$$(114) \quad \sum_{i=0}^I s_i(n) \cdot \prod_{j=0}^{i-1} A(n+j, k) = R(n, k) - R(n, k-1)/B(n, k-1).$$

En général, la vérification de cette relation, une fois déterminés explicitement $s_0(n), \dots, s_I(n)$ et $R(n, k)$, est tout à fait élémentaire ; elle entraîne (sous réserve par exemple que $F(n, k)$ soit nul pour $|k|$ assez grand lorsque n est fixé) que la somme $a(n) = \sum_k F(n, k)$ vérifie l'équation de récurrence

$$(115) \quad \sum_{i=0}^I s_i(n)a(n+i) = 0.$$

Zeilberger dit que $R(n, k)$ "certifie" la validité de l'équation (115).

Pour le calcul de $s_0(n), \dots, s_I(n), R(n, k)$, on procède comme dans l'algorithme de Gosper, en traitant n comme un paramètre, c'est-à-dire en remplaçant le corps de base \mathbf{C} par le corps des fonctions rationnelles $\mathbf{C}(n)$. Donnons simplement les formules essentielles. Comme dans la méthode de Gosper, on introduit une factorisation

$$(116) \quad \frac{F(n+i, k)}{F(n+i, k-1)} = \frac{p_i(n, k)}{p_i(n, k-1)} \frac{q(n, k)}{r(n, k)}$$

avec des polynômes $p_i(n, k), q(n, k)$ et $r(n, k)$ tels que $q(n, k)$ n'ait de facteur commun avec aucun des polynômes $r(n, k+j)$ pour $j = 0, 1, 2, \dots$ [on détermine d'abord $p_0(n, k), q(n, k)$ et $r(n, k)$, ce qui est toujours possible si le numérateur et le dénominateur de $B(n, k)$ sont produits de facteurs linéaires ; on utilise ensuite le fait que $F(n+i, k)/F(n, k)$ est une fonction rationnelle connue d'après (110)]. On détermine ensuite les fonctions rationnelles $s_0(n), \dots, s_I(n)$ et $f_0(n), \dots, f_J(n)$ satisfaisant au système d'équations linéaires (identifier les coefficients des monômes en k)

$$(117) \quad \sum_{j=0}^J \{q(n, k+1)k^j - r(n, k)(k-1)^j\} \cdot f_j(n) = \sum_{i=0}^I p_i(n, k) \cdot s_i(n).$$

On conclut en posant

$$(118) \quad G(n, k) = q(n, k + 1) \sum_{j=0}^J f_j(n) k^j \cdot \frac{\sum_{i=0}^I s_i(n) F(n + i, k)}{\sum_{i=0}^I s_i(n) p_i(n, k)}.$$

Comme $F(n + i, k)/F(n, k)$ est une fonction rationnelle de n et k , il en est bien de même de $G(n, k)/F(n, k)$.

Exemple : $F(n, k) = 1/k!(n - k)!$.

On prend $I = 1$, et l'algorithme de Gosper garantit que $J = 0$ convient. On a alors :

$$(119) \quad \begin{cases} p_0(n, k) = n - k + 1 \\ p_1(n, k) = 1 \\ q(n, k) = n - k + 2 \\ r(n, k) = k . \end{cases}$$

Les inconnues sont s_0, s_1, f_0 et le système (117) s'écrit

$$(120) \quad (n - 2k + 1)f_0 = (n - k + 1)s_0 + s_1 :$$

en séparant les diverses puissances de k , ceci donne le système

$$(121) \quad \begin{cases} (n + 1)f_0 = (n + 1)s_0 + s_1 \\ 2f_0 = s_0 . \end{cases}$$

Une solution est

$$f_0 = 1 , s_0 = 2 , s_1 = -(n + 1),$$

d'où

$$G(n, k) = (n - k + 1) \frac{2F(n, k) - (n + 1)F(n + 1, k)}{2(n - k + 1) - (n + 1)};$$

comme on a

$$(122) \quad \frac{F(n + 1, k)}{F(n, k)} = \frac{1}{n - k + 1} ,$$

on conclut par un calcul de fonctions rationnelles

$$(123) \quad \frac{G(n, k)}{F(n, k)} = 1 .$$

L'équation (111) prend dans ce cas la forme

$$(124) \quad 2F(n, k) - (n + 1)F(n + 1, k) = F(n, k) - F(n, k - 1).$$

Par sommation sur k , on voit que $a(n) = \sum_k F(n, k)$ satisfait à l'équation de récurrence

$$(125) \quad 2a(n) - (n + 1)a(n + 1) = 0 :$$

ceci joint à la condition initiale $a(0) = 1$ donne la solution $a(n) = \frac{2^n}{n!}$, d'où la formule sommatoire

$$(126) \quad \sum_{k=0}^n \frac{1}{k!(n-k)!} = \frac{2^n}{n!}.$$

On a retrouvé *notre exemple favori* $\sum_{k=0}^n \binom{n}{k} = 2^n$.

3.6. Zeilberger et Wilf sont en train de développer des algorithmes semblables pour traiter les q -analogues des fonctions de type hypergéométrique. Avec Almqvist, Zeilberger a aussi décrit des algorithmes du même genre pour déterminer les équations différentielles satisfaites par des intégrales du type $u(x) = \int v(x, y)dy$ où v est de *type hyperexponentiel*.* Enfin, en s'appuyant sur les résultats préliminaires de Galligo [C 7], Takayama [C 14] a développé des techniques d'élimination dans les idéaux d'opérateurs différentiels, imitant les bases de Gröbner [C 5] pour les idéaux de polynômes. Cela lui fournit une alternative aux méthodes de Zeilberger.

* Ceci signifie que $D_x v/v$ et $D_y v/v$ sont des fonctions rationnelles.

4. SÉRIES ET FONCTIONS HYPERGÉOMÉTRIQUES *

4.1. On a déjà rappelé la définition de la série hypergéométrique de Gauss $F = {}_2F_1$:

$$(127) \quad F \left(\begin{matrix} a & b \\ c \end{matrix} \middle| z \right) = \sum_{k \geq 0} \frac{(a)_k (b)_k}{(c)_k} \frac{z^k}{k!};$$

elle a un sens pourvu que c ne soit pas un entier négatif, et la série converge pour $|z| < 1$. Elle satisfait à l'équation différentielle

$$(128) \quad z(1-z) \frac{d^2 F}{dz^2} + \{c - (a+b+1)z\} \frac{dF}{dz} - abF = 0.$$

Dans certains cas particuliers, on retrouve des fonctions élémentaires, par exemple

$$(129) \quad F \left(\begin{matrix} a & b \\ b \end{matrix} \middle| z \right) = (1-z)^{-a}$$

$$(130) \quad F \left(\begin{matrix} a & a+1/2 \\ 1/2 \end{matrix} \middle| z \right) = \frac{1}{2}(1+z^{1/2})^{-2a} + \frac{1}{2}(1-z^{1/2})^{-2a}.$$

Les fonctions correspondant à divers paramètres a, b, c ne sont pas indépendantes ; voici par exemple la relation d'Euler

$$(131) \quad F \left(\begin{matrix} c-a & c-b \\ c \end{matrix} \middle| z \right) = (1-z)^{a+b-c} F \left(\begin{matrix} a & b \\ c \end{matrix} \middle| z \right).$$

Dans certains cas, on peut sommer exactement la série (127) :

$$(132) \quad F \left(\begin{matrix} a & b \\ c \end{matrix} \middle| 1 \right) = \frac{\Gamma(c)\Gamma(c-a-b)}{\Gamma(c-a)\Gamma(c-b)} \quad \text{GAUSS}$$

* Voir [A 16] et [A 2] pour les propriétés de base des fonctions hypergéométriques, et [A 11] (chapitre 5) pour un exposé très détaillé des sommes liées aux coefficients binomiaux.

$$(133) \quad F \left(\begin{matrix} a & b \\ 1+a-b \end{matrix} \middle| -1 \right) = 2^{-a} \frac{\Gamma(1+a-b)\Gamma(\frac{1}{2})}{\Gamma(1-b+\frac{a}{2})\Gamma(\frac{a+1}{2})} \quad \text{KUMMER}$$

$$(134) \quad F \left(\begin{matrix} a & 1-a \\ b \end{matrix} \middle| \frac{1}{2} \right) = 2^{1-b} \frac{\Gamma(\frac{1}{2})\Gamma(b)}{\Gamma(\frac{a+b}{2})\Gamma(\frac{b-a+1}{2})}$$

$$(135) \quad F \left(\begin{matrix} 2a & 2b \\ a+b+\frac{1}{2} \end{matrix} \middle| \frac{1}{2} \right) = \frac{\Gamma(a+b+\frac{1}{2})\Gamma(\frac{1}{2})}{\Gamma(a+\frac{1}{2})\Gamma(b+\frac{1}{2})}.$$

Il y a deux démonstrations classiques de la formule (132) de Gauss. Tout d'abord, on a la représentation intégrale d'Euler

$$(136) \quad F \left(\begin{matrix} a & b \\ c \end{matrix} \middle| z \right) = \frac{\Gamma(c)}{\Gamma(b)\Gamma(c-b)} \int_0^1 t^{b-1} (1-t)^{c-b-1} (1-tz)^{-a} dt$$

qui fournit le prolongement analytique de la fonction hypergéométrique dans le plan coupé le long de l'axe réel de 1 à $+\infty$. Pour $z = 1$, l'intégrale se réduit à l'intégrale bêta d'Euler

$$\int_0^1 t^{b-1} (1-t)^{c-b-a-1} dt,$$

donc à $\Gamma(b)\Gamma(c-b-a)/\Gamma(c-a)$. La même méthode permet d'établir les relations (133) à (135).

La deuxième démonstration utilise une équation de récurrence

$$(137) \quad (c-a)(c-b)zF \left(\begin{matrix} a & b \\ c+1 \end{matrix} \middle| z \right) = \\ = c[(2c-a-b-1)z-c+1]F \left(\begin{matrix} a & b \\ c \end{matrix} \middle| z \right) + c(c-1)(1-z)F \left(\begin{matrix} a & b \\ c-1 \end{matrix} \middle| z \right)$$

(vérification directe par comparaison des coefficients de z^k). En prenant la limite pour z tendant vers 1, le dernier terme disparaît, d'où

$$(138) \quad (c-a)(c-b)F \left(\begin{matrix} a & b \\ c+1 \end{matrix} \middle| 1 \right) = c(c-a-b)F \left(\begin{matrix} a & b \\ c \end{matrix} \middle| 1 \right).$$

Par récurrence sur l'entier $m \geq 0$, on en déduit

$$(139) \quad F \left(\begin{matrix} a & b \\ c & \end{matrix} \middle| 1 \right) = \frac{(c-a)_m (c-b)_m}{(c)_m (c-a-b)_m} F \left(\begin{matrix} a & b \\ c+m & \end{matrix} \middle| 1 \right).$$

On passe ensuite à la limite sur m ; il est facile de voir que $F \left(\begin{matrix} a & b \\ c+m & \end{matrix} \middle| 1 \right)$ tend vers 1 lorsque m tend vers l'infini, et le développement de la fonction gamma en produit infini donne la relation

$$(140) \quad \lim_{m \rightarrow \infty} \frac{(u_1)_m \cdots (u_p)_m}{(v_1)_m \cdots (v_p)_m} = \frac{\Gamma(v_1) \cdots \Gamma(v_p)}{\Gamma(u_1) \cdots \Gamma(u_p)}$$

lorsque $u_1 + \cdots + u_p$ est égal à $v_1 + \cdots + v_p$.

4.2. La formule de Gauss est particulièrement intéressante lorsque b est un entier négatif ; compte tenu de la relation de récurrence

$$(141) \quad \Gamma(s+n) = \Gamma(s)(s)_n ,$$

la formule de Gauss se spécialise en

$$(142) \quad F \left(\begin{matrix} a & -n \\ c & \end{matrix} \middle| 1 \right) = \frac{(c-a)_n}{(c)_n}.$$

Mais la série hypergéométrique se réduit à une somme finie, et après quelques manipulations, on obtient la relation

$$(143) \quad \sum_{k=0}^n (-1)^k \binom{n}{k} \frac{(a)_k}{(c)_k} = \frac{(c-a)_n}{(c)_n} .$$

Il reste à faire la substitution $a \leftarrow -u, c \leftarrow v-n+1$ pour obtenir la *formule de Chu-Vandermonde*

$$(144) \quad \binom{u+v}{n} = \sum_{k=0}^n \binom{u}{k} \binom{v}{n-k} .$$

Voici d'autres spécialisations de la formule (142)

$$(145) \quad \sum_{k=0}^n \binom{r+k}{k} = \binom{r+n+1}{n} \quad (a \leftarrow 1, c \leftarrow -n-r)$$

$$(146) \quad \sum_{k=0}^n (-1)^k \binom{r}{k} = (-1)^n \binom{r-1}{n} \quad (a \leftarrow 1, c \leftarrow -n + r + 1).$$

A son tour, la formule de Chu-Vandermonde contient de nombreuses spécialisations, par exemple*

$$(147) \quad \sum_k \binom{u}{m+k} \binom{v}{n-k} = \binom{u+v}{m+n}.$$

De manière analogue, la formule (133) de Kummer donne par spécialisation

$$(148) \quad F \left(\begin{matrix} 1-r-2n & -2n \\ r & \end{matrix} \middle| -1 \right) = (-1)^n \frac{(2n)!}{n!} \frac{(r-1)!}{(r+n-1)!},$$

et après quelques manipulations simples, on obtient la généralisation suivante de la deuxième formule (2)

$$(149) \quad \sum_{k=0}^{2n} (-1)^k \binom{r}{k} \binom{r}{2n-k} = (-1)^n \binom{r}{n}.$$

On pourrait continuer ce jeu : **pratiquement n'importe quelle formule donnant la somme de produits de deux coefficients binomiaux s'obtient par spécialisation des formules (132) à (135).**

4.3. Pour aller plus loin, il faut introduire les séries hypergéométriques générales (voir le n° 2.1)

$$(150) \quad {}_pF_q \left(\begin{matrix} a_1 \cdots a_p \\ b_1 \cdots b_q \end{matrix} \middle| z \right) = \sum_{k \geq 0} \frac{(a_1)_k \cdots (a_p)_k}{(b_1)_k \cdots (b_q)_k} \frac{z^k}{k!}.$$

L'équation différentielle (128) se généralise ainsi (on pose $D = \frac{d}{dz}$ et $\vartheta = z \frac{d}{dz}$)

$$(151) \quad D(\vartheta + b_1 - 1) \cdots (\vartheta + b_q - 1)F = (\vartheta + a_1) \cdots (\vartheta + a_p)F.$$

* Voir [A 11], page 169 pour une liste plus étendue.

Cela résulte immédiatement des trois formules de dérivation :

$$(152) \quad DF \left(\begin{matrix} a_1 \cdots a_p \\ b_1 \cdots b_q \end{matrix} \middle| z \right) = \frac{a_1 \cdots a_p}{b_1 \cdots b_q} F \left(\begin{matrix} a_1 + 1 \cdots a_p + 1 \\ b_1 + 1 \cdots b_q + 1 \end{matrix} \middle| z \right)$$

$$(153) \quad (\vartheta + a_1) F \left(\begin{matrix} a_1 \cdots a_p \\ b_1 \cdots b_q \end{matrix} \middle| z \right) = a_1 F \left(\begin{matrix} a_1 + 1 & a_2 \cdots a_p \\ b_1 & b_2 \cdots b_q \end{matrix} \middle| z \right).$$

$$(154) \quad (\vartheta + b_1 - 1) F \left(\begin{matrix} a_1 \cdots a_p \\ b_1 \cdots b_q \end{matrix} \middle| z \right) = (b_1 - 1) F \left(\begin{matrix} a_1 & a_2 \cdots a_p \\ b_1 - 1 & b_2 \cdots b_q \end{matrix} \middle| z \right).$$

Dans la formule d'Euler (131), comparons les coefficients de z^n dans les deux membres. On obtient une relation due à PFAFF et SAALSCHÜTZ et qui s'écrit

$$(155) \quad {}_3F_2 \left(\begin{matrix} a & b & -n \\ c & a + b - c - n + 1 \end{matrix} \middle| 1 \right) = \frac{(c-a)_n (c-b)_n}{(c)_n (c-a-b)_n}.$$

Il est facile de retrouver le théorème de Gauss sur ${}_2F_1 \left(\begin{matrix} a & b \\ c \end{matrix} \middle| 1 \right)$ en prenant la limite pour $n \rightarrow \infty$ dans la formule de Pfaff-Saalschütz.

Nous recopions les relations les plus remarquables obtenues par sommation de séries hypergéométriques générales :

DIXON :

$${}_3F_2 \left(\begin{matrix} a & b & c \\ a - b + 1 & a - c + 1 \end{matrix} \middle| 1 \right) = \frac{\Gamma(\frac{a}{2} + 1) \Gamma(a - b + 1) \Gamma(a - c + 1) \Gamma(\frac{a}{2} - b - c + 1)}{\Gamma(a + 1) \Gamma(\frac{a}{2} - b + 1) \Gamma(\frac{a}{2} - c + 1) \Gamma(a - b - c + 1)}$$

WATSON :

$${}_3F_2 \left(\begin{matrix} a & b & c \\ \frac{1}{2}(a + b + 1) & 2c \end{matrix} \middle| 1 \right) = \frac{\Gamma(\frac{1}{2}) \Gamma(c + \frac{1}{2}) \Gamma(\frac{a+b+1}{2}) \Gamma(\frac{1}{2} - \frac{a}{2} - \frac{b}{2} + c)}{\Gamma(\frac{a+1}{2}) \Gamma(\frac{b+1}{2}) \Gamma(c - \frac{a}{2} + \frac{1}{2}) \Gamma(c - \frac{b}{2} + \frac{1}{2})}$$

WHIPPLE :

$${}_3F_2 \left(\begin{matrix} a & b & c \\ e & f \end{matrix} \middle| 1 \right) = \frac{\pi \Gamma(e) \Gamma(f)}{2^{2c-1} \Gamma(\frac{a+e}{2}) \Gamma(\frac{a+f}{2}) \Gamma(\frac{b+e}{2}) \Gamma(\frac{b+f}{2})}$$

(sous les hypothèses $a + b = 1$ et $e + f = 2c + 1$).

L'identité la plus impressionnante est celle de DOUGALL :

$${}_7F_6 \left(\begin{matrix} a & \frac{a}{2} + 1 & b & c & d & e & -m \\ & \frac{a}{2} & a - b + 1 & a - c + 1 & a - d + 1 & a - e + 1 & a + m + 1 \end{matrix} \middle| 1 \right) \\ = \frac{(a+1)_m (a-b-c+1)_m (a-b-d+1)_m (a-c-d+1)_m}{(a-b+1)_m (a-c+1)_m (a-d+1)_m (a-b-c-d+1)_m},$$

sous l'hypothèse que m est un entier positif et que l'on a $2a + 1 = b + c + d + e - m$. Les experts savent déduire de ces formules, par spécialisation, un nombre illimité d'identités entre coefficients binomiaux.

4.4. Toutes ces relations ont des q -analogues.* On a déjà introduit le produit infini $(a; q)_\infty = \prod_{m \geq 0} (1 - aq^m)$, qui est un cousin de la fonction gamma, et le produit partiel

$$(156) \quad (a; q)_n = \frac{(a; q)_\infty}{(aq^n; q)_\infty} = \prod_{m=0}^{n-1} (1 - aq^m).$$

Le q -analogue de la factorielle est

$$(157) \quad [n]_q! = \frac{(q; q)_n}{(1 - q)^n};$$

c'est un polynôme en q de degré $n(n-1)/2$, dont la valeur pour $q = 1$ est égale à $n!$.

Le q -analogue de la fonction hypergéométrique de Gauss est donné par

$$(158) \quad {}_2\Phi_1 \left(\begin{matrix} a & b \\ c \end{matrix} ; q; z \right) = \sum_{k \geq 0} \frac{(a; q)_k (b; q)_k}{(c; q)_k (q; q)_k} z^k,$$

avec la généralisation évidente

$$(159) \quad {}_r\Phi_s \left(\begin{matrix} a_1 \cdots a_r \\ b_1 \cdots b_s \end{matrix} ; q; z \right) = \sum_{k \geq 0} \frac{(a_1; q)_k \cdots (a_r; q)_k}{(b_1; q)_k \cdots (b_s; q)_k} \frac{z^k}{(q; q)_k}.$$

* Voir les livres d'Exton [A 17] et de Gasper et Rahman [A 19].

De même que la série du binôme est un cas particulier de fonction hypergéométrique

$$(160) \quad (1-z)^{-a} = {}_1F_0 \left(\begin{matrix} a \\ - \end{matrix} \middle| z \right)$$

et que l'exponentielle est donnée par

$$(161) \quad e^z = {}_0F_0 \left(\begin{matrix} - \\ - \end{matrix} \middle| z \right),$$

on introduira une série q -binomiale

$$(162) \quad {}_1\Phi_0 \left(\begin{matrix} a \\ - \end{matrix} ; q; z \right) = \sum_{k \geq 0} \frac{(a; q)_k}{(q; q)_k} z^k = \frac{(az; q)_\infty}{(z; q)_\infty}$$

et une série q -exponentielle

$$(163) \quad {}_0\Phi_0 \left(\begin{matrix} - \\ - \end{matrix} ; q; z \right) = \sum_{k \geq 0} \frac{z^k}{(q; q)_k} = \frac{1}{(z; q)_\infty}$$

(pour retrouver le cas classique, remplacer z par $(1-q)x$ et faire tendre q vers 1).

Dans la série ${}_r\Phi_s$, si l'on remplace z par 1 et que l'un des paramètres a_i est de la forme q^{-m} (avec $m \geq 0$ entier), on obtient une somme de coefficients q -binomiaux (voir le n° 2.5). Les formules de sommation hypergéométriques (Gauss, Kummer,...) ont des q -analogues dont voici un échantillon :

GAUSS :

$${}_2\Phi_1 \left(\begin{matrix} a & b \\ c \end{matrix} ; q; c/ab \right) = {}_2\Pi_2 \left(\begin{matrix} c/a & c/b \\ c & c/ab \end{matrix} \middle| q \right)$$

KUMMER :

$${}_2\Phi_1 \left(\begin{matrix} a & b \\ aq/b \end{matrix} ; q; -q/b \right) = {}_1\Pi_2 \left(\begin{matrix} -q \\ -q/b & aq/b \end{matrix} \middle| q \right) {}_2\Pi_0 \left(\begin{matrix} aq & aq^2/b^2 \\ - & - \end{matrix} \middle| q^2 \right)$$

PFAFF-SAALSCHÜTZ :

$${}_3\Phi_2 \left(\begin{matrix} a & b & q^{-n} \\ c & abc^{-1}q^{1-n} \end{matrix} ; q; q \right) = \frac{(c/a; q)_n (c/b; q)_n}{(c; q)_n (c/ab; q)_n}$$

DIXON :

$${}_3\Phi_2 \left(\begin{matrix} a^2 & b & c \\ a^2q/b & a^2q/c \end{matrix} ; q; q^2a/bc \right) = {}_4\Pi_4 \left(\begin{matrix} b/a^2 & c/a^2 & q/a & bc/a \\ b/a & c/a & q/a^2 & bc/a^2 \end{matrix} \middle| q \right).$$

On a adopté la convention suivante

$$(164) \quad {}_r\Pi_s \left(\begin{matrix} a_1 \cdots a_r \\ b_1 \cdots b_s \end{matrix} \middle| q \right) = \frac{(a_1; q)_\infty \cdots (a_r; q)_\infty}{(b_1; q)_\infty \cdots (b_s; q)_\infty}.$$

On peut spécialiser ces formules d'innombrables manières ; citons deux cas particuliers célèbres

JACOBI :

$$\sum_{n=-\infty}^{+\infty} q^{n^2} z^n = \prod_{m=1}^{\infty} \frac{1}{(1 + q^{2m-1}z)(1 + q^{2m-1}z^{-1})(1 - q^{2m})}$$

ROGERS-RAMANUJAN :

$$\sum_{k=0}^{\infty} \frac{q^{k^2}}{(1-q)(1-q^2) \cdots (1-q^k)} = \prod_{n=0}^{\infty} \frac{1}{(1-q^{5n+1})(1-q^{5n+4})}.$$

4.5. Quel est l'apport de Zeilberger ? Je l'illustrerai sur un exemple typique, et je renverrai à [B 16] et [B 18] pour un grand nombre d'applications analogues, y compris la formule de Dougall [B 14].

Tout d'abord, la formule (3) de Dixon est le cas particulier $a = b = c = n$ d'une autre formule de Dixon (1903), à savoir

$$(165) \quad \sum_k (-1)^k \binom{a+b}{a+k} \binom{b+c}{b+k} \binom{c+a}{c+k} = \frac{(a+b+c)!}{a!b!c!}$$

(a, b, c sont des entiers positifs). Ce n'est autre qu'une spécialisation de la formule hypergéométrique de Dixon :

$$(166) \quad {}_3F_2 \left(\begin{matrix} 1-a-2n & 1-b-2n & -2n \\ a & b & \end{matrix} \middle| 1 \right) = (-1)^n \frac{(2n)!}{n!} \frac{(a+b+2n-2)_n}{(a)_n (b)_n} .$$

Voici la démonstration de la formule (165) élaborée par Zeilberger, en collaboration avec son fidèle servant électronique Shalosh Ekhad. Posons

$$(167) \quad F(n, k) = (-1)^k \frac{(n+b)!(n+c)!(b+c)!}{(n+k)!(n-k)!(b+k)!(b-k)!(c+k)!(c-k)!} ,$$

$$(168) \quad R(n) = \frac{(n+b+c)!}{n!b!c!} .$$

La formule à démontrer est le cas $n = a$ de la formule

$$(169) \quad \sum_k F(n, k) = R(n) .$$

Or le cas $n = 0$ est évident, et l'on a l'équation de récurrence

$$(170) \quad (n+1)R(n+1) - (n+b+c+1)R(n) = 0 .$$

On aura réussi si l'on découvre un compagnon $G(n, k)$ de $F(n, k)$ (les deux forment ce que les auteurs Wilf et Zeilberger appellent modestement une WZ -paire) satisfaisant à la relation

$$(171) \quad (n+1)F(n+1, k) - (n+b+c+1)F(n, k) = G(n, k) - G(n, k-1) .$$

L'ordinateur Shalosh Ekhad, utilisant les algorithmes décrits au n° 3.5, recherche une solution sous la forme

$$(172) \quad G(n, k) = R(n, k)F(n, k) ,$$

où $R(n, k)$ est une fonction rationnelle. Par ailleurs, $F(n, k)$ satisfait aux deux équations de récurrence

$$(173) \quad F(n+1, k) = A(n, k)F(n, k) , \quad F(n, k+1) = B(n, k)F(n, k) ,$$

et après division par $F(n, k)$, la relation (171) s'écrit

$$(174) \quad (n+1)A(n, k) - (n+b+c+1) = R(n, k) - \frac{R(n, k-1)}{B(n, k-1)}.$$

Les fonctions $A(n, k)$ et $B(n, k)$ sont évidemment connues :

$$(175) \quad A(n, k) = \frac{(n+b+1)(n+c+1)}{(n+k+1)(n-k+1)}$$

$$(176) \quad B(n, k) = -\frac{(n-k)(b-k)(c-k)}{(n+k+1)(b+k+1)(c+k+1)},$$

et l'ordinateur découvre la solution suivante à l'équation (174) :

$$(177) \quad R(n, k) = \frac{(b-k)(c-k)}{2(n+k+1)}.$$

Une fois découverte la fonction $R(n, k)$, la vérification de l'équation (174) est triviale.

5. VERROUS ET VOIES NOUVELLES

On dispose maintenant d'un dictionnaire précis ramenant la preuve d'identités hypergéométriques (et en particulier de relations entre coefficients binomiaux) à celle d'identités entre fonctions rationnelles. Mais il resterait à organiser la masse de résultats obtenus, et à deviner les structures sous-jacentes. Au vu de quelques exemples, la clé semble être à rechercher dans une cohomologie à la de Rham de formes différentielles à coefficients fonctions rationnelles, de leurs analogues sur un réseau \mathbf{Z}^s et dans la q -machine. Les modules holonomes sont intimement liés aux fibrés vectoriels munis de connexion intégrable, et il y aurait à explorer des sortes de groupes de Galois différentiels.

Dans les récurrences à une variable k , il faut distinguer souvent le *comportement asymptotique* pour k tendant vers $+\infty$, et pour k tendant vers $-\infty$; par exemple, l'équation

$$f(k+1) = (k+1)f(k)$$

a deux solutions

$$f_1(k) = k! \quad \text{pour } k \geq 0$$

$$f_2(k) = \begin{cases} (-1)^{k+1}/(-k-1)! & \text{pour } k < 0 \\ 0 & \text{pour } k \geq 0. \end{cases}$$

Aomoto [D 1, D 2, D 3] a commencé à étudier ces comportements asymptotiques, aussi pour le cas d'équations du type $\frac{F(qz)}{F(z)} = a(q; z)$, et surtout à plusieurs dimensions.

La formule (165) de Dixon est le cas particulier $n = 3$ d'une formule de Dyson affirmant que, si l'on développe en série de Laurent le produit $\prod_{i \neq j} (1 - x_i/x_j)^{a_i}$ (où i, j parcourent les entiers 1 à n), le terme constant est égal à

$$\frac{(a_1 + \dots + a_n)!}{a_1! \dots a_n!}.$$

Ceci est le point de départ de nombreuses identités du même genre, associées aux algèbres de Lie simples : conjectures de Macdonald et Morris-Macdonald. Presque tous les cas sont aujourd'hui traités, mais certains (comme le cas de G_2 et de F_4) n'ont cédé qu'aux ordinateurs utilisant les méthodes de Zeilberger. Ici, on rencontre une difficulté subtile : si $x, x^2, x^3, x_1 + x_2, x_1 + x_2 + x_3$ sont des polynômes, il n'en est plus ainsi de x^n ou de $x_1 + \dots + x_n$ du point de vue algorithmique (pour n indéterminé).

On bute sur un autre verrou lorsque l'on veut prouver des identités ne contenant aucun paramètre ; par exemple, pour prouver à l'ordinateur la formule de Rogers-Ramanujan, il faut passer par un intermédiaire :

$$(178) \quad \sum_k \frac{q^{k^2}}{(q; q)_k (q; q)_{n-k}} = \sum_k \frac{(-1)^k q^{(5k^2-k)/2}}{(q; q)_{n-k} (q; q)_{n+k}}$$

qui contient un paramètre libre n , puis faire tendre n vers l'infini.

On rencontre une difficulté analogue si l'on veut prouver des relations purement numériques. Comment prouver qu'il s'agit du même nombre π dans les formules suivantes :

$$\frac{\pi}{4} = \sum_{n=1}^{\infty} \frac{(-1)^n}{2n+1}, \quad \frac{\pi^2}{6} = \sum_{n=1}^{\infty} \frac{1}{n^2}, \quad \int_{-\infty}^{+\infty} e^{-x^2/2} dx = \sqrt{2\pi} \quad ?$$

Au fond, qu'est-ce qu'un nombre réel du point de vue de l'analyse algébrique au sens de Mikio Sato et de ses émules ? Une remarque de Métropolis et Rota est peut-être pertinente : soit $(a_n)_{n \geq 0}$ une suite de nombres entiers à croissance polynomiale

$$(179) \quad |a_n| \leq Cn^k \quad (n \geq 1)$$

pour deux constantes $C > 0$ et $k \geq 1$. Soit $b \geq 2$ un entier. Alors la série $\sum_{n=0}^{\infty} a_n/b^n$ converge (trivial) ; sa somme est 0 si et seulement s'il existe une autre suite (c_n) à croissance polynomiale et telle que

$$(180) \quad a_n = c_n - bc_{n-1} .$$

Voilà un nouvel exemple de somme télescopique.

Enfin, l'analogie entre sommes de Gauss et fonction gamma est bien connue. Une identité de sommes de Gauss, due à Anna Helversen-Pasotto [E 8], est en fait l'analogie de l'identité de Gauss sur ${}_2F_1 \left(\begin{matrix} a & b \\ c \end{matrix} \middle| 1 \right)$. Comment généraliser ceci pour les relations de Dixon, Saalschütz, etc... ? Quelle est la méthode algorithmique sous-jacente ? Y a-t-il un lien avec les faisceaux-caractères de Lusztig, ou les groupes quantiques ? Place au rêve !

BIBLIOGRAPHIE RAISONNÉE

A. Ouvrages de base

Voici les répertoires classiques pour les séries, les intégrales, les fonctions spéciales :

- [1] M. ABRAMOWITZ et I. STEGUN, *Handbook of Mathematical functions*, Dover, New-York, 1965.
- [2] A. ERDELYI (éditeur), *Higher transcendental functions*, 3 volumes, McGraw-Hill, New York, 1953.

- [3] I. GRADSHTEYN et I. RYZHIK, *Table of integrals, series and products*, Academic Press, New York, 1980.

Quelques traités classiques, comportant une partie importante sur les fonctions hypergéométriques :

- [4] P. APPELL et J. KAMPÉ DE FÉRIET, *Fonctions hypergéométriques et hypersphériques*, Gauthier-Villars, Paris, 1926.
- [5] J.E. RAINVILLE, *Special functions*, MacMillan, New York, 1960.
- [6] L. SLATER, *Confluent hypergeometric functions*, Cambridge University Press, Cambridge, 1960.
- [7] E. WHITTAKER et G. WATSON, *Modern Analysis*, Cambridge University Press, Cambridge, 1946.

Quelques bonnes références sur la combinatoire :

- [8] M. AIGNER, *Combinatorial theory*, Springer, Berlin, 1979 (voir surtout le chapitre 3).
- [9] G. ANDREWS, *The theory of partitions*, Addison-Wesley, Reading, 1976.
- [10] L. COMTET, *Advanced Combinatorics*, Reidel, Dordrecht, 1974. $\dot{\iota}$
- [11] R. GRAHAM, D. KNUTH et O. PATASHNIK, *Concrete Mathematics*, Addison-Wesley, Reading, 1989.
- [12] D. KNUTH, *The art of computer programming*, 3 volumes, Addison-Wesley, Reading, 1968-1973.
- [13] P. MACMAHON, *Combinatory analysis*, 2 volumes, Chelsea, New York, 1960 (réimpression).
- [14] J. RIORDAN, *An introduction to combinatorial analysis*, John Wiley, New York, 1958.

Pour terminer, quelques exposés généraux sur les fonctions q -hypergéométriques (dites aussi “basiques”) :

- [15] G. ANDREWS, *q -series : Their development and application in analysis, number theory, combinatorics, physics and computer algebra*,

- CBMS Regional Conference Lecture Series, 66, Amer. Math. Soc., Providence, 1986.
- [16] W. BAILEY, *Generalized hypergeometric series*, Cambridge University Press, Cambridge, 1935 (réimprimé par Stechert-Hafner, New York, 1964).
- [17] H. EXTON, *q-hypergeometric functions and applications*, Ellis Horwood/John Wiley, New York, 1983.
- [18] N. FINE, *Basic hypergeometric series and applications*, Math. Surv. **27**, Amer. Math. Soc., Providence, 1988.
- [19] G. GASPER et M. RAHMAN, *Basic hypergeometric series*, Cambridge University Press, Cambridge, 1990.
- [20] E. HEINE, *Handbuch der Kugelfunktionen. Theorie und Anwendung*, 2 volumes, Springer, 1898 (=Physica Verlag, Würzburg, 1961).
- [21] L. SLATER, *Generalized hypergeometric functions*, Cambridge University Press, Cambridge, 1966.

B. Articles de Zeilberger et collaborateurs

- [1] D. ZEILBERGER, *Sister Celine's technique and its generalizations*, J. Math. Anal. Appl. **85** (1982), 114-145.
- [2] D. ZEILBERGER, *A Holonomic systems approach to special functions identities*, J. of Computational and Applied Math. **32** (1990), 321-368.
- [3] D. ZEILBERGER, *A Fast Algorithm for proving terminating hypergeometric identities*, Discrete Math. **80** (1990), 207-211.
- [4] D. ZEILBERGER, *The method of creative telescoping*, J. Symbolic Computation **11** (1991), 195-204.
- [5] D. ZEILBERGER, *Closed Form (pun intended !)*, to appear in : "Special volume in memory of Emil Grosswald", M. Knopp, ed., Contemporary Mathematics, AMS.
- [6] D. ZEILBERGER, *Three recitations on Holonomic Systems and Hypergeometric Series*, Proceedings of the Séminaire Lotharingien de combinatoire **24**, IRMA, Strasbourg, à paraître.

- [7] D. ZEILBERGER, *Plain (Lagrange interpolation) proofs of Fancy (representation theory) formulas*, en préparation.
- [8] S.B. EKHAD, *Short proofs of two hypergeometric summation formulas of Karlsson*, Proc. Amer. Math. Soc. **107** (1989), 1143-1144.
- [9] S.B. EKHAD, *A very short proof of Dixon's theorem*, J. Comb. Theo., Series A **54** (1990), 141-142.
- [10] S.B. EKHAD, *A one-line proof of the Habsieger-Zeilberger G_2 constant term identity*, J. Comput. Appl. Math. **34** (1991), 133-134.
- [11] S.B. EKHAD, *Short Proof Of A "Strange" Combinatorial Identity Conjectured by Gosper*, Discrete Math. à paraître.
- [12] S.B. EKHAD, *A Short. Elementary and Easy. WZ proof of the Askey-Gasper inequality that was used by de Branges in his proof of the Bieberbach conjecture*, à paraître.
- [13] S.B. EKHAD and S. TRE, *A purely verification proof of the first Rogers-Ramanujan identity*, J. Comb. The. Ser. A **54** (1990), 309-311.
- [14] S.B. EKHAD and D. ZEILBERGER, *A 21st century proof of Dougall's hypergeometric sum identity*, J. Math. Anal. Appl. **147** (1990), 610-611.
- [15] H.S. WILF and D. ZEILBERGER, *Towards computerized proofs of identities*, Bulletin of the Amer. Math. Soc. **23** (1990), 77-83.
- [16] H.S. WILF and D. ZEILBERGER, *Rational functions certify combinatorial identities*, J. Amer. Math. Soc. **3** (1990), 147-158.
- [17] H.S. WILF and D. ZEILBERGER, *A general theory of multi-variate hypergeometric identities*, en préparation.
- [18] H.S. WILF, *54 computer-generated proofs of binomial coefficient identities*, à paraître.

Voici enfin trois articles donnant des programmes détaillés, sous MAPLE, implémentant les algorithmes théoriques :

- [19] G. ALMKVIST et D. ZEILBERGER, *The method of differentiating under the integral sign*, Journ. Symb. Computation, **10** (1990), 571-591.
- [20] G. ALMKVIST et D. ZEILBERGER, *A MAPLE program that finds, and proves, recurrences and differential equations satisfied by hyperexponential definite integrals*, SIGSAM Bulletin **25** (1991),...
- [21] D. ZEILBERGER, *A MAPLE program for proving hypergeometric series*, SIGSAM Bulletin **25** (1991),...

C. Algèbre des opérateurs différentiels et élimination

- [1] J. BERNSTEIN, *Modules over the ring of differential operators. A study of the fundamental solutions of equations with constant coefficients*, Funk. Analysis, Akademia Nauk CCCR **5** (2) (1971), 1-16.
- [2] J. BERNSTEIN, *The analytic continuation of generalized functions with respect to a parameter*, Funct. Anal. and Appl. **6** (1972), 273-285.
- [3] J.-E. BJÖRK, *Rings of Differential Operators*, North Holland, Amsterdam, 1979.
- [4] A. BOREL et al., *Algebraic D-modules*, Perspectives in Math. **2**, Academic Press, Boston, 1987.
- [5] B. BUCHBERGER, *An algorithmic method in polynomial ideal theory*, N.K. Bose ed. Recent trends in multidimensional systems theory, D. Reidel Publishing Corp., 1985.
- [6] F. EHLERS, *The Weyl Algebra*, Chapitre V de [4], 173-205.
- [7] A. GALLIGO, *Some algorithmic questions on ideals of differential operators*, Lect. Note in Comp. Sci. **204** (1985), 413-421.
- [8] R. GOSPER, *Decision procedure for indefinite hypergeometric summation*, Proc. Nat. Acad. Sci. USA **75** (1978), 40-42.
- [9] M. KASHIWARA, *B-functions and holonomic systems*, Invent. Math. **38** (1976), 33-53.

- [10] M. KASHIWARA, *Vanishing cycles sheaves and holonomic systems of differential equations*, Springer Lect. Notes in Math. **1016** (1983), 134-142.
- [11] B. MALGRANGE, *L'involutivité des caractéristiques des systèmes différentiels et microdifférentiels*, Sémin. Bourbaki, exposé 522, 1977-1978 (Lect. Notes in Math. **710**, 277-289).
- [12] R. RISCH, *The solution of the problem in integrating in finite terms*, Bull. Amer. Math. Soc. **76** (1970), 605-608.
- [13] N. TAKAYAMA, *Gröbner basis and the problem of contiguous relations*, Japan Journ. Appl. Math. **6** (1989), 147-160.
- [14] N. TAKAYAMA, *An algorithm for constructing the integral of a module - an infinite dimensional analog of Gröbner basis*, Proceedings of ISSAC'90, A.C.M. Press.
- [15] N. TAKAYAMA, *An approach to the zero recognition problem by Buchberger algorithm*, Journ. Symb. Computation,

D. Systèmes holonomes aux q -différences

- [1] K. AOMOTO, *A note on holonomic q -difference systems*, Algebraic Analysis (in honor of M. Sato), M. Kashiwara and T. Kawai eds., Academic Press, (1988), 25-28.
- [2] K. AOMOTO, *q -analogue of de Rham cohomology associated with Jackson integrals*, Proc. Japan Acad. **66** (1990), 161-164.
- [3] K. AOMOTO, *Finiteness of a cohomology associated with certain Jackson integrals*, Tohoku J. Math. **43** (1991), 75-101.
- [4] C. SABBAH, *Systèmes holonomes d'équations aux q -différences*, Pré-publication Centre de Math. Ecole Polytechnique, Palaiseau, 1991.

E. Perspectives diverses

Sur la formule de Dixon :

- [1] P. CARTIER et D. FOATA, *Problèmes combinatoires de commutation et réarrangements*, Lect. Notes Math. vol. 85, Springer 1969.

Sur l'irrationalité de $\zeta(3)$:

- [2] R. APÉRY, *Irrationalité de $\zeta(2)$ et $\zeta(3)$* , Astérisque **61** (1979), 11-13.
- [3] A. van des POORTEN, *A proof that Euler missed... Apéry's proof of the irrationality of $\zeta(3)$* , Math. Intelligencer, **1** (1979), 195-203.

Sur les conjectures de Macdonald :

- [4] I.G. MACDONALD, *Affine root systems and Dedekind η -function*, Invent. Math. **15** (1972), 91-143.
- [5] I.G. MACDONALD, *Some conjectures for root systems*, SIAM Journ. Math. Anal. **13** (1982), 988-1007.
- [6] D. ZEILBERGER, *Unified approach to Macdonald's root system conjectures*, SIAM Journ. Math. Anal. **19** (1988), 987-1013.
- [7] F. GARVAN et G. BONNET, *Macdonald's constant term conjectures for exceptional root systems*, Bull. Amer. Math. Soc. **24** (1991), 343-347.

Sur les sommes de Gauss :

- [8] A. HELVERSEN-PASOTTO, *L'identité de Barnes pour les corps finis*, C.R. Acad. Sci. Paris, Série A, **286** (1978), 297-300.
- [9] J. GREENE et D. STANTON, *A character sum evaluation and Gaussian hypergeometric series*, Journ. Number Theory, **23** (1986), 136-148.

Pierre CARTIER
Ecole Normale Supérieure
Département de Mathématiques
et Informatique
45, rue d'Ulm
F- 75230 PARIS Cedex 05

Astérisque

ÉTIENNE GHYS

Dynamique des flots unipotents sur les espaces homogènes

Astérisque, tome 206 (1992), Séminaire Bourbaki, exp. n° 747, p. 93-136

http://www.numdam.org/item?id=SB_1991-1992__34__93_0

© Société mathématique de France, 1992, tous droits réservés.

L'accès aux archives de la collection « Astérisque » (<http://smf4.emath.fr/Publications/Asterisque/>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

**DYNAMIQUE DES FLOTS UNIPOTENTS
SUR LES ESPACES HOMOGÈNES**

par Étienne GHYS

INTRODUCTION

Considérons le flot ϕ^t sur le tore $T^n = \mathbf{R}^n / \mathbf{Z}^n$ défini par : $\phi^t(x \bmod \mathbf{Z}^n) = (x + tw) \bmod \mathbf{Z}^n$, où $w = (w_1, \dots, w_n)$ est un vecteur de \mathbf{R}^n . On sait depuis L. Kronecker que si les w_i sont linéairement indépendants sur \mathbf{Q} , toutes les orbites de ϕ^t sont denses dans le tore. En général, l'adhérence d'une orbite est un sous-tore de T^n , de dimension égale au rang de $\{w_i\}$ sur \mathbf{Q} .

Soit maintenant G un groupe de Lie connexe et Γ un réseau de G , c'est-à-dire un sous-groupe discret tel que le volume (de Haar) du quotient à gauche $\Gamma \backslash G$ soit fini. Un sous-groupe H de G opère naturellement à droite sur $\Gamma \backslash G$. Dans quelles conditions a-t-on un résultat analogue à celui de L. Kronecker garantissant l'homogénéité des adhérences d'orbites ?

Dans une série de quatre articles, M. Ratner vient d'obtenir une réponse très satisfaisante en résolvant une conjecture de M.S. Raghunathan [71-72-73-74]. Pour l'énoncer, convenons de dire qu'un élément g d'un groupe de Lie G est *unipotent* si l'automorphisme adjoint $\text{Ad}(g)$ de l'algèbre de Lie \mathcal{G} est unipotent, c'est-à-dire n'a que 1 comme valeur propre. Un sous-groupe H de G est unipotent si tous ses éléments sont unipotents.

THÉORÈME (M. Ratner, 1990).— *Soit H un sous-groupe unipotent d'un groupe de Lie G et Γ un réseau de G . Pour tout x de $\Gamma \backslash G$, il existe un sous-groupe fermé $H(x)$ de G tel que l'adhérence de l'orbite $x.H$ de x par H dans $\Gamma \backslash G$ coïncide avec l'orbite $x.H(x)$ de x par $H(x)$.*

Cet énoncé n'est pas le meilleur possible ; nous en donnerons d'autres versions plus loin.

Il n'est pas surprenant que ce résultat ait des conséquences concernant les approximations diophantiennes. Avant la preuve du théorème précédent, G.A. Margulis avait développé des méthodes qui lui permirent de démontrer en 1987 la conjecture de A. Oppenheim [50].

THÉORÈME (G.A. Margulis, 1987).— *Soit Q une forme quadratique indéfinie et non dégénérée sur \mathbf{R}^n ($n \geq 3$). On suppose que Q n'est pas un multiple d'une forme rationnelle. Alors, pour tout $\varepsilon > 0$, il existe un vecteur v de \mathbf{Z}^n tel que $0 < Q(v) < \varepsilon$.*

Par la suite, S.G. Dani et G.A. Margulis ont amélioré significativement ce théorème comme nous le verrons plus loin.

Les preuves des théorèmes de S.G. Dani-G.A. Margulis et M. Ratner sont longues et difficiles ; il n'est pas possible de les décrire ici. Dans cet exposé élémentaire, nous nous sommes fixés un but très modeste. Nous nous proposons d'illustrer les méthodes employées dans le cas simple où le groupe G est $SL(2, \mathbf{R})$: il s'agit alors d'étudier le flot horicyclique d'une surface à courbure -1 . Ce cas renferme déjà une bonne partie des difficultés. Nous espérons que son étude préliminaire permettra au lecteur d'aborder plus facilement les preuves complètes qui sont publiées ou en cours de publication. Signalons par ailleurs l'existence de l'article [27] de S.G. Dani et G.A. Margulis qui propose une introduction élémentaire à ces questions d'approximations diophantiennes.

Ce rapport est organisé de la façon suivante. Le premier paragraphe est consacré à des motivations : il retrace l'histoire du flot horicyclique et mène à des énoncés plus précis des théorèmes de M. Ratner. Le paragraphe 2 décrit quelques méthodes utilisées par S.G. Dani et G.A. Margulis pour approcher la conjecture de M.S. Raghunathan. Le paragraphe 3 donne quelques ingrédients de la preuve de M. Ratner, toujours illustrés dans le cas de $SL(2, \mathbf{R})$. Enfin, dans le paragraphe 4, nous discutons trois applications : valeurs des formes quadratiques sur les points entiers, phénomènes de rigidité ergodique et existence de laminations géodésiques sur les variétés hyperboliques.

1. LE FLOT HORICYCLIQUE

1.1. Les flots géodésiques et horicycliques des surfaces à courbure négative sont des exemples très riches de systèmes dynamiques. Leur analyse de plus en plus approfondie s'est faite en parallèle avec le développement de la théorie ergodique. Bien que ces flots aient des comportements topologiques et ergodiques très différents, on ne peut envisager d'étudier l'un sans l'autre.

Rappelons d'abord les définitions. Soit H^2 le demi-plan de Poincaré $\{x + iy \in \mathbf{C} \mid y > 0\}$ muni de la métrique à courbure -1 donnée par $ds^2 = (dx^2 + dy^2)/y^2$. Une géodésique de H^2 est une courbe $\gamma : \mathbf{R} \rightarrow H^2$ qui est un plongement isométrique. L'image d'une géodésique est un demi-cercle (ou une demi-droite) orthogonal à l'axe des x . Une géodésique est entièrement déterminée par son vecteur tangent en $t = 0$, élément du fibré unitaire tangent $T_1 H^2$ à H^2 . Bien sûr, si γ est une géodésique et si $s \in \mathbf{R}$, l'application $\gamma^s : t \in \mathbf{R} \mapsto \gamma(t + s) \in H^2$ est aussi une géodésique. Le groupe à un paramètre de transformations $\gamma \mapsto \gamma^s$ définit donc un flot sur l'espace des géodésiques, c'est-à-dire sur $T_1 H^2$. C'est le *flot géodésique* de H^2 .

Le groupe des isométries directes de H^2 est $\mathrm{PSL}(2, \mathbf{R})$ opérant sur H^2 par $\pm \begin{pmatrix} a & b \\ c & d \end{pmatrix} (z) = \frac{az+b}{cz+d}$. Ce groupe opère librement et transitivement à gauche sur $T_1 H^2$ et commute évidemment avec le flot géodésique. Il en résulte que ce flot géodésique peut aussi être considéré comme un groupe à un paramètre g^s de translations à droite sur $\mathrm{PSL}(2, \mathbf{R})$. On s'assure facilement que ce flot s'identifie à :

$$\begin{aligned} g^s &: \pm \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{PSL}(2, \mathbf{R}) \\ &\mapsto \pm \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} \exp(s/2) & 0 \\ 0 & \exp(-s/2) \end{pmatrix} \in \mathrm{PSL}(2, \mathbf{R}). \end{aligned}$$

Soit $v \in T_1 H^2$ et $\gamma : \mathbf{R} \rightarrow H^2$ la géodésique issue de v . Lorsque T tend vers $+\infty$ (resp. $-\infty$), le cercle de centre (non euclidien) $\gamma(T)$ et passant par $\gamma(0)$ "converge" vers l'*horicycle* positif (resp. négatif) déterminé par v (voir figure 1). Dans le modèle de Poincaré, c'est un cercle tangent à l'axe

des x (privé de son point de tangence) ou une droite parallèle à l'axe des x . En termes intrinsèques, ce sont les courbes de courbure ± 1 .

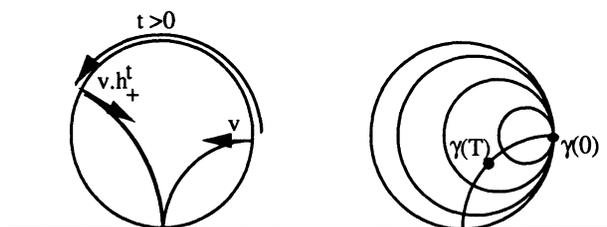


Figure 1

De même qu'avec les géodésiques, les horicycles permettent de définir un flot. Si v est un vecteur unitaire tangent à H^2 en un point z , et si t est un réel, on note $v.h_+^t$ le vecteur unitaire orthogonal à l'horicycle positif déterminé par v , en un point situé à une distance t de z le long de l'horicycle (voir la figure 1 pour les conventions d'orientation). Pour la même raison que précédemment, ce flot s'identifie à un groupe à un paramètre h_+^t de translations à droite de $\text{PSL}(2, \mathbf{R})$.

On vérifie facilement qu'il s'agit de :

$$\pm \begin{pmatrix} a & b \\ c & d \end{pmatrix} . h_+^t = \pm \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix} .$$

En utilisant les horicycles négatifs, on construit de même un troisième flot h_-^t qui correspond bien sûr aux translations à droite par $\begin{pmatrix} 1 & 0 \\ t & 1 \end{pmatrix}$. Les deux flots h_+^t et h_-^t sont conjugués par la symétrie envoyant un vecteur v de $T_1 H^2$ sur son opposé. C'est pour cette raison que nous nous contenterons d'étudier h_+^t .

On notera que tous ces flots préservent la mesure de Haar (bi-invariante) de $\text{PSL}(2, \mathbf{R})$. D'autre part, les champs de vecteurs invariants à gauche correspondant à g^s , h_+^t et h_-^t forment une base de l'algèbre de Lie de $\text{PSL}(2, \mathbf{R})$.

Le rapport entre les flots géodésiques et horicycliques est contenu dans la formule suivante, fondamentale pour la suite :

$$(*) \quad g^s h_{\pm}^t g^{-s} = h_{\pm}^{t \exp(\pm s)}.$$

Ceci est très facile à vérifier par un calcul matriciel. Le flot géodésique normalise donc les flots horicycliques. Il contracte les horicycles positifs et dilate les horicycles négatifs. Cette observation est à la source de la théorie des flots d'Anosov [1] (voir la figure 2) :

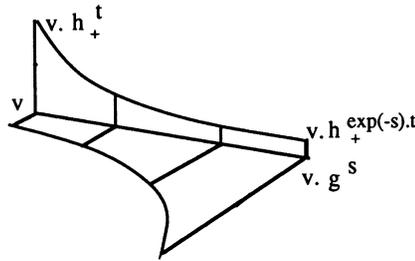


Figure 2

Soit maintenant Γ un sous-groupe discret de covolume fini de $\mathrm{PSL}(2, \mathbf{R})$. Lorsque l'action de Γ sur H^2 est libre, on obtient une surface $\Sigma = \Gamma \backslash H^2$ dont l'aire est finie et dont le fibré unitaire tangent s'identifie à $\Gamma \backslash \mathrm{PSL}(2, \mathbf{R})$. En général, l'action de Γ sur H^2 peut ne pas être libre mais les stabilisateurs sont finis et le quotient $\Sigma = \Gamma \backslash H^2$ est une surface singulière qu'il vaut mieux considérer comme une "orbifold" (au sens de W. Thurston). Un exemple très important est celui de $\Gamma = \mathrm{PSL}(2, \mathbf{Z})$: on obtient l'orbifold modulaire dont l'étude est fondamentale en théorie des nombres.

Les flots h_{\pm}^t et g^s passent bien sûr au quotient sur $\Gamma \backslash \mathrm{PSL}(2, \mathbf{R})$: ce sont les flots horicycliques et géodésiques de la surface (resp. orbifold) $\Gamma \backslash H^2$. Ces flots n'ont pas de point fixes. Ils préservent une forme de volume, de masse totale finie, que l'on peut appeler au choix mesure de Haar ou mesure de Liouville suivant le point de vue adopté.

1.2. Pour illustrer le fait que ces flots géodésiques et horicycliques s'étudient en parallèle, nous allons commencer par quelques exemples simples et

bien classiques.

PROPOSITION.— *Si $\Gamma \backslash \text{PSL}(2, \mathbf{R})$ est compact, le flot horicyclique h_{\pm}^t n'a pas d'orbite périodique.*

Preuve : La relation (*) montre que si γ est une orbite périodique de h_{+}^t de période T , alors l'image $\gamma.g^s$ est une autre orbite périodique, de période $\exp(-s).T$ tendant vers 0 lorsque s tend vers $+\infty$. La proposition résulte du fait qu'un flot sans point fixe sur une variété compacte ne peut avoir des orbites périodiques de périodes arbitrairement petites. •

Dans le cas où $\Gamma \backslash \text{PSL}(2, \mathbf{R})$ est de volume fini mais non compact, la surface $\Sigma = \Gamma \backslash H^2$ (supposée non singulière pour simplifier) se décompose en une partie compacte et un nombre fini de *pointes* (cusps) (voir figure 3). Chacune de ces pointes est isométrique au quotient de $\{x + iy \in \mathbf{C}, y \geq y_0 > 0\}$ par les translations entières $z \mapsto z+n$. La courbe $y = y_0$ donne dans Σ un horicycle fermé, c'est-à-dire une orbite fermée γ pour h_{+}^t . Bien sûr, pour tout s réel, la courbe $\gamma.g^s$ est aussi une orbite périodique. Lorsque s tend vers $+\infty$, cette orbite $\gamma.g^s$ tend vers un bout de Σ (voir figure 3). Il n'est pas difficile de se convaincre que l'on obtient ainsi toutes les orbites périodiques de h_{+}^t qui constituent donc un nombre fini de familles paramétrées par \mathbf{R} . Lorsque s tend vers $-\infty$, il se trouve que les orbites $\gamma.g^s$ sont de plus en plus denses dans $T_1\Sigma$. On trouvera dans [19], [77], [85], une description de ce phénomène et une intéressante reformulation de l'hypothèse de Riemann en ces termes d'horicycles fermés.

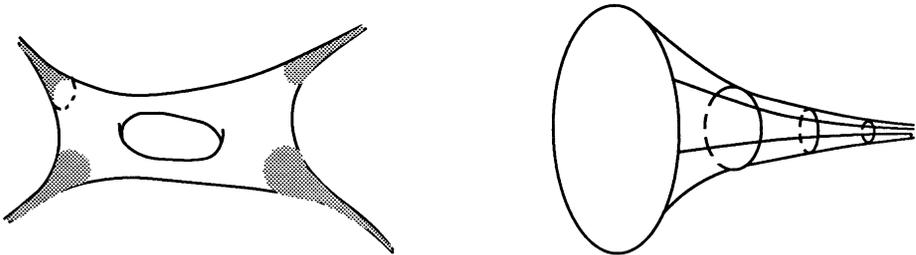


Figure 3

Voici un exemple où le flot horicyclique permet de comprendre le flot géodésique. C'est le fameux théorème de E. Hopf [41-42] :

THÉORÈME.— *Si Γ est un réseau de $\mathrm{PSL}(2, \mathbf{R})$, le flot géodésique sur $\Gamma \backslash \mathrm{PSL}(2, \mathbf{R})$ est ergodique par rapport à la mesure de Haar, i.e. une fonction mesurable constante sur les orbites du flot est constante presque partout.*

Preuve : C'est un exemple d'application du "phénomène de Mautner" (voir [55]). Soit λ la mesure de Haar sur $\Gamma \backslash \mathrm{PSL}(2, \mathbf{R})$ normalisée de façon à être une probabilité. Soit F une fonction de carré λ -intégrable sur $\Gamma \backslash \mathrm{PSL}(2, \mathbf{R})$ qui est invariante par le flot géodésique g^s . Nous allons montrer que F est aussi invariante (λ -presque partout) par le flot horicyclique h_+^t . En effet, l'utilisation de la formule (*) et le fait que g^s préserve λ montrent :

$$\begin{aligned} \int F(v) \cdot F(v \cdot h_+^t) d\lambda &= \int F(v \cdot g^s) \cdot F(v \cdot g^s h_+^t) d\lambda \\ &= \int F(v) \cdot F(v \cdot h_+^{\exp(-s) \cdot t}) d\lambda \xrightarrow{s \rightarrow +\infty} \int F(v) d\lambda. \end{aligned}$$

Ainsi $F(v \cdot h_+^t) = F(v)$ pour λ -presque tout v (cas d'égalité dans l'inégalité de Cauchy-Schwarz). On procède de même avec h_-^t (en faisant tendre s vers $-\infty$). Par conséquent, F est invariante (λ -presque partout) par g^s , h_+^t et h_-^t et donc par l'action à droite de $\mathrm{PSL}(2, \mathbf{R})$ qui est transitive. On a donc bien montré que F est constante λ -presque partout. ●

Réciproquement, le flot géodésique va nous permettre de démontrer le résultat suivant, dû à G. Hedlund [39-40] :

THÉORÈME.— *Si Γ est un réseau de $\mathrm{PSL}(2, \mathbf{R})$, le flot horicyclique h_+^t sur $\Gamma \backslash \mathrm{PSL}(2, \mathbf{R})$ est ergodique par rapport à la mesure de Haar.*

Preuve : Nous allons donner quelques détails car l'idée de la preuve est à la base de beaucoup de techniques que nous verrons plus loin.

Le stabilisateur du point $(1, 0)$ de \mathbf{R}^2 sous l'action linéaire de $\mathrm{SL}(2, \mathbf{R})$ est $H = \{h_+^t\} = \left\{ \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix} \right\}$. L'espace homogène $\mathrm{SL}(2, \mathbf{R})/H$ s'identifie

donc à $\mathbf{R}^2 - \{(0, 0)\}$ et l'action à gauche de H sur $SL(2, \mathbf{R})/H$ est bien sûr par transvections (voir figure 4) :

$$\begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x + ty \\ y \end{pmatrix}.$$

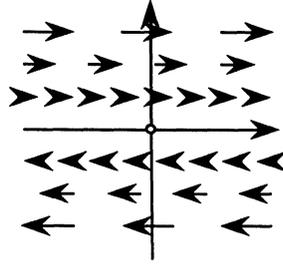


Figure 4

Il est bien clair qu'une fonction continue sur $\mathbf{R}^2 - \{(0, 0)\}$ invariante par ces transvections est une fonction de y . En particulier, une telle fonction est constante sur l'axe des x , et donc sur l'orbite de $(1, 0)$ par $g^s = \begin{pmatrix} \exp(s/2) & 0 \\ 0 & \exp(-s/2) \end{pmatrix}$. Nous avons donc établi le lemme suivant.

Si $\Phi : SL(2, \mathbf{R}) \rightarrow \mathbf{R}$ est une fonction continue, constante sur les doubles classes par H , alors Φ est constante sur le sous-groupe à un paramètre g^s .

Avant d'appliquer ce fait à notre propos, remarquons qu'un espace homogène de $PSL(2, \mathbf{R})$ est aussi un espace homogène de $SL(2, \mathbf{R})$ de sorte que nous pouvons nous placer dans le cas d'un réseau Γ de $SL(2, \mathbf{R})$ (quitte à étendre Γ par $\mathbf{Z}/2\mathbf{Z}$).

Considérons donc une fonction F de carré λ -intégrable sur $\Gamma \backslash SL(2, \mathbf{R})$, invariante par le flot h_+^t . La fonction Φ définie sur $SL(2, \mathbf{R})$ par :

$$\Phi(g) = \int F(v) \cdot F(v.g) d\lambda$$

est continue et vérifie :

$$\Phi(g) = \Phi(h_+^t g) \text{ et } \Phi(g) = \Phi(g h_+^t)$$

car F est invariante par h_+^t et h_+^t préserve λ .

En appliquant l'observation précédente, on trouve :

$$\int F^2(v) d\lambda = \int F(v) \cdot F(v.g^s) d\lambda.$$

Par conséquent, F est aussi invariante (λ -presque partout) par le flot géodésique g^s . Nous savons que ceci entraîne que F est constante λ -presque partout. •

Pour plus d'informations sur le flot géodésique, nous renvoyons à l'exposé de P. Pansu dans ce séminaire [62]. Notons cependant que cette dynamique est très complexe. À titre d'exemples :

i) La réunion des orbites périodiques est dense [1], [6].

ii) Certaines orbites ne sont pas récurrentes : une orbite peut spiraler par exemple sur une orbite périodique (figure 5) ou sur un ensemble limite plus compliqué.

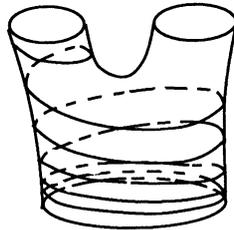


Figure 5

iii) Il existe des ensembles fermés invariants ne contenant aucune orbite périodique [56].

iv) Il existe une infinité non dénombrable de mesures de probabilités invariantes et ergodiques [81].

La situation est assez différente pour le flot horicyclique et, d'une certaine façon, bien plus simple. Le théorème suivant est dû à G. Hedlund [38] : il sera démontré au paragraphe 3 :

THÉORÈME.— *Si $\Gamma \backslash \text{PSL}(2, \mathbf{R})$ est de volume fini, toute orbite du flot horicyclique h_+^t est périodique ou dense.*

Il est naturel d'essayer de décrire la répartition asymptotique des orbites dans $\Gamma \backslash \mathrm{PSL}(2, \mathbf{R})$. Soit ϕ^t un flot sur une variété V préservant une mesure de probabilité μ . Un point v de V est *équiréparti par rapport à μ* si pour toute fonction continue à support compact $F : V \rightarrow \mathbf{R}$, on a :

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T F(v.h_+^t) dt = \int_V F d\mu.$$

Le théorème ergodique de G. Birkhoff garantit que si μ est ergodique, μ -presque tout point v est équiréparti par rapport à μ .

Le théorème suivant est dû à S.G. Dani et J. Smillie [28] : il fait suite à de nombreux travaux (en particulier [7], [8], [32], [35], [37], [46], [84]). Nous esquisserons une preuve d'un résultat plus général au paragraphe 3.

THÉORÈME.— *Si $\Gamma \backslash \mathrm{PSL}(2, \mathbf{R})$ est de volume fini, tout point non périodique pour h_+^t est équiréparti par rapport à la mesure de Haar.*

1.3. Bien sûr, on ne s'est pas contenté longtemps de l'étude des flots géodésiques et horicycliques. Il est naturel d'étudier plus généralement l'action à droite d'un sous-groupe H d'un groupe de Lie G sur un quotient de volume fini $\Gamma \backslash G$ ([2], [9]). Les motivations sont multiples et nous en décrivons quelques-unes au paragraphe 4. Nous renvoyons à S.G. Dani pour un excellent survol de ces travaux [19]. Signalons tout particulièrement l'étude très complète du cas où G est nilpotent [63-64] par W. Parry (et plus récemment [44-45]), ainsi que l'étude des actions de groupes horisphériques par S.G. Dani [21]. C'est sur la base des résultats obtenus par ces auteurs que M.S. Raghunathan conjecture que si H est unipotent, l'adhérence d'une orbite dans $\Gamma \backslash G$ coïncide avec l'orbite d'un sous-groupe de Lie de G contenant H . Cette conjecture entraîne en particulier que ces adhérences d'orbites sont des sous-variétés : une situation bien différente de celle du flot géodésique...

Des progrès décisifs concernant la conjecture sont accomplis par G.A. Margulis et S.G. Dani qui la démontrent (génériquement) lorsque $G = \mathrm{SL}(3, \mathbf{R})$ [26]. Nous décrivons succinctement leurs méthodes au paragraphe 2. En fait, G.A. Margulis formule dans [50-52] une conjecture beaucoup

plus générale que celle de M.S. Raghunathan qui est finalement démontrée par M. Ratner [74-75].

THÉORÈME.— *Soit H un sous-groupe d'un groupe de Lie connexe G , engendré par des éléments unipotents. Soit Γ un réseau de G . Alors, l'adhérence de toute orbite de H sur $\Gamma \backslash G$ coïncide avec l'orbite d'un sous-groupe de Lie de G contenant H .*

L'approche de M. Ratner consiste à étudier d'abord la nature des mesures invariantes par H et d'en déduire, par la suite, la structure des adhérences d'orbites.

Fixons toujours un réseau Γ du groupe de Lie connexe G et soit x un élément de $\Gamma \backslash G$. Considérons un sous-groupe fermé L de G ayant la propriété que $x.L$ est fermé dans $\Gamma \backslash G$. Supposons de plus que l'orbite $x.L \simeq L/x\Gamma x^{-1} \cap L$ soit de volume fini par rapport à la mesure de Haar de L . Comme $x.L$ est contenu dans $\Gamma \backslash G$, on obtient une mesure finie sur $\Gamma \backslash G$. Les mesures obtenues par ce procédé seront appelées *mesures homogènes* sur $\Gamma \backslash G$. Le résultat obtenu par M. Ratner est le suivant [70-71-72-73] :

THÉORÈME.— *Soit H un sous-groupe d'un groupe de Lie G , engendré par des éléments unipotents. Soit Γ un réseau de G . Toute mesure de masse totale finie sur $\Gamma \backslash G$ invariante et ergodique sous l'action de H est une mesure homogène.*

Lorsque H est un groupe à un paramètre ϕ^t , M. Ratner obtient une généralisation du théorème de S.G. Dani et J. Smillie.

THÉORÈME.— *Soit ϕ^t un sous-groupe unipotent à un paramètre du groupe de Lie connexe G et soit Γ un réseau de G . Alors, tout point de $\Gamma \backslash G$ est équidistribué par rapport à une certaine mesure homogène.*

Terminons ce paragraphe par trois remarques.

1. Les résultats de M. Ratner sont en fait plus généraux et couvrent certains cas où Γ n'est pas un réseau. Nous ne donnons pas ici d'énoncé mais nous citerons un exemple en 3.4.
2. La classe des groupes engendrés par des éléments unipotents est bien plus large que celle des sous-groupes unipotents. Par exemple, tout sous-groupe

de Lie de $SL(n, \mathbf{R})$ isomorphe à $SL(2, \mathbf{R})$ est engendré par des éléments unipotents. Un exemple typique de groupe qui n'est pas engendré par des éléments unipotents est le sous-groupe résoluble H de $GL(n, \mathbf{R})$ formé des matrices triangulaires supérieures (au sens large) car les éléments unipotents de H sont triangulaires supérieurs au sens strict.

3. L'hypothèse " H est engendré par des éléments unipotents" est presque optimale. A.N. Starkov montre que si H n'est pas engendré par des éléments h tels que le spectre de $Ad(h)$ est contenu dans le cercle unité, alors il existe des adhérences d'orbites dans $\Gamma \backslash G$ qui ne sont pas des sous-variétés. Par contre, si H est engendré par de tels éléments, il montre (en utilisant les résultats de M. Ratner) que toutes les adhérences d'orbites sont des sous-variétés (mais pas nécessairement orbites d'un sous-groupe de Lie L) [83].

2. DYNAMIQUE TOPOLOGIQUE DES FLOTS UNIPOTENTS

2.1. Dans ce paragraphe, nous proposons une introduction élémentaire aux méthodes développées par G.A. Margulis. Commençons par quelques remarques générales concernant la dynamique des flots unipotents. Fixons un groupe de Lie connexe G , un réseau Γ et un sous-groupe unipotent à un paramètre ϕ^t . Pour simplifier, nous supposons momentanément que $\Gamma \backslash G$ est compact.

Les translations à gauche de G permettent d'identifier les espaces tangents aux divers points de G à l'algèbre de Lie \mathcal{G} (des champs invariants à gauche). Le flot ϕ^t opère à droite sur $\Gamma \backslash G$ et sa différentielle, calculée dans \mathcal{G} , est donnée par l'application adjointe $Ad(\phi^t)$. C'est là une propriété remarquable : la différentielle du flot, évaluée sur une trivialisatation du fibré tangent, ne dépend pas du point où on la considère. On pourra comparer cette propriété à la notion d'"autonomie" dégagée par A. Zeghib [94].

Puisque ϕ^t est unipotent, $Ad(\phi^t)$ est un polynôme en t . Si l'on fixe une structure euclidienne $\| \cdot \|$ sur \mathcal{G} et si $w \in \mathcal{G}$ est non nul, la fonction $t \mapsto \|Ad(\phi^t)(w)\|^2$ est un polynôme qui ne prend que des valeurs strictement positives. En particulier :

$$\inf_{t \in \mathbf{R}} \|Ad(\phi^t)(w)\| > 0.$$

C'est la *distalité infinitésimale* de ϕ^t . Cette propriété est à la source de la plupart des arguments concernant les flots unipotents : elle permet d'exclure l'existence de variétés stables et donc les phénomènes de spiralement. L'étude générale des flots infinitésimalement distaux a été entreprise dans [76] mais il reste beaucoup à faire. Cette propriété entraîne-t-elle par exemple que les adhérences des orbites sont des sous-variétés ?

Dégageons une propriété analogue mais non infinitésimale. Si x et y sont des points suffisamment proches de $\Gamma \backslash G$, il existe un unique petit élément $w(x, y)$ de \mathcal{G} tel que $y = x \cdot \exp w(x, y)$. On a bien sûr : $y \cdot \phi^t = x \cdot \phi^t \cdot \phi^{-t} \cdot \exp w(x, y) \cdot \phi^t$ de sorte que si x et y sont tels que $x \cdot \phi^t$ et $y \cdot \phi^t$ restent proches pour $t \in [0, T]$, alors :

$$w(x \cdot \phi^t, y \cdot \phi^t) = \text{Ad}(\phi^t) w(x, y).$$

On a ainsi une propriété de *divergence polynomiale* des orbites. Insistons sur le fait qu'un flot unipotent n'est pas nécessairement distal, c'est-à-dire que la distance entre $x \cdot \phi^t$ et $y \cdot \phi^t$ dans $\Gamma \backslash G$ peut ne pas être minorée par un nombre strictement positif [36]. Le flot horicyclique par exemple n'est pas distal.

2.2. Densité des horicycles dans le cas cocompact

Nous allons démontrer le théorème de G. Hedlund concernant les adhérences d'orbites avec une preuve inspirée de [50-52-54], elle-même réminiscente de [6] ou [65].

Soit $G = \text{SL}(2, \mathbf{R})$. Notons $H = \{h_+^t\} = \left\{ \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix} \right\}$ et B le sous-groupe des matrices $\begin{pmatrix} \exp(s) & t \\ 0 & \exp(-s) \end{pmatrix}$, normalisant H dans G . Soit Γ un réseau de G . Commençons par le fait très classique suivant :

PROPOSITION.— *Toutes les orbites de B dans $\Gamma \backslash G$ sont denses.*

Preuve : Il s'agit de montrer que toutes les doubles classes $\Gamma \cdot x \cdot B$ sont denses dans G ou, de manière équivalente, que toutes les orbites de Γ dans G/B sont denses. Nous supposons pour simplifier que Γ contient $-\text{id}$, de sorte qu'il suffit de montrer la densité des orbites de Γ dans $\text{PSL}(2, \mathbf{R})/B$,

c'est-à-dire sur le bord $\mathbf{R} \cup \{\infty\}$ du demi-plan de Poincaré. (La preuve est analogue si $-\text{id} \notin \Gamma$.) Par l'absurde, considérons une orbite de Γ qui n'est pas dense et soit I une composante connexe du complémentaire de son adhérence (voir figure 6). La géodésique de H^2 joignant les extrémités de I détermine avec I un demi-espace Π dans H^2 . Le stabilisateur Γ_Π de Π est au plus cyclique et le quotient $\Gamma_\Pi \backslash \Pi$ est donc d'aire infinie. Ceci contredit le fait que Σ est d'aire finie. •

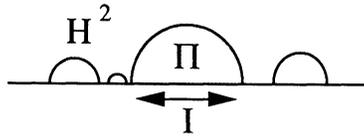


Figure 6

Supposons maintenant Γ cocompact, *i.e.* $\Gamma \backslash G$ compact. Une partie \mathcal{M} de $\Gamma \backslash G$ est un *ensemble minimal* si c'est un fermé non vide invariant par H et minimal pour ces propriétés. Un tel ensemble minimal existe grâce au lemme de Zorn et à la compacité de $\Gamma \backslash G$. Toute orbite par H d'un point de \mathcal{M} est dense dans \mathcal{M} . Nous allons montrer que $\mathcal{M} = \Gamma \backslash G$, c'est-à-dire que toute orbite du flot horicyclique est dense.

Utilisons l'action à droite de G sur $\Gamma \backslash G$ pour définir :

$$M = \{g \in G \mid \mathcal{M}.g \cap \mathcal{M} \neq \emptyset\}.$$

Il est clair que :

- i) M est fermé et contient l'élément neutre id .
- ii) M est une réunion de doubles classes de $\Gamma \backslash G / H$ car \mathcal{M} est invariant par H .

Soit x un point de \mathcal{M} . Nous savons que l'orbite de x par H n'est pas périodique et qu'elle est dense dans \mathcal{M} . Il existe donc une suite g_n d'éléments de G tendant vers l'identité telle que $g_n \notin H$ et $x.g_n \in \mathcal{M}$. En particulier :

- iii) id est adhérent à $M - H$.

Soit $g \in M \cap B$. Comme les éléments de B normalisent H , l'image $\mathcal{M}.g$ est un fermé non vide, invariant par H et intersectant \mathcal{M} non trivialement ; c'est donc que $\mathcal{M}.g = \mathcal{M}$. Ainsi :

iv) $M \cap B$ est un sous-groupe fermé formé d'éléments préservant \mathcal{M} .

Nous avons déjà étudié l'espace $H \backslash G / H$ en 1.2. La projection M' de M dans $G / H \simeq \mathbf{R}^2 - \{(0, 0)\}$ est un fermé s'accumulant sur $(1, 0)$, invariant par les transvections $\begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}$. On en déduit immédiatement que M' contient

une suite de points du type $(\exp(s_n), 0)$ avec $s_n \neq 0$ convergeant vers 0.

Un réel s est tel que $(\exp(s), 0)$ appartient à M' si et seulement si

$\begin{pmatrix} \exp(s) & t \\ 0 & \exp(-s) \end{pmatrix}$ est dans M (pour une valeur quelconque de t). La

remarque iv) montre donc que l'ensemble de ces réels s est un sous-groupe additif fermé de \mathbf{R} , non discret d'après ce qui précède. Il en résulte que M contient B . Toujours grâce à iv), on déduit que le groupe B tout entier préserve \mathcal{M} . Comme nous savons que les orbites de B sont denses dans $\Gamma \backslash G$, nous avons montré que $\mathcal{M} = \Gamma \backslash G$. Ceci achève la preuve du théorème de G. Hedlund dans le cas où $\Gamma \backslash G$ est compact.

2.3. Le cas des réseaux non cocompacts : la non-divergence à l'infini

Plaçons-nous maintenant dans le cas où $\Gamma \backslash G$ est de volume fini mais non compact. Dans la démonstration précédente, la compacité permettait de garantir l'existence d'un ensemble minimal. Une orbite périodique est bien sûr un ensemble minimal mais il n'est plus clair, dans notre cas où $\Gamma \backslash G$ est de volume fini, que tout fermé non vide invariant contient un ensemble minimal. C'est la première propriété à vérifier.

PROPOSITION.— *Tout fermé non vide de $\Gamma \backslash G$ invariant par le flot horicyclique contient une orbite périodique.*

Avant de démontrer la proposition, nous devons énoncer deux lemmes.

Nous avons déjà vu que la surface (orbifold) $\Sigma = \Gamma \backslash H^2$ peut être décomposée en une partie compacte K et un nombre fini de pointes, isométriques au quotient de $\{x + iy \in \mathbf{C} \mid y > y_0 > 0\}$ par les translations entières $z \mapsto z + n$.

Lemme.— *Soit $\gamma : \mathbf{R} \rightarrow \Sigma$ un horicycle non périodique de Σ . Alors les ensembles $\{t \geq 0 \mid \gamma(t) \in K\}$ et $\{t \leq 0 \mid \gamma(t) \in K\}$ ne sont pas bornés.*

Preuve. — Il suffit de constater qu'un demi-horicycle de H^2 contenu dans $\{x+iy \mid y > y_0\}$ est une demi-droite horizontale et produit donc un horicycle périodique dans Σ . ●

Puisque tout fermé non vide de $\Gamma \backslash G$ invariant par H et ne contenant pas d'orbite périodique rencontre le compact \overline{K} des vecteurs unitaires tangents à Σ aux points de K , le lemme de Zorn s'applique. Il en résulte que tout fermé non vide invariant contient un ensemble minimal. Le lemme suivant, dont nous laissons la démonstration (facile) au lecteur, montre que tout ensemble minimal est nécessairement compact. (Voir [53] pour des énoncés plus généraux.)

Lemme. — Soit ϕ^t un flot sur un espace localement compact X . On suppose qu'il existe une partie compacte \overline{K} de X telle que, pour tout x , les parties $\{t \geq 0 \mid x.\phi^t \in \overline{K}\}$ et $\{t \leq 0 \mid x.\phi^t \in \overline{K}\}$ de \mathbf{R} ne sont pas bornées. Alors X est compact.

Nous pouvons maintenant démontrer la proposition. Nous avons vu que tout fermé non vide invariant contient un ensemble minimal compact \mathcal{M} . Si cet ensemble \mathcal{M} n'était pas réduit à une orbite périodique, l'argument du paragraphe 2.2 s'appliquerait sans modification et on aurait $\mathcal{M} = \Gamma \backslash G$ ce qui est absurde puisque nous avons supposé que $\Gamma \backslash G$ n'est pas compact. ●

La proposition que nous venons de montrer est bien sûr plus faible que le théorème de G. Hedlund selon lequel une orbite non périodique est dense. Pour le démontrer, nous pouvons maintenant utiliser la même méthode qu'en 2.2. Soit x un point non périodique de $\Gamma \backslash G$ et F l'adhérence de l'orbite de x dans $\Gamma \backslash G$. Nous savons que F contient un point périodique y , de période $T > 0$. Posons :

$$M = \{g \in G \mid y.g \in F\}.$$

Alors, M est fermé, réunion de classes à droite de G/H , invariant par l'action à gauche de h_+^T , et contient une suite g_n de $G - H$ convergeant vers id . Le fermé M' correspondant dans $\mathbf{R}^2 - \{(0,0)\}$ est invariant par $\begin{pmatrix} 1 & T \\ 0 & 1 \end{pmatrix}$ et s'accumule sur $(1,0)$. Nous savons que les points $y.g$ avec $g \in B$

sont périodiques pour le flot horicyclique et que, d'autre part, l'orbite non périodique de x s'accumule sur y . Il en résulte que $(1, 0)$ est la limite d'une suite de points de M' situés hors de $\mathbf{R} \times \{0\}$.

Toutes ces propriétés montrent que M' contient $\mathbf{R}^* \times \{0\}$. Autrement dit, pour tout g de B , on a $y.g \in F$. Comme toutes les orbites de B sont denses dans $\Gamma \backslash G$, nous avons bien établi que $F = \Gamma \backslash G$, c'est-à-dire que l'orbite non périodique de x est dense dans $\Gamma \backslash G$. ●

2.4. Quelques indications sur le cas général

La plupart des techniques présentées en 2.2 et 2.3 s'étendent au cas d'un sous-groupe unipotent d'un groupe de Lie autre que $\mathrm{SL}(2, \mathbf{R})$. C'est une élaboration de ces idées qui a permis à S.G. Dani et G.A. Margulis de démontrer la conjecture de M.S. Raghunathan dans le cas où $G = \mathrm{SL}(3, \mathbf{R})$, $\Gamma = \mathrm{SL}(3, \mathbf{Z})$ et ϕ^t est un groupe unipotent à un paramètre générique (i.e. tel que $\phi^t - \mathrm{id}$ est de rang 2 pour tout $t \neq 0$). Nous n'entrerons pas dans les détails de la preuve car [27] permet une excellente introduction élémentaire.

Dans la démonstration que nous avons donnée du théorème de G. Hedlund, il importait de constater qu'un horicycle non périodique ne peut séjourner qu'un temps fini dans une pointe sans en sortir. Un énoncé analogue est valable en toute généralité mais sa preuve est considérablement plus difficile. Voici deux résultats dans cette direction.

THÉORÈME.— *Soit G un groupe de Lie semi-simple connexe et Γ un réseau. Étant donné $\varepsilon > 0$, il existe un compact K de $\Gamma \backslash G$ tel que, pour tout point x de $\Gamma \backslash G$, et tout sous-groupe unipotent à un paramètre ϕ^t , on ait deux possibilités :*

i) *il existe un sous-groupe fermé propre L de G contenant ϕ^t tel que l'orbite $x.L$ est fermée et de mesure finie (pour la mesure de Haar de L) ;*

ii) *pour T assez grand, on a : $m(\{t \in [0, T] \mid x.\phi^t \in K\}) \geq (1 - \varepsilon)T$ où m désigne la mesure de Lebesgue dans \mathbf{R} .*

THÉORÈME.— *Soit G un groupe de Lie connexe, Γ un réseau, ϕ^t un sous-groupe unipotent à un paramètre, x un point de $\Gamma \backslash G$ et $\varepsilon > 0$. Il existe*

un compact K de $\Gamma \backslash G$ tel que, pour T assez grand :

$$m(\{t \in [0, T] \mid x \cdot \phi^t \in K\}) \geq (1 - \varepsilon)T.$$

Une première version de ces résultats est due à G.A. Margulis qui montre qu'une orbite d'un flot unipotent dans $SL(n, \mathbf{Z}) \backslash SL(n, \mathbf{R})$ ne tend pas vers l'infini [48]. Ce résultat a été amélioré par S.G. Dani sous la forme du premier des deux théorèmes cités ci-dessus [13-14-16-18-22]. La preuve de S.G. Dani sépare les cas où le \mathbf{R} -rang de G est supérieur ou égal à 2 (où il utilise le théorème d'arithméticité de G.A. Margulis) et où ce rang est égal à 1 (pour lequel la preuve est différente et plus simple). La formulation du second théorème cité est extraite de [74]. Pour une introduction aux méthodes de démonstration, nous renvoyons à l'appendice de [27]. Il serait d'ailleurs agréable de trouver une approche unifiée de cette "non divergence à l'infini", indépendante par exemple des difficiles théorèmes d'arithméticité.

3. COMPORTEMENT ERGODIQUE DES FLOTS UNIPO- TENTS

Nous décrivons dans ce paragraphe quelques-unes des méthodes employées par M. Ratner.

3.1. La dérive dans la direction du centralisateur

L'algèbre de Lie $sl(2, \mathbf{R})$ de $SL(2, \mathbf{R})$ est constituée des matrices (2×2) de trace nulle. Une base est :

$$g = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad ; \quad h_+ = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \quad ; \quad h_- = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}.$$

On a :

$$\begin{aligned} \text{Ad}(h_+^t) \begin{pmatrix} a & c \\ b & -a \end{pmatrix} &= \begin{pmatrix} 1 & -t \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a & c \\ b & -a \end{pmatrix} \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} a - tb & t^2b + 2at + c \\ b & -a + tb \end{pmatrix} \end{aligned}$$

Munissons $sl(2, \mathbf{R})$ de la norme $\sup(|a|, |b|, |c|)$ et fixons $C > 0$ et $\varepsilon > 0$ (très petit par rapport à C). Soit $w = \begin{pmatrix} a & c \\ b & -a \end{pmatrix} \in sl(2, \mathbf{R})$ de norme

$\rho < C$. Alors $\text{Ad}(h_+^t)(w)$ tend vers l'infini sauf si $a = b = 0$, c'est-à-dire si w pointe dans la direction de h_+ . Soit t_0 le premier instant positif tel que $\|\text{Ad}(h_+^{t_0})(w)\| = C$. Un petit exercice montre que si ρ est assez petit, c'est le polynôme du second degré qui atteint le premier la valeur $\pm C$, i.e. $|-t_0^2 b + 2at_0 + c| = C$ et, de plus, $|-a + t_0 b|$ (et $|b|$) sont inférieurs à ε . Autrement dit, $\|\text{Ad}(h_+^t)(w) \pm C h_+\| < \varepsilon$ pour $t = t_0$. On peut même supposer, toujours quitte à prendre ρ inférieur à un réel ρ_0 (ne dépendant que de ε et C) que cette dernière inégalité subsiste pour $t \in [t_0, (1 + \alpha)t_0]$ où $\alpha > 0$ ne dépend que de ε et C .

Supposons maintenant que Γ soit un réseau cocompact dans $G = \text{SL}(2, \mathbf{R})$. La norme choisie sur $sl(2, \mathbf{R})$ munit $\Gamma \backslash G$ d'une distance. Les considérations précédentes montrent que, pour C assez petit, on a :

Pour tout $\varepsilon > 0$, il existe $\rho_0 > 0$ et $\alpha > 0$ tels que si deux points x, y de $\Gamma \backslash G$ sont à distance inférieure à ρ_0 et ne vérifient pas $y = x.h_+^t$ avec $|t| < 2\rho_0$, alors il existe $t_0 > 0$ tel que la distance entre $x.h_+^t$ et $y.h_+^{t+C}$, ou entre $x.h_+^t$ et $y.h_+^{t-C}$, est inférieure à ε pour tout t de $[t_0, (1 + \alpha)t_0]$ (voir figure 7).

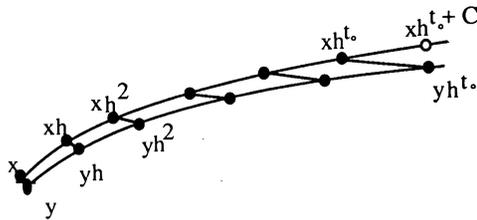


Figure 7

Le lecteur se doutera qu'il s'agit d'un phénomène général pour les flots unipotents ϕ^t . Si w est un petit vecteur et si t_0 désigne le premier instant où $w_{t_0} = \text{Ad}(h_+^{t_0})(w)$ atteint une norme appréciable C , alors w_{t_0} est très proche du centralisateur de ϕ^t dans \mathcal{G} et le reste dans tout un intervalle de temps $[t_0, (1 + \alpha)t_0]$. Cette propriété est introduite par M. Ratner sous le nom de propriété H [66-67-68]. Pour montrer son utilité, nous allons décrire la classification par M. Ratner des quotients du flot horicyclique.

Soit ψ^t un flot mesurable préservant une probabilité μ sur un espace

mesurable (E, \mathcal{A}) (isomorphe à la tribu des boréliens d'un espace métrique séparable). On dit que ψ^t est un *quotient mesurable* du flot horicyclique de $\Gamma \backslash G$ s'il existe une application mesurable $F : \Gamma \backslash G \rightarrow E$ envoyant la mesure de Haar λ (normalisée pour être une probabilité) sur μ et telle que $F(x.h_+^t) = F(x).\psi^t$ pour λ -presque tout x et presque tout t . Nous supposons toujours que F est non trivial, c'est-à-dire que F n'est pas constante presque partout.

Soient $\Gamma \subset \Gamma'$ deux réseaux emboîtés de $G = \mathrm{SL}(2, \mathbf{R})$. Il est clair qu'alors Γ est d'indice fini dans Γ' et le revêtement fini $\Gamma \backslash G \rightarrow \Gamma' \backslash G$ montre que le flot horicyclique de $\Gamma' \backslash G$ est un quotient de celui de $\Gamma \backslash G$. Insistons sur le fait que pour un réseau Γ "générique", il n'existe pas de Γ' contenant strictement Γ .

THÉORÈME.— *Si (E, ψ^t) est un quotient mesurable du flot horicyclique de $\Gamma \backslash G$, alors (E, ψ^t) est isomorphe au flot horicyclique de $\Gamma' \backslash G$ avec $\Gamma \subset \Gamma'$.*

Esquisse de preuve : L'essentiel du théorème consiste à montrer que si $F : \Gamma \backslash G \rightarrow E$ est un quotient mesurable, presque toutes les fibres de F sont finies (et donc de même cardinal par ergodicité). Observons que les fibres de F partitionnent $\Gamma \backslash G$ et sont permutées par h_+^t . Une fibre qui est globalement préservée par h_+^τ produit une orbite périodique de ψ^t , de période τ . Le flot h_+^t étant ergodique, il en est de même pour ψ^t . Par conséquent, la réunion des orbites périodiques de ψ^t est de μ -mesure nulle, à moins que μ ne soit concentrée sur une seule orbite périodique (qui n'est pas un point fixe car F n'est pas constante). Dans les deux cas, on peut affirmer que, pour λ -presque tout x de $\Gamma \backslash G$, les points $F(x)$ et $F(x.h_+^t)$ sont différents pour t assez petit.

Plaçons-nous d'abord dans le cas où $\Gamma \backslash G$ est compact, où E est un espace métrique et où $F : \Gamma \backslash G \rightarrow E$ est continue.

Si une fibre $F^{-1}(e)$ est infinie, elle contient deux points distincts x, y arbitrairement proches. On peut alors leur appliquer la propriété H , car nous venons précisément de voir que y n'est pas de la forme $x.h_+^t$ avec t petit. Il existe donc $t_0 > 0$ tel que si $t \in [t_0, (1 + \alpha)t_0]$, les points

$x.h_+^t$ et $y.h_+^{t\pm C}$ sont très proches. Il en résulte alors, par continuité de F , que les points $F(x.h_+^t) = e.\psi^t$ et $F(y.h_+^{t\pm C}) = e.\psi^{t\pm C}$ sont voisins pour $t \in [t_0, (1 + \alpha)t_0]$. En prenant dans $F^{-1}(e)$ des paires de points distincts x, y de plus en plus proches, on trouve donc des morceaux de l'orbite de e du type $\{e.\psi^t \mid t \in [t_0, (1 + \alpha)t_0]\}$ avec t_0 de plus en plus grand, formés de points de plus en plus fixes par $\psi^{\pm C}$.

Choisissons C de telle sorte que la μ -mesure de l'ensemble des points fixes de $\psi^{\pm C}$ soit nulle. La μ -mesure de l'ensemble des points presque fixes de $\psi^{\pm C}$ est donc petite. Le théorème ergodique montre alors que, pour μ -presque tout e , le morceau d'orbite $\{e.\psi^t \mid t \in [0, T]\}$ avec T grand ne contient qu'une faible densité de points presque fixes par $\psi^{\pm C}$. Par conséquent, pour μ -presque tout e , la fibre $F^{-1}(e)$ ne peut être infinie et ceci établit le théorème.

Lorsque $F : \Gamma \backslash G \rightarrow E$ n'est plus supposé continu, on utilise le théorème de Lusin qui permet de trouver un compact K de $\Gamma \backslash G$ de grande mesure sur lequel F est continu. Le théorème ergodique montre qu'une orbite générique de h_+^t passe une large proportion de temps dans K et ceci permet d'adapter la preuve.

Lorsque $\Gamma \backslash G$ n'est plus supposé compact, on utilise le théorème cité en 2.4 selon lequel une orbite de h_+^t "passe la majorité de son temps" dans un compact de $\Gamma \backslash G$. Ceci permet d'adapter la propriété H et de généraliser la preuve précédente. ●

Dans le cas général d'un flot unipotent ϕ^t sur $\Gamma \backslash G$ où G est un groupe de Lie connexe quelconque, les quotients mesurables ont été classés par D. Witte [92] en utilisant les résultats et méthodes de M. Ratner. Le résultat est le suivant : les quotients sont isomorphes à l'action de ϕ^t sur $L_1 \backslash G / L_2$ où L_1 est un sous-groupe fermé contenant Γ et L_2 un sous-groupe fermé d'applications affines de $L_1 \backslash G$ commutant avec ϕ^t .

3.2. La dérive dans la direction du normalisateur

Fixons un sous-groupe discret Γ de $G = \text{SL}(2, \mathbf{R})$. Pour la première fois dans cet exposé, nous ne supposons pas que Γ est un réseau. Dans le paragraphe précédent, nous avons étudié la séparation entre les points $x.h_+^t$ et $y.h_+^t$. Nous étudions maintenant la séparation entre deux orbites dans

la direction transverse. Notons h_+^\perp le sous-espace de $sl(2, \mathbf{R})$ engendré par h_- et g et $p_\perp : sl(2, \mathbf{R}) \rightarrow h_+^\perp$ la projection parallèlement à h_+ . Munissons toujours $sl(2, \mathbf{R})$ de la norme sup par rapport à la base h_+, h_-, g . Il existe une constante $C_0 > 0$ ayant la propriété suivante. Soient x et $y = x \cdot \exp w$ avec $\|w\| < C_0$; on suppose que $\|p_\perp Ad(h_+^t)(w)\| < C_0$ pour tout $t \in [0, T]$. Alors il existe une fonction $\tau : [0, 2T] \rightarrow \mathbf{R}$ avec $\tau(0) \leq 2C_0$ et $0,99 \leq \frac{d\tau}{dt} \leq 1,01$ et une courbe $t \in [0, 2T] \mapsto w_t \in h_+^\perp$ avec $\|w_0\| < 2C_0$ telles que $y \cdot h_+^t = x \cdot h_+^{\tau(t)} \exp w_t$. Si $\exp w = \begin{pmatrix} a & c \\ b & d \end{pmatrix}$, un calcul élémentaire montre que $\exp w_t = \begin{pmatrix} (bt+d)^{-1} & 0 \\ b & bt+d \end{pmatrix}$. Si $\|w\|$ est petit, c'est-à-dire si $|b|$ et $|d-1|$ sont petits, w_t est donc très proche de $bh_- - (bt + (d-1))g$. On choisit C_0 assez petit de sorte que si v_1 et v_2 sont deux vecteurs de h_+^\perp de norme inférieure à $2C_0$, et si $\exp v_i = \begin{pmatrix} a_i & c_i \\ b_i & d_i \end{pmatrix}$ ($i = 1, 2$), alors $0,99\|v_1 - v_2\| \leq \text{Sup}(|a_1 - a_2|, |b_1 - b_2|, |c_1 - c_2|, |d_1 - d_2|) \leq 1,01\|v_1 - v_2\|$.

Fixons une constante $C < C_0$ et un vecteur w avec $\|w\| = \rho$ petit devant C . La formule donnant w_t montre d'abord que w_t est constant si $b = 0$. Ce n'est pas une surprise et correspond une fois de plus au fait que le flot géodésique g^s normalise h_+^t . Rappelons que nous notons B le sous-groupe de G engendré par h_+^t et g^s . Si $b \neq 0$, soit t_0 le premier instant positif tel que $\|w_{t_0}\| = C$. Ce temps t_0 est de l'ordre de $\frac{C}{|b|}$ et $w_{t_0} - Cg$ ou $w_{t_0} + Cg$ est de norme inférieure à 2ρ . De plus, $\|w_t - w_{t_0}\| \leq 2\eta \cdot C$ si $t \in [(1-\eta)t_0, (1+\eta)t_0]$.

En termes plus vagues, on a établi la propriété suivante. Si deux points x, y sont très proches et ne satisfont pas $y = x \cdot g$ avec g petit élément de B , leurs orbites doivent se séparer transversalement. Dans ce cas, lorsque au temps t_0 , cette séparation w_{t_0} atteint une norme appréciable C , le point $y \cdot h_+^{t_0}$ est très proche d'un point du type $z = (x \cdot h_+^{\tau_+}) \cdot g^{\pm C}$. Enfin les orbites des points $y \cdot h_+^{t_0}$ et z restent à une distance de l'ordre de C pendant un temps linéaire en t_0 .

Ici encore, ce type d'énoncé se généralise à tout flot unipotent. La "dérive transverse" entre deux orbites se fait alors dans la direction du *normalisateur* du flot. C'est la *propriété R*, l'une des idées fondamentales

des travaux de M. Ratner. Nous allons illustrer son usage dans un théorème que nous préciserons d'ailleurs plus loin, et dont la preuve n'est qu'un cas particulier de [71].

THÉORÈME.— *Soit Γ un sous-groupe discret de $SL(2, \mathbf{R})$ (qui n'est pas nécessairement un réseau). Soit μ une mesure de probabilité sur $\Gamma \backslash SL(2, \mathbf{R})$ invariante et ergodique pour le flot h_+^t . Si μ n'est pas concentrée sur une orbite périodique, alors μ est aussi invariante par le flot géodésique g^s .*

Preuve : Soit $C > 0$ petit et supposons que g^C ne préserve pas μ . Comme g^C normalise le flot h_+^t , la probabilité $g_*^C \mu$ est aussi invariante par h_+^t , différente de μ et ergodique. Il en résulte que μ et $g_*^C \mu$ sont étrangères, c'est-à-dire qu'il existe un borélien X tel que $\mu(X) = 1$ et $X.g^C \cap X = \emptyset$. Soit $K \subset X$ un compact de μ -mesure supérieure à $1 - \varepsilon$ (avec $\varepsilon > 0$ arbitrairement petit). Les deux compacts K et $K.g^C$ sont disjoints et il existe donc $\delta > 0$ tel que si $x = y.\exp w$ avec $\|w \pm Cg\| \leq \delta$, on ne peut avoir simultanément $x \in K$ et $y \in K$. Choisissons η suffisamment petit pour que $2\eta C \leq \frac{\delta}{2}$.

Pour μ -presque tout x , l'orbite de x passe une proportion de temps supérieure à $1 - \varepsilon$ dans K . On peut alors trouver $A_1 \subset \Gamma \backslash G$ de μ -mesure supérieure à $1 - \frac{\eta}{10}$ et $T_1 > 0$ tels que si $x \in A_1$ et $T > T_1$, on a

$$\frac{1}{T} m(\{t \in [0, T] \mid x.h_+^t \in K\}) \geq 1 - 2\varepsilon$$

(où m désigne toujours la mesure de Lebesgue).

On peut ensuite trouver $A \subset \Gamma \backslash G$ de mesure arbitrairement proche de 1 et $T_2 > 0$ tels que si $x \in A$ et $T > T_2$, on a :

$$\frac{1}{T} m(\{t \in [0, T] \mid x.h_+^t \in A_1\}) \geq 1 - \frac{\eta}{5}.$$

Considérons maintenant deux points x et $y = x.\exp w$ de A avec $\|r - w\| = \rho$. Nous allons d'abord montrer que, si ρ est assez petit, $\exp r - w$ est dans B , c'est-à-dire que $y = x.h_+^t g^s$ avec s et t petits. Supposons le contraire et soit t_0 le premier instant positif où $\|w_{t_0}\| = C$. On a $t_0 \geq 2T_1$ et $t_0 \geq 2\frac{T_2}{\eta}$ si ρ est assez petit.

Par définition de A , l'intervalle $[0, t_0]$ contient une proportion supérieure à $(1 - \frac{\eta}{5})$ de réels t tels que $x.h_+^t \in A_1$. Soit τ la fonction déjà introduite telle que $y.h_+^t = x.h_+^{\tau(t)} \exp w_t$. Comme $0,99 \leq \frac{d\tau}{dt} \leq 1,01$, l'intervalle $[0, t_0]$ contient une proportion supérieure à $1 - 1,01 \frac{\eta}{5} \geq 1 - \frac{\eta}{4}$ de réels t tels que $x.h_+^{\tau(t)} \in A_1$. Il existe donc $t_1 \in [(1 - \eta)t_0, t_0]$ tel que l'on ait simultanément $x.h_+^{t_1} \in A_1$ et $y.h_+^{\tau(t_1)} \in A_1$.

À son tour, la définition de A_1 montre qu'il existe $t_2 \in [(1 - \eta)t_0, (1 + \eta)t_0]$ tel que $\bar{x} = x.h_+^{t_2}$ et $\bar{y} = y.h_+^{\tau(t_2)}$ sont tous les deux dans K . Par ailleurs, nous savons que si $t \in [(1 - \eta)t_0, (1 + \eta)t_0]$, on a $\|w_t - w_{t_0}\| \leq 2\eta C \leq \frac{\delta}{2}$ et $\|w_{t_0} \pm Cg\| \leq 2\rho$. Si $\rho < \frac{\delta}{4}$, on a donc $\|w_{t_2} \pm Cg\| \leq \delta$.

Mais ceci contredit la définition de δ selon laquelle si \bar{x} et $\bar{y} = \exp w_{t_2}$ sont dans K , on a $\|w_{t_2} \pm Cg\| > \delta$. Nous avons donc bien établi que deux points x et y proches dans A sont du type $y = x.\exp w$ avec $\exp w \in B$ proche de l'identité. En particulier, l'une des orbites de B est de μ -mesure non nulle et l'ergodicité montre en fait que μ est concentrée sur une orbite de B dans $\Gamma \backslash G$.

Il reste à montrer que μ est concentrée sur une orbite périodique de h_+^t . Une orbite $x.B$ est du type B/Λ où Λ est un sous-groupe discret de B . Ces sous-groupes discrets de B sont aisés à décrire. Trois cas sont possibles :

i) Λ est trivial. Le flot h_+^t opérant sur B ne peut préserver de probabilité car toutes les orbites de h_+^t tendent vers l'infini dans B .

ii) Λ est engendré par $\begin{pmatrix} \exp(s_0) & t_0 \\ 0 & \exp(s_0) \end{pmatrix}$ avec $s_0 \neq 0$. L'espace homogène B/Λ est un cylindre hyperbolique. Là encore, toutes les orbites de h_+^t dans B/Λ tendent vers l'infini et il n'y a donc aucune probabilité invariante.

iii) Λ est engendré par $\begin{pmatrix} 1 & t_0 \\ 0 & 1 \end{pmatrix}$ avec $t_0 \neq 0$. L'espace homogène B/Λ est une pointe où toutes les orbites de h_+^t sont périodiques. Une probabilité invariante et ergodique est donc concentrée sur une orbite périodique.

Ceci achève la preuve du théorème. ●

3.3. Classification des probabilités invariantes : cas des réseaux

Le théorème qui suit est dû à S.G. Dani [13-16-17] mais la preuve que nous en présentons s'inspire d'un résultat plus général de M. Ratner [71].

THÉORÈME.— *Soit Γ un réseau de $G = \mathrm{SL}(2, \mathbf{R})$. La mesure de Haar sur $\Gamma \backslash G$ et les mesures concentrées sur les orbites périodiques sont (à un facteur près) les seules mesures finies sur $\Gamma \backslash G$ invariantes et ergodiques pour le flot horicyclique.*

Esquisse de preuve : Supposons d'abord $\Gamma \backslash G$ compact. Soit μ une mesure de probabilité sur $\Gamma \backslash G$ invariante et ergodique pour h_+^t . Soit λ la mesure de Haar, normalisée pour être une probabilité sur $\Gamma \backslash G$. Si τ_+ , τ_- et σ sont trois réels positifs, nous noterons $Q_{\tau_+, \tau_-, \sigma} = \{h_+^{t_+} h_-^{t_-} g^s \in \mathrm{SL}(2, \mathbf{R}) \mid |t_+| \leq \tau_+, |t_-| \leq \tau_-, |s| \leq \sigma\}$. La compacité de $\Gamma \backslash G$ permet de garantir l'existence de

$\rho > 0$ tel que si τ_+ , τ_- et σ sont inférieurs à ρ , la boîte $x.Q_{\tau_+, \tau_-, \sigma}$ est injectivement plongée dans $\Gamma \backslash G$ pour tout x de $\Gamma \backslash G$. Pour simplifier les notations, nous supposerons que $\rho = 2$ convient de sorte que tout arc d'orbite $\{x.h_+^t \mid t \in [0, 1]\}$ est contenu dans l'intérieur d'une boîte $x.Q_{2, 2, 2}$.

Soit $Q_0 = x_0.Q_{\tau_+, \tau_-, \sigma}$ l'une de ces boîtes (τ_+ , τ_- et $\sigma \leq 2$). Nous allons montrer que $\mu(Q_0) = \lambda(Q_0)$, ce qui établira que $\lambda = \mu$.

Comme h_+^t est ergodique pour λ , on a pour λ -presque tout x :

$$\lim \frac{1}{T} m(\{t \in [0, T] \mid x.h_+^t \in Q_0\}) = \lambda(Q_0).$$

Cette convergence est uniforme sur un ensemble de mesure presque totale. Il existe donc $X \subset \Gamma \backslash G$ avec $\lambda(X) \geq 1 - \varepsilon$ ($\varepsilon > 0$ étant arbitraire) et T_0 tels que si $x \in X$ et $T > T_0$, on a :

$$(1 - \varepsilon) \lambda(Q_0) \leq \frac{1}{T} m(\{t \in [0, T] \mid x.h_+^t \in Q_0\}) \leq (1 + \varepsilon) \lambda(Q_0).$$

Considérons maintenant un point y générique pour l'autre mesure μ , c'est-à-dire tel que :

$$\lim \frac{1}{T} m(\{t \in [0, T] \mid y.h_+^t \in Q_0\}) = \mu(Q_0).$$

Pour analyser les passages de h_+^t dans Q_0 , prenons l'image par le flot géodésique $g^{\text{Log } T}$ pour T grand et posons $y_T = y.g^{\text{Log } T}$. Ceci a pour effet de compresser l'arc $\{y.h_+^t \mid t \in [0, T]\}$ sur l'arc de longueur 1 $\{y_T.h_+^t \mid t \in [0, 1]\}$ et d'envoyer la boîte Q_0 sur $\{x_0.g^s h_-^{t_-} h_+^{t_+} \mid |s| \leq \sigma, |t_-| \leq T.\tau_-, |t_+| \leq \frac{\tau_+}{T}\}$. (Voir la figure 8.)

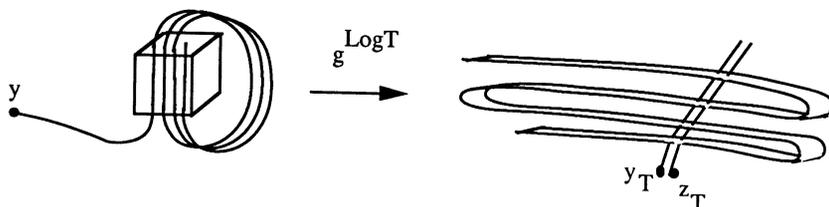


Figure 8

Le flot géodésique préserve λ et on a donc $\lambda(X.g^{\text{Log } T}) \geq 1 - \varepsilon$. Par conséquent, si ε est assez petit, une petite boule centrée sur y_T ne peut éviter $X.g^{\text{Log } T}$. Soit donc $z \in X$ tel que $z_T = z.g^{\text{Log } T}$ et y_T sont voisins. Les segments d'orbites $\{y_T.h_+^t \mid t \in [0, 1]\}$ et $\{z_T.h_+^t \mid t \in [0, 1]\}$ sont donc très proches et contenus dans une même boîte $y_T.Q_{2,2,2}$. La figure 8 montre que les proportions de temps passés par ces deux morceaux d'orbites dans $Q_0.g^{\text{Log } T}$ sont presque égales (et ce n'est pas difficile à justifier). Par définition de z et y , la première de ces deux proportions est proche de $\lambda(Q_0)$ et la seconde de $\mu(Q_0)$. En faisant tendre T vers $+\infty$ et ε vers 0, on établit que $\lambda(Q_0) = \mu(Q_0)$. Ceci termine la preuve lorsque $\Gamma \backslash G$ est compact.

Lorsque $\Gamma \backslash G$ est de volume fini mais non compact, l'existence de $\rho(= 2!)$ n'est plus garantie. La démonstration s'adapte cependant s'il existe une suite T_n tendant vers l'infini telle que y_{T_n} reste dans un compact de $\Gamma \backslash G$. Dans le cas contraire, $y_T = y.g^{\text{Log } T}$ quitte tout compact, c'est-à-dire que la géodésique correspondante tend vers une pointe. Nous avons vu que ce phénomène se produit précisément lorsque l'horicycle $y.h_+^t$ est périodique, c'est-à-dire si μ est concentrée sur une orbite périodique. ●

3.4. Classification des probabilités invariantes : le cas de co-volume infini

Comme nous l'avons déjà indiqué, M. Ratner obtient des résultats valables pour les sous-groupes discrets généraux [72], dont voici un cas particulier.

THÉORÈME.— *Soit Γ un sous-groupe discret de $G = \mathrm{SL}(2, \mathbf{R})$ de co-volume infini. Toute mesure de probabilité sur $\Gamma \backslash G$ invariante et ergodique pour le flot horicyclique est concentrée sur une orbite périodique.*

Preuve : Supposons qu'il existe une mesure de probabilité μ invariante et ergodique pour h_+^t sur $\Gamma \backslash G$, non concentrée sur une orbite périodique. Nous allons montrer que ceci est contradictoire avec le fait que le volume de $\Gamma \backslash G$ est infini.

Nous avons vu en 3.2 que μ est aussi invariante par le flot géodésique g^s . Nous affirmons d'abord que μ est aussi ergodique pour g^s . C'est une conséquence du phénomène de Mautner décrit en 1.2 : toute fonction de carré μ -intégrable qui est invariante par g^s est aussi invariante par h_+^t et donc constante μ -presque partout.

L'invariance de μ par le groupe de Lie B entraîne le fait suivant. Si $X \subset \Gamma \backslash G$ est de μ -mesure totale, alors pour μ -presque tout x , la mesure de Haar (de B) de l'ensemble $\{g \in B \mid x.g \in X\}$ est totale.

Fixons maintenant une fonction $F : \Gamma \backslash G \rightarrow \mathbf{R}$, continue à support compact, positive, d'intégrale 1 par rapport à μ . Posons pour $x \in \Gamma \backslash G$:

$$\overline{F}(x) = \liminf_{S \rightarrow \infty} \frac{1}{S} \int_0^S F(x.g^{-s}) ds.$$

D'après le théorème ergodique, nous savons que $\overline{F}(x) = \int F d\mu = 1$ pour x dans un ensemble X de μ -mesure 1. Par ailleurs, il est évident que \overline{F} est constante sur les orbites du flot horicyclique négatif h_-^t . Il résulte de l'observation précédente que la réunion des parties $X.h_-^t$ avec $t \in \mathbf{R}$ est de λ -mesure de Haar totale dans $\Gamma \backslash G$. Ainsi \overline{F} est λ -presque partout

constante, égale à 1. Le lemme de Fatou donne alors :

$$\begin{aligned} \lambda(\Gamma \backslash G) &= \int_{\Gamma \backslash G} 1 \, d\lambda \leq \liminf \int_{\Gamma \backslash G} \left(\frac{1}{S} \int_0^S F(x.g^{-s}) \, ds \right) d\lambda(g) \\ &= \int F \, d\lambda < \infty. \end{aligned}$$

C'est la contradiction cherchée qui achève la preuve du théorème. ●

3.5. Équirépartition des orbites

Le théorème suivant se réduit au théorème de S.G. Dani et J. Smillie dans le cas des réseaux.

THÉORÈME.— *Soit Γ un sous-groupe discret de $G = \mathrm{SL}(2, \mathbf{R})$, x un point non périodique pour le flot horicyclique h_+^t et $F : \Gamma \backslash G \rightarrow \mathbf{R}$ une fonction continue à support compact. Si T tend vers $+\infty$, la moyenne $\frac{1}{T} \int_0^T F(x.h_+^t) \, dt$ converge :*

- i) *vers $\int_{\Gamma \backslash G} F \, d\lambda$ si $\Gamma \backslash G$ est de volume fini (où λ désigne toujours la mesure de Haar normalisée pour être une probabilité)*
- ii) *vers 0 si $\Gamma \backslash G$ est de volume infini.*

Esquisse de preuve : Soit $\widehat{V} = V \cup \{\infty\}$ le compactifié d'Alexandrov de $\Gamma \backslash G$ et \widehat{h}_+^t l'extension de h_+^t à \widehat{V} . Soit μ_T la probabilité sur \widehat{V} définie par : $\int F \, d\mu_T = \frac{1}{T} \int_0^T F(x.\widehat{h}_+^t) \, dt$. Par compacité de l'espace des probabilités sur \widehat{V} , il existe une limite faible ν d'une suite μ_{T_n} (avec T_n tendant vers $+\infty$).

Plaçons-nous d'abord dans le cas où Γ est un réseau. Il s'agit alors de démontrer que $\nu = \lambda$ car ceci entraîne que μ_T converge faiblement vers λ . Nous affirmons d'abord que $\nu(\{\infty\}) = 0$. Ceci résulte du théorème (cité en 2.4) selon lequel un horicycle passe une faible proportion de temps hors d'un compact. Ainsi, ν est concentrée sur $\Gamma \backslash G$ et c'est, bien sûr, une probabilité invariante par h_+^t .

Nous avons classé les probabilités ergodiques invariantes en 3.3. Pour terminer la preuve, dans le cas des réseaux, il s'agit donc de montrer que

la ν -mesure de la réunion des orbites périodiques est nulle. Nous ne ferons qu'esquisser cette preuve qui utilise, une fois de plus, les propriétés des transvections. Soit γ une orbite périodique, $A_{\alpha,\beta}$ l'anneau réunion des orbites $\gamma.g^s$ avec $\alpha \leq s \leq \beta$, $B_{\alpha,\beta}^\tau$ l'épaississement $\{y.h_-^t \mid y \in A_{\alpha,\beta}, |t| \leq \tau\}$ et enfin $S_{\alpha,\beta}^\tau = \{y_0.g^s.h_-^t, \alpha \leq s \leq \beta, |t| \leq \tau\}$ où y_0 est l'un des points de γ . L'application de premier retour de h_+^t sur la transversale $S_{\alpha,\beta}^\tau$ est une transvection. Soient $[\alpha_1, \beta_1]$ et $[\alpha_2, \beta_2]$ deux intervalles disjoints de même longueur contenus dans $[\alpha, \beta]$. Il est clair que si τ est petit, un long morceau d'orbite passe à peu près le même temps dans $B_{\alpha_1,\beta_1}^\tau$ et $B_{\alpha_2,\beta_2}^\tau$ (voir figure 9). Faisant tendre τ vers 0, on obtient que $\nu(A_{\alpha_1,\beta_1}) = \nu(A_{\alpha_2,\beta_2})$. Par conséquent, tous les anneaux $A_{n,n+1}$ ($n \in \mathbf{Z}$) ont même ν -mesure et sont disjoints deux à deux. Puisque ν est de masse totale finie, nous avons bien établi que la ν -mesure de la réunion des orbites périodiques de h_+^t est nulle.

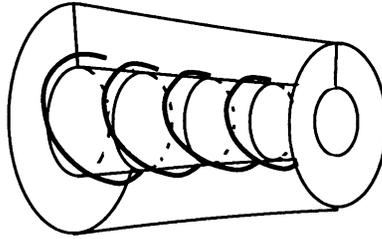


Figure 9

Lorsque le volume de $\Gamma \backslash G$ est infini, il s'agit au contraire de montrer que $\nu(\{\infty\}) = 1$. La preuve précédente montre encore que la ν -mesure de la réunion des orbites périodiques est nulle. La classification des probabilités invariantes faite en 3.4 montre donc bien que ν ne peut que se concentrer sur le point ∞ . ●

3.6. Quelques indications sur le cas général

La classification des mesures invariantes μ par un groupe unipotent H sur un espace homogène de volume fini $\Gamma \backslash G$ utilise toutes les méthodes décrites ci-dessus. Comme indiqué dans l'introduction, nous ne pouvons pas ici donner de détails sur le tour de force que représente cette preuve.

Nous nous contenterons de quelques observations.

L'ensemble des éléments g de G dont l'action à droite préserve μ est un sous-groupe fermé Λ de G et il s'agit de démontrer que μ est concentrée sur une orbite de Λ .

Si G est résoluble, $\Lambda \cap [G, G]$ est unipotent. C'est en appliquant une propriété R adaptée à ce sous-groupe que M. Ratner démontre son théorème dans le cas résoluble [71].

Si G est semi-simple et si ϕ^t est un sous-groupe unipotent à un paramètre, il se trouve qu'il est possible de plonger une copie de $sl(2, \mathbf{R})$ dans l'algèbre de Lie \mathcal{G} en envoyant le flot horicyclique sur ϕ^t . On dispose donc d'un "flot géodésique" g^s dans G qui permet une généralisation de la plupart des techniques décrites plus haut. Bien sûr, la situation est compliquée par le fait que g^s n'est pas de type Anosov car il peut posséder un centralisateur non trivial [72-73].

Pour déduire son résultat sur l'équirépartition et les adhérences des orbites, M. Ratner utilise une méthode dans l'esprit de 3.5 bien que beaucoup plus élaborée [74].

4. TROIS APPLICATIONS

Dans ce paragraphe, nous avons choisi trois exemples d'utilisation des résultats de M. Ratner.

4.1. Rigidité des translations unipotentes

Soit ϕ^t un sous-groupe unipotent à un paramètre d'un groupe de Lie simple G et Γ un réseau de G . Les invariants ergodiques classiques du flot ϕ^t sur $\Gamma \backslash G$ (équipé de la mesure de Haar) ne sont pas très intéressants : entropie nulle et spectre de Lebesgue dénombrable. Il est remarquable que les théorèmes de M. Ratner permettent de classer tous les isomorphismes mesurables entre ces flots. Le théorème suivant est dû à D. Witte [89] ; il fait suite au cas particulier de $SL(2, \mathbf{R})$, dû à M. Ratner [67-68]. Pour simplifier, nous nous limiterons au cas des groupes de Lie simples mais on consultera [89] et [73] pour les énoncés généraux.

THÉORÈME.— Soit ϕ_i^t un sous-groupe unipotent à un paramètre d'un groupe de Lie simplement connexe G_i ($i = 1, 2$) et Γ_i un réseau de G_i . Soit $F : \Gamma_1 \backslash G_1 \rightarrow \Gamma_2 \backslash G_2$ une conjugaison mesurable entre ϕ_1^t et ϕ_2^t envoyant la mesure de Haar λ_1 sur λ_2 . Alors, il existe un isomorphisme α de G_1 sur G_2 envoyant Γ_1 sur Γ_2 et ϕ_1^t sur ϕ_2^t tel que $F(\Gamma_1.g_1) = \Gamma_2.\alpha(g_2)$ pour λ_1 -presque tout g_1 .

On peut en fait dire beaucoup plus. Un lien (joining) entre ϕ_1^t et ϕ_2^t est, par définition, une mesure de probabilité μ sur le produit $\Gamma_1 \backslash G_1 \times \Gamma_2 \backslash G_2$ invariante par le flot diagonal $(\phi_1^t \times \phi_2^t)$ et telle que $\mu(\pi_i^{-1}(X)) = \lambda_i(X)$ où π_i est la projection sur le $i^{\text{ième}}$ facteur et où X est un borélien de $\Gamma_i \backslash G_i$ ($i = 1, 2$). La mesure produit des deux mesures de Haar $\lambda_1 \otimes \lambda_2$ est appelée le lien trivial. Si $F : \Gamma_1 \backslash G_1 \rightarrow \Gamma_2 \backslash G_2$ est une conjugaison mesurable entre ϕ_1^t et ϕ_2^t envoyant λ_1 sur λ_2 , il est clair que l'image de λ_1 par $\text{id} \times F : \Gamma_2 \backslash G_1 \rightarrow \Gamma_1 \backslash G_1 \times \Gamma_2 \backslash G_2$ est un lien non trivial. Il n'est pas difficile d'en déduire que le théorème précédent n'est qu'un cas particulier du suivant (où nous nous limitons au cas des groupes simples).

THÉORÈME.— On garde les notations du théorème précédent. Soit μ un lien ergodique non trivial entre ϕ_1^t et ϕ_2^t . Alors, il existe un isomorphisme $\alpha : G_1 \rightarrow G_2$ et un réseau Γ de G_1 contenu (avec indice fini) dans Γ_1 et $\alpha^{-1}(\Gamma_2)$ tels que μ soit l'image de la mesure de Haar de $\Gamma \backslash G_1$ par l'application $\text{id} \times \alpha : \Gamma \backslash G_1 \rightarrow \Gamma_1 \backslash G_1 \times \Gamma_2 \backslash G_2$.

Preuve : D'après le théorème de M. Ratner, un lien ergodique est une mesure homogène sur $\Gamma_1 \times \Gamma_2 \backslash G_1 \times G_2$, provenant de la mesure de Haar d'un sous-groupe fermé H de $G_1 \times G_2$. Puisqu'il s'agit d'un lien, la projection de H sur chaque facteur G_1 et G_2 est surjective. Par simplicité de G_1 et G_2 , on peut affirmer que H est le graphe d'un isomorphisme entre G_1 et G_2 , à moins que $H = G_1 \times G_2$. Ce dernier cas est bien sûr exclu si μ est non trivial. ●

Signalons deux généralisations de ces théorèmes. Tout d'abord, les conjugaisons mesurables entre flots horicycliques en courbure négative variable sont classés dans [33] (que l'on complétera par [61]). D'autre part, [47] (complété par [61]) étudie les cas d'équivalences topologiques entre flots

horicycliques, c'est-à-dire d'homéomorphismes envoyant orbites sur orbites sans en respecter les paramétrages. Pour des résultats analogues en dimension supérieure, on consultera [34] et [91].

4.2. Valeurs des formes quadratiques sur les points entiers

Le théorème d'approximation diophantienne cité dans l'introduction a été conjecturé par A. Oppenheim en 1929 pour $n \geq 5$ [57-58], puis par H. Davenport pour $n \geq 3$ en 1946 [29]. Il a été démontré pour $n \geq 21$ par B.J. Birch, H. Davenport et H. Ridout. On trouvera dans [3], [29], [43], [58-59], [87-88] d'autres résultats préliminaires, tous obtenus par des méthodes différentes de celles que nous allons décrire. Nous allons montrer ici comment les résultats de M. Ratner permettent d'obtenir un énoncé beaucoup plus fort. La déduction de cette forme forte à partir de la conjecture de M.S. Raghunathan est due à S.G. Dani et G.A. Margulis. Ces derniers démontrent d'ailleurs dans [24-25-26] le cas particulier $n = 3$ du théorème suivant. Il semble aussi que la conjecture de A. Oppenheim ait été l'une des motivations qui ont conduit M.S. Raghunathan à sa conjecture.

THÉORÈME.— *Soit Q une forme quadratique non dégénérée et indéfinie sur \mathbf{R}^n ($n \geq 3$) et $f(x, y) = \frac{1}{4}(Q(x + y) - Q(x - y))$ la forme bilinéaire associée. On suppose que Q n'est pas multiple d'une forme rationnelle. Soit a_{ij} , $1 \leq i \leq n - 1$, $1 \leq j \leq n - 1$ des réels choisis de telle sorte qu'il existe des vecteurs w_1, w_2, \dots, w_{n-1} de \mathbf{R}^n avec $a_{ij} = f(w_i, w_j)$. Alors, pour tout $\varepsilon > 0$, il existe v_1, \dots, v_{n-1} de \mathbf{Z}^n tels que, pour tout i, j : $|f(v_i, v_j) - a_{ij}| < \varepsilon$. On peut de plus supposer que le $(n - 1)$ -uplet (v_1, \dots, v_{n-1}) est primitif dans le sens où il peut être complété en une base de \mathbf{Z}^n .*

Commençons par exprimer ces problèmes diophantiens en termes de dynamique sur les espaces homogènes.

Fixons deux entiers positifs p, q tels que $p + q = n$ et notons $\Phi_{p,q}$ l'ensemble des formes quadratiques sur \mathbf{R}^n de signature (p, q) . Soit Q_0 la forme $x_1^2 + \dots + x_p^2 - x_{p+1}^2 - \dots - x_n^2$ et $O(p, q)$ son groupe orthogonal. Pour simplifier les notations, nous supposons n impair de sorte que toute forme non dégénérée est multiple d'une unique forme de discriminant $+1$. Le groupe $GL(n, \mathbf{R})$ opère transitivement à droite sur $\Phi_{p,q}$ par $Q.g(w) =$

$Q(g^{-1}(w))$ et $\Phi_{p,q}$ s'identifie donc à $GL(n, \mathbf{R})/O(p, q)$.

Deux formes Q et Q' de $\Phi_{p,q}$ sont *équivalentes* s'il existe $c \in \mathbf{R}$ et $g \in SL(n, \mathbf{Z})$ tels que, pour tout w de \mathbf{R}^n , on a : $Q'(w) = c^2 Q(g(w))$. Il est clair que Q et Q' prennent alors, à un facteur multiplicatif près, les mêmes valeurs sur \mathbf{Z}^n .

Nous obtenons alors une bijection naturelle entre :

- i) les *classes d'équivalence de formes* de $\Phi_{p,q}$ et,
- ii) les *doubles classes* de $SL(n, \mathbf{Z}) \backslash SL(n, \mathbf{R}) / SO(p, q)$ ou, de manière équivalente, les *orbites* de $SO(p, q)$ sur $SL(n, \mathbf{Z}) \backslash SL(n, \mathbf{R})$.

Le théorème de M. Ratner va permettre de décrire les adhérences des orbites de $SO(p, q)$. Il suffira alors de traduire ces résultats en termes de formes quadratiques.

Si $n \geq 3$ et $p, q \geq 1$, $SO(p, q)$ contient $SO(2, 1)$, localement isomorphe à $SL(2, \mathbf{R})$, et donc des éléments unipotents. Ces éléments unipotents engendrent la composante neutre $SO_0(p, q)$ de $SO(p, q)$, et on peut donc appliquer le théorème de M. Ratner à $SO_0(p, q)$. Comme $SO_0(p, q)$ est d'indice 2 dans $SO(p, q)$, on en déduit facilement que les adhérences des orbites de $SO(p, q)$ dans $SL(n, \mathbf{Z}) \backslash SL(n, \mathbf{R})$ sont les orbites par un sous-groupe fermé intermédiaire entre $SO(p, q)$ et $SL(n, \mathbf{R})$. Le lemme suivant est laissé en exercice au lecteur. Pour des énoncés analogues, mais beaucoup plus généraux, on peut consulter [30-31].

Lemme. — *Il n'existe aucun sous-groupe fermé strictement compris entre $SO(p, q)$ et $SL(n, \mathbf{R})$.*

On obtient donc immédiatement :

PROPOSITION. — *Une orbite de $SO(p, q)$ dans $SL(n, \mathbf{Z}) \backslash SL(n, \mathbf{R})$ est fermée ou dense.*

Il s'agit maintenant d'interpréter en termes de formes quadratiques la densité ou la fermeture des orbites correspondantes. C'est le but des deux propositions qui suivent et qui achèvent la preuve du théorème.

PROPOSITION. — *Soit Q une forme quadratique dont l'orbite associée est dense. Alors Q satisfait à la conclusion du théorème précédent.*

Preuve : La densité de l'orbite revient à la densité de $SL(n, \mathbf{Z}).SO(Q)$ dans $SL(n, \mathbf{R})$ (où, bien sûr, $SO(Q)$ désigne le groupe spécial orthogonal de Q). Soit (e_1, \dots, e_n) la base canonique de \mathbf{R}^n et g un élément de $SL(n, \mathbf{R})$ tel que $g(e_i) = w_i$ pour $i = 1, \dots, n-1$. Soit $h_\alpha \in SO(Q)$ et $\gamma_\alpha \in SL(n, \mathbf{Z})$ deux suites telles que $\gamma_\alpha \cdot h_\alpha$ converge, lorsque α tend vers $+\infty$, vers g^{-1} . Soit $v_i^\alpha = \gamma_\alpha^{-1}(e_i)$ ($i = 1, \dots, n-1$). Il est clair que $(v_1^\alpha, \dots, v_{n-1}^\alpha)$ est un $(n-1)$ -uplet primitif de \mathbf{Z}^n et que :

$$f(v_i^\alpha, v_j^\alpha) = f(h_\alpha^{-1} \gamma_\alpha^{-1}(e_i), h_\alpha^{-1} \gamma_\alpha^{-1}(e_j)) \xrightarrow{\alpha \rightarrow \infty} f(g(e_i), g(e_j)) = a_{ij}. \bullet$$

La seconde proposition affirme plus que nécessaire. Nous supposons toujours que p et q sont non nuls.

PROPOSITION.— *Une orbite de $SO(p, q)$ est fermée si et seulement si la (classe d'équivalence de) forme quadratique correspondante est multiple d'une forme rationnelle ; elle est compacte si et seulement si elle est multiple d'une forme rationnelle ne représentant pas zéro (dans \mathbf{Q}^n).*

Preuve : L'une des directions (celle que nous n'utilisons pas) est un cas particulier d'un théorème classique de A. Borel et Harish-Chandra [4]. Si Q est une forme rationnelle, $SL(n, \mathbf{Z}) \backslash SO(Q)$ est fermé dans $SL(n, \mathbf{Z}) \backslash SL(n, \mathbf{R})$ et compact si Q ne représente pas zéro.

Soit Q tel que $SL(n, \mathbf{Z}).SO(Q)$ est fermé dans $SL(n, \mathbf{R})$. Nous allons montrer que Q est multiple d'une forme rationnelle en suivant [27] où une partie de la preuve est attribuée à A. Borel. Soit $\Gamma = SL(n, \mathbf{Z}) \cap SO(Q)$. La méthode consiste à prouver que Q et ses multiples sont les seules formes dont le groupe orthogonal contient Γ . Ceci montrera que la droite engendrée par Q dans l'espace des formes quadratiques a l'équation entière $\{Q' \mid Q' \cdot \gamma = Q' \text{ pour tout } \gamma \text{ de } \Gamma\}$ et donc que Q est multiple d'une forme rationnelle. Fixons donc une forme Q' telle que $\Gamma \subset SO(Q')$. Soit ϕ^t un sous-groupe unipotent à un paramètre de $SO_0(Q)$ (où $SO_0(Q)$ est la composante neutre de $SO(Q)$). Si nous montrons que ϕ^t est constitué d'isométries de Q' , nous en déduirons que $SO_0(Q) \subset SO(Q')$, car ces sous-groupes engendrent $SO_0(Q)$. Il sera alors immédiat que $SO(Q) = SO(Q')$ et que Q et Q' sont proportionnelles.

Soit $w \in \mathbf{R}^n$. La fonction $F : g \in \text{SO}(Q) \mapsto Q'(g(w)) \in \mathbf{R}$ passe au quotient en $F : \Gamma \backslash \text{SO}(Q) \rightarrow \mathbf{R}$. Nous avons déjà cité le théorème de G.A. Margulis suivant lequel une orbite de ϕ^t ne peut tendre vers l'infini dans $\text{SL}(n, \mathbf{Z}) \backslash \text{SL}(n, \mathbf{R})$. Il existe donc une suite t_n tendant vers l'infini telle que $\text{SL}(n, \mathbf{Z}) \cdot \phi^{t_n}$ reste dans un compact K de $\text{SL}(n, \mathbf{Z}) \backslash \text{SL}(n, \mathbf{R})$. Puisque nous faisons l'hypothèse que l'orbite de Q , isomorphe à $\Gamma \backslash \text{SO}(Q)$ est fermée dans $\text{SL}(n, \mathbf{Z}) \backslash \text{SL}(n, \mathbf{R})$, elle intersecte le compact K sur un compact. La fonction $t \mapsto F(\phi^t) = Q'(\phi^t(w))$ prend donc des valeurs bornées sur cette suite t_n . Puisque, par ailleurs, cette fonction est polynomiale, elle est constante. Ainsi, on a $Q'(\phi^t(w)) = Q'(w)$ pour tout t , c'est-à-dire que ϕ^t est contenu dans $\text{SO}(Q')$ comme nous l'avions annoncé. Ceci achève la preuve de la proposition et donc du théorème •.

Lorsque $n = 2$, le théorème ne s'applique pas car $\text{SO}(1, 1) \subset \text{SL}(2, \mathbf{R})$ ne contient que des éléments semi-simples. La dynamique de $\text{SO}(1, 1)$ sur $\text{SL}(2, \mathbf{Z}) \backslash \text{SL}(2, \mathbf{R})$ est celle du flot géodésique de l'orbifold modulaire pour laquelle nous avons déjà dit qu'il existe des orbites qui ne sont ni périodiques ni denses. Concrètement, soit θ un nombre de type constant, c'est-à-dire vérifiant les inégalités suivantes, pour un certain $C > 0$:

$$\left| \theta - \frac{y}{x} \right| \geq C x^{-2} \text{ pour tout } (x, y) \in \mathbf{Z}^2 - (0, 0).$$

Par exemple, les nombres irrationnels quadratiques sur \mathbf{Q} sont de type constant. La forme quadratique indéfinie et non dégénérée Q définie sur \mathbf{R}^2 par $Q(x, y) = y^2 - \theta^2 x^2$ n'est pas multiple d'une forme rationnelle si θ^2 est irrationnel. Il est clair cependant que $Q(x, y) = x^2 \left(\frac{y}{x} - \theta \right) \left(\frac{y}{x} + \theta \right)$ ne peut être arbitrairement petit si x et y sont entiers non nuls.

Nous suivons maintenant [54] pour décrire plus précisément les valeurs des formes quadratiques indéfinies et non dégénérées Q sur les points entiers \mathbf{Z}^n . Pour une telle forme, notons $\nu(Q) = \inf \{ |Q(v)| \mid v \in \mathbf{Z}^n - \{0\} \}$, $\text{disc}(Q)$ le discriminant de Q et $M(Q) = \nu(Q) \|\text{disc}(Q)\|^{-\frac{1}{n}}$, de sorte que $M(Q)$ est invariant par équivalence de formes. Soit M_n l'ensemble des valeurs prises par $M(Q)$ lorsque Q parcourt les formes non dégénérées et indéfinies de \mathbf{R}^n . Pour tout $n \geq 2$, c'est un compact de \mathbf{R} .

L'étude de M_2 est fascinante. Voir en particulier [79] et le chapitre de [10] consacré à l'intersection de M_2 avec $] \frac{4}{9}, \frac{4}{5}]$. On sait aussi que M_2 contient un intervalle $[0, \varepsilon]$ ([10], [78]).

Il résulte du théorème de G.A. Margulis que si Q n'est pas multiple d'une forme rationnelle, on a $M(Q) = 0$. Si Q est rationnelle et si $n \geq 5$, le fameux théorème de Meyer affirme que Q représente zéro non trivialement et donc que $M(Q) = 0$ [11]. Ainsi, $M_n = \{0\}$ si $n \geq 5$. Le théorème suivant a été démontré par J.W.S. Cassels et H.P.F. Swinnerton-Dyer conditionnellement à la conjecture de A. Oppenheim [12] et, sous une forme plus forte par L. Vulakh [86]. Nous en présentons ici une preuve légèrement différente de celle de G.A. Margulis [54].

THÉORÈME.— *Si $n = 3$ ou 4 et $\varepsilon > 0$, il n'existe qu'un nombre fini de classes d'équivalence de formes quadratiques non dégénérées et indéfinies Q sur \mathbf{R}^n telles que $M(Q) \geq \varepsilon$.*

Esquisse de preuve : Nous savons qu'une forme indéfinie Q telle que $M(Q) > 0$ est multiple d'une forme rationnelle ne représentant pas zéro et correspond donc à une orbite compacte pour l'action de $SO(p, q)$ sur $SL(n, \mathbf{Z}) \backslash SL(n, \mathbf{R})$. Supposons donc par l'absurde qu'il existe une infinité de formes Q_i non équivalentes deux à deux, de même signature (p, q) , telles que $M(Q_i) \geq \varepsilon$. On obtient ainsi une infinité d'orbites compactes connexes O_i pour l'action de la composante neutre $SO_0(p, q)$ de $SO(p, q)$ (qui est d'indice 2). On vérifie (critère de Mahler) que la condition $M(Q_i) \geq \varepsilon$ signifie que tous les compacts O_i sont contenus dans un même compact de $SL(n, \mathbf{Z}) \backslash SL(n, \mathbf{R})$. Soit O_∞ une limite (au sens de Hausdorff) des compacts O_i . Il est clair que O_∞ est un compact connexe, réunion d'orbites de $SO_0(p, q)$. Comme chaque orbite est dense ou fermée, O_∞ est nécessairement une réunion d'orbites compactes, en nombre dénombrable car il n'existe qu'un nombre dénombrable de formes rationnelles. Il est alors facile de se convaincre que la connexité de O_∞ montre que O_∞ se réduit à une seule orbite compacte. Nous devons donc montrer qu'une suite d'orbites compactes différentes ne peut converger vers une orbite compacte O_∞ .

Soit \mathcal{N} un supplémentaire de l'algèbre de Lie $so(p, q)$ dans $sl(n, \mathbf{R})$ in-

variant par la représentation adjointe de $\mathrm{SO}_0(p, q)$. Ceci détermine une trivialisatation $\mathrm{SO}_0(p, q)$ -équivariante de fibré normal aux orbites de $\mathrm{SO}_0(p, q)$. Si i est assez grand, O_i reste dans un voisinage tubulaire de O_∞ et il existe donc une petite fonction $u : O_i \rightarrow \mathcal{N}$ telle que $O_\infty = \{x. \exp u(x) \mid x \in O_i\}$. Soit $t \in [0, 1]$. Alors, il est clair que $O_\infty^t = \{x. \exp tu(x) \mid x \in O_i\}$ est aussi une orbite compacte de $\mathrm{SO}_0(p, q)$. Ceci contredit la dénombrabilité de l'ensemble des orbites compactes et achève la preuve. •

4.3. Laminations géodésiques

Soit H^n l'espace hyperbolique de dimension $n \geq 2$. Il est bien connu que les sous-variétés totalement géodésiques et complètes sont isométriques à H^p pour $0 \leq p \leq n$. Soit $\mathrm{Rep}(H^n)$ le fibré des repères orthonormés directs de fibré tangent à H^n ; c'est un espace principal homogène sous l'action du groupe des isométries directes de H^n , isomorphe à $\mathrm{SO}_0(n, 1)$.

Soit maintenant V une variété riemannienne complète, à courbure -1 , et de volume fini. Alors V est isométrique au quotient de H^n par un réseau Γ de $\mathrm{SO}_0(n, 1)$. Le quotient $\Gamma \backslash \mathrm{SO}_0(n, 1)$ est alors identifié au fibré $\mathrm{Rep}(V)$ des repères de V . L'action de $\mathrm{SO}_0(p, 1) \times \mathrm{SO}(n-p) \subset \mathrm{SO}_0(n, 1)$ sur $\Gamma \backslash \mathrm{SO}_0(n, 1) \simeq \mathrm{Rep}(V)$ a pour orbites des sous-variétés dont les projections dans V sont précisément les images des immersions isométriques de H^p dans V . Nous sommes donc dans les conditions d'application du théorème de M. Ratner dès que $p > 1$. Nous ne donnerons que le résultat obtenu par N. Shah avec cette méthode [80] :

THÉORÈME.— *Soit V une variété riemannienne complète de dimension n , de volume fini et à courbure -1 . Soit $i : H^p \rightarrow V$ une immersion isométrique avec $p > 1$. Alors l'adhérence de $i(H^p)$ dans V est une sous-variété W de V totalement géodésique, de dimension $\geq p$. De plus, les p -repères orthonormés tangents à $i(H^p)$ sont denses dans les p -repères orthonormés de V tangents à W .*

Rappelons qu'une *lamination géodésique* \mathcal{L} de dimension p de V est un fermé qui est une réunion disjointe de sous-variétés de dimension p , injectivement immergées dans V , complètes, et totalement géodésiques, appelées feuilles. On suppose aussi que les p -repères tangents aux feuilles de \mathcal{L} for-

ment un fermé dans le fibré des p -repères de V . Une lamination est triviale si toutes ses feuilles sont fermées (elles sont alors en nombre fini). Le rôle fondamental des laminations géodésiques de dimension 1 des surfaces a été révélé par W. Thurston. Il contraste avec le corollaire suivant :

COROLLAIRE.— *Toute lamination géodésique de dimension $p \geq 2$ d'une variété hyperbolique V de volume fini est triviale. En particulier, il n'existe pas, pour $1 < p < \dim V$, de feuilletage de classe C^0 dont toutes les feuilles sont totalement géodésiques.*

On consultera [93] pour des généralisations partielles au cas de la courbure variable. La question de l'existence de feuilletages de dimension 1 et de classe C^0 dont les feuilles sont des géodésiques de V est ouverte bien que l'on sache qu'un tel feuilletage ne peut être de classe C^1 .

BIBLIOGRAPHIE

- [1] D.V. ANOSOV - *Geodesic flows on compact Riemannian manifolds of negative curvature*, Trudy Mat. Inst. Steklova, **90** (1967).
- [2] L. AUSLANDER, L. GREEN and F. HAHN - *Flows on homogeneous spaces*, Ann. of Math. Studies, Princeton **53** (1963).
- [3] R.C. BAKER, H.P. SCHLICKEWEY - *Indefinite quadratic forms*, Proc. London Math. Soc. **54** (1987).
- [4] A. BOREL - *Introduction aux groupes arithmétiques*, Hermann, Paris (1969).
- [5] A. BOREL et G. PRASAD - *Valeurs de formes quadratiques aux points entiers*, C.R. Acad. Sci. Paris, Série I **307** (1988), 217-220.
- [6] R. BOWEN - *Periodic orbits for hyperbolic flows*, Amer. J. Math (1972), 1-30.
- [7] R. BOWEN - *Weak mixing and unique ergodicity on homogeneous spaces*, Israel J. Math **23** (1976), 267-273.
- [8] R. BOWEN and B. MARCUS - *Unique ergodicity of horospherical foliations*, Israel J. Math **26** (1977), 43-67.
- [9] J. BREZIN and C. MOORE - *Flows on homogeneous spaces : a new*

- look, Amer. J. Math. **103** (1981), 571-613.
- [10] J.W.S. CASSELS - *An introduction to diophantine approximations*, Cambridge University Press (1957).
- [11] J.W.S. CASSELS - *Rational quadratic forms*, Academic Press, London-New York (1978).
- [12] J.W.S. CASSELS and H.P.F. SWINNERTON-DYER - *On the product of three homogeneous linear forms and indefinite ternary quadratic forms*, Philos. Trans. Roy. Soc. London, Ser. A, **248** (1955), 73-96.
- [13] S.G. DANI - *Invariant measures of horospherical flows on non compact homogeneous spaces*, Invent. Math. **47** (1978), 101-138.
- [14] S.G. DANI - *On invariant measures, minimal sets and a lemma of Margulis*, Invent. Math. **51** (1979), 239-260.
- [15] S.G. DANI - *Dynamics of the horocycle flow*, bullet. A.M.S. **3** (1980), 1037-1039.
- [16] S.G. DANI - *Invariant measures and minimal sets of horospherical flows*, Invent. Math. **64** (1981), 357-385.
- [17] S.G. DANI - *On uniformly distributed orbits of certain horocyclic flows*, Ergodic theory and dynamical systems **2** (1982), 139-158.
- [18] S.G. DANI - *On orbits of unipotent flows on homogeneous spaces*, Ergodic theory and dynamical systems **4** (1984), 25-34.
- [19] S.G. DANI - *Dynamics of flows on homogeneous spaces : a survey*, Proceed. of Colloquio de sistemas dinamicos (Guanajuato, 1983), Aportaciones Mat. **1** (1985), 1-30.
- [20] S.G. DANI - *Divergent trajectories of flows on homogeneous spaces and diophantine approximation*, J. Reine Angew. Math. **359** (1985), 55-89.
- [21] S.G. DANI - *Orbits of horospherical flows*, Duke Math. Journal **53** (1986), 177-188.
- [22] S.G. DANI - *On orbits of unipotent flows on homogeneous spaces II*, Ergodic theory and dynamical systems **6** (1986), 167-182.
- [23] S.G. DANI - *Dense orbits of horospherical flows*, dynamical systems and ergodic theory, Banach center publications, volume **23**, Polish Scientific Publishers, Warsaw (1989).

- [24] S.G. DANI and G.A. MARGULIS - *Values of quadratic forms at primitive integral points*, C.R. Acad. Sci. Paris, Série I, **308** (1989), 199-203.
- [25] S.G. DANI and G.A. MARGULIS - *Values of quadratic forms at primitive integral points*, Invent. Math. **98** (1989), 405-424.
- [26] S.G. DANI and G.A. MARGULIS - *Orbit closures of generic unipotent flows on homogeneous spaces of $SL(3, \mathbf{R})$* , Math. Ann. **286** (1990), 101-128.
- [27] S.G. DANI and G.A. MARGULIS - *Values of quadratic forms at integral points : an elementary approach*, L'enseignement mathématique **36** (1990), 143-174.
- [28] S.G. DANI and J. SMILLIE - *Uniform distribution of horocyclic orbits for Fuchsian groups*, Duke Math. Journal **51** (1984), 185-194.
- [29] H. DAVENPORT and H. HEILBRONN - *On indefinite quadratic in five variables*, J. London Math. Soc. **21** (1946), 185-193.
- [30] E.B. DYNKIN - *Semi-simple subalgebras of semi-simple Lie algebras*, A.M.S. Transl. (Series 2) **6** (1957), 111-244
- [31] E.B. DYNKIN - *Maximal subgroups of the classical groups*, A.M.S. Transl. (Series 2) **6** (1957), 245-378.
- [32] R. ELLIS and W. PERRIZO - *Unique ergodicity of flows on homogeneous spaces*, Israel J. Math. **29** (1978), 276-284.
- [33] J. FELDMAN and D. ORNSTEIN - *Semirigidity of horocycle flows over compact surfaces of variable negative curvature*, Ergodic theory and dynamical systems **7** (1987), 49-72.
- [34] L. FLAMINO - *An extension of Ratner's rigidity theorem to n -dimensional hyperbolic space*, Ergodic theory and dynamical systems **7** (1987), 73-92.
- [35] H. FURSTENBERG - *Strict ergodicity and transformations of the torus*, Amer. J. Math. **83** (1961), 573-601.
- [36] H. FURSTENBERG - *The structure of distal flows*, Amer. J. Math. **85** (1963), 477-515.
- [37] H. FURSTENBERG - *The unique ergodicity of the horocyclic flow*, in Recent Advances in Topical Dynamics, 95-115, Springer (1972).
- [38] G. HEDLUND - *Fuchsian groups and transitive horocycles*, Duke Math. Journal **2** (1936), 530-542.

- [39] G. HEDLUND - *Dynamics of geodesic flows*, bullet. of the A.M.S. **45** (1939), 241-260.
- [40] G. HEDLUND - *Fuchsian groups and mixtures*, Ann. of Math. **40** (1939), 370-383.
- [41] E. HOPF - *Fuchsian groups and ergodic theory*, Trans. A.M.S. **39** (1936), 299-314.
- [42] E. HOPF - *Statistik der Geodätische Linien Manigfaltigkeiten Negativer Krümmung*, Ber. Voch. Sachs. Akad. Wiss. Leipzig **91** (1939), 261-304.
- [43] H. IWANIEC - *On indefinite quadratic forms in four variables*, Acta Arith. **33** (1977), 209-229.
- [44] E. LESIGNE - *Théorèmes ergodiques pour une translation sur une nil-variété*, Ergodic theory and dynamical systems **9** (1989), 115-126.
- [45] E. LESIGNE - *Sur une nil-variété, les parties minimales associées à une translation sont uniquement ergodiques*, Ergodic theory and dynamical systems **11** (1991), 379-391.
- [46] B. MARCUS - *Unique ergodicity of the horocyclic flow : variable curvature case*, Israel J. of Math. **21** (1975), 133-144.
- [47] B. MARCUS - *Topological conjugacy of horocycle flows*, Amer. J. Math. **105** (1983), 623-632.
- [48] G.A. MARGULIS - *On the action of unipotent groups in the space of lattices*, Proc. of the summer school on group representations, Bolyai Janos Math. Soc., Budapest (1971), 365-370.
- [49] G.A. MARGULIS - *Lie groups and ergodic theory*, in Avramov, L.L. (eds) Algebra, some current trends, Proceed. Varna 1986, Springer Lecture Notes in Math. **1352**, 130-146.
- [50] G.A. MARGULIS - *Formes quadratiques indéfinies et flots unipotents sur les espaces homogènes*, C.R. Acad. Sci., Série I, **304** (1987), 249-253.
- [51] G.A. MARGULIS - *Indefinite quadratic forms and unipotent flows on homogeneous spaces*, Banach center publications, volume **23**, Polish Scientific Publishers, Warsaw (1989).
- [52] G.A. MARGULIS - *Discrete subgroups and ergodic theory*, Symposium in honour of A. Selberg, Number theory, trace formulas and discrete

- groups, Academic Press (1989), 377-398.
- [53] G.A. MARGULIS - *Compactnes of minimal closed invariant sets of actions of unipotent groups*, Geometria Dedicata **37** (1991), 1-9.
- [54] G.A. MARGULIS - *Orbits of group actions and values of quadratic forms at integral points*, Preprint IHES.
- [55] C. MOORE - *The Mautner phenomenon for general unitary representations*, Pacific J. Math. **86** (1980), 155-169.
- [56] M. MORSE - *Recurrent geodesics on a surface of negative curvature*, Trans. Amer. Math. Soc. **22** (1921), 84-100
- [57] A. OPPENHEIM - *The minima of indefinite quaternary quadratic forms of signature 0*, Proc. Nat. Acad. Sci. USA **15** (1929), 724-727.
- [58] A. OPPENHEIM - *The minima of indefinite quaternary quadratic forms*, Ann. of Math. **32** (1931), 271-298.
- [59] A. OPPENHEIM - *Values of quadratic forms*, I, II, Quart. J. Math., Oxford Ser. (2) **4** (1953), 54-59, 60-66.
- [60] A. OPPENHEIM - *Values of quadratic forms*, III, Monatsh. Math. Phys. **57** (1953), 97-101.
- [61] J.-P. OTAL - *Le spectre marqué des longueurs des surfaces à courbure négative*, Ann. of Math. **131** (1990), 151-162.
- [62] P. PANSU - *Le flot géodésique des variétés riemanniennes à courbure négative*, Séminaire Bourbaki, exposé n° 738, février 1991.
- [63] W. PARRY - *Ergodic properties of affine transformations and flows on nilmanifolds*, Amer. J. Math. **91** (1969), 757-771.
- [64] W. PARRY - *Metric classification of ergodic nilflows*, Amer. J. Math. **93** (1971), 819-828.
- [65] J. PLANTE - *Anosov flows*, Amer. J. Math. (1972), 729-754.
- [66] M. RATNER - *Factors of horocyclic flows*, Ergodic theory and dynamical systems **2** (1982), 465-489.
- [67] M. RATNER - *Rigidity of horocycle flows*, Ann. of Math. **115** (1982), 597-614.
- [68] M. RATNER - *Horocyclic flows : joinings and rigidity of products*, Ann. of Math. **118** (1983), 277-313.

- [69] M. RATNER - *Ergodic theory in hyperbolic space*, Contemp. Math. **26** (1984), 309-334.
- [70] M. RATNER - *Invariant measures for unipotent translations on homogeneous spaces*, Proc. Nat. Acad. Sci. USA **87** (1990), 4309-4311.
- [71] M. RATNER - *Strict measure rigidity for unipotent subgroups of solvable groups*, Invent. Math. **101** (1990), 449-482.
- [72] M. RATNER - *On measure rigidity of unipotent subgroups of semisimple groups*, Acta Mathematica **165** (1990), 229-309.
- [73] M. RATNER - *On Raghunathan's measure conjecture*, à paraître dans Ann. of Maths.
- [74] M. RATNER - *Raghunathan's topological conjecture and distributions of unipotent flows*, Duke Math. J, **63** (1991), 235-280.
- [75] M. RATNER - *Distribution rigidity for unipotent actions on homogeneous spaces*, bulletin A.M.S. **24** (1991), 321-325.
- [76] M. REES - *Tangentially distal flows*, Israel J. of Math. **15** (1980), 9-31.
- [77] P. SARNAK - *Asymptotic behaviour of periodic orbits of the horocycle flow and Eisenstein series*, Comm. on Pure and Applied Math. **34** (1981), 719-739.
- [78] W.M. SCHMIDT - *Diophantine approximations*, Lecture Notes in Math. **785**, Springer-Verlag.
- [79] C. SERIES - *The geometry of Markov numbers*, Math. Intelligencer **7** (1985), 20-29.
- [80] N.A. SHAH - *Closures of totally geodesic immersions in manifolds of constant negative curvature* in : Proceed. Group theory from a geometrical viewpoint, Trieste 1990, ed. É. Ghys, A. Haefliger, A. Verjovsky, World Scientific (1991).
- [81] K. SIGMUND - *On the space of invariant measures for hyperbolic flows*, Amer. J. Math. **94** (1972), 31-37.
- [82] J.-C. SIKORAV - *Valeurs des formes quadratiques indéfinies irrationnelles (d'après G.A. Margulis)*, Séminaire de théorie des nombres, Paris, 86-87, Ed. C Goldstein, Progress in Math. **81**, 307-315, Birkhauser.

- [83] A.N. STARKOV - *Structure of orbits of homogeneous flows and the Raghunathan conjecture*, Russian Math. Surveys **45** (1990), 227-228.
- [84] W. VEECH - *Unique ergodicity of horospherical flows*, Amer. J. Math. **99** (1977), 827-859.
- [85] A. VERJOVSKY - *Arithmetic, geometry and dynamics in the unit tangent bundle of the modular orbifold*, preprint I.C.T.P.
- [86] L.Ya. VULAKH - *On minima of rational indefinite quadratic forms*, J. Number Theory **21** (1985), 275-285.
- [87] G.L. WATSON - *On indefinite quadratic forms in five variables*, Proc. London Math. Soc. (3) **3** (1953), 170-181.
- [88] G.L. WATSON - *On indefinite quadratic forms in three or four variables*, J. London Math. Soc. **28** (1953), 239-242.
- [89] D. WITTE - *Rigidity of some translations on homogeneous spaces*, Invent. Math. **81** (1985), 1-27.
- [90] D. WITTE - *Zero entropy affine maps on homogeneous spaces*, Amer. J. Math. **109** (1987), 927-961.
- [91] D. WITTE - *Rigidity of horospherical foliations*, Ergodic theory and dynamical systems **9** (1989), 191-205.
- [92] D. WITTE - *Measurable quotients of unipotent translations on homogeneous spaces*, à paraître.
- [93] A. ZEGHIB - *Laminations et hypersurfaces géodésiques des variétés hyperboliques*, Ann. Scient. Éc. Norm. Sup. **24** (1991), 171-188.
- [94] A. ZEGHIB - *Systèmes dynamiques autonomes. Partie I : Ensembles invariants des flots d'Anosov algébriques*, Prépublication E.N.S. Lyon n° 47 (1991).

Étienne GHYS

École Normale Supérieure de Lyon

Unité de Mathématiques Pures et Appliquées

U.M.R. 128 du C.N.R.S.

46, allée d'Italie

F-69364 LYON CEDEX 07

Astérisque

NIGEL HITCHIN

Hyperkähler manifolds

Astérisque, tome 206 (1992), Séminaire Bourbaki,
exp. n° 748, p. 137-166

http://www.numdam.org/item?id=SB_1991-1992__34__137_0

© Société mathématique de France, 1992, tous droits réservés.

L'accès aux archives de la collection « Astérisque » (<http://smf4.emath.fr/Publications/Asterisque/>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

HYPERKÄHLER MANIFOLDS

by Nigel HITCHIN

1. INTRODUCTION

1.1. "I then and there felt the galvanic circuit close; and the sparks which fell from it were the fundamental equations between i , j and k exactly as I have used them ever since" [H].

Hamilton's conviction that the quaternions should play as fundamental a rôle as the complex numbers in mathematics and physics was never realized in his day. There exists now, however, a rich theory of manifolds based on the algebra of quaternions which goes some way towards vindicating his belief. These manifolds moreover arise naturally within the context of the equations of mathematical physics. They are the *hyperkähler manifolds*.

The "fundamental equations between i , j and k " which, on that October day in 1843, Hamilton carved with such enthusiasm on Brougham Bridge were of course

$$i^2 = j^2 = k^2 = ijk = -1$$

A hyperkähler manifold is a manifold (necessarily of dimension a multiple of four) which admits an action on tangent vectors of the same i , j and k in a manner which is compatible with a metric. More precisely,

DEFINITION. *A hyperkähler manifold is a Riemannian manifold with three covariant constant orthogonal automorphisms I , J and K of the tangent bundle which satisfy the quaternionic identities $I^2 = J^2 = K^2 = IJK = -1$*

1.2. Recall that a Riemannian manifold which has just one such automorphism is called a Kähler manifold. The name “hyperkähler”, which originated with E. Calabi [Ca], is a fair description – the metric is Kählerian for several complex structures – even though it does recall Grassmann’s “hyper-complex numbers” rather than Hamilton’s quaternions. There is, however, an essential difference between Kähler and hyperkähler manifolds. A Kähler metric on a given complex manifold can be modified to another one simply by adding a hermitian form $\partial\bar{\partial}f$ for an arbitrary sufficiently small C^∞ function f . Thus the space of Kähler metrics is infinite dimensional. It is moreover easy to find examples of Kähler manifolds. Any complex submanifold of CP^n inherits a Kähler metric and so simply writing down algebraic equations for a projective variety gives a vast number of examples.

By contrast, hyperkähler metrics are much more rigid. On a compact manifold, if one such metric exists, then up to isometry there is only a finite dimensional space of them. Nor is it easy to find examples. Certainly we will never find them as quaternionic submanifolds of the quaternionic projective space HP^n [Gr].

1.3. The concept of a hyperkähler manifold arose first in 1955 with M. Berger’s classification of the holonomy groups of Riemannian manifolds. On a hyperkähler manifold, parallel translation preserves I , J and K (since they are covariant constant) and so the holonomy group is contained in both the orthogonal group $O(4n)$ and the group $GL(n, \mathbf{H})$ of quaternionic invertible matrices (i.e. those linear transformations which commute with right multiplication by i , j and k). The maximal such intersection is $Sp(n)$, the group of $n \times n$ quaternionic unitary matrices. This group appeared in Berger’s list.

1.4. The group $Sp(n)$ is also an intersection of $U(2n)$ and $Sp(2n, \mathbf{C})$, the linear transformations of \mathbf{C}^{2n} which preserve a non-degenerate skew form. Thus a hyperkähler manifold is naturally a complex manifold with a holomorphic symplectic form. One can see this explicitly by taking the three Kähler two-forms

$$\omega_1(X, Y) = g(IX, Y) \quad \omega_2(X, Y) = g(JX, Y) \quad \omega_3(X, Y) = g(KX, Y)$$

defined for the complex structures I , J and K . With respect to the complex structure I , the complex form $\omega_c = \omega_2 + i\omega_3$ is non-degenerate and covariant constant, hence closed and holomorphic.

This point of view provides guidance in the search for examples of hyperkähler manifolds, and elucidates the sort of differential equation which needs to be solved. In the first place the holomorphic volume form ω_c^n must for a hyperkähler manifold give a covariant constant trivialization of the canonical line bundle. The curvature of this bundle for any Kähler metric is the Ricci form and so a hyperkähler metric has in particular vanishing Ricci tensor. In the lowest dimension – four – this means that such metrics satisfy the Riemannian version of the Einstein vacuum equations.

Given a *compact* Kähler manifold with holomorphically trivial canonical bundle, the Calabi-Yau theorem [Y] provides the existence of a Kähler metric with vanishing Ricci tensor. Furthermore, a much older theorem due to S. Bochner [Bo] shows that any holomorphic form on a compact Kähler manifold with zero Ricci tensor is covariant constant. Thus for every compact Kähler manifold with a holomorphic symplectic form, an application of these two theorems yields a hyperkähler metric on the same manifold. This satisfactory state of affairs can be used to prove the existence of hyperkähler metrics on many examples of complex manifolds. The most fundamental is the K3 surface – the only non-trivial example in 4 real dimensions. In higher dimensions, the Hilbert scheme of zero cycles on a K3-surface or a 2-dimensional complex torus yields a natural class of holomorphic symplectic manifolds and hence hyperkähler metrics [Bea].

1.5. In this exposition, however, we shall seek something more than existence. We should like to *construct* solutions in a more explicit manner, in order to gain a better understanding of hyperkähler manifolds and to experience the richness of their geometry. The ability to do this is a relatively recent phenomenon. Indeed, twenty years ago it was hardly possible to write down any non-trivial Riemannian metric with zero Ricci tensor.

There are two main routes to constructing hyperkähler metrics: (a) twistor theory, (b) hyperkähler quotients.

The twistor approach is based on R. Penrose's original work in rela-

tivity [P]. It provides an *encoding* of the data for such a metric in terms of holomorphic geometry. One might say that the differential equations are reduced to just one – the Cauchy-Riemann equation. Breaking that code in order to write down the metric is sometimes a difficult task. Deriving global properties of the metric such as completeness is almost impossible. On the other hand, the quotient construction yields this sort of property quite easily, even if it is not as general as the twistor method.

The hyperkähler quotient construction arose also out of questions of mathematical physics [HKLR], in this case supersymmetry. In practical terms there are two ways of using it. The first is a finite-dimensional construction, whereby determining the actual metric involves solving algebraic equations. The second involves the use of the method in infinite dimensions, even though the quotient itself may be finite-dimensional. Here, one needs to solve differential equations to find the metric. They are, however, equations for which in many cases methods of solution are known so that we have *in principle* more information than an existence theorem.

1.6. We shall illustrate these constructions by a representative collection of examples which are chosen according to our guiding principle of seeking complex symplectic manifolds. These hyperkähler manifolds are all *a priori* complex manifolds with holomorphic symplectic forms:

- (i) resolutions of rational surface singularities
- (ii) coadjoint orbits of complex Lie groups
- (iii) spaces of representations of a surface group in a complex Lie group
- (iv) the space of based rational maps $f : \mathbf{C}P^1 \rightarrow \mathbf{C}P^1$ of degree k
- (v) the space of based loops in a complex Lie group

The construction of hyperkähler metrics on these spaces is contained in the work of P. B. Kronheimer, S. K. Donaldson, and others. What is perhaps remarkable is that these diverse spaces nearly all inherit a hyperkähler metric through special cases of solutions to the anti-self-dual Yang-Mills equations in \mathbf{R}^4 . Those physically motivated equations themselves are ultimately based on the identification of \mathbf{R}^4 with the quaternions. Hamilton's ghost may yet rest content.

2. THE TWISTOR CONSTRUCTION

2.1. A hyperkähler manifold M^{4n} is, by definition, endowed with three complex structures I, J and K . In fact, if $\mathbf{u} = (a, b, c) \in \mathbf{R}^3$ then

$$(aI + bJ + cK)^2 = -(a^2 + b^2 + c^2)$$

so that if $\|\mathbf{u}\| = 1$ we have another covariant constant (and hence integrable) complex structure $I_{\mathbf{u}}$. The hyperkähler metric is Kählerian with respect to all of these structures. The *twistor space* of M is the product $Z = M \times S^2$. The tangent space to the 2-sphere S^2 has a natural complex structure I_0 (considering it as the Riemann sphere) and for $X \in T_m M, Y \in T_{\mathbf{u}} S^2$ we put

$$\mathbf{I}(X, Y) = (I_{\mathbf{u}} X, I_0 Y)$$

which defines a complex structure on the tangent space $T_m M \oplus T_{\mathbf{u}} S^2$ to the twistor space Z . It is a theorem [HKLR],[S],[AHS] that this almost complex structure is *integrable* and so Z is a complex manifold of complex dimension $2n + 1$.

2.2. The twistor space has the following features. Firstly, the projection onto $S^2 \cong \mathbf{C}P^1$ is holomorphic. Secondly, the antipodal map σ on the unit sphere takes $I_{\mathbf{u}}$ to $-I_{\mathbf{u}}$ and I_0 to $-I_0$, so we may consider Z as a *real* complex manifold (a complex manifold with an antiholomorphic involution – like a projective variety defined by equations with real coefficients).

The triple of Kähler forms ω_1, ω_2 and ω_3 on M is determined by the choice of an orthonormal frame in \mathbf{R}^3 . The complex two-form ω_c can then be defined for all complex structures $I_{\mathbf{u}}$ lifted to $M \times SO(3)$. When we obtain S^2 by dividing $SO(3)$ by a circle action, ω_c is no longer a true two-form, but is twisted by the tangent line bundle $\mathcal{O}(2)$ of $\mathbf{C}P^1$. It is holomorphic.

Finally, each point $m \in M$ defines a section $\{m\} \times S^2$ of the projection $p : Z = M \times S^2 \rightarrow \mathbf{C}P^1$ which is both holomorphic and real (i.e. σ -invariant). The normal bundle of this section is isomorphic as a holomorphic bundle to $\mathbf{C}^{2n} \otimes \mathcal{O}(1)$. We then have the basic theorem [HKLR]

THEOREM 1. *Let M^{4n} be a hyperkähler manifold and Z its twistor space. Then*

- (1) *Z is a holomorphic fibre bundle $p^Z \rightarrow \mathbf{CP}^1$ over the projective line*
- (2) *the bundle admits a family of holomorphic sections, each with normal bundle isomorphic to $\mathbf{C}^{2n} \otimes \mathcal{O}(1)$*
- (3) *there exists a holomorphic section ω of $\Lambda^2 T_F^* \otimes \mathcal{O}(2)$ defining a symplectic form on each fibre*
- (4) *Z has a real structure σ compatible with (1), (2) and (3) and inducing the antipodal map on \mathbf{CP}^1 .*

Conversely, the parameter space of real sections of any complex manifold Z^{2n+1} satisfying (1) to (4) is a $4n$ -dimensional manifold with a hyperkähler metric for which Z is the twistor space.

(Here $\mathcal{O}(k)$ denotes the pull-back by p of the corresponding line bundle on \mathbf{CP}^1 and T_F^* the cotangent bundle along the fibres of p . From now on, for a vector bundle E , $E(k)$ will denote the tensor product $E \otimes \mathcal{O}(k)$).

The key to the converse is the fact that the tangent space to the space of sections is

$$H^0(\mathbf{CP}^1; T_F) \cong H^0(\mathbf{CP}^1; T_F(-1)) \otimes H^0(\mathbf{CP}^1; \mathcal{O}(1))$$

and the skew form ω gives the first factor a symplectic form and the Wronskian the second, providing a symmetric form – a metric – on the tensor product.

2.3. This is the “encoding” via the Penrose twistor space of a hyperkähler metric. The differential equations are all contained in the non-linearity of the geometrical construction of a suitable Z . Moreover, Z itself can be considered as a natural object within the realm of holomorphic symplectic geometry. Formally speaking, one may consider the projection p and the structure of the form ω as giving a symplectic manifold defined over the field of rational functions in one variable. The sections are then rational points on this variety. In particular examples, this “Diophantine” aspect is evident, as finding the sections results in solving algebraic equations with polynomials [H1],[H4].

2.4. Example. Let Z be the total space of the rank 2 vector bundle $E = \mathbf{C}^2(1)$ over CP^1 . This is the twistor space for the flat hyperkähler structure on \mathbf{R}^4 .

Now let s be a non-vanishing section of E , then s generates a trivial sub-line bundle with quotient $\mathcal{O}(2)$. Thus translation by a multiple of s describes E as a principal \mathbf{C} -bundle over the total space T of $\mathcal{O}(2)$. Given $\lambda \in \mathbf{R}$ we define a line bundle L^λ over T by

$$L^\lambda = E \times_{\mathbf{C}} \mathbf{C}$$

where $u \in \mathbf{C}$ acts by $w \mapsto e^{\lambda u} w$

If z denotes the tautological section of the pull-back of $\mathcal{O}(2)$ to its total space T , then

$$Z = \{(x, y) \in L^\lambda(1) \oplus L^{-\lambda}(1) : xy = z\}$$

defines a twistor space for a complete hyperkähler metric on \mathbf{R}^4 . This is the *Taub-NUT* metric. A derivation of the explicit metric from the twistor construction can be found in [Be].

3. THE HYPERKÄHLER QUOTIENT

3.1 The twistor construction for hyperkähler metrics has as its starting point the complex structures I, J and K . In fact, it is natural to consider such structures alone on manifolds, without the existence of a compatible metric. This is the more general theory of *hypercomplex manifolds* and the corresponding twistor theory simply involves deleting condition (3) in Theorem 1.

The quotient construction, by contrast, emphasizes not the complex structures but instead the corresponding Kähler forms ω_1, ω_2 and ω_3 . In this case, by contraction with the inverse forms on cotangent vectors, we can recover I, J and K and the metric itself. Put another way, $Sp(n)$ is the stabilizer of $\omega_1, \omega_2, \omega_3$ whereas $GL(n, \mathbf{H})$ is the stabilizer of I, J, K .

A hyperkähler manifold can be characterized in a very straightforward manner using these forms [HKLR]:

THEOREM 2. *Let M^{4n} be a manifold with 2-forms $\omega_1, \omega_2, \omega_3$ whose stabilizer in $GL(4n, \mathbf{R})$ at each point $m \in M$ is conjugate to $Sp(n)$. Then the forms define a hyperkähler structure if and only if they are closed.*

This theorem, which is a straightforward consequence of the Newlander-Nirenberg theorem, places the theory of hyperkähler manifolds firmly within the context of symplectic geometry.

3.2. The hyperkähler quotient is modelled on the Marsden-Weinstein quotient construction in symplectic geometry.

Recall that if (M, ω) is a symplectic manifold with a symplectic action of a Lie group G , then under mild assumptions one can define an equivariant moment map $\mu : M \rightarrow \mathfrak{g}^*$ taking values in the dual of the Lie algebra of G : for each $\xi \in \mathfrak{g}$ the vector field X_ξ generated by the action satisfies $d\mu(\xi) = i(X_\xi)\omega$. The moment map is ambiguous up to the addition of a constant $\zeta \in \mathfrak{z} \subseteq \mathfrak{g}^*$ where \mathfrak{z} is the space of G -invariant elements of \mathfrak{g}^* .

The symplectic quotient construction consists of choosing a regular value $\zeta \in \mathfrak{z}$ for μ and then the form ω restricted to the submanifold $\mu^{-1}(\zeta)$ is invariant and degenerate in the directions of the G -orbits and hence descends to a form $\bar{\omega}$ on the quotient manifold $\mu^{-1}(\zeta)/G$. The form $\bar{\omega}$ is symplectic.

3.3. Suppose now that M is a hyperkähler manifold, with a Lie group G acting so as to preserve the three Kähler forms ω_1, ω_2 and ω_3 . We obtain three moment maps μ_1, μ_2 and μ_3 or equivalently a vector-valued moment map

$$\mu : M \rightarrow \mathfrak{g}^* \otimes \mathbf{R}^3$$

We then have [HKLR]

THEOREM 3. *If $\zeta \in \mathfrak{z} \otimes \mathbf{R}^3$ is a regular value for the hyperkähler moment map μ , then $\mu^{-1}(\zeta)/G$ is a hyperkähler manifold.*

The proof, using Theorem 2, is direct. Each form ω_i descends to a form $\bar{\omega}_i$ just as in the symplectic case. What remains to be checked is that the quaternionic algebraic relations between $\bar{\omega}_1, \bar{\omega}_2$ and $\bar{\omega}_3$ are still satisfied.

Note that

$$\dim \mu^{-1}(\zeta)/G = \dim M - 4 \dim G$$

3.4. The symplectic quotient itself has an important rôle to play in *Kähler geometry*, for if M is a Kähler manifold and G preserves the complex structure as well as the symplectic form, then the symplectic quotient is again Kählerian [Ki],[HKLR]. The complex structure of the quotient is then identified with the complex quotient M^s/G^c of a certain open set M^s of *stable* points in M by a complex group action which holomorphically extends that of G . A point is stable if its G^c -orbit intersects $\mu^{-1}(\zeta)$. This symplectic point of view works very well for projective varieties and correlates effectively with geometric invariant theory [Ki].

3.5. Example. A simple illustration of the Kähler symplectic quotient is the following. Take $M = \mathbf{C}^n$ with its standard hermitian structure and the action of the circle $G = S^1$ by scalar multiplication. The moment map $\mu : \mathbf{C}^n \rightarrow i\mathbf{R}$ is

$$\mu(z) = i\|z\|^2$$

and $\mu^{-1}(i) = S^{2n-1}$. The symplectic quotient is therefore

$$S^{2n-1}/S^1 = \mathbf{C}P^{n-1} = \mathbf{C}^n \setminus \{0\} / \mathbf{C}^*$$

so that $M^s = \mathbf{C}^n \setminus \{0\}$ and $\mathbf{C}P^{n-1}$ inherits a natural Kähler metric – the *Fubini-Study* metric.

3.6. In the hyperkähler situation, since the forms ω_1, ω_2 and ω_3 define I, J and K then if an action of G preserves the symplectic forms, it automatically preserves the complex structures. Fixing attention on one such complex structure I , the function $\mu_c = \mu_2 + i\mu_3$ is actually holomorphic. It is the moment map with respect to the holomorphic symplectic form ω_c of the action of the complex group G^c . Thus $\mu^{-1}(\zeta)$ can be rewritten in the form

$$\mu_1^{-1}(a) \cap \mu_c^{-1}(\alpha)$$

for some $a \in \mathfrak{z}$ and $\alpha \in \mathfrak{z} \otimes \mathbf{C}$.

From this point of view, the hyperkähler quotient is the symplectic quotient of the Kähler manifold $\mu_c^{-1}(\alpha)$. It follows, that when due account is taken of stability, the induced complex structure I on the hyperkähler quotient is simply that of $\mu_c^{-1}(\alpha)/G^c$ – the holomorphic version of the symplectic quotient.

3.7. As the complex structure I varies (equivalently as we look at the different fibres of the twistor space $p : Z \rightarrow \mathbf{CP}^1$) the twistor space of the hyperkähler quotient is now essentially the fibre-wise symplectic quotient of Z by the holomorphic action of G^c . For each complex structure, the complex moment map μ_c is the restriction of a holomorphic section $\tilde{\mu}$ of $\mathfrak{g}^c(2)$, incorporating the twist of the form ω of Theorem 1. The twistor space of the quotient is then $\tilde{\mu}^{-1}(\zeta)^s/G^c$ for $\zeta \in \mathfrak{g} \otimes \mathbf{R}^3 \subset \mathfrak{g}^c \otimes H^0(\mathbf{CP}^1; \mathcal{O}(2)) \subset H^0(Z; \mathfrak{g}^c(2))$.

In the context of (2.3) it is just the symplectic quotient over the field of rational functions.

3.8. In practice it is remarkable that one may obtain interesting examples by starting with the flat hyperkähler manifold $M^{4n} = \mathbf{H}^n$ and taking a quotient by a linear action of a group. The closest analogue of the projective space construction in (3.5) is the following example.

Let V be an n -dimensional hermitian vector space and V^* its dual. Then

$$M^{4n} = V \oplus V^*$$

is a flat hyperkähler manifold. Let $G = S^1$ and let $u \in S^1$ act on M by

$$u(x, \xi) = (ux, u^{-1}\xi)$$

This action preserves the hyperkähler 2-forms and using the complex and real moment maps as in (3.6) one finds

$$\mu_c(x, \xi) = \xi(x) \quad \mu_1(x, \xi) = i(\|x\|^2 - \|\xi\|^2)$$

Thus if $\zeta = (i, 0, 0) \in \mathfrak{g}^* \otimes \mathbf{R}^3$ then

$$\mu^{-1}(\zeta) = \{(x, \xi) : \xi(x) = 0, \|x\|^2 - \|\xi\|^2 = 1\}$$

The second condition determines the set of stable points M^s as the set where $x \neq 0$ and then M^s/\mathbf{C}^* is the *cotangent bundle* of $\mathbf{P}(V) = \mathbf{C}P^{n-1}$. This, however, is only one of the 2-sphere of complex structures on this hyperkähler manifold. Because of the symmetries in this example, the others can be identified with the quotients $\mu_c^{-1}(\alpha)/\mathbf{C}^*$ for $\alpha \neq 0$ which are affine bundles over $\mathbf{C}P^{n-1}$ with group of translations the cotangent bundle. Note that complex cotangent bundles are naturally complex symplectic manifolds.

This particular metric for $n = 2$ is the *Eguchi-Hanson* metric [EH]. The generic complex structure is that of an affine quadric in \mathbf{C}^3 . The higher-dimensional examples were first found by E. Calabi [Ca].

3.9. One general class of hyperkähler moment maps we shall encounter is based on the following example.

Let G be a Lie group with a bi-invariant metric, and put

$$M = \mathfrak{g} \otimes \mathbf{H}$$

where \mathfrak{g} is the Lie algebra. This is a flat hyperkähler manifold, and the adjoint action of G preserves the symplectic forms. The three hyperkähler moment maps are then

$$\mu_1(A) = [A_0, A_1] + [A_2, A_3]$$

$$\mu_2(A) = [A_0, A_2] + [A_3, A_1]$$

$$\mu_3(A) = [A_0, A_3] + [A_1, A_2]$$

where $A \in \mathfrak{g} \otimes \mathbf{H}$ is defined by $A = A_0 + A_1i + A_2j + A_3k$. Note that the complex moment map is given by

$$\mu_c(A) = [A_0 + iA_1, A_2 + iA_3]$$

and putting $\alpha = A_0 + iA_1$ and $\beta = A_2 + iA_3$ the three moment maps can be written as a complex and real moment map

$$\mu_c = [\alpha, \beta]$$

$$\mu_r = [\alpha, \alpha^*] + [\beta, \beta^*]$$

It is worth pointing out here that replacing the A_i by covariant derivatives ∇_i yields the anti-self dual Yang-Mills equations as the single equation $\mu = 0$. We shall see versions of this result in the examples which follow.

4. RATIONAL DOUBLE POINTS

4.1. In the study of surface singularities, the *rational double points* are characterized by the property that their minimal resolution has trivial canonical bundle – these are thus, locally, 2-dimensional complex manifolds with holomorphic symplectic forms. The basic model for such a singularity is the space \mathbf{C}^2/Γ where Γ is a finite subgroup of $SU(2)$. We shall describe here the construction of Kronheimer [K1],[K2] for a hyperkähler metric on the minimal resolution. We have in (3.8) already encountered such a metric – the Eguchi-Hanson metric. This was defined on the cotangent bundle of $\mathbf{C}P^1$, but the zero section of this has self-intersection -2 and so can be blown down to a singularity – the ordinary double point $\mathbf{C}^2/\pm 1$. Kronheimer’s construction proceeds (as in (3.8)) by using hyperkähler quotients.

4.2. Let Γ be a finite subgroup of $SU(2)$ and let $V = L^2(\Gamma)$ be the finite-dimensional Hilbert space of functions on Γ , $U(V)$ the unitary group of V and $\mathfrak{u}(V)$ its Lie algebra. Then, as in (3.9), $\mathfrak{u}(V) \otimes \mathbf{H}$ is a flat hyperkähler manifold – a quaternionic vector space with a compatible inner product.

Now, since $SU(2)$ is isomorphic to the unit quaternions, Γ acts on the left on both the quaternions and $L^2(\Gamma) = V$ and so has an action on $\mathfrak{u}(V) \otimes \mathbf{H}$ commuting with right multiplication by quaternions. Thus

$$M = (\mathfrak{u}(V) \otimes \mathbf{H})^\Gamma$$

the space of invariant elements, is also a flat hyperkähler manifold.

The projective group $PU(V)$ acts by conjugation on $\mathfrak{u}(V) \otimes \mathbf{H}$ preserving the hyperkähler structure and so $G = PU(V)^\Gamma$ acts on M in the same way. This is the context in which we may take the hyperkähler quotient.

4.3. To gain more information, we decompose V into irreducible representations. Now

$$V = L^2(\Gamma) = \bigoplus_{i=0}^l (V_i^L \otimes V_i^R)$$

where the sum is over the irreducible representations of Γ under left and right action, and V_0 is the trivial representation. Thus

$$U(V)^\Gamma = U(V_0^R) \times \dots \times U(V_l^R)$$

Hence if $\dim V_i^R = \dim V_i^L = n_i$, then

$$\dim G = \sum_0^l n_i^2 - 1$$

We can also see clearly the centre of G from this decomposition, and from that the subspace $\mathfrak{z} \in \mathfrak{g}^*$ of invariant elements. We have

$$\dim \mathfrak{z} = l$$

4.4. The structure of the irreducible representations of Γ is conveniently encoded through the McKay correspondence [McK]. To each subgroup Γ (cyclic, or binary dihedral, tetrahedral, octahedral or icosahedral) there corresponds a Dynkin diagram of type A_l, D_l, E_6, E_7, E_8 respectively. Each vertex of the diagram corresponds to a non-trivial irreducible representation of Γ .

In Lie algebra theory, each vertex corresponds to a simple root α_i . The highest root is expressed in terms of the simple roots by

$$\sum_{i=1}^l n_i \alpha_i$$

From McKay the integer n_i is the dimension of the representation space corresponding to the i th vertex. The whole situation is simplified by

introducing the *extended* Dynkin diagram incorporating the highest root. The relation

$$\sum_{i=0}^l n_i \alpha_i = 0$$

with $n_0 = 1$ and α_0 the negative of the highest root then puts the trivial representation on the same footing as the others.

4.5. To find the dimension of M in (4.2) requires a knowledge of the relationship between the tensor product of the defining 2-dimensional representation of $\Gamma \subset SU(2)$ and the other representations V_i . This again is provided by the McKay correspondence:

$$\mathbf{C}^2 \otimes V_i = \bigoplus_j a_{ij} V_j$$

where $a_{ij} = 1$ if the vertices of the extended Dynkin diagram are adjacent and zero otherwise.

As a complex vector space, $M = (\text{Hom}(V, V) \otimes \mathbf{C}^2)^\Gamma$ and so

$$\begin{aligned} \dim_{\mathbf{C}} M &= \dim \bigoplus_{i,j} (\text{Hom}(V_j^L \otimes V_j^R, V_i^L \otimes V_i^R \otimes \mathbf{C}^2))^\Gamma \\ &= \dim \bigoplus_{i,j} a_{ij} \text{Hom}(V_i^L, V_j^L) \\ &= \sum_{i,j} a_{ij} n_i n_j \end{aligned}$$

But now applying the Cartan form to the relation in (4.4) gives

$$-2 \sum_0^l n_i^2 + \sum_{i,j} a_{ij} n_i n_j = 0$$

and so

$$\dim_{\mathbf{R}} M = 4 \sum_0^l n_i^2 = 4(\dim G + 1)$$

from (4.3).

Thus, provided G is shown to act freely, we have a hyperkähler quotient of dimension $\dim M - 4 \dim G = 4$. Note that since \mathfrak{z} is l -dimensional there is a choice in moment map, which must be suitably exercised to obtain a free action.

4.6. The relationship with the resolution of \mathbf{C}^2/Γ may be seen by considering $\zeta = (\zeta_1, 0, 0)$. Since we are essentially in the situation of (3.9), the hyperkähler moment map equations can be written as equations for a pair of complex matrices

$$\begin{aligned} [\alpha, \beta] &= 0 \\ [\alpha, \alpha^*] + [\beta, \beta^*] &= \zeta_1 \end{aligned}$$

but where $\alpha, \beta \in \text{Hom}(V, V)$ define a Γ -invariant element of $\text{Hom}(V, V) \otimes \mathbf{C}^2$.

This invariance means that if $e \in V$ is a common eigenvector of α and β (which exists since they commute) with eigenvalues a, b then $(\alpha, \beta)e = (a, b)e$ and

$$(\alpha, \beta)e^\gamma = (a, b)^\gamma e^\gamma$$

where $(a, b) \mapsto (a, b)^\gamma$ is the defining action of Γ on \mathbf{C}^2 .

Since Γ acts freely on $\mathbf{C}^2 \setminus \{0\} = \mathbf{H} \setminus \{0\}$ (multiplication by a non-zero quaternion) if one eigenvalue pair is non-zero then we have $|\Gamma| = \dim V$ distinct eigenvalue pairs. In particular, the Γ -orbit of one such pair (a, b) is well-defined by (α, β) . This provides a map p , holomorphic with respect to I , to \mathbf{C}^2/Γ . If ζ_1 is chosen appropriately, the quotient is non-singular, so that once one proves that p is biholomorphic outside the origin in \mathbf{C}^2/Γ the map is a resolution.

The four-dimensional metrics produced this way are not only complete hyperkähler manifolds, but they are also *asymptotically locally Euclidean* (ALE) meaning that they approach rapidly the Euclidean metric on \mathbf{R}^4/Γ at infinity.

5. COADJOINT ORBITS

5.1. For any Lie group, an orbit in the dual of the Lie algebra is a symplectic manifold. This is the canonical Kostant-Kirillov symplectic structure. If

the group G is compact, then the orbits are Kähler manifolds. This fact may be exploited to obtain all the irreducible representations of G within the context of geometric quantization (the Borel-Weil theorem).

The existence of a symplectic structure is completely general. In particular, a coadjoint orbit of a complex Lie group G^c is a holomorphic symplectic manifold. These manifolds in many cases possess natural *hyperkähler* metrics due to the work of Kronheimer [K3]. The Eguchi-Hanson metric already provides us with one example – the affine quadric in \mathbf{C}^3 (which is the generic complex structure) is the orbit under $SL(2, \mathbf{C})$ of a semi-simple element in $\mathfrak{g} = \mathfrak{g}^*$. We shall give here the general construction for a *regular* semisimple orbit – one of the form G^c/T^c where G^c is a complex semisimple Lie group and T^c a maximal complex torus. The compact analogue of this is the flag manifold G/T , the orbit type which features prominently in the Borel-Weil theorem.

5.2. The metric is produced by an application of the hyperkähler quotient construction to an infinite-dimensional flat hyperkähler manifold. The setting is that of a special case of the anti-self-dual Yang-Mills equations. We consider a compact semisimple group G and a trivial principal G -bundle P over $\mathbf{R}^4 \setminus \{0\}$, and the space of connections on P . This is an infinite-dimensional affine space modelled on the vector space $\Omega^1(\mathbf{R}^4 \setminus \{0\}; \text{ad}P)$ of 1-forms with values in the Lie algebra bundle. This is itself a quaternionic vector space, inheriting its structure from the identification of \mathbf{R}^4 with \mathbf{H} . The left action of the unit quaternions $Sp(1)$ commutes with the right action and it follows that the space of $Sp(1)$ -invariant connections is also quaternionic. The invariant automorphisms or *gauge transformations* act on this space.

To define a metric requires a closer attention to boundary conditions, but having done that, one may consider the hyperkähler quotient of this affine space by the group of gauge transformations to obtain a hyperkähler quotient.

5.3. The $Sp(1)$ -invariance condition above throws the emphasis onto a radial variable $s \in (-\infty, 0]$. The moment map equations can then be

transformed to the following non-linear system of equations

$$\begin{aligned} \left[\frac{d}{ds} + B_0, B_1\right] + [B_2, B_3] &= 0 \\ \left[\frac{d}{ds} + B_0, B_2\right] + [B_3, B_1] &= 0 \\ \left[\frac{d}{ds} + B_0, B_3\right] + [B_1, B_2] &= 0 \end{aligned}$$

for functions $B_i(s)$ with values in the Lie algebra \mathfrak{g} . Note how these equations compare with those in (3.9) putting $A_0 = d/ds + B_0$ and $A_i = B_i$ for $i > 0$.

The boundary conditions which give rise to this moment map are defined by comparison with a particular configuration given by $B_0 = 0$ and $B_i = \tau_i$ where τ_1, τ_2, τ_3 lie in a fixed Cartan subalgebra \mathfrak{h} . They are to be chosen such that their common centralizer is \mathfrak{h} itself. This is clearly a solution to the equations. The space of operators $d/ds + B_0, B_1, B_2, B_3$ which are close to this model configuration in some exponentially-weighted C^1 norm [K3] then admits a well-defined inner product. (In the 4-dimensional interpretation above this is simply the L^2 inner product).

The adjoint action of the group of smooth functions $g(s)$ with values in G on the four operators $d/ds + B_0, B_1, B_2, B_3$ then gives a hyperkähler group action. (This is the group of invariant gauge transformations in the 4-dimensional formalism).

It is easy to see in this case that the quotient is finite-dimensional, since by a gauge transformation the operator $d/ds + B_0$ can be transformed to d/ds leaving an ordinary differential equation in B_1, B_2 and B_3 with equivalence under the finite-dimensional group G .

5.4. Identifying the complex structures on the quotient involves an extra theorem. Here one chooses the complex structure and rewrites the moment map equations as in (3.9) with a real and complex part

$$\begin{aligned} \frac{d\beta}{ds} + [\alpha, \beta] &= 0 \\ \frac{d}{ds}(\alpha + \alpha^*) + [\alpha, \alpha^*] + [\beta, \beta^*] &= 0 \end{aligned}$$

where $\alpha = (B_0 + iB_1)(s)$ and $\beta = (B_2 + iB_3)(s)$.

One then shows that if the boundary conditions are satisfied, the map

$$(\alpha, \beta) \mapsto \beta(0) \in \mathfrak{g}^c$$

identifies the space of equivalence classes of solutions to the above equations with the adjoint orbit of $\tau_2 + i\tau_3$. For G^c semisimple this is isomorphic to the coadjoint orbit. The proof itself is modelled on a theorem of Donaldson [D1].

5.5. Notice that there is a choice in the moment map reflected this time in the boundary conditions. Part of that choice involves $\tau_2 + i\tau_3$, the particular coadjoint orbit, but the extra choice of τ_1 in the Cartan subalgebra gives a family of hyperkähler metrics.

We may remark also that although finding the metric explicitly involves solving the non-linear equations in (5.3), they are a form of *Nahm's equations*, the general solution of which can be described in terms of the geometry of the Jacobian of an algebraic curve [H3]. There is clearly much more than an existence theorem involved here.

6. REPRESENTATIONS OF SURFACE GROUPS

6.1. If Σ is a compact oriented surface, its fundamental group has an intrinsically symplectic nature [G]. In particular, for any Lie group G with an invariant inner product on its Lie algebra, the moduli space of irreducible representations of $\pi_1(\Sigma)$ is a symplectic manifold. This is the space $\text{Hom}^{\text{irr}}(\pi_1(\Sigma), G)/G$ where G acts by conjugation. The tangent space at a representation can be identified with the cohomology group $H^1(\pi_1(\Sigma); \mathfrak{g})$ and the bilinear form on \mathfrak{g} gives a skew pairing to $H^2(\pi_1(\Sigma))$ which is generated by a fundamental class.

When G is a compact group, a choice of complex structure on Σ makes the moduli space into a Kähler manifold, its holomorphic structure being that of the space of stable holomorphic bundles on the Riemann surface Σ . This is the theorem of Narasimhan and Seshadri [NS].

For a complex semi-simple group G^c , the moduli space

$$\mathrm{Hom}^{\mathrm{irr}}(\pi_1(\Sigma), G^c)/G^c$$

is naturally a complex symplectic manifold and has in fact a natural hyperkähler metric, also determined by a choice of complex structure on Σ .

6.2. The context for this metric is again an infinite-dimensional hyperkähler quotient construction.

Let \mathcal{A}^c be the affine space of all G^c -connections on a principal G^c -bundle over Σ which we assume has a fixed reduction to the maximal compact group G . This provides a metric on any associated vector bundle, and a conjugation operation on the bundle $\mathrm{ad}P$ associated to the adjoint representation of G . Each tangent vector to \mathcal{A}^c may be considered as a 1-form $\alpha \in \Omega^1(\Sigma; \mathrm{ad}P \otimes \mathbf{C})$ and if $\mathrm{Tr}(AB)$ denotes the inner product on \mathfrak{g} , we have a complex symplectic form defined by

$$\omega(\alpha, \beta) = \int_{\Sigma} \mathrm{Tr}(\alpha \wedge \beta)$$

Given a complex structure on Σ , we may decompose $\alpha = \alpha^{1,0} + \alpha^{0,1}$ into forms of different type and then there is a real symplectic form – the Kähler form of the metric

$$\|\alpha\|^2 = \int_{\Sigma} \mathrm{Tr}(\alpha^{1,0} \wedge \alpha^{1,0*}) - \mathrm{Tr}(\alpha^{0,1} \wedge \alpha^{0,1*})$$

This makes \mathcal{A}^c into an infinite-dimensional hyperkähler manifold. The group \mathcal{G} of gauge transformations acts on \mathcal{A}^c preserving this structure.

6.3. The moment maps for this hyperkähler action are expressed in terms of curvature. This is a consequence of a fundamental observation of Atiyah and Bott [AB]. If we write them in terms of a real and complex moment map we obtain

$$\mu_c(A) = F_A \in \Omega^2(\Sigma; \mathrm{ad}P \otimes \mathbf{C})$$

$$\mu_r(A) = F' - F'' \in \Omega^2(\Sigma; \mathrm{ad}P)$$

where F' and F'' are the curvatures of the unique G -connections ∇' and ∇'' such that

$$(\nabla')^{1,0} = \nabla_A^{1,0} \quad (\nabla'')^{0,1} = \nabla_A^{0,1}$$

Thus $\mu_c^{-1}(0)$ is the space of flat connections on the principal G^c -bundle.

A theorem of Donaldson [D2] and Corlette [Co] shows that each irreducible \mathcal{G}^c -orbit contains a solution to $\mu_r = 0$, which is unique modulo \mathcal{G} , so that the complex structure of the hyperkähler quotient can be identified with the space of flat irreducible connections modulo complex gauge equivalence. But the holonomy identifies this with the moduli space of representations $\text{Hom}^{\text{irr}}(\pi_1(\Sigma), G^c)/G^c$.

6.4. The proof of Donaldson and Corlette's theorem involves a reinterpretation of the moment map equations. The set-up for the quotient construction above involves a fixed metric on a principal bundle and an equivalence class of flat connections. Alternatively, we can consider a fixed flat connection and then solve for a metric satisfying $F' = F''$. A metric, compatible with the G -structure, is a section of the associated flat G^c/G -bundle and then the equations to be satisfied are equivalent to the statement that the section is *harmonic*. The non-positivity of the curvature of G^c/G then yields, via the Eells-Sampson theorem [ES], the existence result.

The analysis required to rigorously produce the moduli space with its metric involves Sobolev spaces for compact manifolds and is quite standard (see [H5]).

6.5. This hyperkähler manifold, of dimension $4(g-1) \dim G$, where g is the genus of Σ , shares a number of properties with the simple Eguchi-Hanson metric. The complex structure of the moduli space of representations constructed above is that of an affine variety, like the affine quadric. On the other hand there is one complex structure (and its conjugate) out of the 2-sphere generated by I, J and K which is not affine but which instead contains as a dense open set the cotangent bundle of the moduli space of stable G^c -bundles. This is the analogue of the cotangent bundle of $\mathbf{C}P^1$ for the Eguchi-Hanson metric. With this complex structure, the hyperkähler manifold can be identified as the moduli space of stable *Higgs bundles* (or

stable pairs) on Σ [H5]. In this case the compactness of the surface imposes boundary conditions on the moment map equations which have as yet defied explicit solution.

7. RATIONAL MAPS

7.1. The quotient constructions offered as examples in Sections 5,6 and 7 are based on equations in 0,1 and 2 dimensions. There is a system of equations in \mathbf{R}^3 which again yields moduli spaces which are hyperkählerian. These are the *Bogomolny equations* for magnetic monopoles.

We consider here a trivial principal G -bundle P over \mathbf{R}^3 where G is a compact Lie group, and the space \mathcal{A} of connections on P . We put

$$M = \mathcal{A} \times \Omega^0(\mathbf{R}^3; \text{ad}P)$$

A point of M thus consists of a pair (A, ϕ) where A is a connection and ϕ a section of the adjoint bundle – the Higgs field.

The tangent space at a point is identified with the set of pairs

$$(\alpha, \phi) \in \Omega^1(\mathbf{R}^3; \text{ad}P) \times \Omega^0(\mathbf{R}^3; \text{ad}P)$$

and this has an obvious quaternionic structure by writing a tangent vector as

$$\phi + \alpha_1 i + \alpha_2 j + \alpha_3 k$$

There is an action of the group of gauge transformations \mathcal{G} on M .

If we impose suitable boundary conditions (essentially comparison with a model as in Section 5) on (A, ϕ) and \mathcal{G} , then the L^2 inner product defines a hyperkähler metric on M and the action of \mathcal{G} preserves it. We may then take a hyperkähler quotient.

7.2. If we write ∇_i for the directional covariant derivative defined by the connection A , the moment map equations become (cf 3.9)

$$[\phi, \nabla_1] + [\nabla_2, \nabla_3] = 0$$

$$[\phi, \nabla_2] + [\nabla_3, \nabla_1] = 0$$

$$[\phi, \nabla_3] + [\nabla_1, \nabla_2] = 0$$

or more compactly, using the Hodge star operator,

$$F_A = *\nabla_A\phi$$

which are the Bogomolny equations.

In the simplest case $G = SU(2)$, the boundary conditions on ϕ one usually takes imply that $\|\phi\| \rightarrow 1$ as $r \rightarrow \infty$. This yields an integer k , the degree of the map from a large sphere in \mathbf{R}^3 to the unit sphere in $\mathfrak{g} \cong \mathbf{R}^3$. This is called the *charge* of the solution.

7.3. The moduli space here can be identified by a supplementary theorem (due to Donaldson [D1]) with the space of based rational maps $f : \mathbf{C}P^1 \rightarrow \mathbf{C}P^1$ of degree k . The actual identification involves choosing a direction \mathbf{u} and studying a scattering problem for the ordinary differential equation $(\nabla_{\mathbf{u}} + i\phi)s = 0$.

We may write such a rational map in the form

$$f(z) = \frac{a_0 + \dots + a_{k-1}z^{k-1}}{b_0 + b_1z + \dots + z^k} = \frac{p(z)}{q(z)}$$

where the base-point ∞ is mapped to 0. Here, for f to be of degree k , $p(z)$ and $q(z)$ have no common factor. The space of such maps is clearly a complex manifold of real dimension $4k$. As a hyperkähler manifold it has a complex symplectic form. This is obtained as follows.

First factorize $q(z) = (z - \beta_1)\dots(z - \beta_k)$. Since p and q have no common factor $p(\beta_i) \neq 0$. Then the form

$$\omega = \sum_{i=1}^k \frac{dp(\beta_i) \wedge d\beta_i}{p(\beta_i)}$$

extends to a holomorphic symplectic form on the space of rational maps. It is covariant constant with respect to the hyperkähler metric.

7.4. The rigorous construction of these metrics requires some analysis which was produced by C. H. Taubes. An account may be found in [AH] of the use of these results and properties of the metrics. There is one particular feature which distinguishes this family from the previous ones. The

identification with the space of rational maps involves a choice of direction in \mathbf{R}^3 . On the other hand the Bogomolny equations themselves are $SO(3)$ -invariant. Thus $SO(3)$ acts transitively on the 2-sphere of complex structures on the moduli space which are therefore all equivalent. This is also true of the Taub-NUT metric, whose twistor space as described in (2.4) clearly inherits the action of $SU(2)$ on CP^1 .

These monopole metrics are known insofar as their twistor spaces can be described exactly (see [AH]). In the case of $k = 1$ and 2 they have also been computed explicitly— in the first case it is the flat metric on $S^1 \times \mathbf{R}^3$, and in the second a metric whose description involves elliptic integrals [AH].

8. LOOP GROUPS

8.1. If G is a compact Lie group, and $LG = \text{Map}(S^1; G)$ denotes the space of smooth maps from the circle to G then the quotient LG/G by the subgroup G of constant maps is well-known to have the structure of an infinite-dimensional symplectic manifold [PS]. Indeed it also has a natural Kähler metric. The symplectic form is obtained by translation from the identity of the skew form

$$\omega(f, g) = \int_{S^1} \text{Tr}(fg') d\theta$$

for $f, g : S^1 \rightarrow \mathfrak{g}$.

Clearly this definition holds when G is replaced by a complex semi-simple group G^c , so that the complex manifold LG^c/G^c has a natural holomorphic symplectic form. Donaldson [D3] has shown how to give this space a natural hyperkähler metric.

8.2. The approach is to consider a trivial principal G -bundle over the unit disc $D \subset \mathbf{C}$ and the space \mathcal{A}^c of connections on the associated G^c -bundle P^c which are smooth up to the boundary. The setting is therefore similar to that in Section 6, but now we have a non-trivial boundary, the circle S^1 . However, if we consider the group of gauge transformations \mathcal{G} which restrict

to the identity on the boundary, then the hyperkähler moment maps are the same as in (6.2)

$$\begin{aligned}\mu_c(A) &= F_A \in \Omega^2(D; \text{ad}P \otimes \mathbb{C}) \\ \mu_r(A) &= F' - F'' \in \Omega^2(D; \text{ad}P)\end{aligned}$$

8.3. To solve the moment map equations, the harmonic map formulation is used again. In this case all flat connections are trivial, since the disc is simply-connected, so the problem concerns harmonic maps $f : D \rightarrow G^c/G$. Results of R. Hamilton [Ha] show that the Dirichlet problem can be solved in this case. Thus, given a map on the boundary circle, there is a unique harmonic map extending it to the disc.

In the formulation of connections, this means that given the boundary value of a flat G^c -connection on D , there is a unique solution of the hyperkähler moment map equations, modulo gauge transformations which preserve the boundary value. But since all such connections on the circle are gauge-equivalent to the trivial connection, which has automorphism group G^c , the hyperkähler quotient we are considering is

$$\text{Map}(S^1; G^c)/G^c$$

8.4. The holomorphic form ω of Section 6 restricts to the space of flat connections $\mu_c^{-1}(0)$ to give

$$\omega(\alpha, \beta) = \int_D \text{Tr}(\alpha \wedge \beta)$$

But, to be tangential at A to a flat connection, α and β must satisfy $\alpha = d_A\varphi$ and $\beta = d_A\psi$ for some $\varphi, \psi \in \Omega^0(D; \text{ad}P^c)$ and hence

$$\begin{aligned}\omega(\alpha, \beta) &= \int_D \text{Tr}(d_A\varphi \wedge d_A\psi) \\ &= \int_{S^1} \text{Tr}(\varphi d_A\psi)\end{aligned}$$

which is clearly the skew form defined above on boundary values of maps $f : D \rightarrow G^c$.

8.5. Finally note that the moment map equations for this example can be put in the more familiar form of (3.9) by means of some substitutions. If we set

$$\nabla = (\nabla' + \nabla'')/2 \quad \phi = (\nabla' - \nabla'')/2$$

where $\phi = \phi_1 dx_1 + \phi_2 dx_2$, then the equations may be written as

$$[\nabla_1, \nabla_2] + [\phi_1, \phi_2] = 0$$

$$[\nabla_1, \phi_2] + [\nabla_2, \phi_1] = 0$$

$$[\nabla_1, \phi_1] + [\phi_2, \nabla_2] = 0$$

9. CONCLUSIONS

9.1. The quotient construction yields a vast number of hyperkähler manifolds, in fact such a large number that a secondary task now presents itself to instil some order amongst them. This is particularly important because even those exemplified here have interrelationships – hyperkähler metrics may appear on the same space through different constructions.

Kronheimer's coadjoint orbits provide an example. The construction in Section 6 is given as an infinite-dimensional quotient based on ordinary differential equations. On the other hand D. Burns [Bu] has given an algebraic twistor construction (characteristically manifesting difficulty in showing completeness) for metrics on the same spaces. It is also true that Donaldson's complex loop group metric in Section 8 contains the same coadjoint orbits as fixed point sets of circle actions. Also, some of these spaces (as the Eguchi-Hanson metric shows) can be obtained as finite-dimensional quotients. Proving these to be isometric is not always easy.

Another class of examples are the monopole moduli spaces, which acquired a hyperkähler metric through an infinite-dimensional quotient based on the Bogomolny equations in 3 dimensions. There is however, through the

Nahm transform, a way of defining a hyperkähler metric through Nahm's equations in 1 dimension. Results such as those of Nakajima [N] have shown that these particular metrics coincide. A similar situation holds for instanton moduli spaces where results [BvB], [Ma] show the coincidence of metrics defined by two methods.

9.2. There are also quotient constructions for a class of manifolds related to the quaternions in a slightly more general manner than hyperkähler manifolds. We have already encountered *hypercomplex* manifolds, but there is also the class of *quaternionic Kähler* manifolds, which are Riemannian manifolds with holonomy group $Sp(n).Sp(1)$ like \mathbf{HP}^n and *quaternionic* manifolds, which admit a torsion-free $GL(n, \mathbf{H}).Sp(1)$ connection [S]. For all of these [GL],[J] there are quotient constructions, but so far with a lesser range of examples than the hyperkähler quotient.

9.3. One final question concerns the compact examples like the K3 surface where we still rely entirely on existence theorems. Could they be obtained by the quotient construction? In particular, through a finite-dimensional quotient of a vector space?

The answer is no. Compact examples of finite-dimensional Kähler quotients, like the projective space in (3.5), may exist but the special curvature of a hyperkähler manifold prevents this from happening in the hyperkähler context. Perhaps the easiest way to see this is in the 4-dimensional case.

Suppose that a compact hyperkähler manifold M^4 is obtained by a quotient of a linear hyperkähler action of G on \mathbf{H}^n . Then if μ is the hyperkähler moment map, $\mu^{-1}(\zeta)$ is a principal G -bundle P over M . The flat metric on \mathbf{H}^n induces a G -invariant metric on P and the orthogonal complements to the orbit directions define a connection on P . Now for each complex structure on \mathbf{H}^n (and hence M), the associated principal G^c -bundle is $\mu_c^{-1}(\alpha)$ which is holomorphic. This implies that the connection is compatible with all three complex structures I , J and K and hence (see [AHS]) is a solution to the anti-self dual Yang-Mills equations on M . According to the Atiyah-Ward construction, this corresponds to a holomorphic principal bundle on the twistor space Z of M which is, not surprisingly, $\tilde{\mu}^{-1}(\zeta)$.

The connection has a further property, for we have an embedding of $\tilde{\mu}^{-1}(\zeta)$ in the twistor space $\mathbf{C}^{2n}(1)$ corresponding to the embedding $P = \mu^{-1}(\zeta) \subset \mathbf{H}^n$. This embedding defines an equivariant section of $\mathbf{C}^{2n}(1)$ on the principal G^c -bundle over Z and this (see [H2]) has an interpretation via the Penrose transform as the solution ψ of a differential equation on M – the *twistor equation* coupled to the anti-self-dual Yang-Mills connection. Now for such an object there is a Weitzenböck formula which gives a vanishing theorem on a *compact* manifold. In our case, since the Ricci tensor of a hyperkähler manifold vanishes, then in particular so does the scalar curvature, and the corresponding vanishing theorem [H2] implies covariant constancy of ψ . This in turn means that the map $P \subset \mathbf{H}^n$ is constant in horizontal directions which is a contradiction to it being an embedding.

9.4. As for infinite-dimensional quotients, there remains a hope. Instanton moduli spaces on hyperkähler manifolds have induced hyperkähler metrics through the moment map interpretation of the anti-self-dual Yang-Mills equations, but they tend to be either non-compact or have singularities corresponding to reducible connections. However, introducing singularities into the connections themselves may yet induce more regularity into the moduli space. Ultimately, it is possible that the K3 surface may be brought into the fold of hyperkähler quotients. It does not necessarily mean that we can write down the metric (a glance at [AH] will show that explicitness is not gained easily) but that we shall have in any case a deeper understanding of the pervasiveness of the geometry which has arisen from the quaternions.

REFERENCES

- [AHS] M. F. ATIYAH, N. J. HITCHIN and I. M. SINGER - *Self-duality in four-dimensional Riemannian geometry*, Proc. Roy. Soc. Lond. **A362** (1978), 425-461.
- [AB] M. F. ATIYAH and R. BOTT - *The Yang-Mills equations over Riemann surfaces*, Phil. Trans. Roy. Soc. Lond. **A308** (1982), 523-615.

- [AH] M. F. ATIYAH and N. J. HITCHIN - *"The geometry and dynamics of magnetic monopoles"*, Princeton University Press, Princeton (1988).
- [Bea] A. BEAUVILLE - *Variétés Kählériennes dont la 1ère classe de Chern est nulle*, J. Diff. Geometry **18** (1983), 755-782.
- [Be] A. BESSE - *"Einstein manifolds"*, Springer Verlag, Berlin (1987).
- [Bo] S. BOCHNER - *Curvature in Hermitian metric*, Bull. Amer. Math. Soc. **53** (1947) 179-195.
- [Bu] D. BURNS - *Some examples of the twistor construction*, in "Contributions to several complex variables: in honor of Wilhelm Stoll" (eds. A. Howard and P. -M. Wong), Vieweg, Braunschweig (1986).
- [BvB] P. J. BRAAM and P. van BAAL - *Nahm's transformation for instantons*, Commun. Math. Phys. **122** (1989), 267-280.
- [Ca] E. CALABI - *Métriques kählériennes et fibrés holomorphes*, Ann. Éc. Norm. Sup. **12** (1979) 269-294.
- [Co] K. CORLETTE - *Flat G-bundles with canonical metrics*, J. Diff. Geometry **28** (1988) 351-382.
- [D1] S. K. DONALDSON - *Nahm's equations and the classification of monopoles*, Commun. Math. Phys. **96** (1984) 387-407.
- [D2] S. K. DONALDSON - *Twisted harmonic maps and the self-duality equations*, Proc. London Math. Soc. **55** (1987) 127-131.
- [D3] S. K. DONALDSON - *Boundary value problems for Yang-Mills fields*, preprint (Oxford) (1991).
- [EH] T. EGUCHI and A. J. HANSON - *Asymptotically flat self-dual solutions to Euclidean gravity*, Phys. Lett. **74B** (1978) 249-251.
- [ES] J. EELLS and J. H. SAMPSON - *Harmonic mappings of Riemannian manifolds*, Amer. J. Math. **86** (1964) 109-160.
- [GL] K. GALICKI and H. B. LAWSON - *Quaternionic reduction and quaternionic orbifolds*, Math. Ann. **282** (1988) 1-21.
- [G] W. M. GOLDMAN - *The symplectic nature of fundamental groups of surfaces*, Adv. in Math. **54** (1984) 200-225.
- [Gr] A. GRAY - *A note on manifolds whose holonomy group is a subgroup of $Sp(n).Sp(1)$* , Michigan Math. J. **16** (1969) 125-128.
- [Ha] R. S. HAMILTON - *"Harmonic maps of manifolds with boundary"*, Lecture Notes in Mathematics 471, Springer Verlag, Berlin (1975).

- [H] W. R. HAMILTON - letter to P. G. Tait, October 15th, 1858.
- [H1] N. J. HITCHIN - *Polygons and gravitons*, Math. Proc. Camb. Phil. Soc. **85** (1979) 465-476.
- [H2] N. J. HITCHIN - *Linear field equations on self-dual spaces*, Proc. Roy. Soc. Lond. **A370** (1980) 173-191.
- [H3] N. J. HITCHIN - *On the construction of monopoles*, Commun. Math. Phys. **89** (1983) 145-190.
- [H4] N. J. HITCHIN - *Twistor construction of Einstein metrics*, in "Global Riemannian geometry" (eds. T. J. Willmore and N. J. Hitchin), Ellis Horwood, Chichester (1984)
- [H5] N. J. HITCHIN - *The self-duality equations on a Riemann surface*, Proc. London Math. Soc. **55** (1987) 59-126.
- [HKLR] N. J. HITCHIN, A. KARLHEDE, U. LINDSTRÖM and M. ROČEK - *Hyperkähler metrics and supersymmetry*, Commun. Math. Phys. **108** (1987) 535-59.
- [J] D. JOYCE - *The hypercomplex quotient and the quaternionic quotient*, Math. Ann. **290** (1991) 323-340.
- [Ki] F. KIRWAN - *Cohomology of quotients in symplectic and algebraic geometry*, Princeton University Press, Princeton (1985).
- [K1] P. B. KRONHEIMER - *Instantons gravitationelles et singularités de Klein*, C. R. Acad. Sci. Paris Sér. 1 Math. **303** (1986) 53-55.
- [K2] P. B. KRONHEIMER - *The construction of ALE spaces as hyperkähler quotients*, J. Diff. Geometry **29** (1989) 665-683.
- [K3] P. B. KRONHEIMER - *A hyperkählerian structure on coadjoint orbits of a semisimple Lie group*, J. London Math. Soc. **42** (1990) 193-208.
- [Ma] A. MACIOCIA - *Metrics on the moduli spaces of instantons over Euclidean 4-space*, Commun. Math. Phys. **135** (1991) 467-482.
- [McK] J. MCKAY - *Graphs, singularities and finite groups*, Proc. Sympos. Pure Math. Vol. 37, American mathematical Society, Providence (1980) 183-186.
- [N] H. NAKAJIMA - *Monopoles and Nahm's equations*, preprint (Tokyo) (1991).
- [NS] M. S. NARASIMHAN and C. S. SESHADRI - *Stable and unitary vector bundles on a compact Riemann surface*, Ann. of Math. **82** (1965) 540-

567.

- [P] R. PENROSE - *Nonlinear gravitons and curved twistor theory*, Gen. Relativ. Grav. **7** (1976) 31-52.
- [PS] A. PRESSLEY and G. B. SEGAL - "*Loop groups*", Oxford University Press, Oxford (1986).
- [S] S. SALAMON - *Differential geometry of quaternionic manifolds*, Ann. Sci. Éc. Norm. Sup. Paris **19** (1986) 31-55.
- [Y] S. T. YAU - *On Calabi's conjecture and some new results in algebraic geometry*, Proc. Nat. Acad. Sci. USA **74** (1977) 1798-1799.

Nigel HITCHIN

Mathematics Institute

University of Warwick

GB-COVENTRY CV4 7AL

Astérisque

JEAN-PIERRE SERRE

Revêtements de courbes algébriques

Astérisque, tome 206 (1992), Séminaire Bourbaki,
exp. n° 749, p. 167-182

<http://www.numdam.org/item?id=SB_1991-1992__34__167_0>

© Société mathématique de France, 1992, tous droits réservés.

L'accès aux archives de la collection « Astérisque » (<http://smf4.emath.fr/Publications/Asterisque/>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

REVÊTEMENTS DE COURBES ALGÈBRIQUES

par Jean-Pierre SERRE

1. INTRODUCTION

1.1. Le problème

Soit k un corps algébriquement clos. Soit C une courbe algébrique sur k , supposée irréductible et lisse. On a :

$$C = \overline{C} - S,$$

où \overline{C} est une courbe projective lisse, et S un sous-ensemble fini de $\overline{C}(k)$.

Quels sont les revêtements galoisiens (non ramifiés) de C ?

On peut préciser cette question de deux façons :

(a) On se donne un groupe fini G . On demande à quelle condition il existe un revêtement galoisien connexe $C' \rightarrow C$ de groupe G .

(b) Soit $\pi_C = \pi_1^{\text{alg}}(C, x)$ le groupe fondamental (algébrique) de la courbe C relativement à un point-base x . Ce groupe est un groupe profini. On demande de déterminer sa structure.

Remarque.— 1) Noter que (b) est plus précis que (a) : un groupe fini G satisfait à la condition (a) si et seulement si G est isomorphe à un quotient de π_C .

2) Lorsque l'on remplace le corps de base k par une extension algébriquement close k' , le groupe π_C ne change pas, si la caractéristique est 0. Il n'est va plus de même en caractéristique $p > 0$: il existe des familles non

constantes de revêtements. Toutefois, les quotients finis de π_C sont les mêmes sur k et sur k' , comme le montre un argument de spécialisation. La question (a) est donc “indépendante” du corps de base choisi (pourvu, bien sûr, qu’il soit algébriquement clos).

1.2. Le cas complexe

Lorsque $k = \mathbf{C}$, le *théorème d’existence de Riemann* dit que tout revêtement fini (au sens topologique) de la surface $C(\mathbf{C})$ possède une structure algébrique et une seule compatible avec sa projection sur $C(\mathbf{C})$ (cf. par exemple [5], chap. VI). Cela permet de répondre aux questions (a) et (b) ci-dessus. Pour (b), la réponse est la suivante : si g est le genre de \overline{C} , et $s = |S|$, le groupe π_C est le *complété profini* (i.e. le complété pour la topologie des sous-groupes d’indice fini) du groupe défini par $2g + s$ générateurs a_i, b_i ($i = 1, \dots, g$) et c_j ($j = 1, \dots, s$) liés par la relation

$$a_1 b_1 a_1^{-1} b_1^{-1} \cdots a_g b_g a_g^{-1} b_g^{-1} c_1 \cdots c_s = 1.$$

Lorsque $S \neq \emptyset$ (i.e. lorsque C est affine), ce groupe est un groupe libre de rang $2g + s - 1$. Dans ce cas, la réponse à la question (a) du n° 1.1 s’énonce très simplement : un groupe fini G convient si et seulement si il peut être engendré par $2g + s - 1$ éléments.

(La théorie de Riemann donne en fait un résultat plus précis : les éléments c_1, \dots, c_s peuvent être choisis de telle sorte qu’ils engendrent des groupes d’inertie au-dessus des points de S .)

1.3. Le cas de caractéristique 0

Lorsque la caractéristique de k est 0, le principe de Lefschetz, combiné avec des arguments de spécialisation, montre que *les résultats du n° 1.2 restent valables sans changement*.

Noter que, bien que ces résultats s’énoncent algébriquement, la seule démonstration que l’on en ait repose sur la théorie transcendante pour $k = \mathbf{C}$ (le point essentiel est de montrer qu’une surface de Riemann compacte a suffisamment de fonctions méromorphes).

Une tentative pour attaquer “algébriquement” la classification des revêtements avait été faite par Weil en 1938 (cf. [21], p. 84-86, ainsi que V.

Nori, *Comp. Math.* **33** (1976), p. 29-41). Son point de départ est le suivant : tout revêtement galoisien $C' \rightarrow C$ de groupe G , donne naissance (par l'intermédiaire des représentations linéaires de G) à des *fibrés vectoriels* sur C , qui ont la propriété de satisfaire à des équations algébriques vis-à-vis de la somme directe et du produit tensoriel. Si l'on pouvait expliciter la structure des *variétés de modules* de fibrés vectoriels avec assez de précision, on pourrait en déterminer les éléments algébriques et remonter de là aux revêtements galoisiens. Cette approche "tannakienne" du problème est intéressante, mais n'a jusqu'à présent abouti à rien de concret ; même le cas élémentaire $g = 1$, $s = 0$, semble difficile à traiter par cette méthode.

1.4. Le cas de caractéristique $p > 0$

Ce cas est resté longtemps inexploré. Voici ce qu'en disait Weil en 1946, dans [22] :

"... avant d'aborder la détermination des extensions de corps de nombres par leurs propriétés locales, il conviendrait peut-être de résoudre le problème analogue, déjà fort difficile, au sujet des fonctions algébriques d'une variable sur un corps de base fini, c'est-à-dire d'étendre à ces fonctions les théorèmes d'existence de Riemann. Pour ne citer qu'un cas particulier, le groupe modulaire, dont la structure détermine les corps de fonctions d'une variable complexe ramifiés en trois points seulement, joue-t-il le même rôle, tout au moins en ce qui concerne les extensions de degré premier à la caractéristique, quand le corps de base est fini ? Il n'est pas impossible que toutes les questions de ce genre puissent se traiter par une méthode uniforme, qui permettrait, d'un résultat une fois établi (par exemple par voie topologique) pour la caractéristique 0, de déduire le résultat correspondant pour la caractéristique p ; la découverte d'un tel principe constituerait un progrès de la plus grande importance..."

Ce texte suggérait deux choses :

(i) Pour les groupes d'ordre premier à p , la théorie est la même qu'en caractéristique zéro;

(ii) Il doit être possible de passer de la caractéristique zéro à la caractéristique p (et inversement).

L'assertion (i) a été précisée en 1956 par Abhyankar ([1]) sous la forme de la conjecture suivante :

(i') Soit π'_C le plus grand quotient de π_C dont l'ordre (comme groupe profini) soit premier à p . Alors π'_C est isomorphe au groupe correspondant pour une courbe sur \mathbf{C} ayant mêmes invariants g et s .

2. THÉORÈMES DE GROTHENDIECK ET CONJECTURES D'ABHYANKAR

2.1. Les théorèmes de Grothendieck

L'un des premiers succès de la théorie des schémas de Grothendieck a été la démonstration, en 1958, de la conjecture (i') du n° 1.4 (*cf.* [7], p. 182-27 ainsi que [8], p. 392) ; sa méthode consiste à se ramener à la caractéristique 0 conformément à ce qu'avait prédit Weil.

De façon plus précise, choisissons un anneau de valuation discrète complet A , de corps résiduel k , et dont le corps des fractions K est de caractéristique 0. On procède en trois étapes :

(2.1.1) On peut "relever" la k -courbe projective \overline{C} en un schéma \overline{C}_A projectif et lisse sur A (ce qui donne une courbe \overline{C}_K sur K ayant bonne réduction).

Ce résultat est presque évident lorsque g est < 5 , mais il ne l'est pas pour des valeurs plus grandes de g . Grothendieck le démontre en relevant d'abord \overline{C} en un *schéma formel*, puis en montrant que ce schéma est algébrique puisqu'il possède un fibré de rang 1 qui est ample (à savoir le fibré défini par un point). La démonstration utilise les théorèmes de comparaison "formel \iff algébrique", *i.e.* "GAGA formel".

(2.1.2) Une fois choisi \overline{C}_A , soit S_K un relèvement de S dans $\overline{C}_A(A) = \overline{C}_K(K)$. Si $C' \rightarrow C$ est un revêtement galoisien de C , de groupe G , qui est *modéré* en tous les points de S , on peut relever C' de façon unique en un revêtement de $C_K - S_K$: cela se démontre en utilisant encore "GAGA formel".

(2.1.3) Inversement, soit $C'_K \rightarrow \overline{C}_K - S_K$ un revêtement galoisien absolument irréductible de groupe de Galois G , et supposons que l'ordre de G soit premier à p . On montre alors que, après remplacement éventuel de A par un anneau plus ramifié, le revêtement C'_K "se réduit bien", *i.e.* se prolonge

en un revêtement de $\overline{C}_A - S_K$, et définit donc un revêtement de C . (La démonstration utilise le *lemme d'Abhyankar* ainsi que le *théorème de pureté*.)

En combinant ces résultats, on voit que :

(*) Si p ne divise pas $|G|$, les G -revêtements *sont les mêmes* en caractéristique p et en caractéristique 0, cf. (2.1.2) et (2.1.3).

C'est là un résultat très satisfaisant — à cela près que la condition “ p ne divise pas $|G|$ ” est très restrictive : si $p = 2$, elle n'est satisfaite que par des groupes résolubles, d'après Feit-Thompson.

(**) Si p divise $|G|$, mais si l'on se borne aux revêtements *modérément ramifiés* en tous les points de S , il y a au plus autant (et en général strictement moins) de tels revêtements en caractéristique p qu'en caractéristique 0.

(Pour des formulations plus précises de (*) et (**) en termes de *spécialisation du groupe fondamental*, voir [8].)

2.2. Le cas où C est une courbe complète

C'est le cas $S = \emptyset$. Tous les revêtements de C sont modérés. D'après (**) ci-dessus, le groupe π_C est un quotient du groupe fondamental correspondant en caractéristique 0. *Quel est ce quotient ?* On ne sait le déterminer explicitement que si $g = 0$ ou 1, auquel cas il est commutatif :

pour $g = 0$, on a $\pi_C = \{1\}$;

pour $g = 1$, π_C est isomorphe à $\prod_{\ell \neq p} (\mathbf{Z}_\ell \times \mathbf{Z}_\ell)$ ou à $\mathbf{Z}_p \times \prod_{\ell \neq p} (\mathbf{Z}_\ell \times \mathbf{Z}_\ell)$, suivant que la courbe C est supersingulière ou ordinaire.

Pour $g \geq 2$, on n'a pas de théorème de structure, même conjecturalement. On a seulement des renseignements cohomologiques. On les obtient en remarquant que la cohomologie de π_C est isomorphe à la cohomologie étale de C , tout comme si le “revêtement universel” de C était contractile (cela se démontre en utilisant la suite spectrale de Cartan-Leray pour les revêtements, cf. [14], p. 105, th. 2.20). D'où, d'après [4], exposés IX, X :

$$\mathrm{cd}_p(\pi_C) = 1 \text{ et } \dim H^1(\pi_C, \mathbf{Z}/p\mathbf{Z}) \leq g ;$$

$$\mathrm{cd}_\ell(\pi_C) = 2, \dim H^1(\pi_C, \mathbf{Z}/\ell\mathbf{Z}) = 2g \text{ et } \dim H^2(\pi_C, \mathbf{Z}/\ell\mathbf{Z}) = 1 \text{ si } \ell \neq p.$$

(Rappelons que $\mathrm{cd}_p(\pi)$ désigne la *p-dimension cohomologique* du groupe profini π , cf. [18], p. I-17. L'assertion $\mathrm{cd}_p(\pi_C) \leq 1$ équivaut à dire que les

p -groupes de Sylow de π_C sont des *pro- p -groupes libres* ; ces groupes sont non triviaux d'après Raynaud [16], cor. 4.3.2).

2.3. Le cas où C est une courbe affine : conjecture d'Abhyankar

C'est le cas $S \neq \emptyset$; sur \mathbf{C} , le groupe fondamental correspondant est libre de rang $2g + s - 1$, cf. n° 1.2.

Soit G un groupe fini. Soit $p(G) = O^{p'}(G)$ le sous-groupe de G engendré par ses p -groupes de Sylow. Le groupe $G/p(G)$ est le plus grand quotient de G d'ordre premier à p . Si G est quotient de π_C , il en est de même de $G/p(G)$ et, d'après Grothendieck (cf. (2.1.2)), $G/p(G)$ est quotient du groupe analogue à π_C en caractéristique 0, autrement dit peut être engendré par $2g + s - 1$ éléments. Dans [1], Abhyankar conjecture que cette condition est suffisante :

Conjecture 2.3.1.— *Un groupe fini G est quotient de π_C si et seulement si le quotient $G/p(G)$ peut être engendré par $2g + s - 1$ éléments.*

Si cette conjecture était vraie, elle donnerait une réponse satisfaisante à la question (a) du n° 1.1 (mais pas à la question (b)).

Remarque.— Ici encore, la cohomologie de π_C s'identifie à la cohomologie étale de C , et l'on en déduit :

$$\mathrm{cd}_\ell(\pi_C) = 1 \text{ pour tout } \ell, \text{ et } \dim H^1(\pi_C, \mathbf{Z}/\ell\mathbf{Z}) = \begin{cases} \mathrm{Card}(k) & \text{si } \ell = p \\ 2g + s - 1 & \text{sinon.} \end{cases}$$

En particulier (cf. [18], p. I-74), π_C possède la *propriété de relèvement* : si $f : \tilde{G} \rightarrow G$ est un homomorphisme surjectif de groupes finis, et si φ est un homomorphisme de π_C dans G , il existe $\tilde{\varphi} : \pi_C \rightarrow \tilde{G}$ tel que $f \circ \tilde{\varphi} = \varphi$. (Mais si φ est surjectif, on ne peut pas toujours choisir $\tilde{\varphi}$ surjectif, même si \tilde{G} satisfait à la condition de (2.3.1).)

3. LA CONJECTURE D'ABHYANKAR POUR LA DROITE AFFINE

3.1. Conjecture et résultats

On se restreint maintenant au cas où C est la droite affine $D = \text{Spec } k[X]$, le corps k étant algébriquement clos de caractéristique $p > 0$.

On a alors $g = 0$ et $s = 1$, d'où $2g + s - 1 = 0$, et la conjecture 2.3.1 prend la forme suivante :

Conjecture 3.1.1.— *Un groupe fini G est quotient de π_D si et seulement si l'on a $G = p(G)$, i.e. si et seulement si G est engendré par ses p -groupes de Sylow.*

Un groupe G tel que $G = p(G)$ sera appelé un *quasi- p -groupe*.

Exemple.— Un groupe simple dont l'ordre est divisible par p est un quasi- p -groupe. La conjecture (3.1.1) implique donc que tout groupe simple non abélien est quotient de π_D si $p = 2$.

On verra ci-après que (3.1.1) a été démontrée dans de nombreux cas particuliers :

G est le groupe des points rationnels d'un groupe algébrique semi-simple simplement connexe sur un corps fini de caractéristique p (Nori [10]) ;

G est le groupe alterné A_n , $n \geq p > 2$, ou le groupe symétrique S_n , $n \geq p = 2$ (Abhyankar [2]) ;

G est résoluble ([19]) ;

G est engendré par des sous-groupes vérifiant certaines conditions restrictives (Harbater [9], Raynaud [17]).

Remarque.— La droite D joue un rôle en quelque sorte universel pour les revêtements de variétés affines. De façon plus précise, soit V une variété affine irréductible sur k , de dimension > 0 , et soit G un groupe fini qui soit quotient de π_D . Il existe alors un revêtement galoisien connexe $V' \rightarrow V$ de groupe de Galois G , cf. [19], n° 2.

3.2. Les exemples de Nori

Soit q une puissance de p , et soit Σ un groupe algébrique semi-simple simplement connexe sur \mathbf{F}_q . Soit $G = \Sigma(\mathbf{F}_q)$ le groupe des \mathbf{F}_q -points de Σ .

Le groupe G est un quasi- p -groupe.

THÉORÈME 3.2.1 (Nori).— *Le groupe G est quotient de π_D .*

COROLLAIRE.— *La conjecture 3.1.1 est vraie pour les groupes $\mathrm{SL}_n(\mathbf{F}_q)$, $\mathrm{Sp}_{2n}(\mathbf{F}_q)$, \dots , $E_8(\mathbf{F}_q)$.*

Démonstration (d'après G. Laumon).— Soient B^+ et B^- des sous-groupes de Borel opposés de Σ , et soient U^+ et U^- leurs radicaux unipotents. L'application produit $U^+ \times U^- \rightarrow \Sigma$ est un isomorphisme de $U^+ \times U^-$ sur une sous-variété fermée V de Σ (le fait que V soit fermée n'interviendra d'ailleurs pas). Soit $f : \Sigma \rightarrow \Sigma$ l'isogénie de Lang, définie par :

$$f(x) = x^{-1} \cdot F(x),$$

où F est l'endomorphisme de Frobenius de Σ relativement à \mathbf{F}_q . Cette isogénie définit un revêtement galoisien connexe de Σ , de groupe de Galois G . Soit $W = f^{-1}(V)$. La variété W est connexe. En effet, soit W^0 la composante connexe de W contenant 1, et soit G^0 le sous-groupe de G qui stabilise W^0 . Comme W^0 contient U^+ , le groupe G^0 contient $U^+(\mathbf{F}_q)$; de même, il contient $U^-(\mathbf{F}_q)$. Mais on sait (cf. [20], lemme 64) que G est engendré par $U^+(\mathbf{F}_q)$ et $U^-(\mathbf{F}_q)$. On a donc $G = G^0$, ce qui montre que W est connexe. On a donc obtenu un revêtement galoisien connexe $W \rightarrow V$ de groupe de Galois G . Mais V est isomorphe à $U^+ \times U^-$, donc est un espace affine de dimension > 0 (sauf si $\Sigma = \{1\}$, auquel cas il n'y a rien à démontrer). En restreignant le revêtement $W \rightarrow V$ à une droite Δ de V , on obtient un revêtement de Δ de groupe de Galois G . D'après une variante du théorème de Bertini, ce revêtement est connexe si Δ est assez générale. D'où le résultat cherché, puisque Δ est isomorphe à D .

Remarque.— La démonstration s'applique aussi aux groupes "tordus" du type Suzuki et Ree en caractéristique 2 et 3.

3.3. Les exemples d'Abhyankar

THÉORÈME 3.3.1 (Abhyankar [2]).— *Les groupes ci-dessous sont quotients de π_D :*

le groupe alterné A_n si $p \neq 2$ et $n \geq p$;

le groupe symétrique S_n si $p = 2$.

La méthode consiste à écrire explicitement des équations de degré n qui définissent des revêtements étales de degré n de D , et à montrer que les clôtures galoisiennes de ces revêtements ont pour groupe de Galois un sous-groupe de S_n qui est A_n ou S_n suivant les cas ; cela se fait en prouvant que le groupe en question est suffisamment transitif (dans certains cas, Abhyankar est amené à utiliser le "théorème de classification" des groupes finis simples).

L'une des équations qu'il utilise est la suivante :

$$(I) \quad Y^n - X \cdot Y^t + 1 = 0,$$

avec $n = p + t$, $t \geq 1$ et $t \not\equiv 0 \pmod{p}$.

Il montre (cf. [2], [3]) que le groupe de Galois de cette équation est :

$\mathbf{PSL}_2(\mathbf{F}_p)$	si	$t = 1$;
$\mathbf{SL}_2(\mathbf{F}_8)$	si	$t = 2, p = 7$;
A_n	si	$t = 2, p \neq 2, 7$;
A_n	si	$t > 2, p \neq 2$;
S_n	si	$p = 2$.

Donnons la démonstration dans le cas le plus simple, qui est celui où $t \geq 3$. Soit G le groupe de Galois de (I). C'est un sous-groupe de S_n jouissant des propriétés suivantes :

il est transitif ;

il contient un cycle c d'ordre p (à cause de l'inertie en $X = \infty, Y = \infty$) ;

il contient un élément d'ordre premier à p qui permute circulairement les t points fixes du cycle c (inertie en $X = \infty, Y = 0$).

Comme $n = p + t$, et $(p, t) = 1$, on vérifie facilement que les propriétés ci-dessus entraînent que G est *primitif*. Comme G contient un cycle d'ordre

premier $\leq n - 3$, on a $G = S_n$ ou A_n d'après un théorème de Jordan (cf. [23], p.39). Si $p = 2$, le fait que G contienne une transposition entraîne $G = S_n$. Si $p \neq 2$, le fait que soit un quasi- p -groupe entraîne $G \neq S_n$, donc $G = A_n$, d'où le résultat cherché.

Remarques.— 1) La méthode employée ne semble pas donner le cas de A_n , $n \geq 5$, lorsque $p = 2$. Toutefois Raynaud m'a fait observer que ce cas peut se traiter par "recollement" de groupes A_5 en utilisant le th. 1.1 de [17] (voir n° 3.5).

2) Certaines équations explicites, conduisant à des groupes de Galois intéressants, peuvent s'obtenir par réduction (mod p) à partir d'équations connues en caractéristique zéro. Par exemple :

- (a) On sait depuis Fricke (cf. [6], [24]) qu'il existe un revêtement galoisien de la droite projective de groupe $G = \mathbf{SL}_2(\mathbf{F}_8)$, ramifié en trois points avec ramification d'ordre 2, 3, 7, et défini sur le corps $\mathbf{Q}(\cos(2\pi/7))$; la courbe correspondante est de genre 7. Si l'on représente G comme sous-groupe transitif de A_9 , cela conduit à un revêtement de degré 9 de la droite projective dont l'équation a été écrite par Goursat (cf. [6]). En réduisant cette équation en caractéristique 7 (après des changements de variables convenables), on obtient l'équation $Y^9 - X \cdot Y^2 + 1 = 0$ d'Abhyankar, et l'on retrouve ainsi le résultat principal de [3].
- (b) En réduisant en caractéristique 11 une équation donnée par Matzat [13], on voit que l'équation

$$Y^{11} + 2Y^9 + 3Y^8 - X^8 = 0 \quad (p = 11)$$

définit un revêtement étale de degré 11 de D dont le groupe de Galois est le groupe de Mathieu M_{11} .

- (c) En caractéristique 23, l'équation :

$$Y(Y - 1)^2 (Y + 1)^4 (Y^2 + 17Y + 4)^8 - X^8 = 0 \quad (p = 23)$$

définit un revêtement étale de degré 23 de D . Il semble que le groupe de Galois de ce revêtement soit le groupe de Mathieu M_{23} ; il serait intéressant de le démontrer.

3.4. Extensions par des groupes résolubles

Soit $1 \rightarrow N \rightarrow \tilde{G} \rightarrow G \rightarrow 1$ une suite exacte de groupes finis. Supposons que \tilde{G} soit un *quasi- p -groupe*, et que N soit *résoluble*. Alors, si la conjecture 3.1.1 est vraie pour G , elle l'est pour \tilde{G} . Autrement dit :

THÉORÈME 3.4.1 (cf. [19]).— *Sous les hypothèses ci-dessus, si G est quotient de π_D , il en est de même de \tilde{G} .*

COROLLAIRE 1.— *La conjecture 3.1.1 est vraie pour les groupes résolubles : tout quasi- p -groupe résoluble est quotient de π_D .*

C'est le cas particulier $G = \{1\}$.

COROLLAIRE 2.— *Si G est quotient de π_D , il en est de même de toute extension de G par un p -groupe.*

En effet, une telle extension est un quasi- p -groupe, ce qui permet d'appliquer le th. 3.4.1.

Démonstration de (3.4.1).— On se ramène par dévissage au cas où N est un groupe abélien élémentaire de type (ℓ, \dots, ℓ) , avec ℓ premier, l'action de G sur N étant irréductible.

Par hypothèse, il existe un morphisme surjectif $\varphi : \pi_D \rightarrow G$, et, d'après ce qui a été dit au n° 2.3, on peut relever φ en $\tilde{\varphi} : \pi_D \rightarrow \tilde{G}$. Lorsque \tilde{G} est une extension non scindée de G , $\tilde{\varphi}$ est surjectif, et le théorème est démontré. Le cas où \tilde{G} est produit semi-direct de G par N est plus délicat. Si $\ell = p$, on montre que l'on peut choisir $\tilde{\varphi}$ de telle sorte qu'il soit surjectif. Si $\ell \neq p$, ce n'est pas possible en général. Toutefois, dans ce cas, on peut modifier φ de façon à ce que cela devienne possible. De façon plus précise, choisissons un entier $m > 1$ premier à p . Le morphisme $D \rightarrow D$ défini par $X \mapsto X^m$ induit un endomorphisme $f_m : \pi_D \rightarrow \pi_D$ (on prend pour point-base le point 0). Cet endomorphisme est surjectif. Si on le compose avec φ , on obtient un homomorphisme $\varphi_m : \pi_D \rightarrow G$ qui est encore surjectif, mais qui est "plus ramifié" que φ (son invariant de Swan à l'infini est multiplié par m). On montre qu'il est possible de relever φ_m en un morphisme surjectif $\tilde{\varphi}_m : \pi_D \rightarrow \tilde{G}$ (la démonstration de ce fait repose sur une étude des modules

galoisiens fournis par l'homologie (mod ℓ) des revêtements, cf. [19], th. 2). D'où le théorème.

3.5. Les constructions de Harbater et Raynaud

Remarquons d'abord que la conjecture d'Abhyankar 3.1.1 résulterait (en raisonnant par récurrence sur $|G|$ et en se ramenant au cas où G est cyclique) de l'énoncé suivant :

(3.5.1 γ) *Si G est engendré par deux sous-groupes G_1 et G_2 qui sont quotients de π_D , alors G est quotient de π_D .*

Harbater [9] et Raynaud [17] démontrent (3.5.1 γ) sous diverses hypothèses restrictives. Par exemple :

THÉORÈME 3.5.2 ([9], th. 4 (i)).— *L'assertion (3.5.1 γ) est vraie si G_1 est un p -groupe contenant un p -groupe de Sylow de G_2 .*

L'énoncé de Raynaud fait intervenir les groupes d'inertie à l'infini des revêtements considérés (groupes qui sont définis à conjugaison près) :

THÉORÈME 3.5.3 ([17], th. 1.1).— *Soit Q un p -sous-groupe de G . Supposons que, pour $i = 1, 2$, il existe un revêtement galoisien connexe $\tilde{D}_i \rightarrow D_i = D$, de groupe G_i , ayant comme groupe d'inertie à l'infini un sous-groupe Q_i de Q . Il existe alors un revêtement galoisien connexe de D de groupe G ayant Q comme groupe d'inertie à l'infini.*

Les démonstrations de (3.5.2) et (3.5.3) utilisent respectivement "GAGA formel" et "GAGA rigide" — ce qui d'ailleurs revient à peu près à la même chose, cf. [15].

Voici un résumé (très incomplet) de la méthode esquissée par Raynaud dans [17] :

On utilise la géométrie rigide sur le corps local $K = k((T))$. Dans la droite projective \mathbf{P}_1 sur K , on choisit deux disques fermés Δ_1 et Δ_2 , ne contenant pas ∞ , et à distance > 0 l'un de l'autre (par exemple les disques $|z| \leq 1/p$ et $|z - 1| \leq 1/p$). On choisit des disques un peu plus grands Δ_i^+ les contenant, et satisfaisant aux mêmes conditions. On définit ensuite trois revêtements galoisiens (au sens rigide), de groupe G :

(a₁) Un revêtement de Δ_1^+ ;

(a₂) Un revêtement de Δ_2^+ ;

(a₃) Un revêtement de $\mathbf{P}_1 - \Delta_1 - \Delta_2 - \{\infty\}$, qui se prolonge en un revêtement ramifié à l'infini, de groupe d'inertie Q .

La construction des revêtements (a₁) et (a₂) se fait de la manière suivante : la "réduction mod T " du disque Δ_i peut être identifiée à la k -droite affine D_i . Le G_i -revêtement $\tilde{D}_i \rightarrow D_i$ définit par relèvement à K un G_i -revêtement rigide de Δ_i , et l'on montre que ce revêtement se prolonge à un disque un peu plus grand. On obtient ainsi un G_i -revêtement, et par induction de G_i à G on obtient le revêtement (a_i) cherché. (Noter que ce revêtement n'est pas irréductible en général : ses composantes irréductibles sont en nombre égal à l'indice de G_i dans G .)

Quant à (a₃), on l'obtient par induction de Q à G à partir d'un Q -revêtement (algébrique) de $\mathbf{P}_1 - \{\infty\}$. On le choisit (cf. [17]) de telle sorte que la propriété suivante soit satisfaite :

(b) Pour $i = 1, 2$, la restriction du revêtement (a₃) à la couronne ouverte $\Delta_i^+ - \Delta_i$ est isomorphe à la restriction de (a_i) à cette couronne (du moins si les Δ_i^+ sont assez petits).

Cette propriété permet de *recoller* (au sens de la géométrie rigide) les trois revêtements (a_i) et l'on obtient ainsi un revêtement galoisien de \mathbf{P}_1 , de groupe G , qui est ramifié seulement à l'infini, avec Q pour groupe d'inertie. Par "GAGA rigide" (cf. [11], [12]), ce revêtement est algébrique. De plus, sa construction montre qu'il est géométriquement irréductible (c'est là que sert l'hypothèse que G est engendré par G_1 et G_2). *A priori*, ce revêtement est défini sur K ; par spécialisation (cf. n° 1.1), on en déduit un revêtement sur k ayant les propriétés voulues.

COROLLAIRE 3.5.4.— *Si G est groupe de Galois d'un revêtement connexe de D , ce revêtement peut être choisi tel que les groupes d'inertie à l'infini soient les p -groupes de Sylow de G .*

Soit I un groupe d'inertie à l'infini. En utilisant le lemme d'Abhyankar (*i.e.* un changement de base $X \mapsto X^n$, avec n convenable), on peut supposer que I est un p -groupe. Soit P un p -groupe de Sylow de G contenant I . On applique alors le th. 3.5.3, avec $G_1 = G$, $G_2 = P$, $Q_1 = I$, $Q_2 = P$, $Q = P$ (c'est licite car on sait que tout p -groupe fini est quotient de π_D , cf. par

exemple [19]).

Compte tenu de ce résultat, on voit que (3.5.2) est un cas particulier de (3.5.3).

Compléments

Du th. 3.5.3, Raynaud déduit le résultat suivant :

Soit G un groupe fini qui soit un quasi- p -groupe, et soit P un p -groupe de Sylow de G . Il existe alors un sous-groupe H de G , contenant P , quotient de π_D , et *maximum* pour cette propriété. En particulier, H est stable par tout automorphisme de G préservant P . De plus, tout sous-groupe de G qui est quotient de π_D est conjugué d'un sous-groupe de H . Enfin, Raynaud démontre ([17], th. 1.5) que, si $P \neq G$, on a $H \neq P$ (ce qui permet souvent de prouver que $H = G$, autrement dit que la conjecture d'Abhyankar est vraie pour G).

Application

Indiquons à titre d'exemple comment on peut déduire du th. 3.5.3 le fait que *le groupe alterné A_n , $n \geq 5$, est quotient de π_D si $p = 2$.*

Soit P le 2-groupe de Sylow de A_4 . On va montrer par récurrence sur $n \geq 5$ qu'il existe un revêtement galoisien connexe de D , de groupe A_n , pour lequel les groupes d'inertie à l'infini sont les conjugués de P . C'est vrai pour $n = 5$ car $A_5 = \mathbf{SL}_2(\mathbf{F}_4)$ et l'on applique 3.2.1 et 3.5.4 (on peut aussi, plus simplement, utiliser l'équation $Y^5 + X^3Y + 1 = 0$). Si $n \geq 6$, on applique 3.5.3 au groupe $G = A_n$, avec :

$$G_1 = A_{n-1}, \text{ fixateur de } n \text{ dans } A_n ;$$

$$Q_1 = P ;$$

$$G_2 = \text{fixateur de } n - 1 \text{ dans } A_n \text{ (on a } G_2 \simeq A_{n-1}\text{)};$$

$$Q_2 = P ;$$

$$Q = P.$$

Vu l'hypothèse de récurrence, toutes les conditions nécessaires sont satisfaites.

BIBLIOGRAPHIE

- [1] S. ABHYANKAR - *Coverings of algebraic curves*, Amer. J. Math. **79** (1957), 825-856.
- [2] S. ABHYANKAR - *Galois theory on the line in nonzero characteristic*, Bull. A.M.S. **27** (1992), 68-133.
- [3] S. ABHYANKAR - *Square-root parametrization of plane curves*, Purdue Univ., 1991.
- [4] M. ARTIN, A. GROTHENDIECK et J.-L. VERDIER - *Théorie des Topos et Cohomologie Étale des Schémas* (SGA 4), vol. 3, Lect. Notes in Math. **305**, Springer-Verlag (1973).
- [5] R. et A. DOUADY - *Algèbre et théories galoisiennes ; 2) théories galoisiennes*, CEDIC, F. Nathan, Paris (1979).
- [6] R. FRICKE - *Ueber eine einfache Gruppe von 504 Operationen*, Math. Ann. **52** (1899), 321-339.
- [7] A. GROTHENDIECK - *Géométrie formelle et géométrie algébrique*, Sémin. Bourbaki, exposé n° 182, volume 1958/1959, Benjamin (1966).
- [8] A. GROTHENDIECK - *Revêtement étales et groupe fondamental* (SGA 1), Lect. Notes in Math. **224**, Springer-Verlag (1971).
- [9] D. HARBATER - *Formal patching and adding branch points*, prépublication (1991).
- [10] T. KAMBAYASHI - *Nori's construction of Galois coverings in positive characteristics*, Algebraic and Topological Theories, Tokyo (1985), 640-647.
- [11] R. KIEHL - *Der Endlichkeitssatz für eigentliche Abbildungen in der nicht-archimedischen Funktionentheorie*, Inv. Math. **2** (1967), 191-214.
- [12] U. KÖPF - *Über eigentliche Familien algebraischer Varietäten über affinoiden Räumen*, Schriftenreihe Univ. Münster, 2 Serie, Heft 7 (1974).
- [13] B.H. MATZAT - *Konstruktion von Zahlkörpern mit der Galoisgruppe M_{11} über $\mathbf{Q}(\sqrt{-11})$* , Man. Math. **27** (1979), 103-111.
- [14] J. MILNE - *Étale Cohomology*, Princeton Math. Series 33, Princeton (1980).

- [15] M. RAYNAUD - *Géométrie analytique rigide, d'après Tate, Kiehl, ...*, Bull. Soc. math. France, Mémoire 39-40 (Table ronde anal. non-archim. 1972), 319-327.
- [16] M. RAYNAUD - *Sections des fibrés vectoriels sur une courbe*, Bull. Soc. math. France **110** (1982), 103-125.
- [17] M. RAYNAUD - *Autour d'une conjecture d'Abhyankar*, manuscrit non publié (1991).
- [18] J.-P. SERRE - *Cohomologie galoisienne*, 4ème édit., Lect. Notes in Math. **5**, Springer-Verlag (1973).
- [19] J.-P. SERRE - *Construction de revêtements étales de la droite affine en caractéristique p* , C.R. Acad. Sci. Paris **311** (1990), série I, 341-346.
- [20] R. STEINBERG - *Lectures on Chevalley groups*, Notes prepared by John Faulkner and Robert Wilson, Yale University (1967).
- [21] A. WEIL - *Généralisation des fonctions abéliennes*, J. Liouville **17** (1938), 47-87 (= *Oe.* [1938a]).
- [22] A. WEIL - *L'avenir des mathématiques*, Les Grands Courants de la Pensée Mathématique, éd. F. Le Lionnais, Cahiers du Sud, 1947, 307-320 (= *Oe.* [1947a]).
- [23] H. WIELANDT - *Finite Permutation Groups*, Acad. Press, New York (1964).
- [24] K. WOHLFAHRT - *Macbeath's curve and the modular group*, Glasgow Math. J. **27** (1985), 239-247 ; *Corrigendum, ibid.* **28** (1986), 241.

(Note ajoutée en septembre 1992) La conjecture d'Abhyankar pour la droite affine (3.3.1) a été démontrée par M. Raynaud. Voir M. Raynaud, *Revêtements de la droite affine en caractéristique $p > 0$ et conjecture d'Abhyankar*, à paraître à Inv. Math.

Jean-Pierre SERRE

Collège de France

3, rue d'Ulm

F-75005 PARIS

Astérisque

DANIEL BERTRAND

Groupes algébriques et équations différentielles linéaires

Astérisque, tome 206 (1992), Séminaire Bourbaki,
exp. n° 750, p. 183-204

<http://www.numdam.org/item?id=SB_1991-1992__34__183_0>

© Société mathématique de France, 1992, tous droits réservés.

L'accès aux archives de la collection « Astérisque » (<http://smf4.emath.fr/Publications/Asterisque/>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

GROUPES ALGÉBRIQUES ET ÉQUATIONS DIFFÉRENTIELLES LINÉAIRES

par Daniel BERTRAND

Donnant, à l'aide du théorème de Dirichlet sur les progressions arithmétiques, une description partielle de la liste de Schwarz ([Sc]), Landau notait en 1904 : “*Es ist bemerkenswert, daß das vorliegende funktionentheoretische Problem, welches ja bekanntlich mit den verschiedensten geometrischen, analytischen, und algebraischen Sätzen in Zusammenhang gebracht werden kann, auch zur Anwendung eines rein Zahlentheoretischen Satzes Anlaß gibt ...*” ([La]).

Le problème en question consistait à dresser la liste des équations hypergéométriques de Gauss qui admettent une base de solutions algébriques, c'est-à-dire dont le groupe de monodromie est fini. Plus généralement, on peut, une équation différentielle linéaire à coefficients rationnels étant donnée, chercher à déterminer le degré de transcendance du corps engendré par ses solutions. En présence de singularités irrégulières, le groupe de monodromie ne contrôle plus ce degré, mais la théorie de Galois différentielle permet de remédier à cette difficulté. (De par sa nature algébrique, elle facilite d'ailleurs également les calculs dans le cas fuchsien.)

On sait [Bl] le développement qu'a connu cette théorie entre les mains de Kolchin. Plutôt que de ses généralisations (voir [Ko], [Po]) ou de ses applications (voir par exemple [Ne], [Sn1], [Um]), c'est de techniques récemment introduites, en particulier par J.-P. Ramis et par N. Katz, pour *calculer* explicitement les groupes de Galois différentiels d'équations classiques qu'il s'agira dans cet exposé. Nous n'aurons pour les passer en revue qu'à reprendre l'énumération de Landau : méthodes géométriques au §2, analytiques au §3, algébriques au §4, enfin arithmétiques au §5 (après un §1 de rappels).

1. DÉFINITIONS (cf. [De]).

Soient C un corps algébriquement clos de caractéristique nulle, (K, ∂) un corps différentiel de corps des constantes C , et V un espace vectoriel de dimension finie n sur K , muni d'une connexion⁽¹⁾. Contractée avec ∂ , celle-ci fournit sur V un opérateur différentiel D donné, dans une K -base de V , par $Y \rightarrow DY = \partial Y + AY$, où A est une matrice carrée d'ordre n à coefficients dans K . Pour toute extension différentielle (L, ∂) de (K, ∂) , les solutions dans L^n du système différentiel $DY = 0$ représentent les éléments de l'espace $(V \otimes L)^\partial$ des vecteurs horizontaux de la connexion $(V \otimes_K L, D \otimes \partial)$. On dit que L est une *extension de Picard-Vessiot* de K (associée à V) si les trois conditions suivantes sont réalisées : les corps des constantes de L et de K coïncident ; $(V \otimes L)^\partial \otimes_C L = V \otimes_K L$ (c'est-à-dire : $DY = 0$ admet une matrice fondamentale $U(D)$ de solutions dans L^n) ; enfin, L est engendré sur K par $(V \otimes L)^\partial$ (c'est-à-dire par les coefficients de $U(D)$). Pour une telle extension, on appelle *groupe de Galois différentiel* de L sur K (ou, par un abus de langage que justifie l'alinéa suivant, de V/K) le groupe $\text{Gal}_\partial(L/K)$ (ou $\text{Gal}(D)$) des automorphismes de l'extension L/K qui commutent à ∂ . Puisque D est linéaire, le C -espace vectoriel $V^\partial = (V \otimes L)^\partial$ en est une représentation fidèle. Comme l'aurait fait Galois, on peut interpréter les éléments de $\text{Gal}(D)$ comme ceux des automorphismes de V^∂ qui, pour toute construction tensorielle⁽²⁾ W du vectoriel à connexion V , laissent stables les espaces de vecteurs horizontaux des différents sous-espaces de W stables sous la connexion tensorielle correspondante (voir [Ka2], [De]). En particulier, $\text{Gal}(D)$ est un sous-groupe algébrique du groupe algébrique $\text{Aut}_C(V^\partial)$.

Le point de départ de la théorie de Kolchin est que, grâce à l'hypothèse fondamentale faite sur C , il existe, à isomorphisme différentiel près, une unique extension de Picard-Vessiot associée à V . Je renvoie à [Le] pour une démonstration concise de ce résultat. L'existence est un exercice d'algèbre différentielle. Le fait, plus délicat, que deux extensions de Picard-Vessiot L, L' soient liées par un isomorphisme (non unique) $i(L, L')$ découle de l'interprétation de ces isomorphismes comme C -points d'un torseur sous le C -groupe algébrique $\text{Gal}_\partial(L/K)$. On trouvera une extension (et une nouvelle démonstration) de ces résultats en termes de catégories tannakiennes dans [De]. Retenons de ce point de vue l'existence, triviale, d'un foncteur fibre sur K , qui permet de voir $\text{Gal}(D)$ comme une C -forme

d'un sous-groupe algébrique $\text{Gal}(V/K)$ du K -groupe algébrique $\text{Aut}_K(V)$, défini par les "mêmes" contraintes que supra, alors que tout comme la matrice $U(D)$, les isomorphismes $i(L, L')$ sont de nature transcendante. Les expliciter, c'est-à-dire, dans le langage des groupes de monodromie, rechercher des "formules de connexions", est donc difficile, mais c'est parfois la clef du calcul de $\text{Gal}(D)$. Nous en verrons des exemples aux §§ 2 et 3.

Ces divers résultats entraînent — en le précisant — l'énoncé suivant :

THÉORÈME 1.— i) *La correspondance de Galois usuelle établit une bijection entre les sous-groupes algébriques de $\text{Gal}(D)$ et les extensions différentielles intermédiaires de L/K , où les sous-groupes fermés normaux correspondent aux extensions de Picard-Vessiot de K .*

ii) *La catégorie des C -représentations rationnelles de dimension finie de $\text{Gal}(D)$ est équivalente à celle des connexions sous-quotients des constructions de V . En particulier, la représentation V^∂ de $\text{Gal}(D)$ est irréductible si et seulement si la connexion de V l'est sur K .*

iii) *Le degré de transcendance de L sur K est égal à la dimension de $\text{Gal}(D)$. Plus précisément, $U(D)$ est le point générique d'un K -torseur (K -irréductible) sous le groupe algébrique $G(V/K)$.*

2. AUTOUR DE LA MONODROMIE

On suppose ici que V provient d'un fibré à connexion holomorphe sur le complémentaire d'un ensemble fini Σ de points de la sphère de Riemann, et on désigne, pour tout élément σ de Σ , par $K_{\sigma,an}$ (resp. $K_{\sigma,f} = \widehat{K}$) le corps des fractions des germes de fonctions analytiques (resp. des séries formelles en σ). Soient $G_{\sigma,an}$ et $G_{\sigma,f}$ les groupes de Galois analytique et formel de V en σ , i.e. des connexions $V_{\sigma,an}$, \widehat{V} déduites de V par extension des scalaires de $K = \mathbf{C}(\mathbf{P}_1)$ à $K_{\sigma,an}$, puis à \widehat{K} . Ils s'identifient (voir [Ka3], 2.7) à des sous-groupes algébriques (non normaux, en général) de $\text{Gal}(D)$, et $G_{\sigma,an}$ contient l'adhérence de Zariski $M_{\sigma,an}$ du groupe de monodromie local de D en σ . On déduit du théorème 1, de l'identité entre fonctions méromorphes et fonctions rationnelles sur \mathbf{P}_1 , et de la théorie de Fuchs :

PROPOSITION 1.— *Pour $K = \mathbb{C}(z)$ le groupe de Galois $\text{Gal}(D)$ est engendré ⁽³⁾ (pour la topologie de Zariski) par ses sous-groupes analytiques locaux. De plus, $G_{\sigma,an}$, $G_{\sigma,f}$ et $M_{\sigma,an}$ se confondent lorsque σ est une singularité régulière de D .*

(Rappelons qu'une singularité σ est dite régulière si toute section horizontale est à croissance au plus polynômiale au voisinage de σ). En particulier, si Σ ne contient que des singularités régulières, $\text{Gal}(D)$ est la clôture de Zariski "du" groupe de monodromie attaché au fibré à connexion dont provient V .

Exemple 1.— L'équation hypergométrique de Gauss $H = H(a, b, c)$:

$$z(1-z)\partial^2 y + (c - (a+b+1)z)\partial y - aby = 0 \quad (\partial = d/dz),$$

[ou encore : $(\theta + a)(\theta + b)y - z^{-1} \theta(\theta + c - 1)y = 0$ ($\theta = z\partial$)].

Ses trois singularités $0, \infty, 1$ sont régulières. La méthode de Frobenius y donne donc les sous-groupes analytiques locaux. Or on connaît depuis Riemann une "formule de connexion" permettant de les comparer dans une même réalisation de $\text{Gal}(H)$ (cf. [Va]). Pour a, b, c rationnels, on peut ainsi vérifier, au prix d'un calcul pénible, que la condition d'entrelacement de Katz [Ka1] (voir l'exemple 3 ci-dessous) est nécessaire et suffisante pour que $H(a, b, c)$ appartienne à la liste de Schwarz.

La situation est toute différente dans le cas des singularités irrégulières, qui va servir de test jusqu'à la fin. Rappelons tout d'abord ce qu'il en est de $\hat{V} = V \otimes \hat{K}$. D'après le théorème fondamental de Hukuhara-Turritin-Levelt (voir [Ka3]), il existe une extension \hat{L} de \hat{K} , de degré $N \leq n!$, au-dessus de laquelle \hat{V} devient isomorphe à une extension successive de vectoriels à connexion de dimension 1. (On dispose de nombreuses démonstrations de ce théorème. La dernière en date, due à Babbitt et Varadarajan [BV1], repose sur une double récurrence sur l'ordre et sur la dimension de l'orbite de la partie la plus polaire de la matrice représentative A sous les transformations de jauge de $\text{Aut}(\hat{V})$; elle vient d'être étendue aux équations aux dérivées partielles par C. Sabbah [Sa].) Autrement dit, il existe un isomorphisme horizontal $\hat{\Phi}$, défini sur \hat{L} , de \hat{V} sur une connexion ($\hat{V}_0 = \hat{K}^n, \hat{D}_0$), où

$$\hat{D}_0 = t(d/dt) - \Lambda + t^{-1}[(d/dt)Q](t^{-1}), \text{ — soit } U(\hat{D}_0) = t^\Lambda \exp(Q(t^{-1})) \text{ —,}$$

pour une uniformisante t de \widehat{L} , une matrice diagonale Q de polynômes sans termes constants (appelés facteurs déterminants de \widehat{V} , et permutés sous l'action du groupe d'inertie $\text{Gal}(\widehat{L}/\widehat{K})$), et une matrice de Jordan Λ commutant à Q (voir [Ju], p. 32, pour une forme \widehat{K} -rationnelle de $\widehat{\Phi}$). Les exponentielles des facteurs déterminants engendrent une extension de Picard-Vessiot de \widehat{L} , dont le groupe de Galois différentiel est un *tore* (dit *exponentiel*) $T_\sigma(D)$, isomorphe au groupe des caractères du \mathbf{Z} -module qu'ils engendrent dans $\mathbf{C}[t^{-1}]$. Fixons de plus un "opérateur de monodromie formelle" $t \rightarrow t \cdot \exp(2i\pi/N)$. Il agit sur $T_\sigma(D)$ à travers son image dans $\text{Gal}(\widehat{L}/\widehat{K})$; on note par ailleurs $M_{\sigma,f}(D)$ (*monodromie formelle* de V) la clôture de Zariski du groupe qu'il engendre dans $\text{Aut}(\widehat{V}_0^\partial)$. Dans ces conditions, on peut énoncer ([Ra2], [Mi]; voir également [Ka3], 1.4.5) :

PROPOSITION 2.— *Le groupe de Galois différentiel formel $G_{\sigma,f}(D)$ est engendré par $M_{\sigma,f}(D)$ et $T_\sigma(D)$.*

Comme $\widehat{\Phi}$ est une série en général divergente, les groupes $M_{\sigma,f}(D)$ et $M_{\sigma,an}$ n'ont aucun rapport, et il n'y a *a priori* pas de moyen naturel de ramener $G_{\sigma,f}(D)$ dans $G_{\sigma,an}(D)$. Supposons ainsi que S soit réduit à deux points $\{0, \infty\}$, de sorte que le groupe de monodromie global de D est abélien, et que 0 soit une singularité régulière. (C'est la première situation non triviale où les méthodes décrites plus bas sont couronnées de succès, et, il faut l'avouer, essentiellement la seule pour l'instant si D n'est pas fuchsien (voir [Ka4], [Ka2], 10.3, [Mi]).) Alors $G_{0,f} = G_{0,an} = M_{0,f} = M_{0,an}$ "est" le sous-groupe fermé de $M_{\infty,an}$ du groupe $G_{\infty,an} = \text{Gal}(D)$ mais il n'est pas facile de l'y réaliser :

Exemple 2.— L'équation de Bessel $J = J(v)$:

$$\partial^2 y + z^{-1} \partial y + (1 - v^2 z^{-2})y = 0 \quad (\partial = d/dz).$$

[On l'obtient à partir d'équations hypergéométriques par confluence de deux singularités; malheureusement, les connaissances actuelles sur le comportement des groupes de Galois différentiels par spécialisation ([Ka], 2.4, [Sn3]) ne permettent pas de tirer grand chose de cette remarque.] La sous-construction "puissance extérieure maximale" (wronskien) place $\text{Gal}(J)$ dans $SL(2)$. La méthode de Frobenius fournit un générateur de $G_{0,an}(J)$ unipotent non trivial pour v entier, semi-simple sinon. Par ailleurs (lemme de Hensel, voir [Ro]), les facteurs

déterminants de J en ∞ sont $\pm iz$, de sorte que $T_\infty(J)$ est un tore de dimension 1, et la monodromie formelle $M_{\infty,f}$ vaut $\mathbf{Z}/2\mathbf{Z}$ (qui n'a rien d'unipotent !). Ces seules informations ne suffisent en général pas à déterminer $\text{Gal}(J)$ (voir [Ko2], ou plus bas, pour la réponse).

Certes, on dispose aussi, grâce à Hankel, d'une expression de la base usuelle de solutions de J :

$$J_{\pm v}(z) = z^{-\frac{1}{2}}(c_{\pm v}e^{-iz}H_v(z) + c_{\pm v}^{-1}e^{iz}H_v(-z)),$$

$$Y_v(z) = z^{-\frac{1}{2}}(c'_v e^{iz}H_v(-z) + c'_v{}^{-1}e^{-iz}H_v(-z)) \quad \text{si } v \text{ est entier,}$$

en terme d'une fonction H_v analytique sur $|\text{Arg}(z)| < \pi$, y admettant au voisinage de ∞ un développement asymptotique \widehat{H}_v dans $K_{\infty,f}$. Mais la "prise de développement asymptotique" de $K(H_v)$ vers $K_{\infty,f}$ ne s'étend pas sans précaution en un homomorphisme différentiel de $K(J_v, J'_v, J_{-v}$ (ou Y_v)) vers $K_{\infty,f}(z^{\frac{1}{2}}, e^{iz})$. Ainsi, $2i \sin(z)$ vaut $S(z)e^{iz}$, où S admet $\widehat{S} = 1$ pour développement asymptotique au voisinage de ∞ sur le secteur $\text{Im}(z) < 0$, mais il n'y a pas d'homomorphisme différentiel de $K(\sin(z), \cos(z))$ dans $K_{\infty,f}(e^{iz})$ envoyant $S e^{iz}$ sur $\widehat{S} e^{iz} = e^{iz}$. Rien ne permet donc pour l'instant de regrouper les informations données par $M_{\infty,an}$ et par $G_{\infty,f}$.

La théorie de Ramis, que nous exposons maintenant, donne une réponse à ces différents problèmes, en inversant le sens des isomorphismes recherchés.

3. LE GROUPE DE STOKES (d'après Ramis)

Tout étant local dans ce paragraphe, nous supposons que le corps de base $K = K_{0,an}$ est le corps des fonctions méromorphes au voisinage de $\sigma = 0$. Les notations relatives à son complété formel $\widehat{K} = \mathbf{C}((z))$ sont celles du §2. Nous supposons pour simplifier que les facteurs déterminants $\{Q_1, \dots, Q_n\}$ de $V = V_{0,an}$ n'introduisent pas de ramification (i.e. que $N = 1$), et notons $k(V) = \sup_i(d^0(Q_i))$ le rang de Katz de V . Plus sérieusement, nous nous limitons au cas (dit générique, à une seule pente pour le polygône de Newton de $\text{End}(V)$) où toutes leurs différences non nulles $Q_i - Q_j$ sont de degré égal à $k = k(\text{End}(V))$.

On donne d'abord un aperçu de la méthode sous forme analytique (voir [MR1], [MR2], [RS], [Si]), puis on décrit la version cohomologique qu'en ont suggérée

Deligne et Malgrange (voir [Ma2], [MR2], [BV2], [LR]). L'énoncé suivant rassemble, sous leur forme la plus classique, les clefs d'existence et d'unicité des deux approches (voir [M-R], [RS]). Un isomorphisme, de matrice représentative encore notée $\widehat{\Phi}$, de \widehat{V} sur sa forme normale \widehat{V}_0 y est supposé fixé.

PROPOSITION 3.— Soit B un (germe de) secteur ouvert au voisinage de 0.

i) (Poincaré-Hukuhara-Turritin) : si $\text{Ouv}(B) < \frac{\pi}{k}$, il existe une matrice holomorphe Φ_B , asymptotique à $\widehat{\Phi}$ sur B , telle que

$$U_B(D)(z) = \Phi_B(z) z^\Lambda \exp(Q(z^{-1}))$$

soit une matrice fondamentale de solutions de $DY = 0$.

ii) (Watson-Nevanlinna) : Si $\text{Ouv}(B) > \frac{\pi}{k}$, toute fonction holomorphe sur B , admettant une décroissance exponentielle d'ordre k en 0, est nulle.

[Rappelons qu'une fonction f holomorphe sur B y admet un développement asymptotique en 0 s'il existe une série de Laurent formelle $\widehat{f} = T_B(f) = \sum_{v(\widehat{f}) \leq n} f_n z^n$ (alors unique) telle que pour tout entier $m \geq 0$,

$$R_m f(z) = z^{-m} (f(z) - \sum_{v(\widehat{f}) \leq n \leq m-1} f_n z^n)$$

reste bornée sur tout sous-secteur relativement compact de B . On a par ailleurs noté $\text{Ouv}(B)$ l'angle d'ouverture de B .]

D'après Cauchy, les matrices $\Phi_B(z) = U_B(z) \exp(-Q(z^{-1})) z^{-\Lambda}$ vérifiant l'énoncé i) se prolongent analytiquement sur tout secteur d'angle $< 2\pi$ contenant B . En général, ce prolongement cesse d'être asymptotique à $\widehat{\Phi}$, ou même d'admettre un développement asymptotique (de sorte qu'on ne peut utiliser ii)), mais vu la relation ci-dessus, de telles discontinuités ne peuvent apparaître qu'à la traversée d'une direction d'angle θ telle que près de $z = 0$, $\text{Re}(Q_i - Q_j)(z^{-1})$ change de signe au voisinage de $\text{Arg}(z) = \theta$ pour un couple (i, j) au moins. On appelle *lignes de Stokes* de telles directions, et on leur associe, afin de se placer dans la situation de ii) :

- les *lignes singulières*, d'angles α tels que l'une au moins des expressions $\exp(Q_i - Q_j)(z^{-1})$ présente une décroissance (exponentielle d'ordre k) de type maximal quand z tend vers 0 sur $\text{Arg}(z) = \alpha$;

- des “bons” recouvrements du cercle S des directions autour de 0. Dans le cas générique qui nous occupe (voir [MR2], [LR], pour le cas général), on prend par exemple la réunion, pour les différentes lignes singulières α , d’intervalles du type $B_\alpha =]\alpha - (\frac{\pi}{2k}), \alpha + (\frac{\pi}{2k}) + \delta[$, $B'_\alpha =]\alpha - (\frac{\pi}{2k}) - \delta, \alpha + (\frac{\pi}{2k})[$ bordés par des lignes de Stokes et n’en contenant pas en dehors de $b_\alpha = B'_\alpha \cap B_\alpha$ (voir [Mi], 2.3.6). On étend enfin ces notations au revêtement universel \tilde{S} de S et on identifie les intervalles de \tilde{S} aux secteurs ouverts de centre 0 qu’ils sous-tendent.

3.1. L’approche analytique

Le point de départ de cette approche est que les coefficients de $\hat{\Phi}$ appartiennent à une sous-algèbre différentielle (stricte) de \hat{K} au-dessus de laquelle, pour tout secteur $B(d)$ d’ouverture un peu plus grande que $\frac{\pi}{k}$ et de bissectrice d non singulière, l’application $T_{B(d)}$ admet une section canonique (voir [Si], A.4.1.2) : les fonctions plates de son image présentent une décroissance exponentielle d’ordre k , et la proposition 3.ii) les force à être nulles. Ramis décrit cette section par un procédé de resommation $\Sigma_d = \Sigma_{k,d}$, composé d’une transformation de Borel formelle, et d’une transformation de Laplace (voir [MR2], §1), qui commute donc aux opérations algébriques et à la dérivation. Par conséquent, Σ_d est un isomorphisme d’algèbres différentielles de $K[\hat{\Phi}]$ sur $K[\Sigma_d \hat{\Phi}]$, qu’on étend facilement (voir [MR2], Prop. 9) en un isomorphisme d’algèbres différentielles de $K[\hat{\Phi}, z^\lambda, \exp Q_i(z^{-1}), \text{Log } z]$ sur $K[\Sigma_d \hat{\Phi}, z^\lambda, \exp(Q_i(z^{-1})), \text{Log } z]$. Pour toute direction d non singulière, la matrice $\Sigma_d \hat{\Phi} = \Phi_{B(d)}$ vérifie donc les conditions de la proposition 3.i) sur $B(d)$ (dans notre cas générique, on aurait d’ailleurs pu déduire directement de 3.i) l’existence et l’unicité d’une telle solution, voir [IK]⁽⁴⁾). Associons alors à chaque direction singulière ou, plus exactement (du fait de la matrice Λ), à chacun de ses représentants α dans \tilde{S} les matrices fondamentales

$$U_\alpha(D) = \Sigma_{\alpha+(\frac{\pi}{2})}(\hat{\Phi} z^\Lambda \exp Q(z^{-1})) \quad , \quad U'_\alpha(D) = \Sigma_{\alpha-(\frac{\pi}{2})}(\hat{\Phi} z^\Lambda \exp Q(z^{-1}))$$

de solutions de $DY = 0$ sur B_α, B'_α . Elles sont liées sur le secteur b_α par la relation $U'_\alpha(D) = U_\alpha(D) C_\alpha$, où la matrice unipotente C_α représente un endomorphisme canonique st_α du \mathbf{C} -espace V^∂ des solutions de D holomorphes sur b_α . Mais st_α est composé des isomorphismes différentiels $\Sigma_{\alpha+(\frac{\pi}{2})}^{-1}$ et $\Sigma_{\alpha-(\frac{\pi}{2})}$; c’est donc un élément du groupe de Galois différentiel de V sur K . On appellera *groupe de Stokes*

$St_\sigma(D) = St_0(D)$ (en la singularité irrégulière $\sigma = 0$) le sous-groupe algébrique de $G_{0,an}(D)$ engendré ⁽⁴⁾ par les st_α , où α parcourt l'ensemble des relevés à \tilde{S} des différentes directions singulières de V .

Avec la définition précédente tout au moins, le groupe de Stokes n'est en général pas normal dans $G_{0,an}$, mais $M_{0,f}$ le laisse stable par conjugaison. D'autre part, chaque isomorphisme st_α^{-1} calcule le *prolongement analytique* de la solution $U'_\alpha(D)$ de B'_α à B_α (cf. [DM], Prop. 6.1). En les composant le long de S , et en appliquant l'opérateur de monodromie formelle pour revenir à la détermination de l'argument initial, on obtient donc un générateur du groupe de monodromie analytique $M_{0,an}(D)$. L'incompatibilité remarquée plus haut entre $M_{0,f}$ et $M_{0,an}$ est ainsi élucidée. C'est même ce point qui avait mis Stokes sur la voie de son phénomène, comme il s'en explique dans une lettre ([St], p. 62) à sa fiancée — le mariage a néanmoins convergé. Plus généralement, Ramis établit :

THÉORÈME 2.— *Le groupe de Galois analytique local $G_{0,an}(D)$ est engendré par ses sous-groupes fermés $G_{0,f}(D)$ et $St_0(D)$ (donc aussi, si l'on préfère, par $T_{0,f}(D)$, $M_{0,an}(D)$ et $St_0(D)$).*

La démonstration de ce théorème (voir [MR2], théorème 6) repose sur le théorème 1, et le fait (équivalent à la remarque précédente) qu'une solution invariante sous le groupe de Stokes et la monodromie formelle est uniforme au voisinage de 0. Une démonstration tannakienne a également été proposée par Deligne.

3.2. L'approche cohomologique

Elle s'appuie sur la classification, à K -isomorphisme près, des K -vectoriels à connexion V munis d'un isomorphisme formel $\hat{\Phi}$ de \hat{V} sur une forme normale fixée \hat{V}_0 comme plus haut (théorème de Sibuya-Malgrange ; voir [Ma2], [Si], [BV2]). Soient A le faisceau sur S des germes de fonctions méromorphes admettant un développement asymptotique en 0, et $ST(V)$ le faisceau en groupes (non abéliens) des sections horizontales du faisceau $A \otimes \text{End}(V)$ qui sont asymptotiques à l'identité. On déduit de la variante " $k = \infty$ " (resp. $k = \frac{1}{2}$, ou " $k = 0$ " pour \tilde{S}) de la proposition 3.i) (resp. ii)) que $ST(V)$ est localement isomorphe à $ST(V_0)^{(6)}$, et que l'ensemble classifiant recherché pour les couples $(V, \hat{\Phi})$ s'identifie à $H^1(S, ST(V_0))$, évidemment pointé par $(V_0, (id)^\wedge)$ — c'est à K -isomorphisme près le seul couple à ne pas présenter de phénomène de Stokes. Or $ST(V_0)$ est

facile à décrire ! Ses sections au-dessus d'un secteur B (relevé dans \tilde{S}) sont les matrices P_B holomorphes sur B , y admettant I pour développement asymptotique, et telles que $P_B(z) z^\Lambda \exp(Q(z^{-1})) = z^\Lambda \exp(Q(z^{-1})) C_B$ pour une matrice constante C_B . En particulier, $\exp(Q(z^{-1})) C_B \exp(-Q(z^{-1}))$ doit être asymptotique à I sur B , et $ST(V_0)$ est un faisceau constant par morceaux, dont les discontinuités n'apparaissent qu'aux lignes de Stokes de V . Dans le cas générique (voir [BV2], [MR2], [LR] pour le cas général), les classes de cohomologie de $ST(V_0)$ peuvent donc être décrites univoquement par la donnée, pour chaque secteur b_α , d'une matrice constante unipotente C_α n'ayant, hors de la diagonale, de termes non nuls qu'aux indices (i, j) des facteurs déterminants attachés à la ligne singulière α .

Soit alors $\{C_\alpha\}$ une telle famille de matrices. Il lui correspond, à K -isomorphisme près, un unique couple $(V, \hat{\Phi})$ dans $H^1(S, ST(V_0))$. Comme les "bons" recouvrements sont suffisamment fins (voir [BV2], 3.4.2), chacune des matrices $P_\alpha(z)$ correspondant (dans \tilde{S}) à C_α s'écrit comme un quotient $(\Phi_\alpha)^{-1}(\Phi'_\alpha)$ de sections horizontales de $A \otimes \text{Hom}(V, V_0)$ sur B_α, B'_α , asymptotiques à $\hat{\Phi}$, et ce de façon unique (puisque l'identité est la seule section globale de $ST(V)$). En d'autres termes, on a su associer de façon canonique à $\hat{\Phi}$ et à chaque (relevé de) direction singulière α deux matrices fondamentales de solutions $U_\alpha(D) = \Phi_\alpha(z) z^\Lambda \exp(Q(z^{-1}))$, $U'_\alpha(D)$, de $DY = 0$, telle que $U'_\alpha(D) = U_\alpha(D) C_\alpha$ sur b_α , d'où un élément canonique st_α de $\text{Aut}_{\mathbf{C}}(V^\partial)$ (et l'unicité de la description "analytique" montre que ce sont bien ceux de 3.1). Mais les formes normales W_0 des différentes constructions tensorielles W de V respectent leurs sous-connexions. Il en est donc de même de l'automorphisme que définit (tout au moins si ces W sont encore génériques) st_α sur les espaces W^∂ correspondants. On voit ainsi à nouveau que st_α est un élément du groupe de Galois $\text{Gal}(D)$. Ce point de vue permet aussi de prouver le théorème 3 : pour toute construction W de V , un \mathbf{C} -sous-espace X' de W^∂ stable sous $G_{0,f}(D)$ détermine une sous-connexion X_0 de la construction correspondante W_0 de V_0 ; si de plus X' est stable sous $ST_0(D)$, les restrictions à X des matrices C_α définissent un élément de $H^1(S, ST(X_0))$ que l'isomorphisme de Malgrange-Sibuya permet d'interpréter comme une sous-connexion X de W , d'espace des vecteurs horizontaux $X^\partial = X'$ (voir [LR]⁽⁷⁾).

3.3. Calculs et remarques

Il reste à calculer les matrices de Stokes-Ramis C_α . Dans l'approche analytique, les opérateurs de resommation permettent d'exprimer concrètement leurs coefficients sous forme intégrale, et le théorème des résidus donne immédiatement la réponse (voir [RM1]). Quant à l'approche cohomologique, un de ses avantages est de justifier des passages à la limite : l'espace $H^1(S, ST(V_0))$ est en effet muni d'une structure naturelle de variété affine, qui en fait un espace de modules pour les classes d'isomorphismes de couples $(V, \hat{\Phi})$ (voir [BV2], [LR]).

Exemple 2 (suite) : L'équation de Bessel ($J(v)$).

C'est (en $\sigma = \infty$) un cas générique, avec $k = 1$. Les lignes singulières (resp. les lignes de Stokes) sont les deux demi-axes imaginaires $\alpha_1 = -i$, $\alpha_2 = i$ (resp. les deux demi-axes réels), et un bon recouvrement est donné par

$$B_1 =] - 2\pi, 0[\quad , \quad B'_1 =] - \pi, \pi[\quad , \quad B_2 =] - \pi, \pi[\quad , \quad B'_2 =] 0, 2\pi[,$$

d'où $b_1 =] - \pi, 0[$, $b_2 =] 0, \pi[$. Les matrices de Stokes-Ramis sont

$$C_i = \begin{bmatrix} 1 & 0 \\ c_i(V) & 1 \end{bmatrix} \quad , \quad C_{-i} = \begin{bmatrix} 1 & c_{-i}(v) \\ 0 & 1 \end{bmatrix} ,$$

où

$$c_{\pm i}(v) = \mp 2 \exp(\mp 2i\pi v) \cos \pi v, \text{ d'après [MR1], 3.4.12 ;}$$

$$c_{\pm i}(v) = \mp 2 \cos \pi v, \text{ d'après [BV2], p. 94.}$$

Il ne s'agit pas d'erreur expérimentale, mais d'un choix de base de solutions (ou d'isomorphisme $\hat{\Phi}$) différent. (Les matrices de [MR1] sont écrites dans la base $z^{-\frac{1}{2}} e^{-iz} H_v(z)$, $z^{-\frac{1}{2}} e^{iz} H_v(-z)$; en fait $H_v = \Sigma_0(\hat{H}_v)$, et il n'y avait finalement pas de précaution à prendre dans l'exemple 2 du §2...). Mais peu importe. Il suffit de savoir que $c_i(v)$ et $c_{-i}(v)$ sont non nuls pour déduire immédiatement de ces expressions que dès que v n'est pas congru à $\frac{1}{2}$ modulo \mathbf{Z} , le groupe de Stokes, et donc $\text{Gal}(J(v))$ remplit $SL(2)$. En ce sens, la méthode est peut-être trop riche : elle fournit des "formules de connexions" exactes, là où il suffirait de connaître la position des coefficients non nuls de certaines matrices unipotentes (voir [MR1]).

On trouvera d'autres illustrations du cas générique dans [MR1] (équations hypergéométriques d'ordre 2 confluentes), et dans [DM], [Mi] (pour une vaste

classe de ces équations en ordre supérieur ; voir l'exemple 3 ci-dessous). Je renvoie à [Ra2] et à [LR] pour le cas général, en signalant seulement que, du point de vue théorique, celui-ci se traite à partir d'une décomposition canonique de $\widehat{\Phi}$, indexée par les différentes pentes du polygone de Newton de $\text{End}(V)$. L'approche analytique repose alors sur la notion fondamentale de "multisommabilité" ([MR2], [BBRS]), et nécessite en particulier une version relative de la proposition 3, ii) (voir [M-R], [Ma2]). On en déduit une filtration "Gevrey" du groupe de Galois différentiel ([Ra2]), reflétée, dans l'approche cohomologique, par une filtration du même type ([MR2], [LR]) sur le faisceau $ST(V_0)$.

Enfin, l'un des aspects les plus intéressants de la théorie de Ramis est son lien avec le groupe de *monodromie sauvage* $\pi_{1,s}$ d'un germe de voisinage ouvert épointé de l'origine (voir [MR2], §4, pp. 381, 388). Ce groupe est le produit semi-direct du groupe de monodromie usuel par un groupe qui, en cran fini, attache à un \mathbf{Z} -module de type fini dans l'algèbre des polynômes ramifiés en z^{-1} son groupe des caractères et un "groupe de résurgence" — groupe d'algèbre de Lie libre attaché aux diverses lignes singulières correspondantes —, dont il forme le produit semi-direct. On peut alors voir l'image de $\text{Gal}(D)$ dans $\text{Aut}(V^\partial)$ comme une représentation de $\pi_{1,s}$ et en déduire ([MR2], théorème 18) une correspondance de Riemann-Hilbert généralisée entre la catégorie des $K_{0,an}$ -vectoriels à connexions et celle des représentations de dimension finie de $\pi_{1,s}$. (Voir encore [MR2] pour le lien avec les travaux d'Écalle.)

4. GROUPES RÉDUCTIFS (cf. [Ka4], [Be])

Les résultats de ce § reposent sur les propriétés très rigides des algèbres de Lie semi-simples complexes. L'idée de les exploiter est classique (voir [Kap] et [Ko2] pour $\mathfrak{sl}(2)$, appliqué aux équations d'Airy et de Bessel). Les récents travaux de Beukers, Brownawell, Heckmann ([BBH], [BH]) et de Katz et Gabber ([Ka3], [Ka4]) lui ont donné un nouvel essor. Comme il existe déjà sur le sujet un demi-livre ([Ka4] ; la lecture de l'autre demi est également recommandée) et un rapport [Be], je serai plus bref.

Supposons que le K -vectoriel à connexion V soit irréductible ou, plus généralement, somme directe d'irréductibles. D'après le théorème 1, la représentation

fidèle V^∂ de $G = \text{Gal}(D)$ est alors complètement réductible, et G est un groupe réductif, produit presque direct $Z.DG$ de son centre Z , qui est de type multiplicatif, par son groupe dérivé DG . Comme on est en caractéristique nulle, toutes les représentations de G , et donc toutes les constructions de V , sont alors complètement réductibles, et on peut s'attendre à ce que la recherche des sous-connexions définissant G en soit facilitée. Par exemple (voir [BBH], [Be]), dès que DG n'est pas un groupe fini, il suffit que le carré symétrique $S^2(V)$ de V soit irréductible pour que toutes les puissances de V le soient, et DG est alors le groupe $SL(n)$ ou le groupe symplectique $Sp(n)$ tout entier.

Il est donc important de disposer de critères d'irréductibilité de V . En termes concrets, il s'agit de voir, quand V est défini par une équation différentielle $Hy = 0$, si H est irréductible dans l'anneau $K[\partial]$ des opérateurs différentiels à coefficients dans K . C'est là un problème difficile⁽⁸⁾ (par exemple, on ne dispose pas dans l'anneau $\mathbf{C}[z, d/dz]$ d'un analogue du lemme des Gauss), mais on peut parfois s'en tirer par une analyse locale. Ainsi, avec les notations du §2, si tous les facteurs déterminants de \widehat{V} ont une même valuation, de dénominateur égal à n , \widehat{V} est irréductible sur \widehat{K} , et V l'est donc aussi sur K (voir [Ka3]), 2.2.8 et 4 pour une application aux équations d'Airy généralisées).

Exemple 2 (fin) : l'équation de Bessel $J(v)$

Pour v entier, la présence d'un logarithme dans la solution Y_v , jointe à l'irrationalité de $2\frac{J'}{J}$ montre que son carré symétrique est irréductible sur $K = \mathbf{C}(z)$. Comme $\dim \text{Gal}(J(v)) = \text{deg. tr } K(J, J', Y)/K \geq 2$, DG ne peut être fini, et le marteau concluant l'alinéa précédent plaque $\text{Gal}(J(v))$ sur $SL(2)$. Idem si v n'est pas demi-entier, en adaptant ce qui suit. En revanche, $\text{Gal}(J(\frac{1}{2}))$ est un tore (fonctions d'Hermite).

Par ailleurs, c'est plutôt l'algèbre de Lie de G que ces méthodes permettent de cerner, ce qui conduit à étudier la restriction de la représentation V^∂ à la composante neutre G^0 de G . On dit ainsi que V est *Lie-irréductible* si cette représentation de G^0 est irréductible (autrement dit, si V est "absolument" irréductible). Voici le type d'énoncés, extraits de [Ka4] et [Be], dont on dispose alors, et qu'on déduit de la classification des algèbres de Lie semi-simples, jointe à un résultat de Gabber sur les sous-tores de leurs normalisateurs :

PROPOSITION 4.— Soit G un sous-groupe connexe de $GL(n)$ agissant de façon irréductible sur \mathbf{C}^n , et normalisé par une matrice diagonale μ .

i) Si $\mu = (x, 1, \dots, 1)$ est d'ordre > 2 (resp. $= 2$), alors : $DG = SL(n)$ (resp. $SL(n)$ ou $SO(n)$).

ii) Si $n = 7$, et μ engendre topologiquement le tore $(x, y, xy, x^{-1}, y^{-1}, (xy)^{-1}, 1)$, alors $DG = SL(7)$, ou $SO(7)$, ou G_2 .

Et en voici une illustration :

Exemple 3 ([BBH], [BH], [Ka4]) : l'équation hypergéométrique généralisée H :

$$(\theta + a_1) \cdots (\theta + a_p)y - z^{-1}(\theta + c_1 - 1) \cdots (\theta + c_q - 1)y = 0 \quad (\theta = zd/dz).$$

Elle est irréductible si et seulement si les différences $a_i - c_j$ ne sont jamais entières. Si $q > p$, ses singularités sont $\{0, \infty\}$, et elle ne peut devenir réductible que sur une extension cyclique $\mathbf{C}(z^{\frac{1}{N}})$ de $\mathbf{C}(z)$, ce qui s'énonce encore simplement sur les paramètres (voir [Ka3], 1.2.5 pour un calcul universel de $\text{Gal}(D)/\text{Gal}(D)^0$) ; l'analyse est plus délicate dans le cas fuchsien $q = p$, où $\Sigma = \{0, 1, \infty\}$. Supposons d'abord H Lie-irréductible. La proposition 4.i), appliquée à G^0 et à un générateur μ de $M_{1,an}$, montre que pour $q = p$, $G = \text{Gal}(H)$ contient $SO(n)$ (resp. $SL(n)$) dès $2(\Sigma_i(a_i) - \Sigma_j(c_j))$ n'est pas un entier pair (resp. un entier) ; pour $q > p$, une extension de la proposition 4, appliquée à G^0 et à un générateur topologique μ de $T_{\infty,f}$, ramène l'étude de G à celle de relations de dépendance linéaire entre des racines $(q - p)$ -ièmes de l'unité. L'énoncé ii) entraîne ainsi que $\text{Gal}(H)^0$ vaut G_2 pour $(p, q; \{a\}, \{c\}) = (1, 7; \{\frac{1}{2}\}, \{\frac{j}{7}; 0 \leq j < 6\})$. Noter qu'une partie de ces résultats a également été établie par la méthode du §3 ; voir [DM], et [Mi], 4.8.2, pour une version tordue de ce dernier exemple.

Restreignons-nous maintenant au cas fuchsien $q = p = n$, et supposons au contraire que H ne soit pas Lie-irréductible. En général, la représentation V^∂ de G sera induite de la représentation d'un sous-groupe d'indice fini d divisant n , et l'étude de H se ramène à celle d'une équation hypergéométrique d'ordre $\frac{n}{d}$, qu'un théorème de Levelt de type Riemann-Hilbert permet d'expliciter (voir [Ka4], 3.5.4). Mais dans le cas contraire, la représentation restreinte à G^0 est isotypique non irréductible, et l'étude de $M_{1,an}$ (voir [Ka4], 3.5.7) montre qu'elle est monomiale ! Dès que le wronskien de H est une fonction algébrique, G^0 est alors trivial, et H admet une base de solutions algébriques. Beukers et Heckman,

auxquels on doit cette solution du problème de Schwarz généralisé, montrent dans [BH] que ce cas se produit si et seulement si les paramètres a_i, c_j sont rationnels et si, pour tout automorphisme ξ du groupe de Galois de l'extension cyclotomique de \mathbf{Q} qu'elles engendrent, les racines de l'unité $\{\xi(\mathbf{e}(a_1)), \dots, \xi(\mathbf{e}(a_n))\}$ et $\{\xi(\mathbf{e}(c_1)), \dots, \xi(\mathbf{e}(c_n))\}$ sont entrelacées sur le cercle unité.

Enfin, le cas où H elle-même est réductible se ramène au précédent par une technique de cohomologie galoisienne inspirée de la théorie de Kummer (voir [Bo]). (On peut également le traiter, sans discussion séparée, par les méthodes du §3 [DM].)

Signalons pour conclure que Katz introduit dans [Ka3] une filtration du groupe de Galois formel en une singularité irrégulière, qui permet de préciser l'information donnée par le tore exponentiel $T_{\sigma, f}(D)$. Par souci de symétrie, je n'en parle pas plus que des filtrations Gevrey de Ramis.

5. LA CONJECTURE DE GROTHENDIECK (d'après Katz)

La méthode que nous exposons pour finir donne, avec les notations du §1, la "K-forme" $\text{Gal}(V/K)$ du groupe de Galois différentiel $\text{Gal}(D)$. Elle ne nécessite en particulier aucune connaissance des foncteurs fibres de solutions du type V^∂ . Son principal défaut (ou sa principale qualité, suivant les goûts) réside dans son caractère actuellement conjecturel.

Comme il est bien connu, toute équation différentielle intéressante est à coefficients dans $\mathbf{Q}(z)$ (des techniques standard (voir [Ka2]), permettent en tous cas de supposer les scalaires algébriques sur \mathbf{Q}). Soit donc V un espace vectoriel sur $K = \mathbf{Q}(z)$, de dimension n , muni d'une connexion ∇ (dans les notations du §1, $\nabla(\partial) = D$, avec, disons, $\partial = d/dz$). Pour tout nombre premier p suffisamment grand, il fournit par réduction modulo p un vectoriel à connexion (V_p, ∇_p) sur $(\mathbf{F}_p(z), \partial_p = d/dz)$. Alors (voir [Ka1] pour tout ce qui suit), $(\partial_p)^p$ est encore une dérivation de $\mathbf{F}_p(z)$, et l'élément $\psi_p(\nabla(\partial)) = (\nabla_p(\partial_p))^p - \nabla_p((\partial_p)^p)$ de $\text{End}_{\mathbf{F}_p}(V_p)$ est $\mathbf{F}_p(z)$ -linéaire. On l'appelle (image de ∂ par l') opérateur de p -courbure de la connexion ∇ . Si on définit, par récurrence à partir de la matrice $A(z) = A_1$ représentant D dans une base de V sur $\mathbf{Q}(z)$, une suite de matrices par les relations $A_{k+1} = \partial A_k + A A_k$, la p -courbure est donnée dans la base correspondante de V_p sur $\mathbf{F}_p(z)$ par la réduction modulo p de la matrice $A_p(z)$.

Dans ces conditions, les vecteurs horizontaux de l'extension formelle \widehat{V}_p de V_p à $\mathbf{F}_p((z))$ forment un espace vectoriel de dimension n_p sur le corps des constantes $\mathbf{F}_p((z^p))$, qui s'obtient par extension des scalaires à partir du $\mathbf{F}_p(z^p)$ -espace vectoriel des vecteurs horizontaux de V_p sur $\mathbf{F}_p(z)$ (théorème de Honda), et n_p vaut n (la connexion est "intégrable") si et seulement si la p -courbure $\psi_p((\nabla(\partial)))$ est nulle (théorème de Cartier). D'après le théorème d'Eisenstein, le groupe de Galois différentiel $\text{Gal}(V/K)$ ne peut donc être fini, ou encore, son algèbre de Lie $\mathfrak{gal}(V/K)$ sur K ne peut être nulle, que si les p -courbures $\psi_p((\nabla(\partial)))$ de ∇ sont nulles pour tout nombre premier suffisamment grand (voir [La]). Plus généralement, on peut définir par noéthérianité la plus petite K -sous-algèbre de Lie algébrique $\mathfrak{courb}(V/K)$ de $\mathfrak{gl}(n, K)$ dont les réductions modulo p contiennent $\psi_p((\nabla(\partial)))$ pour tout nombre premier p suffisamment grand. La définition même des p -courbures montre que $\mathfrak{courb}(V/K)$ est contenue dans $\mathfrak{gal}(V/K)$. La conjecture suivante, due à Katz [Ka2], généralise une conjecture de Grothendieck, qui portait sur le cas $\mathfrak{courb}(V/K) = 0$:

Conjecture.— *Les algèbres de Lie $\mathfrak{courb}(V/K)$ et $\mathfrak{gal}(V/K)$ coïncident.*

On notera l'analogie formelle de cette conjecture avec le théorème d'Ambrose-Singer [AS], en vertu duquel l'algèbre de Lie du groupe d'holonomie d'une variété riemannienne M est engendrée par les (conjugués par transport parallèle des différents endomorphismes de) courbures en toutes les places de M . Par ailleurs, Y. André [An2] attache à toute connexion sur $\mathbf{F}_p(z)$ un "schéma en groupe de Galois différentiel" non nécessairement réduit, dont la p -algèbre de Lie est engendrée par la p -courbure de la connexion. La conjecture équivaut alors à dire que $\text{Gal}(V/K)^0$ est le plus petit sous-groupe algébrique de $\text{Aut}_K(V)$ dont les réductions modulo p contiennent ces sous-schémas pour presque tout p .

Considérons notre situation "test" à deux singularités $\Sigma = \{0, \infty\}$. La conjecture est alors démontrée dans [Ka2], 10.3 par réduction au cas $\mathfrak{courb}(V/K) = 0$, où elle découle de [Ka1]. Cela vaudrait la peine de la vérifier directement. Elle fournit en tous cas un lien mystérieux entre les groupes de Stokes et les p -courbures, puisque (pour ne considérer que nos exemples 2 et 3, avec $q > p$) les groupes de Galois formels ne suffisent pas à engendrer $\text{Gal}(V/K)^0$. Noter néanmoins que les logarithmes des automorphismes de Stokes sont nilpotents, alors qu'en présence de singularités irrégulières, la plupart des p -courbures ne peuvent pas l'être ([Ka2],

8.1(1)). Modulo p , les séries divergentes qui apparaissent dans les isomorphismes formels $\widehat{\Phi}$ (et qui sont responsables du phénomène de Stokes) se réduisent d'ailleurs souvent en des polynômes (exemple type : la série d'Euler $\sum_{n \geq 0} (n!)z^n$) ; seule une analyse p -adique permet d'en rendre compte (voir [Dw]).

En fait, en dehors de certaines équations à groupes de Galois unipotents ([An1], 8.1.3, [BR]), les seuls cas concrets où la conjecture de Katz a été vérifiée concernent la conjecture de Grothendieck (voir à ce propos [Ka3], 10.2), ce qui, grâce au théorème de Cartier, ne nécessite pas à proprement parler de calculer $\text{courb}(V/K)$... Outre le cas cité plus haut, il en est ainsi des équations d'ordre 1 sur \mathbf{P}_1 ([Ho]) ou des équations hypergéométriques ([Ka1], [BH] ; voir l'exemple 3 ci-dessus, ainsi que [Dw]). Mais une approche nouvelle a été introduite par D. et G. Chudnovsky ([CC]), à partir d'une version raffinée du critère de transcendance de Schneider-Lang. Elle leur permet de vérifier la conjecture de Grothendieck pour toutes les équations différentielles d'ordre 1 sur une surface de Riemann (de genre arbitraire), et pour les équations de Lamé (qui, comme le souligne Katz, sont des équations d'Airy sur une courbe elliptique — voir [BD], [Ch] pour la recherche des listes de Schwarz correspondantes).

Du point de vue "calcul de $\text{Gal}(D)$ " qui nous intéresse ici, l'aspect le plus frappant de la démonstration de [CC] est qu'elle ne nécessite d'informations locales qu'en un nombre fini, *effectivement calculable*, de nombres premiers p . Grâce aux versions effectives du théorème de Chebotarev ([Se], Théorème 5), la même remarque s'appliquait d'ailleurs déjà à la preuve de Honda [Ho]. Ceci conduit à préciser la conjecture de Katz par la question suivante, qui nous servira de conclusion. Sans chercher à être intrinsèque, appelons hauteur de (V, ∇) le maximum des degrés et des hauteurs logarithmiques des coefficients d'une matrice représentative A de $\nabla(\partial)$ et de leur dénominateur commun T , et disons que ∇ a bonne réduction en p si p ne divise aucun des coefficients de T et est $> n$. La base de V implicitement choisie permet de définir la hauteur (logarithmique) $h(X)$ d'un K -sous-espace vectoriel X de V . Dans ces conditions :

(?) *Existe-t-il deux constantes effectives $c(n)$, $\kappa(n)$ vérifiant la propriété suivante ? Soient $\mathbf{h} \geq 1$ un nombre réel, ∇ une connexion sur V , de hauteur $\leq \mathbf{h}$, et X une K -droite de V telle que, pour tout nombre premier $p \leq c(n)(\mathbf{h} + h(X))^{\kappa(n)}$ de bonne réduction pour ∇ , la réduction X_p de X modulo p soit stable sous $\psi_p((\nabla(\partial)))$. Alors, X est stable sous $\text{Gal}(V/K)^0$.*

BIBLIOGRAPHIE

(A permutations près, les sigles L.O.D.E., S.E.D., ... signifient : équations ou systèmes différentiels linéaires ordinaires.)

- [An1] Y. ANDRÉ - *G-functions and geometry*, Vieweg, Asp. Math **13**, 1989.
- [An2] Y. ANDRÉ - *Notes sur la théorie de Galois différentielle*, Prépubl. IHES (1989).
- [AS] W. AMBROSE and I. SINGER - *A theorem on holonomy*, Trans. AMS **75** (1953), 428-443 (voir aussi : J.-P. Bourguignon, Astérisque **126** (1985), 169-183).
- [BV1] D. BABBIT and V. VARADARAJAN - *Formal reduction of meromorphic D.E. : a group theoretic view*, Pacific J. Math. **108** (1983), 1-80.
- [BV2] D. BABBIT and V. VARADARAJAN - *Local moduli for meromorphic D.E.*, Astérisque **169-170** (1989).
- [BD] F. BALDASSARRI and B. DWORK - *On second order D.E. with algebraic solutions*, Amer. J. Math. **101** (1979), 42-76.
- [BBS] W. BALSER, B. BRAAKSMA, J.-P. RAMIS and Y. SIBUYA - *Multisummability of formal power series solutions of L.O.D.E.*, Asympt. Anal. **5** (1991), 27-45.
- [Be] F. BEUKERS - *Differential Galois theory*, Théorie des Nombres et Physique, Springer, à paraître.
- [BBH] F. BEUKERS, D. BROWNAWELL and G. HECKMAN - *Siegel normality*, Ann. Maths **127** (1988), 279-308.
- [BH] F. BEUKERS and G. HECKMAN - *Monodromy for the hypergeometric function ${}_nF_{n-1}$* , Inv. Math. **95** (1989), 325-354.
- [BR] J.-P. BEZIVIN et P. ROBBA - *Rational solutions of L.D.E.*, J. Austr. Math. Soc. **46** (1989), 184-196.
- [Bl] A. BLANCHARD - *Groupes algébriques et équations différentielles linéaires [d'après E. Kolchin]*, Sémin. Bourbaki n° 17 (1949/50).
- [Bo] K. BOUSSEL - *Groupes de Galois des équations hypergéométriques...*, CRAS Paris **309** (1989), 587-589.
- [Ch] B. CHIARELLOTTO - *On Lamé operators which are pull-backs of hypergeometric ones*, Prépubl. U. Padova (1991).

- [CC] D. and G. CHUDNOVSKY - *Application of Padé approximations to the Grothendieck conjecture on L.D.E.*, Springer L. N. **1135** (1984), 52-100.
- [De] P. DELIGNE - *Catégories tannakiennes*, Birkhäuser Prog. Math., Grothendieck Festschrift II (1990), 111-195.
- [DM] A. DUVAL et C. MITSCHI - *Matrices de Stokes et groupes de Galois des équations hypergéométriques...*, Pacific J. Math. **138** (1989), 25-56.
- [Dw] B. DWORK - *Hypergeometric functions*, Oxford U.P. (1990).
- [Ho] T. HONDA - *Algebraic differential equations*, Symp. Math. **24** (1981), 169-204.
- [IK] Y. ILIANSHENKO and A. KHOVANSKII - *Galois groups, Stokes multipliers and a theorem of Ramis*, Funct. Anal. and Appl. **24** (1990), 286-296.
- [Ju] W. JURKAT - *Meromorphe Differentialgleichungen*, Springer L.N. **637** (1978).
- [Kap] I. KAPLANSKY - *An introduction to differential algebra*, Hermann (1957).
- [Ka1] N. KATZ - *Algebraic solutions of differential equations ; p-curvature and the Hodge filtration*, Inv. Math. **18** (1972), 1-118.
- [Ka2] N. KATZ - *A conjecture in the arithmetic theory of differential equations*, Bull. Soc. Math. France **110** (1982), 203-239 et 347-348.
- [Ka3] N. KATZ - *On the calculation of some differential Galois groups*, Inv. Math. **87** (1987), 13-61.
- [Ka4] N. KATZ - *Exponential sums and differential equations*, Princeton U. Press (1990).
- [Ko1] E. KOLCHIN - *Differential algebra and algebraic groups*, Acad. Press, 1973.
- [Ko2] E. KOLCHIN - *Algebraic groups and algebraic dependence*, Amer. J. Math. **90** (1968), 1151-1164.
- [La] E. LANDAU - *Eine Anwendung des Eisensteinsches Satz auf die Theorie der Gaussche Differentialgleichung*, Crelle **127** (1904), 92-102.
- [Le] A. LEVELT - *Differential Galois theory and tensor products*, Indag. Math. **1** (1990), 439-449.
- [LR] M. LODAY-RICHAUD - *Classification méromorphe locale des S.D.L. ... : phénomène de Stokes et applications*, Thèse, U. Orsay (1991).
- [Ma1] B. MALGRANGE - *La classification des connexions irrégulières à une variable*, Birkhäuser Prog. Math. **37** (1983), 381-399.
- [Ma2] B. MALGRANGE - *Équations différentielles à coefficients polynômiaux*, Birkhäuser Prog. Math. **96** (1991).

- [M-R] B. MALGRANGE et J.-P. RAMIS - *Fonctions multisommables*, Ann. Inst. Fourier **42** (1992).
- [MR1] J. MARTINET et J.-P. RAMIS - *Théorie de Galois différentielle et resommation*, Computer Algebra and Differential Equations, E. Tournier (ed.), Acad. Press (1989).
- [MR2] J. MARTINET et J.-P. RAMIS - *Elementary acceleration and multisummability*, Ann. I.H.P., Phys. Th. **54** (1991), 331-401.
- [Mi] C. MITSCHI - *Groupes de Galois différentiels et G-fonctions*, Computer Algebra and Diff. Eqs., M. Singer (ed.), Acad. Press (1991).
- [Ne] Y. NESTERENKO - *On the algebraic independence of the components of a solution of a S.D.E.*, Math. URSS Izvestia **38** (1974), 492-512.
- [Po] J.-F. POMMARET - *Differential Galois Theory*, Gordon & Breach (1982).
- [Ra1] J.-P. RAMIS - *Théorèmes d'indice Gevrey pour les E.D.O.*, Mem. AMS **296** (1984).
- [Ra2] J.-P. RAMIS - *Filtration Gevrey sur le groupe de Picard-Vessiot d'une E.D. irrégulière*, Prépubl. IMPA **45** (1985), 1-38.
- [RS] J.-P. RAMIS and Y. SIBUYA - *Hukuhara's domains and fundamental existence and uniqueness theorems for asymptotic solutions of Gevrey type*, Asymp. Anal. **2** (1989), 39-94.
- [Sa] C. SABBABH - *Connexions méromorphes formelles de deux variables*, Prépubl. Éc. Polytechnique (1991).
- [Sc] H. SCHWARZ - *Ueber diejenige Fälle in welchen die Gaussche hypergeometrische Reihe einer algebraische Funktion ihres 4.en Elementes darstellt*, Crelle **75** (1873), 292-335.
- [Se] J.-P. SERRE - *Quelques applications du théorème de densité de Chebotarev*, Publ. Math. IHES **54** (1981), 123-201.
- [Si] Y. SIBUYA - *L.D.E. in the complex domain : problems of analytic continuation*, Transl. AMS, vol. **82** (1991).
- [Sn1] M. SINGER - *Algebraic relations among solutions of L.D.E. : Fano's theorem*, Amer. J. Math. **110** (1988), 115-144.
- [Sn2] M. SINGER - *An outline of differential Galois theory*, Computer Algebra and Differential Equations, E. Tournier (ed.), Acad. Press (1989).
- [Sn3] M. SINGER - *Moduli of L.D.E. on the Riemann sphere with fixed Galois group*, Prépubl. N.C.S.U. (1991).

- [St] G. STOKES - *Early letters to Lady Stokes*, Cambridge U.P. (1907) ; cité par J.-P. Ramis : *Séries divergentes et théories asymptotiques*, Journées X-UPS, 1991 (à paraître).
- [Um] H. UMEMURA - *Second proof of the irreducibility of the first D.E. of Painlevé*, Nagoya Math. J. **113** (1990).
- [Va] V. VARADARAJAN - *Meromorphic differential equations*, Expos. Math. **9** (1991), 97-188.

[Ajouté en octobre 1992]

- (1) c'est-à-dire un homomorphisme C -linéaire de V dans $V \otimes \Omega_{K/C}^1$ vérifiant la formule de Leibniz.
- (2) Par "constructions tensorielles", on entend ici les vectoriels à connexion que portent naturellement les sommes directes de V , de son dual et des produits tensoriels de leurs puissances tensorielles. Outre [De], voir à ce propos L. Breen, *Tannakian categories*, Prépubl. U. Paris 13, n° 13, 1992 (à paraître aux Proc. Symp. AMS, "Motives", Seattle, 1991).
- (3) Dans cet énoncé, on suppose fixés des choix compatibles des groupes locaux dans leurs classes de conjugaison dans $\text{Gal}(D)$: par exemple, si on considère la représentation de $\text{Gal}(D)$ (resp. de $G_{\sigma,an}$) attachée à un point ordinaire s de V (resp. un point σ' proche de σ), tout système de chemins de s vers chacun des points σ' fournit un tel choix.
- (4) et vérifier, comme dans [IK], que ce relèvement est compatible aux constructions tensorielles, donc galoisien. Cette méthode a été développée dans le cas général par P. Deligne (lettre à J.-P. Ramis, Fév. 86) en termes des systèmes locaux I -filtrés (cf. [Ma1], [B-V2]) qu'il associe au couple $\{V, \widehat{\Phi}\}$.
- (5) Ici encore, il convient de fixer des choix compatibles des st_α dans leurs classes de conjugaisons : pour la représentation de $G_{0,an}$ attachée à un point s voisin de 0, on peut par exemple fixer des points σ d'arguments approchant par valeurs inférieures chaque relevé à \widetilde{S} des directions singulières α , et un système de chemins reliant s aux σ . La conjugaison par $\Sigma_{\arg(s)}$ permet par ailleurs de transporter $G_{0,f}(D)$ en un sous-groupe de cette représentation de $G_{0,an}$; c'est lui qui intervient dans l'énoncé du théorème 2.
- (6) $(\widehat{V}_0, \widehat{D}_0)$ provient par complétion du K -vectoriel à connexion (V_0, D_0) attaché

à l'opérateur différentiel D_0 " =" \widehat{D}_0 sur $V_0 = K^n$. Noter que l'injection canonique de K dans \widehat{K} se prolonge en des homomorphismes injectifs des extensions de Picard-Vessiot de V_0/K vers celles de $\widehat{V}_0/\widehat{K}$. On a fixé l'un d'eux au cours de cette partie.

- (7) Voir aussi M. Loday-Richaud : *Stokes phenomenon, multisummability and differential Galois groups*, Prépubl. Orsay n° 1371, 1992, à paraître.
- (8) Voir néanmoins les travaux de M. Singer (J. Symb. Comp. **11** (1991), 251-273), F. Ulmer (A.A.E.C.C. **2** (1991), 251-273) et de A. Duval et M. Loday-Richaud (A.A.E.C.C. **3** (1992), 1-36), sur l'algorithme de Kovacic.

Daniel BERTRAND

Université de Paris VI

U.F.R. de Mathématiques

Tour 46

4, place Jussieu

F-75252 PARIS CEDEX 05

Astérisque

JEAN-MARC FONTAINE

Valeurs spéciales des fonctions L des motifs

Astérisque, tome 206 (1992), Séminaire Bourbaki,
exp. n° 751, p. 205-249

<http://www.numdam.org/item?id=SB_1991-1992__34__205_0>

© Société mathématique de France, 1992, tous droits réservés.

L'accès aux archives de la collection « Astérisque » (<http://smf4.emath.fr/Publications/Asterisque/>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

VALEURS SPÉCIALES DES FONCTIONS L DES MOTIFS

par Jean–Marc FONTAINE

Il y a tout juste sept ans que Soulé a exposé dans un séminaire Bourbaki [So85] les conjectures de Deligne et Beilinson¹ donnant, entre autres, une formule pour $L(H^m(X), r)$, où X est une variété projective lisse sur F et où $r \in \mathbf{Z}$, à multiplication par un nombre rationnel près. Les conjectures de Bloch et Kato ([BK90]), qui sont l'objet central de ce rapport, prédisent ce que devrait être ce nombre rationnel.

J'ai fait le choix discutable de privilégier les conjectures aux dépens des résultats. Ces derniers sont encore très partiels (voir à ce sujet les rapports de B. Perrin–Riou [Pe89], Gross [Gr91] et Rubin [Ru91] sur la conjecture de Birch et Swinnerton–Dyer; le livre [RSS88] et les rapports de Ramakrishnan [Ra89] et de Deninger et Scholl [DS91] sur les conjectures de Deligne et Beilinson) mais suffisamment encourageants pour que l'on puisse penser que la situation devrait évoluer rapidement.

§ 1. — Introduction

Dans tout ce qui suit E et F sont deux extensions finies de \mathbf{Q} . On peut espérer avoir un jour une bonne définition de la catégorie $\mathcal{MM}_F(E)$ des motifs mixtes sur F à coefficients dans E . A un tel motif M , on est sensé savoir associer une fonction $L(M, s)$ holomorphe pour $Re(s) \gg 0$ à valeurs dans $\mathbf{C} \otimes E$. On s'attend à ce que cette fonction admette un prolongement méromorphe dans tout le plan complexe et on voudrait donner une interprétation, au moins conjecturale, de ses valeurs aux entiers naturels; autrement dit on voudrait généraliser les conjectures de Birch et Swinnerton–Dyer qui correspondent à $L(M, 0)$ lorsque M est le motif pur de poids -1 associé à une variété abélienne définie sur F (avec $E = \mathbf{Q}$). Citons comme exemples de fonctions L que l'on rencontre dans ce cadre les fonctions L d'Artin associées aux représentations finies de $\text{Gal}(\overline{F}/F)$ à coefficients dans

¹ voir aussi les articles originaux de Bloch [Bl84], Deligne [De79] et Beilinson [Be85].

E , celles que l'on associe aux formes modulaires "nouvelles", les fonctions $L(H^m(X), s)$, où X est une variété projective non singulière sur F et $m \in \mathbf{N}$.

En remplaçant M par son "tordu à la Tate" $M(i)$, on se ramène à étudier la valeur en $s = 0$. Soit r_M l'unique entier tel que $s^{-r_M} L(M, s)$ tende vers une limite finie non nulle $L^*(M, 0)$ lorsque $s \rightarrow 0$.

Notons $\mathbb{1}_{F,E}$ "l'objet unité" de $\mathcal{MM}_F(E)$ et, pour tout $i \in \mathbf{N}$, posons

$$H^i(F, M) = \text{Ext}_{\mathcal{MM}_F(E)}^i(\mathbb{1}_{F,E}, M).$$

On peut définir (cf. § 6 ci-dessous) le sous- E -espace vectoriel $H_f^1(F, M)$ de $H^1(F, M)$ classifiant les extensions de $\mathbb{1}_{F,E}$ par M ayant "bonne réduction en toutes les places finies de F ".

"CONJECTURE" $C_r(M)^2$. — Si $M^*(1)$ est le "dual tordu à la Tate de M ", alors

$$r_M = \dim_E H_f^1(F, M^*(1)) - \dim_E H^0(F, M(1)).$$

On dispose de la **réalisation de Rham** M_{dR} de M , qui est un $E \otimes_{\mathbf{Q}} F$ -module de type fini, muni d'une filtration décroissante $(\text{Fil}^i M_{dR})_{i \in \mathbf{Z}}$ par des sous- $E \otimes F$ -modules, la **filtration de Hodge**. On note t_M et on appelle **espace tangent de M** le quotient $M_{dR}/\text{Fil}^0 M_{dR}$. Pour tout $\mathfrak{p} \in S_{\infty}(F)$, ensemble des places infinies de F , on dispose de la **réalisation Betti** $M_{B,\mu}$ de M qui est un E -espace vectoriel de dimension finie muni d'une action de la conjugaison complexe c si \mathfrak{p} est réelle. On pose $M_{B,\mu}^+ = (M_{B,\mu})^c$ si \mathfrak{p} est réelle et $M_{B,\mu}$ sinon.

Pour tout corps K et tout K -espace vectoriel V de dimension finie, notons $\det_K V = \Lambda_K^{\dim V} V$ le déterminant de V et $\det_K^* V$ le dual de $\det_K V$. Posons

$$L_f(M) = \det_E H^0(F, M) \otimes \det_E^* H_f^1(F, M),$$

$$L_f(M^*(1)) = \det_E H^0(F, M^*(1)) \otimes \det_E^* H_f^1(F, M^*(1))$$

et introduisons enfin la **droite fondamentale de M**

$$\Delta_f(M) = L_f(M) \otimes L_f(M^*(1)) \otimes \det_E^*(\bigoplus_{\mu \in S_{\infty}(F)} M_{B,\mu}^+) \otimes \det_E t_M.$$

On va voir que l'on peut, modulo certaines propriétés "conjecturales" de $\mathcal{MM}_F(E)$,

² Comme, dans la pratique, on ne sait pas toujours que la fonction L admet un prolongement analytique dans \mathbf{C} , il faut comprendre cette conjecture comme disant que la fonction $L(M, s)$, définie pour $\text{Re}(s) \gg 0$, admet un prolongement méromorphe dans un ouvert connexe contenant 0 et que son ordre en $s = 0$ est donné par cette formule.

(a) définir un isomorphisme canonique

$$\iota_M : \mathbf{R} \otimes_{\mathbf{Q}} \Delta_f(M) \longrightarrow \mathbf{R} \otimes_{\mathbf{Q}} E ;$$

(b) pour toute place finie λ de E , utilisant la **réalisation λ -adique** M_λ de M , munir $E_\lambda \otimes_E \Delta_f(M)$ d'une norme canonique $|\cdot|_{\lambda, EP}$.

“CONJECTURE” $C_{DB}(M)$. — *Il existe une base b de $\Delta_f(M)$ (nécessairement unique) telle que*

$$\iota_M(1 \otimes b).L^*(M, 0) = 1 .$$

“CONJECTURE” $C_{BK}(M)$. — $C_{DB}(M)$ est vraie et pour toute place finie λ de E ,

$$|1 \otimes b|_{\lambda, EP} = 1 .$$

Bien sûr $C_{DB}(M)$ (resp. $C_{BK}(M)$) détermine $L^*(M, 0)$ à multiplication par un élément de E^* (resp. \mathcal{O}_E^* près). Lorsque $E = \mathbf{Q}$, $C_{BK}(M)$ détermine donc le nombre réel $L^*(M, 0)$ au signe près (celui-ci est “facile” à déterminer : on s'attend à ce que les zéros et les pôles de $L(M, s)$ sur la droite réelle soient tous entiers ; on a $r_{M(i)} = 0$ pour $i \gg 0$ et le signe devrait être égal à la parité de $\sum_{i>0} r_{M(i)}$).

J'ai mis ces conjectures entre guillemets car elles concernent des objets de la catégorie $\mathcal{MM}_F(E)$ dont je ne connais pas de définition précise. Toutefois, on connaît beaucoup d'objets de cette catégorie.

Soit M un “motif connu”. Dans deux cas au moins, on peut obtenir une “vraie” conjecture

– celui où M est f -clos, i.e. tel que $H_f^1(F, M) = H_f^1(F, M^*(1)) = 0^3$ (lorsque M est pur de poids $\neq -1$, cela revient à dire que M est **critique**);

– celui où M est le motif pur de poids -1 associé à une variété abélienne A sur F , munie d'un plongement de E dans $\mathbf{Q} \otimes \text{End}(A)$. La théorie des 1-**motifs** de Deligne fournit alors une sous-catégorie pleine parfaitement bien définie de $\mathcal{MM}_F(E)$ qui est suffisamment grosse pour tous les calculs que l'on veut faire (cf. § 8).

Sinon, et c'est le point de vue “classique” de Beilinson, on essaie de s'en sortir en utilisant la “**cohomologie motivique**”, i.e. en introduisant des E -espaces vectoriels, définis via la K -théorie algébrique, dont on conjecture qu'ils s'identifient à $H_f^1(F, M)$ et $H_f^1(F, M^*(1))$. On a en outre besoin de

³ Si le formalisme conjectural qui sous-tend la catégorie $\mathcal{MM}_F(E)$ est vraie, il existe une recette universelle pour associer à tout motif M un motif f -clos N et les conjectures sont vraies pour M si et seulement si elles le sont pour N .

connaître certaines applications E -linéaires qui ont pour origine l'un de ces deux espaces vectoriels, ce qui conduit à construire des “régulateurs” et, dans certains cas, l’analogue de la hauteur de Néron–Tate des variétés abéliennes. Le cas le plus important est celui où $M = H^m(X)(r)$, avec X variété projective lisse sur F , $m \in \mathbf{N}$, $r \in \mathbf{Z}$ (et $E = \mathbf{Q}$) (cf. § 9).

Après avoir fixé quelques notations et conventions (§ 2), on attaque la **première partie**, consacrée au **formalisme général**, en faisant quelques “rappels” sur les représentations galoisiennes à coefficients dans E_λ (§ 3); puis on explique (§ 4) comment définir la norme canonique $|\cdot|_{\lambda,EP}$ à partir de la réalisation λ -adique; ce sont plus ou moins les structures de Hodge mixtes qui jouent le rôle de la réalisation λ -adique aux places archimédiennes et on en discute au § 5; puis on décrit (§ 6) les propriétés formelles que devrait satisfaire la catégorie $\mathcal{MM}_F(E)$, on définit $H_f^1(F, M)$ et on construit l’application ι_M utilisée dans la conjecture $C_{BD}(M)$; on discute au § 7 comment ces conjectures se simplifient lorsque l’on fait certaines restrictions sur les poids de M ou bien lorsque l’on suppose $H_f^1(F, M) = H_f^1(F, M^*(1)) = 0$.

Dans la **deuxième partie**, on discute des **cas particuliers** : 1-motifs (§ 8), motifs de la forme $H^m(X)$, avec X projective lisse sur F (§ 9), motifs mixtes de Tate, motifs d’Artin tordus, \dots (§ 10).

La **troisième partie** contient divers compléments : au § 11, on explique comment, pour calculer les nombres qui interviennent dans $C_{BK}(M)$, on est conduit à introduire des groupes de Safarevic et des nombres de Tamagawa. Ceci conduit à une autre formulation de $C_{BK}(M)$ qui est (essentiellement) celle de Birch et Swinnerton–Dyer dans le cas des variétés abéliennes et la conjecture originale de Bloch et Kato ([BK90]) dans les cas qu’ils considèrent. Au § 12 enfin, on discute de la notion de E_λ -représentations et de motifs **semi-stables** (on s’attend à ce que tout motif sur F devienne semi-stable sur une extension finie de F). C’est pour les motifs semi-stables que l’on peut espérer développer des techniques permettant de faire progresser de façon significative cette théorie. C’est aussi dans ce contexte que l’on peut interpréter les valeurs de fonctions L incomplètes, comprendre pourquoi les conjectures discutées ici se comportent bien par suite exacte courte, sont compatibles avec l’équation fonctionnelle et comment elles suggèrent certaines propriétés de la catégorie des motifs mixtes (dimension cohomologique 1, F -semi-simplicité, \dots).

Ces “conjectures” et/ou leur différentes traductions sont le résultat (si l’on ose dire) des efforts de beaucoup. Après le travail initial de Birch et Swinnerton–Dyer ([BS63], [BS65]) étendu par Tate [Ta66], les contributions les plus importantes sont sans doute celles de Bloch [Bl84], de Beilinson [Be85]

et Deligne [De79] pour $C_{BD}(M)$ et de Bloch et Kato [BK90] pour $C_{BK}(M)$. Dans le point de vue développé ici certaines idées de Bloch [Bl80], Deligne [De85], Jannsen ([Ja88], [Ja89], [Ja90]), Kato [Ka91], Lichtenbaum [Li72], Perrin–Riou [FP91] et Scholl [Sc91] ont également joué un grand rôle⁴.

Je voudrais enfin remercier Bernadette Perrin–Riou à qui cet exposé doit beaucoup : cela fait longtemps que nous discutons ensemble de ce sujet et non seulement j’ai bénéficié de ses conseils mais je me suis largement inspiré de nos rédactions communes.

§2. — Conventions et notations

Si G est un groupe profini et λ une place finie de E , une **représentation λ -adique de G** est un E_λ -espace vectoriel V de dimension finie muni d’une action linéaire et continue de G . On a $H^0(G, V) = V^G$ tandis que $H^1(G, V)$ classe les extensions de la représentation triviale E_λ par V . Lorsque $G = \text{Gal}(\bar{K}/K)$ où \bar{K} est une clôture séparable d’un corps K , on écrit aussi $H^i(K, V)$ au lieu de $H^i(G, V)$.

Pour tout corps de nombres K , on note $S(K)$ l’ensemble des places de K . On identifie $S(\mathbb{Q})$ à $\{\text{nombre premiers}\} \cup \{\infty\}$. Pour tout $p \in S(\mathbb{Q})$, on note $S_p(K)$ l’ensemble des places de K au-dessus de p .

Pour fixer les idées, on choisit une clôture algébrique \bar{F} de F et, pour chaque $\mathfrak{p} \in S(F)$, une clôture algébrique $\bar{F}_\mathfrak{p}$ du complété $F_\mathfrak{p}$ et un plongement de \bar{F} dans $\bar{F}_\mathfrak{p}$. On pose $G_F = \text{Gal}(\bar{F}/F)$, $G_\mathfrak{p} = \text{Gal}(\bar{F}_\mathfrak{p}/F_\mathfrak{p})$ ($\subset G_F$). Si \mathfrak{p} est infinie, on choisit une identification de $\bar{F}_\mathfrak{p}$ à \mathbb{C} . Si \mathfrak{p} est une place finie au-dessus du nombre premier p , on note $I_\mathfrak{p}$ le sous-groupe d’inertie de $G_\mathfrak{p}$; pour toute extension L de $F_\mathfrak{p}$ contenue dans $\bar{F}_\mathfrak{p}$, on note L_0 l’extension maximale non ramifiée de \mathbb{Q}_p contenue dans L ; le corps résiduel $\bar{k}_\mathfrak{p}$ de $\bar{F}_\mathfrak{p}$ est le même que celui de $(\bar{F}_\mathfrak{p})_0$ et est une clôture algébrique du corps résiduel $k_\mathfrak{p}$ de $F_\mathfrak{p}$; on note $\sigma_\mathfrak{p} : \bar{k}_\mathfrak{p} \rightarrow \bar{k}_\mathfrak{p}$ l’application $x \mapsto x^p$ ainsi que l’automorphisme continu de $(\bar{F}_\mathfrak{p})_0$ qui le relève; le groupe $G_\mathfrak{p}/I_\mathfrak{p}$ s’identifie à $\text{Gal}(\bar{F}_\mathfrak{p})_0/(F_\mathfrak{p})_0 = \text{Gal}(\bar{k}_\mathfrak{p}/k_\mathfrak{p})$ et est topologiquement engendré par le Frobenius géométrique $f_\mathfrak{p} = \sigma_\mathfrak{p}^{-r}$ où $r_\mathfrak{p} = [k_\mathfrak{p} : \mathbb{F}_p]$.

⁴ L’exposition s’inspire largement de [FP91], [FP92] et de Kato [Ka91]. Le point de vue original de Bloch et Kato est expliqué au § 11. Les conjectures originales concernaient des situations moins générales et le rapporteur revendique la responsabilité de toutes les absurdités qui pourraient résulter d’un changement de point de vue ou de généralisations intempestives.

A – Le formalisme général

§3. — La fonction L et la droite fondamentale d'une représentation λ -adique de G_F .

Dans ce paragraphe et dans le suivant, λ est une place finie de E au-dessus du nombre premier ℓ .

3.1. — Soit \mathfrak{p} une place finie de F au-dessus du nombre premier p . On renvoie à [FI90], [I90] ou [Bures] pour la définition du corps $B_{dR,\mathfrak{p}}$ des périodes p -adiques et de son sous-anneau $B_{cris,\mathfrak{p}}$ correspondant à l'extension $\overline{F}_{\mathfrak{p}}/F_{\mathfrak{p}}$. Le corps $B_{dR,\mathfrak{p}}$ est une $\overline{F}_{\mathfrak{p}}$ -algèbre munie d'une filtration décroissante $(Fil^i B_{dR,\mathfrak{p}})_{i \in \mathbb{Z}}$ et d'une action de $G_{\mathfrak{p}}$. L'anneau $B_{cris,\mathfrak{p}}$ est une sous- $(\overline{F}_{\mathfrak{p}})_0$ -algèbre de $B_{dR,\mathfrak{p}}$ stable par $G_{\mathfrak{p}}$ et munie d'un endomorphisme φ commutant à l'action de $G_{\mathfrak{p}}$ et $\sigma_{\mathfrak{p}}$ -semi-linéaire. On a $(B_{dR,\mathfrak{p}})^{G_{\mathfrak{p}}} = F_{\mathfrak{p}}$, tandis que $(B_{cris,\mathfrak{p}})^{G_{\mathfrak{p}}} = (F_{\mathfrak{p}})_0$.

3.2. — On pose $E_{\lambda,\mathfrak{p}} = E_{\lambda}$ si $\ell \neq p$ et $= (F_{\mathfrak{p}})_0 \otimes_{\mathbb{Q}_p} E_{\lambda}$ sinon.

Soit V une représentation λ -adique de $G_{\mathfrak{p}}$ de dimension $b(V)$. Posons

$$\underline{D}_{\mathfrak{p}}(V) = V^I_{\mathfrak{p}} \text{ si } \ell \neq p \text{ et } \underline{D}_{\mathfrak{p}}(V) = (B_{cris,\mathfrak{p}} \otimes_{\mathbb{Q}_p} V)^{G_{\mathfrak{p}}} \text{ si } \ell = p .$$

C'est un $E_{\lambda,\mathfrak{p}}$ -module libre de rang fini $b_{\mathfrak{p}}(V) \leq b(V)$ muni d'une action **linéaire** de $f_{\mathfrak{p}}$: si $\ell \neq p$, cette action vient de ce que $f_{\mathfrak{p}}$ appartient à $G_{\mathfrak{p}}/I_{\mathfrak{p}}$ qui agit sur $\underline{D}_{\mathfrak{p}}(V)$; si $\ell = p$, φ agit sur $B_{cris,\mathfrak{p}} \otimes V$ via $b \otimes v \mapsto \varphi b \otimes v$, $\underline{D}_{\mathfrak{p}}(V)$ est stable par φ et on fait agir $f_{\mathfrak{p}}$ via $\varphi^{r_{\mathfrak{p}}}$ ⁵.

On dit que V a **bonne réduction** si $b_{\mathfrak{p}}(V) = b(V)$. Si $\ell \neq p$, cela signifie que V est non ramifiée, tandis que, si $\ell = p$, cela signifie que V est cristalline. On note $H^1_f(F_{\mathfrak{p}}, V)$ le sous- E_{λ} -espace vectoriel de $H^1(F_{\mathfrak{p}}, V)$ formé des classes d'extensions W de E_{λ} par V telles que l'application naturelle $\underline{D}_{\mathfrak{p}}(W) \rightarrow \underline{D}_{\mathfrak{p}}(E_{\lambda}) = E_{\lambda,\mathfrak{p}}$ soit surjective; lorsque V a bonne réduction, cela revient à demander que W aussi.

En outre, si $\ell = p$, on pose $\underline{D}_{dR,\mathfrak{p}}(V) = (B_{dR,\mathfrak{p}} \otimes_{\mathbb{Q}_p} V)^{G_{\mathfrak{p}}}$. C'est un $(F_{\mathfrak{p}} \otimes_{\mathbb{Q}_p} E_{\lambda})$ -module muni d'une filtration décroissante, indexée par \mathbb{Z} , par les sous- $F_{\mathfrak{p}} \otimes E_{\lambda}$ -modules $Fil^i \underline{D}_{dR,\mathfrak{p}}(V) = (Fil^i B_{dR,\mathfrak{p}} \otimes_{\mathbb{Q}_p} V)^{G_{\mathfrak{p}}}$. La dimension de $\underline{D}_{dR,\mathfrak{p}}(V)$ comme $F_{\mathfrak{p}}$ -espace vectoriel est $\leq [E_{\lambda} : \mathbb{Q}_p] \cdot b(V)$. On dit que V est **de de Rham** si l'on a l'égalité, auquel cas $\underline{D}_{dR,\mathfrak{p}}(V)$ est libre de rang $b(V)$ sur $F_{\mathfrak{p}} \otimes E_{\lambda}$ et l'application naturelle

$$B_{dR,\mathfrak{p}} \otimes_{F_{\mathfrak{p}}} \underline{D}_{dR,\mathfrak{p}}(V) \longrightarrow B_{dR,\mathfrak{p}} \otimes_{\mathbb{Q}_p} V$$

est un isomorphisme.

⁵ sic, i.e. il faut prendre $f_{\mathfrak{p}} = \varphi^{r_{\mathfrak{p}}}$ et non son inverse!

Enfin, toujours si $\ell = p$, on appelle **espace tangent de V** le $F_{\mathfrak{p}} \otimes E_{\lambda}$ -module $t_{V,\mathfrak{p}} = ((B_{dR,\mathfrak{p}}/Fil^0 B_{dR,\mathfrak{p}}) \otimes_{\mathbb{Q}_p} V)^{G_{\mathfrak{p}}}$. Lorsque V est de de Rham, $t_{V,\mathfrak{p}}$ s'identifie à $\underline{D}_{dR,\mathfrak{p}}(V)/Fil^0 \underline{D}_{dR,\mathfrak{p}}(V)$.

3.3. — Soit maintenant V une représentation λ -adique de G_F . Pour toute place finie \mathfrak{p} de F , on pose

$$P_{\mathfrak{p}}(V, t) = \det_{E_{\lambda,\mathfrak{p}}}(1 - f_{\mathfrak{p}} t \mid \underline{D}_{\mathfrak{p}}(V));$$

c'est un polynôme à coefficients dans E_{λ} (y compris lorsque $\ell = p$); on a $d^{\circ} P_{\mathfrak{p}} \leq b(V)$ avec l'égalité si et seulement si V a bonne réduction en \mathfrak{p} .

On va avoir besoin des trois définitions suivantes :

i) Soit S un ensemble fini de places de F contenant $S_{\infty}(F)$. On dit que V **admet une fonction L_S sur E** si, pour tout $\mathfrak{p} \notin S$, $P_{\mathfrak{p}}(V, t) \in E[t]$ et si le produit infini

$$L_S(V, s) = \prod_{\mathfrak{p} \notin S} P_{\mathfrak{p}}(V, N(\mathfrak{p})^{-s})^{-1}$$

converge dans $\mathbb{C} \otimes_{\mathbb{Q}} E$ pour $Re(s) \gg 0$. Si $S = S_{\infty}(F)$, on dit que V admet une fonction L sur E et on pose $L(V, s) = L_{S_{\infty}(F)}(V, s)$.

ii) On dit que V est **pseudo-géométrique** si elle est non ramifiée en dehors d'un nombre fini de places finies de F et de Rham en toutes les places divisant ℓ .

iii) **On note $H_f^1(F, V)$** le sous- E_{λ} -espace vectoriel de $H^1(F, V)$ formé des éléments tels que, pour tout \mathfrak{p} , l'image dans $H^1(F_{\mathfrak{p}}, V)$ appartient en fait à $H_f^1(F_{\mathfrak{p}}, V)$. C'est un E_{λ} -espace vectoriel de dimension finie.

La représentation triviale E_{λ} est pseudo-géométrique, admet une fonction L sur E , c'est la fonction ζ de Riemann usuelle, et on a $H_f^1(F, E_{\lambda}) = 0$. Le fait d'être pseudo-géométrique est stable par sous-objet, quotient, somme directe, produit tensoriel, dual. Si V est pseudo-géométrique, pour tout $r \in \mathbb{Z}$, $V(r)^6$ l'est aussi et $L(V(r), s) = L(V, r + s)$.

3.4. — Pour toute représentation λ -adique pseudo-géométrique V de G_F , on pose $L_f(V) = \det_{E_{\lambda}} H^0(F, V) \otimes \det_{E_{\lambda}}^* H_f^1(F, V)$, $V^+ = \bigoplus_{\mathfrak{p} \in S_{\infty}(F)} H^0(G_{\mathfrak{p}}, V)$ et $t_V = \bigoplus_{\mathfrak{p} \in S_{\ell}(F)} t_{V,\mathfrak{p}}$. On introduit la **droite fondamentale de V** ,

$$\Delta_f(V) = L_f(V) \otimes L_f(V^*(1)) \otimes \det_{E_{\lambda}}^* V^+ \otimes \det_{E_{\lambda}} t_V .$$

⁶ Comme d'habitude $\mathbb{Z}_{\ell}(1) = \varprojlim_{\mu \in \mathbb{N}} (\overline{\mathbb{Q}})$, si $r \in \mathbb{N}$, $\mathbb{Z}_{\ell}(r) = \text{Sym}_{\mathbb{Z}_{\ell}}^r \mathbb{Z}_{\ell}(1)$ et $\mathbb{Z}_{\ell}(-r)$ est son dual; enfin $V(r) = V \otimes_{\mathbb{Z}_{\ell}} \mathbb{Z}_{\ell}(r)$ pour tout $r \in \mathbb{Z}$.

L'objectif du paragraphe suivant est de munir cette droite d'une norme canonique.

§ 4. — **La norme d'Euler–Poincaré pour une représentation λ -adique pseudo-géométrique**⁷

4.1. — Soit \mathcal{O}_λ l'anneau des entiers de E_λ . On note $D_{par}(\mathcal{O}_\lambda)$ (resp. $D_{par}(E_\lambda)$) la catégorie dérivée de la catégorie des complexes parfaits de \mathcal{O}_λ -modules (resp. E_λ -espaces vectoriels).

Si $d \in \mathbf{N}$, $i \in \mathbf{Z}$ et M est un \mathcal{O}_λ -module libre de rang d ou un E_λ -espace vectoriel de dimension d , on note $\det_\lambda^i M$ la puissance extérieure d -ième de M si i est pair et son dual si i est impair. Si

$$C : \dots \longrightarrow C^{i-1} \longrightarrow C^i \longrightarrow C^{i+1} \longrightarrow \dots$$

est un complexe parfait de \mathcal{O}_λ -modules (resp. E_λ -espaces vectoriels), on pose $\det_\lambda C = \otimes_{i \in \mathbf{Z}} \det_\lambda^i C^i$ et on l'appelle le déterminant de C . Tout quasi-isomorphisme induit un isomorphisme des déterminants, ce qui permet de parler du **déterminant** d'un objet de $D_{par}(\mathcal{O}_\lambda)$ (ou de $D_{par}(E_\lambda)$). En particulier, on peut parler du déterminant d'un \mathcal{O}_λ -module de type fini (si $M = M_0/M_{-1}$, avec M_{-1} et M_0 libres de type fini, $\det_\lambda M = \det_\lambda^0 M_0 \otimes \det_\lambda^{-1} M_{-1}$). Si

$$C' \longrightarrow C \longrightarrow C''$$

est un triangle distingué de $D_{par}(\mathcal{O}_\lambda)$ (ou de $D_{par}(E_\lambda)$), on peut identifier $\det_\lambda C$ à $\det_\lambda C' \otimes \det_\lambda C''$ ⁸. Enfin, l'extension des scalaires définit un foncteur de $D_{par}(\mathcal{O}_\lambda)$ dans $D_{par}(E_\lambda)$ et, pour tout C dans $D_{par}(\mathcal{O}_\lambda)$, $\det_\lambda(E_\lambda \otimes C)$ s'identifie à $E_\lambda \otimes \det_\lambda C$, ce qui permet d'identifier $\det_\lambda C$ à un réseau de $\det_\lambda(E_\lambda \otimes C)$.

4.2. — Dans ce qui suit, S est un ensemble fini de places de F contenant toutes les places infinies et toutes les places divisant ℓ . On note S_f l'ensemble des places finies de S , \mathcal{O}_F l'anneau des entiers de F et U_S l'ouvert de $\text{Spec } \mathcal{O}_F$ qui est le complémentaire de S_f . L'expression "faisceau" (resp. " \mathcal{O}_λ -faisceau",

⁷ Voir [FP91], note I, ou [Ka91], § 3 pour une présentation plus élémentaire en termes de cohomologie galoisienne. Je dois à Deligne le point de vue décrit ici qui me paraît à la fois plus simple et plus naturel.

⁸ Dans ces histoires de déterminants, il y a des problèmes de signes (si $V = \bigoplus_{i \in I} V_i$ est une somme directe finie, l'identification de $\det_\lambda V$ au produit tensoriel des $\det_\lambda V_i$ dépend, par un signe, de l'ordre dans lequel on écrit les V_i) qui sont sans importance pour ce que nous avons en vue et que nous négligerons.

“ E_λ -faisceau”) signifie faisceau (resp. \mathcal{O}_λ -faisceau, E_λ -faisceau) constructible pour la topologie étale⁹.

Pour tout \mathcal{O}_λ -faisceau (resp. E_λ -faisceau) T sur U_S , $R\Gamma_{\acute{e}t}(U_S, T)$ est un objet de $D^+(\mathcal{O}_\lambda)$ ¹⁰ (resp. $D_{par}(E_\lambda)$)¹¹. Pour tout $\mathfrak{p} \in S_f$, il en est de même de $R\Gamma(F_{\mathfrak{p}}, T)$ (défini comme $R\Gamma_{\acute{e}t}(\text{Spec}(F_{\mathfrak{p}}), \nu_{\mathfrak{p}}^* j_* T)$, où $\nu_{\mathfrak{p}} : \text{Spec}(F_{\mathfrak{p}}) \rightarrow \text{Spec}(\mathcal{O}_F)$ et $j : U_S \rightarrow \text{Spec}(\mathcal{O}_F)$ sont les morphismes naturels). La cohomologie relative $R\Gamma_c(U_S, T)$ est $C[-1]$ où C est le cône de $R\Gamma_{\acute{e}t}(U_S, T) \rightarrow \bigoplus_{\mathfrak{p} \in S} R\Gamma(F_{\mathfrak{p}}, T)$. On a $H_c^i(U_S, T) = 0$ si $i \neq 1, 2, 3$ et $R\Gamma_c(U_S, T)$ est un objet de $D_{par}(\mathcal{O}_\lambda)$ (resp. $D_{par}(E_\lambda)$).

On peut donc définir la **droite d’Euler–Poincaré** de T ,

$$\Delta_{\lambda, EP}(T) = \det_{\lambda} R\Gamma_c(U_S, T).$$

Il n’est pas difficile de voir que ce déterminant ne dépend que de la fibre générique de T . Si T est de torsion, on a $\Pi_i \# H_c^i(U_S, T)^{(-1)^i} = 1$ et, par conséquent, $\Delta_{\lambda, EP}(T) = \mathcal{O}_\lambda$. On en déduit que, si V est un E_λ -faisceau sur U_S , si l’on choisit un \mathcal{O}_λ -faisceau T tel que $E_\lambda \otimes T = V$, le réseau $\Delta_{\lambda, EP}(T)$ de $\Delta_{\lambda, EP}(V)$ obtenu est indépendant du choix de T . On note $|\cdot|_{\lambda, EP}$ l’unique **norme sur** $\Delta_{\lambda, EP}(V)$ telle que, si δ est une base de $\Delta_{\lambda, EP}(T)$, alors $|\delta|_{\lambda, EP} = 1$.

Il n’est pas non plus difficile de voir que, si S' est un ensemble fini de places de F contenant S et si V' est un E_λ -faisceau sur $U_{S'}$ qui a la même fibre générale que V , alors $\Delta_{\lambda, EP}(V')$ **s’identifie à** $\Delta_{\lambda, EP}(V)$ **en tant que** E_λ -**droite normée**. Ceci permet en particulier de définir la E_λ -droite normée $\Delta_{\lambda, EP}(V)$ pour n’importe quelle représentation λ -adique de G_F non ramifiée en dehors d’un nombre fini de places : il suffit de choisir un ensemble S comme ci-dessus et un E_λ -faisceau sur U_S de fibre générique V ; le résultat obtenu est indépendant de ces choix.

4.4. — Soit maintenant \mathfrak{p} une place finie de F divisant le nombre premier p et soit V une représentation λ -adique de $G_{\mathfrak{p}}$. On “coupe” le complexe $R\Gamma(F_{\mathfrak{p}}, V)$

⁹ Voir [SGA4], exp. IV, [SGA5], appendice à l’exp. I, exp. V et VI et [SGA4 $\frac{1}{2}$] pour les généralités sur les faisceaux de torsion, \mathcal{O}_λ -faisceaux et E_λ -faisceaux constructibles.

¹⁰ et même de $D_{par}(\mathcal{O}_\lambda)$ si $\ell \neq 2$.

¹¹ Appelons \mathcal{O}_λ -représentation d’un groupe profini G tout \mathcal{O}_λ -module de type fini muni d’une action linéaire et continue de G . Soit G_S le groupe de Galois de l’extension galoisienne maximale de F contenue dans \overline{F} non ramifiée en dehors de S . Si T est un \mathcal{O}_λ -faisceau (resp. un E_λ -faisceau), la fibre générique de T est une \mathcal{O}_λ -représentation (resp. une représentation λ -adique) de G_S . Inversement toute \mathcal{O}_λ -représentation (resp. représentation λ -adique) T de G_S est la fibre générique d’un unique \mathcal{O}_λ -faisceau (resp. E_λ -faisceau) localement constant que nous notons encore T et $R\Gamma_{\acute{e}t}(U_S, T)$ s’identifie au complexe calculant la cohomologie galoisienne continue $H_{cont}^*(G_S, T)$.

en deux pour obtenir un triangle distingué

$$R\Gamma_f(F_p, V) \longrightarrow R\Gamma(F_p, V) \longrightarrow R\Gamma_{/f}(F_p, V)$$

caractérisé par le fait que $H^1(R\Gamma_f(F_p, V))$ est le sous- E_λ -espace vectoriel $H_f^1(F_p, V)$ de $H^1(F_p, V)$ défini au n° 3.2, que $H^0(R\Gamma_f(F_p, V)) = H^0(F_p, V)$ et que

$$H^i(R\Gamma_f(F_p, V)) = 0 \text{ si } i \neq 0, 1.$$

Soient \mathcal{O}_p l'anneau des entiers de F_p , $j_p : \text{Spec } F_p \longrightarrow \text{Spec } \mathcal{O}_p$ et $i_p : \text{Spec } k_p \longrightarrow \text{Spec } \mathcal{O}_p$ les morphismes naturels. Si $\ell \neq p$, $R\Gamma_f(F_p, V) = R\Gamma(k_p, i_p^* j_{p*} V) = R\Gamma(k_p, \underline{D}_p(V))$. On voit aussi que ce complexe s'identifie au complexe concentré en degrés 0 et 1

$$\underline{D}_p(V) \longrightarrow \underline{D}_p(V)$$

(l'application étant $d \longmapsto (f_p - 1)(d)$). En particulier, $\det_\lambda R\Gamma_f(F_p, V) = E_\lambda$.

Si $\ell = p$, et si π désigne la projection de $B_{cris,p} \subset B_{dR,p}$ sur $B_{dR,p}/\text{Fil}^0 B_{dR,p}$, on peut montrer (cf. par exemple, [BK90], prop. 1.7) que la suite

$$(*) \quad 0 \longrightarrow \mathbb{Q}_p \longrightarrow B_{cris,p} \longrightarrow B_{cris,p} \oplus (B_{dR,p}/\text{Fil}^0 B_{dR,p}) \longrightarrow 0$$

(où $\mathbb{Q}_p \longrightarrow B_{cris,p}$ est l'inclusion, tandis que l'application suivante est $b \longmapsto ((\varphi - 1)b, \pi b)$) est exacte. Si l'on tensorise avec V et que l'on prend les invariants par G_p , on trouve¹² une suite exacte

$$0 \longrightarrow H^0(F_p, V) \longrightarrow \underline{D}_p(V) \longrightarrow \underline{D}_p(V) \oplus t_{V,p} \longrightarrow H_f^1(F_p, V) \longrightarrow 0$$

qui nous permet d'identifier $R\Gamma_f(F_p, V)$ au complexe concentré en degrés 0 et 1

$$\underline{D}_p(V) \longrightarrow \underline{D}_p(V) \oplus t_{V,p} .$$

En particulier, $\det_\lambda R\Gamma_f(F_p, V)$ s'identifie à $\det_\lambda^1 t_{V,p}$.

4.5. — Considérons de nouveau un E_λ -faisceau V sur U_S . Définissons la **f -cohomologie de V** comme étant $C[-1]$ où C est le cône $R\Gamma_f(U_S, V)$ du morphisme composé

$$R\Gamma_{\acute{e}t}(U_S, V) \longrightarrow \bigoplus_{p \in S_f} R\Gamma(F_p, V) \longrightarrow \bigoplus_{p \in S_f} R\Gamma_{/f}(F_p, V) .$$

¹² Comme $B_{cris,p}$ et $B_{dR,p}$ ne sont pas des \mathbb{Q}_p -espaces vectoriels de dimension finie, il faut soit prouver que la suite (*) est scindée en tant que suite de \mathbb{Q}_p -espaces vectoriels topologiques, soit remarquer que l'on peut remplacer $B_{cris,p}$ et $B_{dR,p}$ par l'union de leurs sous- \mathbb{Q}_p -espaces vectoriels de dimension finie stables par G_p .

Les deux triangles distingués

$$\begin{array}{l} \text{et} \\ R\Gamma_c(U_S, V) \longrightarrow R\Gamma_{\acute{e}t}(U_S, V) \longrightarrow \bigoplus_{\mathfrak{p} \in S} R\Gamma(F_{\mathfrak{p}}, V) \\ R\Gamma_f(U_S, V) \longrightarrow R\Gamma_{\acute{e}t}(U_S, V) \longrightarrow \bigoplus_{\mathfrak{p} \in S_f} R\Gamma_{/f}(F_{\mathfrak{p}}, V) \end{array}$$

donnent naissance à un troisième

$$R\Gamma_c(U_S, V) \rightarrow R\Gamma_f(U_S, V) \rightarrow (\bigoplus_{\mathfrak{p} \in S_f} R\Gamma_f(F_{\mathfrak{p}}, V)) \oplus (\bigoplus_{\mathfrak{p} \in S_{\infty}(F)} R\Gamma(F_{\mathfrak{p}}, V)) .$$

De celui-ci, on déduit un isomorphisme

$$(a) \quad \det_{\lambda} R\Gamma_f(U_S, V) = \Delta_{\lambda, EP}(V) \otimes \det_{\lambda}^1 t_V \otimes \det_{\lambda} V^+ .$$

Par ailleurs, le calcul des $H_f^i(U_S, V) := H^i(R\Gamma_f(U_S, V))$ est facile : on trouve que ceux-ci sont nuls, si $i \neq 0, 1, 2, 3$. On a $H_f^0(U_S, V) = H^0(F, V)$ et $H_f^1(U_S, V) = H_f^1(F, V)$. On voit aussi que $H_f^3(U_S, V)$ s'identifie au dual de $H^0(F, V^*(1))$; pour cela, on commence par observer que ce sont tous deux, de façon naturelle, des quotients de $\bigoplus_{\mathfrak{p} \in S_f} H^2(F_{\mathfrak{p}}, V)$ (le premier par définition; le second parce que la dualité locale (cf. par exemple, [Mi86], chap. I, § 2) identifiant chaque $H^2(F_{\mathfrak{p}}, V)$ au dual de $H^0(F_{\mathfrak{p}}, V^*(1))$ on peut utiliser la transposée de l'injection naturelle $H^0(F, V^*(1)) \rightarrow \bigoplus_{\mathfrak{p} \in S_f} H^0(F_{\mathfrak{p}}, V^*(1))$) et on vérifie que ces deux quotients sont les mêmes.

Enfin, si $H_0^2(U_S, V)$ désigne le noyau de l'application naturelle

$$H^2(U_S, V) \longrightarrow \bigoplus_{\mathfrak{p} \in S_f} H^2(F_{\mathfrak{p}}, V) ,$$

$H_f^2(U_S, V)$ est la somme directe de $H_0^2(U_S, V)$ et du conoyau $\tilde{H}_f^2(U_S, V)$ de l'application naturelle

$$H^1(U_S, V) \oplus (\bigoplus_{\mathfrak{p} \in S_f} H_f^1(F_{\mathfrak{p}}, V)) \longrightarrow \bigoplus_{\mathfrak{p} \in S_f} H^1(F_{\mathfrak{p}}, V) .$$

La dualité locale (*loc. cit.*) identifie chaque $H^1(F_{\mathfrak{p}}, V)$ au dual de $H^1(F_{\mathfrak{p}}, V^*(1))$, d'où une application de $\bigoplus_{\mathfrak{p} \in S_f} H^1(F_{\mathfrak{p}}, V)$ dans $H_f^1(U_S, V^*(1))^*$ et les théorèmes de dualité globale de Poitou–Tate (cf. par exemple, [Mi86], chap. II, § 1) identifient son conoyau au dual de $H_0^2(U_S, V)$. Lorsque V est pseudo-géométrique, on vérifie que la suite

$$\begin{aligned} H^1(U_S, V) \oplus (\bigoplus_{\mathfrak{p} \in S_f} H_f^1(F_{\mathfrak{p}}, V)) &\rightarrow \bigoplus_{\mathfrak{p} \in S_f} H^1(F_{\mathfrak{p}}, V) \rightarrow \\ &\rightarrow H_f^1(U_S, V^*(1))^* \rightarrow H_0^2(U_S, V) \rightarrow 0 \end{aligned}$$

est exacte. On a donc une suite exacte courte

$$0 \longrightarrow \tilde{H}_f^2(U_S, V) \longrightarrow H_f^1(U_S, V^*(1))^* \longrightarrow H_0^2(U_S, V) \longrightarrow 0 .$$

Comme $H_f^2(U_S, V) = \tilde{H}_f^2(U_S, V) \oplus H_0^2(U_S, V)$, $\det_\lambda^1 H_f^1(U_S, V^*(1))$ s'identifie alors à $\det_\lambda H_f^2(U_S, V)$. Finalement, on obtient un isomorphisme canonique

$$\det_\lambda R\Gamma_f(U_S, V) = L_f(V) \otimes L_f(V^*(1)) ;$$

joint à l'isomorphisme (a) cela fournit un isomorphisme canonique entre $\Delta_f(V)$ et $\Delta_{\lambda, EP}(V)$. On s'en sert pour identifier ces deux droites et $\Delta_f(V)$ hérite de la norme $|\cdot|_{\lambda, EP}$ définie sur $\Delta_{\lambda, EP}(V)$.

§ 5. — Structures de Hodge mixtes

5.1. — Lorsque $\ell \neq \infty$, ce sont les structures de Hodge mixtes qui jouent le rôle des représentations ℓ -adiques.

Si V est un \mathbf{R} -espace vectoriel, posons $V_{\mathbf{C}} = \mathbf{C} \otimes_{\mathbf{R}} V$ et notons $\tau : V_{\mathbf{C}} \rightarrow V_{\mathbf{C}}$ l'application qui envoie $z \otimes v$ sur $\bar{z} \otimes v$. Rappelons (cf. [De71]) qu'une **R-structure de Hodge mixte sur \mathbf{C}** est un \mathbf{R} -espace vectoriel V de dimension finie, muni d'une filtration croissante (**la filtration par le poids**) $(W_m V)_{m \in \mathbf{Z}}$ par des sous- \mathbf{R} -espaces vectoriels, avec $W_i V = 0$ si $i \ll 0$ et $= V$ si $i \gg 0$, et d'une filtration décroissante (**la filtration de Hodge**), $(Fil^i V_{\mathbf{C}})_{i \in \mathbf{Z}}$, de $V_{\mathbf{C}}$ par des sous- \mathbf{C} -espaces vectoriels, telles que, si \overline{Fil}^i désigne la filtration conjuguée de Fil^i (i.e. la filtration de $V_{\mathbf{C}}$ définie par $\overline{Fil}^i V_{\mathbf{C}} = \tau(Fil^i V_{\mathbf{C}})$) alors W_\cdot , Fil^\cdot et \overline{Fil}^\cdot soient opposées (i.e. telles que, pour tout $m \in \mathbf{Z}$, si l'on note encore Fil et \overline{Fil} les filtrations induites sur $gr_m^W V_{\mathbf{C}}$, alors l'application naturelle

$$Fil^r gr_m^W V_{\mathbf{C}} \oplus \overline{Fil}^{m+1-r} gr_m^W V_{\mathbf{C}} \longrightarrow gr_m^W V_{\mathbf{C}}$$

soit un isomorphisme pour tout $r \in \mathbf{Z}$).

Rappelons que les entiers m tels que $gr_m^W V_{\mathbf{C}} \neq 0$ sont les **poinds de V** et que V est **pure** si elle a un et un seul poind.

Les **R-structures de Hodge mixtes sur \mathbf{C}** forment, de manière évidente, une catégorie abélienne \mathbf{R} -linéaire $\underline{SH}_{\mathbf{C}}(\mathbf{R})$. On note $\underline{SH}_{\mathbf{C}}(\mathbf{C})$ la catégorie des **C-structures de Hodge mixtes sur \mathbf{C}** , i.e. la catégorie \mathbf{C} -linéaire déduite de $\underline{SH}_{\mathbf{C}}(\mathbf{R})$ par extension des scalaires. Un objet V de cette catégorie est donc une **R-structure de Hodge mixte sur \mathbf{C}** munie d'un **R-plongement** de \mathbf{C} dans l'anneau de ses endomorphismes (attention que, bien que l'espace vectoriel réel sous-jacent à V soit muni d'une structure d'espace vectoriel sur \mathbf{C} , c'est seulement sur $V_{\mathbf{C}} = \mathbf{C} \otimes_{\mathbf{R}} V$ qu'est définie la filtration de Hodge).

5.2. — Une \mathbf{R} -structure de Hodge mixte sur \mathbf{R} peut être définie comme étant une \mathbf{R} -structure de Hodge mixte sur \mathbf{C} munie d'une action de $\text{Gal}(\mathbf{C}/\mathbf{R})$. Cela revient donc à se donner une décomposition de V en somme directe

$$V = V^+ \oplus V^-$$

compatible avec la filtration par le poids et telle que la filtration de Hodge provienne, par extension des scalaires, d'une filtration définie sur $V_{dR} = (V_{\mathbf{C}})^{\text{Gal}(\mathbf{C}/\mathbf{R})} = V^+ \oplus i.V^-$. On obtient encore une catégorie abélienne $\underline{SH}_{\mathbf{R}}(\mathbf{R})$ et on note aussi $\underline{SH}_{\mathbf{R}}(\mathbf{C})$ la catégorie des \mathbf{C} -structures de Hodge mixtes sur \mathbf{R} , i.e. la catégorie déduite de $\underline{SH}_{\mathbf{R}}(\mathbf{R})$ par l'extension des scalaires $\mathbf{R} \rightarrow \mathbf{C}$.

5.3. — Dans la suite de ce paragraphe, $K, L \in \{\mathbf{R}, \mathbf{C}\}$, $\bar{K} = \mathbf{C}$, $G_K = \text{Gal}(\mathbf{C}/K)$ et V est une L -structure de Hodge mixte sur K^{13} .

Si $K = \mathbf{C}$, on pose $V^+ = V$ et $V_{dR} = V_{\mathbf{C}}$. Dans tous les cas, on a donc $V^+ = V^{G_K}$ et $V_{dR} = (V_{\mathbf{C}})^{G_K}$.

On note $\mathbf{1}_{K,L}$ l'objet-unité de la catégorie $\underline{SH}_K(L)$, i.e. l'unique objet pur de poids 0 vérifiant $\mathbf{1}_{K,L} = (\mathbf{1}_{K,L})^+ = L$. On laisse au lecteur le soin de définir le produit tensoriel de deux objets et le dual d'un objet¹⁴. L'objet de Tate $\mathbf{1}_{K,L}(1)$ est l'unique objet pur de poids -2 tel que $\mathbf{1}_{K,L}(1) = L \cdot (2\pi i)$, avec $\mathbf{1}_{K,L}(1)^- = \mathbf{1}_{K,L}(1)$ si $K = \mathbf{R}$. On note $V^*(1)$ le produit tensoriel du dual de V par $\mathbf{1}_{K,L}(1)$.

On pose $t_V = V_{dR}/\text{Fil}^0 V_{dR}$. L'inclusion de V dans $V_{\mathbf{C}}$ induit en prenant les parties fixes par G_K et en passant au quotient une application L -linéaire

$$\alpha_V : V^+ \rightarrow t_V .$$

5.4. — Pour tout $i \in \mathbf{N}$, on pose $H^i(K, V) = \text{Ext}_{\underline{SH}_K(L)}^i(\mathbf{1}_{K,L}, V)$. Ce sont des L -espaces vectoriels de dimension finie, nuls pour $i \neq 0, 1$. Dans la suite, le noyau et le conoyau de α_V vont jouer un grand rôle alors qu'il semblerait plus naturel de considérer les $H^i(K, V)^{15}$. Ceux-ci sont étroitement liés à ceux-là :

¹³ On appliquera cela à $K = F_{\mathfrak{p}}$ et $L = E_{\lambda}$, avec $\mathfrak{p} \in S_{\infty}(F)$ et $\lambda \in S_{\infty}(E)$ (rappelons que l'on a identifié $\bar{F}_{\mathfrak{p}}$ et \mathbf{C}).

¹⁴ La catégorie $\underline{SH}_K(L)$ est une catégorie tannakienne neutre ([DM82], def. 2.19) sur L .

¹⁵ Si l'on "oublie" la filtration par le poids, on peut voir V comme un objet d'une catégorie additive L -linéaire non abélienne; alors $\text{Ker } \alpha_V$ s'identifie au L -espace vectoriel des morphismes de $\mathbf{1}_{K,L}$ dans V dans cette catégorie, tandis que $\text{Coker } \alpha_V$ classe les "extensions" de $\mathbf{1}_{K,L}$ par V . En particulier, toute suite exacte courte

$$0 \rightarrow V' \rightarrow V \rightarrow V'' \rightarrow 0$$

on a en effet $H^i(K, V) = H^i(K, W_0V)$ et, si $W_0V = V$, $H^0(K, V)$ s'identifie à $\text{Ker } \alpha_V$ et $H^1(K, V)$ à $\text{Coker } \alpha_V$; d'où une suite exacte de L -espaces vectoriels

$$0 \rightarrow H^0(K, V) \rightarrow \text{Ker } \alpha_V \rightarrow \text{Ker } \alpha_{V/W_0V} \rightarrow H^1(K, V) \rightarrow \text{Coker } \alpha_V \rightarrow 0$$

et on voit que $\text{Ker } \alpha_{V/W_0V}$ s'identifie au dual de $H^1(K, (V/W_0V)^*(1))$.

On vérifie facilement que l'on dispose d'une dualité parfaite

$$\text{Coker } \alpha_V \times \text{Ker } \alpha_{V^*(1)} \longrightarrow L .$$

§6. — Les motifs sur F et leurs réalisations

On note $\mathbf{A}_f(\mathbf{Q})$ (resp. $\mathbf{A}_f(E)$) l'anneau des adèles finies de \mathbf{Q} (resp. E). On pose $\mathbf{Q}_\infty = \mathbf{R}$ et, pour tout $\mathfrak{p} \in S_\infty(F)$, $B_{dR, \infty} = \overline{F}_\mathfrak{p} = \mathbf{C}$.

6.1. — Faute de connaître la catégorie $\mathcal{MM}_F(E)$, il est d'usage d'introduire une catégorie "fourre-tout" parfaitement bien définie et jouissant de propriétés agréables, sensée contenir $\mathcal{MM}_F(E)$ et dans laquelle on essaie de faire le maximum du travail (cf. par exemple : [De90], § 1; [Ja90], § 2; [BK90], def. 5.5; [FP91], note II, § 1; ...). Celle d'aujourd'hui, $\mathcal{C}_F(E)$ va être définie au n° suivant. Commençons par noter $\mathcal{C}_{F,0}$ la catégorie suivante :

– Un objet M de $\mathcal{C}_{F,0}$ consiste en les données suivantes :

i) un F -espace vectoriel de dimension finie M_{dR} , muni d'une filtration décroissante $(\text{Fil}^i M_{dR})_{i \in \mathbf{Z}}$ par des sous- F -espaces vectoriels, avec $\text{Fil}^i M_{dR} = M_{dR}$ si $i \ll 0$ et $= 0$ si $i \gg 0$;

ii) pour chaque $\mathfrak{p} \in S_\infty(F)$, un \mathbf{Q} -espace vectoriel $M_{B, \mathfrak{p}}$, muni d'une action de $G_\mathfrak{p}$ (on pose $M_{B, \mathfrak{p}}^+ = (M_{B, \mathfrak{p}})^{G_\mathfrak{p}}$);

iii) un $\mathbf{A}_f(\mathbf{Q})$ -module libre de rang fini M_f , muni d'une action linéaire et continue de G_F ; si, pour ℓ premier, on note M_ℓ sa ℓ -composante, on demande que M_ℓ soit non ramifiée en dehors d'un nombre fini de places;

iv) pour chaque $\mathfrak{p} \in S_\infty(F)$, un isomorphisme de $\mathbf{A}_f(\mathbf{Q})[\text{Gal}(\overline{F}_\mathfrak{p}/F_\mathfrak{p})]$ -modules

$$i_{M, \mathfrak{p}}^B : \mathbf{A}_f(\mathbf{Q}) \otimes_E M_{B, \mathfrak{p}} \longrightarrow M_f ;$$

de $\underline{SH}_K(L)$ donne naissance à une suite exacte longue

$$0 \rightarrow \text{Ker } \alpha_{V'} \rightarrow \text{Ker } \alpha_V \rightarrow \text{Ker } \alpha_{V''} \rightarrow \text{Coker } \alpha_{V'} \rightarrow \text{Coker } \alpha_V \rightarrow \text{Coker } \alpha_{V''} \rightarrow 0.$$

v) pour chaque place $\mathfrak{p} \in S_p(F)$ (p fini ou non), si l'on note $M_{\mathfrak{p}}$ le $\mathbb{Q}_{\mathfrak{p}}[G_{\mathfrak{p}}]$ -module sous-jacent à M_p si p est fini et à $\mathbb{R} \otimes_{\mathbb{Q}} M_{B,\mathfrak{p}}$ sinon, un isomorphisme de $B_{dR,\mathfrak{p}}$ -espaces vectoriels, compatible avec l'action de $G_{\mathfrak{p}}$,

$$i_{M,\mathfrak{p}} : B_{dR,\mathfrak{p}} \otimes_{\mathbb{Q}_{\mathfrak{p}}} M_{\mathfrak{p}} \longrightarrow B_{dR,\mathfrak{p}} \otimes_F M_{dR}^{16}$$

et vérifiant, pour tout $r \in \mathbb{Z}$, lorsque $p \neq \infty$,

$$i_{M,\mathfrak{p}}(Fil^r B_{dR,\mathfrak{p}} \otimes_{\mathbb{Q}_{\mathfrak{p}}} M_{\mathfrak{p}}) = \sum_{s+t=r} (Fil^s B_{dR,\mathfrak{p}} \otimes_F Fil^t M_{dR}).$$

– Si $M = (M_{dR}, (M_{B,\mathfrak{p}})_{\mathfrak{p} \in S_{\infty}(F)}, M_f)$ et $N = (N_{dR}, (N_{B,\mathfrak{p}})_{\mathfrak{p} \in S_{\infty}(F)}, N_f)$ sont deux objets de $\mathcal{C}_{F,0}$ un morphisme $\eta : M \rightarrow N$ est la donnée d'applications $(\eta_{dR}, (\eta_{B,\mathfrak{p}})_{\mathfrak{p} \in S_{\infty}(F)}, \eta_f)$ vérifiant toutes les compatibilités que l'on pense.

6.2. — La catégorie $\mathcal{C}_{F,0}$ est abélienne et même \mathbb{Q} -linéaire. On note $\mathcal{C}_F(E)_0$ la catégorie déduite de $\mathcal{C}_{F,0}$ par extension des scalaires de \mathbb{Q} à E ; un objet de $\mathcal{C}_F(E)_0$ est donc un objet M de $\mathcal{C}_{F,0}$ muni d'un plongement de E dans la \mathbb{Q} -algèbre des $\mathcal{C}_{F,0}$ -endomorphismes de M , tandis qu'un morphisme est un morphisme des objets de $\mathcal{C}_{F,0}$ sous-jacents qui commute à l'action de E .

Si M est un objet de $\mathcal{C}_F(E)_0$, M_{dR} devient un $E \otimes_{\mathbb{Q}} F$ -module libre de rang fini que l'on note $b(M)$, chaque $M_{B,\mathfrak{p}}$ est un E -espace vectoriel de dimension $b(M)$ et M_f devient un $\mathbf{A}_f(E)$ -module libre de rang $b(M)$. Pour chaque place finie λ de E , la E_{λ} -composante M_{λ} de M_f est une représentation λ -adique pseudo-géométrique de G_F de dimension $b(M)$. On appelle **espace tangent de M** le $E \otimes F$ -module $t_M = M_{dR}/Fil^0 M_{dR}$. Si \mathfrak{p} est une place de F et λ une place de E au-dessus du même $p \in S(\mathbb{Q})$, alors

– si p est premier, $\underline{D}_{dR,\mathfrak{p}}(M_{\lambda})$ s'identifie à $(E_{\lambda} \otimes_{\mathbb{Q}_{\mathfrak{p}}} F_{\mathfrak{p}}) \otimes_{E \otimes F} M_{dR}$ et $t_{M_{\lambda,\mathfrak{p}}}$ à $(E_{\lambda} \otimes_{\mathbb{Q}_{\mathfrak{p}}} F_{\mathfrak{p}}) \otimes_{E \otimes F} t_M$;

– si $p = \infty$, posons $M_{\mathfrak{p},\lambda} = E_{\lambda} \otimes_{\mathbb{R}} M_{\mathfrak{p}}$; l'application $i_{M,\mathfrak{p}}$ identifie le \mathbb{C} -espace vectoriel $\mathbb{C} \otimes_{\mathbb{R}} M_{\mathfrak{p},\lambda}$ à $\mathbb{C} \otimes_{E \otimes_{\mathbb{Q}} F} M_{dR}$; en particulier, on peut parler de la filtration de Hodge et de la filtration conjuguée sur $\mathbb{C} \otimes_{\mathbb{R}} M_{\mathfrak{p},\lambda}$.

¹⁶ Supposons $F = \mathbb{Q}$ et soit $\bar{\mathbb{Z}}$ l'anneau des entiers de $\bar{\mathbb{Q}}$. En adaptant la construction habituelle de B_{dR} (essentiellement, en faisant jouer à $\Pi(\bar{\mathbb{Z}}/p\bar{\mathbb{Z}})$ le rôle habituellement dévolu à $\bar{\mathbb{Z}}_p/p\bar{\mathbb{Z}}_p$), on peut définir une $\mathbf{A}_f(\mathbb{Q})$ -algèbre $\mathcal{B}_{dR,f}$ munie d'une filtration indexée par \mathbb{Z} et d'une action de G_F et réénoncer (iv) comme étant la donnée d'un isomorphisme

$$\mathcal{B}_{dR} \otimes_{\mathbb{Q}} M_B = \mathcal{B}_{dR} \otimes_{\mathbb{Q}} M_{dR}$$

(où $\mathcal{B}_{dR} = \mathbb{C} \times \mathcal{B}_{dR,f}$) vérifiant des propriétés convenables.

Si S est un ensemble fini de places de F contenant $S_\infty(F)$, on dit que M **admet une fonction** L_S si chaque M_λ en admet une et si celle-ci est indépendante de λ place finie de E (on la note alors $L_S(M, s)$). Si $S = S_\infty(F)$, on dit que M **admet une fonction** L (et on note celle-ci $L(M, s)$).

6.3. — La catégorie $\mathcal{C}_F(E)$ est ainsi définie :

– **un objet** est un objet M de $\mathcal{C}_F(E)_0$, admettant une fonction L et muni d'une filtration croissante (la **filtration par le poids**) par des sous-objets $(W_i M)_{i \in \mathbf{Z}}$ (vérifiant $W_i M = 0$ si $i \ll 0$ et $= M$ si $i \gg 0$) telle que, si λ (resp. \mathfrak{p}) est une place archimédienne de E (resp. F), $M_{\mathfrak{p}, \lambda}$ devienne une E_λ -structure de Hodge mixte sur $F_{\mathfrak{p}}$ (cela revient à demander que, sur $\mathbf{C} \otimes_{\mathbf{R}} M_{\mathfrak{p}, \lambda}$, alors la filtration par le poids, celle de Hodge et la filtration conjuguée soient opposées).

– **un morphisme** de $\mathcal{C}_F(E)$ est un morphisme des objets de $\mathcal{C}_F(E)_0$ sous-jacents qui est compatible avec la filtration par le poids.

Il n'est pas difficile de vérifier que $\mathcal{C}_F(E)$ est encore une catégorie abélienne E -linéaire. Si M et N sont deux objets de $\mathcal{C}_F(E)$, on laisse au lecteur le soin de définir les objets $M \otimes N$ et $\text{Hom}(M, N)$. On dispose aussi d'un objet unité $\mathbf{1}_{F, E}$ (c'est un objet pur de poids 0; on a $\mathbf{1}_{F, E, dR} = E$, $\mathbf{1}_{F, E, B, \mathfrak{p}} = E \otimes F$, pour tout $\mathfrak{p} \in S_\infty(F)$ et $\mathbf{1}_{F, E, f} = \mathbf{A}_f(E)$, avec action triviale de G_F); le dual M^* de M est alors $\text{Hom}(M, \mathbf{1}_{F, E})$ (et on a $\text{Hom}(M, N) = M^* \otimes N$). On obtient ainsi une catégorie tannakienne neutre sur E ([DM82], § 2). Bien sûr la catégorie $\mathcal{C}_F(E)$ n'est autre que la catégorie tannakienne $\mathcal{C}_F(\mathbf{Q})_{(E)}$ déduite de $\mathcal{C}_F(\mathbf{Q})$ par l'extension des scalaires $\mathbf{Q} \subset E$ ([DM82], § 3).

6.4. — Si F' est une extension finie de F contenue dans \overline{F} , on dispose de foncteurs évidents d'extension des scalaires $\text{Ext}_{F/F'} : \mathcal{C}_F(E) \rightarrow \mathcal{C}_{F'}(E)$ et de son adjoint, la restriction des scalaires à la Weil $\text{Res}_{F'/F} : \mathcal{C}_{F'}(E) \rightarrow \mathcal{C}_F(E)$ (attention que, si λ est finie, $\text{Ext}_{F/F'}(M)_\lambda$ n'est autre que M_λ muni de l'action de $G_{F'} \subset G_F$ tandis que $\text{Res}_{F'/F}(N)_\lambda$ est la représentation de G_F induite par la représentation N_λ de $G_{F'}$). On a $L(\text{Res}_{F'/F}(N), s) = L(N, s)$.

De même, si E' est une extension finie de E on a un foncteur d'extension des scalaires de E à E' allant de $\mathcal{C}_F(E)$ dans $\mathcal{C}_F(E')$ qui admet un adjoint, la restriction des scalaires de E' à E .

On note $\mathbf{1}(1)$ l'**objet de Tate** de la catégorie $\mathcal{C}_{\mathbf{Q}}(\mathbf{Q})$ (il est pur de poids -2 ; on a $\mathbf{1}(1)_{dR} = \mathbf{Q}$, $\mathbf{1}(1)_B = 2\pi i \cdot \mathbf{Q}$; pour tout $n \in \mathbf{N}$, le groupe fini $\mathbf{Z}.2\pi i/n\mathbf{Z}.2\pi i$ s'identifie à $\mu_n(\overline{\mathbf{Q}})$ via l'application qui envoie $2\pi i$ sur $e^{2\pi i/n}$ et ceci donne l'action de $G_{\mathbf{Q}}$ sur

$$\mathbf{1}(1)_f = \mathbf{Q} \otimes_{\widehat{\mathbf{Z}}} \varprojlim_n (\mathbf{Z}.2\pi i/n\mathbf{Z}.2\pi i),$$

les isomorphismes de comparaison sont ceux que l'on pense. Dans la suite, on pose $t = 2\pi i$, que l'on peut donc voir comme un élément de $B_{dR,p}$ pour toute place finie ou non p de \mathbb{Q} .

L'objet de Tate $\mathbf{1}_{F,E}(1)$ est l'objet de $\mathcal{C}_F(E)$ déduit de $\mathbf{1}(1)$ par double extension des scalaires.

Pour $r \in \mathbb{N}$, on note $\mathbf{1}_{F,E}(r)$ la puissance tensorielle r -ième de $\mathbf{1}_{F,E}(1)$ et $\mathbf{1}_{F,E}(-r)$ son dual. Pour tout objet M de $\mathcal{C}_F(E)$ et tout $i \in \mathbb{Z}$, on pose $M(i) = M \otimes \mathbf{1}_{F,E}(i)$. Le **dual tordu à la Tate** $M^*(1)$ va jouer un rôle important dans la suite.

6.5. — Soit \mathcal{M} une sous-catégorie pleine de $\mathcal{C}_F(E)$, stable par sous-objet, quotient, somme directe, $M \mapsto M^*(1)$, contenant $\mathbf{1}_{F,E}$. Pour tout objet M de \mathcal{M} , on pose $H^0(F, M) = \text{Hom}_{\mathcal{C}_F(E)}(\mathbf{1}_{F,E}, M)$ ($= \text{Hom}_{\mathcal{M}}(\mathbf{1}_{F,E}, M)$), on note $H_{\mathcal{M}}^1(F, M)$ le E -espace vectoriel des classes d'isomorphismes d'extensions de $\mathbf{1}_{F,E}$ par M dans \mathcal{M} . Pour chaque place finie λ de E et pour tout $x \in H_{\mathcal{M}}^1(F, M)$, on note x_λ son image dans $H^1(F, M_\lambda)$. On pose

$$H_{\mathcal{M},f}^1(F, M) = \{x \in H^1(F, M) \mid x_\lambda \in H_f^1(F, M_\lambda), \text{ pour tout } \lambda\}$$

et aussi $H_{\mathcal{M},f}^0(F, M) = H^0(F, M)$.

Disons que la catégorie \mathcal{M} est f -**admissible** si, pour tout objet M de \mathcal{M} , toute place λ de E et $i \in \{0, 1\}$, elle satisfait la propriété $C_\lambda^i(M) = C_\lambda^i(M, \mathcal{M})$ suivante :

a) si λ est finie, l'application naturelle $E_\lambda \otimes_E H_{\mathcal{M},f}^i(F, M) \rightarrow H_f^i(F, M_\lambda)$ est un isomorphisme;

b) si $\lambda \in S_\infty(E)$ et si $H_{\mathcal{M},f}^0(F, M^*(1)) = H_{\mathcal{M},f}^1(F, M^*(1)) = 0$, l'application naturelle $E_\lambda \otimes_E H_{\mathcal{M},f}^i(F, M) \rightarrow \bigoplus_{\mathfrak{p} \in S_\infty(F)} H^i(F_{\mathfrak{p}}, M_\lambda)$ est un isomorphisme.

6.6. — Appelons **sous-catégorie tannakienne de $\mathcal{C}_F(E)$** toute sous-catégorie pleine stable par sous-objet, quotient, somme directe, produit tensoriel et dual. Il est raisonnable de penser qu'à tout motif sur F à coefficients dans E , on puisse associer un objet de $\mathcal{C}_F(E)$, sa **réalisation**. On peut alors introduire la catégorie $\underline{SM}_F(E)$ des **structures motiviques** sur F à coefficients dans E comme étant la plus petite sous-catégorie tannakienne de $\mathcal{C}_F(E)$ contenant tout objet qui est isomorphe à la réalisation d'un motif.

Bien sûr, le jour où l'on disposera d'une définition convenable de la catégorie $\mathcal{MM}_F(E)$, les "conjectures standards pour les motifs mixtes" devraient affirmer que le foncteur "réalisation" induit une équivalence entre $\mathcal{MM}_F(E)$ et $\underline{SM}_F(E)$.

"CONJECTURE". — La catégorie $\underline{SM}_F(E)$ est f -admissible

6.7. — **Remarque** : On sait (cf. [De71], [De74], [Ja90]) associer à toute variété quasi-projective lisse U sur F sa réalisation $H(U)$ qui est un objet de $\mathcal{C}_F(\mathbb{Q})$ (par exemple, sa réalisation de Rham est l'hypercohomologie du complexe de de Rham de U) à ceci près

– qu'il faut savoir définir les isomorphismes de comparaison $i_{M,\mathfrak{p}}$ pour les places \mathfrak{p} finies : cela a été fait par Faltings ([Fa89], cf. aussi [I90]) lorsque U est projective ; le cas général doit pouvoir se déduire des travaux de Faltings [Fa90] ;

– que l'on peut exhiber pour chaque U un ensemble fini de places S de F contenant $S_\infty(F)$ tel que M admet une fonction L_S , mais que l'on ne sait pas toujours montrer que l'on peut prendre $S = S_\infty(F)$.

Soit $\underline{SM}_{F,J}(\mathbb{Q})$ la plus petite sous-catégorie tannakienne de $\mathcal{C}_F(\mathbb{Q})$ contenant tout objet qui est isomorphe à un $H(U)$ avec U variété quasi-projective lisse sur F . On veut que $\underline{SM}_F(\mathbb{Q})$ contienne $\underline{SM}_{F,J}(\mathbb{Q})$. Suivant Jannsen ([Ja90], § 4), on peut se demander si $\underline{SM}_F(\mathbb{Q}) = \underline{SM}_{F,J}(\mathbb{Q})$, i.e. si $\underline{SM}_{F,J}(\mathbb{Q})$ est "assez grosse", ce qui peut se traduire, une fois un nombre premier ℓ fixé, par le fait que, pour tout objet M de $\underline{SM}_{F,J}(\mathbb{Q})$, l'application naturelle

$$\mathbb{Q}_\ell \otimes_{\mathbb{Q}} H_{\underline{SM}_{F,J}(\mathbb{Q}),f}^i(F, M) \longrightarrow H_f^i(F, M_\ell)$$

est surjective. Une telle question semble malheureusement totalement inabordable à l'heure actuelle.

6.8. — Dans ce qui suit, on suppose donnée une sous-catégorie pleine \mathcal{M} de $\underline{SM}_F(E)$ qui est f -admissible (si la "conjecture" ci-dessus est vraie, cela revient à dire qu'elle contient $\mathbf{1}_{F,E}$, est stable par sous-objet, quotient, somme directe et que, si elle contient M , alors elle contient $M^*(1)$ et toute extension de $\mathbf{1}_{F,E}$ par M dont la classe appartient à $H_{\underline{SM}_F(E),f}^1(F, M)$). Pour $i = 0, 1$, on pose $H_f^i(F, M) = H_{\mathcal{M},f}^i(F, M)$.

On appelle **structures motiviques** les objets de \mathcal{M} .

6.9. — Pour tout E -espace vectoriel X de dimension finie et tout $\lambda \in S_\infty(E)$, on pose $X_\lambda = E_\lambda \otimes_E X$ et on note X_λ^* son E_λ -dual.

Soient M une structure motivique et $\lambda \in S_\infty(E)$. Si $\mathfrak{p} \in S_\infty(F)$, $M_{\mathfrak{p},\lambda}$ est une E_λ -structure de Hodge mixte sur $F_\mathfrak{p}$. On peut donc considérer (cf. n° 5.3) l'application

$$\alpha_{M_\lambda} = \bigoplus_{\mathfrak{p} \in S_\infty(F)} \alpha_{M_{\mathfrak{p},\lambda}} : (\bigoplus_{\mathfrak{p} \in S_\infty(F)} M_{B,\mathfrak{p}}^+)_\lambda \longrightarrow t_{M,\lambda}$$

(où $M_{B,\mathfrak{p}}^+ = (M_{B,\mathfrak{p}})^{G_\mathfrak{p}}$) et (*loc. cit.*) les E_λ -espaces vectoriels $\text{Ker } \alpha_{M_\lambda}$ et $\text{Coker } \alpha_{M^*(1)_\lambda}$ sont en dualité de même que $\text{Coker } \alpha_{M_\lambda}$ et $\text{Ker } \alpha_{M^*(1)_\lambda}$. Notons

u_{M_λ} le composé des applications naturelles

$$H^0(F, M)_\lambda \longrightarrow \bigoplus_{\mathfrak{p}} H^0(F_{\mathfrak{p}}, M_\lambda) \longrightarrow \bigoplus_{\mathfrak{p}} \text{Ker } \alpha_{M_{\mathfrak{p}, \lambda}} = \text{Ker } \alpha_{M_\lambda} ,$$

v_{M_λ} le composé $H_f^1(F, M)_\lambda \longrightarrow \bigoplus_{\mathfrak{p}} H^1(F_{\mathfrak{p}}, M_\lambda) \longrightarrow \bigoplus_{\mathfrak{p}} \text{Coker } \alpha_{M_{\mathfrak{p}, \lambda}} = \text{Coker } \alpha_{M_\lambda}$, $v'_{M_\lambda} : \text{Ker } \alpha_{M_\lambda} \longrightarrow H_f^1(F, M^*(1))_\lambda^*$ la transposée de $v_{M^*(1)_\lambda}$ et $u'_{M_\lambda} : \text{Coker } \alpha_{M_\lambda} \longrightarrow H^0(F, M^*(1))_\lambda^*$ la transposée de $u_{M^*(1)_\lambda}$.

Soient M une structure motivique et N une extension, dans \mathcal{M} , de $\mathbf{1}_{F, E}$ par M dont la classe appartient à $H_f^1(F, M)$. Si $x \in H_f^1(F, M^*(1))_\lambda^*$, notons x_N son image dans $H_f^1(F, N^*(1))_\lambda^*$ et

$$\Lambda(M, N, \lambda) = \{x \in H_f^1(F, M^*(1))_\lambda^* \mid x_N \in \text{Im } v'_{N_\lambda}\} .$$

Si $x \in \Lambda(M, N, \lambda)$, et si $x_N = v'_{N_\lambda}(y)$, l'image de y dans $\text{Ker } \alpha_{\mathbf{1}_{F, E}}$ est l'image d'un $z \in H^0(F, \mathbf{1}_{F, E})_\lambda$ et l'image $\delta_{M, N, \lambda}(x)$ de z dans $H_f^1(F, M)_\lambda$ ne dépend pas des choix faits.

Il n'est pas difficile de déduire de la f -admissibilité l'existence d'une unique transformation naturelle

$$\delta_\lambda = (\delta_{M, \lambda})_{M \in \text{Ob}(\mathcal{M})} \text{ avec } \delta_{M, \lambda} : H_f^1(F, M^*(1))_\lambda^* \longrightarrow H_f^1(F, M)_\lambda^*$$

telle que, si M et N sont comme ci-dessus et si $x \in \Lambda(M, N, \lambda)$, alors $\delta_{M, \lambda}(x) = \delta_{M, N, \lambda}(x)$.

6.10. — De la f -admissibilité on peut déduire également que, pour toute structure motivique M et tout $\lambda \in S_\infty(E)$, la suite de E_λ -espaces vectoriels de dimension finie

$$\begin{aligned} s_{f, \lambda}(M) \quad 0 \rightarrow H^0(F, M)_\lambda \rightarrow \text{Ker } \alpha_{M_\lambda} \rightarrow (H_f^1(F, M^*(1))_\lambda^* \rightarrow H_f^1(F, M)_\lambda \\ \rightarrow \text{Coker } \alpha_{M_\lambda} \rightarrow H^0(F, M^*(1))_\lambda^* \rightarrow 0 \end{aligned}$$

est exacte.

Avec les notations de l'introduction, on en déduit un isomorphisme canonique de $(L_f(M) \otimes_E L_f(M^*(1)))_\lambda$ sur $\det_{E_\lambda}(\text{Ker } \alpha_{M_\lambda}) \otimes_{E_\lambda} \det_{E_\lambda}^*(\text{Coker } \alpha_{M_\lambda})$. Par ailleurs, la suite exacte tautologique

$$0 \longrightarrow \text{Ker } \alpha_{M_\lambda} \longrightarrow (\bigoplus_{\mathfrak{p} \in S_\infty(F)} M_{B, \mathfrak{p}}^+)_\lambda \longrightarrow t_{M, \lambda} \longrightarrow \text{Coker } \alpha_{M_\lambda} \longrightarrow 0$$

fournit un isomorphisme de $(\det_{E_\lambda}^*(\bigoplus_{\mathfrak{p} \in S_\infty(F)} M_{B, \mathfrak{p}}^+) \otimes_E \det_E t_M)_\lambda$ sur $\det_{E_\lambda}^*(\text{Ker } \alpha_{M_\lambda}) \otimes_{E_\lambda} \det_{E_\lambda}(\text{Coker } \alpha_{M_\lambda})$. Le produit tensoriel de ces deux isomorphismes est un isomorphisme

$$\iota_{M, \lambda} : \Delta_f(M)_\lambda \longrightarrow E_\lambda .$$

On a $\mathbf{R} \otimes_{\mathbf{Q}} \Delta_f(M) = \otimes_{\lambda \in S_{\infty}(E)} \Delta_f(M)_{\lambda}$ tandis que $\mathbf{R} \otimes_{\mathbf{Q}} E = \oplus_{\lambda \in S_{\infty}(E)} E_{\lambda}$ et l'**isomorphisme** ι_M permettant d'énoncer la conjecture $C_{DB}(M)$ est la somme directe des $\iota_{M,\lambda}$.

Enfin, si λ est une place finie de E , la propriété $C_{\lambda}(M)$ permet d'identifier $\Delta_f(M)_{\lambda}$ et $\Delta_f(M_{\lambda})$ (cf. § 3 et 4) et donc de définir la norme $|\cdot|_{\lambda,EP}$, ce qui permet d'énoncer $C_{BK}(M)$.

6.11. — **Remarque** : Si M est une structure motivique sur F , alors $\text{Res}_{F/\mathbf{Q}}(M)$ est un objet de $\mathcal{C}_{\mathbf{Q}}(E)$ qui admet une fonction L et celle-ci est égale à $L(M, s)$. Il est naturel de conjecturer que $\text{Res}_{F/\mathbf{Q}}(M)$ est une structure motivique sur \mathbf{Q} et que, inversement, si N est une structure motivique sur \mathbf{Q} , $\text{Ext}_{\mathbf{Q}/F}(N)$ est une structure motivique sur F et que les foncteurs $\text{Res}_{F/\mathbf{Q}}$ et $\text{Ext}_{\mathbf{Q}/F}$ vérifient les propriétés d'adjonction que l'on pense¹⁷. S'il en est ainsi, les conjectures pour M sont équivalentes aux conjectures pour $\text{Res}_{F/\mathbf{Q}}(M)$ ce qui fait que l'on peut, sans restreindre la généralité, supposer $F = \mathbf{Q}$, ce que nous ferons parfois¹⁸.

§ 7. — Simplifications, réductions

7.1. — Dans la plupart des cas que l'on considère effectivement au moins deux des quatre espaces vectoriels $H^0(F, M)$, $H_f^1(F, M)$, $H^0(F, M^*(1))$ et $H_f^1(F, M^*(1))$ sont nuls, ce qui simplifient la structure des suites exactes $s_{f,\lambda}(M)$ et l'énoncé des conjectures $C_r(M)$ et $C_{DB}(M)$.

– C'est le cas lorsque $W_0 M = 0$ (en particulier si M est **pur de poids** ≥ 1) puisqu'alors $H^0(F, M) = H_f^1(F, M) = H^0(F, M^*(1)) = 0$ (pour les H^0 , cela résulte de $gr_0^W M = gr_0^W M^*(1) = 0$; pour le H_f^1 , on remarque que toute extension N de $\mathbf{1}_{F,E}$ par M est scindée, car $W_0 N = \mathbf{1}_{F,E}$ s'identifie à un sous-objet de N , d'où la nullité de $H^1(F, M)$ et a fortiori celle de $H_f^1(F, M)$). Si $\lambda \in S_{\infty}(E)$, la suite $s_{f,\lambda}(M_{\lambda})$ se réduit à un isomorphisme

$$\text{Ker } \alpha_{M_{\lambda}} \longrightarrow (H_f^1(F, M^*(1))_{\lambda})^*$$

(en particulier, on doit avoir $r_M = \dim_E H_f^1(F, M^*(1)) = \dim_{E_{\lambda}} \text{Ker } \alpha_{M_{\lambda}}$, qui se calcule à partir de la seule connaissance des structures de Hodge mixtes $M_{\mathfrak{p},\lambda}$, pour $\mathfrak{p} \in S_{\infty}(F)$).

– Inversement, si $W_{-3} M = M$ (en particulier si M est **pur de poids** ≤ -3), on a $W_0(M^*(1)) = 0$; on doit donc avoir $r_M = 0$ et, si $\lambda \in S_{\infty}(E)$,

¹⁷ On s'attend à ce que la même chose se produise avec les foncteurs $\text{Res}_{F'/F}$ et $\text{Ext}_{F'/F}$ pour tout sous-corps F' de F .

¹⁸ Il est cependant parfois commode de travailler avec des motifs "semi-stables" et des fonctions L incomplètes (cf. § 12). On ne peut plus alors remplacer F par \mathbf{Q} .

$s_{f,\lambda}(M_\lambda)$ se réduit à un isomorphisme

$$H_f^1(F, M)_\lambda \longrightarrow \text{Coker } \alpha_{M_\lambda} .$$

– On s'attend (bien que cela ne résulte pas du formalisme développé jusqu'à présent) qu'il n'y ait pas d'extension non triviale entre deux motifs purs de même poids. S'il en est ainsi, on voit que, si $W_{-1}M = 0$ (en particulier, si M est pur de poids 0) $s_{f,\lambda}(M_\lambda)$ se réduit à

$$0 \longrightarrow H^0(F, M)_\lambda \longrightarrow \text{Ker } \alpha_{M_\lambda} \longrightarrow (H_f^1(F, M^*(1))_\lambda^* \longrightarrow 0$$

(et donc encore à un isomorphisme de $\text{Ker } \alpha_{M_\lambda}$ sur le dual de $(H_f^1(F, M^*(1))_\lambda$ si $H^0(F, M) = 0$), tandis que, si $W_{-2}M = M$ (en particulier, si M est pur de poids -2), on doit avoir $r_M = -\dim_E H^0(F, M^*(1))$ et $s_{f,\lambda}(M_\lambda)$ se réduit à

$$0 \longrightarrow H_f^1(F, M)_\lambda \longrightarrow \text{Coker } \alpha_{M_\lambda} \longrightarrow H^0(F, M^*(1))_\lambda^* \longrightarrow 0 .$$

– Enfin, si M est pur de poids -1 , $\text{Ker } \alpha_M = \text{Coker } \alpha_{M_\lambda} = H^0(F, M) = H^0(F, M^*(1)) = 0$ et $s_{f,\lambda}(M_\lambda)$ se réduit à un isomorphisme

$$H_f^1(F, M^*(1))_\lambda^* \longrightarrow H_f^1(F, M)_\lambda ,$$

autrement dit à une dualité parfaite

$$H_f^1(F, M)_\lambda \times H_f^1(F, M^*(1))_\lambda \longrightarrow E_\lambda .$$

7.2. — La situation est particulièrement agréable lorsque M est f -close, i.e. lorsque $H_f^1(F, M) = H_f^1(F, M^*(1)) = 0$ (il suffit de savoir que $H_f^1(F, M_\lambda) = H_f^1(F, M^*(1))_\lambda = 0$ pour une place finie λ). Toutes les applications intervenant dans la suite exacte $s_{f,\lambda}(M)$ sont alors triviales, ce qui rend $C_{DB}(M)$ particulièrement simple à énoncer, surtout si $H^0(F, M) = H^0(F, M^*(1)) = 0$: soit \mathcal{B} (resp. \mathcal{B}') une base de $\bigoplus_{\mathfrak{p} \in S_\infty(F)} M_{B,\mathfrak{p}}^+$ (resp. t_M) sur E ; pour tout $\lambda \in S_\infty(E)$,

$$\alpha_{M_\lambda} : (\bigoplus_{\mathfrak{p} \in S_\infty(F)} M_B^+)_\lambda \longrightarrow t_{M,\lambda}$$

doit être un isomorphisme et, si b_λ est le déterminant de la matrice de α_{M_λ} relativement aux bases choisies, il doit exister $c \in E^*$ tel que $L(M, 0) = c \cdot (b_\lambda)_{\lambda \in S_\infty(E)}$.

Appelons f -équivalence sur \mathcal{M} la relation d'équivalence la plus restrictive sur $\text{Obj}(\mathcal{M})$ telle que,

i) si N est une extension de $\mathbf{1}_{F,E}$ par M dont la classe appartient à $H_f^1(F, M)$, alors M et N sont f -équivalents,

ii) si M et N sont f -équivalents, $M^*(1)$ et $N^*(1)$ le sont aussi.

Si M et N sont f -équivalents, $C_r(M)$ (resp. $C_{DB}(M)$, $C_{BK}(M)$) est vraie si et seulement si $C_r(N)$ (resp. $C_{DB}(N)$, $C_{BK}(N)$) l'est. Comme tout objet de \mathcal{M} est f -équivalent à un objet f -clos, il suffirait théoriquement de vérifier ces conjectures pour les objets f -clos.

7.3. — Une structure motivique M pure de poids w est dite **critique** s'il existe $\lambda \in S_\infty(E)$ telle que α_{M_λ} soit un isomorphisme (et α_{M_μ} est alors un isomorphisme pour tout $\mu \in S_\infty(E)$). Si M n'est pas de poids -1 , on voit que cela implique que M est f -close. Si M est de poids -1 et si M est critique, conjecturalement $r_M \geq 0$ et M est f -close si et seulement si $r_M = 0$, i.e. si et seulement si $L(M, 0) \neq 0$.

B – Cas particuliers

§8. — Le cas des 1-motifs

8.1. — Rappelons ([De74], §10) qu'une variété semi-abélienne A sur F est une extension d'une variété abélienne par un tore et qu'un 1-motif lisse sur F est un complexe de schémas en groupes commutatifs sur F , concentré en degré -1 et 0 ,

$$\mathcal{X} \xrightarrow{u} S,$$

où S est une variété semi-abélienne et \mathcal{X} est localement constant pour la topologie étale, $\mathcal{X}(\overline{F})$ étant un \mathbf{Z} -module libre de rang fini.

Les 1-motifs lisses sur F forment de manière naturelle une catégorie additive $\mathcal{MM}_{F,1}$. Notons $\mathcal{MM}_{F,1}(\mathbf{Q})$ la catégorie des 1-motifs à isogénie près, i.e. la catégorie suivante :

- les objets de $\mathcal{MM}_{F,1}(\mathbf{Q})$ sont les 1-motifs lisses,
- on a $\mathrm{Hom}_{\mathcal{MM}_{F,1}(\mathbf{Q})}(M_1, M_2) = \mathbf{Q} \otimes_{\mathbf{Z}} \mathrm{Hom}_{\mathcal{MM}_{F,1}}(M_1, M_2)$.

Il n'est pas difficile de voir que la catégorie \mathbf{Q} -linéaire $\mathcal{MM}_{F,1}(\mathbf{Q})$ est abélienne. On note $\mathcal{MM}_{F,1}(E)$ la catégorie des 1-motifs sur F à coefficients dans E , i.e. la catégorie E -linéaire déduite de $\mathcal{MM}_{F,1}(\mathbf{Q})$ par l'extension des scalaires $\mathbf{Q} \rightarrow E$. Il est commode de voir un objet M de $\mathcal{MM}_{F,1}(E)$ comme un complexe $[X \xrightarrow{u} S]$, où X est un E -espace vectoriel de dimension finie muni d'une action discrète de G_F , S une variété semi-abélienne sur F , munie d'un plongement de E dans $\mathbf{Q} \otimes_{\mathbf{Z}} \mathrm{End}_F(S)$ et $u : X \rightarrow \mathbf{Q} \otimes_{\mathbf{Z}} S(\overline{F})$ une application E -linéaire, G_F -équivariante.

Pour tout 1-motif $M = [X \xrightarrow{u} S]$ sur F à coefficients dans E , on sait (*loc. cit.*) définir sa filtration par le poids, lui associer sa réalisation de de Rham M_{dR} (c'est l'algèbre de Lie de l'extension universelle de M par un groupe vectoriel, tandis que l'espace tangent de M est l'algèbre de Lie de S), sa

réalisation Betti $M_{B,\mathfrak{p}}$ pour chaque place infinie \mathfrak{p} de F (appelée réalisation de Hodge dans *loc. cit.*) et sa réalisation galoisienne M_f (on a $M_f = \varinjlim_n M_n$, avec

$$M_n = \{(x, s) \in X \times S(\overline{F}) \mid u(x) = ns\} / \{(nx, u(x)) \mid x \in X\} .$$

On sait aussi ([Bures]) définir des isomorphismes de comparaison et montrer que l'on obtient bien ainsi un objet de $\mathcal{C}_F(E)$. Cette construction est fonctorielle. De la conjecture de Tate pour les variétés abéliennes (démontrée par Faltings, cf. [De83]), on déduit facilement que ce foncteur

$$\underline{\text{Réal}} : \mathcal{M}\mathcal{M}_{F,1}(E) \longrightarrow \mathcal{C}_F(E)$$

est pleinement fidèle et induit donc une équivalence entre $\mathcal{M}\mathcal{M}_{F,1}(E)$ et son image essentielle que nous notons $\underline{SM}_{F,1}(E)$. Celle-ci contient $\mathbf{1}_{F,E}$ (qui est la réalisation de $[E \rightarrow 0]$) et est stable par $M \mapsto M^*(1)$.

Si $M = \underline{\text{Réal}}(X \xrightarrow{u} S)$, tout $s \in \mathbb{Q} \otimes S(F)$ fournit une extension de $\mathbf{1}_{F,E} = \underline{\text{Réal}}([E \rightarrow 0])$ par M : c'est $\underline{\text{Réal}}([X \oplus E \xrightarrow{(u,v)} S])$, avec $v(1) = s$. Cette extension est scindée si et seulement si $s \in u(X)$. D'où une suite exacte

$$X \longrightarrow \mathbb{Q} \otimes S(F) \longrightarrow H_{\underline{SM}_{F,1}(E)}^1(F, M) \longrightarrow 0 .$$

8.2. — Si M est un objet de $\underline{SM}_{F,1}(E)$, on a $h^{r,s}(M) = 0^{19}$ sauf, peut-être, si $r, s \in \{-1, 0\}$. Il semble raisonnable de conjecturer que $\underline{SM}_{F,1}(E)$ est exactement la sous-catégorie de $\underline{SM}_F(E)$ formée des objets M tels que $h^{r,s}(M) = 0$ si $r \notin \{-1, 0\}$ ou si $s \notin \{-1, 0\}$.

En particulier $\underline{SM}_{F,1}(E)$ devrait être une sous-catégorie de $\underline{SM}_F(E)$ stable par sous-objet, quotient, somme directe et telle que, si elle contient M , elle contient toute extension de $\mathbf{1}_{F,E}$ par M . Elle devrait donc être f -admissible²⁰, i.e. tout objet $M = \underline{\text{Réal}}([X \xrightarrow{u} S])$ devrait satisfaire $C_\lambda^i(M)$ pour toute place λ de E . Il n'est pas difficile de voir que c'est le cas lorsque λ est archimédienne. Lorsque λ est fini, on vérifie que l'application $E_\lambda \otimes_E H^0(F, M) \longrightarrow H^0(F, M_\lambda)$ est un isomorphisme tandis que $E_\lambda \otimes_E H_f^1(F, M) \longrightarrow H_f^1(F, M_\lambda)$ est injective.

¹⁹ On peut définir les nombres de Hodge d'une structure motivique M en posant $h^{r,s}(M) = \dim_E(\text{Fil}^r M / \text{Fil}^{r+1} M)$ si M est pur de poids $r + s$ et $h^{r,s}(M) = h^{r,s}(gr_{r+s}^W M)$ en général.

²⁰ Aussi nous écrirons $H^1(F, M) = H_{\underline{SM}_{F,1}(E)}^1(F, M)$ et $H_f^1(F, M) = H_{\underline{SM}_{F,1}(E),f}^1(F, M)$.

La surjectivité pour toutes les places λ divisant un nombre premier ℓ fixé équivaut (cf. [BK90], prop. 5.4), si S est extension de la variété abélienne A par le tore T , à ce que la partie ℓ -primaire $\mathfrak{III}(A)(\ell)$ du groupe de Safarevic de A ,

$$\mathfrak{III}(A) = \text{Ker}(H^1(F, A(\overline{F})) \longrightarrow \prod_{\mathfrak{p} \in S(F)} H^1(F_{\mathfrak{p}}, A(\overline{F}_{\mathfrak{p}})))$$

soit finie. Par conséquent :

PROPOSITION. — *La catégorie $\underline{SM}_{F,1}(\mathbb{Q})$ est f -admissible si et seulement si, pour toute variété abélienne A sur F et tout nombre premier ℓ , $\mathfrak{III}(A)(\ell)$ est finie. S'il en est ainsi, $\underline{SM}_{F,1}(E)$ est aussi f -admissible.*

En outre, sans même connaître la finitude des groupes de Safarevic, les propriétés établies de $\underline{SM}_{F,1}(E)$ permettent de construire la suite exacte $s_{f,\lambda}(M)$ pour tout objet M de $\underline{SM}_{F,1}(E)$ et tout $\lambda \in S_{\infty}(E)$.

8.3. — Intéressons nous au cas $E = \mathbb{Q}$ et regardons quelques cas particuliers :

a) **Le cas $M = \mathbf{1}_{F,\mathbb{Q}}$** , réalisation de $[\mathbb{Q} \rightarrow 0]$: alors $M^*(1) = \mathbf{1}_{F,\mathbb{Q}}(1)$, réalisation de $[0 \rightarrow \mathbf{G}_m]$. On a $H^0(F, M) = \mathbb{Q}$, $H_f^1(F, M) = H^1(F, M) = 0$, $H^0(F, M^*(1)) = 0$, $H^1(F, M^*(1)) = \mathbb{Q} \otimes F^*$ tandis que $H_f^1(F, M^*(1))$ s'identifie au sous- \mathbb{Q} -espace vectoriel $\mathbb{Q} \otimes U_F$ où U_F désigne le groupe des unités de F . Pour tout $\mathfrak{p} \in S_{\infty}(F)$, on a $M_{B,\mathfrak{p}} = \mathbb{Q}$, avec action triviale de G_F et $t_M = 0$, de sorte que $\alpha_{M_{\infty}}$ est l'application 0, $\text{Ker } \alpha_{M_{\infty}} = \bigoplus_{\mathfrak{p} \in S_{\infty}(F)} \mathbb{R}$ et $\text{Coker } \alpha_{M_{\infty}} = 0$. La suite exacte $s_{f,\infty}(M)$ se réduit à

$$0 \longrightarrow \mathbb{R} \longrightarrow \bigoplus_{\mathfrak{p} \in S_{\infty}(F)} \mathbb{R} \longrightarrow (\mathbb{R} \otimes_{\mathbb{Z}} U_F)^* \longrightarrow 0,$$

où $\mathbb{R} \longrightarrow \bigoplus_{\mathfrak{p} \in S_{\infty}(F)} \mathbb{R}$ est l'application diagonale tandis que l'application suivante est la transposée de l'application \mathbb{R} -linéaire, déduite par extension des scalaires de l'application

$$U_F \longrightarrow \bigoplus_{\mathfrak{p} \in S_{\infty}(F)} \mathbb{R}$$

qui est $u \longmapsto (\log \|u\|_{\mathfrak{p}})_{\mathfrak{p} \in S_{\infty}(F)}$ ²¹.

La fonction $L(M, s)$ est la fonction $\zeta(F, s)$ du corps de nombres F . La conjecture $C_r(M)$ dit que, si $r = \#S_{\infty}(F)$, $\zeta(F, s)$ a un pôle en $s = 0$ de multiplicité $r - 1$, $C_{DB}(M)$ dit que, si u_1, u_1, \dots, u_{r-1} sont des unités de F linéairement indépendantes sur \mathbb{Z} , si $\mathfrak{p}_1, \mathfrak{p}_2, \dots, \mathfrak{p}_{r-1}$ sont des éléments distincts de $S_{\infty}(F)$ et si $R = \det(\log \|u_i\|_{\mathfrak{p}_j})$, $\zeta^*(F, 0)/R \in \mathbb{Q}^*$, tandis que les

²¹ Où $\| \cdot \|_{\mathfrak{p}}$ est la valeur absolue usuelle si \mathfrak{p} est réelle et son carré si \mathfrak{p} est imaginaire.

techniques esquissées au § 11 ci-dessous permettent de montrer que $C_{BK}(M)$ dit que, si e est l'indice dans U_F du groupe engendré par les u_i et si h est le nombre de classes de F ,

$$\zeta^*(F, 0) = \pm Rh/e ,$$

ce qui est bien connu (cf. par exemple [Ta84], Chap. I).

b) **Le cas $\mathbf{1}_{F, \mathbf{Q}}(1) = \underline{\text{Réal}}[0 \rightarrow \mathbf{G}_m]$** : c'est le cas "dual du précédent". Si l'on pose $t = 2\pi i \in \mathbf{C}$, on a $M_{B, \mathfrak{p}} = \mathbf{Q} \cdot t$, pour tout $\mathfrak{p} \in S_{\infty}(F)$ et donc $M_{B, \mathfrak{p}}^+ = 0$ si \mathfrak{p} est réelle, $\mathbf{Q} \cdot t$ sinon. On a $t_M = F$. Si l'on note $S_{\mathbf{R}}(F)$ (resp. $S_{\mathbf{C}}(F)$) l'ensemble des places réelles (resp. imaginaires) de F , l'application

$$\alpha_{\mathbf{1}_{F, \mathbf{Q}}(1)_{\infty}} : \bigoplus_{\mathfrak{p} \in S_{\mathbf{C}}(F)} \mathbf{R} \cdot t \longrightarrow \mathbf{R} \otimes_{\mathbf{Q}} F = \left(\bigoplus_{\mathfrak{p} \in S_{\mathbf{R}}(F)} \mathbf{R} \right) \oplus \left(\bigoplus_{\mathfrak{p} \in S_{\mathbf{C}}(F)} \mathbf{C} \right)$$

est l'application qui envoie $(u_{\mathfrak{p}} \cdot t)_{\mathfrak{p} \in S_{\mathbf{C}}(F)}$ sur $(0, (u_{\mathfrak{p}} \cdot 2\pi i)_{\mathfrak{p} \in S_{\mathbf{C}}(F)})$. On a $\text{Ker } \alpha_{M_{\infty}} = 0$ et $\text{Coker } \alpha_{M_{\infty}}$ s'identifie à $\bigoplus_{\mathfrak{p} \in S_{\infty}(F)} \mathbf{R}$, canoniquement isomorphe à son dual, qui est aussi $\text{Ker } \alpha_{(\mathbf{1}_{F, \mathbf{Q}})_{\infty}}$. La suite exacte $s_{F, \infty}(M)$ est la transposée de $s_{F, \infty}(\mathbf{1}_{F, \mathbf{Q}})$.

On a $L(M, s) = \zeta(F, s+1)$. La conjecture $C_r(M)$ dit que $\zeta(F, s)$ a un pôle simple en $s = 1$, $C_{DB}(M)$ dit que, si d est le discriminant de F , $\zeta^*(F, 1) \cdot \sqrt{d}/R \in \mathbf{Q}^*$ et $C_{BK}(M)$ peut, via les techniques du § 11, se traduire en disant que

$$\zeta^*(F, 1) = \pm 2^{r_1} \cdot (2\pi)^{r_2} \cdot d^{-1/2} \cdot hR/e ,$$

ce qui est encore bien connu (cf. par exemple, [Ta84], chap. I).

c) **Le cas d'une extension non triviale de $\mathbf{1}_{F, \mathbf{Q}}$ par $\mathbf{1}_{F, \mathbf{Q}}(1)$** : nous allons, pour simplifier, supposer en outre $F = \mathbf{Q}$. Alors il existe q entier naturel ≥ 2 tel que $M = \underline{\text{Réal}}([\mathbf{Q} \xrightarrow{u} \mathbf{G}_m])$, avec $u(1) = q$. On a $W_{-1}M = \mathbf{1}_{\mathbf{Q}, \mathbf{Q}}(1)$. On peut écrire $M_{B, \infty} = \mathbf{Q} \cdot t \oplus \mathbf{Q}u$, avec $M_{B, \infty}^+ = \mathbf{Q}u$, $t_M = t_{W_{-1}M} = \mathbf{Q}$. L'application $\alpha_{M_{\infty}}$ envoie u sur $\log(q)$ et est un isomorphisme. La structure motivique $M^*(1)$ s'identifie canoniquement à M . On a $H^0(\mathbf{Q}, M) = 0$; le \mathbf{Q} -espace vectoriel $H^1(\mathbf{Q}, M)$ s'identifie au quotient de $\mathbf{Q} \otimes \mathbf{Q}^*$ par la droite engendrée par $1 \otimes q$. Posons $q = p_1^{r_1} \cdot p_2^{r_2} \cdots p_a^{r_a}$, avec les p_i des nombres premiers distincts et les r_i des entiers ≥ 1 . On vérifie que $H_f^1(\mathbf{Q}, M)$ s'identifie au sous- \mathbf{Q} -espace vectoriel de $H^1(\mathbf{Q}, M)$ engendré par les images des $1 \otimes p_i$. La suite exacte $s_{f, \infty}(M)$ se réduit à un isomorphisme de $H_f^1(\mathbf{Q}, M)_{\infty}^*$ sur $H_f^1(\mathbf{Q}, M)_{\infty}$, ce qui équivaut à une forme bilinéaire non dégénérée

$$h : H_f^1(\mathbf{Q}, M)_{\infty} \times H_f^1(\mathbf{Q}, M)_{\infty} \longrightarrow \mathbf{R} ;$$

si l'on pose $\varepsilon_{i,i} = \log(p_i)/r_i$ et $\varepsilon_{i,j} = 0$ si $i \neq j$, celle-ci est caractérisée par

$$h(p_i, p_j) = (\log(p_i) \cdot \log(p_j)) / \log(q) - \varepsilon_{i,j} .$$

On a $L(M, s) = \zeta(s) \cdot \zeta(s+1) \cdot \prod_{1 \leq i \leq a} (1 - p_i^{-s})$. Les conjectures $C_r(M)$, $C_{DB}(M)$ et $C_{BK}(M)$ sont alors des théorèmes qui se déduisent facilement des résultats sur la valeur de la fonction zéta de Riemann en $s = 0$ et $s = 1$.

d) **Le cas d'une variété abélienne**, i.e. le cas où $M = \text{Réal}[0 \rightarrow A]$, où A est une variété abélienne sur F . Quitte à remplacer A par sa restriction des scalaires à la Weil, on peut supposer $F = \mathbb{Q}$. On vérifie que $M^*(1)$ s'identifie à M , que α_{M_∞} est un isomorphisme, que $H^0(\mathbb{Q}, M) = 0$ et que $H_f^1(\mathbb{Q}, M) = H^1(\mathbb{Q}, M) = \mathbb{Q} \otimes A(\mathbb{Q})$, de sorte que, ici encore, la suite exacte $s_{f, \infty}(M)$ se réduit à un isomorphisme de $H_f^1(\mathbb{Q}, M)_\infty^*$ sur $H_f^1(\mathbb{Q}, M)_\infty$, ce qui équivaut à une forme bilinéaire non dégénérée

$$h : \mathbb{R} \otimes A(\mathbb{Q}) \times \mathbb{R} \otimes A(\mathbb{Q}) \longrightarrow \mathbb{R} .$$

On peut montrer (cf. [Bl80]) que celle-ci n'est autre que la hauteur de Néron-Tate.

On a $L(M, s) = L_A(s+1)$, où L_A est la fonction L de Hasse-Weil de A . Les techniques du § 11 permettent de montrer que la conjonction de $C_r(M)$ et $C_{BK}(M)$ équivaut à la conjecture de Birch et Swinnerton-Dyer²². On sait qu'il y a eu récemment des progrès importants sur cette conjecture (cf. [Ko90], [Pe89], [Gr91], [Ru91]).

§ 9. — Motifs des variétés projectives lisses

9.1. — Soit X une variété projective lisse sur \mathbb{Q}^{23} . On se propose d'étudier ce qui se passe pour les structures motiviques $H^i(X)(m)^{24}$, où $i, m \in \mathbb{Z}$ avec $i \geq 1$. Si $M = H^i(X)(m)$, la dualité de Poincaré jointe au théorème de Lefschetz "vache" permet d'identifier $M^*(1)$ à $H^i(X)(i+1-m)$. Nous allons étudier simultanément les conjectures pour M et $M^*(1)$, i.e. les valeurs de $L(H^i(X), s)$ en $s = m$ et en $s = n$, où $n = i+1-m$, ce qui nous permet de supposer $m \leq (i+1)/2 \leq n$. Remarquons que M est pur de poids $i-2m$ et $M^*(1)$ de poids $i-2n$.

9.2. — **Supposons d'abord** $m < i/2$, ce qui fait que M est de poids ≥ 1 et $M^*(1)$ de poids ≤ -3 . On doit donc avoir (n° 7.1) $H_f^1(\mathbb{Q}, M) = 0$, $r_{M^*(1)} = 0$

²² Une fois supposée la finitude de $\mathfrak{III}(A)$ qui est nécessaire pour donner un sens à $C_{BK}(M)$ aussi bien qu'à la conjecture de Birch et Swinnerton-Dyer.

²³ Le cas d'une variété projective lisse Y sur F se ramène à celui-ci en prenant pour X la restriction des scalaires à la Weil de Y .

²⁴ Je triche un peu : lorsque l'on veut associer à $H^i(X)(m)$ un objet de $\mathcal{C}_F(\mathbb{Q})$, on peut avoir des problèmes aux places de mauvaise réduction.

et $r_M = \dim_{\mathbf{Q}} H_f^1(\mathbf{Q}, M^*(1)) = \dim_{\mathbf{R}} \text{Ker } \alpha_{M_\infty} = \dim_{\mathbf{R}} \text{Coker } \alpha_{M^*(1)_\infty}$. On constate que $\text{Coker } \alpha_{M^*(1)_\infty} = H_{dR}^i(X)_{\mathbf{R}} / (\text{Fil}^n H_{dR}^i(X)_{\mathbf{R}} + H_B^i(X(\mathbf{C}), \mathbf{Q}(n))_{\mathbf{R}})$ s'identifie au groupe de cohomologie de Deligne $H_{\mathcal{D}}^{i+1}(X/\mathbf{R}, \mathbf{R}(n))$ (cf. par exemple [Sc88], [Ja88], § 4). Modulo les conjectures habituelles sur l'équation fonctionnelle vérifiée par $L(H^m(X), s)$ (*op. cit.*, § 1), les conjectures $C_r(M)$ et $C_r(M^*(1))$ sont vraies, tandis que $C_{DB}(M)$ et $C_{DB}(M^*(1))$ sont équivalentes (la situation est un peu plus compliquée pour la relation entre $C_{BK}(M)$ et $C_{BK}(M^*(1))$, cf. § 12).

On voit aussi que M est critique si et seulement si $H_{\mathcal{D}}^{i+1}(X/\mathbf{R}, \mathbf{R}(n)) = 0$ et qu'alors $C_{DB}(M)$ et $C_{DB}(M^*(1))$ sont des cas particuliers de la conjecture de Deligne ([De79], conj. 1.8).

9.3. — Pour comprendre le lien avec la conjecture de Beilinson, rappelons (cf. par exemple [So85]) que, si $r \in \mathbf{N}$, $K_r(X) \otimes \mathbf{Q}$ est muni d'opérations d'Adams $(\psi^a)_{a \in \mathbf{N}^*}$ et que $K_r(X) \otimes \mathbf{Q} = \bigoplus_{j \in \mathbf{N}} K_r(X)^{(j)}$, où $K_r(X)^{(j)}$ désigne le sous- \mathbf{Q} -espace vectoriel de $K_r(X) \otimes \mathbf{Q}$ où ψ^a opère par multiplication par a^j pour tout a . Rappelons aussi que Beilinson a introduit le **groupe de cohomologie motivique** $H_{\mathcal{M}}^{i+1}(X, \mathbf{Q}(n)) = K_{n-m}(X)^{(n)}$ et construit une application

$$r_{i,n,\infty} : H_{\mathcal{M}}^{i+1}(X, \mathbf{Q}(n)) \longrightarrow H_{\mathcal{D}}^{i+1}(X/\mathbf{R}, \mathbf{R}(n)).$$

On sait aussi (au moins lorsque X admet un modèle régulier sur \mathbf{Z} , sinon il faut encore une conjecture de plus, cf. *op. cit.*) définir un sous- \mathbf{Q} -espace vectoriel $H_{\mathcal{M}}^{i+1}(X, \mathbf{Q}(n))_{\mathbf{Z}}$ de $H_{\mathcal{M}}^{i+1}(X/\mathbf{R}, \mathbf{R}(n))$; la conjecture de Beilinson dit, dans ce cas, que $r_{i,n,\infty}$ induit un isomorphisme

$$c_{i,n,\infty} : \mathbf{R} \otimes_{\mathbf{Q}} H_{\mathcal{M}}^{i+1}(X, \mathbf{Q}(n))_{\mathbf{Z}} \longrightarrow H_{\mathcal{D}}^{i+1}(X/\mathbf{R}, \mathbf{R}(n)),$$

et donc aussi un isomorphisme $d_{i,n,\infty}$ des déterminants de ces deux \mathbf{R} -espaces vectoriels; et que, si $b_{\mathcal{D}}$ est une base bien choisie de $\det_{\mathbf{R}} H_{\mathcal{D}}^{i+1}(X/\mathbf{R}, \mathbf{R}(n))^{25}$, il existe une base $b_{\mathcal{M}}$ de $H_{\mathcal{M}}^{i+1}(X, \mathbf{Q}(n))_{\mathbf{Z}}$ telle que $d_{i,n,\infty}(b_{\mathcal{M}}) = L^*(H^i(X), n) \cdot b_{\mathcal{D}}$.

Bien sûr, on pense que $H_{\mathcal{M}}^{i+1}(X, \mathbf{Q}(n))_{\mathbf{Z}}$ doit pouvoir s'identifier à $H_f^1(\mathbf{Q}, M^*(1))$ et $c_{i,n,\infty}$ à l'application naturelle de $H_f^1(\mathbf{Q}, M^*(1))_{\mathbf{R}}$ sur $\text{Coker } \alpha_{M^*(1)_\infty}$. Modulo cette interprétation et l'équation fonctionnelle, c'est un exercice de vérifier que $C_{DB}(M)$ équivaut à la conjecture de Beilinson.

9.4. — En utilisant les classes de Chern en cohomologie ℓ -adique, on peut aussi définir [So81], pour tout nombre premier ℓ , des applications

$$r_{i,n,\ell} : H_{\mathcal{M}}^{i+1}(X, \mathbf{Q}(n)) \longrightarrow H^1(\mathbf{Q}, H_{\text{ét}}^i(X \otimes \overline{\mathbf{Q}}, \mathbf{Q}_{\ell}(n))).$$

²⁵ De façon précise, $\det_{\mathbf{R}} H_{\mathcal{D}}^{i+1}(X/\mathbf{R}, \mathbf{R}(n))$ s'identifie à $\mathbf{R} \otimes_{\mathbf{Q}} V$ avec $V = \det_{\mathbf{Q}} H^i(X(\mathbf{C}), \mathbf{Q}(n-1))^{(-1)^{n-1}} \otimes \det_{\mathbf{Q}}^* \text{Fil}^n H_{dR}^i(X)$ et il faut prendre $b_{\mathcal{D}} = 1 \otimes b$, avec b une base de V .

On conjecture que l'image de $H_{\mathcal{M}}^{i+1}(X, \mathbb{Q}(n))_{\mathbf{Z}}$ est contenue dans $H_f^1(\mathbb{Q}, H_{\acute{e}t}^i(X \otimes \overline{\mathbb{Q}}, \mathbb{Q}_{\ell}(n)))$ (et cela a été démontré sous des hypothèses assez larges par Gros [Gr90])²⁶. Bien sûr, on s'attend à ce que cette application induise un \mathbb{Q}_{ℓ} -isomorphisme

$$c_{n,i,\ell} : \mathbb{Q}_{\ell} \otimes H_{\mathcal{M}}^{i+1}(X, \mathbb{Q}(n))_{\mathbf{Z}} \longrightarrow H_f^1(\mathbb{Q}, H_{\acute{e}t}^i(X \otimes \overline{\mathbb{Q}}, \mathbb{Q}_{\ell}(n)))^{27}$$

et que celui-ci ne soit autre que l'isomorphisme naturel de $\mathbb{Q}_{\ell} \otimes H_f^1(\mathbb{Q}, M^*(1))$ sur $H_f^1(\mathbb{Q}, M^*(1)_{\ell})$. Autrement dit, l'utilisation du **groupe de cohomologie motivique** $H_{\mathcal{M}}^{i+1}(X, \mathbb{Q}(n))_{\mathbf{Z}}$ et des **régulateurs** $c_{n,i,\ell}$ pour ℓ place finie ou non de \mathbb{Q} , nous permet de supprimer les guillemets dans l'énoncé des conjectures $C_{BK}(M)$ et $C_{BK}(M^*(1))$ i.e. de la réinterpréter comme des conjectures où la catégorie \mathcal{M} n'intervient plus.

9.5. — **Le cas** $m = i/2$, avec i pair est très semblable, à ceci près que, comme M est de poids 0, il se peut que $H^0(\mathbb{Q}, M) \neq 0$. La fonction $L(H^{2m}(X), s)$ doit encore avoir un zéro d'ordre $= \dim_{\mathbb{R}} H_{\mathcal{D}}^{i+1}(X/\mathbb{R}, \mathbb{R}(n))$ en $s = m$, mais elle doit avoir un pôle d'ordre $\dim_{\mathbb{Q}} H^0(\mathbb{Q}, H^{2m}(X)(m))$ en $s = n = m + 1$. Via les mêmes réinterprétations que dans le cas $m < i/2$, la suite exacte $s_{f,\infty}(M^*(1))$ s'écrit

$$0 \longrightarrow \mathbf{R} \otimes_{\mathbb{Q}} H_{\mathcal{M}}^{i+1}(X, \mathbb{Q}(n))_{\mathbf{Z}} \longrightarrow H_{\mathcal{D}}^{i+1}(X/\mathbb{R}, \mathbb{R}(n)) \longrightarrow \mathbf{R} \otimes_{\mathbb{Q}} H^0(\mathbb{Q}, M)^* \longrightarrow 0$$

alors que, si $C_h^m(X)$ désigne le groupe des cycles de codimension m modulo équivalence homologique, la conjecture de Beilinson fait intervenir l'application (conjecturée être un isomorphisme)

$$\mathbf{R} \otimes_{\mathbb{Q}} H_{\mathcal{M}}^{i+1}(X, \mathbb{Q}(n))_{\mathbf{Z}} \oplus (\mathbf{R} \otimes C_h^m(X)) \longrightarrow H_{\mathcal{D}}^{i+1}(X/\mathbb{R}, \mathbb{R}(n)) ,$$

où $\mathbf{R} \otimes C_h^m(X) \longrightarrow H_{\mathcal{D}}^{i+1}(X/\mathbb{R}, \mathbb{R}(n))$ est induite par l'application cycle en cohomologie de de Rham. Modulo les mêmes interprétations que pour $m < i/2$ et la conjecture standard qui dit que l'application naturelle $C_h^m(X)_{\mathbb{Q}} \longrightarrow$

²⁶ Dans le contexte de Bloch–Kato, il paraît plus simple de définir $H_{\mathcal{M}}^{i+1}(X, \mathbb{Q}(n))_f$ comme le sous-groupe de $H_{\mathcal{M}}^{i+1}(X, \mathbb{Q}(n))$ formé des x tels que $r_{n,i,\ell}(x) \in H_f^1(\mathbb{Q}, H_{\acute{e}t}^i(X \otimes \overline{\mathbb{Q}}, \mathbb{Q}_{\ell}(n)))$ pour tout ℓ et de remplacer dans l'énoncé des conjectures $H_{\mathcal{M}}^{i+1}(X, \mathbb{Q}(n))_{\mathbf{Z}}$ par $H_{\mathcal{M}}^{i+1}(X, \mathbb{Q}(n))_f$ quitte à conjecturer, par ailleurs, que $H_{\mathcal{M}}^{i+1}(X, \mathbb{Q}(n))_{\mathbf{Z}} = H_{\mathcal{M}}^{i+1}(X, \mathbb{Q}(n))_f$.

²⁷ Le fait que $c_{n,i,\ell}$ est un isomorphisme a été démontré dans certains cas particuliers par Jannsen ([Ja89], [Ja90]).

$H^0(\mathbb{Q}, M)$ est un isomorphisme, on vérifie que la conjecture $C_{DB}(M)$ est encore équivalente à la conjecture de Beilinson²⁸.

9.6. — Considérons enfin le cas où $m = n = (i + 1)/2$ (avec i impair). Alors M est pur de poids -1 et s'identifie à $M^*(1)$ ce qui fait (cf. n° 7.1) que la suite exacte $s_{f,\infty}(M^*(1))$ correspond à l'existence d'une forme bilinéaire non dégénérée

$$\langle, \rangle_M: \mathbf{R} \otimes H_f^1(\mathbb{Q}, M) \times \mathbf{R} \otimes H_f^1(\mathbb{Q}, M) \longrightarrow \mathbf{R},$$

la dimension du \mathbb{Q} -espace vectoriel $H_f^1(\mathbb{Q}, M)$ devant être égale à l'ordre du zéro de $L(H^{2m-1}(X), s)$ en $s = m$.

Rappelons ([Be87], [Bl84b]) que, dans ce cas, si $CH^m(X)^0$ désigne le groupe des cycles de codimension m homologiquement équivalents à 0 modulo l'équivalence rationnelle, et si d désigne la dimension de X , il existe un accouplement

$$\langle, \rangle_m: CH^m(X)^0 \times CH^{d+1-m}(X)^0 \longrightarrow \mathbf{R}.$$

On conjecture (*loc. cit.*) que $\mathbb{Q} \otimes CH^m(X)^0$ est de dimension finie sur \mathbb{Q} , que les \mathbf{R} -espaces vectoriels $\mathbf{R} \otimes CH^m(X)^0$ et $\mathbf{R} \otimes CH^{d+1-m}(X)^0$ sont mis en dualité par cet accouplement et que, si $L \in CH^1(X)$ est la classe d'une section hyperplane, alors (si $m \leq (d + 1)/2$; sinon il n'y a qu'à échanger m et $d + 1 - m$) la flèche naturelle

$$L^{d+1-2m}: \mathbb{Q} \otimes CH^m(X)^0 \longrightarrow \mathbb{Q} \otimes CH^{d+1-m}(X)^0$$

est un isomorphisme.

Si on l'utilise pour identifier ces deux \mathbb{Q} -espaces vectoriels, \langle, \rangle_m induit une forme bilinéaire sur $\mathbf{R} \otimes CH^m(X)^0$. On s'attend à ce qu'il existe un isomorphisme naturel de $\mathbb{Q} \otimes CH^m(X)^0$ sur $H_f^1(\mathbb{Q}, M)$ identifiant \langle, \rangle_m à \langle, \rangle_M . S'il en est ainsi, $C_{DB}(M)$ est équivalente à la conjecture de Beilinson.

On peut construire [Ja90], pour chaque nombre premier ℓ , l'**application d'Abel–Jacobi ℓ -adique**

$$CH^m(X)^0 \longrightarrow H^1(\mathbb{Q}, H_{\text{ét}}^{2m-1}(X \otimes \overline{\mathbb{Q}}, \mathbb{Q}_\ell(m))) = H^1(\mathbb{Q}, M_\ell).$$

²⁸ Pour le démontrer, il faut comprendre ce qu'est l'application composée $\mathbf{R} \otimes C_h^m(X) \rightarrow H_{\mathcal{D}}^{i+1}(X/\mathbf{R}, \mathbf{R}(n)) \rightarrow \mathbf{R} \otimes H^0(\mathbb{Q}, M)^*$. Celle-ci se factorise par l'application naturelle de $\mathbf{R} \otimes C_h^m(X)$ dans $\mathbf{R} \otimes H^0(\mathbb{Q}, M)$. L'application $\mathbf{R} \otimes H^0(\mathbb{Q}, M) \rightarrow \mathbf{R} \otimes H^0(\mathbb{Q}, M)^*$ qui en résulte n'est autre que celle que l'on déduit par extension des scalaires de l'isomorphisme $H^0(\mathbb{Q}, M) \rightarrow H^0(\mathbb{Q}, M)^*$ qui résulte de l'identification de M à son dual (Lefschetz "vache" plus dualité de Poincaré).

On s'attend encore à ce que l'image soit contenue dans le H_f^1 et que cette application s'identifie à l'application naturelle de $H_f^1(\mathbb{Q}, M)$ dans $H_f^1(\mathbb{Q}, M_\ell)$.

§ 10. — **Motifs d'Artin tordus**

10.1. — Pour les valeurs aux entiers positifs de la fonction zêta de Riemann, Bloch et Kato démontrent (à peu de choses près) leur conjecture ([BK90], § 6). Leur méthode donne un peu plus : soit d un entier ≥ 1 et χ un caractère de Dirichlet, i.e. un homomorphisme de $\text{Gal}(\mathbb{Q}(\mu_d)/\mathbb{Q}) = (\mathbb{Z}/d\mathbb{Z})^*$ dans E^* . Il définit un motif d'Artin $E(\chi)$ sur \mathbb{Q} à coefficients dans E (que l'on peut voir comme la réalisation du 1-motif $[E \rightarrow 0]$, où $G_{\mathbb{Q}}$ agit sur E via χ). On s'intéresse au comportement en $s = m$ de $L(\chi, s)$, i.e. au comportement en $s = 0$ de $L(M, s)$ avec $M = E(\chi)(m)$. Disons que χ est pair (resp. impair) si $\chi(-1) = 1$ (resp. -1).

a) Si $m = 0$ ou 1 , on est dans le cadre des 1-motifs et les conjectures sont vraies.

b) Si $m < 0$ et si les parités de m et de χ sont opposées, $C_r(M)$ et $C_{DB}(M)$ disent que $L(\chi, m)$ est un nombre rationnel non nul, ce qui est bien connu ; on est dans un cas critique et, si ℓ est un nombre premier, on peut déduire de la “**conjecture principale**” (pour ℓ) de la théorie d'Iwasawa que, si b est comme dans $C_{DB}(M)$, on a bien $|1 \otimes b|_{\ell, EP} = 1$; grâce à Mazur et Wiles [MW84], c'est un théorème si ℓ^2 ne divise pas d^{29} .

c) Si $m \geq 2$ et si m et χ ont la même parité, on est encore dans un cas critique (dual du précédent si l'on remplace χ par χ^{-1}) ; l'équation fonctionnelle montre que $C_r(M)$ et $C_{DB}(M)$ sont vraies. Un calcul local assez délicat permet en outre de déduire du cas précédent que si b est comme dans $C_{DB}(M)$, alors on a bien $|1 \otimes b|_{\ell, EP} = 1$, au moins si ℓ ne divise pas d .

d) Si $m < 0$ et si m et χ ont la même parité, la fonction $L(M, s)$ a un pôle simple en $s = 1$ et on doit donc avoir $\dim_E H_f^1(\mathbb{Q}, M^*(1)) = 1$, où $M^*(1) = E(\chi^{-1})(1 - m)$. Deligne a construit une extension non triviale de $\mathbb{1}_{\mathbb{Q}, E}$ par $M^*(1)$ dans la catégorie $\mathcal{C}_{\mathbb{Q}}(E)$ ([De89], n° 3.18 ; en toute rigueur, il faudrait vérifier l'existence d'isomorphisme de comparaison pour les p divisant d). On s'attend bien sûr à ce que cette extension soit une structure motivique et à ce que sa classe engendre $H_f^1(\mathbb{Q}, M^*(1))$. S'il en est ainsi, alors $C_{DB}(M)$ est vraie et, avec toujours les mêmes notations, on peut vérifier que $|1 \otimes b|_{\ell, EP} = 1$, au moins si ℓ ne divise pas $\varphi(d)$.

e) Enfin le cas $m \geq 2$ avec parités différentes se déduit de (d) comme

²⁹ Autrement dit, $C_{BK}(M)$ est vraie à multiplication près par un nombre rationnel qui est une unité en dehors des ℓ tels que ℓ^2 divise d .

(c) de (b) : modulo le fait que l'extension construite par Deligne est "la bonne", on trouve que $C_{DB}(M)$ est vraie et que $|1 \otimes b|_{\ell, EP} = 1$, au moins si ℓ ne divise pas $\varphi(d)$.

10.2. — **Remarque** : Plaçons nous dans le cas (e); on peut de nouveau définir un groupe de cohomologie motivique $K_{2m-1}(\chi)^{(m)}$ qui est un E -espace vectoriel de dimension 1 et des régulateurs ([So79]) permettant, comme aux n° 9.3 et 9.4 de réexprimer les conjectures sans faire référence à des extensions de motifs. Il semble que Beilinson ait montré [Be90] que ces constructions sont compatibles avec celles de Deligne.

10.3. — Supposons le corps F totalement réel et considérons la fonction $\zeta(F, s)$ au voisinage de $s = m$. Bien sûr, c'est le cas particulier de la situation considérée au § 9 où l'on prend $X = \text{Spec } F$ et $i = 0$. On voit que les entiers m négatifs impairs sont critiques. La conjecture $C_{DB}(M)$ dit que pour un tel m , $\zeta(F, m)$ est un nombre rationnel non nul et on peut montrer que $C_{BK}(M)$ équivaut à une conjecture de Lichtenbaum ([Li72]). Avec les mêmes notations que ci-dessus, $C_{BK}(M)$ revient à prouver que, pour tout ℓ , $|1 \otimes b|_{\ell, EP} = 1$. On sait [BN78] que, pour ℓ fixé, cette condition résulte de la "conjecture principale" correspondante. En particulier, si F/\mathbb{Q} est abélienne, c'est vrai si ℓ^2 ne divise pas le conducteur de F ³⁰.

10.4. — **Remarques** : *i*) Considérons la catégorie des **motifs mixtes de Tate sur F** , i.e. la sous-catégorie pleine de la catégorie des motifs mixtes sur F à coefficients dans \mathbb{Q} dont les objets sont ceux dont tous les quotients simples sont isomorphes à un tordu à la Tate de l'objet-unité. Bien sûr, la seule fonction L intéressante produite par cette catégorie est la fonction $\zeta(F, s)$. Cependant les travaux de Deligne sur $\mathbb{P}^1 - \{0, 1, \infty\}$ ([De89], voir aussi [Ih89], [BGSV90]) semblent fournir les moyens de réaliser cette catégorie comme sous-catégorie pleine de $\mathcal{C}_F(\mathbb{Q})$; il serait intéressant de vérifier qu'il en est bien ainsi et que la catégorie obtenue est bien f -admissible. On peut se poser la même question avec la catégorie un peu plus grosse formée des motifs qui sont extensions successives de motifs d'Artin tordus.

ii) Nous n'avons fait qu'évoquer la **cohomologie motivique** ([BMS87], [Li87], [BGSV90],...) qui figure pourtant en bonne place parmi les ingrédients qui sont amenés à jouer un rôle important dans les développements liés à ces conjectures.

iii) Un brillant avenir semble aussi promis à l'**utilisation des variétés de Shimura pour la construction de motifs mixtes** (permettant en

³⁰ Cela peut aussi se déduire (mais c'est la même démonstration) du n° 10.1 en décomposant la fonction $\zeta(F, s)$ en produit de fonctions L de Dirichlet).

particulier, dans certains cas, de vérifier que lorsque $L(M, 0) = 0$, il existe une extension non triviale du motif unité par $M^*(1)$: voir le rapport [Ra90] et les travaux récents de Harder ([Ha89], ?) et de Franke (travail en préparation ?).

iv) En dehors des résultats sur les conjectures de Birch et Swinnerton-Dyer et de ceux sur les fonctions L de Dirichlet, il commence à y avoir quelques résultats partiels pour d'autres motifs. Disons seulement :

a) Que Bloch et Kato ([BK90], § 7) obtiennent des résultats sur $L(H^1(X), 2)$, lorsque X est une courbe elliptique à multiplication complexe définie sur E ; en gros, ils montrent que $C_{DB}(M)$ est vraie dans ce cas et, utilisant les résultats de Kolyvagin et Rubin sur la "conjecture principale" dans le cas elliptique, que $|1 \otimes b|_{\ell, EP} = 1$ si ℓ est "régulier pour E " [So87].

b) Que Nekovar [Ne92] obtient des résultats "à la manière de Kolyvagin" qui confortent la conjecture de Bloch et Kato pour $L(M(f), m)$ où M est le motif de rang 2 associé à une forme modulaire nouvelle, à coefficients rationnels, de poids $2m$ pour $\Gamma_0(N)$.

v) Tous ces résultats ont en commun l'utilisation des idées de Kolyvagin (en particulier l'utilisation des systèmes d'Euler, cf. [Pe89], [Ko90]) et la **théorie d'Iwasawa** y joue un grand rôle. A chaque fois on a besoin d'utiliser la "conjecture principale" de cette théorie pour le motif correspondant et il faut commencer par comprendre ce qu'elle doit être. Dans le cas "ordinaire", cela a été fait par Greenberg [Gre89] et Schneider [Sc89].

Dans le cas général, Kato a commencé à le faire [Ka91]. Très en gros, il s'agit de voir comment se comporte un motif M sur F dans une extension finie abélienne L de F , ce qui consiste essentiellement à refaire ce que l'on a fait ici pour des motifs à coefficients dans la \mathbb{Q} -algèbre étale $E[\text{Gal}(L/F)]$. Parallèlement, il faut développer la théorie p -adique, i.e. construire et étudier les fonctions L p -adiques, comprendre les analogues p -adiques des conjectures de Bloch et Kato (cf. entre autres [Co89], [Co91], [CP89], [Gre91], [Ne92b] [Pa91], [Pe92],...).

C – Compléments

§11. — Groupes de Safarevic et nombres de Tamagawa

Tout comme la conjecture de Birch et Swinnerton-Dyer, la conjecture de Bloch et Kato, sous sa forme originale fait intervenir des groupes de Safarevic et des nombres de Tamagawa. Faute d'avoir fait les vérifications nécessaires, nous allons, comme Bloch et Kato, nous limiter au cas où $E = \mathbb{Q}$.

11.1. — Considérons tout d'abord un \mathbb{Z}_ℓ -module de type fini T muni d'une action linéaire et continue de G_F , non ramifiée en dehors d'un nombre fini de places, tel que la représentation $V = \mathbb{Q}_\ell \otimes_{\mathbb{Z}_\ell} T$ soit pseudo-géométrique.

Notons $H_f^1(F, T)$ l'image inverse de $H_f^1(F, V)$ dans $H^1(F, T)$. C'est un \mathbf{Z}_ℓ -module de type fini et on peut parler de son déterminant $\det_\ell H_f^1(F, T)$ qui est un \mathbf{Z}_ℓ -réseau de $\det_\ell H_f^1(F, V)$; de même $\det_\ell H^0(F, T)$ est un réseau de $\det_\ell H^0(F, V)$ et $L_f(T) = \det_\ell H^0(F, T) \otimes \det_\ell^* H_f^1(F, T)$ est un réseau de $L_f(V)$; si $T^+ = \bigoplus_{\mathfrak{p} \in S_\infty(F)} H^0(G_{\mathfrak{p}}, T)$, $\det_\ell T^+$ est un réseau de $\det_\ell V^+$; enfin si $T^*(1)$ est le "dual tordu à la Tate" de T , $L_f(T^*(1))$ est un réseau de $L_f(V^*(1))$. Par conséquent $\Delta'_f(T) = L_f(T) \otimes L_f(T^*(1)) \otimes \det_\ell^* T^+$ est un réseau de $\Delta'_f(V) = L_f(V) \otimes L_f(V^*(1)) \otimes \det_\ell^* V^+$. Choisissons une base ω'_f de $\Delta'_f(T)$ et une base ω de $\det_\ell t_V$. Le problème est de calculer $|\omega'_f \otimes \omega|_{\ell, EP} \in \ell^\times$ (qui dépend, a priori, des choix de T et ω mais pas de ω'_f).

11.2. — Pour toute place \mathfrak{p} de F , notons encore $H_f^1(F_{\mathfrak{p}}, T)$ l'image inverse de $H_f^1(F_{\mathfrak{p}}, V)$ dans $H^1(F_{\mathfrak{p}}, T)$. Si \mathfrak{p} est une place finie, l'application naturelle

$$(\mathbf{Q}_\ell/\mathbf{Z}_\ell) \otimes_{\mathbf{Z}_\ell} H_f^1(F_{\mathfrak{p}}, T) \longrightarrow H^1(F_{\mathfrak{p}}, V/T)$$

est injective et nous notons $H_{f, \mathfrak{p}}^1(F_{\mathfrak{p}}, V/T)$ son conoyau. Si \mathfrak{p} est infinie, on pose $H_{f, \mathfrak{p}}^1(F_{\mathfrak{p}}, V/T) = H^1(F_{\mathfrak{p}}, V/T)$. On choisit un ensemble fini S de places de F contenant toutes les places à l'infini, toutes les places divisant ℓ et toutes les places où V est ramifiée. Bloch et Kato définissent alors le groupe de Safarevic $\mathfrak{III}(T)$ comme le premier groupe de cohomologie du complexe

$$(\mathbf{Q}_\ell/\mathbf{Z}_\ell) \otimes_{\mathbf{Z}_\ell} H_f^1(F, T) \longrightarrow H^1(U_S, V/T) \longrightarrow \prod_{\mathfrak{p} \in S} H_{f, \mathfrak{p}}^1(F_{\mathfrak{p}}, V/T).$$

Il est facile de voir que ce groupe est fini et indépendant de S ³¹.

11.3. — Pour tout \mathbf{Q}_ℓ -espace vectoriel W de dimension finie, si ω est une base de $\det_\ell W$, notons μ_ω l'unique mesure de Haar sur W telle que, si Λ_0 est un réseau de W vérifiant $\det_\ell \Lambda_0 = \mathbf{Z}_\ell \cdot \omega$, alors $\mu_\omega(\Lambda_0) = 1$. Si Λ est un \mathbf{Z}_ℓ -module de type fini muni d'une identification de $\mathbf{Q}_\ell \otimes_{\mathbf{Z}_\ell} \Lambda$ à W et si $\det_\ell \Lambda = \mathbf{Z}_\ell \cdot a\omega$, avec $a \in \mathbf{Q}_\ell$, on a $\mu_\omega(\Lambda) = |a|_\ell$.

Soit \mathfrak{p} une place finie de F ; convenons que $t_{V, \mathfrak{p}} = 0$ si \mathfrak{p} ne divise pas ℓ . On dispose (cf. n° 4.4) d'une suite exacte

$$0 \longrightarrow H^0(F_{\mathfrak{p}}, V) \longrightarrow \underline{D}_{\mathfrak{p}}(V) \longrightarrow \underline{D}_{\mathfrak{p}}(V) \oplus t_{V, \mathfrak{p}} \longrightarrow H_f^1(F_{\mathfrak{p}}, V) \longrightarrow 0$$

permettant d'identifier $\det_\ell H_f^1(F_{\mathfrak{p}}, V) \otimes \det_\ell^* H^0(F_{\mathfrak{p}}, V)$ à $\det_\ell t_{V, \mathfrak{p}}$. Si $\omega_{\mathfrak{p}}$ est une base de $\det_\ell t_{V, \mathfrak{p}}$, et si ω_1 et ω_0^* sont des bases de $\det_\ell H_f^1(F_{\mathfrak{p}}, V)$

³¹ Lorsque $T = T_\ell(A)$, module de Tate d'une variété abélienne A , on vérifie que, si $\mathfrak{III}(A)(\ell)$ est fini, alors $\mathfrak{III}(A)(\ell) = \mathfrak{III}(T)$.

et $\det_\ell^* H^0(F_p, V)$ respectivement telles que $\omega_p = \omega_1 \otimes \omega_0^*$, le nombre $\mu_{\omega_1}(H_f^1(F_p, T))/\mu_{\omega_0}(H^0(F_p, T)) \in \ell^\times$ ne dépend que de ω_p et nous le notons $Tam_{p, \omega_p}^0(T)$. Si p ne divise pas ℓ , $\det_\ell t_{V, p} = \mathbb{Q}_\ell$ et on pose $Tam_p^0(T) = Tam_{p, 1}^0(T)$. On montre que $Tam_p^0(T) = 1$ pour presque tout p fini ne divisant pas ℓ . Le produit

$$\prod_{p \nmid \ell} Tam_p^0(T) \times \prod_{p|\ell} Tam_{p, \omega_p}^0(T)$$

a donc un sens et ne dépend que de la base $\omega = \pm \otimes_{p|\ell} \omega_p$ de $\det_\ell t_V$. On le note $Tam_\omega^0(T)$.

11.4. — On démontre

PROPOSITION. — Avec les hypothèses et notations ci-dessus, on a $\#\mathfrak{III}(T) = \#\mathfrak{III}(T^*(1))$ et

$$(|\omega'_f \otimes \omega|_{\ell, EP})^{-1} = Tam_\omega^0(T) \times \#\mathfrak{III}(T^*(1)).$$

11.5. — Soit maintenant p une place finie de F telle que $P_p(V, 1) \neq 0$. On voit qu'alors $H^0(F_p, V) = 0$ et que

- si p ne divise pas ℓ , $H_f^1(F_p, V) = 0$ donc $H_f^1(F_p, T) = H^1(F_p, T)_{tor}$;
- si p divise ℓ , l'exponentielle de Bloch–Kato, qui est l'application naturelle définie au n° 4.4,

$$t_{V, p} \longrightarrow H_f^1(F_p, V)$$

est un isomorphisme; en particulier, le choix d'une base ω_p de $\det_\ell t_{V, p}$ munit $H_f^1(F_p, V)$ par transport de structure d'une mesure de Haar μ_{BK, p, ω_p} . Si p ne divise pas ℓ et si $Tam_p(T) = \#H^1(F_p, T)_{tor}$, on vérifie que $Tam_p^0(T) = |L_p(V, 0)|_\ell \cdot Tam_p(T)$. De même, si p divise ℓ et si $Tam_{p, \omega_p}(T) = \mu_{BK, p, \omega_p}(H_f^1(F_p, T))$, on a $Tam_{p, \omega_p}^0(T) = |L_p(V, 0)|_\ell \cdot Tam_{p, \omega_p}(T)$.

11.6. — Supposons $F = \mathbb{Q}$ et venons-en à la conjecture originale de Bloch et Kato. Celle-ci est énoncée à la manière des théorèmes sur les nombres de Tamagawa des groupes semi-simples [Cl88] et nous allons avoir besoin d'introduire quelques notations supplémentaires.

Les couples (M, Θ) formés d'une structure motivique M sur \mathbb{Q} à coefficients dans \mathbb{Q} et d'un \mathbb{Z} -module de type fini Θ avec une action linéaire et continue de $G_{\mathbb{Q}}$ sur $\hat{\Theta} = \hat{\mathbb{Z}} \otimes \Theta$, munis d'un isomorphisme de $\mathbb{Q} \otimes \Theta$ sur $M_{B, \infty}$, avec les conditions de compatibilité que l'on pense, forment, de manière naturelle une catégorie abélienne. Notons $H^1(\mathbb{Q}, \Theta) = H^1(\mathbb{Q}, (M, \Theta))$ le groupe des classes d'extension de $(\mathbf{1}_{\mathbb{Q}, \mathbb{Q}}, \mathbb{Z})$ par (M, Θ) dans cette catégorie et $H_f^1(\mathbb{Q}, \Theta) \subset H^1(\mathbb{Q}, \Theta)$ l'image inverse de $H_f^1(\mathbb{Q}, M)$.

Considérons une structure motivique M telle que (a) $W_{-3}M = M$ et (b) $P_p(M, 1) (= (P_p(M_\ell, 1),$ pour un ℓ quelconque) $\neq 0$ pour tout nombre premier p ³². Choisissons Θ comme ci-dessus, sans torsion. Posons $A_\Theta(\mathbb{Q}) = H_f^1(\mathbb{Q}, \Theta)$, $A_\Theta(\mathbb{Q}_p) = H_f^1(\mathbb{Q}_p, \hat{\Theta})$ pour p premier, $A_\Theta(\mathbb{C}) = \mathbb{C} \otimes t_M / \text{Im}(\Theta)$ et $A_\Theta(\mathbb{R}) = (A_\Theta(\mathbb{C}))^{\text{Gal}(\mathbb{C}/\mathbb{R})}$ ^{33, 34}.

Si $\hat{\Theta} = \Pi \Theta_\ell$, et si p est premier, $A_\Theta(\mathbb{Q}_p) = \Pi_\ell H_f^1(\mathbb{Q}_p, \Theta_\ell)$ et (b) implique que, si $\ell \neq p$, $H_f^1(\mathbb{Q}_p, \Theta_\ell)$ est un groupe fini, trivial pour presque tout ℓ . Alors, $A_\Theta(\mathbb{Q}_p)$ est compact si p est fini, localement compact pour $p = \infty$, tandis que le quotient

$$\left(\prod_{p \in S(\mathbb{Q})} A_\Theta(\mathbb{Q}_p) \right) / A_\Theta(\mathbb{Q})$$

est compact.

Choisissons une base ω sur \mathbb{Q} de $\det_{\mathbb{Q}} t_M$. Pour chaque $p \in S(\mathbb{Q})$, ω peut être considérée comme une base de $\det_{\mathbb{Q}_p} t_{M_p}$. Si $p = \infty$, elle définit une mesure de Haar $\mu_{BK, \infty, \omega}$ sur $A_\Theta(\mathbb{R})$. Si p est fini, la mesure $\mu_{BK, p, \omega}$ sur $H_f^1(\mathbb{Q}_p, M_p) = \mathbb{Q} \otimes_{\mathbb{Z}} A_\Theta(\mathbb{Q}_p)$ peut être considérée comme une mesure sur $A_\Theta(\mathbb{Q}_p)$. Moyennant une hypothèse un peu technique sur les isomorphismes de comparaison p -adiques (qui dans la pratique est toujours satisfaite), Bloch et Kato montrent que, pour presque tout p ,

$$\mu_{BK, p, \omega}(A_\Theta(\mathbb{Q}_p)) = P_p(M, 1)$$

et le produit infini $\prod_{p \in S_f(\mathbb{Q})} \mu_{BK, p, \omega}(A_\Theta(\mathbb{Q}_p))$ converge. Ceci permet de définir la mesure produit $\mu_{BK} = \prod \mu_{BK, p, \omega}$ sur le groupe localement compact $\prod_{p \in S(\mathbb{Q})} A_\Theta(\mathbb{Q}_p)$ qui, grâce à la formule du produit, est indépendante de ω et que Bloch et Kato appelle **la mesure de Tamagawa**. Ils posent

$$Tam(\Theta) = \mu_{BK} \left(\left(\prod_{p \in S(\mathbb{Q})} A_\Theta(\mathbb{Q}_p) \right) / A_\Theta(\mathbb{Q}) \right)$$

³² Conjecturalement (b) résulte de (a) puisque l'on s'attend à ce que, dans $\mathbb{C}[t]$, $P_p(M, t) = \Pi(1 - \alpha_i t)$ avec $|\alpha_i| \leq p^{w/2}$ si $W_w M = M$, résultat qui, grâce à Deligne (conjectures de Weil), est vrai pour presque tout p .

³³ Bloch et Kato utilisent le point de vue "Beilinson", i.e. ils voient le $H_f^1(\mathbb{Q}, M)$ comme un groupe défini via la cohomologie motivique, muni de régulateurs.

³⁴ Ces notations suggèrent que l'on devrait savoir définir un faisceau de groupes abéliens A_Θ pour une certaine topologie. Lorsque $M = \text{Réal}[0 \rightarrow A]$ où A est une variété abélienne sur \mathbb{Q} , si on prend $\Theta = H_1(A(\mathbb{C}), \mathbb{Z})$, on peut montrer que $A_\Theta(K) = A(K)$, pour $K = \mathbb{Q}, \mathbb{Q}_p, \mathbb{C}$ ou \mathbb{R} .

(c'est aussi $\mu_{BK,\infty,\omega}(A_\Theta(\mathbb{R})/A_\Theta(\mathbb{Q})) \cdot \prod_{p \neq \infty} \mu_{BK,p,\omega}(A_\Theta(\mathbb{Q}_p))$); ils **conjecturent** que le groupe fini $\mathfrak{III}(\Theta_\ell)$ est nul pour presque tout ℓ , et que, si l'on pose $\mathfrak{III}(\Theta) = \prod_{\ell \text{ premier}} \mathfrak{III}(\Theta_\ell)$, alors

$$Tam(\Theta) = \#H^0(\mathbb{Q}, M^*(1)/\Theta^*(1)) / \#\mathfrak{III}(\Theta).$$

Pour $p \in S(\mathbb{Q})$, posons $\mu_{BK,p,\omega}^0 = \mu_{BK,\infty,\omega}$ si $p = \infty$ et $= |P_p(M, 1)|^{-1} \cdot \mu_{BK,p,\omega}$ sinon (c'est donc la mesure considérée au n° 11.3). Notons μ_{BK}^0 la mesure produit sur $\prod A_\Theta(\mathbb{Q}_p)$ et posons

$$Tam^0(M) = \mu_{BK}^0 \left(\left(\prod_{p \in S(\mathbb{Q})} A_\Theta(\mathbb{Q}_p) \right) / A_\Theta(\mathbb{Q}) \right).$$

La conjecture de Bloch et Kato équivaut à dire que

$$C_{Tam}(M) \quad L(M, 0) = Tam^0(\Theta) \cdot \#\mathfrak{III}(\Theta) / \#H^0(\mathbb{Q}, M^*(1)/\Theta^*(1)).$$

Il n'est pas difficile de déduire de la proposition 11.4 que cette conjecture équivaut à $C_{BK}(M)$.

11.7. — **Remarque** : Il semble que l'on puisse traduire la conjecture $C_{BK}(M)$ dans le style de $C_{Tam}(M)$ pour une structure motivique quelconque M . Mais il est alors indispensable de travailler avec la mesure μ_{BK}^0 (en dehors du domaine de convergence, la mesure μ_{BK} n'existe pas) et il faut remplacer le "groupe" A_Θ par un complexe.

§ 12. — Représentations λ -adiques et motifs semi-stables, équations fonctionnelles, fonctions L incomplètes³⁵

12.1. — Rappelons ([Se70]) que, pour tout $s \in \mathbb{C}$, on pose $\Gamma_{\mathbb{R}}(s) = \pi^{-s/2} \Gamma(s/2)$. Soient $K \in \{\mathbb{R}, \mathbb{C}\}$ et V une structure de Hodge mixte sur K à coefficients réels. On définit une filtration décroissante sur V par des sous \mathbb{R} -espaces vectoriels en posant, pour tout $r \in \mathbb{Z}$,

$\gamma^r V$ = le plus grand sous-espace vectoriel de V tel que $(\gamma^r V)_{\mathbb{C}} \subset Fil^r V_{\mathbb{C}}$. Lorsque $K = \mathbb{R}$, on remarque que $\gamma^r V$ est stable par $\text{Gal}(\mathbb{C}/\mathbb{R})$ qui agit donc sur $gr_\gamma^r V$; on pose alors $n_r^+ = \dim_{\mathbb{R}}(gr_\gamma^r V)^+$ et $n_r^- = \dim_{\mathbb{R}}(gr_\gamma^r V)^-$. Si $K = \mathbb{C}$, on pose $n_r^+ = n_r^- = \dim_{\mathbb{R}} gr_\gamma^r V$. Dans tous les cas, on pose³⁶

$$L(V, s) = \prod_{r \in \mathbb{Z}} \Gamma_{\mathbb{R}}(s + \varepsilon_r - r)^{n_r^+} \cdot \Gamma_{\mathbb{R}}(s + 1 - \varepsilon_r - r)^{n_r^-},$$

³⁵ cf. [FP92]. Pour simplifier, on se place dans le cas $E = \mathbb{Q}$, mais la généralisation ne présente pas de difficulté.

³⁶ On vérifie facilement que dans le cas d'une structure de Hodge pure, on retrouve la définition usuelle ([Se70]).

où $\varepsilon_r \in \{0, 1\}$ est défini par $\varepsilon_r \equiv r \pmod{2}$.

On peut également (cf. [FP92]) par des recettes du même genre définir le “facteur ε ” $\varepsilon(V, \psi_0, \mu_0)$.

On remarque que $(\gamma^0 V)^+$ s’identifie à $\text{Ker } \alpha_V$ et que $L(V, s)$ a un pôle en $s = 0$ d’ordre $\dim_{\mathbb{R}} \text{Ker } \alpha_V$.

12.2. — **Remarque** : On ne change pas les “bonnes propriétés” des facteurs locaux archimédiens si l’on remplace $\Gamma_{\mathbb{R}}(s)$ par $a \cdot \Gamma_{\mathbb{R}}(s)$. Les travaux récents de Deninger ([Den90], [Den91]) suggèrent de prendre $a = 1/\sqrt{2}$.

12.3. — Soient maintenant \mathfrak{p} une place finie de F et $B_{st, \mathfrak{p}}$ la $B_{cris, \mathfrak{p}}$ -algèbre B_{st} relative à l’extension $\overline{F}_{\mathfrak{p}}/F_{\mathfrak{p}}$ (cf. par exemple [II90], n° 1.2.3). Pour tout nombre premier ℓ , on dit qu’une représentation ℓ -adique V de $G_{\mathfrak{p}}$ est **semi-stable**

a) lorsque ℓ ne divise pas \mathfrak{p} , si l’action de l’inertie $I_{\mathfrak{p}}$ est unipotente;

b) lorsque ℓ divise \mathfrak{p} si $\dim_{(F_{\mathfrak{p}})_0} (B_{st} \otimes_{\mathbb{Q}_{\ell}} V)^{G_{\mathfrak{p}}} = \dim_{\mathbb{Q}_{\ell}} V$.

On dit que V est **potentiellement semi-stable** ou *pst* s’il existe une extension finie $F'_{\mathfrak{p}}$ de $F_{\mathfrak{p}}$ telle que V soit semi-stable en tant que représentation de $\text{Gal}(\overline{F}_{\mathfrak{p}}/F'_{\mathfrak{p}})$. Lorsque ℓ ne divise pas \mathfrak{p} , c’est toujours le cas (c’est le théorème de monodromie ℓ -adique de Grothendieck, [ST68], Appendix). Lorsque ℓ divise \mathfrak{p} ce n’est plus vrai³⁷.

A toute représentation ℓ -adique *pst* V de $G_{\mathfrak{p}}$, on sait associer une représentation linéaire de $W'_{\mathfrak{p}}$, groupe de Weil–Deligne de $F_{\mathfrak{p}}$, à coefficients dans \mathbb{Q}_{ℓ} si ℓ ne divise pas \mathfrak{p} (cf. [De73], [Ta79], § 4)³⁸, dans $(\overline{F}_{\mathfrak{p}})_0$ si ℓ divise \mathfrak{p} (cf. [Bures]). Ce qui fait que l’on peut associer à V non seulement un facteur $L_{\mathfrak{p}}(V, s) = (P_{\mathfrak{p}}(V, N(\mathfrak{p})^{-s})^{-1}$ défini comme au n° 3.3 mais aussi un conducteur $a_{\mathfrak{p}}(V)$ et un “facteur ε ” $\varepsilon_{\mathfrak{p}}(V, \psi, dx, s)$ ([Ta79], § 4).

12.4. — Disons qu’une représentation ℓ -adique de G_F est **géométrique** si elle est *pst* en toutes les places finies et non ramifiées en dehors d’un nombre fini d’entre elles.

“CONJECTURE” (cf. [FM92]). — *Soit ℓ un nombre premier. Le foncteur $M \mapsto M_{\ell}$ induit une \otimes -équivalence entre la catégorie \mathbb{Q}_{ℓ} -linéaire déduite de la catégorie $\underline{SM}_F(\mathbb{Q})$ par l’extension des scalaires $\mathbb{Q} \rightarrow \mathbb{Q}_{\ell}$ et la sous-catégorie pleine de la catégorie des représentations ℓ -adiques de G_F qui sont géométriques.*

³⁷ Toute représentation potentiellement semi-stable est de de Rham; j’ignore si la réciproque, qui constituerait une sorte de théorème de monodromie p -adique, est vraie.

³⁸ La construction de cette représentation dépend de certains choix, mais pas sa classe d’isomorphisme.

En particulier, si M est un objet de $\underline{SM}_F(\mathbb{Q})$, alors pour tout nombre premier ℓ , sa réalisation ℓ -adique est géométrique³⁹. S'il en est ainsi et si \mathfrak{p} est une place de F au-dessus du nombre premier p , on dispose d'une famille, indexée par les nombres premiers ℓ , de représentations de $W_{\mathfrak{p}}'$ dont on conjecture qu'elles sont définies sur \mathbb{Q} et compatibles ([De73], n° 8.7). Ceci permet de définir $a_{\mathfrak{p}}(M)$ et $\varepsilon_{\mathfrak{p}}(M, \psi, dx)$ comme étant $a_{\mathfrak{p}}(M_{\ell})$ et $\varepsilon_{\mathfrak{p}}(M_{\ell}, \psi, dx)$ pour n'importe quel choix de ℓ , y compris $\ell = p$. Si $\mathfrak{p} \in S_{\infty}(F)$, on note $\varepsilon_{\mathfrak{p}}(M, \psi_0, \mu_0, s)$ la constante $\varepsilon(M_{\mathfrak{p}}, \psi_0, \mu_0)$ où $M_{\mathfrak{p}}$ est la structure de Hodge mixte sur F associée à M . Si ψ désigne un caractère additif non trivial de \mathbf{A}_F/F et $dx = \otimes dx_{\mathfrak{p}}$ la mesure de Tamagawa sur \mathbf{A}_F , on peut alors définir le facteur $\varepsilon(M, s)$ par la formule usuelle

$$\varepsilon(M, s) = \prod_{\mathfrak{p}} \varepsilon_{\mathfrak{p}}(M, \psi_{\mathfrak{p}}, dx_{\mathfrak{p}}, s) .$$

Dans ce cadre général, si $\Lambda(M, s) = \prod_{\mathfrak{p} \in S(F)} L_{\mathfrak{p}}(M, s)$ est la fonction L "complète" de M , l'équation fonctionnelle s'écrit

"CONJECTURE" $C_{EF}(M)$. — *La fonction $\Lambda(M, s)$ admet un prolongement méromorphe dans \mathbb{C} et vérifie $\Lambda(M, s) = \varepsilon(M, s) \cdot \Lambda(M^*(1), -s)$.*

12.5. — Il semble que l'on puisse vérifier que, si

$$0 \longrightarrow M' \longrightarrow M \longrightarrow M'' \longrightarrow 0$$

est une suite exacte courte de $\underline{SM}_F(\mathbb{Q})$ et si deux des trois conjectures $C_{EF}(M')$, $C_{EF}(M)$, $C_{EF}(M'')$ sont satisfaites, il en est de même de la troisième.

Si $C_{EF}(M)$ est satisfaite, on peut vérifier que $C_r(M)$ et $C_r(M^*(1))$ sont équivalentes, de même que $C_{DB}(M)$ et $C_{DB}(M^*(1))$. Pour C_{BK} , c'est un peu plus subtil : B. Perrin-Riou a énoncé [FP92] une formule conjecturale $C_{EP, \mathfrak{p}}(V)$, de nature élémentaire mais un peu longue à expliquer, pour le calcul de la "caractéristique d'Euler-Poincaré" d'une représentation p -adique pst V de $G_{\mathfrak{p}}$ (où p divise \mathfrak{p}). On montre que, si le motif $\det(M)$ est un tordu à la Tate d'un motif d'Artin, si $C_{EP, \mathfrak{p}}(M_{\mathfrak{p}})$ est vraie pour tout \mathfrak{p} et si $C_{EF}(M)$ est vraie, alors $C_{BK}(M)$ et $C_{BK}(M^*(1))$ sont équivalentes⁴⁰.

³⁹ Les "bons" motifs seraient alors les motifs **semi-stables**, i.e. les M tels que, pour tout ℓ et tout \mathfrak{p} , M_{ℓ} munie de l'action de $G_{\mathfrak{p}}$, est semi-stable. On devrait avoir un **théorème de réduction semi-stable** disant que toute structure motivique sur F devient semi-stable après une extension finie des scalaires. Pour les 1-motifs, cela se déduit facilement du théorème de réduction semi-stable pour les variétés abéliennes.

⁴⁰ Il est amusant de remarquer que, dans la démonstration de Bloch et Kato de leur

12.6. — Soit S un ensemble fini de places de F contenant les places à l'infini. On peut alors énoncer des conjectures $C_{r,S}(M)$, $C_{DB,S}(M)$, $C_{BK,S}(M)$ sur le comportement en $s = 0$ de la fonction $L_S(M, s)$ via un formalisme tout à fait analogue à ce qu'on a développé pour $L(M, s)$ et nous ne le ferons pas, faute de place. Disons seulement que les groupes $H_f^1(F, M)$ et $H_f^1(F, M^*(1))$ doivent être remplacés par $H_{f,S}^1(F, M)$ et $H_{f,S}^1(F, M^*(1))$ où

$$H_{f,S}^1(F, M) = \{x \in H^1(F, M) \mid x_{\ell, \mathfrak{p}} \in H_f^1(F_{\mathfrak{p}}, M_{\ell}) \forall \ell \text{ premier et } \mathfrak{p} \notin S\},$$

que l'ordre du zéro éventuel en $s = 0$ de $L_S(M, s)$ est égal à $\dim_{\mathbb{Q}} H_{f,S}^1(F, M^*(1)) - \dim_{\mathbb{Q}} H^0(F, M)$ et que l'application naturelle de $H_{f,S}^1(F, M)$ dans $H^1(F, M_{\ell})$ doit induire un isomorphisme de $\mathbb{Q}_{\ell} \otimes H_{f,S}^1(F, M)$ sur

$$H_g^1(F, M_{\ell}) = \left\{ x \in H^1(F, M_{\ell}) \mid x_{\mathfrak{p}} \in \begin{cases} H_f^1(F_{\mathfrak{p}}, M_{\ell}) & \text{si } \mathfrak{p} \notin S \\ H_g^1(F_{\mathfrak{p}}, M_{\ell}) & \text{si } \mathfrak{p} \in S \end{cases} \right\}$$

où $H_g^1(F_{\mathfrak{p}}, M_{\ell})$ désigne le sous-groupe de $H^1(F_{\mathfrak{p}}, M_{\ell})$ classifiant les extensions de \mathbb{Q}_{ℓ} par M_{ℓ} qui sont *pst*.

Il est intéressant d'observer que la seule façon raisonnable d'arriver à rendre toutes ces conjectures compatibles entre elles conduit aux conjectures suivantes :

“CONJECTURE”. — *La catégorie $\underline{SM}_F(\mathbb{Q})$ est de dimension cohomologique 1.*

CONJECTURE. — *Pour tout nombre premier ℓ , la catégorie des représentations ℓ -adiques géométriques de F est de dimension cohomologique 1. En outre, pour toute représentation géométrique V et toute place finie \mathfrak{p} de F , la représentation du groupe $W'_{\mathfrak{p}}$ associée à V est F -semi-simple (au sens de [De73], déf. 8.6).*

Modulo ces conjectures, il semble que l'on puisse vérifier que les conjectures $C_{BK}(M)$ et $C_{BK,S}(M)$ sont équivalentes. En outre, il est facile de voir que, si

$$0 \longrightarrow M' \longrightarrow M \longrightarrow M'' \longrightarrow 0$$

est une suite exacte courte de $\underline{SM}_F(\mathbb{Q})$ qui “a bonne réduction en dehors de S ” (auquel cas $L_S(M, s) = L_S(M', s).L_S(M'', s)$) et si deux des trois

conjecture pour la fonction zéta de Riemann qu'ils ne regardent qu'aux entiers positifs, ce qu'ils font revient à la prouver d'abord aux entiers négatifs, puis à prouver la conjecture $C_{EP, \mathfrak{p}}(\mathbb{Q}_p(i))$ et à utiliser l'équation fonctionnelle.

conjectures $C_{BK,S}(M')$, $C_{BK,S}(M)$, $C_{BK,S}(M'')$ sont vraies, il en est de même de la troisième. Ces deux "résultats" impliquent le bon comportement des conjectures $C_{BK}(M)$ par suite exacte courte.

12.7. — Un autre intérêt des fonctions L incomplètes est que, étant donné un motif M "connu", et avec la terminologie du n° 6.2, on ne sait pas toujours que M admet une fonction $L(M, s)$, mais que l'on sait montrer en général que M admet une fonction $L_S(M, s)$ pourvu que S soit assez grand. La démonstration de la "conjecture" C_{BK} devrait pouvoir se décomposer en deux parties :

a) l'existence d'un bon formalisme;

b) la démonstration de $C_{BK,S}(M)$ pour un motif M "ayant bonne réduction en dehors de S " (que l'on doit même pouvoir supposer vérifier $H_{f,S}^1(F, M) = H_{f,S}^1(F, M^*(1)) = 0$).

Mais peut-être ne faut-il pas trop rêver!

RÉFÉRENCES

- [BN78] P. BAYER et J. NEUKIRCH. — On Values of Zeta Functions and ℓ -adic Euler Characteristics, *Invent. Math.* 50 (1978), 35–64.
- [Be85] A. BEILINSON. — Higher regulators and values of L -functions, *J. Soviet. Math.* 30 (1985), 2036–2070.
- [Be87] A. BEILINSON. — Height pairing between algebraic cycles, *Contemporary Math.* 67 (1987), 1–24.
- [Be90] A. BEILINSON. — Polylogarithms and cyclotomic elements, *Preprint* 1990.
- [BGSV90] A. BEILINSON, A. GONCHAROV, V. SCHECHTMAN and A. VARCHENKO. — Aomoto Dilogarithms, Mixed Hodge structures and Motivic Cohomology of Pairs of triangles on the Plane in **The Grothendieck Festschrift**, vol. I, *Prog. in Math.* 86, Birkhäuser (1990), 135–172.
- [BMS87] A. BEILINSON, R. MACPHERSON and V. SCHECHTMAN. — Notes on motivic cohomology, *Duke Math. J.* 54 (1987), 679–710.
- [BS63] B. BIRCH and H. SWINNERTON-DYER. — Notes on elliptic curves I, *Journal de Crelle* 212 (1963), 7–25.
- [BS65] B. BIRCH and H. SWINNERTON-DYER. — Notes on elliptic curves II, *Journal de Crelle* 218 (1965), 79–108.

- [Bl80] S. BLOCH. — A note on height pairings, Tamagawa numbers and Swinnerton–Dyer conjecture, *Invent. Math.* 58 (1980), 65–76.
- [Bl84] S. BLOCH. — Algebraic cycles and values of L -functions, *Journal de Crelle* 350 (1984), 94–108.
- [Bl84b] S. BLOCH. — Height pairings for algebraic cycles, *Journal of Pure and Applied Algebra* 34 (1984), 119–145.
- [Bl86] S. BLOCH. — Algebraic cycles and the Beilinson conjectures, *Contemporary Math.* 58 (1986), 65–79.
- [BK90] S. BLOCH et K. KATO. — L functions and Tamagawa numbers of motives, in **The Grothendieck Festschrift**, vol. 1, *Prog. in Math.* 86, Birkhäuser, Boston (1990), 333–400.
- [CL88] L. CLOZEL. — Nombres de Tamagawa des groupes semi-simples (d’après Kottwitz), *Séminaire Bourbaki*, exp. 702, novembre 1988.
- [Co89] J. COATES. — On p -adic L -functions, *Séminaire Bourbaki*, exp. 701, novembre 1988.
- [Co91] J. COATES. — Motivic p -adic L -functions, in **L -functions and Arithmetic**, *Proc. of the Durham Symposium*, *London Math. Soc. L.N.S.* 153, Cambridge University Press (1991), 141–172.
- [CP89] J. COATES and B. PERRIN–RIOU. — On p -adic L -functions attached to motives over \mathbb{Q} , *Adv. Stud. Pure Math.* 17, (1989), 23–54.
- [De71] P. DELIGNE. — Théorie de Hodge II, *Pub. Math. I.H.E.S.* 40 (1971), 5–57.
- [De73] P. DELIGNE. — Les constantes des équations fonctionnelles des fonctions L in **Modular Functions of One Variable**, vol. 2, *Lecture Notes in Math.* 349, Springer 1973, 501–595.
- [De74] P. DELIGNE. — Théorie de Hodge III, *Pub. Math. I.H.E.S.* 44 (1974), 5–77.
- [De79] P. DELIGNE. — Valeurs de fonctions L et périodes d’intégrales, *AMS, Proc. Symp. Pure Math.* 33 (1979), 313–346.
- [De83] P. DELIGNE. — Preuve des Conjectures de Tate et de Shafarevitch (d’après G. Faltings), *Séminaire Bourbaki*, exp. 616, novembre 1983.
- [De85] P. DELIGNE. — Lettre à C. Soulé, 1985.
- [De89] P. DELIGNE. — Le groupe fondamental de la droite projective moins trois points in **Galois Groups over \mathbb{Q}** , *MSRI publications* 16, Springer 1989, 79–297.

- [DM82] P. DELIGNE and J. MILNE. — Tannakian categories in **Hodge Cycles, Motives and Shimura Varieties**, Lecture Notes in Math. 900, Springer 1982, 101–228.
- [Den91] C. DENINGER. — On the Γ -factors attached to motives, *Invent. Math.* 104 (1991), 245–261.
- [Den92] C. DENINGER. — Local L -factors of motives and regularized determinants, *Invent. Math.* 107 (1992), 137–150.
- [DS91] C. DENINGER and A. SCHOLL. — The Beilinson conjectures, in **L -functions and Arithmetic**, Proc. of the Durham Symposium, London Math. Soc. L.N.S. 153, Cambridge University Press (1991), 173–209.
- [Fa89] G. FALTINGS. — Crystalline cohomology and p -adic Galois representations, in **Algebraic analysis, geometry and number theory**, Johns Hopkins University Press, Baltimore (1989), 25–80.
- [Fa90] G. FALTINGS. — F -isocrystals on Open Varieties : Results and Conjectures, in **The Grothendieck Festschrift**, vol. 2, Prog. in Math. 87, Birkhäuser, Boston (1990), 219–248.
- [FI90] J.-M. FONTAINE et L. ILLUSIE. — p -adic periods : a survey, prépublication d’Orsay, 1990.
- [FM92] J.-M. FONTAINE et B. MAZUR. — Représentations ℓ -adiques géométriques, in [Bures], en préparation.
- [FP91] J.-M. FONTAINE et B. PERRIN-RIOU. — Autour des conjectures de Bloch et Kato : I : Cohomologie galoisienne, II : Structures motiviques f -closes, III : Le cas général, C.R. Acad. Sci. Paris, 313, série I, 1991, 189–196, 349–356 et 421–428.
- [FP92] J.-M. FONTAINE et B. PERRIN-RIOU. — Autour des conjectures de Bloch et Kato : cohomologie galoisienne et valeurs de fonctions L , *prépublication d’Orsay*, 1992.
- [Gr90] M. GROS. — Régulateurs syntomiques et valeurs de fonctions L p -adiques (avec un appendice de M. Kurihara : Computation of the syntomic regulator in the cyclotomic case), *Invent. Math.* 99 (1990), 293–320.
- [Gr91] B. GROSS. — Kolyvagin’s work for modular elliptic curves, in **L -functions and Arithmetic**, London Math. Soc. L. Note Ser. vol. 153, Cambridge Univ. Press 1991, 235–256.
- [Gre89] R. GREENBERG. — Iwasawa theory for p -adic representations, *Adv. Stud. Pure Math.* 17, (1989), 97–137.

- [Gre91] R. GREENBERG. — Iwasawa theory for motives, in *L -functions and Arithmetic*, Proc. of the Durham Symposium, London Math. Soc. L.N.S. 153, Cambridge University Press (1991), 373–392.
- [Ha89] G. HARDER. — Arithmetische Eigenschaften von Eisenstein Klassen. die modulare Konstruktion von gemischten Motiven und von Erweiterungen endlicher Galoismoduln, *Preprint* 1989.
- [Il90] L. ILLUSIE. — Cohomologie de de Rham et cohomologie étale p -adique, *Séminaire Bourbaki*, exp. 726, juin 1990.
- [Ja88] U. JANNSEN. — Deligne Homology, Hodge- D -conjecture, and Motives, in [RSS88], 305–372.
- [Ja89] U. JANNSEN. — On the ℓ -adic cohomology of varieties over number fields and its Galois cohomology, in *Galois groups over \mathbb{Q}* , MSRI Publications vol. 16, Springer 1989.
- [Ja90] U. JANNSEN. — Mixed motives and algebraic K -theory, *Lecture Notes in Math.* 1400, Springer, Berlin (1990).
- [Ka91] K. KATO. — Iwasawa theory and p -adic Hodge theory, *preprint* 1991.
- [Ko90] V.A. KOLYVAGIN. — Euler Systems, in the Grothendieck Festschrift, vol. II, *Prog. in Math.* 87, Birkhäuser (1990), 435–483.
- [Li72] S. LICHTENBAUM. — On the values of zeta and L -functions : I, *Annals of Math.* 96 (1972), 338–360.
- [Li84] S. LICHTENBAUM. — Values of zeta functions at non-negative integers in *Number Theory*, Noordwijkerhout, *Lecture Notes in Math.* 1068, Springer, Berlin (1984), 127–138.
- [Li87] S. LICHTENBAUM. — The Construction of Weight-two Arithmetic Cohomology, *Invent. Math.* 88 (1987), 183–215.
- [Mi86] J.S. MILNE. — **Arithmetic duality theorems**, *Perspectives in Mathematics*, vol. 1, Academic Press (1986).
- [MW84] B. MAZUR and A. WILES. — Class fields of abelian extensions of \mathbb{Q} , *Invent. Math.* 76 (1984), 179–330.
- [Ne92] J. NEKOVAR. — Kolyvagin’s method for Chow groups of Kuga-Sato varieties, *Invent. Math.* 107 (1992), 99–125.
- [Ne92b] J. NEKOVAR. — On p -adic height pairing, *Séminaire de Théorie des Nombres de Paris*, à paraître.
- [Pa91] A. PANCHISHKIN. — Non Archimedean L -Functions of Siegel and Hilbert Modular Forms, *Lecture Notes in Math.* 1471, Springer 1991.
- [Pe89] B. PERRIN-RIOU. — Travaux de Kolyvagin et Rubin, *Séminaire Bourbaki*, exp. 717, novembre 1989.

- [Pe92] B. PERRIN-RIOU. — Théorie d'Iwasawa et hauteur p -adique, *Invent. Math.* 109 (1992), 137–185.
- [Ra89] D. RAMAKRISHNAN. — Regulators, Algebraic Cycles, and Values of L -Functions, in **Algebraic K -theory and Algebraic Number Theory**, Contemporary Mathematics, vol. 83, Amer. Math. Soc., Providence (1989), 183–310.
- [Ra90] D. RAMAKRISHNAN. — Problems arising from the Tate and Beilinson Conjectures in the context of Shimura Varieties, in **Automorphic Forms, Shimura Varieties, and L -functions**, vol. II, Perspectives in Mathematics, vol. 11, Academic Press (1990), 227–252.
- [RSS88] M. RAPOPORT, N. SCHAPPACHER and P. SCHNEIDER (éd.). — **Beilinson's Conjectures on Special Values of L -Functions**, Perspectives in Mathematics, vol. 4, Academic Press, Boston (1988).
- [Ru91] K. RUBIN. — The “main conjectures” of Iwasawa theory for imaginary quadratic fields, *Invent. Math.* 103 (1991), 25–68.
- [Sc88] P. SCHNEIDER. — Introduction to the Beilinson conjectures, in [RSS88], 1–35.
- [Sc89] P. SCHNEIDER. — Motivic Iwasawa theory, *Adv. Stud. Pure Math.* 17, (1989), 421–456.
- [Sc90] A.J. SCHOLL. — Motives for modular forms, *Invent. Math.* 100 (1990), 419–430.
- [Sc91] A.J. SCHOLL. — Remarks on special values of L -functions, in **L -functions and Arithmetic**, Proc. of the Durham Symposium, London Math. Soc. L.N.S. 153, Cambridge University Press (1991), 373–392.
- [Se70] J.-P. SERRE. — Facteurs locaux des fonctions zêta des variétés algébriques (définitions et conjectures), in **Collected Papers**, vol. 2, Springer Berlin (1986), 581–592.
- [ST68] J.-P. SERRE and J.-P. TATE. — Good reduction of abelian varieties, *Ann. of Math.* 88 (1968), 492–517.
- [So79] C. SOULÉ. — K -théorie des anneaux d'entiers de corps de nombres et cohomologie étale, *Invent. Math.* 55 (1979), 251–295.
- [So81] C. SOULÉ. — On higher p -adic regulators, in **Algebraic K -theory**, Lecture Notes in Math. 854, Springer, Berlin (1981), 371–401.
- [So83] C. SOULÉ. — K -théorie et zéros aux points entiers de fonctions zêta, Proc. ICM. Warszawa (1983).
- [So85] C. SOULÉ. — Régulateurs, *Séminaire Bourbaki*, exp. 644, février 1985.

- [So87] C. SOULÉ. — p -adic K -theory of elliptic curves, *Duke Math. Journal* 54 (1987), 249–269.
- [Ta65] J. TATE. — Algebraic cycles and poles of zeta functions, in **Arithmetical Algebraic Geometry**, Harper and Row, New York (1965).
- [Ta66] J. TATE. — On the Conjectures of Birch and Swinnerton–Dyer and a geometric analog, *Séminaire Bourbaki*, exp. 306, février 1966.
- [Ta79] J. TATE. — Number theoretic Background, in **Proceedings of Symposia in Pure Mathematics**, vol. 33 (1979), part 2, 3–26.
- [Ta84] J. TATE. — **Les Conjectures de Stark sur les Fonctions L d’Artin en $s = 0$** , Birkhäuser, Boston (1984).
- [Bures] Séminaire sur les périodes p -adiques (tenu à Bures en 1988), actes en prépartition.
- [SGA4] **Théorie des topos et cohomologie étale des schémas**, dirigé par M. Artin, A. Grothendieck et J.-L. Verdier, *Lecture Notes in Math.* 269, 270 et 305, Springer, Berlin (1972–73).
- [SGA $\frac{1}{2}$] **Cohomologie étale**, par P. Deligne, *Lecture Notes in Math.* 569, Springer, Berlin (1977).
- [SGA5] **Cohomologie ℓ -adique et Fonctions L** , dirigé par A. Grothendieck, *Lecture Note in Math.* 589, Springer, Berlin (1977).

Jean-Marc FONTAINE
Université de Paris-Sud
Arithmétique et Géométrie Algébrique
C.N.R.S. URA D0752
Bâtiment 425
F-91405 ORSAY CEDEX

Astérisque

JEAN-LOUIS LODAY

Excision en k -théorie algébrique

Astérisque, tome 206 (1992), Séminaire Bourbaki,
exp. n° 752, p. 251-271

<http://www.numdam.org/item?id=SB_1991-1992__34__251_0>

© Société mathématique de France, 1992, tous droits réservés.

L'accès aux archives de la collection « Astérisque » (<http://smf4.emath.fr/Publications/Asterisque/>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

EXCISION EN K -THÉORIE ALGÈBRIQUE,

d'après A. Suslin et M. Wodzicki

par Jean-Louis LODAY

Pour tout anneau A les groupes de K -théorie algébrique $K_n(A)$, $n \geq 0$, sont des groupes abéliens définis de la manière suivante. Pour $n \geq 0$, $K_0(A)$ est le groupe de Grothendieck des modules projectifs de type fini sur A . Pour $n \geq 1$, $K_n(A) := \pi_n(BGL(A)^+)$ où $GL(A)$ est la réunion des groupes linéaires $GL_n(A)$ ($n \times n$ matrices inversibles) et $BGL(A)$ est le classifiant du groupe discret $GL(A)$. La construction plus de Quillen donne l'espace $BGL(A)^+$ qui est un H -espace ayant même homologie que $BGL(A)$. En basses dimensions on trouve

$$K_1(A) = H_1(GL(A)) = GL(A)/[GL(A), GL(A)] = GL(A)/E(A),$$

où $E(A)$ est le sous-groupe de $GL(A)$ engendré par les matrices élémentaires. Pour $n = 2$ on a $K_2(A) = H_2(E(A))$.

Il est important pour les calculs de pouvoir comparer $K_n(A)$ et $K_n(A/I)$ lorsque I est un idéal bilatère. On définit abstraitement des groupes de K -théorie relatifs $K_n(A, I)$ (voir (1.3)) qui s'insèrent dans une suite exacte

$$(0) \quad \cdots \rightarrow K_{n+1}(A/I) \rightarrow K_n(A, I) \rightarrow K_n(A) \\ \rightarrow K_n(A/I) \rightarrow \cdots \rightarrow K_0(A/I).$$

La question naturelle qui se pose est alors la suivante : les groupes $K_n(A, I)$ ne dépendent-ils que de I et pas de l'anneau ambiant A ? En fait pour tout pseudo-anneau (=anneau sans élément unité) I on peut définir des groupes $K_n(I)$ de la façon suivante.

La loi $(n, u)(m, v) = (nm, nv + mu + uv)$ fait de $\mathbf{Z} \rtimes I$ un anneau unitaire (unité = $(1, 0)$) contenant I comme idéal bilatère. On pose

$$K_n(I) := \text{Ker}(K_n(\mathbf{Z} \rtimes I) \rightarrow K_n(\mathbf{Z})).$$

Remarquons que l'on a $K_n(I) = K_n(\mathbf{Z} \rtimes I, I)$. L'homomorphisme naturel d'anneaux $\mathbf{Z} \rtimes I \rightarrow A$ induisant l'identité sur I permet de reformuler la question précédente de la façon suivante :

– l'homomorphisme naturel $K_n(I) \rightarrow K_n(A, I)$ est-il un isomorphisme pour tout n et tout anneau A ?

Si la réponse est positive on dit que I est *K-excisif*.

H. Bass [B] a montré que la réponse est oui pour $n = 0$. Par contre ce n'est plus vrai pour $n > 0$ (voir la discussion en section 6). Dans [S-W] Suslin et Wodzicki donnent une condition nécessaire et suffisante simple pour l'excisivité de I en K -théorie algébrique rationnelle.

Définition [W3]. — On dit que la \mathbf{Q} -algèbre (non nécessairement unitaire) I est *H-unitaire* si le complexe

$$B_*(I) \quad \dots \rightarrow I^{\otimes n+1} \xrightarrow{b'} I^{\otimes n} \rightarrow \dots \rightarrow I \otimes_{\mathbf{Q}} I \xrightarrow{b'} I \rightarrow 0$$

où $\otimes = \otimes_{\mathbf{Q}}$ et $b'(a_0 \otimes \dots \otimes a_n) = \sum_{i=0}^{n-1} (-1)^i (a_0 \otimes \dots \otimes a_i a_{i+1} \otimes \dots \otimes a_n)$, est acyclique (i.e. $H_n(B_*(I)) = 0, n \geq 0$).

Remarquons que l'on a alors $I^2 = I$. Toute \mathbf{Q} -algèbre unitaire est *H-unitaire* car $s : I^{\otimes n-1} \rightarrow I^{\otimes n}$, $s(a_1 \otimes \dots \otimes a_{n-1}) = (1 \otimes a_1 \otimes \dots \otimes a_{n-1})$, vérifie $sb' + b's = Id$.

Supposons que pour tout ensemble fini $\{a_0, \dots, a_n\}$ d'éléments de I il existe un élément $u \in I$ tel que $ua_i = a_i u = a_i$ pour tout i . Alors I , qui

est dit *localement unitaire*, est H -unitaire. On trouvera d'autres exemples dans [W2].

Pour tout groupe abélien M on note $M_{\mathbb{Q}}$ le localisé $M \otimes_{\mathbb{Z}} \mathbb{Q}$.

THÉORÈME 1. — *Pour tout pseudo-anneau I les propriétés suivantes sont équivalentes :*

- (a) I est $K_{\mathbb{Q}}$ -excisif,
- (b) $I_{\mathbb{Q}}$ est H -unitaire.

On en déduit aisément le résultat suivant

THÉORÈME 2. — *Pour toute \mathbb{Q} -algèbre I (non nécessairement unitaire) les propriétés suivantes sont équivalentes*

- (a) I est K -excisif,
- (b) I est H -unitaire.

A. Suslin m'a annoncé qu'il était actuellement en train d'écrire la démonstration du résultat suivant pour $\Lambda = \mathbb{Z}, \mathbb{Z}/n\mathbb{Z}$ ou \mathbb{Q} :

“ I est K_{Λ} -excisif si et seulement si $\mathrm{Tor}_*^{\mathbb{Z} \ltimes I}(\mathbb{Z}, \Lambda) = 0$.”

Le but de cet exposé est d'indiquer les principales étapes de la démonstration du théorème 1 dans le cas où I est une \mathbb{Q} -algèbre. Dans la section 1 on traduit la K -excisivité en termes d'homologie du groupe GL . Dans la section 2 on relie l'hypothèse de H -unitarité à l'homologie cyclique, puis à l'homologie de l'algèbre de Lie gl . Le plan général de la preuve est discuté dans la section 3. Les sections 4 et 5 sont consacrées à la comparaison des homologies de GL et de gl .

Dans la dernière section (6) on aborde le problème du calcul de l'obstruction à l'excision, qui est évidemment plus général que de donner des conditions pour la nullité de cette obstruction.

Dans toute la suite on entend par “algèbre” une algèbre associative non nécessairement unitaire.

Avant de passer aux démonstrations, indiquons une généralisation et une application données dans le même article [S-W].

THÉORÈME 3. — *Supposons que pour tout m -uplet d'éléments a_1, \dots, a_m du pseudo-anneau I il existe des éléments $b_1, \dots, b_m, c, d \in I$ tels que $a_i = b_i cd$, $1 \leq i \leq m$, et tels que les anneaux gauches de cd et c coïncident. Alors I est K -excisif.*

L'une des applications remarquables est la preuve du résultat suivant, connu sous le nom de "conjecture de Karoubi".

THÉORÈME 4. — *Pour toute C^* -algèbre \mathcal{A} il y a un isomorphisme canonique*

$$K_n(\mathcal{A} \widehat{\otimes} \mathcal{K}) \cong K_n^{\text{top}}(\mathcal{A} \widehat{\otimes} \mathcal{K}), \quad n \geq 0.$$

Dans cette formule \mathcal{K} désigne la C^* -algèbre des opérateurs compacts d'un espace de Hilbert infini dénombrable, $\widehat{\otimes}$ est le produit tensoriel dans la catégorie des C^* -algèbres, K_n^{top} désigne la K -théorie topologique, i.e. $\pi_n(BGL(-))$ où la topologie de GL est prise en compte pour former l'espace classifiant.

Dans la même veine voici une application due à M. Wodzicki :

THÉORÈME 5. — *Pour toute C^* -algèbre A les groupes de K -théorie algébrique $K_n(A \otimes \mathcal{K})$ sont périodiques de période 2.*

Je remercie Andrei Suslin et Mariusz Wodzicki pour leurs commentaires durant la rédaction de cet article.

1. Groupe général linéaire et K -théorie algébrique

Pour tout anneau A le groupe linéaire $GL_n(A)$ est le groupe des $n \times n$ -matrices inversibles à coefficients dans A . En bordant la matrice par des 0 et un 1 on obtient une inclusion $i_n : GL_n(A) \hookrightarrow GL_{n+1}(A)$. La limite inductive des i_n est le groupe $GL(A)$. Considéré comme un groupe discret $GL(A)$ admet un classifiant $BGL(A)$ (espace d'Eilenberg-Mac Lane de type $K(GL(A), 1)$), dont l'homologie est notée $H_*(BGL(A))$ ou encore $H_*(GL(A))$.

Quillen a montré qu'on peut modifier $BGL(A)$ en lui ajoutant des cellules de manière à faire du nouvel espace $BGL(A)^+$ un H -espace sans en changer l'homologie (cf. [Q], ou [L1]). Par définition la K -théorie algébrique de l'anneau A est

$$(1.1) \quad K_n(A) := \pi_n(BGL(A)^+), \quad n \geq 1.$$

La principale propriété de ces groupes dont nous aurons besoin est la suivante. Le \mathbb{Q} -espace vectoriel gradué $H_*(GL(A), \mathbb{Q}) = \bigoplus_{n \geq 0} H_n(GL(A), \mathbb{Q})$ est muni d'une comultiplication, grâce à la diagonale, et d'une multiplication, grâce à la somme directe des matrices (car cette opération est interne sur $GL(A)$), qui en font une \mathbb{Q} -algèbre de Hopf graduée commutative et co-commutative. En conséquence $H_*(GL(A), \mathbb{Q})$ est complètement déterminé par sa partie primitive, qui est précisément la K -théorie de A . Plus précisément on a un isomorphisme

$$(1.2) \quad \Lambda(\bigoplus_{n \geq 1} K_n(A)_{\mathbb{Q}}) \cong H_*(GL(A), \mathbb{Q}),$$

où Λ est le foncteur "algèbre symétrique graduée".

Soit I un idéal bilatère de A . La surjection $A \rightarrow A/I$ induit un homomorphisme de groupes $GL(A) \rightarrow GL(A/I)$ dont on note $\overline{GL}(A/I)$ l'image. L'application induite $BGL(A)^+ \rightarrow B\overline{GL}(A/I)^+$ admet une fibre homotopique connexe notée $F(A, I)$. Notons que $\pi_n(BGL(A/I)^+) = \pi_n(\overline{BGL}(A/I)^+)$ pour $n > 1$ et $\pi_1(\overline{BGL}(A/I)^+) = \text{Im}(K_1(A) \rightarrow K_1(A/I))$. Par définition les *groupes de K -théorie relatifs* de la paire (A, I) sont les groupes

$$(1.3) \quad K_n(A, I) := \pi_n(F(A, I)), \quad n \geq 1.$$

La suite exacte (0) de l'introduction est donc tout simplement la longue suite exacte d'homotopie d'une fibration.

Signalons ici que les groupes K_n existent aussi pour $n < 1$. En fait ils se définissent à partir du groupe de Grothendieck K_0 pour lequel le problème d'excision est résolu.

La première étape pour aborder la question de l'excisivité en K -théorie algébrique rationnelle consiste à comparer l'homologie du groupe

$$GL(I) := \text{Ker}(GL(\mathbf{Z} \ltimes I) \rightarrow GL(\mathbf{Z}))$$

avec celle du groupe $\widetilde{GL}(I) := \lim_n \widetilde{GL}_n(I)$ où

$$\begin{aligned} \widetilde{GL}_n(I) &:= GL_n(I) \ltimes M_{n1}(I) \\ &= \left\{ \begin{pmatrix} \alpha & v \\ 0 & 1 \end{pmatrix} \in GL_{n+1}(I) \mid \alpha \in GL_n(I), v \in M_{n1}(I) \right\}. \end{aligned}$$

Ici $M_{n1}(I)$ désigne les $n \times 1$ -matrices à coefficients dans I . Notons que pour tout sous-groupe G de $GL_n(I)$ (resp. $GL(I)$) on a un groupe \widetilde{G} défini par $\widetilde{G} := G \ltimes M_{n1}(I)$. De manière analogue le groupe $\widetilde{\widetilde{GL}}_n(I)$ (et plus généralement $\widetilde{\widetilde{G}}$) est défini par

$$\widetilde{\widetilde{GL}}_n(I) := M_{1n}(I) \ltimes GL_n(I).$$

1.4. PROPOSITION. — *Considérons le diagramme commutatif de fibrations homotopiques suivant*

$$\begin{array}{ccccc} BGL(I) & \longrightarrow & BGL(A) & \longrightarrow & B\widetilde{GL}(A/I) \\ \downarrow \varphi & & \downarrow & & \downarrow \\ F(A, I) & \longrightarrow & BGL(A)^+ & \longrightarrow & B\widetilde{GL}(A/I)^+. \end{array}$$

Si les inclusions $GL(I) \rightarrow \widetilde{GL}(I)$ et $GL(I) \rightarrow \widetilde{\widetilde{GL}}(I)$ induisent des isomorphismes en homologie, alors l'application φ induit une équivalence d'homotopie :

$$\varphi^+ = BGL(I)^+ \longrightarrow F(A, I).$$

Preuve (Esquisse) (cf. Corollary 1.7 de [S-W]). Les deux applications verticales de droite induisent des isomorphismes en homologie par la propriété fondamentale de la construction plus. On a alors envie d'appliquer le théorème de comparaison des suites spectrales de Zeeman pour démontrer que φ induit un isomorphisme en homologie. Encore faut-il que les hypothèses soient remplies, à savoir que le π_1 de la base opère trivialement sur l'homologie de la fibre. Pour la fibration du bas c'est immédiat car on a à faire à des H -espaces. Pour la fibration du haut il suffit de montrer que tout élément provenant de $\pi_1(BGL_n(A)) = GL_n(A)$ opère trivialement (pour l'action par conjugaison) sur $H_*(GL(I))$. Une manipulation de matrices (cor. 1.6 of loc. cit.) permet de se ramener à montrer que $GL(\mathbf{Z})$ opère trivialement sur $H_*(GL(I))$. La première hypothèse (avec \widetilde{GL}) permet de montrer qu'il en est bien ainsi pour les matrices élémentaires $E_{i+1}(1)$ et la seconde (avec $\widetilde{\widetilde{GL}}$) pour les $E_{i+1,i}(1)$. Comme $GL(\mathbf{Z})$ est engendré par ces matrices élémentaires et -1 (qui opère trivialement), le tour est joué. \square

On a donc ainsi une condition homologique sur GL qui implique la K -excisivité de I . Remarquons qu'il suffit de montrer que cette condition est remplie pour \widetilde{GL} , la condition pour $\widetilde{\widetilde{GL}}$ s'en déduisant par dualité.

2. Algèbres de Lie des matrices et homologie cyclique

Si, dans le paragraphe précédent, on remplace le groupe $GL(A)$ par l'algèbre de Lie $gl(A)$ (où A est maintenant une algèbre associative unitaire sur un corps k de caractéristique 0), on a aussi que l'homologie de $gl(A)$ à coefficients dans k est une algèbre de Hopf graduée commutative et cocommutative. Elle est donc complètement déterminée par sa partie primitive que l'on peut appeler la " K -théorie additive" de A .

La grande différence avec la situation précédente est que l'on peut "calculer" cette partie primitive, en le sens qu'on peut exhiber un complexe *sans matrices* dont cette partie primitive est l'homologie :

2.1. THÉORÈME (Loday-Quillen [L-Q], Tsygan [T]). — *Pour toute algèbre associative unitaire A sur un corps k de caractéristique 0 il y a un isomorphisme d'algèbres de Hopf graduées :*

$$H_*(gl(A), k) \cong \Lambda HC_{*-1}(A),$$

où $HC_n(A), n \geq 0$, est l'homologie cyclique de A .

Dans ce cadre l'homologie cyclique se définit de la manière suivante. Reprenons le complexe $B_*(A)$ de l'introduction utilisé pour définir la notion de H -unitarité. Lorsque A est unitaire c'est une résolution de A par des A -bimodules (mettre $A^{\otimes n+2}$ en degré n). Notons A^{op} l'anneau opposé de A , A^e le produit tensoriel $A \otimes_k A^{op}$ et (a_0, \dots, a_n) le générateur $(a_0 \otimes \dots \otimes a_n)$ de $A^{\otimes n+1}$. Par définition le *complexe de Hochschild* $(C_*(A), b)$ est $(A \otimes_{A^e} B_*(A), id_A \otimes b')$. Explicitement on a $C_n(A) = A^{\otimes n+1}$ car $a \otimes (a_0, \dots, a_{n+1}) \mapsto (a_{n+1}aa_0, a_1, \dots, a_n)$, $A \otimes_{A^e} B_n(A) \cong A^{\otimes n+1}$ est un isomorphisme. L'application $b : C_n(A) \rightarrow C_{n-1}(A)$ est alors donnée par

$$b(a_0, \dots, a_n) = \sum_{i=0}^{n-1} (-1)^i (a_0, \dots, a_i a_{i+1}, \dots, a_n) + (-1)^{n+1} (a_n a_0, a_1, \dots, a_{n-1}).$$

L'homologie du complexe $(C_*(A), b)$ est notée $HH_n(A), n \geq 0$, et est appelée *l'homologie de Hochschild* de A .

Introduisons maintenant *l'opérateur cyclique* $t : C_n(A) \rightarrow C_n(A)$ défini par

$$t(a_0, \dots, a_n) = (-1)^n (a_n, a_0, \dots, a_{n-1}).$$

A. Connes [C] a remarqué que l'on a l'identité

$$b(1-t) = (1-t)b'.$$

En conséquence, si l'on pose $C_n^\lambda(A) := A^{\otimes n+1}/(1-t)$, on a un nouveau complexe $(C_*^\lambda(A), b)$ dont l'homologie, notée $HC_n(A)$, $n \geq 0$, est appelée *l'homologie cyclique* de A (cette définition est la bonne lorsque k contient \mathbb{Q} , cas qui nous intéresse ici. Sinon voir [L-Q], [L2]). L'une des propriétés principales de l'homologie cyclique est la longue suite exacte de périodicité de Connes

$$(2.2) \quad \cdots \rightarrow HH_n(I) \rightarrow HC_n(I) \rightarrow HC_{n-2}(I) \rightarrow HH_{n-1}(I) \rightarrow \cdots$$

qui est valable, avec les définitions données ici, pour toute algèbre H -unitaire I (cf. [C] pour le cas unitaire, rationnel et [L-Q] pour le cas général).

Bien que nous n'en ayons pas besoin, signalons, pour fixer les idées, le calcul suivant de l'homologie cyclique d'une \mathbb{Q} -algèbre (commutative, unitaire) lisse :

$$HC_n(A) \cong \Omega_A^n / d\Omega_A^{n-1} \oplus H_{DR}^{n-2}(A) \oplus H_{DR}^{n-4}(A) \oplus \cdots$$

Ici Ω_A^n désigne le module des n -formes différentielles absolues et H_{DR}^n est la cohomologie de de Rham.

Le problème de l'excision se pose aussi pour les théories HH et HC . Il est évidemment plus facile à résoudre que pour la K -théorie algébrique.

Tout épimorphisme d'algèbres $A \rightarrow A/I$ donne naissance à un épimorphisme de complexes $C(A) \rightarrow C(A/I)$ (resp. $C^\lambda(A) \rightarrow C^\lambda(A/I)$). Par définition l'homologie de Hochschild (resp. cyclique) relative $HH_n(A, I)$ (resp. $HC_n(A, I)$) est l'homologie du complexe noyau. On dit que I est *HH-excisé* (resp. *HC-excisé*) si $HH_n(I) := HH_n(\mathbb{Z} \times I) \rightarrow HH_n(A, I)$ (resp. $HC_n(\mathbb{Z} \times I, I) \rightarrow HC_n(A, I)$) est un isomorphisme pour tout n .

2.3. PROPOSITION [W1, thm 3]. — *Pour toute \mathbb{Q} -algèbre I les énoncés suivants sont équivalents :*

- a) I est H -unitaire,

- b) I est excisif pour HH ,
- c) I est excisif pour HC .

Commentaire sur la preuve. L'équivalence entre b) et c) est immédiate à partir de la suite exacte de Connes. De b) vers a) il suffit de calculer un cas particulier d'algèbre contenant un idéal nilpotent. De a) vers b) Wodzicki remarque astucieusement que le complexe dont on veut démontrer l'acyclicité est en fait le complexe total d'un "multi-complexe". L'hypothèse de H -unitarité implique l'acyclicité du multi-complexe lorsqu'on prend l'homologie dans une direction. Ceci suffit à impliquer l'acyclicité du complexe total. \square

Au vu de ce qu'on a fait pour la K -théorie et du lien entre $HC_*(A)$ et $H_*(gl(A))$, il est assez naturel de traduire la propriété d'excision pour HC en termes d'homologie de gl .

2.4. PROPOSITION. — *Pour toute \mathbb{Q} -algèbre H -unitaire I l'homomorphisme*

$$H_*(gl(I)) \rightarrow H_*(\tilde{gl}(I))$$

est un isomorphisme.

Preuve. Le même type d'arguments que dans la preuve ci-dessus permet de démontrer que l'algèbre $I_1 = \left\{ \begin{pmatrix} a & b \\ 0 & 0 \end{pmatrix} \mid a, b \in I \right\}$ est H -unitaire et que l'inclusion $I \hookrightarrow I_1$ induit un isomorphisme en homologie de Hochschild (cf. [W1]). Par la suite exacte de Connes c'est encore un isomorphisme en homologie cyclique. Pour transférer ce résultat à l'homologie de gl on utilise l'extension du théorème de Loday-Quillen-Tsygan (cf. 2.1) aux algèbres H -unitaires faite par P. Hanlon [H]. On en conclut un isomorphisme

$$H_*(gl(I)) \xrightarrow{\sim} H_*(gl(I_1)).$$

Il est aisé de comparer les homologies de $gl(I_1)$ et de $\tilde{gl}(I)$ pour conclure. \square

2.5. *Remarque.* Signalons que la réciproque de la proposition 2.4 est vraie (M. Wodzicki, communication personnelle).

3. Le théorème principal

On va se contenter de traiter le cas suivant du théorème de Suslin et Wodzicki.

3.1. THÉORÈME. — *Pour toute \mathbb{Q} -algèbre (non nécessairement unitaire) I les énoncés suivants sont équivalents.*

- a) I est H -unitaire,
- b) I est excisif pour HC (avec $k = \mathbb{Q}$),
- c) I est excisif pour $K_{\mathbb{Q}}$.

Les implications (a) \Leftrightarrow (b) et (c) \Rightarrow (a) sont déjà dans [W1] (cf. section 2 et 5.5). Le nouveau résultat qu'apporte [S-W] est l'implication a) \Rightarrow c), que nous allons maintenant détailler.

Nous avons d'ores et déjà traduit la propriété de K -excision (resp. HC -excision) en termes d'homologie de GL (resp. gl), cf. 1.4 (resp. cf. 2.4). Il nous faut donc disposer d'un moyen de comparer ces deux homologies. L'idée est d'utiliser le "modèle de Volodin" pour décrire l'homologie de GL . Ce modèle utilise un recouvrement de $GL(I)$ par des sous-groupes $T_n^\sigma(I)$ de matrices triangulaires. Plus précisément on va construire un espace $\underline{T}(I)$ (resp. $\tilde{\underline{T}}(I)$) à partir des classifiants $BT_n^\sigma(I)$ (resp. $B\tilde{T}_n^\sigma(I)$), puis une application continue $\underline{T}(I) \rightarrow \tilde{\underline{T}}(I)$.

De manière analogue on va associer, dans le cadre des algèbres de Lie, à $gl(I)$ (resp. $\tilde{gl}(I)$) un complexe de chaînes $\underline{t}(I)$ (resp. $\tilde{\underline{t}}(I)$) construit à partir des algèbres de Lie $t_n^\sigma(I)$ correspondant aux $T_n^\sigma(I)$. Puis on construit un morphisme de complexes $\underline{t}(I) \rightarrow \tilde{\underline{t}}(I)$.

Le point-clé, utilisé pour la première fois dans un contexte analogue par T. Goodwillie [G], est que, $T_n^\sigma(I)$ étant un groupe nilpotent, on sait comparer son homologie rationnelle à celle de l'algèbre de Lie $t_n^\sigma(I)$. Ceci permet de comparer $H_*(\underline{T}(I))$ à $H_*(\underline{t}(I))$ et $H_*(\tilde{\underline{T}}(I))$ à $H_*(\tilde{\underline{t}}(I))$.

En résumé le plan de la démonstration de l'implication $a) \Rightarrow c)$ du théorème 3.1 est :

I est H – unitaire

$$\Updownarrow (0)$$

I est HC –excisif $\stackrel{(1)}{\Leftrightarrow} H_*(gl(I)) \cong H_*(\widetilde{gl}(I)) \stackrel{(2)}{\Leftrightarrow} H_*(\underline{t}(I)) \cong H_*(\widetilde{t}(I))$

$$\Updownarrow (3)$$

I est $K_{\mathbf{Q}}$ –excisif $\stackrel{(1')}{\Leftarrow} H_*(GL(I)) \cong H_*(\widetilde{GL}(I)) \stackrel{(2')}{\Leftrightarrow} H_*(\underline{T}(I)) \cong H_*(\widetilde{T}(I))$.

Les étapes (0) et (1) ont été traitées en section 2. L'étape (1') a été traitée en section 1. Voici quelques détails sur les étapes (2), (2') et (3).

4. Constructions de Volodin

Soit I un pseudo-anneau. Pour tout ordre partiel σ de l'ensemble $\{1, 2, \dots, n\}$ on note $T_n^\sigma(I)$ le sous-groupe de matrices σ -triangulaires de $GL_n(I)$:

$$T_n^\sigma(I) = \{(a_{ij}) \in GL_n(I) \mid a_{ij} = 0 \text{ si } i \not\prec_\sigma j \text{ et } a_{ii} = 1 ; 1 \leq i \neq j \leq n\}.$$

En particulier si σ est l'ordre usuel $1 < 2 < \dots < n$, alors $T_n^\sigma(I)$ est simplement le sous-groupe des matrices triangulaires supérieures au sens habituel.

L'espace classifiant $BT_n^\sigma(I)$ est un sous-espace de $BGL_n(I)$ et l'espace $\underline{T}(I) := \cup_{n,\sigma} BT_n^\sigma(I)$ est un sous-espace de $BGL(I)$ (note : ce n'est *pas* une réunion disjointe).

Le lien entre l'homologie de $\underline{T}(I)$ et l'homologie de $BGL(I)$ s'obtient sous la forme d'une suite spectrale de la manière suivante.

Pour tout groupe discret G l'espace classifiant BG admet un revêtement universel EG qui est contractile. Si $\{G_j\}_{j \in J}$ est une famille de sous-groupes de G , il existe un sous-espace connexe $V(G, \{G_j\})$ de EG au-dessus de $\bigcup_j BG_j$ sur lequel G opère et pour lequel on a une suite spectrale

$$(4.1) \quad E_{pq}^2 = H_p(BG; H_q(V(G, \{G_j\}))) \Rightarrow H_{p+q}(\bigcup_j BG_j).$$

Cet espace, appelé *espace de Volodin*, possède en outre la propriété suivante. Pour toute autre donnée $(G', \{G'_j\}_{j \in J})$ et tout homomorphisme $\varphi : G \rightarrow G'$ tels que $\varphi(G_j) \subset G'_j$ et $\varphi : G/G_{i_1} \cap \cdots \cap G_{i_p} \xrightarrow{\sim} G'/G'_{i_1} \cap \cdots \cap G'_{i_p}$ est un isomorphisme pour tout p -uplet d'indices,

(4.2) l'application induite $V(G, \{G_j\}) \rightarrow V(G', \{G'_j\})$ est une équivalence d'homotopie.

Dans la pratique on fait ces constructions dans le cadre simplicial.

4.3. PROPOSITION. — *Pour tout pseudo-anneau I tel que $I = I^2$ les conditions suivantes sont équivalentes :*

- (a) $H_*(GL(I)) \cong H_*(\widetilde{GL}(I))$,
- (b) $H_*(\underline{T}(I)) \cong H_*(\widetilde{\underline{T}}(I))$.

Esquisse de démonstration. La propriété (4.2) implique que les deux espaces de Volodin $V(I)$ et $\widetilde{V}(I)$ sont homotopiquement équivalents. Puis on montre que $GL(I^2)$ (resp. $\widetilde{GL}(I^2)$) opère trivialement sur $H_*(V(I))$ (resp. $H_*(\widetilde{V}(I))$). La proposition est alors une conséquence de la comparaison des suites spectrales (4.1) pour $\underline{T}(I) = \bigcup_{n,\sigma} BT_n^\sigma(I)$ et $\widetilde{\underline{T}}(I) = \bigcup_{n,\sigma} B\widetilde{T}_n^\sigma(I)$. \square

Pour l'étape (2) on répète le raisonnement ci-dessus, mais dans le cadre des algèbres de Lie. La différence technique est que les espaces classifiants

sont remplacés par des complexes de chaînes. Explicitement l'objet de base est le complexe de Chevalley-Eilenberg

$$C_*(\mathfrak{g}, M) \quad \dots \rightarrow M \otimes \Lambda^n \mathfrak{g} \xrightarrow{d} M \otimes \Lambda^{n-1} \mathfrak{g} \rightarrow \dots \rightarrow M,$$

où \mathfrak{g} est une algèbre de Lie sur un anneau commutatif k , $\Lambda^n \mathfrak{g}$ sa n -ième puissance extérieure, M un \mathfrak{g} -module, et d est donné par

$$d(g_0, g_1, \dots, g_n) = \sum_{0 \leq i < j \leq n} (-1)^j (g_0, \dots, g_{i-1}, [g_i, g_j], g_{i+1}, \dots, \hat{g}_j, \dots, g_n),$$

pour $g_0 \in M, g_1, \dots, g_n \in \mathfrak{g}$.

Par définition l'homologie de \mathfrak{g} à coefficients dans M , notée $H_n(\mathfrak{g}, M)$, est l'homologie de $C_*(\mathfrak{g}, M)$.

Pour toute famille $\{\mathfrak{g}_j\}_{j \in J}$ de sous-algèbres de Lie de \mathfrak{g} , l'espace de Volodin est remplacé par un certain *complexe de Volodin* $v_*(\mathfrak{g}, \{\mathfrak{g}_j\}) \subset C_*(\mathfrak{g}, U(\mathfrak{g}))$ ($U(\mathfrak{g})$ est l'algèbre enveloppante de \mathfrak{g}).

L'analogue de l'espace $\underline{T}(I)$ est le complexe $\underline{t}_*(I)$ construit de la manière suivante. Soit $t_n^\sigma(I)$ l'algèbre de Lie des matrices σ -triangulaires de $gl_n(I)$. Le complexe $C_*(t_n^\sigma(I))$ est un sous-complexe de $C_*(gl_n(I))$. Par définition $\underline{t}_*(I)$ est le sous-complexe de $C_*(gl(I))$ suivant

$$\underline{t}_*(I) := \sum_{n, \sigma} C_*(t_n^\sigma(I)).$$

Les propriétés du complexe de Volodin, en tout point analogue à celle de l'espace de Volodin, permettent de démontrer la

4.4. PROPOSITION. — *Pour toute k -algèbre I telle que $I^2 = I$ les conditions suivantes sont équivalentes*

- (a) $H_*(gl(I)) \cong H_*(\tilde{gl}(I))$,
- (b) $H_*(\underline{t}(I)) \cong H_*(\tilde{\underline{t}}(I))$. \square

4.5. *Remarque.* En fait les complexes $\underline{t}(I)$ et $\tilde{t}(I)$ sont acycliques lorsque I est H -unitaire (cf. section 9 de [S-W]). Dans une récente prépublication [Su2] A. Suslin donne un argument direct pour démontrer cette acyclicité rendant caduque les étapes (0), (1) et (2). Néanmoins celles-ci sont très intéressantes en vue d'un calcul complet de l'obstruction à l'excision (cf. section 6).

5. Comparaison homologie du groupe linéaire et de son algèbre de Lie

La théorie de Malcev met en bijection les algèbres de Lie nilpotentes \mathfrak{n} sur \mathbb{Q} et les groupes nilpotents N uniquement divisibles. Par cette bijection $T_n^\sigma(I)$ correspond à $t_n^\sigma(I)$. Le point principal de cette dernière étape est le résultat suivant.

5.1. THÉORÈME (cf. [P]). — *Pour toute algèbre de Lie nilpotente \mathfrak{n} sur \mathbb{Q} , il y a un isomorphisme canonique (de coalgèbres graduées) :*

$$H_*(\mathfrak{n}, \mathbb{Q}) \cong H_*(N, \mathbb{Q}).$$

A gauche c'est l'homologie de l'algèbre de Lie, à droite celle du groupe discret.

Esquisse de preuve. L'algèbre enveloppante $U(\mathfrak{n})$ de \mathfrak{n} admet un complété pour son idéal d'augmentation, que l'on note $U(\mathfrak{n})^\wedge$. De même l'algèbre $\mathbb{Q}[N]$ du groupe discret N admet un complété noté $\mathbb{Q}[N]^\wedge$. On a alors les propriétés suivantes, dues à la condition de nilpotence (cf. [P]) :

(a) L'inclusion $N \subset (U(\mathfrak{n})^\wedge)^\times$ induit un isomorphisme d'algèbres de Hopf $\mathbb{Q}[N]^\wedge \cong U(\mathfrak{n})^\wedge$,

(b) $\mathbb{Q}[N]^\wedge$ est plat sur $\mathbb{Q}[N]$,

(c) $U(\mathfrak{n})^\wedge$ est plat sur $U(\mathfrak{n})$.

Ces deux dernières propriétés sont dues à la propriété d'Artin-Rees.

L'isomorphisme de 5.1 peut se construire en comparant abstraitement les résolutions donnant respectivement le complexe de Chevalley-Eilenberg (pour \mathfrak{n}) et le complexe de Mac Lane (pour N). Mais par

cette méthode la naturalité est délicate à prouver (cf. [S-W,§5]). Une autre méthode consiste à comparer ces deux groupes d'homologie avec $\text{Tor}^{U(n)^\wedge}(\mathbb{Q}, \mathbb{Q}) = \text{Tor}^{\mathbb{Q}[N]^\wedge}(\mathbb{Q}, \mathbb{Q})$ comme dans [L2,XI.3.14]. \square

Un argument classique de suites exactes de Mayer-Vietoris permet de prouver le

5.2. THÉORÈME. — *Pour tout pseudo-anneau I on a un isomorphisme*

$$H_*(\underline{T}(I)) := H_*\left(\bigcup_{n,\sigma} BT_n^\sigma(I)\right) \cong H_*\left(\sum_{n,\sigma} C_*(t_n^\sigma(I))\right) =: H_*(\underline{t}(I))$$

et pareillement pour $\tilde{\underline{T}}(I)$ et $\tilde{\underline{t}}(I)$. \square

On en déduit alors le dernier maillon de la preuve (3) :

5.3. COROLLAIRE. — *Pour toute \mathbb{Q} -algèbre I les énoncés suivants sont équivalents*

- (a) $H_*(\underline{T}(I)) \cong H_*(\tilde{\underline{T}}(I))$,
- (b) $H_*(\underline{t}(I)) \cong H_*(\tilde{\underline{t}}(I))$. \square

5.4. *Remarque.* Le théorème 1 (I pseudo-anneau, $I_{\mathbb{Q}}$ H -unitaire) résulte essentiellement du théorème 3.1. (valable pour $I_{\mathbb{Q}}$) et du fait que $H_*(T_n^\sigma(I); \mathbb{Q})$ est isomorphe à $H_*(T_n^\sigma(I_{\mathbb{Q}}); \mathbb{Q})$ (et idem avec \tilde{T} à la place de T).

5.5. *Commentaire sur la preuve de "I $K_{\mathbb{Q}}$ -excisif \Rightarrow I H -unitaire".*

Ce résultat avait été précédemment obtenu par M. Wodzicki [W1] en utilisant le même type de technique. Le principe est de choisir, à partir de I , une situation d'excision qui se ramène à calculer un groupe de K -théorie relatif $K_*(R, J)_{\mathbb{Q}}$ où J est un idéal nilpotent. Le point crucial est le théorème de Goodwillie [G] affirmant qu'on a alors un isomorphisme

$$(5.6) \quad K_*(R, J)_{\mathbb{Q}} \cong HC_{*-1}(R, J)_{\mathbb{Q}}.$$

L'hypothèse de K -excision (nullité d'un certain groupe de K -théorie relative) implique alors la nullité d'un certain groupe d'homologie cyclique

relative. L'exemple (R, J) est choisi de telle manière que le groupe d'homologie cyclique relative contienne $H_*(B_*(I), b')$ en facteur direct.

Signalons que les idées exposées ci-dessus concernant les espaces et complexes de Volodin sont issues de la preuve de (5.6) par T. Goodwillie [G] et de l'article [Su1] de A. Suslin.

6. Obstruction à l'excision en K -théorie algébrique

Soit I un idéal bilatère de l'anneau A . Soit $\varphi : A \rightarrow B$ un homomorphisme d'anneau tel que φ restreint à I soit injectif et $\varphi(I)$, que l'on notera encore I , soit un idéal bilatère de B . Le diagramme commutatif

$$(6.0) \quad \begin{array}{ccc} A & \xrightarrow{\varphi} & B \\ \downarrow & & \downarrow \\ A/I & \xrightarrow{\varphi'} & B/I \end{array}$$

est cartésien. Si I est K -excisif on a un isomorphisme $K_n(A, I) \cong K_n(B, I)$, dont on déduit l'exactitude de la suite de Mayer-Vietoris

$$(6.1) \quad \begin{aligned} \cdots \rightarrow K_n(A) \rightarrow K_n(B) \oplus K_n(A/I) \rightarrow K_n(B/I) \\ \rightarrow K_{n-1}(A) \rightarrow \cdots \rightarrow K_0(B/I). \end{aligned}$$

Comme indiqué déjà dans l'introduction $K_0(A, I) \rightarrow K_0(B, I)$ est toujours un isomorphisme. Si, de plus, φ est surjectif, on peut montrer qu'on a aussi un isomorphisme $K_1(A, I) \xrightarrow{\sim} K_1(B, I)$ (cf. [M, lemme 6.3]). Si on note $J = \text{Ker } \varphi$, cette donnée est équivalente à la donnée d'un anneau A et de deux idéaux bilatères I et J vérifiant

$$I \cap J = \{0\}.$$

Il est alors naturel de construire des groupes de K -théorie birelatifs $K_n(A; I, J)$ s'inscrivant dans une suite exacte

$$(6.2) \quad \cdots \rightarrow K_n(A; I, J) \rightarrow K_n(A, I) \rightarrow K_n(A/J, I + J/J) \rightarrow K_{n-1}(A; I, J) \rightarrow \cdots$$

Ces groupes sont l'*obstruction à l'exactitude* de la suite de Mayer-Vietoris.

Les résultats précédents se traduisent par

$$K_0(A; I, J) = K_1(A, I, J) = 0.$$

Le premier exemple pour lequel $K_2(A; I, J) \neq 0$ a été construit par Swan [Sw] : $A = \mathbf{Z}[x, y]/(xy)$, $I = (x)$, $J = (y)$. Dans cette situation $K_2(A; I, J)$ est un facteur direct dans $K_2(A)$. L'élément, dont Swan démontre la non-trivialité, est l'image du générateur standard de $H_2(\mathbf{Z}^2)$ (classe fondamentale du tore) par l'homomorphisme $\rho_* : H_2(\mathbf{Z}^2) \rightarrow H_2(E(A)) \cong K_2(A)$ induit par

$$\rho : \mathbf{Z}^2 \rightarrow E(A), \rho(1, 0) = \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix}, \rho(0, 1) = \begin{pmatrix} 1 & 0 \\ y & 1 \end{pmatrix}.$$

En fait on peut calculer explicitement ce premier groupe d'obstruction.

6.3. THÉORÈME. — [GW-L, K]. *Si $I \cap J = \{0\}$ alors*

$$K_2(A; I, J) \cong I \otimes_{A^e} J \text{ où } A^e = A \otimes A^{op}.$$

On remarquera que ce théorème ne comporte pas d'hypothèse de rationalité.

Si l'on remplace la K -théorie par l'homologie cyclique (se rappeler qu'il y a un décalage de degré dans les notations), on prouve aisément que

$$HC_0(A; I, J) = 0 \text{ et } HC_1(A; I, J) = I \otimes_{A^e} J.$$

Ces calculs, joints à l'isomorphisme de Goodwillie (cf. 5.6) comparant K -théorie et homologie cyclique pour un idéal nilpotent et à la preuve du théorème de Suslin-Wodzicki, suggèrent la

6.4. *Conjecture.* Si $I \cap J = \{0\}$, alors

$$K_n(A; I, J)_{\mathbf{Q}} \cong HC_{n-1}(A; I, J)_{\mathbf{Q}}.$$

(En fait, s'il en est ainsi, l'hypothèse $I \cap J$ nilpotent est suffisante). Bien que plusieurs fois annoncée cette conjecture n'a pas encore été démontrée.

On trouvera dans [G-W] des résultats sur l'obstruction à l'excision en basses dimensions dans le cas où φ (cf. 6.0) n'est pas surjectif.

Bibliographie

[B] BASS, H., *Algebraic K-theory*, Benjamin, 1968.

[C] CONNES, A., *Non-commutative differential geometry*, Publ. Math. IHES 62 (1985), 257-360.

[G-W] GELLER, S. and WEIBEL, C., $K(A, B, I) : II$, J. *K-theory* 2 (1989), 753-760.

[G] GOODWILLIE, T.G., *Relative algebraic K-theory and cyclic homology*, Ann. of Math. 124 (1986), 347-402.

[GW-L] GUIN-WALÉRY, D. et LODAY, J.-L., *Obstruction à l'excision en K-théorie algébrique*, Springer Lect. Notes in Math. 854 (1981), 179-216.

[H] HANLON, P., *On the complete $GL(n, C)$ -decomposition of the stable cohomology of $gl_n(A)$* , Trans. AMS 308 (1988), 209-225.

[K] KEUNE, F., *The relativization of K_2* , J. Algebra 54 (1978), 159-177.

[L1] LODAY, J.-L., *K-théorie algébrique et représentations de groupes*, Ann. Sc. Ec. Norm. Sup. 9 (1976), 309-377.

[L2] LODAY, J.-L., *Cyclic Homology*, Grund. Math. Wiss. 301, 1992, Springer-Verlag.

[L-Q] LODAY, J.-L. and QUILLEN, D., *Cyclic homology and the Lie algebra homology of matrices*, Comment. Math. Helvetici 59 (1984), 565-591.

[M] MILNOR, J., *Introduction to algebraic K-theory*, Ann. Math. Studies 72, Princeton U. Press. 1974.

[P] PICKEL, P.F., *Rational cohomology of nilpotent groups and Lie algebras*, Comm. in Alg. 6(1978), 409-419.

[Q] QUILLEN, D., *Cohomology of groups*, Actes Congrès International des Mathématiciens, t. II, 1970, 47-51.

[Su1] SUSLIN, A., *On the equivalence of K-theories*, Comm. in alg. 9 (1981), 1559-1566.

[Su2] SUSLIN, A., *On the acyclicity of the sum of triangular complexes*, preprint, University of Utrecht, (1991), 6 p.

[S-W] SUSLIN, A. and WODZICKI, M., *Excision in algebraic K-theory*, Ann. of Maths. 136 (1992), 51-122.

[Sw] SWAN, R.G., *Excision in algebraic K-theory*, J. Pure Applied Algebra 1 (1971), 221-252.

[T] TSYGAN, B.L., *The homology of matrix Lie algebras over rings and the Hochschild homology (en russe)*, Uspekhi Mat. Nauk 38 (1983), 217-218 – Russ. Math. Survey 38 n°2 (1983), 198-199.

[W1] WODZICKI, M., *Excision in cyclic homology and in rational algebraic K-theory*, Ann. Math. 129 (1989), 591-639.

[W2] WODZICKI, M., *Homological properties of rings of functional analytic type*, Proc. Nat. Acad. Sci. USA 87 (1990), 4910-4911.

[W3] WODZICKI, M., *The long exact sequence in cyclic homology associated with an extension of algebras*, Comptes Rendus Acad. Sci. Paris Sér. A-B 307 (1988), 249-254.

Jean-Louis LODAY
Institut de Recherche Mathématique Avancée
Université Louis Pasteur et C.N.R.S.
7, rue René-Descartes
67084 Strasbourg Cedex

Astérisque

RICARDO PÉREZ-MARCO

**Solution complète au problème de Siegel de linéarisation
d'une application holomorphe au voisinage d'un point fixe**

Astérisque, tome 206 (1992), Séminaire Bourbaki,
exp. n° 753, p. 273-310

http://www.numdam.org/item?id=SB_1991-1992__34__273_0

© Société mathématique de France, 1992, tous droits réservés.

L'accès aux archives de la collection « Astérisque » (<http://smf4.emath.fr/Publications/Asterisque/>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

**SOLUTION COMPLÈTE AU PROBLÈME DE SIEGEL
DE LINÉARISATION D'UNE APPLICATION HOLOMORPHE
AU VOISINAGE D'UN POINT FIXE
(D'après J.-C. YOCCOZ)**

par Ricardo PÉREZ-MARCO

On veut présenter ici les nouvelles techniques géométriques dans les problèmes de petits diviseurs en dynamique holomorphe inventées par J.-C. Yoccoz ([Y1]) pour résoudre le problème de Siegel. Il a aussi appliqué avec succès ces techniques révolutionnaires pour découvrir les conditions diophantiennes optimales dans les problèmes de linéarisation des difféomorphismes analytiques du cercle ([Y2]).

Le sujet étant abordable sans aucune difficulté à tous les non spécialistes, on a cru bon de faire pour eux (avec l'indulgence des autres) une introduction élémentaire et un petit aperçu historique. Pour une introduction plus exhaustive et une bibliographie complète on se référera au texte de M. Herman [He1].

Je remercie P. Sentenac pour son aide précieuse dans la préparation de ce texte.

1. Dynamique holomorphe au voisinage d'un point fixe.

Considérons un système dynamique, i.e. une application d'un espace dans lui même $f : X \rightarrow X$. L'étude de la dynamique de f consiste, sous sa forme la plus naïve, dans la compréhension des itérés de tout point. Une étude plus poussée concerne la classification des classes de conjugaison de f par les morphismes de X . Une première démarche serait la compréhension

de la structure des points prépériodiques (ceux qui ont une orbite finie) et la dynamique au voisinage des orbites périodiques. Or une orbite périodique n'est rien d'autre qu'un point fixe d'un certain itéré. Ceci nous mène d'une façon naturelle à l'étude de la dynamique locale en un point fixe.

En dynamique holomorphe l'espace X est muni d'une structure complexe qui est préservée par f , i.e. f est holomorphe. Ce premier programme d'étude n'est même pas achevé dans le cas où $\dim_{\mathbf{C}} X = 1$. Essentiellement on se limitera dans la suite à ce cas là. On placera notre point fixe à l'origine de \mathbf{C} et f sera définie dans son voisinage. Ainsi, on écrira $f(z) = \lambda z + \sum_{n \geq 2} f_n z^n$.

1.1. Notions préliminaires.

La notion dynamique la plus importante au voisinage d'un point fixe est certainement la stabilité :

Définition (Stabilité). *L'application f est stable en 0 si pour tout voisinage V de 0 il existe un voisinage U de 0 tel que pour $n \geq 0$, f^n est défini dans U et $f^n(U) \subset V$.*

Clairement ceci est équivalent à l'existence d'un voisinage de 0 où les itérés positifs de f sont bien définis et forment une famille normale.

Lorsque la partie linéaire de f en 0 est non nulle ($\lambda \neq 0$), f est très bien approché au voisinage de 0 par $z \mapsto \lambda z$. Il est naturel de se demander si la dynamique est la même. Lorsque ceci se produit on dit que f est linéarisable.

Définition (Linéarisabilité). *L'application f est linéarisable en 0 si $\lambda \neq 0$ et s'il existe un changement de variables holomorphe $z \mapsto h(z) = z + \sum_{n \geq 2} h_n z^n$ qui conjugue f à sa partie linéaire*

$$(h^{-1} \circ f \circ h)(z) = \lambda z.$$

Lorsque λ n'est pas racine de l'unité, l'application h est formellement uniquement déterminée par l'équation fonctionnelle. Dans ce cas, on parlera de la linéarisante de f . On notera $R(f) \geq 0$ son rayon de convergence, avec $R(f) = 0$ dans le cas non linéarisable, et $h(\mathbf{D}_{R(f)})$ sera le domaine de Siegel ou de linéarisation de f (on note $\mathbf{D}_r = \{z \in \mathbf{C}; |z| < r\}$).

Un fait remarquable dans le cadre holomorphe est que les notions de stabilité pour f ou f^{-1} et de linéarisabilité sont équivalentes.

Proposition. *Pour $\lambda \neq 0$ et $|\lambda| \leq 1$, f est linéarisable si et seulement si f est stable en 0.*

Démonstration. Si f est linéarisable les itérés positifs de f forment une famille normale dans le domaine de linéarisation de f . Inversement, soit U un voisinage de 0 où les itérés $(f^n)_{n \geq 0}$ forment une famille normale. Soit

$$k_n = \frac{1}{n} \sum_{i=0}^{n-1} \lambda^{-i} f^i.$$

On a

$$k_n \circ f = \lambda k_n + \frac{\lambda}{n} (\lambda^{-n} f^n - \text{id}).$$

La famille $(k_n)_{n \geq 0}$ est normale et on peut extraire une sous-suite convergente vers k et $h = k^{-1}$ linéarise f . \diamond

1.2. Point fixe super-attractif.

On désigne ainsi un point fixe qui est aussi un point critique, c'est à dire on a $\lambda = 0$. La dynamique est simple : tous les points dans un voisinage de 0 sont attirés vers l'origine. En fait on démontre facilement :

Théorème (L. E. Böttcher [Bot]). *Soit $f(z) = f_n z^n + \mathcal{O}(z^{n+1})$ avec $n \geq 2$ et $f_n \neq 0$. Il existe un changement de variables holomorphe $z \mapsto h(z)$, $h(0) = 0$, qui conjugue f à $z \mapsto z^n$. i.e.*

$$(h^{-1} \circ f \circ h)(z) = z^n.$$

1.3. Point fixe attractif ou répulsif.

Ce cas correspond à $0 < |\lambda| < 1$ (attractif) ou $1 < |\lambda|$ (répulsif), et on a toujours linéarisabilité (appliquer le critère de stabilité à f ou à f^{-1}).

Théorème (G. Koenigs [Ko]). *Soit $f(z) = \lambda z + \mathcal{O}(z^2)$ avec $|\lambda| \neq 0, 1$, alors f est linéarisable.*

1.4. Point fixe indifférent rationnel.

Ce cas correspond à $|\lambda| = 1$ et λ racine de l'unité, $\lambda = e^{2\pi i\alpha}$ avec $\alpha = p/q \in \mathbf{Q}$, $p \in \mathbf{Z}$, $q \geq 1$ et $p \wedge q = 1$. L'étude de la dynamique est plus subtile et la linéarisabilité est plutôt rare comme le montre la proposition suivante.

Proposition. *L'application $f(z) = e^{2\pi ip/q}z + \mathcal{O}(z^2)$ est linéarisable si et seulement si $f^q = \text{id}$.*

Démonstration. Si $f^q = \text{id}$ alors $h = \left(\frac{1}{q} \sum_{i=0}^{q-1} \lambda^{-i} f^i \right)^{-1}$ linéarise f . Inversement, si $h^{-1} \circ f \circ h(z) = \lambda z$ alors $h^{-1} \circ f^q \circ h(z) = \lambda^q z = z$ donc $f^q = h \circ h^{-1} = \text{id}$. \diamond

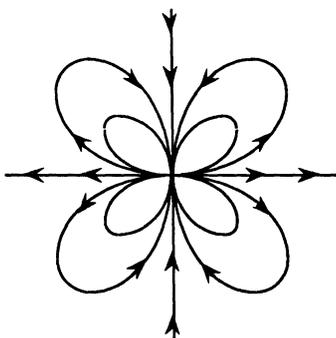


Figure 0

Par exemple, le polynôme quadratique $P_\lambda(z) = \lambda z(1 - z)$ sera non linéarisable car P_λ^q est un polynôme de degré 2^q . Néanmoins la dynamique topologique de f est complètement comprise depuis les travaux de Fatou et L. Leau ([Le]). Pour la classification des classes de conjugaison topologiques on se référera au travail de C. Camacho [Ca]. Les points, tout en tournant (f est proche d'une rotation rationnelle), décrivent une "fleur" comme celle de la figure. Il existe un entier $n \geq 1$ et $2nq$ domaines paraboliques (pétales) qui sont permutés par f avec un nombre de rotation p/q en formant $2n$ cycles de q éléments. Sur la figure on a dessiné la dynamique de f^q .

D'après les travaux de J. Ecalle et S. M. Voronin (voir [Ec], [Vo] et

[Ma]) on a une classification complète des classes de conjugaison analytiques qui, contrairement aux cas déjà vus, est plus riche que la classification topologique.

1.5. Point fixe indifférent irrationnel. Le problème de Siegel.

C'est le cas qui reste à traiter, où $\lambda = e^{2\pi i\alpha}$ avec $\alpha \in \mathbf{R} - \mathbf{Q}$. L'application linéarisante existe toujours formellement mais cette série peut diverger à cause des problèmes de petits diviseurs causés par les bonnes approximations rationnelles de α . En effet, on peut résoudre formellement $h^{-1} \circ f \circ h(z) = \lambda z$,

et on trouve, avec $h(z) = \sum_{n=1}^{+\infty} h_n z^n$, $h_1 = 1$,

$$h_n = \frac{1}{\lambda^n - \lambda} \left[f_n + \sum_{p=2}^{n-1} f_p \sum_{\substack{j_1 + \dots + j_p = n \\ j_i \geq 1}} h_{j_1} \dots h_{j_p} \right].$$

On soupçonne la possibilité d'instabilité de f pour des nombres α très "Liouvillesques" lorsque l'on considère f comme limite d'applications $f_n(z) = e^{2\pi i p_n/q_n} z + \mathcal{O}(z^2)$, où $(p_n/q_n)_{n \geq 0}$ sont de bonnes approximations de α . On a vu que "génériquement" un point fixe indifférent rationnel est instable. Le problème de Siegel, dit aussi problème du centre ou de l'équation fonctionnelle de Schröder, consiste à étudier l'ensemble des valeurs de α pour lesquelles on a toujours linéarisabilité.

Comme on le verra, le polynôme quadratique $P_\lambda(z) = \lambda z(1-z)$ (unique à conjugaison linéaire près) joue un rôle important dans le problème de Siegel : sa dynamique reflète très fidèlement la dynamique d'un germe générique à multiplicateur λ fixé.

2. Résultats classiques sur le problème de Siegel.

Bien que l'existence d'applications non linéarisables semble être connue des mathématiciens du siècle dernier, la première trace écrite semble être due à G. A. Pfeiffer en 1917 ([Pf]). D'ailleurs, il est facile de montrer que génériquement le polynôme quadratique est non linéarisable, i.e. il existe un G_δ dense de valeurs de α pour lesquelles la linéarisante diverge (cf. 5.1).

Le premier à avoir mis l'accent sur l'importance de l'arithmétique de α a été H. Cremer (voir [Cr]). Si $(p_n/q_n)_{n \geq 0}$ désigne la suite des réduites de α (cette suite est définie dans l'annexe consacrée à l'arithmétique), il montre :

Théorème (H. Cremer [Cr]). *Si $\alpha \in \mathbf{R} - \mathbf{Q}$ satisfait la condition de Cremer,*

$$(C) \quad \sup_{n \geq 0} \frac{\log q_{n+1}}{q_n} = +\infty,$$

alors il existe $f(z) = e^{2\pi i \alpha} z + \mathcal{O}(z^2)$ non linéarisable.

Sa démonstration, élémentaire, consiste à construire f à partir de ses coefficients. On choisit $|f_n| = 1$ et par récurrence l'argument de f_n égal à celui du deuxième terme de la somme entre crochets de la relation de récurrence donnant la suite $(h_n)_{n \geq 2}$ (voir 1.5). On minore ainsi

$$|h_n| \geq \frac{1}{|\lambda^n - \lambda|}.$$

Même dans le cas où la condition \mathcal{C} n'est pas vérifiée, ceci donne une majoration du rayon de convergence de h . Il montre aussi, par un argument simple et joli :

Théorème (H. Cremer [Cr]). *Si $d \geq 2$ et $\alpha \in \mathbf{R} - \mathbf{Q}$ satisfait à la condition de Cremer de degré d ,*

$$(C_d) \quad \sup_{n \geq 0} \frac{\log q_{n+1}}{d^{q_n}} = +\infty,$$

alors tout germe polynômial $P(z) = e^{2\pi i\alpha}z + \dots + a_d z^d$, $a_d \neq 0$, est non linéarisable.

Noter que $\mathcal{C}_\infty = \bigcap_{d \geq 2} \mathcal{C}_d$, qui est la condition qui permet d'appliquer le théorème à tout polynôme indépendamment du degré, s'écrit

$$(C_\infty) \quad \sup_{n \geq 1} \frac{\log \log q_{n+1}}{q_n} = +\infty.$$

On a

$$\mathcal{C}_\infty \subset \dots \subset \mathcal{C}_3 \subset \mathcal{C}_2 \subset \mathcal{C} \subset \mathbf{R} - \mathbf{Q}.$$

Le problème de l'existence de valeurs de α pour lesquelles tout germe $f(z) = e^{2\pi i\alpha}z + \mathcal{O}(z^2)$ serait linéarisable est délicat et est resté longtemps ouvert. Même G. Julia ([Ju]) a donné une démonstration incorrecte de leur inexistence. Le problème a été résolu de façon remarquable en 1942 par C.-L. Siegel.

Théorème (C.-L. Siegel [Si], amélioré par A. Bruno [Br]). *Si $\alpha \in \mathbf{R} - \mathbf{Q}$ satisfait la condition arithmétique,*

$$(B) \quad \sum_{n \geq 0} \frac{\log q_{n+1}}{q_n} < +\infty,$$

alors toute application $f(z) = e^{2\pi i\alpha}z + \mathcal{O}(z^2)$ est linéarisable.

L'ensemble des nombres qui satisfont la condition arithmétique du théorème de Siegel de 1942 est de mesure de Lebesgue totale. Ce résultat est le premier à remonter les difficultés liées aux petits diviseurs. La démonstration repose sur des estimations des coefficients de la linéarisante directement à partir de la relation de récurrence de 1.5. On minore les petits diviseurs grâce à la condition diophantienne. Plus tard, dans les années 50 et 60, les nouvelles techniques, qui donnèrent les théorèmes K.A.M, ont été découvertes. Le lecteur pourra se référer au séminaire Bourbaki de J.-B. Bost [Bo] sur ce sujet.

Ces techniques s'appliquent aussi au problème de Siegel mais donnent des résultats plus grossiers quand aux estimations des domaines de linéarisation, et les conditions arithmétiques sont plus restrictives.

A propos des conditions diophantiennes, Siegel démontra la linéarisation pour des $\alpha \in \mathbf{R} - \mathbf{Q}$ satisfaisant à une condition diophantienne, i.e. il existe $\gamma, \tau > 0$ tels que pour tout rationnel p/q on a

$$|\alpha - p/q| \geq \frac{\gamma}{q^\tau}.$$

En 1965, A. Bruno [Br] améliora cette condition arithmétique en raffinant la démonstration de Siegel. Simultanément et de façon indépendante, T. M. Cherry [Ch] conjectura ce résultat et l'optimalité de la condition \mathcal{B} . Il n'a pas pu démontrer sa conjecture avant sa mort mais son travail [Ch], bien que très vague, montre une remarquable compréhension de la dynamique non linéarisable.

Ce n'est qu'en 1987 que J.-C. Yoccoz dans son mémoire [Y1] a montré cette conjecture.

3. Nouveaux résultats.

3.1. Sur le problème de Siegel.

3.1.a. Préliminaire.

La partie I.3 de [Y1] contient une nouvelle démonstration du théorème de Siegel et de Bruno. Cette nouvelle démonstration géométrique donne des estimations optimales, à une constante multiplicative près, du rayon de convergence des applications linéarisantes. Le problème étant invariant par conjugaison par des homothéties, il faut, pour énoncer les résultats, normaliser les applications.

Ainsi, quitte à conjuguer f par une homothétie, on supposera $f \in S_\alpha$, où S_α est l'ensemble des applications univalentes (holomorphes et injectives) sur le disque unité $\mathbf{D} = \mathbf{D}_1$ et telles que $f(z) = e^{2\pi i \alpha} z + \mathcal{O}(z^2)$. Il est bien connu que l'espace S_α est compact pour la topologie de la convergence uniforme sur les compacts. On avait noté $R(f) \geq 0$ le rayon de convergence de l'application linéarisante, on posera

$$R(\alpha) = \inf_{f \in S_\alpha} R(f).$$

Lorsque $\alpha \in \mathcal{B}$ on notera

$$B(\alpha) = \sum_{n=0}^{+\infty} \frac{\log q_{n+1}}{q_n}.$$

Si $\alpha \notin \mathcal{B}$, $B(\alpha) = +\infty$. La différence entre la fonction B et la fonction Φ définie dans l'annexe arithmétique est universellement bornée aux points où elles restent finies.

3.1.b. Les théorèmes pour les germes.

Le théorème principal de [Y1] est :

Théorème.

1. Si $\alpha \in \mathbf{R} - \mathbf{Q}$ est telle que

$$\sum_{n=0}^{+\infty} \frac{\log q_{n+1}}{q_n} = +\infty,$$

alors il existe $f(z) = e^{2\pi i \alpha} z + \mathcal{O}(z^2)$ non linéarisable.

2. Il existe une constante universelle $C > 0$ telle que si $B(\alpha) < +\infty$ alors $R(\alpha) > 0$ et

$$|B(\alpha) - \log(R(\alpha)^{-1})| \leq C.$$

La partie 1 démontre l'optimalité de la condition \mathcal{B} et la partie 2 améliore le résultat de Siegel et Bruno qui ne donne que

$$\log(R(\alpha)^{-1}) \leq C_1 B(\alpha) + C_2,$$

où C_1 et C_2 sont des constantes universelles et on peut prendre $C_1 = 2$ (d'après [He1]).

Comme on le verra, les applications non linéarisables de la première partie du théorème sont construites en "renversant" la démonstration géométrique

du théorème de Siegel. Ceci permet d'étudier certains points de la dynamique : ces germes non linéarisables ont une suite d'orbites périodiques qui tend vers l'origine et il est démontré que les points de ces orbites se trouvent à une distance essentiellement optimale de 0 (i.e. la plus petite possible compte tenu de la période). La construction est d'une grande flexibilité. Elle permet de fixer avec liberté la dynamique de retour au voisinage de ces orbites périodiques. En particulier, on peut faire que des itérés de f soient l'identité au voisinage des orbites périodiques (mais jamais au voisinage de 0 !). Ceci lui permet de démontrer,

Théorème. *Si $\alpha \notin \mathcal{B}$ il existe un ensemble non dénombrable de classes de conjugaison de germes holomorphes $f(z) = e^{2\pi i\alpha}z + \mathcal{O}(z^2)$ sans aucune fonction entière.*

Une des motivations pour prouver l'optimalité de la condition \mathcal{B} se trouve dans l'appendice 2 de [Y1] où il est montré que la démonstration de Bruno ne peut s'appliquer au polynôme quadratique lorsque $\alpha \notin \mathcal{B}$. Dans l'appendice 1 on trouve un résultat isolé et intéressant qui montre que le théorème de linéarisation de Siegel (qui est valable en dimension supérieure lorsque la différentielle est diagonalisable et les valeurs propres sont distinctes, de module 1, non résonantes et satisfont des conditions diophantiennes) n'est plus vrai si la différentielle n'est pas diagonalisable.

Théorème. *Soient $n \geq 2$ et $A \in GL_n(\mathbb{C})$. On suppose que A possède une valeur propre λ de module 1, et que le sous-espace caractéristique associé à λ diffère du sous-espace propre. Alors il existe des germes de difféomorphismes holomorphes de $(\mathbb{C}^n, 0)$, dont la différentielle en 0 est A , qui ne sont pas linéarisables.*

3.1.c. Les résultats sur le polynôme quadratique.

Avant son travail sur le théorème de Siegel, Yoccoz s'était intéressé au cas du polynôme quadratique. Déjà les résultats de H. Cremer montraient que le polynôme quadratique était en un certain sens "moins linéarisable" que les autres germes polynômiaux. Ceci a été précisé par Yoccoz :

Théorème. *Si $\alpha \in \mathbb{R} - \mathbb{Q}$ et le polynôme quadratique $e^{2\pi i\alpha}z(1-z)$ est linéarisable alors tout germe $f(z) = e^{2\pi i\alpha}z + \mathcal{O}(z^2)$ est linéarisable.*

Par ailleurs, il a aussi montré d'une façon élémentaire et indépendante du théorème de Siegel :

Théorème. *Le polynôme quadratique $e^{2\pi i\alpha}z(1-z)$ est linéarisable pour un ensemble de mesure totale de valeurs de α .*

Noter que ces deux derniers théorèmes ont comme corollaire le théorème de Siegel pour un ensemble de valeurs de α de mesure totale. On trouvera leur démonstration dans [Y1] et dans [Do]. La démonstration du premier théorème conjointement avec les résultats pour les germes donne ([Y1] partie II) :

Théorème.

1. *Si $\alpha \notin \mathcal{B}$, le polynôme quadratique P_λ , $\lambda = e^{2\pi i\alpha}$, est non linéarisable.*
2. *Il existe une constante universelle $C_1 > 0$, et pour tout $\varepsilon > 0$, une constante $C(\varepsilon)$, telle que, pour $\alpha \in \mathcal{B}$,*

$$(1 - \varepsilon)B(\alpha) - C(\varepsilon) \leq \log (R(P_\lambda)^{-1}) \leq B(\alpha) + C_1.$$

Ainsi la famille quadratique fournit des germes non linéarisables lorsque $\alpha \notin \mathcal{B}$, et si $\alpha \in \mathcal{B}$ le rayon de convergence est presque (au facteur ε près) le plus petit possible. D'où la question naturelle ([Y1] II.1.9) :

Est-ce que $\alpha \mapsto B(\alpha) - \log (R(P_{e^{2\pi i\alpha}})^{-1})$ est bornée sur \mathcal{B} ?

En 1989, Yoccoz a précisé la partie 1 du théorème.

Théorème. *Si $\alpha \notin \mathcal{B}$ le polynôme quadratique a des orbites périodiques distinctes de 0 contenues dans tout voisinage de 0.*

On pourra trouver une démonstration dans [PM1]. Son argument s'applique aussi aux fractions rationnelles structurellement stables (en fixant le multiplicateur λ en 0).

Lorsque $|\lambda| < 1$, 0 est un point fixe attractif de P_λ qui est linéarisable. L'application linéarisante h_λ a une unique singularité $U(\lambda)$ sur le bord de son disque de convergence qui vérifie $h_\lambda(U(\lambda)) = 1/2$.

Les paragraphes II.2 et II.3 de [Y1] sont consacrés à l'étude de la fonction $\lambda \mapsto U(\lambda)$, holomorphe et bornée sur \mathbf{D} . Il est naturel de penser que les

limites radiales de U lorsque $\lambda \rightarrow \lambda_0 \in S^1 = \partial\mathbf{D}$ donnent des informations sur P_{λ_0} . En fait il montre :

Théorème. *Pour tout $\lambda_0 \in S^1$ le module de la fonction U admet en λ_0 une limite radiale qui est égale au rayon de convergence de h_{λ_0} .*

Ceci démontre la linéarisabilité pour p. t. valeur α en appliquant le théorème de Fatou qui montre que la fonction U , qui est holomorphe et bornée dans \mathbf{D} , admet une limite radiale non nulle pour p. t. valeur du bord. Des questions très intéressantes sont posées dans [Y1] au sujet de la fonction U qui semble posséder des propriétés arithmétiques profondes.

3.2. Sur les difféomorphismes analytiques du cercle.

Le problème de Siegel est à rapprocher du problème de linéarisation des difféomorphismes analytiques du cercle. La relation heuristique est résumée par V. Arnold dans l'introduction de son article [Ar1] :

"... Le problème du centre est un cas singulier d'une application du cercle dont le rayon, dans le cas singulier, est égal à zéro ..."

(Nous décrirons en 6.1 une construction qui rend rigoureuse cette relation heuristique des deux problèmes)

Yoccoz a appliqué ses techniques géométriques aux problèmes de linéarisation des difféomorphismes analytiques du cercle.

3.2.a. Préliminaire.

Soit $f : S^1 \rightarrow S^1$ un homéomorphisme du cercle préservant l'orientation; son principal invariant dynamique est son nombre de rotation, $\rho(f) \in \mathbf{R}/\mathbf{Z}$ défini par H. Poincaré. Lorsque f est un difféomorphisme C^2 , par exemple, et $\rho(f)$ est irrationnel alors f est topologiquement linéarisable par le théorème de Denjoy ([He2]). Il se pose la question de la régularité de la linéarisante. Lorsque f est analytique un théorème de V. Arnold ([Ar1]) (amélioré par H. Rüssmann [Ru]) montre que la linéarisante est analytique pour presque tout nombre de rotation si l'on suppose f proche d'une rotation (résultat local). M. Herman ([He2]) démontra le théorème global (sans supposer de proximité à la rotation) en classe C^∞ et analytique, résolvant ainsi une conjecture de V. Arnold. La condition diophantienne en classe C^∞ fut améliorée par

J.-C. Yoccoz ([Y3]) ce qui donna la condition diophantienne optimale. En analytique la condition arithmétique optimale était inconnue.

Dans la suite on travaillera dans le revêtement universel de S^1 donné par $E : \mathbf{R} \rightarrow S^1$, $E(z) = e^{2\pi iz}$. Pour $\Delta > 0$ on note $S(\alpha, \Delta)$ l'espace des difféomorphismes analytiques de S^1 relevés à \mathbf{R} , de nombre de rotation $\alpha \in \mathbf{R}$, et qui s'étendent en une application univalente sur la bande $\{z \in \mathbf{C}; |\operatorname{Im} z| < \Delta\}$. L'ensemble des relevés des difféomorphismes analytiques du cercle de nombre de rotation α est noté $S(\alpha, 0) = \bigcup_{\Delta > 0} S(\alpha, \Delta)$.

3.2.b. Les théorèmes sur les difféomorphismes du cercle.

On renvoie le lecteur à l'annexe arithmétique pour les définitions des conditions arithmétiques \mathcal{H} , $\mathcal{H}(\Delta)$ et \mathcal{B} . Rappelons simplement qu'elles sont complètement explicites, que $\mathcal{H} \subset \bigcap_{\Delta > 0} \mathcal{H}(\Delta)$ et $\mathcal{B} = \bigcup_{\Delta > 0} \mathcal{H}(\Delta)$. Dans ce qui suit, linéarisabilité signifie linéarisabilité analytique. On retrouve trois types de résultats dans [Y2] :

Théorème global.

1. Si $\alpha \in \mathcal{H}$ alors tout difféomorphisme analytique du cercle de nombre de rotation α est linéarisable.

2. Si $\alpha \notin \mathcal{H}$ il existe des difféomorphismes analytiques du cercle de nombre de rotation α qui ne sont pas linéarisables.

Théorème semi-local. Il existe C_0 ($= 3$) constante universelle telle que :

1. Si $\alpha \in \mathcal{H}(\Delta - C_0)$, $\Delta > C_0$, et $f \in S(\alpha, \Delta)$ alors f est linéarisable.

2. Si $\alpha \notin \mathcal{H}(\Delta - C_0)$ alors il existe $f \in S(\alpha, \Delta)$ non linéarisable.

Théorème local.

1. Si $\alpha \in \mathcal{B}$, $\Delta > \frac{1}{2\pi}\Phi(\alpha) + 3.15$ et $f \in S(\alpha, \Delta)$ alors f est linéarisable et la linéarisante est dans $S(1, \Delta - \frac{1}{2\pi}\Phi(\alpha) - 3.15)$.

2. Si $\alpha \in \mathcal{B}$, pour tout $0 < \delta' < \delta$, il existe $\varepsilon > 0$ tel que si pour $|\operatorname{Im} z| < \delta$, $|f(z) - z - \alpha| < \varepsilon$; alors f est linéarisable et la linéarisante est dans $S(1, \delta')$.

3. Si $\alpha \notin \mathcal{B}$ la fraction de Blaschke $f_{a,\mu}(z) = \mu z^2 \frac{z+a}{1+\mu z}$, $a > 3$, $\rho(f_{a,\mu}) = \alpha$, est non linéarisable.

La partie 3 de ce théorème montre que la condition \mathcal{B} est ici encore optimale. La non linéarisabilité des fractions de Blaschke se déduit de la non linéarisabilité du polynôme quadratique à l'aide d'une construction de E. Ghys (voir [Gh] et [Do]).

4. Les techniques.

4.1. Démonstration géométrique du théorème de Siegel et de Bruno.

La démonstration est basée sur le critère dynamique de linéarisabilité : la stabilité.

A divers endroits des constantes universelles apparaîtront. Leur valeur explicite n'a aucune importance dans la discussion. On les notera C, C_0, C_1, \dots en donnant parfois le même nom à des constantes différentes.

4.1.a. L'application de retour dans un secteur.

Le point de départ est la remarque suivante de A. Douady et E. Ghys (voir [Do]) : l'ensemble

$$\{\alpha \in \mathbf{R} - \mathbf{Q}; \text{ tout } f(z) = e^{2\pi i\alpha} z + \mathcal{O}(z^2) \text{ est linearisable } \},$$

$$\text{est invariant sous l'action de } SL(2, \mathbf{Z}), \begin{pmatrix} a & b \\ c & d \end{pmatrix}. x = \frac{ax + b}{cx + d}.$$

Il suffit de montrer que cet ensemble de nombres de Siegel est invariant par $\alpha \mapsto -\alpha^{-1}$. La démonstration est géométrique. Considérons $f(z) = e^{2\pi i\alpha} z + \mathcal{O}(z^2)$ et un segment $l = [0, \varepsilon]$ issu de 0. Les arcs l et $f(l)$ forment, au voisinage de 0, un secteur que l'on peut refermer par un segment l' joignant un point de l à son image (voir figure 1).

Si l'on recolle l et $f(l)$ par f on obtient une variété topologique à bord qui est homéomorphe à $\overline{\mathbf{D}}$. Il est facile de montrer que, pour la structure complexe induite, l'intérieur est biholomorphe à \mathbf{D} . L'application de retour dans le secteur recollé est bien définie et holomorphe au voisinage de 0. On obtient ainsi un germe $g(z) = e^{2\pi i\beta} z + \mathcal{O}(z^2)$. Il est facile de voir que $\beta = -\alpha^{-1}$ (pour se convaincre penser au cas où f est une rotation).

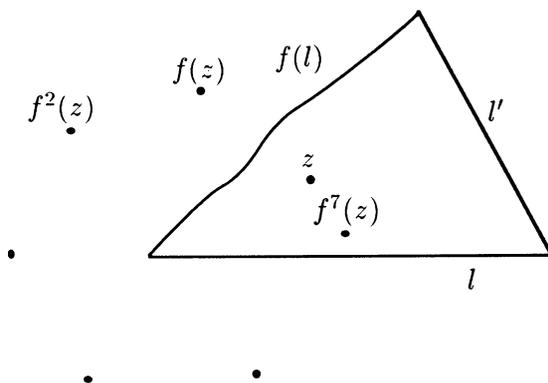


Figure 1

La dynamique de l'application de retour g est, au nombre d'itérations près, celle de f . On a f linéarisable si et seulement si g l'est. Ceci prouve l'invariance par $SL(2, \mathbf{Z})$. On peut dire davantage : si f a une orbite périodique proche de 0 alors g aura une orbite périodique dont on peut déterminer la période et le nombre de rotation à partir de celle de f . Aussi, si f a un "point d'instabilité", c.à.d. un point proche de 0 qui s'échappe, alors il lui correspondra un "point d'instabilité" qui s'échappera en moins d'itérations. C'est cette remarque et une réalisation quantitative de la construction précédente qui donnent le théorème de stabilité, i.e. de linéarisabilité.

On commence par normaliser $f \in S_\alpha$. On travaille plus aisément sur le revêtement universel de $\mathbf{D} - \{0\}$ donné par $E : \mathbf{H} = \{z \in \mathbf{C}; \text{Im } z > 0\} \rightarrow \mathbf{D} - \{0\}$, $E(z) = e^{2\pi iz}$. L'application $f \in S_\alpha$ se relève en F telle que $E \circ F = f \circ E$ et

1. $F : \mathbf{H} \rightarrow \mathbf{C}$ est univalente.
2. $F(z) = z + \alpha + \phi(z)$ où ϕ est \mathbf{Z} -périodique et $\lim_{\text{Im } z \rightarrow +\infty} \phi(z) = 0$, c'est à dire, le développement de Fourier s'écrit

$$\phi(z) = \sum_{n \geq 1} \phi_n e^{2\pi inz}.$$

On peut montrer que ϕ décroît exponentiellement quand $\text{Im } z \rightarrow +\infty$. On note $S(\alpha)$ l'espace de telles applications F . $S(\alpha)$ est compact et est le

revêtement universel de S_α . Pour construire l'application de retour, il faut qu'une demi-droite verticale et son image par F forment les bords d'une demi-bande verticale. En notant $\mathbf{H}_r = \mathbf{H} + ir$, pour r grand, F est proche de la translation $z \mapsto z + \alpha$ dans \mathbf{H}_r . Si $\alpha \neq 0$, par compacité de $S(\alpha)$, il existe une hauteur $t(\alpha)$ telle que, pour $z \in \mathbf{H}_{t(\alpha)}$, on ait l'estimation fondamentale,

$$(*) \quad |F(z) - z - \alpha| \leq \frac{\alpha}{4}.$$

On voit facilement que l'on peut prendre toujours

$$t(\alpha) = \frac{1}{2\pi} \log \alpha^{-1} + C.$$

La plus petite valeur de $t(\alpha)$ que l'on peut prendre de façon à satisfaire (*) représente la hauteur des fortes non linéarités de F . Au-delà de cette hauteur F est proche de la translation $z \mapsto z + \alpha$. Des fortes non linéarités pour F sont, par exemple, des points fixes. A cause de l'estimation (*) leur partie imaginaire est inférieure à $t(\alpha)$.

Maintenant on peut définir l'application de retour dans la demi-bande B délimitée par $l = [it(\alpha), +i\infty[$, $F(l)$ et $[it(\alpha), F(it(\alpha))]$. Si $z \in B$ on itère par F jusqu'à avoir $\operatorname{Re} F^n(z) > 1$. Si $\operatorname{Im} z \geq t(\alpha) + C_0$ on a $z' = F^n(z) - 1 \in B$ et $z \mapsto z'$ est l'application de retour dans la demi-bande B . On remarquera que l'on ne s'intéresse à la valeur de $t(\alpha)$ qu'à une constante universelle additive près, i.e. si l'on considère $t(\alpha) + C$ au lieu de $t(\alpha)$ la démonstration ne change pas.

En recollant les bords l et $F(l)$ de B par F on obtient une surface de Riemann \mathcal{S} correspondant à \mathring{B} biholomorphe à $\mathbf{D} - \{0\}$. En considérant une uniformisation convenable de \mathcal{S} sur un voisinage de \mathbf{D} , la dynamique de retour définit une application $g \in S_{\alpha^{-1}}$ (pour échanger $-\alpha^{-1}$ en α^{-1} on change l'orientation) qui se relève en $G \in S(\alpha^{-1})$. On peut alors montrer qu'aux points $z \in \mathbf{H}$ qui s'échappent de \mathbf{H} par itération par F , il leur correspond des points $z' \in \mathbf{C}$ qui s'échappent de \mathbf{H} par itération par G , et on contrôle $\operatorname{Im} z'$ en fonction de $\operatorname{Im} z$. Plus précisément :

Proposition. Soit $0 < \alpha < 1$, $F \in S(\alpha)$ et $t(\alpha) > 0$ tel que pour $\text{Im } z \geq t(\alpha)$, $|F(z) - z - \alpha| \leq \alpha/4$. Il existe $G \in S(\alpha^{-1})$ tel que si $z \in \mathbf{H}$, $\text{Im } z \geq t(\alpha)$, et pour $0 \leq i \leq n$, $F^i(z) \in \mathbf{H}$, mais $F^n(z) \notin \mathbf{H}$, alors il existe $z' \in \mathbf{C}$ vérifiant :

1. $\text{Im } z' \geq \alpha^{-1}(\text{Im } z - t(\alpha) - C_0)$.
2. Il existe m , $0 \leq m < n$, tel que $G^m(z') \notin \mathbf{H}$.

Les estimations du point 1 s'obtiennent en étudiant l'uniformisante de \mathcal{S} par des méthodes de fonctions univalentes ou par des estimations de modules d'anneaux inclus dans \mathcal{S} . Elles ne sont pas surprenantes : ce sont celles qu'on aurait si F était la translation $z \mapsto z + \alpha$, dans ce cas $t(\alpha) = 0$ et l'uniformisation est explicite. Or à la hauteur $t(\alpha)$, F est proche de cette translation, ...

4.1.b. L'argument de descente infinie.

On applique maintenant, par récurrence, la construction précédente. Partons de $\alpha \in \mathbf{R} - \mathbf{Q}$, $0 < \alpha < 1$, et $F \in S(\alpha)$. On pose (voir annexe) $\alpha_0 = \alpha$ et pour $n \geq 0$,

$$\alpha_n^{-1} = a_{n+1} + \alpha_{n+1},$$

avec $a_{n+1} \in \mathbf{N}^*$ et $0 < \alpha_{n+1} < 1$. Puis $\beta_{-1} = 1$ et $\beta_n = \alpha_0 \dots \alpha_n$.

On construit une suite $(F_n)_{n \geq 0}$ avec $F_n \in S(\alpha_n)$. Posons $F_0 = F$. Supposons construit F_n . Soit $G_n \in S(\alpha_n^{-1})$ l'application de retour de F_n construite en 4.2.a. On pose $F_{n+1}(z) = G_n(z) - a_{n+1}$. Une récurrence immédiate en utilisant la proposition de 4.2.a donne :

Proposition. Il existe C_0 constante universelle telle que si $z_0 \in \mathbf{H}$,

$$\text{Im } z_0 \geq \sum_{i=0}^m \beta_{i-1} t(\alpha_i) + C_0,$$

et $F^{n_0}(z_0) \notin \mathbf{H}$ alors il existe des points $(z_i)_{1 \leq i \leq m}$, $z_i \in \mathbf{H}$, et $(n_i)_{1 \leq i \leq m}$ tels que $F_i^{n_i}(z_i) \notin \mathbf{H}$ et

$$n_0 > n_1 > \dots > n_m \geq 0.$$

D'où l'on déduit, par un argument de descente infinie,

Théorème. *Si l'on peut choisir la suite $(t(\alpha_n))_{n \geq 0}$ telle que*

$$\sum_{n=0}^{+\infty} \beta_{n-1} t(\alpha_n) < +\infty,$$

alors f est stable en 0, donc linéarisable.

Noter que l'on obtient un disque de rayon $C \exp\left(-2\pi \sum_{n=0}^{+\infty} \beta_{n-1} t(\alpha_n)\right)$ contenu dans le domaine de stabilité.

Rappelons que l'on peut toujours prendre

$$t(\alpha_n) = \frac{1}{2\pi} \log \alpha_n^{-1} + C.$$

Puisqu'on a $\sum_{n \geq 0} \beta_{n-1} \leq C$ (voir annexe) une constante universelle rajoutée aux $t(\alpha_n)$ n'a aucune importance et la condition arithmétique précédente devient :

$$\sum_{n=0}^{+\infty} \beta_{n-1} \log \alpha_n^{-1} < +\infty.$$

C'est la condition \mathcal{B} (voir l'annexe arithmétique), d'où le théorème de Siegel et de Bruno. La figure 2 résume schématiquement l'idée de la démonstration.

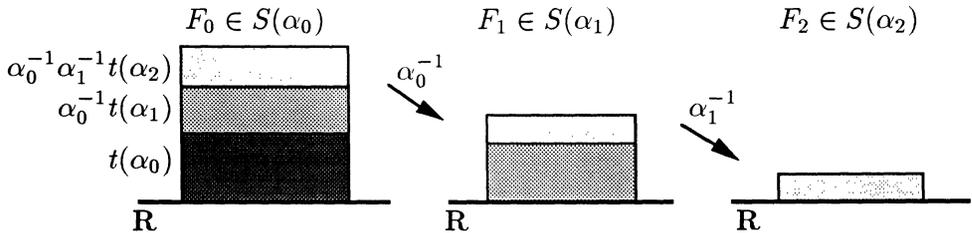


Figure 2

A chaque étape l'application de retour F_{n+1} ne se souvient de la dynamique de F_n qu'au-dessus de $t(\alpha_n)$. Les régions marquées, où l'information sur la dynamique est perdue, correspondent, dans le demi-plan où agit $F = F_0$, à des régions qui s'empilent et dont l'épaisseur est majorée par $\beta_{n-1}(t(\alpha_n) + C)$. Ainsi, si ces régions empilées ne dépassent pas une hauteur fixée, un point qui serait au-dessus et qui quitterait \mathbf{H} en n_0 itérations par F devrait les traverser toutes. Il lui correspondrait un point qui quitterait \mathbf{H} en n_1 itérations par F_1 , etc ... avec

$$n_0 > n_1 > \dots \geq 0,$$

absurde.

4.2. Constructions de germes non linéarisables.

4.2.a. Le collage d'une non linéarité.

Le principe pour obtenir des conditions diophantiennes optimales est de renverser la construction de l'application de retour. Dans le passage de F_{n+1} à F_n il faut réintroduire les non linéarités de F_n au-dessous de $t(\alpha_n)$ qui empêchaient de prendre $t(\alpha_n)$ plus petit. La non linéarité "générique" est celle créée par des points fixes avec une dynamique comme celle de la figure 3.

Rappelons que l'estimation (*) de 4.1.a montrait que $t(\alpha_n) > 0$ était supérieur à la partie imaginaire des points fixes de F_n . Nous allons décrire une façon de "coller" des points fixes à F_{n+1} . Mais d'abord décrivons comment revenir de F_{n+1} à F_n sans coller de points fixes.

On part de $G \in S(\alpha)$ et on veut construire $F \in S(\beta)$ avec $\beta^{-1} = a + \alpha$, $0 < \alpha, \beta < 1$ et $a \in \mathbf{N}^*$ (on rappelle que $\alpha_n^{-1} = a_{n+1} + \alpha_{n+1}$). On pose $\tilde{G}(z) = G(z) + a$, $\tilde{G} \in S(a + \alpha) = S(\beta^{-1})$. On a $\beta^{-1} > 1$ donc au-dessus d'une hauteur $C > 0$, \tilde{G} est proche de la translation $z \mapsto z + \beta^{-1}$, i.e. on a une demi-bande B' délimitée par $l_1 = [iC, +i\infty[$, $\tilde{G}(l_1) = l_2$ et $[iC, \tilde{G}(iC)]$. Lorsqu'on recolle l_1 et l_2 par \tilde{G} on obtient une surface de Riemann \mathcal{S}' correspondant à \hat{B}' qui est un demi-cylindre, i.e. biholomorphe à $\mathbf{D} - \{0\}$. La translation $T(z) = z + 1$, qui commute à \tilde{G} , induit une application univalente sur une

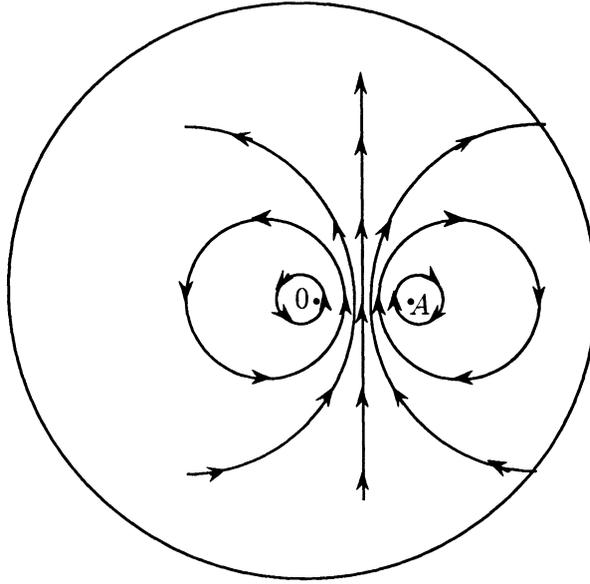


Figure 3

partie de S' qui est un voisinage de 0. Ainsi en choisissant une uniformisation $k : S' \rightarrow \mathbf{D} - \{0\}$, en conjuguant la dynamique, en normalisant (univalence sur \mathbf{D}) et en relevant à \mathbf{H} on obtient une dynamique $F' \in S(\beta)$.

Cette construction que l'on vient de décrire renverse la construction de l'application de retour mais ne colle pas de point fixe. Pour cela il faut prolonger le demi-cylindre S' en un cylindre S . On le construit en considérant la bande B délimitée par $l_1, l_2, l_3 =] -i\infty, iC]$ et $l_4 = l_3 + (\tilde{G}(iC) - iC)$, voir figure 4.

Dans la construction on a une certaine liberté : on pourrait choisir pour l_4 n'importe quelle image de l_3 par une application univalente proche de $z \mapsto z + \beta^{-1}$ et qui commute à T . Ceci permet de fixer la classe de conjugaison analytique de l'application de retour au voisinage du point fixe que l'on va construire.

Maintenant, en recollant l_1 et l_2 comme auparavant, puis l_3 et l_4 par $z \mapsto z + (\tilde{G}(iC) - iC)$ on montre facilement que l'on obtient une surface de Riemann qui est un cylindre, i.e. biholomorphe à $\mathbf{C} - \{0\}$. La translation T

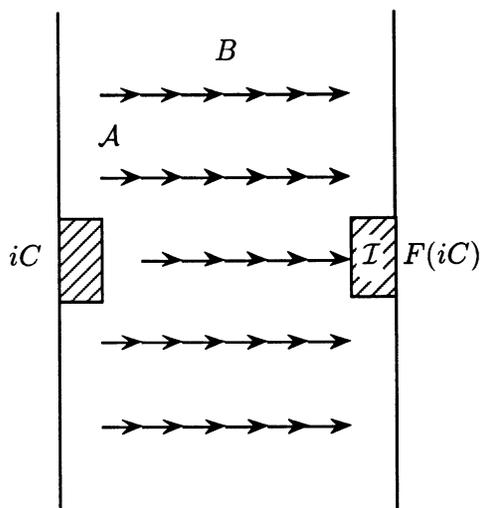


Figure 4

défini sur \mathcal{S} une dynamique univalente en dehors d'un voisinage \mathcal{I} autour du point qui correspond à iC , soit $f : \mathcal{S} - \mathcal{I} \rightarrow \mathcal{S}$ (figure 5).

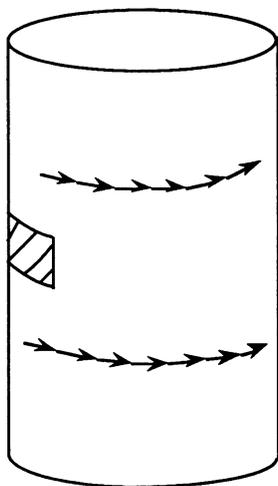


Figure 5

Comme auparavant, on uniformise \mathcal{S} par $k : \mathcal{S} \rightarrow \mathbb{C} - \{0, A\}$ de façon que $k(\mathcal{I})$ reste en dehors de \mathbb{D} et que le point A soit le plus proche possible

de 0. Le point 0 correspond au bout $+i\infty$ de S et le point A au bout $-i\infty$. Ces conditions sont nécessaires pour que :

1. La dynamique conjuguée $k \circ f \circ k^{-1}$ soit définie et univalente sur D .
2. La non linéarité, qui correspond au point fixe A , soit collée le plus proche possible de l'origine.

A ce point des estimations sur l'uniformisante k sont nécessaires. Pour cela on considère l'anneau A dessiné sur la figure 4. Une des composantes connexes de son complémentaire contient \mathcal{I} et l'autre 0 et A . Par des estimations de longueur extrémale de lacets joignant les deux bords, on montre que son module est minoré par $\frac{1}{2\pi} \log \beta^{-1} - C$. Si l'on choisit k telle que $k(\mathcal{I}) \cap D \neq \emptyset$ le problème extrémale de Teichmüller donne l'estimation $|A| \leq C\alpha$. Ainsi, lorsqu'on relève l'application $k \circ f \circ k^{-1}$ à \mathbf{H} , on obtient $F \in S(\beta)$ qui a un point fixe de partie imaginaire de l'ordre de $\frac{1}{2\pi} \log \beta^{-1} - C$. Ceci est, à une constante universelle près, la hauteur optimale à laquelle on pouvait espérer coller le point fixe. En effet, on a vu dans la partie directe que l'on pouvait toujours prendre $t(\beta) = \frac{1}{2\pi} \log \beta^{-1} + C_0$.

Remarquer que les estimations obtenues pour l'uniformisante ne sont pas surprenantes : ce sont celles que l'on obtiendrait si on était parti de $G(z) = z + \alpha$, dans ce cas on a des formules explicites pour k .

D'autre part on a complété la première construction, et G est une application de retour de F on obtient ainsi :

Proposition. *Soit $0 < \alpha < 1$, $a \in \mathbf{N}^*$, $\beta = \frac{1}{a+\alpha}$, et $G \in S(\alpha)$. Il existe $F \in S(\beta)$ telle que :*

1. F a un point fixe $z_0 \in \mathbf{H}$ avec $\text{Im } z_0 \geq \frac{1}{2\pi} \log \beta^{-1} - C$.
2. Si G a une orbite périodique de nombre de rotation p/q , i.e. $G^q(z) = z + p$ avec $q \geq 1$ et $p \in \mathbf{Z}$, avec

$$\text{Min}_{1 \leq i \leq q} \text{Im } G^i(z) \geq C_0,$$

alors F a une orbite périodique de nombre de rotation $\frac{1}{a+p/q}$ et

$$\text{Min}_{1 \leq j \leq aq+p} \text{Im } F^j(z') \geq \beta \text{Min}_{1 \leq i \leq q} \text{Im } G^i(z) + \frac{1}{2\pi} \log \beta^{-1} - C_0.$$

4.2.b. La récurrence et le passage à la limite.

Comme dans la partie directe, il ne reste qu'à faire une récurrence. A partir de $0 < \alpha < 1$ on a défini la suite $(\alpha_n)_{n \geq 0}$. On part d'un ordre $n \geq 0$ et on pose $F_{n,n+1} = z + \alpha_{n+1} \in S(\alpha_{n+1})$. A cette translation on lui colle un point fixe de la façon décrite en 4.2.a pour aboutir à $F_{n,n} \in S(\alpha_n)$, et par récurrence on obtient $F_{n,l} \in S(\alpha_l)$ à partir de $F_{n,l+1}$. En faisant une récurrence la proposition de 4.2.a donne :

Proposition. *Soit $m \geq 0$. Si $0 \leq n \leq m$ et p_n/q_n est une bonne réduite (cf. annexe arithmétique), alors $F_{m,0}$ a une orbite périodique de nombre de rotation p_n/q_n telle que*

$$\text{Min}_{1 \leq i \leq q_n} \text{Im } F_{m,0}^i(z_n) \geq \frac{1}{2\pi} \sum_{i=0}^n \beta_{i-1} \log \alpha_i^{-1} - C.$$

On doit faire l'hypothèse que p_n/q_n est une bonne réduite pour récupérer le point fixe de $F_{m,n} \in S(\alpha_n)$ en une orbite p_n/q_n de $F_{m,0}$. On rappelle que $\alpha \notin \mathcal{B}$ si et seulement si l'ensemble des bonnes réduites est infini.

Finalement, on extrait une sous-suite convergente de $(F_{n,0})_{n \geq 0} \subset S(\alpha)$. La proposition montre que, lorsque $\alpha \notin \mathcal{B}$, la limite F à une suite d'orbites périodiques qui accumule 0, et est non linéarisable.

4.3. Modifications pour les difféomorphismes du cercle.

Ici les démonstrations sont plus techniques et on ne va donner que quelques indications sur les nouveautés par rapport aux germes. Pour les difféomorphismes du cercle on travaille dans les espaces $S(\alpha, \Delta)$ et on aura intérêt à penser à la droite réelle "près de $+\infty$ " par analogie avec le cas des germes. En effet, dans les constructions, lorsque Δ sera grand, tout se passera comme si on travaillait dans les espaces $S(\alpha)$. Il faudra simplement les réaliser de façon symétrique par rapport à l'axe réel.

Pour les difféomorphismes analytiques du cercle, la stabilité de S^1 est encore équivalente à la linéarisabilité. Les démonstrations de linéarisabilité sont du même style que celle du théorème de Siegel et Bruno, du moins pour ce qui concerne le théorème local. On laisse au lecteur deviner la construction

de l'application de retour. Pour le théorème semi-local et global on est heurté à un nouveau problème : Que faire lorsque Δ est petit ?

En effet, dans ce cas défavorable on n'a plus d'estimations sur les parties imaginaires des correspondances orbitales de F et de son application de retour. Dans le cas des germes holomorphes on "coupait" à une hauteur de l'ordre de $\frac{1}{2\pi} \log \alpha^{-1} + C_0$ lorsque $F \in S(\alpha)$. Pour les difféomorphismes du cercle on ne peut plus procéder de la même façon si $\Delta \leq \frac{1}{2\pi} \log \alpha^{-1} + C_0$. Par contre, si $\Delta \geq \frac{1}{2\pi} \log \alpha^{-1} + C_0$ on obtient une application de retour $G \in S(\beta, \Delta')$ avec $\Delta' = \alpha^{-1}(\Delta - \log \alpha^{-1})$ et $\alpha^{-1} = a + \beta$, $a \in \mathbf{N}^*$, $0 < \beta < 1$. Si on peut procéder indéfiniment ainsi et obtenir une suite infinie $(F_n)_{n \geq 0}$, la même démonstration de stabilité que pour le théorème de Siegel s'applique car on contrôle les parties imaginaires des correspondances orbitales. Les conditions diophantiennes de linéarisabilité sont celles pour lesquelles on est heurté à la difficulté précédente un nombre fini de fois. Après ce nombre fini d'écueils on peut poursuivre la preuve habituelle.

Pour traiter les cas défavorables, Yoccoz démontre un lemme de Denjoy complexe qu'il utilise pour la construction de l'application de retour et pour obtenir une grande largeur de bande Δ . Ainsi, dans les cas défavorables à partir de $F_n \in S(\alpha_n, \Delta_n)$, $\Delta_n < \frac{1}{2\pi} \log \alpha_n^{-1}$, il obtient l'application de retour $F_{n+1} \in S(\alpha_{n+1}, \Delta_{n+1})$, avec $\Delta_{n+1} = C e^{2\pi \Delta_n}$. Ces formules expliquent la définition de la condition $\mathcal{H}(\Delta)$.

Tout ceci n'est possible que si $F \in S(\alpha, \Delta)$ et Δ est supérieur à une constante universelle. La première étape consiste à démontrer que l'on peut se ramener à ce cas là et partir toujours avec $\Delta_0 \geq 1000$. Il montre :

Proposition. *Soit $f \in S(\alpha, 0)$ et $\Delta > 0$. Il existe $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in PGL(2, \mathbf{Z})$ et une uniformisation réelle analytique $h : \mathbf{R}/\mathbf{Z} \rightarrow \mathbf{R}/(f^a \circ T^b)$, (quotient de \mathbf{R} par l'action de $f^a \circ T^b$) telle que $h^{-1} \circ (f^c \circ T^d) \circ h \in S(\frac{a\alpha+b}{c\alpha+d}, \Delta)$.*

Le théorème semi-local et cette proposition montrent le théorème global. La démonstration de la proposition comporte deux étapes :

-On montre d'abord l'existence d'un h telle que $h^{-1} \circ (f^c \circ T^d) \circ h$ soit C^2 proche d'une translation sur \mathbf{R} .

-Ensuite on utilise le lemme de Denjoy complexe, dont nous épargnons

l'énoncé, qui permet de contrôler dans le plan complexe, les orbites des difféomorphismes, et obtenir de cette façon une grande largeur de bande $\Delta > 0$.

Dans la partie de constructions de difféomorphismes non linéarisables on a aussi à distinguer divers cas. Cependant la géométrie est toujours la même que pour les germes holomorphes : on colle des points fixes avec une surface de Riemann \mathcal{S} cylindrique.

5. Autres applications et extensions de ces techniques.

5.1. Une obstruction à la linéarisabilité : l'accumulation par des orbites périodiques.

Si $f(z) = e^{2\pi i\alpha}z + \mathcal{O}(z^2)$, $\alpha \in \mathbf{R} - \mathbf{Q}$, est linéarisable il n'existe aucune orbite périodique, distincte de 0, dans un voisinage de l'origine. Ceci fournit dans certains cas un critère de non linéarisabilité. Illustrons ceci par un exemple.

Lorsque $\lambda = e^{2\pi i p/q}$, $p \in \mathbf{Z}$, $q \geq 1$, $p \wedge q = 1$, le polynôme quadratique P_λ a en 0 un point fixe d'ordre q , c'est à dire : une petite perturbation de λ en $e^{2\pi i\alpha}$ fait naître une orbite périodique, de période q , au voisinage de 0. Ainsi, pour $n \geq 1$, il existe un voisinage $U(p/q, n)$ de p/q dans \mathbf{R} tel que pour $\alpha \in U(p/q, n)$, P_λ ait une orbite périodique, de période q , contenue dans $\mathbf{D}_{1/n}$. L'ensemble $U_n = \bigcap_{p/q \in \mathbf{Q}} U(p/q, n)$ est un ouvert dense. Par le théorème de Baire on obtient un G_δ dense, $U = \left(\bigcap_{n \geq 1} U_n \right) \cap (\mathbf{R} - \mathbf{Q})$, de valeurs α telles que le polynôme quadratique P_λ , $\lambda = e^{2\pi i\alpha}$, a une suite d'orbites périodiques qui convergent vers 0 et n'est pas linéarisable.

Ceci laisse penser que cette obstruction topologique est peut-être la seule qui empêche la linéarisabilité, c'est à dire que tout germe non linéarisable aurait une suite d'orbites périodiques qui accumulent 0. Ceci a été conjecturé par P. Sad ([Sa]) et par V. Arnold ([Ar2]) pour les difféomorphismes du cercle. J.-C. Yoccoz a dirigé ma thèse sur ce sujet. On a :

Théorème (R. P.-M. [PM1]).

1. Si $\alpha \in \mathbf{R} - \mathbf{Q}$ satisfait $\sum_{n \geq 1} \frac{\log \log q_{n+1}}{q_n} < +\infty$ et $f(z) = e^{2\pi i \alpha} z + \mathcal{O}(z^2)$

est non linéarisable, alors f a une suite d'orbites périodiques qui accumulent 0 dont les périodes sont une sous-suite $(q_{n_k})_{k \geq 0}$ telle que

$$\sum_{k \geq 0} \frac{\log q_{n_k+1}}{q_{n_k}} = +\infty.$$

2. Si $\alpha \in \mathbf{R} - \mathbf{Q}$ satisfait $\sum_{n \geq 1} \frac{\log \log q_{n+1}}{q_n} = +\infty$ il existe $f \in S_\alpha$ telle

que toute orbite positive $(f^n(z))_{n \geq 0}$ contenue dans \mathbf{D} accumule 0. Donc f est non linéarisable et n'a pas d'orbites périodiques dans $\mathbf{D} - \{0\}$.

Dans [PM1] on a construit aussi des difféomorphismes du cercle non linéarisables et sans orbites périodiques. La condition diophantienne sur le nombre de rotation sous laquelle elle est faite est assez artificielle. Cependant, il se trouve qu'on peut récupérer de tels difféomorphismes du cercle à partir des germes holomorphes :

Théorème (R. P.-M. [PM1], complété récemment). Si $\alpha \in \mathbf{R} - \mathbf{Q}$ satisfait

$$\sum_{n \geq 1} \frac{\log \log q_{n+1}}{q_n} = +\infty,$$

pour tout $\Delta > 0$, il existe un difféomorphisme analytique du cercle f dont un relevé est dans $S(\alpha, \Delta)$ tel que toute orbite positive $(f^n(z))_{n \geq 0}$ restant dans $U = \mathbf{D}_{e^{2\pi\Delta}} - \overline{\mathbf{D}_{e^{-2\pi\Delta}}}$ accumule S^1 . Donc f est non linéarisable et sans orbites périodiques dans U .

La démonstration du point 1 du théorème pour les germes holomorphes suit les mêmes lignes que la démonstration de Yoccoz du théorème de Siegel et Bruno. Il faut faire deux remarques en plus. Premièrement, à un point fixe de $F_n \in S(\alpha_n)$ (cf. 4.1.b) correspond une orbite périodique de F de nombre de rotation p_n/q_n dont on sait minorer la partie imaginaire des points. Et en second lieu, si $F_n \in S(\alpha_n)$ n'a pas de points fixes la hauteur des fortes non

linéarités est plus basse. On obtient l'estimation fondamentale (*) de 5.1.a dès que

$$t(\alpha_n) = \frac{1}{2\pi} \log^+ \log \alpha_n^{-1} + C_0.$$

Le point 2 du théorème correspond à la construction de germes non linéarisables lorsque $\alpha \notin \mathcal{B}$. Cependant, il faut changer complètement la géométrie de la non linéarité collée. On ne cherche plus à coller des points fixes, la non linéarité voulue est synthétisée par la surface \mathcal{S} de la figure 6.

On modifie le cylindre de Yoccoz (figure 5) en faisant une coupure et collant un infini de plans d'un côté et d'autre. De façon très approximative, on peut dire que l'on a éclaté le point fixe A en un point de ramification infini.

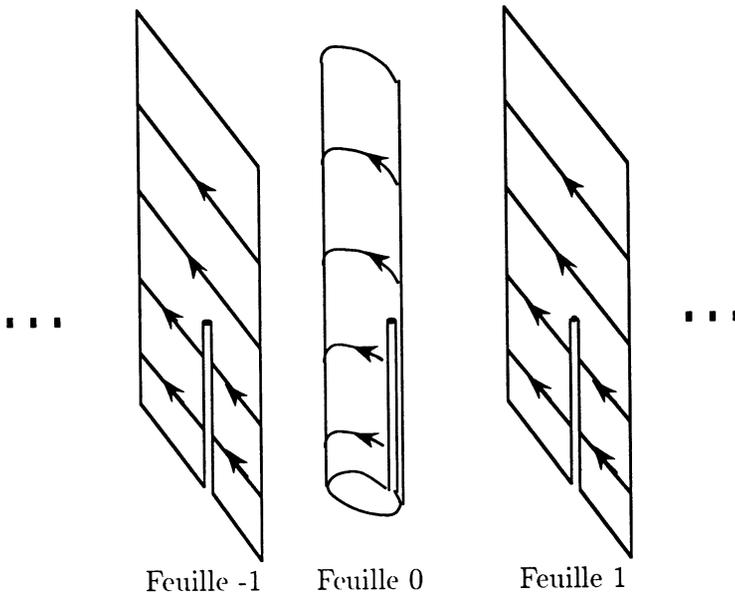


Figure 6

5.2. Symétries des germes non linéarisables.

Considérons le groupe de symétrie d'un germe de difféomorphisme holomorphe f ,

$$\text{Cent}(f) = \{g(z) = \mu z + \mathcal{O}(z^2); \mu \in \mathbf{C}^*, g \circ f = f \circ g\}.$$

Noter que le centralisateur de f , $\text{Cent}(f)$, contient le sous-groupe formé par les itérés de f et noté $f^{\mathbb{Z}}$. Lorsque $Df(0) \in \mathbb{C}^*$ n'est pas racine de l'unité, $\text{Cent}(f)$ se plonge dans (\mathbb{C}^*, \cdot) par le morphisme injectif,

$$\begin{aligned} \text{Cent}(f) &\longrightarrow \mathbb{C}^* \\ g &\longmapsto \mu = Dg(0) \end{aligned}$$

L'étude du groupe de symétries de f généralise le problème de Siegel comme le montre la proposition suivante.

Proposition. *Si $Df(0) \in \mathbb{C}^*$ n'est pas racine de l'unité, f est linéarisable si et seulement si*

$$\text{Cent}(f) = \mathbb{C}^*.$$

Dans le cas non linéarisable on a facilement :

Proposition. *Considérons f non linéarisable, $f(z) = e^{2\pi i\alpha}z + \mathcal{O}(z^2)$, $\alpha \in \mathbb{R} - \mathbb{Q}$. Alors $f^{\mathbb{Z}} \subset \text{Cent}(f) \subset S^1$ et $\text{Cent}(f)$ est*

-dense dans S^1 ,

- F_σ donc borélien dans S^1 ,

-de mesure de Haar nulle dans (S^1, \cdot) .

-Si $g \in \text{Cent}(f)$ et $Dg(0)$ n'est pas racine de l'unité alors g est non linéarisable.

-Si $g \in \text{Cent}(f)$ et $Dg(0)$ est racine q^{eme} de l'unité alors $g^q = \text{id}$.

Il se pose naturellement le problème de la structure de $\text{Cent}(f)$ pour f non linéarisable : Quels sont les sous-groupes de S^1 de la forme $\text{Cent}(f)$?

J. Moser a donné certaines restrictions :

Théorème (J. Moser [Mo]). *Soit $f(z) = e^{2\pi i\alpha}z + \mathcal{O}(z^2)$ et $g \in \text{Cent}(f)$ avec $Dg(0) = e^{2\pi i\beta}$. Supposons qu'il existe $\gamma, \tau > 0$ tels que pour $q \geq 1$ et $p \in \mathbb{Z}$,*

$$\text{Max}(|q\alpha - p|, |q\beta - p|) \geq \frac{\gamma}{q^\tau},$$

alors f est linéarisable.

M. Herman, qui a construit des centralisateurs non dénombrables pour des difféomorphismes du cercle non linéarisables en classe différentiable ([He2], voir aussi [Y4] pour d'autres résultats en classe différentiable), a demandé si de tels objets existaient en classe analytique. Ceci montrerait que des conditions diophantiennes sont nécessaires dans le théorème de Moser et donnerait des exemples de germes f tels que $f^{\mathbf{Z}}$ ne soit pas d'indice fini dans $\text{Cent}(f)$. Il se trouve que la géométrie des non linéarités collées en 5.1 permet la construction de tels exemples :

Théorème (R. P.-M. [PM2]). *Il existe un ensemble de Cantor explicite $K \subset \mathbf{T} = \mathbf{R}/\mathbf{Z}$ et une famille commutative $(f_t)_{t \in K}$, tels que $f_t \in S(t)$ et si $t \in K \cap (\mathbf{R} - \mathbf{Q})/\mathbf{Z}$, f_t est non linéarisable.*

On obtient $K \cap (\mathbf{Q}/\mathbf{Z})$ dense dans K . Le Cantor K engendre un sous-groupe $G \subset \mathbf{T}$ non dénombrable et pour $t \in K$, $G \subset \text{Cent}(f)$. Les constructions sont assez souples pour imposer à tous les $(f_t)_{t \in K}$ d'avoir une même suite d'orbites périodiques qui accumule 0 ou de n'avoir aucune orbite périodique distincte de 0 lorsque $t \in K \cap (\mathbf{R} - \mathbf{Q})/\mathbf{Z}$.

La construction se réalise en partant du flot des rotations et en collant des non linéarités du "type" de celles de la figure 6. A chaque collage on perd des intervalles du flot et en passant à la limite on ne garde que le compact K . Les recours utilisés pour éviter les problèmes liés au collage simultané des non linéarités à plusieurs éléments du flot, font que les nombres de $K \cap (\mathbf{R} - \mathbf{Q})/\mathbf{Z}$ sont "très Liouville".

6. Equivalence de certains problèmes.

Les constructions de ce dernier paragraphe datent des derniers mois de 1991. Bien que non reliés directement aux techniques géométriques précédentes, on a cru bon de les mentionner puisqu'ils unifient trois problèmes de petits diviseurs. Les résultats mentionnés auparavant se transportent des uns aux autres.

6.1. Comment obtenir des difféomorphismes du cercle à partir de germes de difféomorphismes holomorphes de $(\mathbf{C}, 0)$.

Lorsque $f(z) = e^{2\pi i\alpha}z + \mathcal{O}(z^2)$ est linéarisable ou $\alpha \in \mathbf{Q}$ on voit facilement qu'au voisinage de 0 on a des compacts connexes K distincts de $\{0\}$

tels que $f(K) = K$. Il se trouve que ceci est général :

Théorème (R. P.-M.). *Soit $f \in S(\alpha)$, $\alpha \in \mathbf{R}$ et $0 \leq r \leq 1$. Il existe un compact connexe plein $0 \in K \subset \overline{\mathbf{D}}_r$ tel que $K \cap \partial\mathbf{D}_r \neq \emptyset$ et*

$$f(K) = K.$$

Le cas le plus intéressant se présente lorsque f est non linéarisable, dans ce cas la topologie de K peut être compliquée. G. Birkhoff ([Bi]) a démontré l'existence d'un tel K qui vérifie uniquement $f(K) \subset K$. Le théorème se démontre pour f linéarisable (ou $\alpha \in \mathbf{Q}$) et puis par passage à la limite en considérant des perturbations de f .

Une fois ceci établi, on remarque que, si $r \neq 0$, $\mathbf{D} - K$ est un anneau topologiquement biholomorphe à $\mathbf{D}_R - \overline{\mathbf{D}}$ (figure 7), soit $h : \mathbf{D}_R - \overline{\mathbf{D}} \rightarrow \mathbf{D} - K$ une représentation conforme.

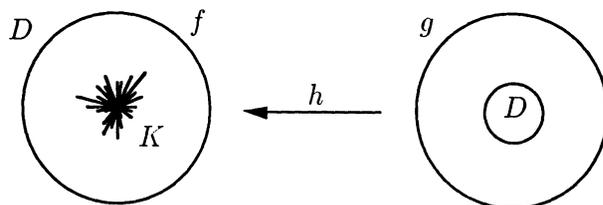


Figure 7

On conjugue f par h et on obtient $g = h^{-1} \circ f \circ h$ qui est un homéomorphisme d'un voisinage de S^1 dans $\mathbf{D}_R - \overline{\mathbf{D}}$. On voit facilement que g s'étend continument à S^1 , $g(S^1) = S^1$, puis par le principe de réflexion de Schwarz, g définit un difféomorphisme analytique de S^1 . On peut montrer que $\rho(g) = \alpha = \rho(f)$. Il est clair que si g est linéarisable alors f l'est aussi, et si f n'a pas d'orbites périodiques distinctes de 0, g n'en a pas. Aussi, si l'on choisit $r > 0$ très petit le difféomorphisme g obtenu s'étend de façon univalente dans un anneau autour de S^1 de module aussi grand que l'on veut.

On peut ainsi transporter les résultats relatifs aux germes aux difféomorphismes du cercle et inversement. Plus précisément, les résultats directs

pour les difféomorphismes du cercle donnent les résultats correspondant pour les germes holomorphes. Exemple : le théorème local de Yoccoz pour les difféomorphismes du cercle implique le théorème de Siegel et de Bruno. Les contre-exemples construits pour les germes holomorphes en donnent d'autres pour les difféomorphismes du cercle. Exemple : si f est non linéarisable sans orbites périodiques distinctes de 0, alors g l'est aussi. On obtient que les problèmes locaux pour les difféomorphismes du cercle sont équivalents à ceux des germes holomorphes. Aussi on peut lire les théorèmes globaux pour les difféomorphismes du cercle dans le cadre des germes holomorphes ...

6.2. Comment obtenir des singularités de feuilletages holomorphes à partir de germes de difféomorphismes holomorphes de $(\mathbf{C}, 0)$.

Le procédé inverse est classique. Pour $\alpha > 0$ on considère le feuilletage d'un voisinage de $(\mathbf{C}^2, 0)$, singulier à l'origine, défini par la forme

$$\omega = x(1 + \dots)dy + \alpha y(1 + \dots)dx,$$

où les points de suspensions désignent des termes d'ordre supérieur. L'axe pointé $\{y = 0\} - \{0\}$ est une feuille. Pour $\varepsilon > 0$ assez petit, le lacet $\gamma : [0, 1] \rightarrow \{y = 0\}$, $\gamma(t) = (\varepsilon e^{2\pi it}, 0)$, définit une holonomie (application de retour) sur un petit disque transverse à $\{y = 0\}$ en $(\varepsilon, 0)$. On obtient ainsi une classe de conjugaison d'un germe holomorphe $f(z) = e^{2\pi i\alpha}z + \mathcal{O}(z^2)$, et on définit une application H_α des classes de conjugaison par des difféomorphismes holomorphes de $(\mathbf{C}^2, 0)$ des feuilletages de ce type dans les classes de conjugaison des germes de difféomorphismes de $(\mathbf{C}, 0)$.

J.-F. Mattei et R. Moussu ont montré ([M-M]) que H_α était toujours injective. J. Martinet et J.-P. Ramis ont montré la surjectivité lorsque $\alpha \in \mathbf{Q}$ et les résultats de linéarisation de A. Bruno ([Br]) donnent aussi ceci lorsque $\alpha \in \mathcal{B}$ (il y a une seule classe de conjugaison). En fait on a :

Théorème (J.-C. Yoccoz, R. P.-M. [Y-P-M]). H_α est surjective.

C'est à dire que l'on réalise tout germe $f(z) = e^{2\pi i\alpha}z + \mathcal{O}(z^2)$ comme une holonomie du type précédent. La démonstration se situe au niveau de \mathbf{C}^2 , revêtement universel de $\mathbf{C}^2 - (\{x = 0\} \cup \{y = 0\})$ donné par $E_2 : (z_1, z_2) \mapsto (e^{2\pi iz_1}, e^{2\pi iz_2})$. On considère le feuilletage linéaire de \mathbf{C}^2 donné

par $z_2 + \alpha z_1 = \text{constante}$. En quotientant \mathbf{C}^2 par l'action de $T_1 : (z_1, z_2) \mapsto (z_1 + 1, z_2)$ et $T_2 : (z_1, z_2) \mapsto (z_1, z_2 + 1)$, l'image par E_2 est le feuilletage linéaire donné par $\omega = xdy + \alpha ydx$, d'holonomie linéarisable.

Soit $f(z) = e^{2\pi i \alpha z} + \mathcal{O}(z^2) \in S_\alpha$ le germe que l'on veut réaliser et $F \in S(\alpha)$ son relevé à \mathbf{H} , $F(z) = z + \alpha + \phi(z)$. On réalise ce germe comme holonomie du feuilletage, induit par le feuilletage linéaire, sur la variété complexe M obtenue en recollant un domaine fondamental pour l'action de $F_1 : (z_1, z_2) \mapsto (z_1 + 1, z_2 + \phi(z_2 + \alpha z_1))$ et $F_2 : (z_1, z_2) \mapsto (z_1, z_2 + 1)$. On ne considère ces applications que dans une zone $\{(z_1, z_2) \in \mathbf{C}^2; \text{Im}(z_2 + \alpha z_1) \geq C_0\} \subset \mathbf{C}^2$ où F_1 et F_2 sont proches de T_1 et T_2 . Tout le problème consiste à montrer que la variété abstraite M est encore biholomorphe à un voisinage de l'origine dans \mathbf{C}^2 en retirant les axes coordonnés.

Ceci est vrai pour la structure C^∞ : on a une uniformisation C^∞ de ce type car on peut conjuguer F_1 et F_2 à T_1 et T_2 de façon C^∞ . Le défaut d'holomorphie se trouve dans une seule des coordonnées de l'uniformisation et est petit (on a un $\bar{\partial}$ petit). Si on réalise le collage dans un domaine fondamental bien choisi de façon que M soit une variété de Stein, on peut appliquer les techniques de L. Hörmander ([Ho]) pour résoudre l'équation $\bar{\partial}$ avec des estimées. On peut ainsi rectifier l'uniformisation C^∞ pour qu'elle devienne holomorphe. On veut conserver un feuilletage holomorphe ce qui conduit de nouveau à résoudre une deuxième équation $\bar{\partial}$ pour la forme qui le définit.

7. Annexe arithmétique.

7.1. Développement en fraction continue usuel.

On rappelle tout d'abord le développement en fraction continue usuel (voir [La]). Si x est un nombre réel on note $[x]$ sa partie entière et $\{x\} = x - [x]$ sa partie décimale.

Soit $\alpha \in \mathbf{R} - \mathbf{Q}$. On pose $a_0 = [\alpha] \in \mathbf{Z}$, $\alpha_0 = \{\alpha\}$, et pour $n \geq 1$,

$$a_n = [\alpha_{n-1}^{-1}],$$

$$\alpha_n = \{\alpha_{n-1}^{-1}\}.$$

Puis

$$q_{-2} = p_{-1} = 1,$$

$$q_{-1} = p_{-2} = 0,$$

et pour $n \geq 0$,

$$q_n = a_n q_{n-1} + q_{n-2},$$

$$p_n = a_n p_{n-1} + p_{n-2}.$$

La suite de rationnels $(p_n/q_n)_{n \geq 0}$ est constituée des réduites de α . La suite des dénominateurs croît d'autant plus vite que α est bien approché par des nombres rationnels. Ensuite, on définit la suite $(\beta_n)_{n \geq -1}$ par $\beta_{-1} = 1$ et pour $n \geq 0$,

$$\beta_n = \prod_{i=0}^n \alpha_i.$$

La suite $(\beta_n)_{n \geq 0}$ décroît toujours au moins de façon géométrique. On vérifie sans peine, pour $n \geq -2$,

$$\beta_n = (-1)^n (q_n \alpha - p_n),$$

$$q_{n+1} \beta_n + q_n \beta_{n+1} = 1.$$

J.-C. Yoccoz introduit et utilise dans [Y1] et [Y2] une variante de ce développement classique. On pose $\|x\| = \inf_{p \in \mathbf{Z}} |x - p|$, et $\alpha_0 = \|\alpha\|$ et pour $n \geq 1$,

$$\alpha_n = \|\alpha_{n-1}^{-1}\|.$$

Les β_n sont définis de la même façon à partir des α_n . Cette variante a l'avantage qu'on a $0 < \alpha_n < 1/2$ ce qui facilite les constructions géométriques. Si tous les α_n usuels vérifient $0 < \alpha_n < 1/2$, les deux développements en fraction continue coïncident. Pour alléger l'exposé, on ne considère que le développement usuel. Ceci n'a de relevance que pour la définition des conditions arithmétiques pour les difféomorphismes du cercle (où l'on pensera à

prendre les α_n modifiés), et pour les constantes explicites dans les énoncés des théorèmes des difféomorphismes du cercle.

On définit la fonction

$$\Phi(\alpha) = \sum_{i=0}^{+\infty} \beta_{i-1} \log \alpha_i^{-1}.$$

Et l'application,

$$\tau_\alpha(x) = \begin{cases} e^x & \text{pour } 0 \leq x \leq \log \alpha^{-1} \\ \alpha^{-1}(x - \log \alpha^{-1} + 1) & \text{pour } x \geq \log \alpha^{-1} \end{cases}.$$

7.2. Les conditions arithmétiques.

7.2.a. Pour les germes.

Il est facile de voir que la condition

$$\sum \frac{\log q_{n+1}}{q_n} < +\infty,$$

est équivalente à

$$\Phi(\alpha) < +\infty.$$

On note \mathcal{B} les nombres vérifiant ces deux conditions. Cet ensemble est un F_σ de \mathbf{R} .

Pour $C_0 > 0$ et $0 \leq l \leq n$ on pose

$$B_{n,l} = \beta_{l-1}^{-1} \left[\sum_{j=l}^n \beta_{j-1} \log \alpha_j^{-1} - C_0 \sum_{j=l}^n \beta_{j-1} \right].$$

On dit que p_n/q_n est une bonne réduite si pour $0 \leq l \leq n$ on a

$$B_{n,l} \geq \log \alpha_l^{-1}.$$

Ceci signifie que p_n/q_n approche bien α par rapport aux réduites précédentes. On montre sans peine que $\alpha \notin \mathcal{B}$ si et seulement si α possède une infinité de bonnes réduites.

7.2.b. Pour les difféomorphismes du cercle.

Soit $\Delta > 0$. On définit la suite $(\Delta_n(\alpha))_{n \geq 0}$ par

$$\begin{aligned}\Delta_0(\alpha) &= 2\pi\Delta, \\ \Delta_{n+1} &= \tau_{\alpha_n}(\Delta_n(\alpha)).\end{aligned}$$

L'ensemble $\mathcal{H}(\Delta)$ est formé par les nombres $\alpha \in \mathbf{R} - \mathbf{Q}$ pour lesquels à partir d'un certain rang on a $\Delta_n(\alpha) \geq \log \alpha_n^{-1}$. Cet ensemble est un F_σ , et on peut vérifier que

$$\bigcup_{\Delta > 0} \mathcal{H}(\Delta) = \mathcal{B}.$$

Finalement \mathcal{H} est formé des nombres α tels que pour tout $n \geq 0$, on ait $\alpha_n \in \mathcal{H}(\Delta)$. Cet ensemble ne dépend plus de la constante $\Delta > 0$ choisie et est invariant sous l'action de $SL(2, \mathbf{Z})$, i.e. il définit une condition diophantienne. \mathcal{H} est un $F_{\sigma, \delta}$, i.e. une intersection dénombrable de F_σ .

BIBLIOGRAPHIE

- [Ar1] V.I. ARNOLD, *Small denominators I: on the mapping of a circle into itself*, Ivestijia Akad. Nauk., serie Math, **25**, I (1961) p. 21-86 = Translations Amer. Math. Soc., 2nd series, **46**, p. 213-284.
- [Ar2] V.I. ARNOLD, *Chapitres supplémentaires de la théorie des équations différentielles ordinaires*, Editions MIR, (1980), p.120.
- [Bi] G. D. BIRKHOFF, *Surface transformation and their dynamical applications*, Oeuvres complètes t.II p. 111-129.
- [Bo] J.-B. BOST, *Tores invariants des systèmes dynamiques hamiltoniens (d'après Kolmogorov, Arnold, Moser, Rüssmann, Zehnder, Herman, Pöschel, ...)*, Sémin. Bourbaki 639, 1984-85; Astérisque 133-134 (1985), p. 113-157.

- [Bot] L. E. BÖTKHER, *Les principales lois de convergence des itérés et leurs applications à l'analyse*, Izv. Kazan. Fiz-Mat. **14** (1904) (en russe).
- [Br] A.D. BRUNO, *Analytical form of differential equations*, Transactions Moscow Math. Soc. **25** (1971), p. 131-288; **26** (1972), p. 199-239.
- [Ca] C. CAMACHO, *On the local structure of conformal mappings and holomorphic vector fields*, Astérisque 59-60 (1978), p. 83-84.
- [Cr1] H. CREMER, *Über die Iteration rationaler Funktionen*, Jahresber. Deutsch Math. Verein **33** (1925), p. 185-210.
- [Cr2] H. CREMER, *Zum Zentrumproblem*, Math. Ann. **98** (1928), p. 151-153.
- [Cr3] H. CREMER, *Über das Zentrumproblem (Mit besonderer Berücksichtigung der Lückenreihen)*, Ber. Math. Phys. Klasse der Säch. Akad. Wiss. Leipzig **82** (1930), p.243-250.
- [Cr4] H. CREMER, *Über die Schrödersche Funktionalgleichung und das Schwarzsche Eckenabbildungsproblem*, *ibid.* **84** (1932), p.291-324.
- [Cr5] H. CREMER, *Über die Häufigkeit der Nichtzentren*, Math. Ann. **115** (1938), p. 573-580.
- [Ch] T. M. CHERRY, *A singular case of iteration of an analytic functions : a contribution to the small divisor problem*, Nonlinear problems of engineering, ed. W. F. Ames, Academic Press, New York (1964), p. 29-50.
- [Do] A. DOUADY, *Disques de Siegel et anneaux de Herman*, Sémin. Bourbaki n° 677, 1986-87.
- [Ec] J. ECALLE, *Les fonctions réurgentes et leurs applications*, t. I, II, III, Publications mathématiques d'Orsay 81-05, 81-06, 85-05.
- [He1] M. R. HERMAN, *Recent results and some open questions on Siegel's linearisation theorem of germs of complex analytic diffeomorphisms of \mathbb{C}^n near a fixed point*, Proceedings VIIIth Int. Conf. Math. Phys., World Scientific Publishers, Singapour (1987).

- [He2] M. R. HERMAN, *Sur la conjugaison différentiable des difféomorphismes du cercle à des rotations*, Publ. Math. I.H.E.S. **49** (1979).
- [Ju] G. JULIA, *Oeuvres*, vol. I, Gauthiers-Villars, Paris (1969), p.231-232.
- [Gh] E. GHYS, *Transformations holomorphes au voisinage d'une courbe de Jordan*, C. R. Acad. Sc. t. 298 série I **16** (1984), p. 385-388.
- [Ko] G. KOENIGS, *Recherches sur les intégrales de certaines équations fonctionnelles*, Ann. Sc. E.N.S. (3eme série) **1** (1884) supplém. p. 1-14.
- [Ho] L. HÖRMANDER, *An introduction to complex analysis in several variables*, Van Nostrand (1966).
- [La] S. LANG, *Introduction to diophantine approximations*, Addison-Wesley series in mathematics (1966).
- [Le] L. LEAU, *Etude sur les équations fonctionnelles à une ou plusieurs variables*, Ann. Fac. Sci. Toulouse **11** (1897).
- [Ma] B. MALGRANGE, *Travaux d'Ecalte et de Martinet-Ramis sur les systèmes dynamiques*, Sémin. Bourbaki, exp 582 (1981), Astérisque 92-93 (1982), p. 59-73.
- [Mo] J. MOSER, *On commuting circle mappings and simultaneous Diophantine approximations.*, Math. Zeitschrift **205** (1990), p. 105-121.
- [M-M] J.-F. MATTEI, R. MOUSSU, *Holonomie et intégrales premières*, Ann. Sc. E.N.S. **13** (1980), p. 469-523.
- [M-R] J. MARTINET, J.-P. RAMIS, *Problème de modules pour les équations différentielles non linéaires du premier ordre*, Publ. Math.IHES, **55** (1982), p. 63-164
- [Pf] G. A. PFEIFFER, *On the conformal mapping of curvilinear angles. The functional equation $\phi[f(x)] = a_1\phi(x)$* , Math. Ann. 1917
- [PM1] R. PÉREZ MARCO, *Sur la dynamique des germes de difféomorphismes holomorphes de $(\mathbb{C}, 0)$ et des difféomorphismes analytiques du cercle*, Thèse Université de Paris-Sud, décembre 1990.

- [PM2] R. PÉREZ-MARCO, *Centralizers I : Uncountable centralizers for non linearizable holomorphic germs of $(\mathbf{C}, 0)$* , préprint IHES (1991).
- [Ru] H. RÜSSMANN, *Kleine Nener II : Bemerkungen zur Newtonschen Methode*, Nach. Akad. Wiss. Göttingen, Math. Phys. Kl. (1972), p. 1-10.
- [Sa] P. SAD, *A note on non linearizable analytic functions*, Bol.Soc. Bras. Mat. (1981).
- [Si] C.L. SIEGEL, *Iterations of analytic functions*, Ann. Math.43 (1942), p. 807-812.
- [Vo] S. M. VORONIN, *Classification analytique des germes d'applications conformes $(\mathbf{C}, 0) \rightarrow (\mathbf{C}, 0)$ tangentes à l'identité*, Functional Analysis 15/1 (1981), p.1-17.
- [Y1] J.-C. YOCCOZ, *Théorème de Siegel, polynômes quadratiques et nombres de Brjuno*, préprint (1987); à paraître dans Astérisque.
- [Y2] J.-C. YOCCOZ, *Conjugaison des difféomorphismes analytiques du cercle*, A paraître dans Acta Mathematica.
- [Y3] J.-C. YOCCOZ, *Conjugaison différentiable des difféomorphismes du cercle dont le nombre de rotation vérifie une condition diophantienne*, Ann. Sc. E.N.S. 4eme série t. 17 (1984), p. 333-359.
- [Y4] J.-C. YOCCOZ, *Centralisateurs et conjugaison différentiable des difféomorphismes du cercle*, Thèse d'État Université de Paris-Sud, juin 1985.
- [Y5] J.-C. YOCCOZ, *Linéarisation des germes de difféomorphismes holomorphes de $(\mathbf{C}, 0)$* , C. R. Acad. Sci. Paris 306, p. 55-58.
- [Y-P-M] J.-C. YOCCOZ, R. PÉREZ-MARCO, *Germes de feuilletages holomorphes à holonomie prescrite*, preprint Université de Paris-Sud 92-24, à paraître dans Astérisque.

Ricardo PÉREZ-MARCO

Université Paris-Sud

Math. 425 - U.R.A. 1169 du C.N.R.S.

F-91405 ORSAY CEDEX

Astérisque

JEAN-CHRISTOPHE YOCCOZ

Travaux de Herman sur les tores invariants

Astérisque, tome 206 (1992), Séminaire Bourbaki,
exp. n° 754, p. 311-344

<http://www.numdam.org/item?id=SB_1991-1992__34__311_0>

© Société mathématique de France, 1992, tous droits réservés.

L'accès aux archives de la collection « Astérisque » (<http://smf4.emath.fr/Publications/Asterisque/>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

TRAVAUX DE HERMAN SUR LES TORES INVARIANTS

par Jean-Christophe YOCCOZ

0. Cet exposé se veut une introduction aux spectaculaires récents résultats de M. Herman en dynamique symplectique ([H1] à [H7]).

Après quelques rappels dans les deux premiers paragraphes, le troisième est consacré à la construction d'un contre-exemple, dans la catégorie des champs hamiltoniens, au C^∞ closing lemma.

Le paragraphe 4 tourne tout entier autour de l'existence de tores diophantiens, invariants ou translatés, de codimension 1. De même que pour les courbes translattées de Rüssmann ([R1], [H9]), ces tores apparaissent déjà dans des situations non conservatives. Dans un cadre conservatif (flots hamiltoniens, difféomorphismes préservant le volume), ils produisent des contre-exemples à plusieurs conjectures : Hypothèse quasi-ergodique, conjecture de Pesin, ...

Dans le paragraphe 5 on présentera quelques uns des résultats de Herman sur les tores invariants *lagrangiens*. La géométrie de ces tores dépend alors très fortement de la signature de la torsion; on possède aussi de nombreuses informations sur la dynamique sur le tore générique.

Pour alléger l'exposition, nous nous placerons fréquemment, pour présenter les résultats, dans un cadre simple, mais pas optimal. En particulier, on se place *toujours, sauf mention contraire* en classe C^∞ .

1. COURBES INVARIANTES PAR LES DIFFÉOMORPHISMES DU CYLINDRE

1.1. On pose $\mathbf{A} = \mathbf{T}^1 \times \mathbf{R}$; on note $\theta \in \mathbf{T}^1$, $r \in \mathbf{R}$ les coordonnées canoniques sur \mathbf{A} . On considère la 1-forme $v = rd\theta$ sur \mathbf{A} et la 2-forme symplectique $\omega = dv = dr \wedge d\theta$.

Soit $f : (\theta, r) \mapsto (\Theta(\theta, r), R(\theta, r))$ un difféomorphisme de \mathbf{A} , isotope à l'identité. On écrit

$$\Theta(\theta, r) = \theta + \phi(\theta, r), \quad \phi \in C^\infty(\mathbf{A}).$$

DÉFINITIONS.— *Le difféomorphisme F est*

- monotone positif (resp. négatif) si $\partial_r \phi > 0$ (resp. $\partial_r \phi < 0$) sur \mathbf{A} ;
- symplectique si $F^* \omega = \omega$;
- exact symplectique si la 1-forme $F^* v - v$ est exacte ;
- complètement intégrable si $R(\theta, r) = r$, et $\phi(\theta, r) = l(r)$ ne dépend pas de θ .

1.2. Théorie de Birkhoff ([B1],[H9],[F],[L3])

Soit F un difféomorphisme monotone positif de \mathbf{A} .

Premier théorème de Birkhoff : *Supposons que F laisse invariant le graphe d'une fonction $\psi \in C^0(\mathbf{T}^1)$. Alors ψ est lipschitzienne et on a, pour presque tout $\theta \in \mathbf{T}^1$*

$$\frac{\partial_r R}{\partial_r \phi} (F^{-1}(\theta, \psi(\theta))) > D\psi(\theta) > -\frac{\partial_\theta \Theta}{\partial_r \phi} (\theta, \psi(\theta)).$$

On suppose maintenant de plus que F est symplectique (il suffirait que F préserve une mesure de Radon chargeant tous les ouverts).

Deuxième théorème de Birkhoff: *Supposons que F préserve un ouvert U de \mathbf{A} ayant les propriétés suivantes :*

- Il existe $r_0 < r_1$ tels que $\mathbf{T}^1 \times (-\infty, r_0) \subset U \subset \mathbf{T}^1 \times (-\infty, r_1)$;
- U est difféomorphe à \mathbf{A} ;
- U est l'intérieur de son adhérence.

Alors la frontière de U est le graphe d'une fonction $\psi \in C^0(\mathbf{T}^1)$ (lipschitzienne d'après le premier théorème).

Conséquences : Soit F un difféomorphisme symplectique monotone positif de \mathbf{A} ; pour $r_0 > 0$, considérons l'ensemble $\mathcal{C}(r_0)$ des courbes simples, homotopes à $\{r = 0\}$, de \mathbf{A} qui sont invariantes par F , et contenues dans $\mathbf{T}^1 \times [-r_0, r_0]$. Toute courbe $C \in \mathcal{C}(r_0)$ est graphe d'une application ψ_C , lipschitzienne de rapport inférieur ou égal à $K(r_0)$. L'union des courbes $C \in \mathcal{C}(r_0)$ est compacte, de même que l'ensemble des ψ_C (pour la C^0 topologie). Si l'on écrit

$$F(\theta, \psi_C(\theta)) = (f_C(\theta), \psi_C(f_C(\theta))),$$

et qu'on note ρ_C le nombre de rotation de f_C , l'ensemble des ρ_C est compact. De plus, si ρ_C est irrationnel et $C' \in \mathcal{C}(r_0)$ est distincte de C , on a soit $\rho_{C'} < \rho_C$, $\psi_{C'} < \psi_C$ soit $\rho_{C'} > \rho_C$, $\psi_{C'} > \psi_C$.

Remarque : Pour que $\mathcal{C}(r_0)$ soit non vide, il faut que F soit *exact* symplectique. Mais cette condition n'est pas suffisante comme le montre l'exemple suivant :

$$F_0(\theta, r) = (\theta + r, r + \cos 2\pi(\theta + r)), \quad F_0^n(0, 0) = (0, n).$$

1.3. Courbes diophantiennes : forme normale de Birkhoff

Pour $\gamma > 0$, $\tau \geq 0$, on pose :

$$CD(\gamma, \tau) = \{\alpha \in \mathbf{R}; \forall q \geq 1, \forall p \in \mathbf{Z}, |q\alpha - p| \geq \gamma q^{-1-\tau}\},$$

$$CD(\tau) = \bigcup_{\gamma > 0} CD(\gamma, \tau), \quad CD = \bigcup_{\tau \geq 0} CD(\tau).$$

Un nombre $\alpha \in CD$ est dit *diophantien*. On dit qu'une courbe (simple) C , invariante par un difféomorphisme F est diophantienne si F/C préserve l'orientation et si le nombre de rotation de F/C est diophantien. Si F et C sont lisses, d'après le théorème de conjugaison global des difféomorphismes

du cercle ([H8],[Y]), il existe sur toute courbe diophantienne C une coordonnée $\theta \in \mathbf{T}^1$ telle que F/C soit une translation dans cette coordonnée.

PROPOSITION ([B2]).— Soient F un difféomorphisme symplectique de \mathbf{A} , isotope à l'identité, et C une courbe invariante diophantienne, homotope à $\{r = 0\}$. Pour tout entier $n \geq 0$, il existe un difféomorphisme symplectique H de \mathbf{A} , isotope à l'identité, tel qu'on ait $H(C) = \{r = 0\}$ et :

$$H \circ F \circ H^{-1}(\theta, r) = \left(\theta + \sum_1^n b_i r^i + \alpha, r \right) + \mathcal{O}(r^{n+1}),$$

où $\alpha \in \mathbf{CD}$ est le nombre de rotation de F/C et les nombres b_i , $i \geq 1$ sont uniquement déterminés (invariants de Birkhoff).

On dit que la torsion de F le long de C est positive si $b_1 > 0$, négative si $b_1 < 0$, dégénérée si $b_1 = 0$.

Exercice : Si le difféomorphisme symplectique F est monotone positif, la torsion de F le long de toute courbe invariante diophantienne, homotope à $\{r = 0\}$, est positive. Ceci n'est plus nécessairement vrai si la courbe n'est pas homotope à $\{r = 0\}$.

1.4. Le théorème de la courbe translátée.

Le théorème de la courbe translátée, du à H. Rüssmann ([R1],[H9]), est une forme plus flexible, car valable sans hypothèse de conservativité, du théorème de la courbe invariante de J. Moser ([Mo1]).

Soit $L(\theta, r) = (\theta + l(r), r)$ un difféomorphisme complètement intégrable de \mathbf{A} . On suppose que $l(0) = \alpha$ est diophantien, et que $l'(0) \neq 0$. Pour $\mu \in \mathbf{R}$ on note T_μ la translation : $(\theta, r) \mapsto (\theta, r + \mu)$ de \mathbf{A} .

THÉORÈME.— Si F est un difféomorphisme de \mathbf{A} , assez proche de L dans la C^∞ -topologie, il existe un réel μ_F proche de 0, et une fonction $\psi_F \in C^\infty(\mathbf{T}^1)$ petite dans la C^∞ -topologie, tels que le graphe de ψ_F soit une courbe invariante, de nombre de rotation α , par $T_{\mu_F} \circ F$.

Remarques. — 1) L'application $F \mapsto (\mu_F, \psi_F)$ définie au voisinage de L , est une bonne application de classe C^∞ au sens de Hamilton ([Ha1],[H10],[Bo]).

2) Soient $\tau \geq 0$, $\gamma > 0$, $\delta > 0$. Supposons que $l'(r) \neq 0$ pour $r \in [-\delta, \delta]$; posons $K = l([-\delta, \delta]) \cap CD(\gamma, \tau)$. Il existe alors, pour tout $k > 2\tau + 3$, un voisinage de L dans la C^k -topologie telle que la conclusion du théorème soit valide pour tout nombre de rotation dans K . La dépendance de ψ_F , μ_F par rapport au nombre de rotation est lisse au sens de Whitney ([La],[Po]).

3) Soit F un difféomorphisme exact symplectique monotone positif de \mathbf{A} . Soit α un nombre réel diophantien .

- Si F ne préserve pas de courbe , homotope à $\{r = 0\}$, de nombre de rotation α , il en est de même pour toute perturbation de F .

- Si F préserve une courbe *lisse* , homotope à $\{r = 0\}$, de nombre de rotation α , il en est de même pour toute perturbation *exacte symplectique* de F .

1.5. Zones d'instabilité.

Soit F un difféomorphisme exact symplectique monotone positif de \mathbf{A} . L'union des courbes invariantes par F homotopes à $\{r = 0\}$ est une partie fermée de \mathbf{A} . Les composantes connexes du complémentaire qui sont bordées par deux telles courbes sont appelées *zones d'instabilité* de F ; la dynamique de F dans ces zones est loin d'être parfaitement comprise, mais de nombreux progrès ont été enregistrés récemment, dont il ne nous est pas possible de rendre compte ici; nous renvoyons le lecteur aux références : [An], [B-H], [Ch], [L1] à [L4], [M1] à [M5].

2. QUELQUES RAPPELS DE DYNAMIQUE SYMPLECTIQUE

2.1. Une variété symplectique est une paire (M, ω) , où M est une variété (de dimension paire $2n$) et ω une 2-forme fermée partout non dégénérée. Soit N une variété de dimension moitié. Une immersion $j : N \rightarrow M$ est lagrangienne si $j^*\omega = 0$. Soit $H \in C^\infty(M)$; le champ X_H défini par:

$$\omega(X_H, \cdot) = dH(\cdot)$$

est dit champ hamiltonien associé au hamiltonien H .

Si une hypersurface V de M est variété de niveau de deux hamiltoniens H_1, H_2 sur M , les champs associés X_{H_1}, X_{H_2} sont colinéaires sur V et les flots qu'ils engendrent ont, sur V , les mêmes orbites.

2.2. Le fibré cotangent $p : T^*V \rightarrow V$ d'une variété V est canoniquement une variété symplectique : on pose $\omega = dv$, où la 1-forme v au point ζ de T^*M est définie par:

$$v(Y) = \zeta(Tp(Y)), \quad Y \in T_\zeta(T^*M).$$

Une section $s : V \rightarrow T^*V$, c'est-à-dire une 1-forme sur V , est alors un plongement lagrangien si et seulement si s est fermée.

PROPOSITION (Weinstein)[W].— *Si N est une sous-variété lagrangienne plongée d'une variété symplectique M , il existe un difféomorphisme symplectique d'un voisinage de N dans M sur un voisinage de la section nulle dans T^*N (envoyant N sur la section nulle).*

2.3. Pour $2n \geq k \geq 0$, on munit $\mathbf{T}^{2n-k} \times \mathbf{R}^k$ des coordonnées $\theta_1, \theta_2, \dots, \theta_{2n-k} \in \mathbf{T}^1, r_1, \dots, r_k \in \mathbf{R}$. Pour $H \in C^\infty(\mathbf{T}^{2n-k} \times \mathbf{R}^k)$, on pose

$$\nabla H = {}^t(\partial_{\theta_1} H, \dots, \partial_{r_k} H).$$

Soit $A \in GL(2n, \mathbf{R})$ une matrice antisymétrique ; la forme symplectique (à coefficients constants) telle que

$$\omega_A(A\nabla H, \cdot) = dH(\cdot)$$

est dite associée à A .

2.4. On identifiera $T^*(\mathbf{T}^n)$ à $\mathbf{T}^n \times \mathbf{R}^n$ qu'on notera \mathbf{A}^n . La forme symplectique canonique est alors $\omega = dv = \sum dr_i \wedge d\theta_i$, $v = \sum r_i d\theta_i$.

Soit F un difféomorphisme de \mathbf{A}^n , homotope à l'identité. On écrit:

$$F(\theta, r) = (\Theta(\theta, r), R(\theta, r)) = (\theta + \phi(\theta, r), R(\theta, r)).$$

DÉFINITIONS.— *On dit que F est*

- symplectique si $F^*\omega = \omega$;

- exact symplectique si $F^*v - v$ est exacte ;
- monotone si $\partial_r \phi$ est partout inversible ;
- fortement monotone si pour tout $\theta \in \mathbf{T}^n, r \mapsto \phi(\theta, r)$ est un difféomorphisme de \mathbf{R}^n ;
- complètement intégrable si $R(\theta, r) = r$ et $\phi(\theta, r) = l(r)$ ne dépend pas de θ .

Si $L : (\theta, r) \mapsto (\theta + l(r), r)$ est complètement intégrable et symplectique la différentielle $Dl(r)$ est pour tout $r \in \mathbf{R}^n$ une matrice symétrique. On a $l = D\hat{l}$, pour une certaine fonction $\hat{l} \in C^\infty(\mathbf{R}^n)$.

Si le graphe d'une fonction $\psi \in C^\infty(\mathbf{T}^n, \mathbf{R}^n)$ est une sous-variété lagrangienne, on a $\psi = [\psi] + d\hat{\psi}$ avec $\hat{\psi} \in C^\infty(\mathbf{T}^n), [\psi] = \int \psi dm$ (où on notera dm la mesure de Lebesgue sur \mathbf{T}^n).

2.5. Conditions diophantiennes.

Soient $\gamma > 0, \tau \geq 0, n \geq 1$. On définit:

$$CDH_n(\gamma, \tau) = \{(\alpha_1, \dots, \alpha_n) \in \mathbf{R}^n, \forall k \in \mathbf{Z}^n - \{0\}, \left| \sum k_i \alpha_i \right| \geq \gamma (\sup |k_i|)^{1-n-\tau}\},$$

$$CDH_n(\tau) = \bigcup_{\gamma > 0} CDH_n(\gamma, \tau), \quad CDH_n = \bigcup_{\tau \geq 0} CDH_n(\tau)$$

$$CD_n(\gamma, \tau) = \{(\alpha_1, \dots, \alpha_n) \in \mathbf{R}^n, \forall k \in \mathbf{Z}^n - \{0\}, \forall l \in \mathbf{Z}, |l + \sum k_i \alpha_i| \geq \gamma (\sup |k_i|)^{-n-\tau}\},$$

$$CD_n(\tau) = \bigcup_{\gamma > 0} CD_n(\gamma, \tau), \quad CD_n = \bigcup_{\tau \geq 0} CD_n(\tau).$$

Pour $\tau > 0$, $CD_n(\tau)$ et $CDH_n(\tau)$ sont de mesure de Lebesgue pleine, mais maigres au sens de Baire (pour $n \geq 2$ dans le cas de CDH_n). Les translations entières préservent les $CD_n(\gamma, \tau)$; on note de la même façon leurs images dans \mathbf{T}^n .

2.6. Théorèmes de conjugaison locaux sur les tores

Pour $\alpha \in \mathbf{T}^n$, on note R_α la translation $\theta \mapsto \theta + \alpha$ de \mathbf{T}^n . Pour $\alpha \in \mathbf{R}^n$, on note X_α le champ de vecteurs $\sum \alpha_i \partial_{\theta_i}$. On note $\text{Diff}_0^\infty(\mathbf{T}^n)$ l'ensemble des difféomorphismes h de \mathbf{T}^n qui sont homotopes à l'identité et vérifient $\int (h(\theta) - \theta) dm = 0$

THÉORÈME.— *Soit $\alpha \in CD_n$. Tout difféomorphisme f de \mathbf{T}^n suffisamment proche de R_α s'écrit de façon unique $f = R_\lambda \circ h \circ R_\alpha \circ h^{-1}$, avec $\lambda \in \mathbf{T}^n$ proche de 0 et $h \in \text{Diff}_0^\infty(\mathbf{T}^n)$ proche de l'identité.*

THÉORÈME.— *Soit $\alpha \in CDH_n$. Tout champ de vecteurs X sur \mathbf{T}^n suffisamment proche de X_α s'écrit de façon unique $X = X_\lambda + h_* X_\alpha$, avec $\lambda \in \mathbf{R}^n$ proche de 0 et $h \in \text{Diff}_0^\infty(\mathbf{T}^n)$ proche de l'identité.*

L'application $f \mapsto (\lambda, h)$ (resp. $X \mapsto (\lambda, h)$) définie au voisinage de R_α (resp. de X_α) est une bonne application de classe C^∞ au sens de Hamilton ([Ha1],[H10],[Bo]).

Si f préserve la mesure de Lebesgue (resp. si X est de divergence nulle) la conjugaison h préserve la mesure de Lebesgue, car R_α (resp. X_α) est uniquement ergodique (c'est-à-dire la mesure de Lebesgue est l'unique mesure de probabilité invariante).

De façon beaucoup plus élémentaire, on a :

PROPOSITION.— *Soient $\alpha \in CDH_n$, $\psi \in C^\infty(\mathbf{T}^n)$, $\psi > 0$. Il existe un unique nombre réel $c > 0$ et un unique difféomorphisme $h \in \text{Diff}_0^\infty(\mathbf{T}^n)$ tels qu'on ait $\psi X_\alpha = h_* X_{c\alpha}$.*

2.7. Forme normale de Birkhoff

Soit F un difféomorphisme symplectique d'une variété symplectique (M, ω) . Un tore T , de dimension k , invariant par F , est dit *diophantien* s'il existe $\alpha \in CD_k$ et un difféomorphisme $h : \mathbf{T}^k \rightarrow T$ tels que $h \circ R_\alpha \circ h^{-1} = F|_T$.

Remarques.— 1) Si on remplace h par $h \circ A$, $A \in GL(k, \mathbf{Z})$, on remplace α par $A^{-1}\alpha$.

2) La 2-forme $h^*(\omega/T)$, étant invariante par la translation ergodique R_α , est à coefficients constants. Si la forme symplectique ω est exacte (par exemple $M = T^*V$), on doit avoir $h^*(\omega/T) = 0$; ceci implique $k \leq \frac{1}{2} \dim M$.

Soit T un tore invariant diophantien *lagrangien*.

PROPOSITION.— *Pour tout $n \geq 0$, il existe un plongement symplectique H d'un voisinage de $\mathbf{T}^n \times \{0\}$ dans $\mathbf{T}^n \times \mathbf{R}^n$ dans M tel que $H(\mathbf{T}^n \times \{0\}) = T$ et*

$$H^{-1} \circ F \circ H(\theta, r) = \left(\theta + \sum_1^m l_i(r) + \alpha, r \right) + \mathcal{O}(\|r\|^{m+1}),$$

où, pour $1 \leq i \leq m$, $l_i(r)$ est une application polynômiale homogène de degré i de \mathbf{R}^n dans \mathbf{R}^n de la forme $D\hat{l}_i$ (où \hat{l}_i est un polynôme homogène de degré $i + 1$ dans \mathbf{R}^n).

Remarque.— Si on remplace H par $H \circ T^* A$, $A \in GL(n, \mathbf{Z})$, on remplace \hat{l}_i par $\hat{l}_i \circ {}^t A^{-1}$. La signature de la forme quadratique \hat{l}_2 reste inchangée.

DÉFINITION.— *On dit que le tore T a une torsion dégénérée, positive, négative, indéfinie, nulle, suivant que la forme quadratique \hat{l}_2 est dégénérée, positive, négative, indéfinie, nulle.*

Ce qui précède possède un analogue pour les champs hamiltoniens: au voisinage d'un tore lagrangien diophantien invariant, un champ hamiltonien est approché, à un ordre arbitraire, par un champ hamiltonien complètement intégrable.

2.8. Théorèmes KAM

Nous renvoyons à [Bo] pour un exposé très complet et une bibliographie très détaillée sur les différentes variantes de théorèmes KAM. Nous nous contentons ici d'en citer une.

THÉORÈME.— *Soient $(M^{2n}, \omega = dv)$ une variété exacte symplectique et F_0 un difféomorphisme exact symplectique de (M, ω) . Soient $\alpha \in CD_n$ et $h_0 : \mathbf{T}^n \rightarrow M$ un plongement (automatiquement lagrangien) tels que*

$T_0 = h_0(\mathbf{T}^n)$ soit invariant par F_0 et qu'on ait :

$$h_0 \circ R_\alpha \circ h_0^{-1} = F_0/T_0.$$

Pour tout difféomorphisme exact symplectique F suffisamment proche de F_0 , il existe un plongement (lagrangien) $h_F : \mathbf{T}^n \rightarrow M$, proche de h_0 , tel que $T_F = h_F(\mathbf{T}^n)$ soit invariant par F et qu'on ait

$$h_F \circ R_\alpha \circ h_F^{-1} = F/T_F.$$

L'application $F \mapsto h_F$ est une bonne application de classe C^∞ au sens de Hamilton.

3. UN CONTRE-EXEMPLE AU C^∞ -CLOSING LEMMA POUR LES CHAMPS HAMILTONIENS

3.1. Le "closing lemma"

Étant donné une variété M et un difféomorphisme f de M (ou un flot $(f^s)_{s \in \mathbf{R}}$ associé à un champ de vecteurs X), on dit qu'un point $x \in M$ est (positivement) récurrent si

$$\liminf_{s \rightarrow +\infty} d(f^s(x), x) = 0.$$

Le problème du "closing lemma" est le suivant : peut-on perturber f (ou X) de manière que le point x devienne *périodique* ? Ce problème se pose pour les difféomorphismes les plus généraux, mais aussi à l'intérieur de classes plus restreintes de difféomorphismes ou de champs de vecteurs : difféomorphismes préservant le volume, difféomorphismes symplectiques, champs de vecteurs hamiltoniens, difféomorphismes biholomorphes, ...

Les réponses à ces questions dépendent de façon essentielle de la topologie dans laquelle sont permises les perturbations. En ce qui concerne la classe de tous les difféomorphismes, la réponse est trivialement positive en C^0 -topologie; en C^1 -topologie, la réponse est encore positive, mais c'est un théorème difficile de Pugh ([Pu]). En C^2 -topologie, Gutierrez [Gu] a construit un champ de vecteurs sur \mathbf{T}^2 possédant un point récurrent $x \neq 0$

qu'on ne peut rendre périodique par une C^2 -perturbation à support compact dans $\mathbf{T}^2 - \{0\}$. La question est totalement ouverte pour la C^2 -topologie sur une variété compacte.

Le résultat de Pugh a été étendu à la classe des champs de vecteurs hamiltoniens par Pugh et Robinson ([P-R]) : si x est récurrent pour le champ hamiltonien X_{H_0} , on peut perturber H_0 dans la C^2 -topologie en un hamiltonien H tel que x soit périodique pour X_H .

L'exemple ci-dessous a été introduit par Zehnder ([Z1]) pour montrer qu'une surface d'énergie compacte d'un hamiltonien ne contient pas nécessairement d'orbites périodiques. Herman a démontré que ce même exemple fournit un contre-exemple au "closing lemma" en topologie C^k (k assez grand), dans la classe des champs de vecteurs hamiltoniens.

3.2. Soient $\alpha \in CDH_{2n-1}$, et A une matrice antisymétrique dans $GL(2n, \mathbf{R})$ de la forme :

$$A = \begin{pmatrix} B & {}^t\alpha \\ -\alpha & 0 \end{pmatrix}, \quad B \in M(2n-1, \mathbf{R}).$$

Munissons $\mathbf{T}^{2n-1} \times [-\delta, \delta]$ de la forme symplectique constante ω_A associée à A (cf. 2.3).

Posons $H_0(\theta, r) = r$. Sur chaque surface d'énergie $\{r = r_0\}$, le champ hamiltonien X_{H_0} coïncide avec le champ de vecteurs constant X_α sur \mathbf{T}^{2n-1} .

Considérons maintenant un hamiltonien $H \in C^\infty(\mathbf{T}^{2n-1} \times [-\delta, \delta])$, C^∞ -proche de H_0 et $r_0 \in [-\delta/2, \delta/2]$. La surface de niveau $\{H = r_0\}$ est le graphe T d'une fonction $\psi \in C^\infty(\mathbf{T}^{2n-1}, [-\delta, \delta])$.

Posons $K(\theta, r) = r - \psi(\theta)$, de sorte que $T = \{K = 0\}$. Les champs X_H et X_K sont colinéaires sur T .

Le champ $Z = -B\nabla_\theta\psi + X_\alpha$ sur \mathbf{T}^{2n-1} est la projection de X_K/T . Comme B est antisymétrique, Z a une divergence nulle. Comme Z est proche de X_α , on peut l'écrire (cf. 2.6)

$$Z = X_\lambda + h_*X_\alpha,$$

où $h \in \text{Diff}_0(\mathbf{T}^{2n-1})$ préserve la mesure de Lebesgue.

Or on a

$$\int Z dm = \alpha = \int h_* X_\alpha dm,$$

donc $\lambda = 0$.

Le champ X_K sur T est donc conjugué à X_α ; d'après 2.6, le champ X_H sur T est conjugué à $X_{c\alpha}$, pour un certain c proche de 1.

Ceci démontre le :

THÉORÈME ([H1], [H2]).— *Pour tout hamiltonien H suffisamment proche de H_0 , et toute surface d'énergie $\{H = r_0\}$, $|r_0| \leq \delta/2$, le champ de vecteurs X_H sur $\{H = r_0\}$ est conjugué à un champ diophantien $X_{c\alpha}$, $c > 0$. En particulier, la région $\{|H| \leq \delta/2\}$ ne contient pas d'orbites périodiques de X_H .*

Remarques.— 1) On peut plonger l'exemple précédent dans \mathbf{T}^{2n} , puis éclater symplectiquement des points de \mathbf{T}^{2n} , pour obtenir une famille dénombrable de variétés symplectiques compactes sur lesquelles se produit le phénomène précédent. On ne peut cependant plonger cet exemple dans une variété symplectique exacte (cf. 2.7), sauf si $n = 1$. Le "closing lemma" pour les champs hamiltoniens sur un fibré cotangent (par exemple) demeure une question ouverte.

2) En prenant $\alpha \in CDH_{2n-1}(0)$, il suffit d'exiger dans le théorème précédent que H soit C^k -proche de H_0 , pour tout k donné $> 2n$.

3) On verra au prochain paragraphe une illustration plus subtile de ce phénomène de rigidité du nombre de rotation.

4. TORES INVARIANTS DE CODIMENSION 1

4.1. Le théorème du tore translaté ([H7])

Pour $n \geq 1, \delta > 0$, on pose $\mathbf{B}^n = \mathbf{T}^n \times \mathbf{R}$, $\mathbf{B}^n(\delta) = \mathbf{T}^n \times [-\delta, +\delta]$, et on note $p_1 : \mathbf{B}^n \rightarrow \mathbf{T}^n$ et $p_2 : \mathbf{B}^n \rightarrow \mathbf{R}$ les deux projections.

Pour $(\lambda, \mu) \in \mathbf{B}^n$, on note $R_{\lambda, \mu}$ la translation $(\theta, r) \mapsto (\theta + \lambda, r + \mu)$.

Considérons un plongement $F_0 : \mathbf{B}^n(\delta) \hookrightarrow \mathbf{B}^n$, de classe C^∞ , qui vérifie $F_0(\theta, 0) = (\theta + \alpha, 0)$, où α appartient à CD_n .

THÉORÈME.— Pour tout plongement $F : \mathbf{B}^n(\delta) \hookrightarrow \mathbf{B}^n$ suffisamment proche de F_0 , il existe une translation $R_{\lambda,\mu}$ de \mathbf{B}^n , une application $\psi \in C^\infty(\mathbf{T}^n)$ de moyenne nulle, et un difféomorphisme $h \in \text{Diff}_0^\infty(\mathbf{T}^n)$ tels que:

- 1) le plongement $\tilde{F} = R_{\lambda,\mu} \circ F$ préserve le graphe de ψ ;
- 2) pour $\theta \in \mathbf{T}^n$, on a

$$\tilde{F}(\theta, \psi(\theta)) = (f(\theta), \psi(f(\theta))) \text{ avec } f = h \circ R_\alpha \circ h^{-1}.$$

Compléments.— 1) Pour $F = F_0$, on a $\lambda = \mu = 0$, $\psi = 0$, $h = id$; pour F proche de F_0 , si l'on exige que λ, μ, ψ soient petits et h proche de l'identité, ils sont uniquement déterminés par les propriétés du théorème.

2) Le bon cadre (du moins, le plus agréable et le plus flexible) pour formuler de nombreux théorèmes de petits diviseurs est celui de Hamilton ([Ha1],[Ha2],[H10],[Bo]) ; il définit des notions de “bon espace de Fréchet” et de “bonnes applications de classe C^r ” entre bons espaces de Fréchet, pour lesquelles le théorème d'inversion locale (ou des fonctions implicites) reste valide (il faut cependant vérifier l'inversibilité de la différentielle sur tout un voisinage du point considéré). Je renvoie à [Bo] pour de plus amples renseignements. La démonstration du théorème, dont nous donnons une vague indication ci-dessous, montre qu'alors l'application $F \mapsto (\lambda, \mu, \psi, h)$ est une bonne application de classe C^∞ au sens de Hamilton.

3) Fixons $\tau \geq 0, \gamma > 0$. Pour tout $\alpha' \in CD_n(\gamma, \tau)$, on peut appliquer le théorème à $R_{\alpha'-\alpha,0} \circ F_0$; on obtient alors pour F dans un voisinage $U_{\alpha'}$ de F_0 , $\lambda_{\alpha'}, \mu_{\alpha'}, \psi_{\alpha'}, h_{\alpha'}$ tels que

$$F(\theta, \psi_{\alpha'}(\theta)) = (f_{\alpha'}(\theta), \psi_{\alpha'}(f_{\alpha'}(\theta))) - (\lambda_{\alpha'}, \mu_{\alpha'}),$$

$$f_{\alpha'} = h_{\alpha'} \circ R_{\alpha'} \circ h_{\alpha'}^{-1}.$$

Les deux faits suivants sont vitaux pour les applications :

— L'intersection des $U_{\alpha'}$, lorsque α' décrit $CD_n(\gamma, \tau)$, est encore un voisinage de F_0 .

— L'application $\alpha' \mapsto (\lambda_{\alpha'}, \mu_{\alpha'}, \psi_{\alpha'}, h_{\alpha'})$ est lisse au sens de Whitney.

Pour démontrer ceci à peu de frais, la remarque suivante de Bost est fondamentale : considérons la démonstration du théorème, pour α fixé,

dans le cadre de Hamilton ; elle fait intervenir de bons espaces de Fréchet F_1, F_2, \dots et de bonnes applications $L_1 : F_1 \rightarrow F_2, \dots$. Considérant ensuite α comme un paramètre, remplaçons F_1, F_2 par $\tilde{F}_1 = C^p(CD_n(\gamma, \tau), F_1)$, $\tilde{F}_2 = C^p(CD_n(\gamma, \tau), F_2)$ et L_1 par $\tilde{L}_1 : \tilde{F}_1 \rightarrow \tilde{F}_2$. Alors \tilde{F}_1, \tilde{F}_2 sont encore de bons espaces de Fréchet et \tilde{L}_1 une bonne application, ce qui permet de conclure.

4.2. Quelques indications sur la démonstration

Soit F_1 un difféomorphisme de $\mathbf{T}^n \times \mathbf{T}^1 = \mathbf{T}^{n+1}$ qui coïncide avec F_0 au voisinage de $\mathbf{T}^n \times \{0\}$. Posons

$$C_1 = \{\psi \in C^\infty(\mathbf{T}^n), \int \psi dm = 0\},$$

$$C_2 = C_1^n \subset C^\infty(\mathbf{T}^n, \mathbf{R}^n),$$

$$C_3 = \{g \in C^\infty(\mathbf{T}^{n+1}, \mathbf{R}^{n+1}), g|_{\mathbf{T}^n \times \{0\}} \equiv 0\}.$$

Pour $\psi \in C_1$, définissons

$$K_\psi(\theta, r) = (\theta, r + \psi(\theta)).$$

Pour $\varphi \in C_2$, définissons

$$H_\varphi(\theta, r) = (\theta + \varphi(\theta), r).$$

Pour $g \in C_3$, définissons

$$J_g(\theta, r) = (\theta, r) + g(\theta, r).$$

Nous voulons montrer que l'application

$$\Phi : \mathbf{T}^n \times \mathbf{R} \times C_1 \times C_2 \times C_3 \rightarrow \text{Diff}^\infty(\mathbf{T}^{n+1})$$

$$(\lambda, \mu, \psi, \varphi, g) \mapsto R_{\lambda, \mu} \circ K_\psi \circ H_\varphi \circ J_g \circ F_1 \circ H_\varphi^{-1} \circ K_\psi^{-1},$$

définie au voisinage de l'origine $(0, 0, 0, 0, 0)$, est un bon difféomorphisme de classe C^∞ au sens de Hamilton d'un voisinage de l'origine sur un voisinage de F_1 dans $\text{Diff}^\infty(\mathbf{T}^{n+1})$. Il suffit pour ceci de vérifier les hypothèses du

théorème d'inversion locale de Hamilton. Nous n'indiquons ici que la raison pour laquelle la différentielle $D_0\Phi$ de Φ à l'origine est inversible.

On a

$$\begin{aligned} D_0\Phi(\lambda, \mu, \psi, \varphi, g) &= (\lambda, \mu) + g \circ F_1 + S_1 + S_2, \\ S_1(\theta, r) &= (\varphi \circ p_1 \circ F_1(\theta, r), 0) - D_{(\theta, r)}F_1(\varphi(\theta), 0), \\ S_2(\theta, r) &= (0, \psi \circ p_1 \circ F_1(\theta, r)) - D_{(\theta, r)}F_1(0, \psi(\theta)). \end{aligned}$$

En posant $a(\theta) = \partial_r(p_2 \circ F_1)(\theta, 0)$, on a :

$$\begin{aligned} S_1(\theta, 0) &= (\varphi(\theta + \alpha) - \varphi(\theta), 0), \\ S_2(\theta, 0) &= (0, \psi(\theta + \alpha) - a(\theta)\psi(\theta)). \end{aligned}$$

Donnons-nous $\eta_1 \in C^\infty(\mathbf{T}^{n+1}, \mathbf{R}^n)$, $\eta_2 \in C^\infty(\mathbf{T}^{n+1}, \mathbf{R})$, et considérons l'équation

$$D_0\Phi(\lambda, \mu, \psi, \varphi, g) = (\eta_1, \eta_2).$$

On doit avoir

$$\eta_1(\theta, 0) = \lambda + \varphi(\theta + \alpha) - \varphi(\theta)$$

ce qui détermine de façon unique $\lambda \in \mathbf{R}^n$, $\varphi \in C_2$.

On doit aussi avoir

$$\eta_2(\theta, 0) = \mu + \psi(\theta + \alpha) - a(\theta)\psi(\theta)$$

ce qui détermine de façon unique $\mu \in \mathbf{R}$, $\psi \in C_1$: on écrit $a(\theta) = a^* \frac{b(\theta)}{b(\theta + \alpha)}$, $a^* \neq 0$, $b > 0$, $\log b \in C_1$, puis on résout (cf.[H9])

$$b(\theta + \alpha)\eta_2(\theta, 0) - \mu b(\theta + \alpha) = b\psi(\theta + \alpha) - b\psi(\theta)a^*.$$

Finalement on doit avoir

$$g \circ F_1(\theta, r) = (\eta_1(\theta, r) - \eta_1(\theta, 0), \eta_2(\theta, r) - \eta_2(\theta, 0)),$$

ce qui détermine uniquement g .

On a ainsi montré que $D_0\Phi$ est inversible.

4.3. Rigidité du nombre de rotation

Le lemme ci-dessous est une étape cruciale dans l'application du théorème 4.1 qui sera donnée en 4.5.

Soient $0 < k < n$ deux entiers, U un ouvert de \mathbf{R}^k . Munissons $\mathbf{T}^{2n-k} \times U$ d'une forme symplectique constante ω ; notons $p_1 : \mathbf{T}^{2n-k} \times U \rightarrow \mathbf{T}^{2n-k}$ la projection canonique.

Soient $H \in C^\infty(\mathbf{T}^{2n-k} \times U)$ un hamiltonien, X_H le champ associé, $\lambda \in \mathbf{R}^{2n-k}$, $X_\lambda = \sum \lambda_i \partial / \partial \theta_i$, $Y = X_H + X_\lambda$.

Lemme : Supposons que :

1) Y soit tangent au graphe T d'une fonction lisse ψ de \mathbf{T}^{2n-k} dans U ;

2) on ait $p_{1*}(Y/T) = h_*(X_\alpha)$, où h est un difféomorphisme homotope à l'identité de \mathbf{T}^{2n-k} , et X_α un champ constant ergodique sur \mathbf{T}^{2n-k} .

Alors $\lambda - \alpha$ est ω -orthogonal à $\mathbf{R}^{2n-k} \times \{0\}$.

Démonstration. — Identifions \mathbf{T}^{2n-k} à $\mathbf{T}^{2n-k} \times \{0\}$; posons $j(\theta) = (\theta, \psi(\theta))$, $\omega_1 = \omega / \mathbf{T}^{2n-k}$, $\omega_2 = j^* \omega$.

Comme j est homotope à l'inclusion : $\mathbf{T}^{2n-k} \times \{0\} \hookrightarrow \mathbf{T}^{2n-k} \times \mathbf{R}^k$, les formes ω_1 et ω_2 sont cohomologues ; comme h est homotope à l'identité, ω_2 et $h^* \omega_2$ sont cohomologues. Les formes ω_1 et $h^* \omega_2$ sont à coefficients constants (puisque X_α est ergodique) et cohomologues, donc égales.

Posons $Y_1 = p_{1*}(Y/T)$. Le champ Y est symplectique, donc $i_Y \omega$ est fermée ; on a $j^*(i_Y \omega) = i_{Y_1} \omega_2$, et cette forme est cohomologue à $h^*(i_{Y_1} \omega_2) = i_{X_\alpha} \omega_1$. D'autre part $i_Y \omega = i_{X_H} \omega + i_{X_\lambda} \omega$, où $i_{X_H} \omega = dH$ est exacte, et $j^*(i_{X_\lambda} \omega)$ est cohomologue à $i_{X_\lambda} \omega_1$.

Donc $i_{X_\lambda} \omega_1$ et $i_{X_\alpha} \omega_1$ sont cohomologues ; étant à coefficients constants, elles sont égales, ce qu'on voulait démontrer.

4.4. L'hypothèse quasi-ergodique

L'hypothèse quasi-ergodique est une tentative (désespérée, on va le voir) d'assurer les fondements mathématiques de la mécanique statistique.

L'hypothèse ergodique de Boltzmann ("les moyennes temporelles sont égales aux moyennes spatiales"), en langage moderne (Birkhoff, Koopman

[B-K]) signifie que pour (presque) tout hamiltonien propre H sur une variété symplectique (M, ω) et (presque) toute valeur de l'énergie, le champ X_H est ergodique sur (chaque composante connexe de) ce niveau d'énergie. Cette hypothèse est mise en défaut par les théorèmes KAM : pour un ouvert de Hamiltoniens H et un ouvert de valeurs de l'énergie, on trouve sur chaque niveau d'énergie un ensemble de tores diophantiens lagrangiens invariants de mesure positive.

Au lieu d'exiger que le champ X_H soit ergodique sur chaque composante connexe de presque tout niveau d'énergie, on peut ne demander que la propriété (considérablement plus faible) suivante : que, pour un ensemble dense de valeurs de l'énergie, le champ X_H ait une orbite dense sur chaque composante connexe de ce niveau d'énergie. On obtient ainsi l'hypothèse quasi-ergodique ([Eh],[P]).

Cette hypothèse est trivialement vraie si $\dim M = 2$; si $\dim M = 4$, elle est fautive d'après les théorèmes KAM, puisque les tores diophantiens invariants séparent la surface d'énergie. Au numéro suivant, nous expliquerons, suivant Herman, pourquoi, sur certaines variétés symplectiques (non exactes), il existe un ouvert de hamiltoniens, et un ouvert de niveaux d'énergie pour lesquels un ensemble de Cantor de tores invariants diophantiens de codimension 1 exclut l'existence d'une orbite dense dans la surface d'énergie.

4.5. Un contre-exemple à l'hypothèse quasi-ergodique ([H7])

Munissons $\mathbf{T}^{2n} \times [-\delta, +\delta]^2$ (coordonnées $\theta_1 \dots \theta_{2n}, r_1, r_2$) de la forme symplectique constante ω_A associée à une matrice $A \in GL(2n + 2, \mathbf{R})$ antisymétrique de la forme :

$$A = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & B & {}^t\beta & {}^t\beta' \\ 0 & -\beta & 0 & 0 \\ -1 & -\beta' & 0 & 0 \end{pmatrix}$$

où $B \in M(2n - 1, \mathbf{R})$, $\beta' \in CD_{2n-1}$, $\beta \in \mathbf{R}^{2n-1}$, $\beta \neq 0$.

Soit $H_0 \in C^\infty([-\delta, +\delta]^2)$ un hamiltonien ne dépendant que de r_1, r_2 .

On suppose que

$$\begin{aligned} H_0(0, 0) &= \partial_{r_1} H_0(0, 0) = 0, \\ \partial_{r_2} H_0(0, 0) &\neq 0 \neq \partial_{r_1}^2 H_0(0, 0). \end{aligned}$$

THÉORÈME.— Soient $H \in C^\infty(\mathbf{T}^{2n} \times [-\delta, +\delta]^2)$ un hamiltonien suffisamment proche de H_0 , et b une valeur de l'énergie proche de 0. Le champ X_H , sur la surface d'énergie $\{H = b\}$ préserve un ensemble de Cantor de tores diophantiens de dimension $2n$, qui sont C^∞ voisins de $\mathbf{T}^{2n} \times (0, 0)$.

Esquisse de démonstration.— 1) On choisit $\gamma > 0, \tau > 0$ de façon que $s = 0$ soit point de densité de l'ensemble des $s \in \mathbf{R}$ tels que $\beta' - \beta s \in CD_{2n-1}(\gamma, \tau)$.

2) La surface d'énergie $\{H = b\}$, au voisinage de $\mathbf{T}^{2n} \times (0, 0)$, a pour équation

$$K(\theta, r_1, r_2) = r_2 - \eta(\theta, r_1) = 0.$$

Les champs X_H, X_K sur $\{H = b\}$ sont colinéaires. La projection sur $\{r_2 = 0\}$ de X_K est le champ Z sur $\mathbf{T}^{2n} \times [-\varepsilon, +\varepsilon]$ ($0 < \varepsilon < \delta$) défini par :

$$Z = (1, -B\nabla_\theta \eta - \beta \partial_{r_1} \eta + \beta', \langle \beta, \nabla_\theta \eta \rangle).$$

Le champ Z est une petite perturbation C^∞ du champ constant horizontal $(1, \beta', 0)$.

Le temps un du flot de Z préserve $\{\theta_1 = 0\}$ et y induit un difféomorphisme qu'on notera F .

On identifie $\mathbf{T}^{2n-1} \times \mathbf{R}$ à $\{0\} \times \mathbf{T}^{2n-1} \times \mathbf{R}$.

3) Pour $\lambda \in [-1, +1]^{2n-1}$, notons Z_λ le champ $Z + (0, \lambda, 0)$ et F_λ la restriction à $\mathbf{T}^{2n-1} \times \mathbf{R}$ du temps un de ce flot.

Faisons r_1 proche de 0 et s proche de 0 tel que $\alpha = \beta' - \beta s \in CD_{2n-1}(\gamma, \tau)$. D'après le théorème du tore translaté (4.1), il existe $\Lambda(\lambda) \in \mathbf{R}^{2n-1}$, $\mu(\lambda) \in \mathbf{R}$ tels que $\tilde{F}_\lambda = R_{\Lambda(\lambda), \mu(\lambda)} \circ F_\lambda$ préserve le graphe d'une fonction lisse de \mathbf{T}^{2n-1} dans \mathbf{R} , de moyenne r_1 , et y soit C^∞ -conjugué à la rotation R_α .

4) On vérifie aisément que $\lambda \mapsto \Lambda(\lambda)$ est un difféomorphisme au voisinage de l'origine, et qu'il existe donc un unique $\lambda_0 = \lambda_0(r_1, s)$ tel que $\Lambda(\lambda_0) = 0$.

On a $|\partial_{r_1}^2 \eta(\theta, 0)| > k > 0$; on peut donc choisir $r_1 = r_1(s)$ de façon que $\lambda_0(r_1(s), s)$ soit (euclidiennement) orthogonal à β .

5) Montrons que $\mu(\lambda_0(r_1(s), s)) = 0$.

Soit $\omega_{A'}$ la forme symplectique constante sur $\mathbf{T}^{2n-1} \times \mathbf{R}$ associée à :

$$A' = \begin{pmatrix} B & -\beta \\ \beta & 0 \end{pmatrix}.$$

Le difféomorphisme $\widehat{F} = F_{\lambda_0(r_1(s), s)}$ est exact symplectique (θ_1 est le paramètre d'une isotopie hamiltonienne reliant l'identité à \widehat{F}). Si $d\Omega$ est la forme volume associée à $\omega_{A'}$, la forme $\widehat{F}^*\Omega - \Omega$ est exacte. Mais alors, un tore translaté est en fait invariant.

6) Montrons finalement que $\lambda_0(r_1(s), s) = 0$. On applique le lemme de rigidité (4.3) à $Y = Z_{\lambda_0(r_1(s), s)}$: on conclut que le vecteur $(1, \alpha) - (0, \lambda_0(r_1(s), s))$ est $\omega_{A'}$ -orthogonal à $\mathbf{R}^{2n} \times (0, 0)$, donc combinaison linéaire de $(1, \beta')$ et $(0, \beta)$. Donc $\lambda_0(r_1(s), s)$ et β sont colinéaires. Comme il sont aussi orthogonaux, $\lambda_0(r_1(s), s) = 0$.

4.6. Tores invariants de codimension 1 dans un cadre non hamiltonien

4.6.1. Courbes non planaires

Soit $\varphi : [-\delta, +\delta] \rightarrow \mathbf{R}^n$ une application de classe C^∞ . On munit \mathbf{R}^n de la norme euclidienne. Soient $0 < a < b$. On dit que φ est (a, b) *non planaire* si, pour $|x| \leq \delta$, $v \in \mathbf{S}^{n-1}$, on a :

$$a \leq \sup_{1 \leq j \leq n} |\langle v, D^j \varphi(x) \rangle| \leq b.$$

En particulier, pour tout $x \in [-\delta, +\delta]$, les vecteurs $D\varphi(x), \dots, D^n\varphi(x)$ forment une base de \mathbf{R}^n .

Exemple. — $\varphi(t) = (t, t^2, \dots, t^n)$.

De nombreux mathématiciens russes, autour de Sprindzük, se sont intéressés à la question de savoir quelles approximations diophantiennes vérifient presque tout point d'une courbe non planaire. Nous n'avons ici besoin que du résultat élémentaire suivant, dû à Pyartli ([Py]).

PROPOSITION.— Soit $\tau > n^2 - 1$. Il existe une constante $C = C(a, b, \tau, \delta) > 0$ telle que, si $\varphi : [-\delta, +\delta] \rightarrow \mathbf{R}^n$ est (a, b) non planaire, on ait :

$$m\{x \in [-\delta, +\delta], \varphi(x) \notin CD_n(\gamma, \tau)\} \leq C\gamma^{1/n}.$$

Remarque.— Sprindzük conjecture (lorsque φ est analytique) que l'estimation précédente vaut encore pour tout $\tau > 0$. Même obtenir $\tau = o(n^2)$ semble difficile !

4.6.2. Considérons un plongement F_0 de $\mathbf{B}^n(\delta)$ dans \mathbf{B}^n (cf. 4.1) de la forme

$$F_0(\theta, r) = (\theta + \ell(r), r).$$

Supposons que, pour certaines constantes $0 < a < b$, l'application ℓ soit (a, b) non planaire.

Faisons $\tau > n^2 - 1$ et $\gamma > 0$ suffisamment petit. Soit $F : \mathbf{B}^n(\delta) \hookrightarrow \mathbf{B}^n$ un plongement proche de F_0 . Appliquons le théorème du tore translaté : pour $\alpha \in CD_n(\gamma, \tau)$ et $r \in [-\delta/2, +\delta/2]$, il existe une translation $R_{\lambda(r, \alpha), \mu(r, \alpha)}$ de \mathbf{B}^n et une fonction ψ sur \mathbf{T}^n de moyenne r telles que $R_{\lambda(r, \alpha), \mu(r, \alpha)} \circ F$ préserve le graphe $T_{\alpha, r}$ de ψ et y soit conjugué à R_α .

Pour $F = F_0$, on aurait $\mu_0(\alpha, r) \equiv 0$, $\lambda_0(\alpha, r) = \alpha - \ell(r)$. L'application $(\alpha, r) \mapsto \lambda(\alpha, r)$, définie sur $CD_n(\gamma, \tau) \times [-\delta/2, \delta/2]$, s'étend en une application de classe C^∞ de $\mathbf{R}^n \times [-\delta/2, \delta/2]$ dans \mathbf{R}^n (cf. 4.1), C^∞ -proche de λ_0 . L'équation $\lambda(\alpha, r) = 0$ définit alors implicitement une fonction $\ell_F : [-\delta/2, \delta/2] \rightarrow \mathbf{R}^n$, telle que $\lambda(\ell_F(r), r) = 0$. La fonction ℓ_F étant C^n proche de ℓ , elle est $(a/2, 2b)$ non planaire.

Si $r_0 \in [-\delta/2, +\delta/2]$ est tel que $\ell_F(r_0) \in CD_n(\gamma, \tau)$, le tore $T_{\ell_F(r_0), r_0}$ est verticalement translaté par F ; ceci se produit, d'après la proposition 4.6.1, pour un ensemble K de valeurs de r_0 qui a une grande mesure relative (pour γ petit) dans $[-\delta/2, \delta/2]$.

4.6.3. Difféomorphismes préservant le volume

Dans la situation précédente, supposons que F préserve le volume $d\Omega = d(r d\theta_1 \wedge \dots \wedge d\theta_n)$, et que le flux de F soit nul, c'est-à-dire que la forme $F^*\Omega - \Omega$ soit exacte. Chacun des tores translétés doit alors être invariant. On obtient donc :

THÉORÈME ([H7]).— *Pour tout plongement $F : \mathbf{B}^n(\delta) \hookrightarrow \mathbf{B}^n$, préservant le volume, de flux nul, suffisamment proche de F_0 , il existe un ensemble de Cantor $K \subset [-\delta/2, \delta/2]$, de mesure positive, et une application $r \mapsto \psi_r$ de K dans $C^\infty(\mathbf{T}^n)$ tels que :*

- (i) $\forall r \in K, \int_{\mathbf{T}^n} \psi_r dm = r; \|D\psi_r\| < \varepsilon ;$
- (ii) $\forall r \in K, \text{ le graphe de } \psi_r \text{ est un tore diophantien } T_r \text{ invariant par } F ;$
- (iii) *l'union des T_r est de mesure de Lebesgue positive dans $\mathbf{T}^n \times \mathbf{R}$.*

L'existence de ces tores invariants de codimension un, lorsque $n = 2$ et F est analytique, a été démontrée par Cheng et Sun ([C-S]). Un résultat similaire semble avoir été obtenu indépendamment (mais postérieurement) par De la Llave. Le résultat plus général et plus précis ci-dessus est dû à Herman ([H7]).

4.6.4. Plongeons $\mathbf{T}^n \times [-\delta, +\delta]$ dans un ouvert de carte d'une variété compacte arbitraire M^{n+1} . Considérons un difféomorphisme G_0 de M^{n+1} , dont la classe d'isotopie peut être prescrite arbitrairement, qui coïncide avec F_0 sur $\mathbf{T}^n \times \left[-\frac{3\delta}{4}, \frac{3\delta}{4}\right]$. Observons que $\mathbf{T}^n \times \{0\}$ sépare M .

Pour tout difféomorphisme G dans un voisinage U de G_0 dans $\text{Diff}^\infty(M)$, nous avons la remarquable alternative suivante :

— Soit G laisse invariant un ensemble de Cantor de tores invariants, dont l'union est de mesure de Lebesgue positive, sur lesquels il induit des rotations diophantiennes.

— Soit G n'est pas *transitif par chaînes*, c'est-à-dire qu'il existe des points x, x' de M qui ne sont dans la même orbite d'aucune C^0 -perturbation de G .

Si la perturbation G préserve une mesure de probabilité μ chargeant tous les ouverts, le second terme de l'alternative ne peut se produire. Observons qu'il est essentiel de considérer un voisinage de G_0 dans la C^∞ -topologie : Oxtoby a en effet montré ([O]) que, dans le groupe $\text{Homeo}_\mu(M)$ des homéomorphismes d'une variété compacte M préservant une mesure de probabilité μ (sans atomes, chargeant tous les ouverts), la propriété d'avoir une orbite dense est générique au sens de Baire.

Dans le cas où la mesure μ est équivalente à la mesure de Lebesgue, l'union des tores invariants est une partie de μ -mesure positive sur laquelle tous les exposants de Lyapunov de G sont nuls : ceci donne un contre-exemple à une conjecture de Pesin ([Pe]).

5. GÉOMÉTRIE ET DYNAMIQUE DES TORES INVARIANTS LAGRANGIENS

5.1. Le premier théorème de Birkhoff

Soit F un difféomorphisme exact symplectique, fortement monotone de \mathbf{A}^n (cf. 2.4). Un relèvement \widehat{F} de F à $\mathbf{R}^n \times \mathbf{R}^n$ s'écrit

$$\widehat{F}(\theta, r) = (\Theta, R) = (\theta + \varphi_1(\theta, r), \varphi_2(\theta, r)),$$

où φ_1, φ_2 sont \mathbf{Z}^n -périodiques en θ .

Écrivons $\widehat{F}^*v - v = dH$, et choisissons θ, Θ comme coordonnées globales sur $\mathbf{R}^n \times \mathbf{R}^n$; la fonction H , définie à une constante additive près, s'appelle la *fonction génératrice* de \widehat{F} . Elle vérifie

$$H(\theta + k, \Theta + k) = H(\theta, \Theta), \quad k \in \mathbf{Z}^n$$

$$\widehat{F}(\theta, r) = (\Theta, R) \iff \begin{cases} \partial_\theta H(\theta, \Theta) = -r \\ \partial_\Theta H(\theta, \Theta) = R. \end{cases}$$

DÉFINITION.— On dit que \widehat{F} (ou F) est globalement positif si H vérifie l'une des conditions équivalentes suivantes :

- (i) H est minorée :
- (ii) $\forall \theta \in \mathbf{R}^n, \quad \Theta \rightarrow H(\theta, \Theta)$ est minorée :
- (iii) $\forall \Theta \in \mathbf{R}^n, \quad \theta \rightarrow H(\theta, \Theta)$ est minorée :
- (iv) $\lim_{\|\theta - \Theta\| \rightarrow +\infty} \|\theta - \Theta\|^{-1} H(\theta, \Theta) = +\infty.$

DÉFINITION.— Le graphe d'une application $\psi \in C^0(\mathbf{T}^n, \mathbf{R}^n)$ est dit C^0 -lagrangien s'il existe $c \in \mathbf{R}^n, \eta \in C^1(\mathbf{T}^n)$ tels que $\psi = c + d\eta$ (de façon

équivalente, si ψ , vue comme section du fibré cotangent $T^*(\mathbf{T}^n)$, est fermée au sens des distributions).

THÉORÈME ([H6]).— Soit F un difféomorphisme globalement positif de \mathbf{A}^n . Supposons que F préserve le graphe C^0 -lagrangien d'une fonction $\psi \in C^0(\mathbf{T}^n, \mathbf{R}^n)$. Alors ψ est lipschitzienne ; de plus, si on pose $\widehat{F}(\theta, \psi(\theta)) = (\Theta, \psi(\Theta))$, on a, pour presque tout $\theta \in \mathbf{T}^n$:

$$\begin{aligned} D\psi(\theta) &\geq -\partial_{\theta}^2 H(\theta, \Theta), \\ D\psi(\Theta) &\leq \partial_{\Theta}^2 H(\theta, \Theta) \end{aligned}$$

(relation d'ordre entre matrices symétriques).

Esquisse de démonstration.— Posons $\widehat{\psi}(\theta) = \langle c, \theta \rangle + \eta(\theta)$ (cf. déf. ci-dessus) et notons f l'homéomorphisme de \mathbf{R}^n tels que

$$\widehat{F}(\theta, \psi(\theta)) = (f(\theta), \psi(f(\theta))).$$

Posons $K(\theta, \Theta) = H(\theta, \Theta) + \widehat{\psi}(\theta) - \widehat{\psi}(\Theta)$.

On a

$$\partial_{\theta} K(\theta, \Theta) = 0 \iff \Theta = f(\theta) \iff \partial_{\Theta} K(\theta, \Theta) = 0.$$

Comme $\lim_{\|\Theta - \theta\| \rightarrow +\infty} K(\Theta, \theta) = +\infty$, la variété critique $\{\Theta = f(\theta)\}$ réalise le *minimum* de K . Les distributions $\theta \mapsto \partial_{\Theta}^2 K|_{\Theta=f(\theta)}$ et $\Theta \mapsto \partial_{\theta}^2 K|_{\theta=f^{-1}(\Theta)}$, à valeurs dans les matrices symétriques, sont donc *positives*. Le théorème en résulte.

5.2. Deuxième théorème de Birkhoff (version perturbative)

La grassmannienne lagrangienne $\Lambda(n)$ de $\mathbf{C}^n = \mathbf{R}^n \oplus i\mathbf{R}^n \simeq T^*\mathbf{R}^n$ s'identifie à $U(n)/O(n)$, le point base étant le sous-espace horizontal $h_0 = \mathbf{R}^n \times \{0\}$. L'application $\delta : U(n)/O(n) \rightarrow S^1$

$$\delta : u \mapsto (\det_{\mathbf{C}} u)^2$$

induit des isomorphismes : $\pi_1(\Lambda(n)) \rightarrow \pi_1(S^1)$, $H^1(S^1, \mathbf{Z}) \rightarrow H^1(\Lambda(n), \mathbf{Z})$.

DÉFINITION.— Soit $j : N \rightarrow \mathbf{A}^n$ une immersion lagrangienne. L'indice de Maslov de (N, j) est la classe $J^*(\tau) \in H^1(N, \mathbf{Z})$, où τ est le générateur canonique de $H^1(\Lambda(n), \mathbf{Z})$ et $J : x \mapsto Tj(T_x N)$ est l'application de N dans $\Lambda(n)$ déduite de j .

THÉORÈME (Viterbo [V]).— Un tore lagrangien plongé homotope à $\{r = 0\}$ dans \mathbf{A}^n a un indice de Maslov nul.

DÉFINITION.— Soit f une application continue d'un espace métrique (X, d) dans lui-même. Un point $x \in X$ est récurrent par chaînes si pour tout $\varepsilon > 0$ il existe $n > 0$ et $x = x_0, x_1, \dots, x_n = x$ tels que $d(fx_i, x_{i+1}) < \varepsilon$ pour $0 \leq i < n$. L'application f est récurrente par chaînes si tout point est récurrent par chaînes.

Pour $R > 0$ posons $\mathbf{A}^n(R) = \mathbf{T}^n \times [-R, +R]^n$.

THÉORÈME ([H6]).— Soit $L(\theta, r) = (\theta + \ell(r), r)$ un difféomorphisme symplectique complètement intégrable de $\mathbf{A}^n(R)$. On suppose que $D\ell(r) > 0$ pour tout $r \in [-R, +R]^n$.

Soit $\varepsilon > 0$. Tout plongement symplectique $F : \mathbf{A}^n(R) \hookrightarrow \mathbf{A}^n$, C^1 suffisamment voisin de L , a la propriété suivante : si $N \subset \mathbf{A}^n(R)$ est une sous-variété compacte, connexe, lagrangienne, d'indice de Maslov nul, invariante par F , et si F/N est récurrent par chaînes, alors N est un tore, graphe d'une application $\psi \in C^1(\mathbf{T}^n, \mathbf{R}^n)$ vérifiant $\|D\psi\| < \varepsilon$.

Le point crucial de la démonstration est le suivant : notons $\tilde{L} : \mathbf{A}^n(R) \times \tilde{\Lambda}(n) \rightarrow \mathbf{A}^n(R) \times \tilde{\Lambda}(n)$ (où $\tilde{\Lambda}(n)$ est le revêtement universel de $\Lambda(n)$) l'application déduite de TL ; si $(z, \tilde{h}) \in \mathbf{A}^n(R) \times \tilde{\Lambda}(n)$ est récurrent par chaînes pour une petite perturbation C^0 de \tilde{L} , alors la projection de \tilde{h} dans $\Lambda(n)$ est proche du plan horizontal h_0 .

Pour des généralisations non perturbatives du deuxième théorème de Birkhoff, voir [Bi],[B-P2],[Pol].

5.3. Dynamique générique ([H5])

Considérons comme précédemment, un difféomorphisme symplectique

complètement intégrable de $\mathbf{A}^n(R)$:

$$L(\theta, r) = (\theta + \ell(r), r)$$

et supposons encore qu'on a $D\ell(r) > 0$ pour tout $r \in [-R, +R]^n$.

Pour un plongement symplectique $F : \mathbf{A}^n(R) \hookrightarrow \mathbf{A}^n$, C^1 -proche de L , notons $IT^\infty(F)$ l'ensemble des tores invariants lisses diophantiens qui sont homotopes à $\{r = 0\}$. Un tel tore est automatiquement lagrangien (cf. 2.7) et son indice de Maslov est nul d'après le théorème de Viterbo. D'après le théorème 5.2 ci-dessus, il existe donc un voisinage \mathcal{U} de L dans la C^1 -topologie tel que, pour $F \in \mathcal{U}$, tout tore T de $IT^\infty(F)$ est graphe d'une fonction $\psi_T \in C^\infty(\mathbf{T}^n, [-R, +R]^n)$ vérifiant $\|D\psi_T\| \leq 1$.

Munissons l'espace S des applications lipschitziennes de \mathbf{T}^n dans $[-R, +R]^n$, de constante de Lipschitz ≤ 1 , de la C^0 -topologie. L'espace S est compact. Considérons, via l'application $T \rightarrow \psi_T$, l'ensemble $IT^\infty(F)$ comme une partie de S , et notons $IT(F)$ son adhérence dans S . Pour tout $\psi \in IT(F)$, le graphe T_ψ de ψ est un tore invariant par F , C^0 -lagrangien.

Pour tout $F \in \mathcal{U}$, et tout $\psi \in IT(F)$, la restriction F/T_ψ est *récurrente par chaînes*: en effet T_ψ est limite de tores diophantiens, et la propriété d'être récurrent par chaînes est fermée.

Pour tout $F \in \mathcal{U}$, et un ensemble résiduel de $\psi \in IT(F)$:

- l'application ψ est de classe C^1 ;
- la restriction de F à T_ψ est minimale, uniquement ergodique, et son nombre de rotation est un nombre de Liouville.

Finalement, pour une application générique $F \in \mathcal{U}$, et un ensemble résiduel de $\psi \in IT(F)$, la restriction de F à T_ψ a une entropie topologique nulle, n'est pas mélangeante, est faiblement mélangeante, et son unique mesure invariante est singulière et même supportée par un borélien de dimension de Hausdorff 0.

Ces dernières propriétés, spécialement le mélange faible, montrent combien la dynamique sur le tore générique diffère de celle d'une rotation !

Nous renvoyons [H5] pour des propriétés génériques supplémentaires et pour les démonstrations, qui montrent toute la puissance des méthodes basées sur la catégorie de Baire.

Remarque.— Un tore lagrangien invariant par le difféomorphisme symplectique générique d'une variété symplectique M^{2n} peut-il contenir une orbite périodique? Il est facile de voir que non si $n = 1$; M.C.Arnaud ([A]) a prouvé que c'est impossible pour $n = 2$. La question est ouverte pour $n > 2$.

5.4. Torsion indéfinie ([H4])

5.4.1. Les hypothèses sur F dans les numéros précédents 5.2, 5.3 sont en particulier satisfaites par tout difféomorphisme symplectique au voisinage d'un tore lisse invariant lagrangien diophantien, quand la torsion de ce tore est *positive* (cf. 2.7) : cela résulte de la forme normale de Birkhoff. Un énoncé analogue vaut bien sûr également lorsque la torsion est *négative*.

Lorsqu'on perturbe un difféomorphisme complètement intégrable dont la torsion est dégénérée ou indéfinie, la situation change dramatiquement.

5.4.2. Exemple

Considérons

$$L(\theta_1, \theta_2, r_1, r_2) = (\theta_1 + r_2, \theta_2 + r_1, r_1, r_2)$$

$$F(\theta_1, \theta_2, r_1, r_2) = (\theta_1 + r_2, \theta_2 + r_1, r_1 + \varepsilon\varphi(\theta_1 + r_2), r_2),$$

où

$$0 < \varepsilon \ll 1, \quad \varphi \in C^\infty(\mathbf{T}^1), \quad \int_{\mathbf{T}^1} \varphi dm = 0.$$

Pour tout $\alpha \in \mathbf{R}$, la sous-variété $M_\alpha = \{r_2 = \alpha\}$ est invariante par F .

Pour $n \geq 0$, on a : $F^n(\theta_1, \theta_2, r_1, \alpha) = (\Theta_1^{(n)}, \Theta_2^{(n)}, R_1^{(n)}, \alpha)$ avec

$$\begin{cases} \Theta_1^{(n)} = \theta_1 + n\alpha \\ \Theta_2^{(n)} = \theta_2 + nr_1 + \varepsilon \sum_1^n (n-i)\varphi(\theta_1 + i\alpha) \\ R_1^{(n)} = r_1 + \varepsilon \sum_1^n \varphi(\theta_1 + i\alpha) \end{cases}$$

1er cas : l'équation

$$\psi(\theta_1 + \alpha) - \psi(\theta_1) = \varphi(\theta_1 + \alpha)$$

admet une solution $\psi \in C^0(\mathbf{T}^1)$, $\int_{\mathbf{T}^1} \psi dm = 0$. C'est le cas (avec $\psi \in C^\infty(\mathbf{T}^1)$) si α est diophantien. Les tores T_c d'équation $\{r_1 = \varepsilon\psi(\theta_1) + c, r_2 = \alpha\}$ sont invariants par F et leur union est M_α . Sur T_c , F est C^0 -conjugué à la rotation $R_{\alpha,c}$ si et seulement si l'équation

$$\eta(\theta_1 + \alpha) - \eta(\theta_1) = \psi(\theta_1)$$

a une solution $\eta \in C^0(\mathbf{T}^1)$.

2e cas : l'équation

$$\psi(\theta_1 + \alpha) - \psi(\theta_1) = \varphi(\theta_1 + \alpha)$$

n'a pas de solution $\psi \in C^0(\mathbf{T}^1)$. D'après Gottschalk et Hedlund, pour tout $(\theta_1, \theta_2, r_1, \alpha) \in M_\alpha$, on a

$$\sup_{n \geq 1} |R_1^{(n)}| = +\infty.$$

Supposons qu'une infinité de coefficients de Fourier de φ soient non nuls :

— pour presque tout $\alpha \in \mathbf{R}$, M_α est union de tores invariants lisses lagrangiens,

— pour un ensemble G_δ -dense de α , l'orbite de tout point de M_α n'est pas bornée.

Il est facile de vérifier qu'aucune inégalité a priori ne peut être satisfaite par les tores invariants lagrangiens. En conjuguant F par un difféomorphisme exact symplectique proche de l'identité, on peut obtenir des tores invariants lisses lagrangiens diophantiens, homotopes à $\{r_1 = r_2 = 0\}$, qui ne sont pas des graphes.

5.4.3. Instabilité de la torsion

Si F est une perturbation d'un difféomorphisme symplectique complètement intégrable L à torsion *positive* comme en 5.2, 5.3, il n'est pas difficile de voir que tout tore invariant $T \in IT^\infty(F)$ a encore une torsion positive ; c'est en particulier le cas lorsque F décrit (via la forme normale de Birkhoff) la dynamique d'un difféomorphisme symplectique au voisinage d'un tore lagrangien diophantien invariant de torsion positive.

La situation est radicalement différente si la torsion est indéfinie. Soient F un difféomorphisme symplectique d'une variété symplectique (M, ω) , et T un tore diophantien lagrangien (lisse) invariant par F . Supposons que la torsion de F le long de T soit dégénérée ou indéfinie.

THÉORÈME.— *On peut perturber, dans la C^∞ -topologie, F en un difféomorphisme \tilde{F} de façon que \tilde{F} préserve un tore diophantien lagrangien \tilde{T} C^∞ -proche de T , et que la torsion de \tilde{F} le long de \tilde{T} ait une signature, non identiquement nulle, arbitrairement prescrite.*

On ne sait qu'il est possible d'annuler complètement la torsion par petite perturbation C^∞ .

5.4.4. Instabilité de la géométrie des tores

Une partie de la pathologie de l'exemple 5.4.2 persiste, sous l'hypothèse de torsion indéfinie, dans un cadre plus général.

Considérons un difféomorphisme symplectique complètement intégrable de \mathbf{A}^n $L(\theta, r) = (\theta + \ell(r), r)$. Supposons que $D\ell(0)$ soit indéfinie, ou dégénérée ; soit $v \in \mathbf{R}^n, v \neq 0$ un vecteur tel que $\langle v, D\ell(0)v \rangle = 0$.

Quitte à perturber L (parmi les difféomorphismes symplectiques complètement intégrables), on peut supposer que $\ell(0) \in \mathbf{Q}^n, v \in \mathbf{Q}^n, \langle v, D_0^2 \ell(v, v) \rangle \neq 0$. À changement de coordonnées linéaire et revêtement fini près, on se ramène à $\ell(0) = 0, v = (1, 0, \dots, 0)$.

La fonction $\hat{\ell}(r)$ telle que $D\hat{\ell} = \ell$ satisfait alors :

$$\hat{\ell}(r_1, 0) = ar_1^3 + O(r_1^4), \quad a \neq 0.$$

Considérons le hamiltonien :

$$H(\theta, r) = \hat{\ell}(r) - \varepsilon \cos 2\pi\theta_1 + \sum_1^n \beta_i r_i,$$

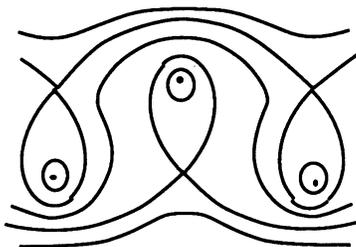
où $0 < |\beta_1|, \dots, |\beta_n|, |\varepsilon| \ll 1$.

Le champ associé X_H possède n intégrales en involution : r_2, \dots, r_n, H . Lorsqu'on fixe les valeurs $r_2 = \dots = r_n = 0$ des $n-1$ premières intégrales, et qu'on fait varier la valeur de la dernière intégrale H , les variétés invariantes

obtenues sont produit de \mathbf{T}^{n-1} (coordonnées $\theta_2, \dots, \theta_n$) par les courbes de niveau de la fonction : $(\theta_1, r_1) \mapsto \widehat{\ell}(r_1, 0) - \varepsilon \cos 2\pi\theta_1 + \beta_1 r_1$. On a représenté les courbes de niveau sur la figure ci-dessous (pour $a > 0, \beta_1 < 0$).

Le temps un du flot de X_H est une petite perturbation F de L (qui correspond à $\varepsilon = \beta_1 = \dots = \beta_n = 0$) dans la C^∞ -topologie. Le difféomorphisme F possède des tores invariants lagrangiens lisses diophantiens (pour un bon choix des β_i) qui ne sont pas des graphes, et qui s'accumulent même, pour la topologie de Hausdorff, sur un compact invariant qui n'est pas une variété.

Qui plus est, on peut faire en sorte qu'un tel tore ait à nouveau une torsion indéfinie, ce qui permet d'itérer la construction précédente...



BIBLIOGRAPHIE

- [A.M.] R. ABRAHAM, J. MARSDEN - *Foundations of Mechanics*, 2nd edition, Benjamin Cummings, Reading, Mass.(1970).
- [An] S.B ANGENENT - *Monotone recurrence relations, their Birkhoff orbits and topological entropy*, Erg. Th. Dyn. Sys. 1 (1990) 15-41.
- [A] M.-C. ARNAUD - *Points périodiques et tores lagrangiens invariants des difféomorphismes symplectiques*, Thèse Univ. Paris VII (1990).
- [Ar1] V.I. ARNOLD - *Méthodes mathématiques de la mécanique classique*, Mir, Moscou (1974).
- [Ar2] V.I. ARNOLD - *Small denominators I: On the mapping of a circle into itself*, Trans l. Am. Math. Soc., série 2, 46 (1965) 213-284.

- [Ar3] V.I. ARNOLD -*Small denominators II: Proof of a theorem of A.N. Kolmogorov on the invariance of quasiperiodic motion under small perturbation of the hamiltonian*, Russ. Math. Surv., 18, 5 (1963) 9-36.
- [Ar4] V.I. ARNOLD (editor) -*Dynamical Systems (III)*, E.M.S 3, Springer Verlag (1988).
- [Bi] M.L BIALYI -*Aubry-Mather sets and Birkhoff's theorem for geodesic flows on the two-dimensional torus*, Comm. Math. Phys. 126 (1989) 13-24.
- [B-P1] M.L BIALYI, L.V.POLTEROVICH -*Geodesic flows on the two-dimensional torus and phase-transitions "commensurability-non commensurability"*, Func. Anal. Appl. 20 (1986) 260-266.
- [B-P2] M.L BIALYI, L.V.POLTEROVICH -*Hamiltonian diffeomorphisms and Lagrangian distributions*, preprint (1991).
- [B1] G.D.BIRKHOFF -*Surface transformations and their dynamical applications*, Acta Math. 43 (1920), 1-119.
- [B2] G.D.BIRKHOFF -*Note sur la stabilité*, J. Math., 15 (1936) 339-344. Coll. Math. Papers, vol.2, 662-667.
- [B-K] G.D. BIRKHOFF, B.O. KOOPMAN -*Recent contributions to the ergodic theory*, Proc. Nat. Ac. Sc. 18 (1932), 279-282. Coll. Math. Papers, vol.2, 462-466.
- [B-H] P. BOYLAND, G.R. HALL -*Invariant circles and the order structure of periodic orbits in monotone twist maps*,Topology 26 (1987), 21-35.
- [Bo] J.B. BOST -*Tores invariants des systèmes dynamiques hamiltoniens*, Sem.Bourbaki n° 639, Astérisque 133-134 (1986), 113-157.
- [Ch] A. CHENCINER -*La dynamique au voisinage d'un point fixe elliptique conservatif : de Poincaré et Birkhoff à Aubry et Mather*, Sémin. Bourbaki n° 622, Astérisque 121-122 (1985), 147-170.
- [C-S] C.Q. CHENG, Y.S. SUN -*Existence of invariant tori in three-dimensional measure preserving mappings*, Cel. Mech. 47 (1990), 275-293.
- [D] R. DOUADY -*Stabilité et instabilité des points fixes elliptiques*, Ann. Sc. E.N.S.,4^e série 21 (1988), 1-46.
- [Eh] P. EHRENFEST, T. EHRENFEST -*The conceptual foundations of statistical mechanics* (1912), Engl. trans. repr., Dover publ. (1990).

- [E] H. ELIASSON - *Invariant Tori for Hamiltonian systems*, Ann. Sc. Norm. Pisa 15 (1988), 115-147.
- [F] A. FATHI - *Une interprétation plus topologique de la démonstration du théorème de Birkhoff*, App. au Ch.1 de [H9, vol.1].
- [G] C. GUTTIEREZ - *A counterexample to a C^2 closing lemma*, Erg. Th. Dyn. Sys. 7 (1987), 509-530.
- [Ha1] R.S. HAMILTON - *The inverse function theorem of Nash and Moser*, preprint (1974).
- [Ha2] R.S. HAMILTON - *The inverse function theorem of Nash and Moser*, Bull. A. M. S. 7 (1982), 65-222.
- [H1] M.R. HERMAN - *Exemple de flots hamiltoniens dont aucune perturbation en topologie C^∞ n'a d'orbites périodiques sur un ouvert de surfaces d'énergie*, C.R.A.S, t. 312, série I (1991), 989-994.
- [H2] M.R. HERMAN - *Différentiabilité optimale et contre-exemples à la fermeture en topologie C^∞ des orbites récurrentes des flots hamiltoniens*, C.R.A.S, t. 313, série I (1991), 49-51.
- [H3] M.R. HERMAN - *Stabilité topologique des systèmes dynamiques conservatifs*, à paraître dans Atas do 18° Col. Bras. Mat.(1992).
- [H4] M.R. HERMAN - *Dynamics connected to indefinite torsion*, à paraître dans Proc. Conf. "Twist maps and its applications", I.M.A, Univ. Minnesota, Springer Verlag.
- [H5] M.R. HERMAN - *On the dynamics on Lagrangian tori invariant under symplectic diffeomorphisms*, preprint École Polytechnique (1990).
- [H6] M.R. HERMAN - *Inégalités a priori pour des tores lagrangiens invariants par des difféomorphismes symplectiques*, Publ. Math. I.H.E.S. 70 (1989), 47-101.
- [H7] M.R. HERMAN - *Théorème des tores translétés et quelques applications à la stabilité topologique des systèmes dynamiques conservatifs*, en préparation.
- [H8] M.R. HERMAN - *Sur la conjugaison différentiable des difféomorphismes du cercle à des rotations*, Publ. Math. I.H.E.S. 49 (1979), 5-233.
- [H9] M.R. HERMAN - *Sur les courbes invariantes par les difféomorphismes de l'anneau*, Astérisque, vol.1, 103-104 (1983), vol.2, 144 (1986).

- [H10] M.R. HERMAN -*Démonstration du théorème des courbes translitées par les difféomorphismes de l'anneau ; démonstration du théorème des tores invariants*, manuscrit (1980).
- [Ho] L. HORMANDER -*The boundary problem of physical geodesy*, Arch. Rat. Mech. Anal. 62 (1976), 1-52.
- [K] A.N. KOLMOGOROFF -*Théorie générale des systèmes dynamiques et mécanique classique* (en russe), Proc.1954, Int. Cong., North Holland, Amsterdam, 315-333.
- [La] V.F. LAZUTKIN -*The existence of caustics for a billiard problem in a convex domain*, Math. USSR Izv. 7 (1973), 185-214.
- [L1] P. LE CALVEZ -*Propriétés dynamiques des zones d'instabilité*, Ann. Sc. E.N.S., 4^e série, 20 (1987), 443-464.
- [L2] P. LE CALVEZ -*Propriétés générales des applications déviant la verticale*, Bull. S.M.F., 117 (1989), 69-102.
- [L3] P. LE CALVEZ -*Étude topologique des applications déviant la verticale*, Ensaies Mat., Vol. 2 (1990).
- [L4] P. LE CALVEZ -*Sur les difféomorphismes de l'anneau et du tore*, soumis à Astérisque.
- [M1] J. MATHER -*The existence of quasiperiodic orbits for twist homeomorphisms of the annulus*, Topology 21 (1982), 457-467.
- [M2] J. MATHER -*Minimal measures*, Comm. Math. Helv. 64 (1989), 375-394.
- [M3] J. MATHER -*Variational construction of orbits of twist diffeomorphisms*, preprint (1990).
- [M4] J. MATHER -*A criterion for the non existence of invariant circles*, Publ. Math. I.H.E.S. 63 (1986), 153-204.
- [M5] J. MATHER -*Destruction of invariant circles*, Erg. Th. Dyn. Sys. 8 (1988), 199-214.
- [Mo1] J. MOSER -*On invariant curves of area preserving mappings of an annulus*, Nachr. Ak. Wiss. Gott., Math. Phys. Kl. (1962), 1-20.
- [Mo2] J. MOSER -*A rapidly convergent iteration method and nonlinear partial differential equations 2*, Ann. Sc. Norm. Pisa 20 (1966), 499-535.

- [O] J.C. OXTOBY - *Note on transitive transformations*, Proc. Nat. Ac. Sc. (1937), 443-446.
- [Pe] Y.B. PESIN - *Characteristic Lyapunov exponents and smooth ergodic theory*, Russ. Math. Surv., 92, 4 (1977), 60.
- [P] H. POINCARÉ - *Les méthodes nouvelles de la mécanique céleste*, 3 volumes, Gauthier-Villars, Paris (1892).
- [Pol] L.V. POLTEROVICH - *The second Birkhoff theorem for optical Hamiltonian systems*, à paraître dans Proc. A.M.S.
- [Po] J. POSCHEL - *Integrability of Hamiltonian systems on Cantor sets*, Comm. Pure Appl. Math. 35 (1982), 653-695.
- [Pu] C. PUGH - *The closing Lemma*, Amer. J. Math. 89 (1967), 956-1009.
- [P-R] C. PUGH, C. ROBINSON - *The C^1 closing lemma including Hamiltonians*, Erg. Th. Dyn. Sys. 3 (1983), 261-314.
- [Py] A.S. PYARTLI - *Diophantine approximation on submanifolds of euclidean space*, Funct. Anal. App. 3 (1969), 303-306.
- [R1] H. RUSSMANN - *Kleine Nenner I: Über invarianten Kurven differenzierbarer Abbildungen eines Kreisringes*, Nachr. Ak. Wiss. Gott., Math. Phys. Kl. (1970), 67-105.
- [R2] H. RUSSMANN - *On the existence of invariant curves of twist mappings of an annulus*, Lect. Notes in Math. 1007 (1983), 677-712.
- [R3] H. RUSSMANN - *Non-degeneracy in the perturbation theory of integrable dynamical systems*, in Number Theory and Dynamical Systems, Dodson-Vickers (eds), L.N. 134, Lond. Math. Soc. (1989), 1-18.
- [S-Z] D. SALAMON, E. ZEHNDER - *KAM theory in configuration space*, Comm. Math. Helv. 64 (1989), 84-142.
- [V] C.VITERBO - *A new obstruction to embedding Lagrangian tori*, Inv. Math. 100 (1990), 301-320.
- [Y] J.-C. YOCCOZ - *Conjugaison différentiable des difféomorphismes du cercle dont le nombre de rotation vérifie une condition diophantienne*, Ann. Sc. E.N.S., 4^e série 17 (1984), 333-359.
- [Z1] E. ZEHNDER - *Remarks on periodic solutions on hypersurfaces*, in "Periodic solutions of Hamiltonian systems and related topics", P.H. Rabinowitz et coll., D. Reidel Co., Dordrecht (1987), 91-140.

J.-C. YOCCOZ

- [Z2] E. ZEHNDER - *Generalized implicit function theorems with application to some small divisors problems I*, Comm. Pure Appl. Math. 28 (1975), 91-140.

Jean-Christophe YOCCOZ
Université de Paris-Sud
Département de Mathématiques
Bâtiment 425
F-91405 ORSAY CEDEX

Astérisque

MARTIN BARLOW

Harmonic analysis on fractal spaces

Astérisque, tome 206 (1992), Séminaire Bourbaki,
exp. n° 755, p. 345-368

<http://www.numdam.org/item?id=SB_1991-1992__34__345_0>

© Société mathématique de France, 1992, tous droits réservés.

L'accès aux archives de la collection « Astérisque » (<http://smf4.emath.fr/Publications/Asterisque/>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

HARMONIC ANALYSIS ON FRACTAL SPACES

par Martin BARLOW

1. INTRODUCTION

The initial interest in this area came from mathematical physicists in the early 1980s, who were studying *transport properties* of disordered media. As there are good reasons for believing that fractals can provide a good model for such media, this led to their interest in such questions on the study of the wave and heat equations on fractal spaces – see for example [RT].

Mathematical work has, so far, largely centered on the more easily treated Laplace and heat equations. It began with a probabilistic treatment, by Kusuoka [K1], Goldstein [G] and Barlow–Perkins [BP], but an analytic treatment, making a substantial use of Dirichlet forms has been developed, mainly in Japan – see [Kig1, Kig2, F2, K2].

Most of the problems and difficulties arise, naturally, on the microscopic scale, from absence of any kind of Euclidean structure. However, given a regular fractal, such as the Sierpinski gasket, one can define a *pre-fractal* manifold or graph, whose large scale structure mimics that of the true fractal. These pre-fractals are entirely classical, but classical techniques fail to give, for example, the right bounds on the heat kernel on such an object.

Let me now be more specific about the kinds of problem that will be considered.

Let $F \subseteq \mathbb{R}^d$ be a connected self-similar fractal, let $d_f(F)$ be the Hausdorff dimension of F , and let μ_F be Hausdorff $x^{d_f(F)}$ -measure on

F . The heat equation on F should take the form

$$\Delta_F u = \frac{\partial u}{\partial t}, \quad u(x, 0) = u_0(x), \quad x \in F$$

where $u : F \times \mathbb{R}_+ \rightarrow \mathbb{R}$, $u_0 \in C(F)$, and Δ_F is a Laplacian operator acting on a subspace $\mathcal{D}(\Delta_F) \subset C(F)$. The following problems arise immediately:

- (i) *Existence.* The construction of a suitable operator Δ_F which is F -isotropic, that is, locally invariant with respect to the local isometries of F .
- (ii) *Uniqueness.* Is Δ_F characterised by the property of being F -isotropic?
- (iii) *Properties.* The properties of the solutions to the Laplace and heat equation associated with Δ_F , and the form of the spectrum of Δ_F .

To these we add questions about the Markov process X_t with generator Δ_F and semigroup $P_t = \exp(t\Delta_F)$.

2. DEFINITION OF THE DIRICHLET FORM

The first work was done on what appears to be the simplest non-trivial connected fractal, the Sierpinski gasket. Later works have extended many of these results to larger classes of sets: to *nested fractals* ([L]), *p.c.f. self-similar sets* ([Kig2]), and a still more general class which includes the Sierpinski carpets ([KZ]). However in this survey I will for the most part restrict myself to a subset of nested fractals which is large enough to capture the essential features of the subject. Indeed, it seems clear that more general sets exhibit the same kind of behaviour as the simpler ones: it is just that the results are more difficult to prove.

Let $F_0 = \{a_1, \dots, a_k\}$ be the vertices of a regular k -sided polygon in \mathbb{R}^2 , and let H_0 be the closed convex hull of F_0 . Let $\lambda \geq 1$, $M \geq k$, $a_{k+1}, \dots, a_M \in \mathbb{R}^2$ and let ψ_i , $1 \leq i \leq M$ be defined by

$$\psi_i(x) = a_i + \lambda^{-1}(x - a_i).$$

We assume:

- (A1) (Symmetry) The set $\{a_i, 1 \leq i \leq M\}$ has the same symmetries as F_0 ,
- (A2) (Nesting) $\psi_i(H_0) \subseteq H_0$,
- (A3) (Connectedness) The set $H_1 = \bigcup_{i=1}^M \psi_i(H_0)$ is connected.
- (A4) (Open set condition) $\{\text{Int}(\psi_i(H_0)), 1 \leq i \leq M\}$ are disjoint.
- (A5) (Finitely ramified) If $x, y \in H_0$ and $\psi_i(x) = \psi_j(y)$ for $i \neq j$ then $x, y \in F_0$.

Remark. Note that (A5) rules out fractals such as the Sierpinski carpet. This assumption is in fact very strong, and, as we will see, enables the different ‘levels’ of the fractal to be treated separately. Most of the work done so far has used this hypothesis in an essential way. Nevertheless, it is possible to handle the Sierpinski carpet, at least in two–dimensions, (see [BB1, BB2, BB3], [KZ]), at the cost of working rather harder.

Let $\Psi(\cdot) = \bigcup_{i=1}^M \psi_i(\cdot)$; and set $F = \bigcap_{n=0}^{\infty} \Psi^n(H_0)$, $F_n = \Psi^n(H_0)$. The set F is a *nested fractal* with dimension $d_f(F) = \log M / \log \lambda$, and is approximated by the finite sets F_n .

It is helpful to introduce co–ordinates via the associated abstract sequence spaces. Let $I = \{1, \dots, M\}$, and for $w = (i_1, \dots, i_n) \in I^n$ let

$$\psi_w = \psi_{i_1, i_2, \dots, i_n} = \psi_{i_1} \circ \dots \circ \psi_{i_n};$$

we will call the sets $\psi_w(F_0), w \in I^n$, *n–cells*, and the sets $\psi_w(H_0), w \in I^N$, *n–complexes*. For $w \in I^N$ write $w|n = (w_1, \dots, w_n) \in I^n$, and note that $\psi_w(H_0) = \bigcap_{n=0}^{\infty} \psi_{w|n}(H_0)$ consists of a single point, $\phi(w)$ say. It is easily checked that the map $\phi : I^N \rightarrow F$ is continuous and surjective, but not injective. In the case of the Sierpinski gasket, for example, we have $\phi(1\dot{2}) = \phi(2\dot{1})$. In fact, one can define a fractal in an entirely abstract way, by considering the set I^N under a quotient map which identifies suitable sequences of this kind – and this is done in [Kig2].

We now proceed to define Dirichlet forms on the finite sets F_n .

Let $\Lambda = \{\alpha \in \mathbb{R}^{k-1} : \alpha_i = \alpha_{k-i}, 1 \leq i \leq k-1\}$, and define, for $x, y \in F_n, \alpha \in \Lambda$,

$$\alpha_n(\alpha, x, y) = \begin{cases} \alpha_i, & \text{if } x, y \text{ belong to the same } n\text{-cell and are} \\ & i \text{ steps apart on the circumference,} \\ 0, & \text{otherwise,} \end{cases}$$

$$\mu_n(\alpha, x) = \sum_y a_n(\alpha, x, y).$$

Note that $\mu_n(\alpha, F_n) = M^n \sum \alpha_i$, and that $\mu_n(\alpha, x)$ depends on α only through $\sum \alpha_i$. Set, for $f \in B(F_n) = \mathbb{R}^{F_n}$,

$$(2.1) \quad \mathcal{E}_\alpha^n(f, f) = \sum_x \sum_y a_n(\alpha, x, y) (f(x) - f(y))^2.$$

We may interpret the Dirichlet form \mathcal{E} in a number of ways.

1. If we regard $(F_n, a_n(\alpha))$ as an electric network where the wire between x and y has conductivity $a_n(\alpha, x, y)$ then $\mathcal{E}_\alpha^n(f, f)$ is the energy dissipation if a potential f is maintained on the network.
2. The discrete Laplacian Δ_α^n on F_n is defined by

$$\Delta_\alpha^n f(x) = \mu_n(\alpha, x)^{-1} \sum_y a_n(\alpha, x, y) (f(y) - f(x)),$$

and, writing (f, g) for the inner product with respect to $\mu(\alpha, \cdot)$, Δ_α^n satisfies

$$\mathcal{E}_\alpha^n(f, g) = -(\Delta_\alpha^n f, g), \quad f, g \in C(F_n).$$

The random walk $X_r^n, r \geq 0$ with generator Δ_α^n and Dirichlet form \mathcal{E}_α^n is defined by

$$P(X_r^n = y | X_{r-1}^n = x) = p_\alpha^n(x, y) = a_n(\alpha, x, y) / \mu_n(\alpha, x).$$

We write P_α^n for the law of this random walk.

3. It will sometimes be helpful to write (2.1) in matrix terms. Let A ($= A^n(\alpha)$) be the matrix defined by

$$A_{xy} = \delta_{xy}\mu_n(\alpha, x, y) - a_n(\alpha, x, y);$$

then

$$(2.2) \quad \mathcal{E}_\alpha^n(f, f) = f^T A f.$$

The fundamental property of finitely ramified fractals is that the operation of ‘decimation’ works in a particularly straightforward fashion. We now introduce this. Given \mathcal{E}_α^n , we define, for $g \in B(F_{n-1})$, the decimated Dirichlet form

$$(2.3) \quad \tilde{\mathcal{E}}_\alpha^n(g, g) = \inf\{\mathcal{E}_\alpha^n(f, f) : f|_{F_{n-1}} = g\}.$$

THEOREM 2.1. $\tilde{\mathcal{E}}_\alpha^n$ is a Dirichlet form and there exists $\tilde{\alpha}(\alpha) \in \Lambda$ such that

$$\tilde{\mathcal{E}}_\alpha^n = \mathcal{E}_{\tilde{\alpha}}^{n-1}.$$

It is straightforward to check this result. Note that the minimising function f in (2.3) satisfies

$$(2.4) \quad \Delta_n^\alpha f(x) = 0, \quad x \in F_n - F_{n-1},$$

and that f inside an $n - 1$ cell $\Psi_w(H_0)$ depends only on the values of g on $I_w(F_0)$. Thus the proof reduces immediately to the case $n = 1$.

We now comment on the various interpretations of this operation.

1. In terms of electrical networks we can view $(F_n, a_n(\alpha))$ as a ‘black box’, with only the nodes in F_{n-1} being accessible. Then the network $(F_n, a_n(\alpha))$ has the same response (in terms of energy dissipation, current flows etc.) to inputs in F_{n-1} as the network $(F_{n-1}, a_{n-1}(\tilde{\alpha}))$.

2. We may ‘decimate’ the random walk $(X_r^n, r \geq 0)$ by defining

$$\begin{aligned} T_0 &= \min\{r \geq 0 : X_r^n \in F_{n-1}\}, \\ T_m &= \min\{r \geq T_{m-1} : X_r^n \in F_{n-1} - \{X_{T_{m-1}}\}\} \\ \tilde{X}_r^n &= X_{T_r}^n, r \geq 0. \end{aligned}$$

Then the random walk $(\tilde{X}^n, P_\alpha^n)$ is equal in law to the random walk $(X^{n-1}, P_\alpha^{n-1})$

3. In matrix terms we write vectors $f \in \mathbb{R}^{F_n}$ with the coordinates in F_{n-1} first. So

$$A_n(\alpha) = \begin{pmatrix} B_n(\alpha) & C_n(\alpha) \\ C_n(\alpha)^T & D_n(\alpha) \end{pmatrix}$$

and the matrix corresponding to the Dirichlet form $\tilde{\mathcal{E}}_\alpha^n$ is given by

$$\tilde{A}_n(\alpha) = B_n(\alpha) - C_n(\alpha)D_n(\alpha)^{-1}C_n(\alpha)^T.$$

The decimation operation enables us to reduce many questions concerning the Dirichlet forms \mathcal{E}_α^n to the behaviour of iterates of the map $\tilde{\alpha}$. Of particular significance are (affine) fixed points of the function $\tilde{\alpha}$, that is $\beta \in \Lambda$ such that

$$(2.5) \quad \tilde{\alpha}(\beta) = \beta/\rho \quad \text{for some } \rho > 0.$$

These fixed points correspond to processes and operators on F with the correct scaling properties with respect to the maps ψ_i . As $\tilde{\alpha}$ is continuous, the fixed point theorem applied to the simplex $\Lambda' = \{\alpha \in \Lambda : \sum \alpha_i = 1\}$ shows that at least one fixed point exists. A reflection argument [L, Ch. V] shows that for nested fractals there is at least one fixed point that is non-degenerate in the sense that $\alpha_i > 0$ for all i , so that the electrical network $(F_1, a_1(a))$ is connected.

The general problem of the existence and uniqueness of (non-degenerate) fixed points seems to be a hard one. The natural approach is to

show that $\tilde{\alpha}$ is a contraction on Λ' in some suitable metric d , but finding such a metric does not appear to be easy. There are some partial results in [B1], which introduces an approach which may be effective for general nested fractals, but which will not work for general p.c.f. self-similar sets. It also seems likely that any non-degenerate fixed point β will also be stable; again this would be settled if a suitable contraction could be found.

Problem. Can the map $\tilde{\alpha}$ have more than one non-degenerate fixed point? Are non-degenerate fixed points always stable?

Remark. It is easy to see that $\tilde{\alpha}$ can have one or more degenerate fixed points – examples are given in [L] and [B1]. It is also possible that for a p.c.f self – similar set there is no non-degenerate fixed point – see the example in [HHW].

The scalar ρ in (2.5) will play a fundamental role in the analytic properties of the fractal F . In terms of the electrical network model, replacing the network $(F_0, a_0(\beta))$ by the network $(F_1, a_1(\beta))$ gives rise, to an observer only able to access the nodes in F_0 , to a network equivalent to $(F_0, a_0(\rho^{-1}\beta))$: since resistance = 1/ conductivity it is natural to call ρ the *resistance scale factor*.

The close connection between electric networks and symmetric random walks also enables us to give a probabilistic interpretation of ρ . Let G be a finite set, $a = (a_{xy}, x, y \in G)$ be a ‘conductivity’ matrix satisfying

$$a_{xx} = 0, \quad a_{xy} = a_{yx} \geq 0,$$

and let $Y_n, n \geq 0$ be the Markov chain with transition probabilities given by $p_{xy} = a_{xy}/\mu_x$, where $\mu_x = \sum_y a_{xy}$. Then writing $S_x = \min\{n \geq 0 : Y_n = x\}$ and $R(x, y)$ for the effective resistance of the network (G, a) between the points x and y , we have ([C])

$$(2.6) \quad E^x S_y + E^y S_x = R(x, y) \sum_x \mu_x.$$

Since the final term may be interpreted as the ‘mass’ of the network, we have the informal relation

Time = Resistance \times Mass.

With this motivation, it is easy to prove that, for the random walk X^n on $(F_n, a_n(\beta))$ we have

$$ET_m = M\rho.$$

We call M (the number of 1-cells in F_1) the *mass scaling factor* of the fractal F , $\tau = M\rho$ the *time scaling factor*, and λ (the contraction factor in the similitudes ψ_i) the *length scaling factor*. From these we define two new indices connected with F :

$$d_w(F) = \frac{\log \tau}{\log \lambda}$$

$$d_s(F) = -\frac{2 \log M}{\log \tau};$$

following the physics literature we call these the *walk* and *spectral* dimensions of F . (The reasons for this terminology will become apparent later).

Most of the analytic properties of F can be summarised in terms of these two ‘dimensions’, together with the Hausdorff dimension $d_f(F)$. Of course there are really only two independent quantities, since we have

$$d_w(F) = \frac{2d_f(F)}{d_s(F)}.$$

Remarks 1. For the fractals considered here, it is easy to show that $\rho \geq 1$, or equivalently that $\tau \geq M$, so that $d_s(F) \leq 2$. We also have that $d_s(F) \leq d_f(F)$, giving $d_w(F) \geq 2$.

2. Though it does not fall into the class of nested fractals considered here, it is possible to define these dimensions for the unit cube $C \subset \mathbb{R}^d$. We then have

$$d_f(C) = d, \quad d_w(C) = 2, \quad d_s(C) = d.$$

3. LIMITING PROCESSES AND REGULARITY.

We now fix a non-degenerate fixed point α , with resistance scaling factor ρ , and drop the α from expressions like Δ_α^n . We rescale α such that $\mu_n(\alpha, F_n) = M^n$.

We wish to obtain a limit of the (suitably rescaled) Δ^n . One method is probabilistic. If X^n is the random walk on F_n with generator Δ^n , then by decimation we obtain random walks on F_{n-1}, F_{n-2} , etc. with generators $\Delta^{n-1}, \Delta^{n-2}$, etc. Taking projective limits, we obtain a sequence $X^n, n \geq 0$, tied together by this decimation property. If we rescale time by considering the processes

$$Y_t^n = X_{[r^n t]}^n, \quad t \geq 0,$$

then each of the Y^n crosses F_0 in mean time 1, and it is not too hard to see that in fact $Y_t^n \rightarrow Y_t$ a.s., and that Y is a continuous F -valued process. Unfortunately, establishing the Markov property for Y is rather tiresome and technical. There are therefore some advantages in using the more analytic approach outlined in [F2].

For $f \in B(F_m)$, and $n \geq m$, let $L_{n,m}f \in B(F_n)$ be the unique function such that

$$(3.1) \quad \begin{aligned} L_{n,m}f(x) &= f(x), & x \in F_m \\ \Delta^n L_{n,m}f(x) &= 0 & x \in F_n - F_m. \end{aligned}$$

We call a function f such that $\Delta^n f(x) = 0$ for $x \in F_n - F_m$ m -harmonic. It is easily verified that $L_{m,m-1}f$ is the function g which attains the minimum in (2.3); therefore

$$(3.2) \quad \mathcal{E}^m(L_{m,m-1}f, L_{m,m-1}f) = \rho^{-1} \mathcal{E}^{m-1}(f, f).$$

Thus, since $\mathcal{E}^m(f, f - L_{m,m-1}(f|_{F_{m-1}})) = 0$, we have

$$(3.3) \quad \rho^m \mathcal{E}^m(f, f) \geq \rho^{m-1} \mathcal{E}^{m-1}(f|_{F_{m-1}}, f|_{F_{m-1}}).$$

From now on we will often avoid notation like $f|_{F_{m-1}}$ by extending \mathcal{E}^{m-1} etc. to functions on F . (3.3) gives an easy definition of the limiting Dirichlet form on F : we set, for $f \in C(F)$,

$$\mathcal{E}(f, f) = \lim_n \rho^n \mathcal{E}^n(f, f), \quad \mathcal{D} = \{f \in C(F) : \mathcal{E}(f, f) < \infty\}.$$

We now consider the regularity properties of $(\mathcal{E}, \mathcal{D})$.

If $f \in B(F_0)$, then, as $\alpha_i > 0$ for all i , we have

$$\sup_{x,y \in F_0} |f(x) - f(y)|^2 = \text{Osc}(f, F_0)^2 \leq c_0 \mathcal{E}^0(f, f).$$

As any pair of points in F_1 can be connected by a chain of at most M 1-cells, it follows that, for $f \in B(F_1)$,

$$(3.4) \quad \text{Osc}(f, F_1)^2 \leq c_0 M \mathcal{E}^1(f, f).$$

A standard argument, linking x, y by a suitable tower of r -cells, with $0 \leq r \leq n$, establishes the following Sobolev inequality:

$$(3.5) \quad \text{Osc}(f, F_n) \leq c_1 \rho^n \mathcal{E}^n(f, f), \quad f \in B(F_n).$$

If x, y belong to the same m -complex, then $|x - y| < c\lambda^{-m}$, and the chain need only use r -cells with $m \leq r \leq n$; it follows that

$$(3.6) \quad |f(x) - f(y)|^2 \leq c_2 |x - y|^{d_w - d_f} \rho^n \mathcal{E}^n(f, f).$$

Let $\mathcal{H}_{n,m} = \{f \in B(F_n) : \Delta^n f(x) = 0, x \in F_n - F_m\}$ be the set of m -harmonic functions on F_n . The decimation property implies that if $f \in \mathcal{H}_{n,m}$ then $f|_{F_{n-1}} \in \mathcal{H}_{n-1,m}$ if $n - 1 \geq m$. (Perhaps the quickest way to see this is in probabilistic terms - $f(X^n)$ is a martingale up to the first time X^n hits F_m , and therefore the process $f(\tilde{X}^n)$ is also). Thus if $n \geq j \geq m$ then $L_{n,j} L_{j,m} f = L_{n,m} f$. If $g \in B(F_m)$, then for $x \in F_\infty = \cup_{n=0}^\infty F_n$, define

$$L_m g(x) = L_{n,m} g(x), \text{ where } x \in F_n.$$

Since $\rho^n \mathcal{E}^n(L_{n,m} g, L_{n,m} g) = \rho^m \mathcal{E}^m(g, g)$ it follows using (3.6) that $L_m g$ is uniformly continuous on F_∞ , and so may be extended to a function (also denoted $L_m g$) in $C(F)$, satisfying $\mathcal{E}(L_m g, L_m g) = \rho^m \mathcal{E}^m(g, g)$.

This shows that

$$\bigcup_{m=0}^\infty \{L_m g, \quad g \in B(F_m)\} \subseteq \mathcal{D},$$

so that \mathcal{D} is dense in $C(F)$, and this, together with the fact that $\mathcal{D} \subset C(F)$ proves that the Dirichlet form $(\mathcal{E}, \mathcal{D})$ is regular.

Let μ be the weak limit of the measures $M^{-n}\mu_n$: then $\mu(F) = 1$, and μ assigns mass M^{-n} to any n -complex. (μ is a multiple of the Hausdorff x^{d_f} - measure on F , and is also the image measure under the mapping ϕ of the uniform product measure on the abstract sequence space $I^{\mathbb{N}}$). Then since

$$\rho^n \mathcal{E}^n(f, g) = - \sum_{x \in F_n} M^{-n} \mu_n(x) \tau^n \Delta^n f(x) g(x),$$

if we define Δ_F to be the self-adjoint operator on $L^2(F, \mu)$ associated with $(\mathcal{E}, \mathcal{D})$ by $\mathcal{E}(f, g) = -(\Delta_F f, g)$, then Δ_F is approximated by the sequence $\tau^n \Delta^n$.

Let us summarise the properties of \mathcal{E} and Δ_F and the associated diffusion process X in the following

THEOREM 3.1 (a) $(\mathcal{E}, \mathcal{D})$ is a regular local Dirichlet form on $L^2(F, \mu)$.

(b) For $f \in \mathcal{D}$, $x, y \in F$,

$$(3.7) \quad |f(x) - f(y)|^2 \leq c|x - y|^{d_w - d_f} \mathcal{E}(f, f),$$

(c) For $f \in \mathcal{D}$

$$\rho^n \mathcal{E}^n(f|_{F_n}, f|_{F_n}) \uparrow \mathcal{E}(f, f).$$

(d) Δ_F is a self-adjoint operator on $L^2(E, \mu)$ with $\mathcal{D}(\Delta_F) \subset \mathcal{D}$, satisfying

$$\begin{aligned} \mathcal{E}(f, g) &= -(\Delta_F f, g) \\ \Delta_F f(x) &= \lim_{n \rightarrow \infty} \tau^n \Delta^n f(x), \quad x \in F_\infty \end{aligned}$$

(e) If $(X_t, t \geq 0, P^x, x \in F)$ is the diffusion (continuous strong Markov process) with semigroup $P_t = e^{t\Delta_F}$, then X is the weak limit of the processes $Y_t^n = X_{[\tau^n t]}^n$. The process X is symmetric with respect to μ .

The standard Laplacian on \mathbb{R}^d can be characterised, up to a multiplicative constant, as the unique second order which is invariant with respect to the isometries of \mathbb{R}^d . It is natural to ask for a similar characterisation of Δ_F . In general the answer is not known – the problem is related to that of the uniqueness of fixed points mentioned earlier. However for the Sierpinski gasket (where there is evidently only one fixed point), it is proved in [BP] that, up to a deterministic time change, X is the unique process which is locally invariant with respect to the local isometries of F : a corresponding uniqueness for Δ_F follows.

4. ANALYTIC PROPERTIES OF Δ_F .

We begin by introducing the potential kernel densities $u_\alpha(x, y)$, $\alpha > 0$, $x, y \in F$. Formally these are the solutions to

$$(\alpha - \Delta_F)u_\alpha(\cdot, y) = \delta_y(\cdot).$$

They may be defined via the process X as the density with respect to μ of the α -resolvent

$$(4.1) \quad U_\alpha f(x) = E^x \int_0^\infty e^{-\alpha t} f(X_t) dt;$$

of course it requires some work to prove that $U_\alpha(\cdot, x) \ll \mu$. A quicker approach is via the Dirichlet form $\mathcal{E}_\alpha(f, g) = \alpha \cdot (f, g) + \mathcal{E}(f, g)$; from (3.7) it follows that

$$\|f\|_\infty^2 \leq c_\alpha \mathcal{E}_\alpha(f, f), \quad f \in \mathcal{D},$$

so that the map $T_y : \mathcal{D} \rightarrow \mathbb{R}$ defined by $T_y(g) = g(y)$ is bounded. It follows, as in [F1, 3.3.3] that \mathcal{E}_α admits a reproducing kernel $u_\alpha(x, y)$, so that

$$(4.2) \quad \mathcal{E}_\alpha(u_\alpha(\cdot, y), g) = g(y), \quad g \in \mathcal{D}.$$

THEOREM 4.1 (a) $u_\alpha(x, y) = u_\alpha(y, x)$, $x, y \in F$.

(b) u_α is Hölder continuous of order $d_w - d_f$ on $F \times F$.

(c) $U_\alpha f(x) = \int_F u_\alpha(x, y) f(y) \mu(dy)$, $f \in C(F)$.

Various probabilistic properties of X follow from this result. The boundedness of the $u_\alpha(\cdot, y)$ implies that, writing $T_y = \inf\{t > 0 : X_t = y\}$ we have

$$u_\alpha(x, y) = E^x e^{-\alpha T_y} u_\alpha(y, y),$$

so that $P^y(T_y = 0) = 1$, and y is regular for $\{y\}$. The Hölder continuity of u_α implies that X has a jointly continuous local time (occupation density) $(L_t^x, x \in F, t \geq 0)$ and from this it follows that X is 'space-filling':

$$P^x(\{X_t, 0 \leq t \leq N\} = F \text{ for some } N < \infty) = 1.$$

Some properties of X can be expressed more simply if we consider instead the process on an unbounded version of F . Let us take $a_1 = 0$, so that the similitude ψ_1 is given by $\psi_1(x) = \lambda^{-1}x$. Then set

$$\tilde{F} = \bigcup_{n=0}^{\infty} \lambda^n F.$$

The construction of the Dirichlet form \mathcal{E} , the Laplacian Δ and the diffusion X on \tilde{F} can be done by a straightforward patching argument. Then \tilde{X} satisfies the scaling relation

$$(4.3) \quad (X_t, t \geq 0, P^0) \stackrel{(d)}{=} (\lambda^{-1} X_{\tau t}, t \geq 0, P^0),$$

which may be compared with the relation $\theta^{-1} B_{\theta^2 t} \stackrel{(d)}{=} B_t$ for standard Brownian motion on \mathbb{R}^d . Since $|X_1 - X_0| = 0(1)$ it follows from (4.3) that $|X_t - X_0| = 0(t^{1/d_w})$ for $t > 0$. (Recall $1/d_w = \log \lambda / \log \tau$). More precisely we have

PROPOSITION 4.2 (a) X is Hölder continuous of order $d_w^{-1} - \varepsilon$ for any $\varepsilon > 0$, and is not Hölder continuous of order d_w^{-1} .

(b) There exist constants c_1, c_2 such that for $t \geq 0$

$$c_1 t^{2/d_w} \leq E^x |X_t - x|^2 \leq c_2 t^{2/d_w}.$$

More precise results, giving the exact modulus of continuity of the paths of X can be found in [BP, BB3, Kum]. This result gives the most intuitive explanation of the meaning of the ‘dimension’ d_w : it governs the space–time scaling of the diffusion X in the same way that the number 2 governs the space–time scaling of ordinary Brownian motion.

COROLLARY 4.3 (a) X is not a semimartingale.

(b) If $f \in C^1(\mathbb{R}^d)$ and $f|_F \in \mathcal{D}(\Delta_F)$ then $f|_F$ is constant.

(a) follows from the Proposition and the fact that the paths of any semimartingale have finite quadratic variation. For (b), if $f \in \mathcal{D}(\Delta)$ then

$$M_t^f = f(X_t) - f(X_0) - \int_0^t \Delta f(X_s) ds$$

is a martingale, so $f(X_t)$ is a semimartingale, which is impossible unless $f|_F$ is constant.

We thus see that the ‘smooth functions’ in $\mathcal{D}(\Delta_F)$ have nothing to do with the ordinary smooth functions on \mathbb{R}^d . Not a lot is known at present about the structure and form of functions in $\mathcal{D}(\Delta_F)$ or \mathcal{D} . From Theorem 3.1 we have that they are Hölder continuous of order $d_w - d_f$ everywhere; they are also Hölder continuous of order $\frac{1}{2}d_w$ μ -a.e. on F . (Note that as $\frac{1}{2}d_w > 1$ this last result cannot hold everywhere). Some more precise results are obtained in [K2]: if ν_f is the local measure associated with $\mathcal{E}(f, f)$, so that

$$\mathcal{E}(f, f) = \int_F \nu_f(dx), \quad f \in \mathcal{D},$$

then there exists a measure ν on F such that $\mu \perp \nu$, and $\nu_f \ll \nu$ for all $f \in \mathcal{D}$. Informally, we may say that for $f \in \mathcal{D}$, $|\nabla f| = 0$ μ -a.e., while $\|\nabla f\|_\infty = \infty$.

As the $U_\alpha(x, y)$ are continuous U_α is compact, and we can write

$$u_\alpha(x, y) = \sum_{i=1}^{\infty} (\alpha + \lambda_i)^{-1} \varphi_i(x) \varphi_i(y),$$

where $0 \leq \lambda_1 \leq \lambda_2 \leq \dots$ are the eigenvalues of $-\Delta_F$, and the φ_i are the normalised eigenfunctions. Note that the λ_i and φ_i satisfy

$$\mathcal{E}(\varphi_i, f) = \lambda_i(f, g) \text{ for all } g \in \mathcal{D}.$$

We now examine the asymptotic behaviour of the eigenvalues. Define

$$\mathcal{D}_i = \{f \in \mathcal{D} : f = 0 \text{ on } F_i\}, \quad i = 0, 1,$$

and write Δ_i for the associated Laplacian (which corresponds to Dirichlet boundary conditions on F_i). Note that if $(\mathcal{E}_i, \mathcal{D}_i)$, $i = 1, 2$ are Dirichlet forms with $\mathcal{D}_1 \subseteq \mathcal{D}_2$ and $\mathcal{E}_2 = \mathcal{E}_1$ on \mathcal{D}_1 , then the max-min principle (see [Ch]) shows that, writing $\lambda_n^i, n \geq 1$ for the eigenvalues of the operators $-\Delta_i$, we have $\lambda_n^1 \geq \lambda_n^2$. Hence if $N_i(x) = \#\{\lambda_n^i : \lambda_n^i \leq x\}$ we have $N_1(x) \leq N_2(x)$. We call the functions N_i the *integrated density of states* (for $(\mathcal{E}_i, \mathcal{D}_i)$).

Write N, N_i for the integrated density of states for $(\mathcal{E}, \mathcal{D}), (\mathcal{E}, \mathcal{D}_i)$ respectively. If $f, g \in B(F_n)$ then

$$\mathcal{E}^n(f, g) = \sum_{i=1}^M \mathcal{E}^{n-1}(f \circ \psi_i, g \circ \psi_i),$$

so for $f, g \in \mathcal{D}$ it follows that

$$(4.4) \quad \mathcal{E}(f, g) = \sum_{i=1}^M \rho \mathcal{E}(f \circ \psi_i, g \circ \psi_i).$$

As $\mathcal{D}_1 \subseteq \mathcal{D}_0$, $N_1(x) \leq N_0(x)$. However, if g is an eigenfunction of $-\Delta_0$, with $\Delta_0 g + \lambda g = 0$, then for $1 \leq i \leq M$ define $h_i \in C(F)$ by

$$h_i(x) = \begin{cases} g \circ \psi_i^{-1}(x) & \text{for } x \in \psi_i(F). \\ 0 & \text{for } x \notin \psi_i(F). \end{cases}$$

Then for $v \in \mathcal{D}_1$, by (4.4), and as $h_i \circ \psi_i = g$,

$$\mathcal{E}(h_i, v) = \rho \mathcal{E}(g, v \circ \psi_i) = \rho \lambda(g, v \circ \psi_i) = M \rho \lambda(h_i, v).$$

So $\tau\lambda$ is an eigenvalue of Δ_1 with multiplicity M and it follows that $N_1(x) = MN_0(x/\tau)$.

We perform a similar kind of surgery for the Neumann eigenvalues on F ; this time it involves cuts at the points in F_1 , so replacing F by a disjoint union of the spaces $\psi_i(F), 1 \leq i \leq M$. Writing \mathcal{D}' for the associated Dirichlet space, and N' for the integrated density of states, we have (taking a suitable embedding) $\mathcal{D} \subseteq \mathcal{D}'$. The same argument as for the Dirichlet case gives (writing $y = x/\tau$)

$$\begin{aligned} N_0(y) &= M^{-1}N_1(\tau y) \leq M^{-1}N_0(\tau y) \leq M^{-1}N(\tau y) \\ &\leq M^{-1}N'(Ty) = N(y). \end{aligned}$$

Thus there exist c_1, c_2 such that

$$(4.5) \quad c_1 x^{\frac{1}{2}d_s} \leq N_0(x) \leq N(x) \leq c_2 x^{\frac{1}{2}d_s} \quad \text{for } x \geq x_0.$$

It is now natural to ask about finer details of the asymptotics of $N(\cdot)$. However, Fukushima and Shima [FS] have proved that, for the Sierpinski gasket,

$$\liminf_{x \rightarrow \infty} x^{-\frac{1}{2}d_s} N(x) < \limsup_{x \rightarrow \infty} x^{-\frac{1}{2}d_s} N(x);$$

thus one cannot hope for an asymptotic expansion of N of the kind that occurs for domains in \mathbb{R}^d .

Remarks 1. [S] and [FS] give a very detailed description of the spectrum of Δ for the Sierpinski gasket.

2. Equation (4.5) explains the term *spectral dimension* for the number d_s .

5. HEAT KERNEL BOUNDS

So far we have had to pay little attention to the detailed geometry of F . However, when we consider the form of the heat kernel the metric $|x - y|$

is inadequate: it is better to consider an intrinsic metric $d(x, y)$, related to the shortest path between x and y in F . Let

$$H_n = \Psi^n(H_0)$$

be the closed connected set obtained after n iterations of the set map Ψ defined by (1.2), and $d_n(x, y)$ be the length of the shortest path in H_n connecting x and y .

For fractals such as the Sierpinski gasket and carpet one has $|x - y| \leq d_n(x, y) \leq c|x - y|$ for all n , but for more general fractals, where the limiting set does not contain straight line segments, this fails. In general there exists $b \geq \lambda$ such that points in F_0 may be connected by a chain of $O(b^n)$ n -cells, so that, writing

$$d_c = \log b / \log \lambda ,$$

one has

$$d_n(x, y) \approx |x - y|^{d_c} (b/\lambda)^n .$$

Taking limits along a subsequence one obtains a metric d on F satisfying

$$d(x, y) \approx |x - y|^{d_c} .$$

This metric is called the *chemical distance* in the physics literature, and d_c is known as the *chemical exponent*. (Contrary to what a mathematician might expect, physicists do not usually measure the length of a polymer with this metric).

It is helpful to redefine the fractal and walk dimensions in terms of this new metric, setting

$$d_f^l = \log M / \log b = d_f / d_c ,$$

$$d_w^l = \log \tau / \log b = d_w / d_c .$$

Note that the spectral dimension satisfies $\frac{1}{2}d_s = d_f / d_w = d_f^l / d_w^l$ and is unaffected by the change of metric.

Now consider X on the unbounded fractal \tilde{F} , and let $p(t, x, y)$ be the associated transition density. (The existence of such a function follows from the μ -symmetry of X and the existence of a resolvent density U_α). p is the solution to the heat equation on \tilde{F} :

$$\begin{aligned} \frac{\partial}{\partial t} p(t, x, y) &= \Delta_x p(t, x, y) \\ p(0, x, y) &= \delta_y(x) \end{aligned}$$

Upper bounds on $p(t, x, x)$ can be obtained from the scaling of $\mathcal{E}(\cdot, \cdot)$, and the general theory of [CKS]. The argument here is due to Fitzsimmons and Hambly. Since (see [F, p21])

$$\mathcal{E}(f, f) = \sup_\alpha (\alpha f, f - \alpha U_\alpha f) \geq (f, f - U_1 f),$$

we have

$$\|f\|_2^2 \leq \|u_1\|_\infty \|f\|_1^2 + \mathcal{E}(f, f).$$

Replacing f by $\tilde{f}(x) = f(\lambda^n x)$, using scaling, and optimising over n , one obtains the *Nash inequality*

$$(5.1) \quad \|f\|_2^{2+4/d_s} \leq c \mathcal{E}(f, f) \cdot \|f\|_1^{4/d_s}.$$

This implies ([CKS]) that

$$(5.2) \quad p(t, x, x) \leq c_1 t^{\frac{1}{2}d_s}, \quad 0 < t < \infty.$$

On-diagonal lower bounds are also quite easy, and follow from the estimate

$$\limsup_{\theta \uparrow \infty} \sup_{x, t} P^x(|X_t - x| > \theta t^{1/d_w}) = 0.$$

This implies that, for a suitable $r > 0$, writing $z = r t^{1/d_w}$,

$$\frac{1}{2} < \int_{B(x, z)} p(t, x, y) \mu(dy) \leq \mu(B(x, z)) p(t, x, x)^{\frac{1}{2}} \sup_y p(t, y, y)^{\frac{1}{2}}.$$

Using $\mu(B(x, z)) \geq cz^{d_f}$ gives the lower bound

$$(5.3) \quad p(t, x, x) \geq ct^{\frac{1}{2}d_s}, \quad 0 < t < \infty.$$

Remarks. 1. The link between the asymptotics of the heat kernel and the number of eigenvalues means that one would expect bounds of this form for small t from (4.5). The scaling relation (4.2) then implies that these bounds hold for all t .

2. Note that the exponent in (5.2) and (5.3) is not related to the isoperimetric dimension, which is 0 for finitely ramified fractals. The situation here is therefore rather different from the groups studied in [V], where an isoperimetric inequality yields a Nash inequality with the correct exponent. Compare also the bounds on the heat kernel given in [O] and [BB3].

The off-diagonal bounds are more tricky. There is a general ‘machine’ for obtaining off-diagonal upper bounds – see the account of ‘Davies’ method’ in [CKS]. This essentially involves looking at the semigroup $e^{-\psi} P_t e^{\psi}$, using the global upper bound given by the on-diagonal upper bound, and varying ψ suitably. Unfortunately this does not seem to work in the fractal case. The essential reason is that, on one hand, one requires $\psi \in \mathcal{D}$, while the upper bound obtained involves, in a more or less explicit form, expressions involving $\|\nabla\psi\|_{\infty}$, which as is proved in [Kus2], is $+\infty$ for nested fractals.

The technique one adopts, for both upper and lower bounds, is that of *chaining*. For the lower bound this is classical. Suppose we have that

$$(5.4) \quad p(t, x, y) \geq ct^{-d_s/2} \text{ for } t > 0, \quad |x - y| < c_2 t^{1/d_w}.$$

Writing $D = |x - y|$, we have that x and y may be connected by a chain of $n = c(D/\varepsilon)^{d_c}$ balls of Euclidean radius ε ; denote these $B_i = B(x_i, \varepsilon)$,

$1 \leq i \leq n$, where $x_0 = x, x_n = y$. Choose ε so that $\varepsilon = c_2(t/n)^{1/d_w}$; then

$$\begin{aligned}
 p(t, x, y) &\geq \int_{B_1} \dots \int_{B_{n-1}} p(t/n, x_0, x_1) \dots p(t/n, x_{n-1}, y) \mu(dx_1) \dots \mu(dx_{n-1}) \\
 &\geq \prod_1^{n-1} \mu(B_i) c_1^n (t/n)^{-nd_s/2} \\
 &\geq (c\varepsilon^{d_f})^{n-1} c_1^n (t/n)^{-d_s/2} \cdot (\varepsilon/c_2)^{-d_f(n-1)} \\
 &\geq t^{-d_s/2} c_3^n \\
 &= t^{-d_s/2} \exp(-c_4 n).
 \end{aligned}$$

Substituting for n , one obtains

$$(5.5) \quad p(t, x, y) \geq t^{-d_s/2} \exp(-c(|x - y|^{d_w}/t)^{d_c/(d_w-d_c)});$$

the exponent in (5.5) may be rewritten as

$$c(d(x, y)^{d_w^l}/t)^{1/(d_w^l-1)}.$$

Thus the chaining argument transforms the crude lower bound of (5.4) into a global lower bound which is, up to constants, of the best possible form.

It remains to prove (5.4). However, it holds for $y = x$, and the Hölder continuity of functions in \mathcal{D} allows this to be extended to a ball.

A similar technique works for the upper bound, but the argument here is more probabilistic. For $x \in \tilde{F}$, $\varepsilon > 0$ write

$$S(x, \varepsilon) = \inf\{t \geq 0 : |X_t - x| < \varepsilon\}.$$

The space-time scaling of X suggests that $\varepsilon^{-1/d_w} S(x, \varepsilon)$ should be $O(1)$, and using this and the elementary inequality

$$P(\xi \leq t) \leq (\text{Var}(\xi) + 2tE\xi)/E\xi^2$$

which holds for any non-negative random variable ξ , one obtains

$$(5.6) \quad P(S(x, \varepsilon) < t) \leq p + a\varepsilon^{-1/d_w} t, \quad t \geq 0.$$

The following lemma ([BB1, Lemma 1.1]) is the key to the chaining argument.

LEMMA 5.1 *Let T, S_1, \dots, S_n be non-negative random variables satisfying, for some $p < 1, a > 0$,*

- (a) $T \geq \sum_1^n S_i$,
- (b) $P(S_i < t \mid \sigma(S_1, \dots, S_{i-1})) \leq p + at$.

Then $P(T \leq t) \leq \exp(2(ant/p)^{\frac{1}{2}} - n \log p^{-1})$.

Now fix x, y, t , and set $A = \{Z \in F : |z - y| < |z - x|\}$. If we set $S_0 = 0, S_{i+1} = S(X_{S_i}, \varepsilon)$, then since the shortest path from x to A in F crosses $n = c(|x - y|/\varepsilon)^{d_c}$ balls of radius ε , we have, writing $T_A = \inf\{t \geq 0 : X_t \in A\}$,

$$\begin{aligned} \log P^x(T_A \leq t) &\leq (2a\varepsilon^{-1/d_w} nt/p)^{\frac{1}{2}} - n \log p^{-1} \\ &\leq c(|x - y|^{d_w}/t)^{d_c/(d_w - d_c)}. \end{aligned}$$

Combining this with the global upper bound (5.1), one obtains:

THEOREM 5.2 *Let $\phi(z, t) = (z^{d_w}/t)^{1/(d_w - 1)}$. The heat kernel density $p(t, x, y)$ on \tilde{F} satisfies, for $0 < t < \infty, x, y \in \tilde{F}$,*

$$c_1 t^{-d_s/2} \exp(-c_2 \phi(d(x, y), t)) \leq p(t, x, y) \leq c_3 t^{-d_s/2} \exp(-c_4 \phi(d(x, y), t)).$$

Remark. Bounds of this type are quite unfamiliar in classical situations, but they do nevertheless arise in setups less unfamiliar than the fractals treated here. For example, let \tilde{G} be the unbounded two-dimensional Sierpinski gasket (for which $d_c = 1, d_w = \log 5/\log 2$ and $d_f = \log 3/\log 2$). Embed \tilde{G} in \mathbb{R}^3 , set

$$M' = \bigcup_{x \in \tilde{G}} B(x, 1),$$

and let $M = \partial M'$. Then M is a manifold with a large scale structure which mimics that of the Sierpinski gasket. If $p(t, x, y)$ is the heat kernel on M , then for small t , and $|x - y| = 0(1)$ its behaviour is similar to that of the heat kernel on \mathbb{R}^2 . However one expects the estimates of Theorem 5.2 to hold in the region $1 \leq t \leq |x - y|$. (For $|x - y| > t$ one is back in the classical large deviation situation).

REFERENCES

- [B1] M.T. BARLOW: Random walks, electrical resistance and nested fractals. To appear in *Asymptotic Problems in Probability Theory*, ed. K.D. Elworthy, N. Ikeda, Pitman.
- [BB1] M. T. BARLOW and R. F. BASS: The construction of Brownian motion on the Sierpinski carpet. *Ann. Inst. H. Poincaré* **25** (1989) 225–257.
- [BB2] M. T. BARLOW and R. F. BASS: On the resistance of the Sierpinski carpet. *Proc. R. Soc. London A.* **431** (1990) 345–360.
- [BB3] M.T. BARLOW and R.F. BASS: Transition densities for Brownian motion on the Sierpinski carpet. *Probab. Th. Rel. Fields* **91** (1992) 307–330.
- [BP] M. T. BARLOW and E. A. PERKINS: Brownian motion on the Sierpinski gasket. *Probab. Th. Rel. Fields* **79** (1988) 543–623.
- [CKS] E.A. CARLEN, S. KUSUOKA and D.W. STROOCK: Upper bounds for symmetric Markov transition functions. *Ann. Inst. H. Poincaré Sup no. 2* (1987) 245–287.
- [C] A.K. CHANDRA, P. RAGHAVEN, W.L. RUZZO, R. SMOLENSKY, P. TIWARI: The electrical resistance of a graph captures its commute and cover times. Proceedings of the 21st ACM Symposium on theory of computing, 1989.
- [Ch] I. CHAVEL: *Eigenvalues in Riemannian Geometry*. Academic Press, 1984.
- [F1] M. FUKUSHIMA: *Dirichlet forms and Markov processes*, North Holland 1980.

- [F2] M. FUKUSHIMA: Dirichlet forms, diffusion processes, and spectral dimensions for nested fractals. To appear in "Ideas and methods in stochastic analysis, stochastics and applications", ed. S Albeverio et. al., Cambridge Univ. Press., Cambridge.
- [FS] M. FUKUSHIMA and T. SHIMA: On a spectral analysis for the Sierpinski gasket. To appear *J. of Potential Analysis*.
- [G] S. GOLDSTEIN: Random walks and diffusion on fractals. In: Kesten, H. (ed.) *Percolation theory and ergodic theory of infinite particle systems* (IMA Math. Appl., vol.8.) Springer, New York, 1987, pp.121–129.
- [HHW] K. HATTORI, T. HATTORI and H. WATANABE: Gaussian field theories and the spectral dimensions. *Prog. Th. Phys. Supp. No. 92* (1987) 108-143.
- [Kig1] J. KIGAMI: A harmonic calculus on the Sierpinski space. *Japan J. Appl. Math.* **6** (1989) 259–290.
- [Kig2] J. KIGAMI: A harmonic calculus for p.c.f. self-similar sets. To appear *Trans. A.M.S.*
- [Kum] T. KUMAGAI: Estimates of the transition densities for Brownian motion on nested fractals. Preprint 1991.
- [K1] S. KUSUOKA: A diffusion process on a fractal. In: Ito, K., N. Ikeda, N. (ed.) *Symposium on Probabilistic Methods in Mathematical Physics*, Taniguchi, Katata. Academic Press, Amsterdam, 1987, pp.251–274
- [K2] S. KUSUOKA: Dirichlet forms on fractals and products of random matrices. *Publ. RIMS Kyoto Univ.*, **25** (1989) 659-680.
- [KZ] S. KUSUOKA and X.Y.ZHOU: Dirichlet form on fractals: Poincaré constant and resistance. *Probab. Th. Rel. Fields* **93**, (1992) 169-186.
- [L] T. LINDSTRØM: Brownian motion on nested fractals. *Mem. A.M.S.* **420**, 1990.
- [O] H. OSADA: Isoperimetric dimension and estimates of heat kernels of pre-Sierpinski carpets. *Probab. Th. Rel. Fields* **86** (1990) 469–490.

- [RT] R. RAMMAL and G. TOULOUSE: Random walks on fractal structures and percolation clusters, *J. Physique Lettres* **44** (1983) L13–L22.
- [S] T. SHIMA: On eigenvalue problems for the random walk on the Sierpinski pre-gaskets. *Japan J. Appl. Ind. Math.*, **8** (1991) 127-142.
- [V] N. Th. VAROPOULOS: Isoperimetric inequalities and Markov chains. *J. Funct. Anal.* **63** (1985) 215-239.

M.T. BARLOW
Department of Mathematics
University of British Columbia
Vancouver
British Columbia
Canada V6T 1Z2

Astérisque

HENRI CARAYOL

Variétés de Drinfeld compactes

Astérisque, tome 206 (1992), Séminaire Bourbaki,
exp. n° 756, p. 369-409

<http://www.numdam.org/item?id=SB_1991-1992__34__369_0>

© Société mathématique de France, 1992, tous droits réservés.

L'accès aux archives de la collection « Astérisque » (<http://smf4.emath.fr/Publications/Asterisque/>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

VARIÉTÉS DE DRINFELD COMPACTES,
d'après Laumon, Rapoport et Stuhler
par Henri Carayol

0. INTRODUCTION

0.1. Considérons une forme modulaire parabolique $f = \sum a_n q^n$ normalisée ($a_1 = 1$) de poids 2 pour le groupe $\Gamma_0(N)$, dont on suppose qu'elle est vecteur propre des opérateurs de Hecke. La théorie désormais classique d'Eichler-Shimura, (généralisée par Deligne aux formes de poids > 2) associe alors à f des $\overline{\mathbb{Q}}_\ell$ -représentations σ_ℓ de dimension 2 du groupe de Galois $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$; la représentation σ_ℓ est caractérisée par le fait d'être non ramifiée en les $p \nmid N\ell$ et de vérifier les relations :

$$\text{tr } \sigma_\ell(\text{Frob}_p) = a_p, \quad \det \sigma_\ell(\text{Frob}_p) = p,$$

où Frob_p désigne un élément de Frobenius en p (on a implicitement choisi un plongement dans $\overline{\mathbb{Q}}_\ell$ du corps engendré par les a_p). On sait que les σ_ℓ s'obtiennent en décomposant la cohomologie ℓ -adique des *courbes modulaires* sous l'action des opérateurs de Hecke.

Une théorie plus générale, qui a connu ces dernières années des développements profonds (cf. [Ko]), consiste à faire de même pour la cohomologie des *variétés de Shimura* : on veut associer des représentations galoisiennes ℓ -adiques (en fait des motifs) à certaines représentations automorphes de

groupes réductifs sur des corps de nombres. Ce programme, qui s'insère dans la "philosophie de Langlands" (voir [La], [Cl]), n'a pour l'instant abouti complètement que pour certains types de groupes, pour l'essentiel dans des cas où les variétés de Shimura associées sont *compactes*.

0.2. Vers 1973, Drinfeld a commencé à développer une théorie semblable pour les *corps de fonctions* d'une variable sur un corps fini : c'est-à-dire des corps $F = \mathbf{F}_q(X)$ de fonctions rationnelles de courbes algébriques X/\mathbf{F}_q . L'analogie entre ces corps de fonctions et les corps de nombres est bien connue des arithméticiens, qui ont baptisé *corps globaux* l'un et l'autre type de corps. La "philosophie" de Langlands, de même que la théorie du corps de classes qu'elle généralise, doit s'appliquer à tous les corps globaux.

0.3. La première notion inventée par Drinfeld est celle de "*module elliptique de rang d* ". Cette notion dépend du choix d'une place, notée ∞ , de F , et fait intervenir le sous-anneau A de F constitué des fonctions rationnelles qui n'admettent pas de pôle en dehors de ∞ . Les objets considérés vérifient des propriétés formellement analogues à celles des courbes elliptiques. Munis de *structures de niveau J* convenables (pour J un idéal de A), ils sont classifiés par un *espace de modules M_J^d* , affine et lisse de dimension relative $(d - 1)$ sur $A[J^{-1}]$. En particulier, pour $d = 2$, on obtient des courbes sur F qui constituent l'analogie des courbes modulaires sur \mathbf{Q} . Considérant la cohomologie de ces courbes (convenablement compactifiées), Drinfeld a su développer une théorie parallèle à celle d'Eichler-Shimura : le résultat est qu'on peut associer à toute *représentation automorphe parabolique* Π du groupe adélique $\mathrm{GL}_2(\mathbf{A}_F)$, telle que la composante Π_∞ soit une représentation spéciale, des *représentations ℓ -adiques* de dimension 2 du groupe $\mathrm{Gal}(\overline{F}/F)$, où \overline{F} désigne une clôture séparable de F (ces représentations sont caractérisées en termes de Π par des relations analogues à celles qui déterminent les σ_ℓ ci-dessus). Dans un travail ultérieur [Dr 2], Drinfeld a construit

un système de revêtements étales des M_J^d , au moyen desquels (pour $d = 2$) il a pu associer des représentations galoisiennes aux représentations automorphes Π telles que Π_∞ soit cuspidale.

0.4. La définition d'un module elliptique est, ainsi qu'on le verra au cours de cet exposé, une définition algébrique élémentaire. Vers 1976, Drinfeld a introduit d'autres objets de nature plus géométrique, nommés "*faisceaux elliptiques*", qui consistent en certains diagrammes de fibrés vectoriels sur la courbe X . Il a prouvé que cette notion était équivalente à celle de module elliptique, et donc que les variétés de modules obtenues sont les mêmes. L'année suivante, c'est ce concept de faisceau elliptique qu'il généralisera pour obtenir des objets appelés "*Shtukas*" ou F -faisceaux, pour lesquels la place ∞ ne joue plus de rôle distinct. Utilisant ces derniers (en rang 2) et leurs variétés de modules, il parviendra à attacher une représentation galoisienne ℓ -adique à toute représentation automorphe parabolique de $\mathrm{GL}_2(\mathbf{A}_F)$.

0.5. Pour $d \geq 3$, les variétés de modules (ou faisceaux) elliptiques de rang d sont de dimension relative $(d - 1)$, et jouent encore pour le groupe $\mathrm{GL}_{d,F}$ le rôle de variétés de Shimura (lesquelles n'existent d'ailleurs pas pour GL_d ($d \geq 3$) sur un corps de nombres). L'étude de la cohomologie de ces variétés devrait permettre d'associer des représentations galoisiennes ℓ -adiques de dimension d aux représentations automorphes paraboliques Π de $\mathrm{GL}_d(\mathbf{A}_F)$ telles que Π_∞ soit spéciale. Mais cette étude est rendue extrêmement ardue par le fait que les variétés considérées ne sont pas propres, ce qui pose des problèmes difficiles tant du côté géométrique (questions de compactification, formule des traces de Lefschetz) que du côté automorphe (la formule des traces d'Arthur-Selberg n'a pas encore été établie pour les corps de fonctions). Signalons toutefois que Laumon [Lau 2] a su, en admettant certaines conjectures, calculer le module virtuel $\sum (-1)^i H_c^i(M_J^d \otimes_F \bar{F}, \bar{\mathbb{Q}}_\ell)$ où agissent $\mathrm{Gal}(\bar{F}/F)$ et les opérateurs de Hecke (les conjectures admises sont celle de Deligne, donnant une version simple

de la formule de Lefschetz pour une correspondance tordue par une grande puissance de Frobenius, et la formule des traces d'Arthur-Selberg). Pour les Shtukas, ces problèmes liés à la non-compacité sont encore pires, et même déjà sérieux — bien que résolus — pour $d = 2$.

0.6. Afin de contourner ces difficultés, Deligne avait, dès la parution du premier article de Drinfeld, posé la question de savoir définir et étudier des analogues propres des variétés modulaires de Drinfeld : ces nouvelles variétés devaient être associées, non plus au groupe $GL_{d,F}$, mais au groupe multiplicatif G d'un *corps gauche* D de centre F et de dimension d^2 sur F , qui se déploie au-dessus de la place ∞ . C'est cette question qui vient d'être résolue dans l'article [L-R-S] et qui fait l'objet du présent exposé.

Un point essentiel était de trouver une bonne définition de l'analogue d'un module elliptique : il se trouve que c'est en partant de la notion de faisceau elliptique, et non pas de celle plus élémentaire de module elliptique, qu'on aboutit à la définition la plus viable : celle d'un " *\mathcal{D} -faisceau elliptique*". Cependant, ces objets peuvent aussi — si la base est le spectre d'un corps parfait — être interprétés comme certains "fibrés vectoriels sur la droite projective non commutative", un concept qui généralise plus directement celui de module elliptique, dans l'esprit du travail d'Anderson sur les " *t -motifs*" [An].

0.7. Laumon, Rapoport et Stuhler montrent que les \mathcal{D} -faisceaux elliptiques, munis de structures de niveau I convenables (où I est un sous-schéma fermé fini de $X - \{\infty\}$) admettent des schémas de modules $\mathcal{E}ll_{X,\mathcal{D},I}$, propres et lisses de dimension relative $(d-1)$ sur $X - I - \mathcal{R} - \{\infty\}$ (où \mathcal{R} désigne le lieu de ramification de D). Ils décomposent ensuite la cohomologie $H^{d-1}(\mathcal{E}ll_{X,\mathcal{D},I} \times_X \text{Spec } \overline{F}, \overline{\mathbb{Q}}_\ell)$ sous l'action des opérateurs de Hecke, et prouvent que cette construction permet d'associer des représentations galoisiennes aux représentations automorphes Π de $G(\mathbf{A}_F)$ dont la composante Π_∞ est spéciale. Les représentations de $\text{Gal}(\overline{F}/F)$ ainsi obtenues sont celles que l'on espère (en particulier de dimension d) lors-

que la représentation Π intervient avec multiplicité 1 dans l'espace des formes automorphes, ce que l'on conjecture être toujours vrai : cela résulterait de l'existence d'une correspondance à la Jacquet-Langlands entre représentations automorphes de $G(\mathbf{A}_F)$ et de $GL_d(\mathbf{A}_F)$; malheureusement, aucune preuve complète de ce fait n'a été écrite pour les corps de fonctions (voir [De-Ka-Vi] et [Ro]) : il en résulte, outre la frustration que la multiplicité $m(\Pi)$ apparaisse dans l'énoncé du théorème principal de [L-R-S], de nombreuses complications techniques tout au long de l'article.

0.8. L'étude de la cohomologie $H^*(\mathcal{E}ll_{X,\mathcal{D},I} \times_X \text{Spec } \overline{F}, \overline{\mathbb{Q}}_\ell)$ et de sa décomposition se fait en utilisant l'arsenal familier aux spécialistes des variétés de Shimura : description de l'ensemble des points de la variété sur un corps fini, calcul d'intégrales orbitales, comparaison entre les formules des traces de Lefschetz et de Selberg... On utilise de plus, pour montrer que l'essentiel de la cohomologie se situe en degré $(d - 1)$, des arguments fondés sur le théorème de Lefschetz difficile et sur le théorème de Deligne de pureté de la filtration de monodromie. Cette partie de la preuve se simplifierait notablement si l'on disposait de la correspondance de Jacquet-Langlands généralisée.

0.9. Finalement, on prouve dans [L-R-S] la *conjecture de Langlands locale pour GL_d sur un corps local d'égale caractéristique* : cela utilise la construction globale de représentations galoisiennes (appliquée à des Π de multiplicité 1), ainsi que plusieurs résultats dus à Henniart. Deligne, dans les années 70, avait obtenu ce résultat pour $d = 2$, en s'appuyant sur la théorie de Drinfeld.

0.10. Nous nous bornerons dans cet exposé à décrire la théorie des modules elliptiques, puis la théorie équivalente des faisceaux elliptiques, et à expliquer ensuite l'essentiel du contenu de [L-R-S]. Nous ne dirons rien des revêtements des M_J^d construits dans [Dr 2], ni des revêtements analogues des $\mathcal{E}ll_{X,\mathcal{D},I}$ définis dans [L-R-S], ni des Shtukas.

0.11. Notations. Dans tout ce qui suit, X désigne une courbe algébrique

lisse, projective et géométriquement connexe sur le corps fini \mathbb{F}_q (avec q une puissance p^e d'un nombre premier p). On note $F = \mathbb{F}_q(X)$ le corps des fonctions rationnelles sur X ; \bar{F} désigne une clôture séparable de F .

Pour tout point fermé $x \in |X|$ (correspondant à une place de F), on désigne par \mathcal{O}_x et F_x les complétés respectifs de $\mathcal{O}_{X,x}$ et F en x . Le corps résiduel de \mathcal{O}_x est noté $\kappa(x)$, de cardinal $q_x = q^{\deg(x)}$.

La valuation $v_x : F_x \rightarrow \mathbb{Z}$ est telle que $v_x(\varpi_x) = 1$, pour ϖ_x une uniformisante de \mathcal{O}_x ; la valeur absolue $|\cdot|_x$ est $q_x^{-v_x(\cdot)}$.

On fixe un tel point fermé ∞ , et l'on dénote par $A = H^0(X - \{\infty\}, \mathcal{O}_X)$ le sous-anneau de F constitué des éléments qui sont entiers à toutes les places $x \neq \infty$. En un tel x , le complété \hat{A}_x s'identifie à \mathcal{O}_x .

Quelques simplifications, bien que mineures, apparaissent avec l'hypothèse que $\deg(\infty) = 1$; nous ferons cette hypothèse pour alléger l'exposé.

0.12. Dans l'analogie, qu'il est utile de garder à l'esprit, entre corps de fonctions et corps de nombres, F correspond au corps \mathbb{Q} , ∞ à la place réelle de \mathbb{Q} , et $|X| - \{\infty\}$ à l'ensemble des nombres premiers. L'analogie de A est l'anneau \mathbb{Z} . Les différents complétés \mathcal{O}_x (resp. F_x), pour $x \neq \infty$, correspondent aux \mathbb{Z}_p (resp. \mathbb{Q}_p), tandis que l'analogie de F_∞ est le corps \mathbb{R} .

0.13. On désignera enfin par \mathbf{A}_F , ou plus simplement par \mathbf{A} , l'anneau des adèles de F (produit restreint des F_x). Pour S un ensemble fini de places, \mathbf{A}^S désigne le produit restreint des F_x pour $x \notin S$. On aura souvent à considérer $\mathbf{A}^\infty = \mathbf{A}^{\{\infty\}}$.

1. MODULES ELLIPTIQUES

1.1. Définition d'un A -module elliptique sur un A -schéma S .

1.1.1. Commençons par examiner le cas particulier où S est le spectre d'un corps K , lequel est donc muni d'un homomorphisme $i : A \rightarrow K$.

Rappelons que, pour K un corps de caractéristique $p > 0$, le groupe

additif $G_{a,K} = \text{Spec } K[T]$ admet comme *endomorphismes*, outre les homothéties $\lambda : T \rightarrow \lambda T$ (pour $\lambda \in K$), l'homomorphisme $\tau : T \rightarrow T^p$ et ses puissances. Il est élémentaire de vérifier que tout endomorphisme f de $G_{a,K}$ peut s'exprimer, de façon unique, comme une somme :

$$f = \lambda_0 + \lambda_1 \tau + \cdots + \lambda_n \tau^n, \quad \lambda_i \in K.$$

D'autre part, cet anneau d'endomorphismes est en général *non commutatif*, car on a la relation évidente :

$$\tau \lambda = \lambda^p \tau.$$

Drinfeld note $K\{\tau\}$ l'ensemble des expressions polynomiales en τ à coefficients dans K comme ci-dessus, muni de l'addition naturelle et du produit qui résulte de la relation de commutation précédente : cet anneau $K\{\tau\}$ peut ainsi être identifié à $\text{End}(G_{a,K})$. Nous dirons que l'élément $f \in K\{\tau\}$ écrit ci-dessus est de *degré* p^n si λ_n est non nul. D'autre part, la dérivée (de l'endomorphisme correspondant) est donnée par : $f' = \lambda_0$.

1.1.2. DÉFINITION. — *Un A -module elliptique (rigidifié) sur K est un homomorphisme d'anneaux : $\phi : A \rightarrow K\{\tau\} = \text{End}(G_{a,K})$, tel que l'on ait pour tout $a \in A$: $(\phi(a))' = i(a)$.*

On montre qu'il existe un entier $d \geq 0$, par définition le *rang* du A -module elliptique, tel que l'on ait pour tout $a \in A - \{0\}$: $\deg \phi(a) = |a|_\infty^d$. On exclut en général le cas trivial $d = 0$, qui correspond à : $\phi(a) = i(a)$.

Il est facile de voir que ϕ prend en fait ses valeurs dans le sous-anneau $K\{\tau^e\} \subset K\{\tau\}$ engendré par τ^e .

EXEMPLE. — Pour $X = \mathbf{P}_{\mathbf{F}_q}^1$ et ∞ le "point à l'infini", A s'identifie à un anneau de polynômes $\mathbf{F}_q[Y]$. La donnée d'un A -module elliptique de rang d sur K revient à la donnée de $\phi(Y)$, qui doit être de la forme :

$$\phi(Y) = i(Y) + \mu_1 \tau^e + \mu_2 \tau^{2e} + \cdots + \mu_d \tau^{de} \quad (\mu_i \in K, \mu_d \neq 0).$$

1.1.3. La définition qui précède se globalise naturellement au cas d'un A -schéma S : remarquons que si L est un fibré en droites sur S (i.e. un \mathcal{O}_S -module localement libre de rang 1), muni d'une action de A sur le S -groupe sous-jacent à L , alors les fibres de L sont isomorphes à des groupes additifs (isomorphismes bien définis à des homothéties près), et on peut alors vérifier si ces fibres sont ou non des modules elliptiques de rang d .

DÉFINITION. — *Un A -module elliptique de rang d sur S est un fibré en droites L sur S , muni d'un homomorphisme $\phi : A \rightarrow \text{End}_{S\text{-gr}}(L)$, tel que pour tout $a \in A$ la différentielle de $\phi(a)$ soit donnée par l'homomorphisme structural $A \rightarrow \mathcal{O}_S$, et tel que pour tout point $s : \text{Spec } K \rightarrow S$ la fibre L_s soit un A -module elliptique de rang d au sens précédent.*

On définit de façon évidente la notion d'homomorphisme (resp. d'isomorphisme) entre modules elliptiques.

1.2. Exemple : description analytique des modules elliptiques au-dessus de F_∞ .

Notons \widehat{F}_∞ le complété d'une clôture séparable de F_∞ , et donnons-nous un réseau Γ de dimension d dans \widehat{F}_∞ , c'est-à-dire un sous- A -module discret (projectif) de rang d . On considère alors le produit infini :

$$\eta_\Gamma(z) = z \prod_{\gamma \in \Gamma - \{0\}} (1 - z/\gamma),$$

et on vérifie qu'il converge pour tout $z \in \widehat{F}_\infty$. On obtient ainsi une application η_Γ de \widehat{F}_∞ dans lui-même, qui se trouve être un endomorphisme additif, surjectif et de noyau Γ : d'où une bijection entre le quotient $\widehat{F}_\infty/\Gamma$ et \widehat{F}_∞ . Transportant alors par cette bijection l'action naturelle de A sur $\widehat{F}_\infty/\Gamma$, Drinfeld prouve qu'on obtient ainsi une structure de A -module elliptique de rang d , ϕ_Γ , sur \widehat{F}_∞ ; par définition, on a donc, pour $a \in A$:

$$\phi_\Gamma(a)(\eta_\Gamma(z)) = \eta_\Gamma(az).$$

De plus, tout module elliptique sur \widehat{F}_∞ provient de cette façon d'un réseau Γ , lequel est bien déterminé à homothétie près. Pour $d = 2$ (resp. $d = 1$), cette description est analogue à la description analytique habituelle des courbes elliptiques (resp. du groupe multiplicatif) sur \mathbf{C} . Une différence essentielle réside en ceci : tandis qu'un \mathbf{Z} -réseau discret dans \mathbf{C} ne peut être que de rang ≤ 2 , ici le rang d peut être quelconque, car \widehat{F}_∞ est de degré infini sur F_∞ .

1.3. Espaces de modules.

1.3.1. Comme dans le cas classique des courbes elliptiques, on commence par introduire la notion de *structure de niveau* J , où $\{0\} \neq J \subsetneq A$ est un idéal. Nous supposons pour simplifier que (L, ϕ) est un module elliptique de rang d sur un $A[J^{-1}]$ -schéma S . Notons alors L_J le sous-groupe de L défini par les équations $\phi(a)(x) = 0$, pour a décrivant J : on peut montrer que ce sous-groupe (qui est muni d'une action de A) est *étale* sur S à fibres isomorphes à $(J^{-1}/A)^d$. Par définition, une structure de niveau J sur (L, ϕ) consiste en la donnée d'un isomorphisme A -linéaire : $\psi : (J^{-1}/A)_S^d \xrightarrow{\sim} L_J$.

1.3.2. THÉORÈME (Drinfeld). — *Soit $\{0\} \neq J \subsetneq A$ un idéal. Alors le foncteur qui à un $A[J^{-1}]$ -schéma S associe l'ensemble des classes d'isomorphisme de modules elliptiques de rang d sur S , munis de structures de niveau J , est représentable par un $A[J^{-1}]$ -schéma affine M_J^d , lisse et purement de dimension relative $(d - 1)$.*

1.3.3. Pour $d = 1$, Drinfeld montre que M_J^1 est le spectre du normalisé de $A[J^{-1}]$ dans l'extension abélienne de F qui a pour conducteur J , qui est totalement décomposée en la place ∞ , et qui est maximale pour ces propriétés (c'est là un résultat semblable au théorème de Kronecker-Weber).

Lorsque $d = 2$, les M_J^2 constituent des analogues des courbes modulaires $Y(N)$. Cette analogie est rendue plus frappante par un résultat d'*uniformisation p -adique*, semblable à la description usuelle de $Y(N)_{\mathbf{C}}$

comme quotient du demi-plan de Poincaré : pour $d \geq 2$, notons Ω^d le sous-ensemble de $\mathbf{P}^{d-1}(\widehat{F}_\infty)$ constitué des points qui n'appartiennent à aucun hyperplan défini sur F_∞ (par exemple, $\Omega^2 = \mathbf{P}^1(\widehat{F}_\infty) - \mathbf{P}^1(F_\infty)$ est l'analogie du "double" demi-plan de Poincaré $\mathbf{P}^1(\mathbf{C}) - \mathbf{P}^1(\mathbf{R})$). De la théorie analytique (1.2) des modules elliptiques, on peut déduire la description suivante de $M_J^d(\widehat{F}_\infty)$:

$$M_J^d(\widehat{F}_\infty) \simeq \mathrm{GL}_d(F) \setminus [\Omega^d \times (\mathrm{GL}_d(\mathbf{A}^\infty)/U_J)],$$

où U_J désigne le sous-groupe de $\mathrm{GL}_d(\widehat{A}) = \prod_{x \neq \infty} \mathrm{GL}_d(A_x)$ formé des éléments qui sont congrus à 1 modulo J . On peut préciser cette description en définissant sur Ω^d une structure d'espace analytique rigide : la bijection ci-dessus est alors un isomorphisme d'espaces rigide-analytiques.

1.4. Cohomologie de l'espace de modules ($d = 2$).

1.4.1. Les courbes affines $M_{J,F}^2 = M_J^2 \otimes_A F$ sur F constituent, pour J variant, un système projectif : si $J \subset J'$, alors $((J')^{-1}/A)^2$ est inclus dans $(J^{-1}/A)^2$, et on obtient par restriction de la structure de niveau un morphisme (étales) de $M_{J,F}^2$ sur $M_{J',F}^2$. D'autre part, ces courbes admettent des compactifiées canoniques (des courbes projectives et lisses sur F), que nous notons $\overline{M}_{J,F}^2$, et qui constituent encore un système projectif.

Fixons un nombre premier $\ell \neq p$, et intéressons-nous à la limite inductive des groupes de cohomologie ℓ -adique :

$$H^1 = \varinjlim_J H^1(\overline{M}_{J,F}^2 \otimes_F \overline{F}, \overline{\mathbb{Q}}_\ell).$$

Sur chacun de ces groupes, et donc sur la limite, on dispose d'une action du groupe de Galois $\mathrm{Gal}(\overline{F}/F)$. D'autre part, le groupe $\mathrm{GL}_2(\widehat{A}) = \varprojlim \mathrm{GL}_2(A/J)$ agit à droite, par composition sur la structure de niveau, sur les $M_{J,F}^2$, et cette action s'étend aux compactifiées. D'une manière

similaire à la façon dont on définit les correspondances de Hecke sur les courbes modulaires, on étend l'action de $\mathrm{GL}_2(\widehat{\mathbf{A}})$ en une action du groupe $\mathrm{GL}_2(\mathbf{A}^\infty)$ tout entier sur le système projectif des $\overline{M}_{J,F}^2$ (mais non plus sur les $\overline{M}_{J,F}^2$ individuels) : voir [Dr 1] pour cette construction un peu fastidieuse, ou infra (4.2) pour l'analogie dans le cadre des \mathcal{D} -faisceaux elliptiques.

Ainsi, l'espace H^1 se trouve être une représentation du groupe produit $\mathrm{Gal}(\overline{F}/F) \times \mathrm{GL}_2(\mathbf{A}^\infty)$.

1.4.2. On renvoie par exemple à [Bo-Ja] pour la définition précise de ce qu'est une *représentation automorphe parabolique* du groupe $\mathrm{GL}_2(\mathbf{A})$. Disons ici, de façon vague, qu'il s'agit d'une représentation admissible irréductible qui intervient comme sous-quotient de certains espaces de fonctions sur l'espace homogène $\mathrm{GL}_2(F) \backslash \mathrm{GL}_2(\mathbf{A})$. On considère d'habitude des représentations à valeurs complexes, mais la définition garde un sens, pour les corps de fonctions, sur tout corps algébriquement clos de caractéristique 0, par exemple $\overline{\mathbb{Q}_\ell}$ (variante : on peut aussi utiliser un isomorphisme $\overline{\mathbb{Q}_\ell} \simeq \mathbb{C}$).

Une telle représentation Π se décompose en produit tensoriel restreint, étendu à toutes les places x de F , de composantes locales Π_x (des représentations admissibles irréductibles des groupes $\mathrm{GL}_2(F_x)$). Nous noterons $\Pi = \Pi^\infty \otimes \Pi_\infty$, où Π^∞ désigne le produit tensoriel restreint des Π_x pour $x \neq \infty$.

Rappelons aussi que la *représentation spéciale* (ou *de Steinberg*) St_∞ du groupe $\mathrm{GL}_2(F_\infty)$ (voir [Ca 2]) peut se réaliser comme le quotient de l'espace des fonctions localement constantes sur $\mathbf{P}^1(F_\infty)$ — espace où $\mathrm{GL}_2(F_\infty)$ opère par translation à droite — par le sous-espace des fonctions constantes.

Enfin, presque toutes les composantes Π_x d'une représentation automorphe parabolique de $\mathrm{GL}_2(\mathbf{A})$ sont dans la *série principale non ramifiée*, i.e. proviennent, par induction unitaire à partir du sous-groupe des ma-

trices triangulaires supérieures, de deux quasi-caractères non ramifiés μ_x et μ'_x de F_x^* .

1.4.3. Le théorème principal de [Dr 1] décrit complètement la représentation H^1 du groupe $\text{Gal}(\bar{F}/F) \times \text{GL}_2(\mathbf{A}^\infty)$.

THÉORÈME (Drinfeld). — *La représentation H^1 définie ci-dessus est isomorphe à une somme directe : $\bigoplus \Pi^\infty \otimes \sigma(\Pi^\infty)$, étendue aux (classes d'équivalence de) $\bar{\mathbf{Q}}_\ell$ -représentations admissibles irréductibles Π^∞ du groupe $\text{GL}_2(\mathbf{A}^\infty)$ telles que $\Pi = \Pi^\infty \otimes \text{St}_\infty$ soit automorphe parabolique. Alors $\sigma(\Pi^\infty)$ est une $\bar{\mathbf{Q}}_\ell$ -représentation de degré 2 de $\text{Gal}(\bar{F}/F)$, caractérisée en fonction de Π^∞ par la propriété suivante : en toute place x de F telle que Π_x soit principale non ramifiée, associée à deux quasi-caractères μ_x et μ'_x de F_x^* , la représentation $\sigma(\Pi^\infty)$ est non ramifiée ; sa restriction à un groupe de décomposition en x est la somme des deux quasi-caractères non ramifiés qui envoient un élément de Frobenius (géométrique) respectivement sur $\mu(\varpi_x)^{-1}q^{1/2}$ et $\mu'(\varpi_x)^{-1}q^{1/2}$.*

Ce théorème constitue l'exact analogue des résultats d'Eichler-Shimura ; la preuve qu'en donne Drinfeld est proche de cette théorie classique : calcul analytique (rigide) de la cohomologie — dans l'esprit de "l'isomorphisme de Shimura" — et utilisation d'une "relation de congruence".

2. FAISCEAUX ELLIPTIQUES

Rappelons que nous avons supposé pour simplifier que $\text{deg}(\infty) = 1$. Afin d'alléger les notations dans ce qui suit, pour S et T deux \mathbf{F}_q -schémas, nous noterons simplement $S \times T$ leur produit $S \times_{\mathbf{F}_q} T$.

2.1. DÉFINITION. — *Soit S un \mathbf{F}_q -schéma. Un faisceau elliptique de rang*

d (et de pôle ∞) sur S consiste en la donnée d'un diagramme commutatif :

$$\begin{array}{cccccccc}
 \dots & \xleftarrow{j} & \mathcal{F}_{i-1} & \xleftarrow{j} & \mathcal{F}_i & \xleftarrow{j} & \mathcal{F}_{i+1} & \xleftarrow{j} & \dots \\
 & & \nearrow t & & \nearrow t & & \nearrow t & & \nearrow t \\
 \dots & \xleftarrow{\tau j} & \tau \mathcal{F}_{i-1} & \xleftarrow{\tau j} & \tau \mathcal{F}_i & \xleftarrow{\tau j} & \tau \mathcal{F}_{i+1} & \xleftarrow{\tau j} & \dots
 \end{array}$$

où pour chaque $i \in \mathbb{Z}$, \mathcal{F}_i est un $\mathcal{O}_{X \times S}$ -module localement libre (= un fibré vectoriel) de rang d , où

$$\tau \mathcal{F}_i = (\text{id}_X \times \text{Frob}_{S/\mathbb{F}_q})^* \mathcal{F}_i$$

(pull-back par le Frobenius de S), et où les j et t sont des injections $\mathcal{O}_{X \times S}$ -linéaires. Ces données sont astreintes à satisfaire aux conditions suivantes :

- (i) [Périodicité] : $\mathcal{F}_{i+d} = \mathcal{F}_i(\{\infty\} \times S)$, le composé de d morphismes j consécutifs étant l'injection naturelle (induite par $\mathcal{O}_X \hookrightarrow \mathcal{O}_X(\infty)$).
- (ii) [Pôle] : $\mathcal{F}_i/j(\mathcal{F}_{i-1})$ est l'image directe $(\Gamma_\infty)_* \mathcal{A}_i$ d'un \mathcal{O}_S -module inversible \mathcal{A}_i par la section ∞ :

$$\Gamma_\infty : S \rightarrow X \times S, \quad s \mapsto (\infty, s).$$

- (iii) [Zéro] : $\mathcal{F}_i/t(\tau \mathcal{F}_{i-1})$ est l'image directe $(\Gamma_z)_* \mathcal{B}_i$ d'un \mathcal{O}_S -module inversible \mathcal{B}_i par une section $\Gamma_z : S \rightarrow X \times S$, graphe d'un morphisme de \mathbb{F}_q -schémas $z : S \rightarrow X - \{\infty\}$.

- (iv) [Normalisation] : Pour tout point géométrique s de S , la caractéristique d'Euler-Poincaré $\chi(\mathcal{F}_0|_{X \times s})$ est nulle.

2.2. Remarques.

2.2.1. En vertu des conditions (ii) ou (iii), le passage de \mathcal{F}_i à \mathcal{F}_{i+1} fait croître de 1 la caractéristique d'Euler-Poincaré. D'où, pour $i \in \mathbf{Z}$: $\chi(\mathcal{F}_i|_{X \times_S}) = i$.

2.2.2. Le morphisme z de (iii) est uniquement déterminé. On associe ainsi à tout faisceau elliptique sur S un S -point de $X - \{\infty\}$, qu'on appelle son *zéro*.

2.3. Structures de niveau. Donnons-nous un sous-schéma fini non vide $I \subset X - \{\infty\}$. Soit d'autre part (\mathcal{F}_i, j, t) un faisceau elliptique sur S tel que I soit disjoint de l'image du morphisme z (zéro) associé. Dans ces conditions, la restriction $\mathcal{F}_i|_{I \times_S}$ est indépendante (via j) de i : notons la \mathcal{F}_I ; on voit que t induit un isomorphisme ${}^\tau \mathcal{F}_I \simeq \mathcal{F}_I$. Définissons alors une *structure de niveau I* comme la donnée d'un isomorphisme $\mathcal{O}_{I \times_S}$ -linéaire $\iota : \mathcal{O}_{I \times_S}^d \simeq \mathcal{F}_I$ tel que le diagramme suivant soit commutatif :

$$\begin{array}{ccc}
 & & {}^\tau \mathcal{F}_I \\
 & \nearrow \tau_\iota & \downarrow t \\
 \mathcal{O}_{I \times_S}^d & & \mathcal{F}_I \\
 & \searrow \iota &
 \end{array}$$

2.4. L'équivalence.

Drinfeld a montré, dans le bref mais profond article [DR 3], que les notions de faisceau elliptique et de module elliptique étaient équivalentes. Il s'inspirait d'idées provenant du travail de Krichever [Kr] sur les équations différentielles, mettant ainsi en évidence une connexion troublante entre deux sujets éloignés en apparence. Voir aussi l'article [Mu].

Soit $\{0\} \neq J \subsetneq A$ l'idéal de A qui correspond au fermé I de $X - \{\infty\} = \text{Spec}(A)$. Le résultat de Drinfeld peut s'énoncer comme suit :

THÉOREME. — Soit $S \xrightarrow{z} \text{Spec}(A) - I$ un $A[J^{-1}]$ -schéma. Alors il existe une bijection, fonctorielle en S , entre l'ensemble des classes d'isomorphisme de A -modules elliptiques de rang d sur S , munis d'une structure de niveau J , et l'ensemble des classes d'isomorphisme de faisceaux elliptiques de rang d sur S , de zéro z , et munis d'une structure de niveau I .

Le foncteur qui à S associe l'ensemble des classes d'isomorphisme de faisceaux elliptiques de rang d sur S , munis de structures de niveau I , est donc relativement représentable par un $A[J^{-1}]$ -schéma isomorphe à M_J^d .

2.5. Nous allons maintenant donner une idée de la construction de Drinfeld dans un sens, en supposant pour simplifier que S est le spectre d'un corps K .

2.5.1. Partons d'un faisceau elliptique (\mathcal{F}_i, j, t) sur $\text{Spec } K$, de zéro donné par $z : A \rightarrow K$. Considérons alors les K -vectoriels $M_i \stackrel{\text{déf}}{=} H^0(X \otimes K, \mathcal{F}_i)$, lesquels constituent (via j) une suite croissante. Leur réunion M s'identifie pour tout i à $H^0((X - \{\infty\}) \otimes K, \mathcal{F}_i)$ (noter que j est un isomorphisme sur $X - \{\infty\}$). Cette réunion est de façon évidente un $(A \otimes_{\mathbb{F}_q} K)$ -module.

On dispose de plus d'une application $t : M \rightarrow M$ induite par $t : {}^r\mathcal{F}_i \rightarrow \mathcal{F}_{i+1}$. Cette application envoie M_i dans M_{i+1} ; elle est A -linéaire et K -semi-linéaire :

$$\begin{aligned} t(am) &= at(m) \quad \text{pour } a \in A, \\ t(\lambda m) &= \lambda^q(m) \quad \text{pour } \lambda \in K. \end{aligned}$$

2.5.2. Utilisant les notations de (1.1), la semi-linéarité de t fait qu'on peut voir M comme un $K\{\tau^e\}$ -module à gauche, où τ^e opère par t (on oublie pour le moment l'action de A). Montrons que ce $K\{\tau^e\}$ module est libre de rang 1 :

(a) Le fait que le zéro et le pôle soient disjoints entraîne que t induit pour chaque i une injection (isomorphisme si K est parfait) :

$t : \tau(\mathcal{F}_i/j(\mathcal{F}_{i-1})) \rightarrow \mathcal{F}_{i+1}/j(\mathcal{F}_i)$, d'où il résulte que $t : M_i/M_{i-1} \rightarrow M_{i+1}/M_i$ est injectif.

(b) On a : $M_0 = \{0\}$. Dans le cas contraire en effet, il existerait $i \leq 0$ et un élément $x \in M_i - M_{i-1}$. Utilisant (a), on aurait alors pour tout $m \geq 0$: $t^m x \in M_{i+m} - M_{i+m-1}$, d'où le fait que $\dim_K M_{i+m} \geq m+1$ (car $x, tx, \dots, t^m x$ sont linéairement indépendants). Mais pour m grand, cela contredirait (2.2.1), car on a alors

$$\dim_K M_{i+m} = \chi(\mathcal{F}_{i+m}) = i + m.$$

(c) Essentiellement le même argument qu'en (b) prouve que, pour $i \geq 0$, les M_{i+1}/M_i sont de dimension 1 sur K . On voit alors par récurrence que, si μ est une base de M_1 , alors $\{\mu, t\mu, \dots, t^{n-1}\mu\}$ constituent une base de M_n sur K . D'où il résulte bien que M est libre de rang 1 sur $K\{\tau^e\} \simeq K\{t\}$, de base μ .

2.5.3. Fixons donc une telle base, d'où un isomorphisme $M \simeq K\{t\}$ (on identifie t et τ^e). Regardons maintenant l'action de A sur M : le fait que cette action commute à celle de $K\{t\}$ définit un homomorphisme :

$$\phi : A \rightarrow \text{End}_{K\{t\}}(K\{t\}) = K\{t\}^{\text{opp}},$$

qui est aussi un homomorphisme $A \rightarrow K\{t\} = K\{\tau^e\}$ car A est commutatif. La dérivée $(\phi(a))'$ se calcule en considérant l'action de A sur $M/K \cdot t(M) \simeq M_1$: cet espace est naturellement inclus dans la fibre au-dessus du point zéro du faisceau \mathcal{F}_1 , d'où : $(\phi(a))' = z(a)$.

Finalement, on a pour $a \in A - \{0\}$

$$aM_i \subset H^0(X \otimes K, \mathcal{F}_i(-v_\infty(a))) = M_{i-dv_\infty(a)},$$

avec $(-dv_\infty(a))$ le plus petit entier tel que cette inclusion soit toujours valide. Il en résulte que le degré de $\phi(a)$ est égal à : $p^{e(-dv_\infty(a))} = (q^{-v_\infty(a)})^d = |a|_\infty^d$.

On a donc bien obtenu un module elliptique de rang d .

2.5.4. Continuons à identifier M et $K\{t\}$. La donnée d'une structure de niveau I sur le faisceau elliptique est la donnée d'un isomorphisme $(A \otimes K)$ -linéaire :

$$\iota : (A/J)^d \otimes_{\mathbb{F}_q} K \xrightarrow{\sim} M/MJ,$$

tel que l'action du Frobenius sur le membre de gauche corresponde à droite à l'action de t .

Si $\lambda \in [(A/J)^d]^*$ est un élément du \mathbb{F}_q -dual de $(A/J)^d$, on obtient par composition avec ι^{-1} un homomorphisme K -linéaire h_λ de $M/MJ = K\{t\}/K\{t\} \cdot \phi(J)$ dans K , qui vérifie : $h_\lambda(tm) = h_\lambda(m)^q$.

Un tel homomorphisme est déterminé par $\psi_\lambda = h_\lambda(1) \in K$ (d'où $h_\lambda(t^m) = \psi_\lambda^{q^m}$). Cet élément ψ_λ doit satisfaire, pour tout $a \in J$ avec $\phi(a) = u_0 + u_1 t + \dots + u_n t^n$, l'équation :

$$\phi(a)(\psi_\lambda) = u_0 \psi_\lambda + u_1 \psi_\lambda^q + \dots + u_n \psi_\lambda^{q^n} = h_\lambda(\phi(a)) = 0.$$

Autrement dit, ψ_λ doit être un point de J -torsion du module elliptique ϕ . L'application $\lambda \rightarrow \psi_\lambda$ est ainsi une injection (donc en fait une bijection) A -linéaire :

$$\psi : [(A/J)^d]^* \xrightarrow{\sim} (\phi)_{J\text{-tors}}.$$

D'où une structure de niveau J sur ϕ (cela dépend du choix d'un isomorphisme $(A/J)^* \simeq J^{-1}/A$).

On vérifie sans mal que les constructions précédentes sont, à un isomorphisme près, indépendantes du choix de μ .

3. \mathcal{D} -FAISCEAUX ELLIPTIQUES

3.1. Fixons-nous désormais une algèbre centrale simple D de dimension d^2 sur son centre F , *déployée à l'infini*, c'est-à-dire que $D \otimes_F F_\infty \simeq M_d(F_\infty)$. Choisissons d'autre part un faisceau cohérent \mathcal{D} localement libre en \mathcal{O}_X -algèbres, de fibre générique D , tel qu'en tout point fermé x de X , $\mathcal{D}_x = \mathcal{D} \otimes_{\mathcal{O}_X} \mathcal{O}_x$ soit un *ordre maximal* de $D_x = D \otimes_F F_x$: le choix de \mathcal{D} revient à celui de chacun des \mathcal{D}_x , choix qui doit être cohérent en ce sens que, pour presque tout x , \mathcal{D}_x doit être engendré sur \mathcal{O}_x par une base fixée de D sur F .

On note \mathcal{R} l'ensemble des mauvaises places, où l'algèbre D_x n'est pas déployée : pour $x \notin \mathcal{R}$, le couple (D_x, \mathcal{D}_x) est donc isomorphe à $(M_d(F_x), M_d(\mathcal{O}_x))$.

Nous continuons à nous placer sous l'hypothèse simplificatrice que $\deg(\infty) = 1$. La définition que donnent Laumon, Rapoport et Stuhler d'un \mathcal{D} -module elliptique est alors la suivante (nous notons toujours $S \times T$ le produit sur \mathbf{F}_q) :

3.1.1. DÉFINITION. — *Soit S un \mathbf{F}_q -schéma. Un \mathcal{D} -faisceau elliptique (de pôle ∞) sur S consiste en la donnée d'un diagramme commutatif :*

$$\begin{array}{ccccccc}
 \dots & \xrightarrow{j} & \mathcal{E}_{i-1} & \xrightarrow{j} & \mathcal{E}_i & \xrightarrow{j} & \mathcal{E}_{i+1} & \xrightarrow{j} & \dots \\
 & & \nearrow t & & \nearrow t & & \nearrow t & & \nearrow t \\
 \dots & & \xrightarrow{\tau_j} & \tau \mathcal{E}_{i-1} & \xrightarrow{\tau_j} & \tau \mathcal{E}_i & \xrightarrow{j} & \tau \mathcal{E}_{i+1} & \xrightarrow{\tau_j} & \dots
 \end{array}$$

où pour chaque $i \in \mathbf{Z}$, \mathcal{E}_i est un $\mathcal{O}_{X \times S}$ -module localement libre de rang d^2 , muni d'une action à droite de \mathcal{D} compatible à l'action de \mathcal{O}_X , où $\tau \mathcal{E}_i$ a la même signification qu'au paragraphe précédent, et où les j et t sont des injections $\mathcal{O}_{X \times S}$ -linéaires compatibles à l'action de \mathcal{D} . Ces données sont astreintes à satisfaire aux conditions suivantes :

(i) [Périodicité] : $\mathcal{E}_{i+d} = \mathcal{E}_i(\{\infty\} \times S)$, le composé de d morphismes j consécutifs étant l'injection naturelle.

(ii) [Pôle] : $\mathcal{E}_i/j(\mathcal{E}_{i-1})$ est isomorphe comme $\mathcal{O}_{X \times S}$ -module à l'image directe $(\Gamma_\infty)_* \mathcal{A}_i$, par la section ∞ , d'un \mathcal{O}_S -module \mathcal{A}_i localement libre de rang d .

(iii) [Zéro] : $\mathcal{E}_i/t({}^\tau \mathcal{E}_{i-1})$ est isomorphe comme $\mathcal{O}_{X \times S}$ -module à l'image directe $(\Gamma_z) \mathcal{B}_i$ d'un \mathcal{O}_S -module \mathcal{B}_i localement libre de rang d par une section $\Gamma_z : S \rightarrow X \times S$, graphe d'un morphisme de \mathbb{F}_q -schémas $z : S \rightarrow X - \{\infty\} - \mathcal{R}$.

3.1.2. Normalisation. Les conditions (ii) ou (iii) ci-dessus entraînent que les caractéristiques d'Euler-Poincaré des faisceaux \mathcal{E}_i en un point géométrique s de S augmentent de d lorsqu'on passe de \mathcal{E}_i à \mathcal{E}_{i+1} . On peut imposer une condition de normalisation analogue à (2.1) (iv), par exemple la suivante : $\chi(\mathcal{E}_0|_{X \times s}) \in [0, d[$; dans la suite de cet exposé, nous supposerons les \mathcal{D} -faisceaux elliptiques ainsi normalisés.

Dans l'article [L-R-S] au contraire, aucune condition de normalisation n'est imposée : on y considère l'action naturelle (par décalage) de \mathbf{Z} sur le champ classifiant les \mathcal{D} -faisceaux elliptiques, puis on passe au quotient, ce qui revient au même.

3.1.3. De même que pour les faisceaux elliptiques, le morphisme z ("zéro") est bien déterminé : à tout \mathcal{D} -faisceau elliptique sur S correspond donc un S -point de $X - \{\infty\} - \mathcal{R}$.

3.1.4. Exemple : cas où l'algèbre \mathcal{D} est déployée. Supposons que $\mathcal{D} = M_d(F)$ et que $\mathcal{D} = M_d(\mathcal{O}_X)$. A tout faisceau elliptique (\mathcal{F}_i, j, t) on peut associer le triple $(\mathcal{F}_i \boxtimes \mathcal{O}_X^d, j \boxtimes \text{id}, t \boxtimes \text{id})$: c'est un \mathcal{D} -faisceau elliptique (\mathcal{E}_i, j', t') où, pour chaque i , \mathcal{E}_i est la somme directe de d copies de \mathcal{F}_i , sur laquelle \mathcal{D} opère à droite par multiplication matricielle. Réciproquement, il n'est pas difficile de vérifier, en décomposant un \mathcal{D} -faisceau elliptique sous l'action des idempotents diagonaux de \mathcal{D} , que tout \mathcal{D} -faisceau elliptique normalisé est ainsi obtenu (équivalence de Morita).

Dans le cas déployé, on n'obtient donc rien de nouveau.

3.2. Structures de niveau.

Soit $I \neq \emptyset$ un sous-schéma fermé fini de $X - \{\infty\}$; soit d'autre part (\mathcal{E}_i, j, t) un \mathcal{D} -faisceau elliptique sur un schéma S , tel que l'image $z(S)$ du morphisme zéro associé soit disjointe de I . Dans ces conditions, la restriction $\mathcal{E}_I = \mathcal{E}_i|_{I \times S}$ est indépendante (via j) de i , et t induit un isomorphisme ${}^\tau \mathcal{E}_I \xrightarrow{\sim} \mathcal{E}_I$.

On définit alors une *structure de niveau I* sur (\mathcal{E}_i, j, t) comme la donnée d'un isomorphisme $\mathcal{O}_{I \times S}$ -linéaire $\iota : \mathcal{D}_I \boxtimes \mathcal{O}_S \xrightarrow{\sim} \mathcal{E}_I$, compatible à l'action à droite de \mathcal{D}_I sur les deux membres, et rendant commutatif le diagramme :

$$\begin{array}{ccc}
 & & {}^\tau \mathcal{E}_I \\
 & \nearrow^{\tau \iota} & \downarrow t \\
 \mathcal{D}_I \boxtimes \mathcal{O}_S & & \mathcal{E}_I \\
 & \searrow_{\iota} &
 \end{array}$$

3.3. Espace de modules.

Fixons un sous-schéma fini non vide I de $X - \{\infty\}$. Pour $z : S \rightarrow X - \{\infty\} - I - \mathcal{R}$ un schéma au-dessus de $X - \{\infty\} - I - \mathcal{R}$, notons $\mathcal{E}ll_{X, \mathcal{D}, I}(S)$ l'ensemble des classes d'isomorphisme de \mathcal{D} -faisceaux elliptiques sur S , de zéro z , munis d'une structure de niveau J , et normalisés (3.1.2).

Le premier résultat essentiel démontré dans [L-R-S] est le :

3.3.1 THÉORÈME. — *Le foncteur $\mathcal{E}ll_{X, \mathcal{D}, I}$ est représentable par un schéma quasi-projectif sur $X - \{\infty\} - I - \mathcal{R}$:*

$$\mathcal{E}ll_{X, \mathcal{D}, I} \xrightarrow{\text{zéro}} X - \{\infty\} - I - \mathcal{R}.$$

Ce schéma est lisse et purement de dimension relative $(d - 1)$ au-dessus de $X - \{\infty\} - I - \mathcal{R}$. Lorsque D est une algèbre à division, c'est un $(X - \{\infty\} - I - \mathcal{R})$ -schéma projectif.

Par exemple, lorsque D est déployée, on voit, compte tenu de (3.1.4) et du dictionnaire (2.4), que l'on retrouve les schémas de modules de modules elliptiques : $\mathcal{E}ll_{X,\mathcal{D},I}$ est isomorphe à M_J^d pour J l'idéal de A correspondant à I .

3.4. Donnons quelques indications succinctes sur la façon dont ce théorème de représentabilité est établi.

3.4.1. Parmi les ingrédients essentiels de la preuve, on démontre dans [L-R-S] un résultat de (presque) stabilité pour les fibrés \mathcal{E}_i qui interviennent : à savoir, qu'il existe une constante C telle que, pour tout \mathcal{D} -faisceau elliptique (\mathcal{E}_i, j, t) sur un corps, et tout sous-fibré vectoriel non nul \mathcal{F} de \mathcal{E}_0 , on ait :

$$\frac{\deg \mathcal{F} - C}{\text{rg } \mathcal{F}} < \frac{\deg \mathcal{E}_0 - C}{\text{rg } \mathcal{E}_0}$$

(où \deg désigne le degré et rg le rang). On utilise pour cela les propriétés de la filtration canonique de Harder-Narasimhan-Quillen [Ha-Na].

3.4.2. Lorsque I est de degré $\geq C$ (ce que l'on peut toujours supposer, quitte à passer ensuite au quotient par un groupe fini), et que (\mathcal{E}_i, j, t) est muni d'une structure de niveau I , l'inégalité ci-dessus signifie que \mathcal{E}_0 est stable en tant que fibré muni d'une structure de niveau. Or il existe un schéma sur \mathbf{F}_q (quasi-projectif et lisse) classifiant de tels objets. On montre alors qu'il existe un \mathbf{F}_q -schéma $\text{Vect}_{X,\mathcal{D},I}^\bullet$ classifiant les suites :

$$\dots \xleftarrow{j} \mathcal{E}_{-1} \xleftarrow{j} \mathcal{E}_0 \xleftarrow{j} \mathcal{E}_1 \xleftarrow{j} \mathcal{E}_2 \xleftarrow{j} \dots$$

de fibrés vectoriels de rang d^2 sur $X \times S$, munis d'actions compatibles de \mathcal{D} et de structures de niveau I (compatibles) $\mathcal{D}_I \boxtimes \mathcal{O}_S \simeq \mathcal{E}_I$, telles que : les conditions (i) et (ii) de (3.1.1) soient satisfaites, \mathcal{E}_0 soit stable (en tant que fibré à structure de niveau) et normalisé comme en (3.1.2).

D'autre part, il existe aussi un \mathbf{F}_q -schéma Hecke $_{X,\mathcal{D},I}$ classifiant les

diagrammes commutatifs :

$$\begin{array}{ccccccc}
 \dots & \xrightarrow{j} & \mathcal{E}_{i-1} & \xrightarrow{j} & \mathcal{E}_i & \xrightarrow{j} & \mathcal{E}_{i+1} & \xrightarrow{j} & \dots \\
 & & \nearrow & & \nearrow & & \nearrow & & \nearrow \\
 \dots & \xrightarrow{j'} & \mathcal{E}'_{i-1} & \xrightarrow{j'} & \mathcal{E}'_i & \xrightarrow{j'} & \mathcal{E}'_{i+1} & \xrightarrow{j'} & \dots
 \end{array}$$

dont les deux rangs sont des suites de $\text{Vect}_{X,\mathcal{D},I}^\bullet$, et tels que les t soient des injections $\mathcal{O}_{X \times S}$ -linéaires compatibles aux actions de \mathcal{D} et aux structures de niveau, et qui vérifient la condition (3.1.1) (iii).

3.4.3. On prouve que le schéma $\text{Vect}_{X,\mathcal{D},I}^\bullet$ est lisse et quasi-projectif sur \mathbf{F}_q , et que le morphisme :

$$\text{Hecke}_{X,\mathcal{D},I} \xrightarrow{(z,r_1)} (X - \{\infty\} - I - \mathcal{R}) \times \text{Vect}_{X,\mathcal{D},I}^\bullet,$$

donné par le zéro et le premier rang, est lisse de dimension relative $(d-1)$.

On conclut alors à la représentabilité de $\mathcal{E}ll_{X,\mathcal{D},I}$ par un schéma quasi-projectif lisse de dimension relative $(d-1)$ en remarquant que c'est le produit fibré :

$$\begin{array}{ccc}
 \mathcal{E}ll_{X,\mathcal{D},I} & \xrightarrow{r_1} & \text{Vect}_{X,\mathcal{D},I}^\bullet \\
 \downarrow & & \downarrow (\text{id}, \text{Frob}) \\
 \text{Hecke}_{X,\mathcal{D},I} & \xrightarrow{(r_1, r_2)} & \text{Vect}_{X,\mathcal{D},I}^\bullet \times \text{Vect}_{X,\mathcal{D},I}^\bullet \\
 \downarrow z & & \\
 X - \{\infty\} - I - \mathcal{R} & &
 \end{array}$$

3.4.4. La projectivité dans le cas où D est une algèbre à division s'obtient en vérifiant le critère valuatif de propreté : on fait usage pour cela d'un *théorème de réduction semi-stable* développé par Drinfeld dans un travail sur la compactification des espaces de modules de Shtukas [Dr 6].

4. COHOMOLOGIE DE L'ESPACE DE MODULES : LE THÉORÈME FONDAMENTAL

Nous supposons désormais que D est une *algèbre à division*. Pour alléger les notations, nous notons simplement $\mathcal{E}ll_I$ les schémas $\mathcal{E}ll_{X,D,I}$, et $\mathcal{E}ll_{I,F} = \mathcal{E}ll_I \times_X \text{Spec } F$.

4.1. Les variétés $\mathcal{E}ll_{I,F}$ sont projectives et lisses de dimension $(d - 1)$ sur F . Pour $I' \supset I$, la restriction de la structure de niveau définit une projection (finie étale) de $\mathcal{E}ll_{I',F}$ sur $\mathcal{E}ll_{I,F}$. On s'intéresse aux espaces de cohomologie ℓ -adique (pour $\ell \neq p$ un nombre premier fixé, et $0 \leq n \leq 2d - 2$) :

$$H_I^n = H^n(\mathcal{E}ll_{I,F} \otimes_F \bar{F}, \bar{\mathcal{Q}}_\ell),$$

ainsi qu'à leur limite inductive $H^n = \varinjlim_I H_I^n$. Sur chacun de ces espaces, on dispose d'une action du groupe $\text{Gal}(\bar{F}/F)$.

4.2.1. Désignons par G le groupe algébrique sur F défini par : $G(B) = (D \otimes_F B)^*$, pour B une F -algèbre : en particulier, $G(F) = D^*$. Nous notons $G(\mathbf{A})$ (resp. $G(\mathbf{A}^\infty)$) le groupe de ses points adéliques (resp. à valeurs dans les adèles "finies"). On définit une *action à droite* de $G(\mathbf{A}^\infty)$ sur le système projectif des $\mathcal{E}ll_{I,F}$, d'où il résulte une action à gauche (commutant à celle de Galois) sur l'espace H^n . Pour décrire cette action, il suffit de dire comment agissent les éléments "entiers" $g \in G(\mathbf{A}^\infty) \cap \mathcal{D}_{\hat{A}}$, avec $\hat{A} = \prod_{x \neq \infty} \mathcal{O}_x$ et $\mathcal{D}_{\hat{A}} = D \otimes_A \hat{A} = \varprojlim_I \mathcal{D}_I = \prod_{x \neq \infty} \mathcal{D}_x$.

4.2.2. Pour $I = \text{Spec}(A/J)$ un fermé, notons U_I le sous-groupe de $\mathcal{D}_{\hat{A}}^*$ constitué des éléments congrus à 1 modulo J . Donnons-nous g entier, et I et I' tels que $gU_I g^{-1} \subset U_{I'}$.

La donnée d'une structure de niveau I sur un \mathcal{D} -faisceau elliptique (\mathcal{E}_i, j, t) , défini sur un corps séparablement clos K extension de F , revient à la donnée d'une classe modulo U_I de structures de niveau $\iota : \mathcal{D}_{\widehat{A}} \otimes K \xrightarrow{\sim} \mathcal{E}_i \otimes_{A \otimes K} (\widehat{A} \otimes K)$.

L'élément g définit un endomorphisme de $\mathcal{D}_{\widehat{A}}$, et donc (via ι) de $\mathcal{E}_i \otimes_{A \otimes K} (\widehat{A} \otimes K)$.

A ces données, on associe alors le \mathcal{D} -faisceau elliptique (\mathcal{E}'_i, j', t') qui s'insère dans des diagrammes cartésiens :

$$\begin{array}{ccc} \mathcal{E}'_i & \longrightarrow & \mathcal{E}_i \otimes_{A \otimes K} (\widehat{A} \otimes K) \\ \downarrow & & \downarrow g \\ \mathcal{E}_i & \longrightarrow & \mathcal{E}_i \otimes_{A \otimes K} (\widehat{A} \otimes K). \end{array}$$

En termes plus concrets, les sections locales de \mathcal{E}'_i sont les sections locales de \mathcal{E}_i qui vérifient certaines conditions (être dans l'image de g) en les x où g_x n'est pas bijectif.

On munit (\mathcal{E}'_i, j', t') de la structure de niveau ι' (bien définie modulo $U_{I'}$) qui rend commutatif le diagramme suivant :

$$\begin{array}{ccc} \mathcal{D}_{\widehat{A}} \otimes K & \xrightarrow{\iota'} & \mathcal{E}'_i \otimes_{A \otimes K} (\widehat{A} \otimes K) \\ \downarrow g & & \downarrow \\ \mathcal{D}_{\widehat{A}} \otimes K & \xrightarrow{\iota} & \mathcal{E}_i \otimes_{A \otimes K} (\widehat{A} \otimes K). \end{array}$$

On peut alors vérifier que l'application $(\mathcal{E}_i, j, t, \iota) \rightarrow (\mathcal{E}'_i, j', t', \iota') \bmod \mathbf{Z}$ (i.e. au décalage près pour lequel la condition de normalisation (3.1.2) est satisfaite) provient d'un morphisme $g : \mathcal{E}ll_{I, F} \rightarrow \mathcal{E}ll_{I', F}$.

4.2.3. La construction qui précède définit l'action de $G(\mathbf{A}^\infty)$ sur le système projectif, ou si l'on préfère sur la limite projective $\mathcal{E}ll_F$ des

$\mathcal{E}ll_{I,F}$. On peut aussi parler de *correspondances de Hecke* : un élément $g \in G(\mathbf{A}^\infty)$ — ou plutôt sa double classe modulo U_I — induit sur $\mathcal{E}ll_{I,F}$ la correspondance :

$$\begin{array}{ccc}
 & \mathcal{E}ll_F / (U_I \cap g^{-1}U_I g) & \\
 \alpha \swarrow & & \searrow \beta \\
 \mathcal{E}ll_F / U_I & \leftarrow \text{-----} & \mathcal{E}ll_F / U_I
 \end{array}$$

avec α le morphisme qui provient de l'inclusion de $U_I \cap g^{-1}U_I g$ dans U_I , tandis que β provient de l'action de g^{-1} sur $\mathcal{E}ll_F$ et de l'injection $u \rightarrow gug^{-1}$ de $U_I \cap g^{-1}U_I g$ dans U_I .

Pour $x \neq \infty$ une place telle que la composante g_x soit triviale, on peut vérifier que cette correspondance s'étend en une correspondance sur $\mathcal{E}ll_I \times_X \text{Spec}(\mathcal{O}_{X,x})$, avec des morphismes α et β étales finis.

4.3.1. Considérons l'espace $L^2 = L^2(G(F) \backslash G(\mathbf{A}) / \varpi_\infty^{\mathbf{Z}})$ des fonctions de carré intégrable sur l'espace (compact) $G(F) \backslash G(\mathbf{A}) / \varpi_\infty^{\mathbf{Z}}$. Cet espace L^2 , où le groupe $G(\mathbf{A})$ opère par la représentation régulière, se décompose discrètement comme une somme hilbertienne :

$$L^2 = \widehat{\bigoplus}_{\Pi} m(\Pi)\Pi,$$

avec des multiplicités $m(\Pi)$ finies. Les représentations (irréductibles) Π qui interviennent dans cette somme sont par définition les *représentations automorphes* (de caractère central trivial sur ϖ_∞) du groupe $G(\mathbf{A})$. On peut donner un sens plus algébrique à cette notion en considérant, plutôt que l'espace L^2 , l'espace des fonctions localement constantes, dont on regarde tous les sous-quotients irréductibles (cf. [Bo-Ja]). Cela garde un sens lorsque le corps \mathbb{C} est remplacé par un quelconque corps

algébriquement clos de caractéristique 0, par exemple $\overline{\mathbb{Q}_\ell}$. C'est de ce point de vue algébrique que nous nous placerons dans la suite.

Toute représentation automorphe Π se décompose en un produit tensoriel (restreint) $\otimes \Pi_x$ de représentations (admissibles irréductibles) des groupes $G(F_x)$. Nous noterons encore, comme en (1.4.2),

$$\Pi = \Pi^\infty \otimes \Pi_\infty.$$

4.3.2. Parmi les représentations automorphes de $G(\mathbf{A})$, notons les *caractères* $\chi \circ \text{Nr}$, factorisés via la norme réduite $\text{Nr} : G(\mathbf{A}) \rightarrow \mathbf{A}^*$ en un caractère de Hecke χ de \mathbf{A}^* : ils interviennent avec multiplicité 1. Toutes les autres représentations automorphes sont de dimension infinie (mieux : leurs composantes Π_x aux places $x \notin \mathcal{R}$ sont de dimension infinie).

4.3.3. Pour $x \notin \mathcal{R}$, une représentation (admissible irréductible) Π_x du groupe $G(F_x) \simeq \text{GL}_d(F_x)$ est *non ramifiée* si elle admet un vecteur fixe sous le sous-groupe compact maximal $\text{GL}_d(\mathcal{O}_x)$ (voir [Ca 1]). Elle intervient alors comme constituant d'une induite $\text{Ind}(\mu_1, \dots, \mu_d)$ (induction unitaire) de d quasi-caractères non ramifiés de F_x^* . Les $z_i = \mu_i(\varpi_x)$ sont bien déterminés à l'ordre près : ce sont les *valeurs propres de Hecke* de Π_x .

4.3.4. Enfin, la *représentation de Steinberg* St_∞ de $G(F_\infty) \simeq \text{GL}_d(F_\infty)$ est l'unique quotient irréductible de l'induite

$$\text{Ind}(| \cdot |_\infty^{-\frac{d-1}{2}}, | \cdot |_\infty^{-\frac{d-3}{2}}, \dots, | \cdot |_\infty^{-\frac{d-3}{2}}, | \cdot |_\infty^{-\frac{d-1}{2}})$$

(induction unitaire à partir du groupe des matrices triangulaires supérieures). C'est une représentation de la série discrète.

4.4.1. Considérons les semi-simplifiées $(H^n)^{ss}$ des H^n . Ces représentations du produit $G(\mathbf{A}^\infty) \times \text{Gal}(\overline{F}/F)$ se décomposent en composantes isotypiques suivant les (classes d'équivalence de) représentations admissibles irréductibles Π^∞ de $G(\mathbf{A}^\infty)$

$$(H^n)^{ss} = \bigoplus_{\Pi^\infty} \Pi^\infty \otimes V_{\Pi^\infty}^n,$$

où les $V_{\Pi^\infty}^n$ sont des représentations de dimension finie de $\text{Gal}(\bar{F}/F)$. Le résultat central de [L-R-S] décrit ces décompositions : il affirme en bref que seules interviennent non trivialement dans les sommes ci-dessus des Π^∞ qui sont la composante “finie” de représentations automorphes Π qui sont ou bien des caractères, ou bien telles que $\Pi_\infty \simeq \text{St}_\infty$. Celles du premier type sont associées à des caractères galoisiens (intervenant en les degrés pairs de cohomologie) tandis que celles du second type produisent des représentations galoisiennes plus intéressantes, liées aux conjectures globales de Langlands; ces dernières représentations interviennent en degré médian $(d - 1)$.

4.4.2 THÉORÈME.

(i) *Seules peuvent intervenir non trivialement, dans les décompositions isotypiques des $(H^n)^{ss}$, des représentations Π^∞ telles que, ou bien $\Pi^\infty \otimes \mathbf{1}$, ou bien $\Pi^\infty \otimes \text{St}_\infty$, soit automorphe. Dans le premier cas, Π^∞ se factorise sous la forme $\chi^\infty \circ \text{Nr}$, où χ^∞ est un caractère de $(\mathbf{A}^\infty)^*$ tel que $\chi = \chi^\infty \otimes \mathbf{1}$ soit un caractère de Hecke de \mathbf{A}^* .*

(ii) *Dans le premier cas ($\Pi^\infty = \chi^\infty \circ \text{Nr}$), $V_{\Pi^\infty}^0$ est le caractère galoisien qui correspond à χ par la théorie du corps de classes (normalisation “géométrique”); on a, pour $0 \leq i \leq d - 1$: $V_{\Pi^\infty}^{2i} = V_{\Pi^\infty}^0(-i)$ (twist à la Tate), tandis que les $V_{\Pi^\infty}^{2i+1}$ sont nuls.*

(iii) *Dans le second cas (où $\Pi = \Pi^\infty \otimes \text{St}_\infty$ est automorphe), les $V_{\Pi^\infty}^n$ sont nuls pour $n \neq d - 1$, tandis que $V_{\Pi^\infty}^{d-1}$ est une représentation semi-simple de dimension $m(\Pi)d$ de $\text{Gal}(\bar{F}/F)$, caractérisée par la propriété suivante : pour toute place $x \notin \mathcal{R} \cup \{\infty\}$ telle que la composante Π_x soit non ramifiée, alors $V_{\Pi^\infty}^{d-1}$ est non ramifiée en x , et l'on a :*

$$\text{tr}(\text{Frob}_x^r; V_{\Pi^\infty}^{d-1}) = m(\Pi) q_x^{r(d-1)/2} (z_1(\Pi_x)^r + \cdots + z_d(\Pi_x)^r),$$

où Frob_x désigne un élément de Frobenius (géométrique) et où les $z_i(\Pi_x)$ sont les valeurs propres de Hecke de Π_x .

(iv) *De plus — toujours dans le second cas — la Frob_∞ -semisimplifiée (cf. [De]) de $V_{\Pi_\infty}^{d-1} |_{\text{Gal}(\bar{F}_\infty/F_\infty)}$ est isomorphe à la somme de $m(\Pi)$ copies de $\text{Sp}_\infty(-d+1)$, i.e. de la représentation spéciale tordue à la Tate.*

4.4.3. Comme corollaire de (iii) — et de l'ex-conjecture de Weil — on obtient une *preuve de la conjecture de Ramanujan-Petersson* pour les représentations automorphes Π de $G(\mathbf{A})$ telles que $\Pi_\infty \simeq \text{St}_\infty$: c'est-à-dire que *les valeurs propres de Hecke de Π_x (en une place où Π_x est non ramifiée) sont de valeur absolue 1 pour tout plongement complexe.*

4.4.4. Commentaire. On conjecture que la multiplicité $m(\Pi)$ qui intervient dans l'énoncé du théorème est toujours égale à 1 ; cela doit résulter de la théorie de la correspondance de Jacquet-Langlands généralisée, entre représentations automorphes de $G(\mathbf{A})$ et certaines représentations automorphes de $\text{GL}_d(\mathbf{A})$: cf. [De-Ka-Vi] et [Ro]. Cette théorie n'est pas encore au point pour les corps de fonctions (sauf pour $d = 2$ ou 3). Lorsque $m(\Pi) = 1$, la représentation $V_{\Pi_\infty}^{d-1}$ est (à torsion près par un quasi-caractère) celle que les conjectures de Langlands associent à Π .

5. INDICATIONS SUR LA PREUVE DU THÉORÈME PRINCIPAL

Les différentes étapes de cette preuve couvrent une grande partie de l'article [L-R-S] ; nous n'en donnerons ici qu'une idée schématique. Le principe en est de calculer d'abord, au moyen de formules des traces — et c'est la tâche la plus longue — les représentations virtuelles $H_I^* = \sum (-1)^n H_I^n$, et ensuite d'en extraire, par des arguments géométriques, les H_I^n individuels.

5.1. Fixons un point fermé $x \in |X| - \{\infty\} - \mathcal{R}$, et soit I un sous-schéma fermé de $X - \{\infty, x\}$. Considérons alors l'*algèbre de Hecke* $\mathcal{H}_I^{x,\infty}$ des fonctions localement constantes (à valeurs dans $\bar{\mathbb{Q}}_\ell$) à support compact sur $G(\mathbf{A}^{x,\infty})$, bi-invariantes par U_I (défini en (4.2.2)). Cette algèbre opère sur H_I^n : pour $g^{x,\infty} \in G(\mathbf{A}^{x,\infty})$, la fonction caractéristique de la double

classe $U_I g^{x,\infty} U_I$ opère via la correspondance de Hecke (4.2.3) associée à $g^{x,\infty}$ (auquel on a ajouté la composante 1 en x).

Or la cohomologie H_I^n de $\mathcal{E}ll_I \times_X \text{Spec } \overline{F}$ s'identifie à la cohomologie de la fibre spéciale $\mathcal{E}ll_I \times_X \text{Spec } \overline{\kappa(x)}$, car le schéma considéré est propre et lisse sur $\mathcal{O}_{X,x}$. De plus, comme on l'a vu plus haut (4.2.3), les correspondances de Hecke ont "bonne réduction" en x . On peut donc faire usage de la *formule des traces de Lefschetz*, appliquée à la fibre spéciale en x , afin de calculer, pour $f^{x,\infty} \in \mathcal{H}_I^{x,\infty}$ et $r > 0$:

$$\text{Lef}_r(f^{x,\infty}) \stackrel{\text{déf}}{=} \text{tr}(\text{Frob}_x^r \times f^{x,\infty}; H_I^*).$$

Ce calcul est possible à partir du moment où l'on dispose d'une description précise de l'ensemble des points de la fibre spéciale, ainsi que de l'action du Frobenius et des correspondances de Hecke.

5.2. La description des points de la fibre spéciale se fait en deux temps, de façon similaire à celle dont on décrit des ensembles de variétés abéliennes sur un corps fini : on commence par partitionner l'ensemble considéré en *classes d'isogénie* (à la Honda-Tate), puis on paramétrise chaque classe d'isogénie.

Si (\mathcal{E}_i, j, t) est un \mathcal{D} -faisceau elliptique sur $k = \overline{\kappa(x)}$ (de zéro $z = x : \text{Spec } k \rightarrow X$), sa *fibre générique* V est la fibre de \mathcal{E}_0 au point générique $\text{Spec}(F \otimes_{\mathbf{F}_q} k)$ de $X \otimes_{\mathbf{F}_q} k$: c'est donc un $(F \otimes k)$ -espace vectoriel muni d'une action de D ; d'autre part, on peut voir V (via j) comme la fibre générique de chacun des \mathcal{E}_i , et t induit alors un isomorphisme semi-linéaire $\varphi : V \rightarrow V$: suivant une terminologie due à Drinfeld, on dit que V est un φ -*espace*, muni d'une action de D . Deux \mathcal{D} -faisceaux elliptiques sont *isogènes* si leurs fibres génériques sont isomorphes.

Utilisant une classification, due à Drinfeld, des φ -espaces, on montre dans [L-R-S] que les classes d'isogénie de \mathcal{D} -faisceaux elliptiques sont en bijection avec les classes d'isomorphisme de " (D, x, ∞) -types", qui sont des couples $(\tilde{F}, \tilde{\pi})$, avec \tilde{F} une extension finie séparable de F et $\tilde{\pi} \in \tilde{F}^* \otimes \mathbb{Q}$, vérifiant les propriétés suivantes :

- Pour $\tilde{F}' \subsetneq \tilde{F}$ un sous-corps, $\tilde{\pi} \notin (\tilde{F}')^* \otimes \mathbb{Q}$.
- $[\tilde{F} : F]$ divise d .
- \tilde{F} n'a qu'une place $\tilde{\infty}$ divisant ∞ ; on a :

$$\deg(\tilde{\infty}) \cdot v_{\tilde{\infty}}(\tilde{\pi}) = -[\tilde{F} : F]/d.$$

• Il n'existe qu'une place $\tilde{x} \neq \tilde{\infty}$ de \tilde{F} où $v_{\tilde{x}}(\tilde{\pi}) \neq 0$; cette place divise x .

- Pour toute place \tilde{y} de \tilde{F} divisant une place y de F , on a :

$$(d[\tilde{F}_{\tilde{y}} : F_y]/[\tilde{F} : F]) \operatorname{inv}_y(D) \in \mathbb{Z}$$

(on a noté $v_{\tilde{\infty}}$ (resp. $v_{\tilde{x}}$) la valuation de F qui étend v_{∞} (resp. v_x)).

Ensuite, on donne une description combinatoire de la classe d'isogénie (munie de l'action de Frobenius et des correspondances de Hecke) associée à un tel (D, x, ∞) -type.

5.3. L'étape suivante consiste à appliquer la formule des traces de Lefschetz, c'est-à-dire à dénombrer des points fixes. Après un travail combinatoire qui repose sur la description des points esquissée ci-dessus, on parvient à une expression de $\operatorname{Lef}_r(f^{x, \infty})$ ($r > 0$) du type suivant :

(5.3.1)

$$\operatorname{Lef}_r(f^{x, \infty}) = \sum_{\gamma} \operatorname{vol}(G_{\gamma}(F) \backslash G_{\gamma}(\mathbf{A}) / \varpi_{\infty}^{\mathbf{Z}}) \cdot \frac{\varepsilon_{\infty}(\gamma)}{\operatorname{vol}(\bar{D}_{\infty, \gamma}^* / \varpi_{\infty}^{\mathbf{Z}})} \cdot O_{\gamma}(f_x^{(r)} f^{x, \infty})$$

où la somme est étendue aux classes de conjugaison γ de $G(F) = D^*$ qui sont elliptiques à la place ∞ . On a noté G_{γ} le centralisateur de γ . Par définition, $\varepsilon_{\infty}(\gamma) = (-1)^{d'-1}$, avec $d' = d/[F_{\infty}(\gamma) : F_{\infty}]$, et $\bar{D}_{\infty, \gamma}$ est le corps gauche de centre $F_{\infty}(\gamma)$ et d'invariant $1/d'$ (de sorte que $\bar{D}_{\infty, \gamma}^*$ est une forme intérieure de $G_{\gamma}(F_{\infty}) \simeq \operatorname{GL}_{d'}(F_{\infty}(\gamma))$). Comme d'habitude, pour f^{∞} une fonction localement constante à support compact sur

$G(\mathbf{A}^\infty)$, $O_\gamma(f^\infty)$ désigne l'intégrale orbitale :

$$O_\gamma(f^\infty) = \int_{G_\gamma(\mathbf{A}^\infty) \backslash G(\mathbf{A}^\infty)} f^\infty((h^\infty)^{-1} \gamma h^\infty) \frac{dh^\infty}{dh_\gamma^\infty}$$

(la formule (5.3.1) dépend d'un choix compatible, que nous ne préciserons pas ici, des différentes mesures de Haar qui interviennent).

Finalement, $f_x^{(r)} \in \mathcal{C}_c^\infty(D_x^* // \mathcal{D}_x^*) \simeq \mathcal{C}_c^\infty(\mathrm{GL}_d(F_x) // \mathrm{GL}_d(\mathcal{O}_x))$ est la fonction localement constante à support compact sur D_x^* , bi-invariante par \mathcal{D}_x^* , de transformée de Satake (cf. [Ca 1]) :

$$(f_x^{(r)})^\vee(z_1, \dots, z_d) = q_x^{r(d-1)/2} (z_1^r + \dots + z_d^r).$$

Cette fonction est telle que, pour Π_x une représentation admissible irréductible de $G(F_x)$, on ait :

(5.3.2)

$$\mathrm{tr} \Pi_x(f_x^{(r)}) = \begin{cases} 0 & \text{si } \Pi_x \text{ est ramifiée,} \\ q_x^{r(d-1)/2} (z_1^r + \dots + z_d^r) & \text{si } \Pi_x \text{ est non ramifiée, de} \\ & \text{valeurs propres de Hecke } z_1, \dots, z_d \end{cases}$$

(pour la mesure qui donne le volume 1 à \mathcal{D}_x^*).

Dans des travaux antérieurs, Drinfeld avait donné une expression explicite pour $f_x^{(r)}$, et calculé ses intégrales orbitales, résultats qui jouent un rôle crucial dans la preuve de (5.3.1).

5.4. Soit $f = \otimes f_v$ une fonction localement constante à support compact sur $G(\mathbf{A})/\varpi_\infty^{\mathbf{Z}}$, produit de fonctions f_v sur les $G(F_v)$, presque partout égales à la fonction caractéristique de \mathcal{D}_v^* . La formule des traces de Selberg exprime la trace de l'opérateur induit par f sur l'espace $L^2(G(F) \backslash G(\mathbf{A})/\varpi_\infty^{\mathbf{Z}})$, c'est-à-dire (cf. (4.3.1)) la somme $\sum m(\Pi) \mathrm{tr} \Pi(f)$ pour toutes les représentations automorphes de caractère central trivial sur ϖ_∞ . Cette formule est simple car on travaille ici sur un quotient compact; elle s'écrit

$$(5.4.1) \quad \sum_{\Pi} m(\Pi) \mathrm{tr} \Pi(f) = \sum_{\gamma} \mathrm{vol}(G_\gamma(F) \backslash G_\gamma(\mathbf{A})/\varpi_\infty^{\mathbf{Z}}) O_\gamma(f),$$

où la somme du membre de droite est étendue à toutes les classes de conjugaison γ de $G(F)$, et où O_γ désigne l'intégrale orbitale

$$O_\gamma(f) = \int_{G_\gamma(\mathbf{A}) \backslash G(\mathbf{A})} f(h^{-1}\gamma h) \frac{dh}{dh_\gamma}.$$

5.5. Il existe une "fonction d'Euler-Poincaré" f_∞ , localement constante sur le groupe $G(F_\infty)$, à support compact modulo le centre, et dont on peut donner une expression explicite comme combinaison linéaire de fonctions caractéristiques de parahoriques [Lau 2], qui vérifie les propriétés suivantes :

5.5.1. Ses intégrales orbitales $O_\gamma(f_\infty) = \int_{G_\gamma(F_\infty) \backslash G(F_\infty)} f_\infty(h_\infty^{-1}\gamma h_\infty) \frac{dh_\infty}{dh_{\gamma,\infty}}$ sont données par :

$$O_\gamma(f_\infty) = \begin{cases} 0 & \text{si } \gamma \text{ n'est pas elliptique dans } G(F_\infty), \\ \frac{\varepsilon_\infty(\gamma)}{\text{vol}(\overline{D}_{\infty,\gamma}^*/\varpi_\infty^{\mathbf{Z}})} & \text{si } \gamma \text{ est elliptique dans } G(F_\infty). \end{cases}$$

5.5.2. Si Π_∞ est une représentation admissible irréductible unitarisable de $G(F_\infty)/\varpi_\infty^{\mathbf{Z}}$, on a :

$$\text{tr } \Pi_\infty(f_\infty) = \begin{cases} 0 & \text{si } \Pi_\infty \not\simeq \mathbf{1}, \text{ St}_\infty, \\ 1 & \text{si } \Pi_\infty \simeq \mathbf{1}, \\ (-1)^{d-1} & \text{si } \Pi_\infty \simeq \text{St}_\infty \end{cases}$$

(autrement dit, f_∞ "isole" $\mathbf{1}$ et St_∞ dans le dual unitaire).

5.6. Replaçant les formules (5.5.1) ci-dessus dans l'expression (5.3.1) de Lef_r , on obtient :

$$\text{Lef}_r(f^{x,\infty}) = \sum_\gamma \text{vol}(G_\gamma(F) \backslash G_\gamma(\mathbf{A})/\varpi_\infty^{\mathbf{Z}}) O_\gamma(f_x^{(r)} f_\infty f^{x,\infty})$$

où la somme est étendue à toutes les classes de conjugaison γ de $G(F)$. Cette somme s'interprète au moyen de la formule de Selberg : compte

tenu de (5.5.2), on arrive à :

$$\text{Lef}_r(f^{x,\infty}) = \sum_{\Pi_\infty \simeq 1} m(\Pi) \text{tr} \Pi^\infty(f^\infty) + (-1)^{d-1} \sum_{\Pi_\infty \simeq \text{St}_\infty} m(\Pi) \text{tr} \Pi^\infty(f^\infty)$$

(avec $f^\infty = f_x^{(r)} f^{x,\infty}$). Utilisons maintenant la formule (5.3.2) qui donne $\text{tr} \Pi_x(f_x^{(r)})$, le fait que $m(\Pi) = 1$ pour Π un caractère, ainsi que le suivant : pour Π_x un caractère non ramifié de $G(F_x)$, ses valeurs propres de Hecke sont les $\Pi_x(\varpi_x) q_x^{(1-d)/2+i}$ pour $0 \leq i \leq d-1$. On trouve finalement :

$$(5.6.1) \quad \text{Lef}_r(f^{x,\infty}) = \sum_{\substack{\Pi \text{ caractère,} \\ \Pi_\infty \simeq 1, \Pi_x \text{ non ramifié}}} \Pi^{x,\infty}(f^{x,\infty}) \cdot \Pi_x(\varpi_x)^r \cdot (1 + q_x^r + \dots + q_x^{r(d-1)})$$

$$+ (-1)^{d-1} \sum_{\substack{\Pi_\infty \simeq \text{St}_\infty \\ \Pi_x \text{ non ramifiée}}} \text{tr} \Pi^{x,\infty}(f^{x,\infty}) \cdot m(\Pi) \cdot q_x^{r(d-1)/2} \cdot (z_1(\Pi_x)^r + \dots + z_d(\Pi_x)^r).$$

5.7. La formule que l'on vient d'obtenir signifie que la représentation virtuelle $H^* = \sum (-1)^n H^n$ admet la trace que prédit le théorème (4.4.2). Il est aisé d'en déduire les parties (i) et (ii) du théorème : pour (ii), on utilise la "conjecture" de Weil, qui force les différentes valeurs propres $\Pi_x(\varpi_x) q_x^i$ à provenir des espaces de cohomologie de degré $2i$. Un argument facile analogue marcherait aussi pour la partie (iii) si l'on disposait déjà de la conjecture de Ramanujan-Petersson ; il suffirait même de disposer d'un renseignement beaucoup plus faible, à savoir que la composante Π_x est *générique* (i.e. admet un modèle de Whittaker), ce que l'on conjecture être toujours vrai : cela résulterait de la correspondance de Jacquet-Langlands généralisée. En effet, on sait [Ja-Sh] que les valeurs propres de Hecke d'une représentation générique non ramifiée unitaire vérifient : $|z_i(\Pi_x)| < q_x^{1/2}$.

Comme on ne sait pas encore à l'heure actuelle prouver la généralité de Π_x , Laumon, Rapoport et Stuhler utilisent un argument plus complexe, de nature géométrique, pour prouver les parties (iii) et (iv) du théorème.

Cet argument (où on étudie la représentation galoisienne restreinte à $\text{Gal}(\overline{F}_\infty/F_\infty)$) repose sur le théorème de Lefschetz difficile, sur le théorème (de Deligne) de pureté de la filtration de monodromie, et utilise aussi les propriétés des fonctions L .

6. LA CONJECTURE DE LANGLANDS LOCALE POUR LES CORPS D'ÉGALE CARACTÉRISTIQUE

6.1. Soit K un corps local non archimédien et $d \geq 1$ un entier. Désignons par $\mathcal{A}_K(d)$ l'ensemble des (classes d'équivalence de) représentations complexes admissibles irréductibles du groupe $\text{GL}_d(K)$, et par $\mathcal{G}_K(d)$ l'ensemble des (classes d'équivalence de) représentations continues complexes ϕ -semi-simples de dimension d du groupe de Weil-Deligne W'_K . La *conjecture de Langlands locale* (cf. [He 1]) prédit l'existence pour chaque d d'une bijection entre $\mathcal{A}_K(d)$ et $\mathcal{G}_K(d)$, ces bijections étant astreintes à vérifier de nombreuses propriétés (préservation des facteurs L et ε qu'on sait définir de chaque côté pour les paires de représentations, compatibilité à la torsion, au passage à la contragrédiente...). Ces bijections doivent se restreindre en des bijections entre $\mathcal{A}_K^0(d)$, ensemble des représentations *cuspidales* irréductibles à caractère central d'ordre fini, et $\mathcal{G}_K^0(d)$, ensemble des représentations *irréductibles* de degré d de $\text{Gal}(\overline{K}/K)$.

Réciproquement, si on dispose pour tout d de bijections convenables entre $\mathcal{A}_K^0(d)$ et $\mathcal{G}_K^0(d)$, il existe une façon canonique de les étendre en des bijections entre $\mathcal{A}_K(d)$ et $\mathcal{G}_K(d)$ (cf. [He 1], [He 4]).

Kutzko, pour $d = 2$, puis Henniart, pour $d = 3$, avaient prouvé cette conjecture locale. Le dernier résultat fondamental de [L-R-S] est une preuve de la conjecture de Langlands locale (pour tout d) pour K un corps local d'égale caractéristique p (autrement dit les corps $K \simeq \mathbb{F}_q((T))$).

6.2. THÉORÈME. — Soit K un corps local d'égale caractéristique p . Il existe une famille de bijections :

$$\mathcal{A}_K^0(d) \rightarrow \mathcal{G}_K^0(d), \quad \pi \mapsto \sigma_\pi,$$

telles que les propriétés suivantes soient satisfaites :

(i) Pour chaque $\pi \in \mathcal{A}_K^0(d)$, $\pi' \in \mathcal{A}_K^0(d')$, on a :

$$L(\sigma_\pi \otimes \sigma_{\pi'}, s) = L(\pi \times \pi', s), \quad \varepsilon(\sigma_\pi \otimes \sigma_{\pi'}, s) = \varepsilon(\pi \times \pi', s).$$

(ii) Pour chaque $\pi \in \mathcal{A}_K^0(d)$: $\sigma_{\bar{\pi}} = \check{\sigma}_\pi$ (contragrédiente).

(iii) Le déterminant de σ_π correspond par la théorie locale du corps de classes (normalisation "géométrique") au caractère central de π .

(iv) Si $\pi \in \mathcal{A}_K^0(d)$ et si χ est un caractère d'ordre fini de K^* (d'où un caractère, noté aussi χ , de $\text{Gal}(\bar{K}/K)$), on a :

$$\sigma_{(\pi \otimes \chi)} = \sigma_\pi \otimes \chi \quad (\text{compatibilité à la torsion}).$$

Henniart prouve dans l'appendice [He 3] à [L-R-S] que les propriétés ci-dessus caractérisent la famille des bijections obtenues. Voir aussi [He 4] pour l'extension en des bijections $\mathcal{A}_K(d) \rightarrow \mathcal{G}_K(d)$.

6.3. A vrai dire, le théorème précédent est démontré dans [L-R-S] non pas pour des représentations à valeurs complexes, mais pour des représentations ℓ -adiques, i.e. à valeurs dans $\bar{\mathbb{Q}}_\ell$. Cela revient au même : tant du côté automorphe que galoisien, les objets considérés admettent des définitions purement algébriques, et on peut transporter la bijection au moyen d'un isomorphisme $\bar{\mathbb{Q}}_\ell \simeq \mathbb{C}$ de sorte que les facteurs L et ε se correspondent (cf. [He 4]). Changeant légèrement de notations, nous fixons donc un nombre premier $\ell \neq p$, et nous notons désormais $\mathcal{A}_K^0(d)$ l'ensemble des classes d'équivalence de $\bar{\mathbb{Q}}_\ell$ -représentations cuspidales irréductibles de $\text{GL}_d(K)$, à caractère central d'ordre fini, et $\mathcal{G}_K^0(d)$ l'ensemble des classes d'équivalence de $\bar{\mathbb{Q}}_\ell$ -représentations irréductibles de dimension d de $\text{Gal}(\bar{K}/K)$, à déterminant d'ordre fini. Nous allons maintenant décrire dans ce cadre les principales étapes de la preuve du théorème (6.2).

6.4. L'idée de la construction des applications $\pi \rightarrow \sigma_\pi$ est de procéder par voie globale en faisant usage du théorème (4.4.2). Choisissons pour cela

un corps global F de caractéristique p ainsi qu'une place x_0 de F tels que le complété F_{x_0} soit isomorphe à K (nous fixons un tel isomorphisme). Choisissons ensuite deux autres places ∞ et x_1 (distinctes et distinctes de x_0). Pour tout entier $d \geq 1$, notons D_d l'algèbre à division de centre F , de dimension d^2 , et d'invariants :

$$\text{inv}_x(D_d) = \begin{cases} 1/d & \text{si } x = x_0, \\ -1/d & \text{si } x = x_1, \\ 0 & \text{sinon.} \end{cases}$$

Le lemme suivant, démontré dans [L-R-S] au moyen d'une formule des traces de Selberg simplifiée, permet de voir toute $\pi \in \mathcal{A}_K^0(d)$ comme composante en x_0 d'une représentation automorphe de $\text{GL}_d(\mathbf{A}_F)$:

6.4.1. LEMME. — *Donnons-nous $\pi \in \mathcal{A}_K^0(d)$ et une place $x_2 \notin \{x_0, x_1, \infty\}$ de F . Il existe alors une représentation automorphe parabolique $\Pi = \otimes \Pi_v$ de $\text{GL}_d(\mathbf{A}_F)$ telle que $\Pi_\infty \simeq \text{St}_\infty$, $\Pi_{x_0} \simeq \pi$, et que Π_{x_1} et Π_{x_2} soient cuspidales.*

Puis un lemme de Henniart permet de transférer Π au groupe $G_d(\mathbf{A}_F) = (D_d \otimes \mathbf{A}_F)^*$; c'est un cas particulier de cette correspondance de Jacquet-Langlands généralisée, qu'on aimerait bien faire marcher en général. Noter que l'on a $G_d(F_y) \simeq \text{GL}_d(F_y)$ pour tout $y \notin \{x_0, x_1\}$.

6.4.2. LEMME. — *Soit Π comme dans le lemme précédent. Il existe alors une (unique) représentation automorphe $\tilde{\Pi}$ de $G_d(\mathbf{A}_F)$ telle que, pour tout $y \notin \{x_0, x_1\}$, on ait : $\tilde{\Pi}_y \simeq \Pi_y$. De plus la multiplicité $m(\tilde{\Pi})$ vaut 1.*

6.5. On applique maintenant le théorème (4.4.2) : il correspond à $\tilde{\Pi}$ une représentation $V_{\tilde{\Pi}_\infty}^{d-1}$ de dimension d du groupe $\text{Gal}(\bar{F}/F)$. Notons $\Sigma(\tilde{\Pi})_{x_0}$ la restriction de cette représentation à un groupe de décomposition en x_0 , tordue par le quasi-caractère non ramifié qui envoie Frob_{x_0} sur $q_{x_0}^{(1-d)/2}$ (cette opération rend le déterminant d'ordre fini) : on obtient ainsi une représentation du groupe $\text{Gal}(\bar{K}/K)$.

Notre construction dépend de choix (celui de x_2 , celui de Π) et il n'est pas clair a priori que $\Sigma(\tilde{\Pi})_{x_0}$ n'en dépend pas. Nous notons provisoirement $\Sigma(\pi)$ l'ensemble des représentations $\Sigma(\tilde{\Pi})_{x_0}$ (à équivalence près) obtenues pour tous les choix possibles. Un raisonnement qui utilise la théorie des fonctions L globales (lesquelles admettent des équations fonctionnelles, dont on sait, tant du côté automorphe que galoisien, exprimer les constantes comme produits de constantes locales) permet de déduire de (4.4.2) le :

6.5.1. LEMME.

(i) Soient $\pi \in \mathcal{A}_K^0(d)$ et $\sigma \in \Sigma(\pi)$. Alors $\det \sigma$ correspond au caractère central de π .

(ii) Si $\pi \in \mathcal{A}_K^0(d)$ et $\sigma \in \Sigma(\pi)$, alors $\check{\sigma} \in \Sigma(\check{\pi})$.

(iii) Pour $\pi \in \mathcal{A}_K^0(d)$, $\sigma \in \Sigma(\pi)$, et pour un caractère χ d'ordre fini de K^* , on a : $\sigma \otimes \chi \in \Sigma(\pi \otimes \chi)$.

(iv) Pour $\pi \in \mathcal{A}_K^0(d)$, $\pi' \in \mathcal{A}_K^0(d')$, $\sigma \in \Sigma(\pi)$, $\sigma' \in \Sigma(\pi')$, on a :

$$\begin{cases} L(\sigma \otimes \sigma', s) = L(\pi \times \pi', s), \\ \varepsilon(\sigma \otimes \sigma', s) = \varepsilon(\pi \times \pi', s). \end{cases}$$

6.6. Le lemme suivant permet alors de conclure que les ensembles $\Sigma(\pi)$ se réduisent à un élément σ_π :

LEMME. — Soit $\pi \rightarrow \Sigma(\pi)$ une application qui à tout $\pi \in \mathcal{A}_K^0(d)$ (d fixé) associe un ensemble non vide $\Sigma(\pi)$ de classes d'équivalence de représentations de dimension d de $\text{Gal}(\bar{K}/K)$, à déterminant d'ordre fini. Supposons que soient satisfaites les propriétés suivantes :

(i) $\forall \pi, \pi' \in \mathcal{A}_K^0(d)$, $\forall \sigma \in \Sigma(\pi)$, $\forall \sigma' \in \Sigma(\pi') : L(\sigma \otimes \sigma', s) = L(\pi \times \pi', s)$.

(ii) $\sigma \in \Sigma(\pi) \Rightarrow \check{\sigma} \in \Sigma(\check{\pi})$.

Alors pour tout $\pi \in \mathcal{A}_K^0(d)$, $\Sigma(\pi)$ est réduit à un élément $\sigma_\pi \in \mathcal{G}_K^0(d)$. De plus, l'application $\pi \rightarrow \sigma_\pi$ est injective.

6.7. On a donc obtenu à ce stade des *injections* $\pi \rightarrow \sigma_\pi$ qui satisfont à toutes les propriétés énoncées dans le théorème (6.2). Pour conclure, il nous reste à faire appel à la *conjecture de Langlands locale numérique*, prouvée par Henniart :

THÉORÈME [He 2]. — *Toute application injective $A_K^0(d) \rightarrow \mathcal{G}_K^0(d)$ qui préserve les conducteurs et qui est compatible à la torsion par les caractères non ramifiés est une bijection.*

RÉFÉRENCES

- [An] ANDERSON, G. — *t-motives*, Duke Math. Journal 53 (1986), 457–502.
- [Bo-Ja] BOREL, A. et JACQUET, H. — *Automorphic forms and automorphic Representations*, dans : Automorphic forms, Representations and L-functions, Proc. Symp. Pure Math., Vol 33, Tome 1, A.M.S., Providence (1979), 189–202.
- [Ca 1] CARTIER, P. — *Representations of p-adic groups : a survey*, dans Automorphic forms, Representations and L-functions, Proc. Symp. Pure Math., Vol 33, Tome 1, A.M.S., Providence (1979), 111–156.
- [Ca 2] CARTIER, P. — *La conjecture locale de Langlands pour GL(2) et la démonstration de Ph. Kutzko*, Sémin. Bourbaki, Fev 1980, exp. 550, Lecture Notes in Math. 842, Springer-Verlag (1981), 112–138.
- [Cl] CLOZEL, L. — *Motifs et formes automorphes : applications du principe de fonctorialité*, dans : Automorphic Forms, Shimura Varieties and L-functions, Proceedings of the Ann Arbor Conference, Academic Press (1990), T.1, 77–159.
- [De] DELIGNE, P. — *Les constantes des équations fonctionnelles des fonctions L*, Modular functions of one variable II, Lecture Notes in Math. 349, Springer-Verlag (1973), 501–597.

- [De-Hu] DELIGNE, P. et HUSEMÖLLER, D. — *Survey of Drinfeld modules*, dans : Current Trends in Arithmetical Algebraic Geometry, Contemporary Mathematics 67, A.M.S. (1987), 25–91.
- [De-Ka-Vi] DELIGNE, P., KAZHDAN, D., VIGNÉRAS, M.-F. — *Représentations des algèbres centrales simples p -adiques*, dans Représentations des groupes réductifs sur un corps local (Bernstein, Deligne, Kazhdan, Vignéras), Hermann (1984), 33–118.
- [Dr 1] DRINFELD, V.G. — *Elliptic modules*, Math. USSR Sbornik 23 (1974), 561–592.
- [Dr 2] DRINFELD, V.G. — *Elliptic modules II*, Math. USSR Sbornik 31 (1977), 159–170.
- [Dr 3] DRINFELD, V.G. — *Commutative subrings of certain non commutative rings*, Funct. Anal. and Appl. 10 (1976), 107–115.
- [Dr 4] DRINFELD, V.G. — *Varieties of modules of F -sheaves*, Funct. Anal. and Appl. 21 (1987), 107–122.
- [Dr 5] DRINFELD, V.G. — *The proof of Petersson's conjecture for $GL(2)$ over a global field of characteristic p* , Funct. Anal. and Appl. 22 (1988), 28–43.
- [Dr 6] DRINFELD, V.G. — *Cohomology of compactified manifolds of modules of F -sheaves of rank 2*, Journal of Soviet math. 46, n° 1 (1989), 1789–1821.
- [Ha-Na] HARDER, G. et NARASIMHAN, M.S. — *On the cohomology of moduli spaces of vector bundles on curves*, Math. Annalen 212 (1975), 215–248.
- [He 1] HENNIART, G. — *Le point sur la conjecture de Langlands pour $GL(n)$ sur un corps local*, Séminaire de théorie des nombres, Paris 1983–1984, Progress in Math, Birkhäuser (1985), 115–131.
- [He 2] HENNIART, G. — *La conjecture de Langlands locale numérique pour $GL(n)$* , Ann. Sci. E.N.S. (IV), 21 (1988), 145–203.

- [He 3] HENNIART, G. — *Caractérisation de la correspondance de Langlands locale par les facteurs ε de paires*, appendice à [L-R-S].
- [He 4] HENNIART, G. — *La correspondance de Langlands locale : caractérisation et propriétés fonctorielles* (en préparation).
- [Ja-Sh] JACQUET, H. et SHALIKA, J. — *On Euler products and the classification of automorphic representations, I*, Amer. J. of Math. 103 (1981) 499–558, II, Amer. J. of Math. 103 (1981), 777–815.
- [Ko] KOTTWITZ, R. — *On the λ -adic representations associated to some simple Shimura varieties*, preprint (1989).
- [Kr] KRICHEVER, I.M. — *Algebro-geometric constructions of the Zakharov-Shabat equations and their periodic solutions*, Dokl. Akad. Nauk SSSR 227, n° 2 (1976), 291–294.
- [La] LANGLANDS, R.P. — *Automorphic representations, Shimura varieties, and motives. Ein Märchen*. dans : Automorphic Forms, Representations, and L -functions, Proc. Symp. Pure Math., Vol 33, Tome 2, A.M.S. Providence (1979), 205–246.
- [Lau 1] LAUMON, G. — *Transformation de Fourier, constantes d'équations fonctionnelles et conjecture de Weil*, Publ. Math. I.H.E.S. 65 (1987), 131–210.
- [Lau 2] LAUMON, G. — *Cohomology with compact support of Drinfeld modular varieties*, part. I et II, prépublications de l'Université de Paris-Sud, 91–01 et 92–12.
- [L-R-S] LAUMON, G., RAPOPORT, M., STUHLER, U. — *\mathcal{D} -elliptic sheaves and the Langlands correspondence*, prépublication (1991).
- [Mu] MUMFORD, D. — *An algebro-geometric construction of commuting operators and of solutions to the Toda lattice equations, Korteweg – de Vries equation and related non-linear equations*, Internat. Symp. on Algebraic geometry (Kyoto 1977), Kinokuniya, Tokyo 1977, 115–153.

[Ro] ROGAWSKI, J. — *Representations of $GL(n)$ and division algebras over a p -adic field*, Duke Math. J. 50 (1983), 161–196.

Henri CARAYOL
Institut de Recherche Mathématique Avancée
Université Louis Pasteur et C.N.R.S.
7, rue René-Descartes
67084 Strasbourg Cedex

Astérisque

PATRICK GÉRARD

**Résultats récents sur les fluides parfaits
incompressibles bidimensionnels**

Astérisque, tome 206 (1992), Séminaire Bourbaki,
exp. n° 757, p. 411-444

http://www.numdam.org/item?id=SB_1991-1992__34__411_0

© Société mathématique de France, 1992, tous droits réservés.

L'accès aux archives de la collection « Astérisque » (<http://smf4.emath.fr/Publications/Asterisque/>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

**RÉSULTATS RÉCENTS SUR LES FLUIDES PARFAITS
INCOMPRESSIBLES BIDIMENSIONNELS**

[d' après J.-Y. CHEMIN et J.-M. DELORT]

par **Patrick GÉRARD**

L'équation d'Euler pour les fluides parfaits incompressibles est l'une des plus importantes équations d'évolution non linéaires de la physique mathématique, et l'étude de ses solutions a suscité de nombreuses recherches, notamment dans les trente dernières années, parallèlement au développement de méthodes d'analyse non linéaire. Récemment, Chemin et Delort ont démontré deux résultats concernant les solutions singulières de cette équation en dimension 2 d'espace. Dans une première partie, nous rappelons le cadre mathématique et essayons de situer ces deux résultats dans l'approche générale du problème de Cauchy pour l'équation d'Euler bidimensionnelle. Dans la seconde partie, nous donnons une démonstration de deux résultats classiques importants, dûs respectivement à Wolibner et à Yudovitch, puis nous présentons plus en détail les travaux de Chemin et Delort.

Je remercie J.-B. Bost, J.-Y. Chemin, J.-M. Delort et X. Saint-Raymond pour l'aide qu'ils m'ont apportée dans la préparation de cet exposé.

0.1. Préliminaires

Soit (M, g) une variété riemannienne compacte connexe orientée de dimension d .¹ Les notations suivantes de géométrie différentielle seront utilisées au cours de cet exposé :

¹ *La plupart des résultats discutés ici ont été obtenus dans l'espace \mathbf{R}^d plutôt que sur une variété compacte. Nous avons préféré ce cadre car il évite une discussion délicate des conditions à l'infini, et permet de mieux montrer les aspects invariants des techniques utilisées.*

Si X est un champ de vecteurs sur M , on désigne par L_X la dérivée de Lie suivant X , et par ∇_X la dérivée covariante suivant X . On désigne par τ la forme volume positive sur M . La divergence du champ X est la fonction $\operatorname{div}X$ définie par

$$L_X\tau = (\operatorname{div}X)\tau.$$

La norme de X sera notée $|X| = g(X, X)^{1/2}$. La dérivée covariante induit un opérateur différentiel d'ordre 1 ∇ des champs de vecteurs dans les sections du fibré $\operatorname{End}(TM)$, ou champs de tenseurs. On a alors $\operatorname{div}X = \operatorname{tr}(\nabla X)$.

On note $u \mapsto \tilde{u}$ l'isomorphisme de TM dans T^*M induit par la métrique g . Le gradient ∇f d'une fonction f est le champ de vecteurs correspondant à df par cet isomorphisme, et le laplacien de f est défini par $\Delta f = \operatorname{div}\nabla f$. La divergence d'un champ de tenseurs T est le champ de vecteurs $\operatorname{div}T$ correspondant à $-\operatorname{tr}\nabla T$ par cet isomorphisme, $\operatorname{tr}\nabla$ désignant l'opérateur différentiel transposé de ∇ pour la densité τ . Si u est un champ de vecteurs à divergence nulle, on a $\operatorname{div}(u \otimes \tilde{u}) = \nabla_u u$.

Enfin, on désigne par $*$ l'opérateur de Hodge sur les formes.

Lorsque M est de dimension 2, on appelle J l'opérateur sur les champs de vecteurs défini par

$$(JX)\tilde{} = *\tilde{X}.$$

On a $J^2 = -1$. Le rotationnel d'un champ de vecteurs X est la fonction $\operatorname{rot}X$ définie par $d\tilde{X} = (\operatorname{rot}X)\tau$. On a aussi $\operatorname{rot}u = -\operatorname{tr}(J\nabla X)$. Enfin, si u est un champ de vecteurs à divergence nulle, on a $\operatorname{rot}\nabla_u u = L_u \operatorname{rot}u$.

0.2. L'équation d'Euler.

En 1755, pour décrire le mouvement d'un fluide parfait incompressible occupant M , Euler [E] introduit le système d'équations suivant. Si l'on désigne par $u(t, x)$ la vitesse à l'instant $t \in \mathbf{R}$ de la particule de fluide occupant la position $x \in M$, le principe fondamental de la dynamique et la condition d'incompressibilité s'écrivent successivement

$$(E_1) \quad \partial_t u + \nabla_u u = -\nabla p,$$

$$(E_2) \quad \operatorname{div}u = 0,$$

où p désigne le champ scalaire de pression. Dans la suite, nous désignerons par (E) le système ci-dessus. Notons que le gradient de pression n'est pas une donnée

du problème, mais une inconnue supplémentaire, que l'on élimine en prenant la divergence de l'équation (E_1) ,

$$-\Delta p = \operatorname{div} \nabla_u u,$$

ce qui détermine ∇p en fonction du champ de vecteurs u .

On peut également voir (E) comme l'équation des géodésiques sur le groupe de Lie de dimension infinie des difféomorphismes de M préservant le volume, muni de la métrique définie par l'énergie cinétique

$$E_c = \int_M |u(t, x)|^2 \tau(x).$$

Dans ce cadre, p s'interprète comme un multiplicateur de Lagrange. Nous ne développerons pas ici ce point de vue et renvoyons à Arnold [Ar], Ebin–Marsden [EB], Shnirelman [Sh] et Brenier [Br] pour une étude détaillée. Retenons seulement de cette interprétation que, si u est solution de (E) , l'énergie cinétique est constante au cours du temps—ce que l'on peut d'ailleurs aisément retrouver en prenant le produit scalaire L^2 des deux membres de (E_1) par u et en utilisant (E_2) .

Un problème fondamental en mécanique des fluides est bien sûr le problème de Cauchy associé à (E) , c'est-à-dire la détermination de u vérifiant (E) et la condition $u(0, x) = u^0(x)$, u^0 étant un champ de vecteurs donné sur M à divergence nulle. L'unicité et l'existence sur un petit intervalle de temps semblent avoir été démontrées pour la première fois par Lichtenstein [Li] en 1925. L'existence globale d'une telle solution est toujours un problème ouvert en dimension 3, mais a été démontrée en dimension 2 par Wolibner [W] en 1933 (*cf.* aussi Kato [Ka]). Néanmoins le résultat de Wolibner nous renseigne peu sur les propriétés qualitatives de la solution qu'il met en évidence ; les estimations obtenues laisseraient même présager un comportement très chaotique des lignes de courant lorsque t devient grand (*cf.* §1).

Pour obtenir des renseignements sur les propriétés qualitatives des solutions d'un tel problème, une approche couramment employée est d'essayer de résoudre le même problème avec des données singulières et de décrire comment évoluent les singularités. Ces solutions singulières (par exemple, un champ de vitesses discontinu) sont en général des modèles "limites" de phénomènes physiques mettant en présence des échelles de grandeur très différentes (par exemple, l'évolution d'une

interface). Les deux résultats qui font l'objet de cet exposé se rattachent directement à cette problématique pour un fluide bidimensionnel. Avant de les présenter, il nous faut introduire une quantité fondamentale en mécanique des fluides : le tourbillon.

0.3. Le tourbillon.

Dans toute la suite de l'exposé, la variété M est supposée de dimension 2. On définit le tourbillon ω associé au champ de vitesses u par la formule

$$\omega = \operatorname{rot} u.$$

Notons que $\int_M \omega(x) \tau(x) = 0$. Outre son interprétation physique évidente, l'intérêt de cette notion est qu'elle permet une formulation équivalente de (E), de la façon suivante. En appliquant l'opérateur rot à (E₁), on obtient l'équation de transport

$$(T) \quad \partial_t \omega + L_u \omega = 0,$$

traduisant que ω est une fonction constante le long des lignes de courant. Par ailleurs, on peut—presque—retrouver u à partir de ω . Pour cela, on utilise le fait que le laplacien est un automorphisme de l'espace vectoriel des fonctions C^∞ de moyenne nulle (cf. par exemple [R]); désignons par Δ^{-1} son inverse. Soit d'autre part \mathcal{H} l'espace vectoriel de dimension finie des champs de vecteurs harmoniques sur M , et soit H le projecteur orthogonal de l'espace des champs de vecteurs sur M à valeurs dans \mathcal{H} , pour le produit scalaire L^2 associé à τ . En tenant compte de (E₂), la décomposition de Hodge–de Rham [R] du champ u donne la “relation de Biot et Savart”

$$(B) \quad u = B\omega + h,$$

où $B = J\nabla\Delta^{-1}$ est la restriction aux fonctions de moyenne nulle d'un opérateur pseudodifférentiel d'ordre -1 , et $h = H(u)$. Enfin, l'évolution de h dans l'espace \mathcal{H} est obtenue en appliquant H à l'équation (E₁),

$$(H) \quad \partial_t h + Q(u, u) = 0,$$

où l'on a posé

$$Q(v, w) = - \sum_j \left(\int_M g(\nabla_v X_j, w) \tau \right) X_j,$$

(X_j) désignant une base orthonormale de \mathcal{H} . Puisque h varie dans un espace de dimension finie et que sa norme L^2 est majorée par celle de u donc bornée au cours du temps, on peut prévoir que l'évolution de h n'aura pas d'influence sur les propriétés qualitatives de la solution u . On notera que, dans deux cas simples, cette évolution est triviale : si M est une sphère, $\mathcal{H} = \{0\}$, tandis que si M est un tore plat, h est constant au cours du temps.

Il est aisé de vérifier que le système (T, B, H) est équivalent à (E) .

0.4. Poches de tourbillon.

Venons en aux solutions singulières de (E) . Le résultat de base est ici celui de Yudovitch [Y] en 1963, et s'applique notamment aux fluides à tourbillon discontinu. Plus précisément, Yudovitch montre que, si la donnée initiale a un tourbillon essentiellement borné, il existe une solution et une seule au problème de Cauchy dont le tourbillon soit essentiellement borné pour tout temps—les équations devant alors être comprises en un sens faible. Pour apprécier ce résultat, remarquons que la relation de Biot et Savart (B) ne permet pas de conclure que le champ des vitesses est lipschitzien; en effet, un opérateur pseudodifférentiel d'ordre -1 n'envoie pas L^∞ dans les fonctions lipschitziennes, mais dans un espace de fonctions légèrement moins régulières. Yudovitch montre néanmoins qu'un champ ayant cette régularité admet un flot, et que le tourbillon à l'instant t se déduit du tourbillon à l'instant initial par l'action de ce flot, conformément à l'équation (T) . Hélas la régularité de ce flot se dégrade exponentiellement au cours du temps, ce qui rend délicat un contrôle de l'évolution des singularités (*cf.* §2).

Dans son excellent article de survol consacré à la fonction de tourbillon en mécanique des fluides [M1], Majda pose en 1985 le problème suivant, dit des "poches de tourbillon". Supposons que le tourbillon à l'instant initial ne prenne que deux valeurs, l'ensemble de discontinuité étant une courbe de Jordan γ^0 de classe C^∞ . D'après le résultat de Yudovitch, le tourbillon à l'instant t ne prendra que ces deux valeurs, l'ensemble de discontinuité γ^t se déduisant de γ^0 par le flot peu régulier ci-dessus. Que peut-on dire de la régularité de γ^t ?

Dans le même article, après avoir décrit l'équation intégrodifférentielle que doit vérifier une paramétrisation régulière de γ^t , Majda annonce que, pour un petit intervalle de temps, le contour restera régulier; de plus, en se basant sur des si-

mulations numériques [Z], il conjecture l'apparition de cusps et la non-rectifiabilité de γ^t au bout d'un temps fini.

Dès lors, plusieurs travaux se succèdent dans le but de confirmer cette conjecture ; outre des travaux numériques [D], [DI], signalons l'approche proposée par Constantin et Titi [CT], qui étudient la perturbation d'une poche de tourbillon stationnaire (par exemple, un disque dans le plan). Dans ce contexte, Alinhac [A1] démontre ensuite que l'approximation quadratique de l'équation des poches de tourbillon développe effectivement des singularités.

Toutes ces contributions rendent d'autant plus remarquable le résultat suivant, démontré par Chemin en 1991 :

THÉORÈME A. (Chemin [C4]). *Le contour d'une poche de tourbillon reste C^∞ pour tout temps.*

La démonstration de Chemin repose sur deux idées principales.

La première consiste à remarquer que, si l'on désigne par X^0 un champ de vecteurs C^∞ sur M tangent non nul à la courbe γ^0 , et si X^t est le champ — singulier — transporté de X^0 par le flot du fluide à l'instant t , la régularité de γ^t découlera du fait que, bien que X^t ne soit pas C^∞ , X^t et toutes les dérivées covariantes $\nabla_{X^t}^k X^t$ appartiendront à un même espace de Hölder C^α . Ce type de régularité, dite "conormale" ou "deux-microlocale", est au centre des travaux sur la propagation des singularités dans les équations hyperboliques non linéaires développés par Bony et son école dans les dix dernières années. Pour une introduction à ces résultats, nous renvoyons le lecteur à l'article de survol de Bony [B2] et à l'exposé de Lebeau [L] dans ce séminaire. En utilisant les méthodes utilisées dans ces travaux, et auxquelles il a lui-même contribué [C5], Chemin démontre en 1989 [C1,3] le résultat de régularité locale en temps du contour de la poche de tourbillon, tel qu'il avait été annoncé par Majda. Mentionnons aussi que, par les mêmes méthodes, Chemin montre le remarquable résultat suivant, qui, lui, concerne l'équation d'Euler en toute dimension: "*Toute solution C^r ($r > 1$) de l'équation d'Euler sur $[0, T] \times M$ possède des lignes de courant C^∞ .*" [C2]

La deuxième idée est le pas décisif vers la régularité globale en temps et fait l'objet de l'article [C4]. Elle réside dans une estimation de la norme Lipschitz du champ des vitesses en fonction de la norme de ω dans L^∞ et de quantités "conormales" mesurant la régularité de la géométrie, le point essentiel étant que

ces quantités apparaissent comme argument d'un logarithme (*cf.* §3, proposition 8).

Remarquons que ces estimations sont en fait très générales, et que le résultat de Chemin s'étend à des situations géométriques beaucoup plus compliquées que la simple poche de tourbillon (*cf.* §3).

Signalons que les travaux de Chemin mentionnés ci-dessus ont donné lieu à diverses améliorations ou reformulations : ainsi, la régularité locale des poches de tourbillon a été prouvée indépendamment par Serfati [S], qui a travaillé directement sur l'équation intégrodifférentielle du contour, et a prouvé de plus une régularité analytique en temps de ce contour. Une approche directe de la régularité locale a également été développée récemment par Bertozzi [Be] dans sa thèse. Enfin, il y a trois mois, Constantin et Bertozzi [BC] ont donné une démonstration un peu plus courte du théorème A, également basée sur une estimée logarithmique, mais dépendant plus fortement de la structure géométrique particulière d'une poche de tourbillon.

Pour conclure l'introduction à ce problème, indiquons que les estimations sur la géométrie du contour obtenues par Chemin sont doublement exponentielles par rapport au temps ; il serait intéressant de savoir si ces estimations sont optimales ; si tel est le cas, cela pourrait expliquer les prédictions d'apparition de singularités découlant des simulations numériques. Un autre problème ouvert est l'évolution de contours présentant au départ des singularités (coins, cusps, etc...).

0.5. Nappes de tourbillon.

Passons maintenant à des solutions plus singulières encore. Supposons donné un champ de vitesses initial u^0 discontinu à travers une courbe régulière de M ; le tourbillon est alors une mesure de Radon dont la partie singulière est une densité sur cette courbe. Peut-on encore résoudre le problème de Cauchy pour (E) dans ce cas (on notera que le théorème de Yudovitch ne s'applique plus)?

Plus encore que le précédent, ce problème modélise un grand nombre de situations physiques, par exemple l'écoulement près d'une aile d'avion, et a donné lieu à une abondante littérature physique et numérique ; nous renvoyons à l'article de Saffman et Baker [SB] pour une introduction détaillée. Contentons-nous ici de dire qu'il s'agit d'un problème très instable, comme l'avaient déjà remarqué

Helmholtz [H], Kelvin [Ke] et Taylor [T]. Contrairement au cas des poches de tourbillon, l'équation intégrodifférentielle censée décrire l'évolution de l'interface est mal posée en temps petit, en ce sens que le problème *linéarisé* associé n'a pas de solutions dans C^∞ (cf. Birkhoff [Bi]). On dispose néanmoins de résultats d'existence locale pour ce problème lorsque l'interface et la restriction de u^0 à son complémentaire sont analytiques, les méthodes étant alors basées sur des formes abstraites du théorème de Cauchy–Kowaleska (cf. [BF], [DR], [CO]). Mais les simulations numériques (cf. notamment Krasny [Kr]) laissent présager un comportement beaucoup plus complexe au bout d'un temps fini.

Plus généralement, on peut se poser la question suivante : étant donné un champ de vitesses initial dans L^2 dont le tourbillon est une mesure bornée, existe-t-il une solution au problème de Cauchy? Les méthodes de recherche de solutions faibles pour des équations aux dérivées partielles non linéaires suggèrent la stratégie suivante : en régularisant la donnée de Cauchy et en utilisant le théorème de Wolibner, on obtient une suite (u_n) de solutions régulières dont les énergies cinétiques sont bornées, ainsi que les normes L^1 des tourbillons correspondants. Une sous-suite $(u_{n'})$ convergera donc faiblement vers un champ de vitesses u de carré intégrable dont le tourbillon est une mesure. Le problème est de savoir si u est solution de l'équation d'Euler

$$(E') \quad \partial_t u + \operatorname{div}(u \otimes \tilde{u}) = -\nabla p, \quad \operatorname{div} u = 0,$$

(E_1) étant écrite ici sous sa forme conservative pour garder un sens lorsque u est seulement de carré intégrable. La difficulté est que l'on ignore si la convergence de $u_{n'}$ vers u est forte, et que l'on ne peut donc pas passer à la limite dans le terme $u_{n'} \otimes \tilde{u}_{n'}$. Dans leurs trois articles [DM1,2,3], DiPerna et Majda sont les premiers à aborder ce problème. Ils montrent en particulier que la suite converge fortement en dehors d'un ensemble de dimension de Hausdorff plus petite ou égale à 1 ; un peu plus tard, Greengard et Thomann [GT] montrent que, si cette dimension est < 1 , la convergence est forte. DiPerna et Majda suggèrent également la possibilité d'un phénomène de "concentration évanescence" selon lequel, bien que la convergence ne soit pas forte, u serait solution de l'équation d'Euler ; Alinhac [A2] montre ensuite que ce phénomène a lieu si les valeurs d'adhérence de la suite $(|u_{n'} - u|^2)$ pour la topologie vague sont concentrées sur des ensembles "suffisamment petits".

Ce n'est qu'en 1990 qu'une réponse positive est donnée à ce problème, dans un cas particulier important :

THÉORÈME B. (Delort [D1,2]). Soit (u_n) une suite de solutions régulières de (E) dont la suite des données initiales (u_n^0) converge en moyenne quadratique vers (u^0) ; on suppose que $\omega^0 = \text{rot} u^0$ est une mesure de Radon dont la partie singulière est positive (resp. négative), et que la suite $(\text{rot} u_n^0)$ converge vers ω_0 pour la topologie vague. Alors toute valeur d'adhérence de (u_n) est solution de (E') . En particulier, le problème de Cauchy pour (E') avec la donnée (u^0) admet une solution.

Mentionnons que le théorème de Delort contient en fait aussi un résultat de stabilité par passage à la limite faible pour les solutions de (E') qu'il met en évidence, à savoir les solutions d'énergie finie dont le tourbillon est à tout instant une mesure de Radon à partie singulière positive (resp. négative). En revanche, l'unicité pour de telles solutions n'est pas connue ; de même, la conservation de l'énergie cinétique au cours du temps est un problème ouvert — on dispose simplement d'une inégalité, et l'égalité équivaudrait à la convergence forte des suites de solutions introduites plus haut.

La démonstration du théorème B repose sur deux arguments. Tout d'abord, en éliminant la fonction de pression, Delort établit une forme équivalente de (E) , qui lui permet de constater qu'il suffit de passer à la limite dans *certaines* expressions quadratiques en u_n . Ce dernier fait avait d'ailleurs été préalablement constaté par les auteurs cités plus haut, en lien avec la "concentration évanescence." Il se trouve alors que l'hypothèse de positivité sur la partie singulière de ω^0 induit un remarquable phénomène de compensation dans les expressions quadratiques pseudodifférentielles en $\omega_n = \text{rot} u_n$, qui permet de passer effectivement à la limite (cf. §4). Indiquons enfin que ce phénomène de compensation a été interprété plus récemment par Evans et Müller [EM] en termes d'espace de Hardy, dans l'esprit des résultats de Coifman–Lions–Meyer–Semmes [CL].

0.6. Rappels sur les espaces fonctionnels utilisés.

Si E est un fibré vectoriel sur M et $k \in \mathbb{N} \cup \{\infty\}$, on désigne par $C^k(M, E)$ l'espace des sections de classe C^k de E . Pour chaque fibré E utilisé dans la suite, et pour tout $k \neq \infty$, on fait le choix d'une norme sur $C^k(M, E)$ qui définit la

topologie de la convergence C^k ; on désignera cette norme par $[\]_k$.

Si $\alpha \in]0, 1[$, on désigne par $C^\alpha(\mathbf{R}^n)$ l'espace des fonctions f bornées sur \mathbf{R}^n telles qu'il existe $C > 0$ satisfaisant à

$$|f(x) - f(y)| \leq C|x - y|^\alpha.$$

Si α est un réel positif non entier, $C^\alpha(\mathbf{R}^n)$ désigne l'espace des fonctions bornées de classe $C^{[\alpha]}$ dont les dérivées d'ordre $[\alpha]$ appartiennent à $C^{\alpha-[\alpha]}(\mathbf{R}^n)$. On munit chacun de ces espaces de sa structure naturelle d'espace de Banach.

Si $\alpha - m$ est positif non entier, un opérateur pseudodifférentiel d'ordre m est borné de $C^\alpha(\mathbf{R}^n)$ dans $C^{\alpha-m}(\mathbf{R}^n)$. On définit alors les espaces de Hölder pour les autres valeurs de α de telle sorte que cette propriété reste vraie pour tous réels α et m . On note encore $C^\alpha(\mathbf{R}^n)$ ces espaces, sauf dans le cas où $\alpha = k$ est un entier naturel, où l'on préfère noter $C_*^k(\mathbf{R}^n)$, pour distinguer cet espace de celui des fonctions de classe C^k bornées à dérivées bornées, qu'il contient strictement. Là encore, ces espaces sont munis d'une structure naturelle d'espace de Banach, et on notera $\| \cdot \|_\alpha$ une norme définissant cette structure. Par exemple, $C_*^1(\mathbf{R}^n)$ est la "classe de Zygmund", c'est-à-dire l'espace des fonctions bornées satisfaisant à

$$|f(x+h) + f(x-h) - 2f(x)| \leq C|h|.$$

On décrit aisément les espaces de Hölder ainsi définis en utilisant la décomposition dyadique dans l'espace de Fourier comme suit (cf. par exemple Coifman-Meyer [CM]). Soit $\varphi = \varphi(\xi)$ une fonction C^∞ sur \mathbf{R}^n , à support compact dans $\mathbf{R}^n \setminus \{0\}$, telle que

$$\chi(\xi) = 1 - \sum_{p=0}^{\infty} \varphi(2^{-p}\xi)$$

soit supportée dans un compact de \mathbf{R}^n . Si f est une distribution tempérée sur \mathbf{R}^n , on pose, pour $p \geq 0$,

$$\Delta_p f(x) = \varphi(2^{-p}D)f(x), \quad S_p f(x) = \chi(2^{-p}D)f(x)$$

et $\Delta_{-1}f(x) = \chi(D)f(x)$, de sorte que

$$S_p f = \sum_{-1 \leq q \leq p-1} \Delta_q f, \quad f = \sum_{-1 \leq q} \Delta_q f.$$

On notera que $S_p f$ est une fonction de classe C^∞ , et que, en désignant par $[\]_k$ la norme C^k sur \mathbf{R}^n , on a

$$[S_p f]_k \leq C_k 2^{kp} [S_p f]_0,$$

où C_k ne dépend que de k . La distribution tempérée f appartient alors à $C^\alpha(\mathbf{R}^n)$ pour $\alpha \in \mathbf{R} \setminus \mathbf{N}$ (resp. $C_*^\alpha(\mathbf{R}^n)$ pour $\alpha \in \mathbf{N}$) si et seulement si il existe une constante C telle que, pour tout $p \geq -1$, on ait

$$[\Delta_p f]_0 \leq C 2^{-p\alpha},$$

la meilleure constante C définissant alors une norme sur l'espace de Hölder d'ordre α qui induit sa structure d'espace de Banach.

Les espaces de Hölder ainsi définis sont stables par multiplication par une fonction C^∞ à support compact, et localement invariants par difféomorphisme. Si E est un fibré vectoriel sur la variété compacte M , on peut donc définir l'espace des sections hölderiennes d'ordre α de E au-dessus de M , que l'on note $C^\alpha(M, E)$ si α n'est pas un entier naturel et $C_*^\alpha(M, E)$ sinon; on note en outre $\| \cdot \|_\alpha$ une norme définissant la topologie naturelle de cet espace.

Si Ω est un ouvert de M et si f est une fonction hölderienne d'ordre α dans Ω , on note

$$\|f\|_{\alpha, \Omega} = \inf\{\|g\|_\alpha, g \in C^\alpha(M), g|_\Omega = f\}.$$

Enfin, on désigne par $L^2(M, E)$ l'espace des sections L^2 du fibré E au-dessus de M .

1. SOLUTIONS RÉGULIÈRES.

Théorème 1. (Wolibner [W], 1933 ; Kato [Ka], 1967). *Soit u^0 un champ de vecteurs de classe C^∞ sur M , à divergence nulle. Il existe une solution unique $u \in C^\infty(\mathbf{R} \times M, TM)$ à l'équation (E), telle que $u(0, x) = u^0(x)$.*

Démonstration. Elle s'effectue en trois temps: on montre d'abord l'existence et l'unicité d'une solution sur un petit intervalle de temps; puis on montre que, si la solution ne peut être prolongée au-delà d'un temps $T > 0$, on a nécessairement

$$\limsup_{t \rightarrow T^-} [u(t)]_1 = +\infty.$$

On conclut alors en estimant $[u(t)]_1$ sur tout intervalle de temps borné. Si les deux premiers arguments sont généralisables à tous les systèmes hyperboliques non linéaires d'ordre 1, le troisième argument utilise la forme particulière de l'équation (E) et le fait que M est de dimension 2 ; rappelons que le problème analogue est ouvert en dimension 3.

1.1. Existence et unicité à temps petit.

On se contente de résoudre le problème pour $t \geq 0$, le cas $t \leq 0$ étant analogue. Si u et v sont des solutions régulières de (E) sur l'intervalle de temps $[0, T]$ avec la même donnée initiale et des pressions respectives p et q , on a

$$\partial_t(u - v) + \nabla_u(u - v) = -\nabla(p - q) + \nabla_{v-u}v.$$

En égalant le produit scalaire dans $L^2(M, TM)$ des deux membres de cette équation et en utilisant que $\operatorname{div} u = \operatorname{div} v = 0$, il vient

$$\frac{1}{2} \frac{d}{dt} \|u(t) - v(t)\|_{L^2}^2 = (\nabla_{v(t)-u(t)}v(t), u(t) - v(t)).$$

Puisque $[v(t)]_1 \leq C$, on en déduit

$$\frac{d}{dt} \|u(t) - v(t)\|_{L^2}^2 \leq C \|u(t) - v(t)\|_{L^2}^2,$$

et le lemme de Gronwall permet de conclure que $u = v$.

Passons à l'existence locale. On utilise la formulation de (E) à l'aide du système $(T), (H), (B)$; on définit par récurrence trois suites $(u_n), (\omega_n), (h_n)$ de la façon suivante: $u_0(t, x) = u^0(x), \omega_0 = \operatorname{rot} u_0, h_0 = H(u_0)$, et, pour tout $n \geq 0$,

$$\begin{aligned} \partial_t \omega_{n+1} + L_{u_n} \omega_{n+1} &= 0, & \omega_{n+1}(0) &= \operatorname{rot} u^0, \\ \partial_t h_{n+1} + Q(u_n, B\omega_{n+1} + h_{n+1}) &= 0, & h^{n+1}(0) &= H(u^0), \\ u_{n+1} &= B\omega_{n+1} + h_{n+1}. \end{aligned}$$

Notons que la première équation est une équation de transport linéaire, et que la seconde équation est un système différentiel linéaire dans l'espace de dimension finie \mathcal{H} des champs harmoniques ; de plus, $Q(h_n, h_{n+1}) = 0$ car h_n et h_{n+1} sont harmoniques. De l'équation de transport, on déduit

$$[\omega_n(t)]_0 = [\operatorname{rot} u^0]_0$$

pour tous t et n , et, en utilisant cette estimation et la seconde équation, on obtient que $h_n(t)$ est bornée dans l'espace de dimension finie \mathcal{H} . On montre alors facilement qu'il suffit que $[\omega_n(t)]_1$ soit uniformément bornée sur $[0, T]$ pour que les suites (ω_n) (resp. (h_n)) soient de Cauchy dans l'espace $C^0([0, T] \times M)$ (resp. $C^0([0, T], \mathcal{H})$). L'opérateur pseudodifférentiel B étant d'ordre -1 , il est borné de C^ϵ dans C^1 pour tout $\epsilon > 0$, a fortiori de C^1 dans C^1 . Comme par ailleurs h_n est bornée, il existe une constante C_1 telle que, pour tous n et t , on ait

$$[u_n(t)]_1 \leq C_1([\omega_n(t)]_1 + 1).$$

On estime alors $[\omega_n(t)]_1$ en intégrant par la méthode des caractéristiques l'équation de transport définissant ω_n . Il vient

$$\frac{d}{dt}[\omega_n(t)]_1 \leq C_2[\omega_n(t)]_1[u_{n-1}(t)]_1.$$

De cette dernière équation, on déduit que, si $[u_{n-1}(t)]_1 \leq M$ pour $t \in [0, T]$, on a

$$[\omega_n(t)]_1 \leq [\text{rot}u^0]_1 e^{C_2MT},$$

d'où

$$[u_n(t)]_1 \leq C_1(1 + [\text{rot}u^0]_1 e^{C_2MT}),$$

et il suffit de choisir T et M tels que

$$C_1(1 + [\text{rot}u^0]_1 e^{C_2MT}) \leq M,$$

pour conclure par récurrence que $[u_n(t)]_1$ est bornée sur $[0, T]$, donc aussi $[\omega_n(t)]_1$. On estime enfin les dérivées d'ordre k de $\omega_n(t)$ à l'aide de l'équation de transport,

$$\frac{d}{dt}[\omega_n(t)]_k \leq C_k([u_{n-1}(t)]_1[\omega_n(t)]_k + [u_{n-1}(t)]_k[\omega_n(t)]_1).$$

En tenant compte de l'estimée

$$[u_{n-1}(t)]_k \leq C_k(\|\omega_{n-1}(t)\|_{k-1/2} + 1) \leq C_k([\omega_{n-1}(t)]_k^{1/2}[\omega_{n-1}(t)]_{k-1}^{1/2} + 1),$$

on conclut aisément par récurrence sur n et k que $[\omega_n(t)]_k \leq M_k$ pour tout $t \in [0, T]$, ce qui assure que (u_n) converge dans C^∞ vers une solution de (E) . Notons

en outre que le temps d'existence T ainsi obtenu ne dépend que d'un majorant de $[u_0]_2$.

1.2. Existence globale.

Supposons que u soit une solution de (E) sur l'intervalle $[0, T[$. A partir des estimées fournies par l'équation de transport (T) ,

$$(1) \quad \frac{d}{dt}[\omega(t)]_k \leq C_k ([u(t)]_1[\omega(t)]_k + [u(t)]_k[\omega(t)]_1),$$

et par la relation de Biot et Savart,

$$[u(t)]_k \leq C_k ([\omega(t)]_k + |h(t)|),$$

on constate que, si $[u(t)]_1$ est bornée sur $[0, T[$, alors il en est de même de $[u(t)]_k$, quel que soit k , et le résultat d'existence locale précédent permet de prolonger la solution u au-delà du temps T .

Il nous reste donc à montrer que $[u(t)]_1$ est bornée sur $[0, T[$. Pour cela, le point crucial est que, d'après (T) ,

$$[\omega(t)]_0 = [\operatorname{rot} u^0]_0$$

pour tout t . Hélas l'opérateur B n'agit pas de C^0 dans C^1 , mais seulement de C^0 dans C_*^1 . Pour remédier à cette imperfection, on utilise l'estimation non linéaire suivante:

Lemme 2. *Soit B un opérateur pseudodifférentiel d'ordre -1 sur une variété compacte M . Il existe $C > 0$ tel que, pour toute fonction régulière f , on ait*

$$[Bf]_1 \leq C[f]_0 \log\left(1 + \frac{[f]_1}{[f]_0}\right).$$

Ce type d'estimation étant également un des ingrédients principaux du résultat de Chemin, nous y reviendrons au paragraphe 3. Pour l'instant, voyons comment elle permet d'achever la démonstration du théorème 1. De l'équation (E) , on déduit la conservation de l'énergie,

$$\|u(t)\|_{L^2} = \|u^0\|_{L^2}$$

qui entraîne en particulier que $h(t)$ est bornée dans \mathcal{H} . De la relation de Biot et Savart, on déduit donc

$$[u(t)]_1 \leq C \log(2 + [\omega(t)]_1),$$

tandis que la relation (1) pour $k = 1$ et le lemme de Gronwall donnent

$$[\omega(t)]_1 \leq [\omega(0)]_1 (1 + e^{\int_0^t [u(s)]_1 ds}).$$

En comparant ces deux estimées et en appliquant à nouveau le lemme de Gronwall, on conclut qu'il existe une constante A telle que

$$(2) \quad [u(t)]_1 \leq Ae^{At},$$

ce qui achève la démonstration du théorème 1.

Remarques.

a) Il est frappant de comparer les croissances respectives par rapport à t des normes—apparemment très proches— C_*^1 et C^1 du champ $u(t)$; alors que la première est bornée, les seules estimées dont on dispose sur la seconde sont exponentielles (cf. l'inégalité (2)). On ignore si cette croissance exponentielle a lieu effectivement ; notons qu'elle induirait une croissance doublement exponentielle pour la première dérivée du tourbillon ω . On ne connaît d'ailleurs à l'heure actuelle aucun exemple explicite de solution régulière non stationnaire de (E). Malgré le théorème 1, on peut donc dire que l'étude des solutions régulières de (E) reste un problème largement ouvert.

b) Lorsque M est de dimension 3, le tourbillon ω n'est plus une fonction mais une deux-forme différentielle sur M . L'équation de transport (B) ne permet donc plus d'estimer $[\omega(t)]_0$ sans faire apparaître la quantité $[u(t)]_1$, ce qui rend la démonstration ci-dessus inapplicable.

2. SOLUTIONS QUASILIPSCHITZIENNES.

2.1. Une fonction f sur \mathbf{R}^n est dite *quasilipschitzienne* s'il existe une constante $C > 0$ telle que, pour tous x, y ,

$$|f(x) - f(y)| \leq C|x - y| \left(1 + \log_+ \frac{1}{|x - y|} \right),$$

où $\log_+ t$ désigne le plus grand des deux nombres 0 et $\log t$. Une telle fonction appartient donc à tous les espaces de Hölder d'ordre $\alpha < 1$. Un exemple non trivial de fonction quasilipschitzienne est donné par le lemme suivant :

Lemme 3. *Toute fonction appartenant à C_*^1 est quasilipschitzienne.*

Démonstration. On écrit, pour tout entier N ,

$$f(x) - f(y) = \sum_{p \leq N} (\Delta_p f(x) - \Delta_p f(y)) + \sum_{p > N} (\Delta_p f(x) - \Delta_p f(y)).$$

On utilise alors l'estimation $[\Delta_p f]_k \leq C2^{(k-1)p}$ avec $k = 1$ dans la première somme et $k = 0$ dans la seconde somme. Il vient

$$|f(x) - f(y)| \leq C(N|x - y| + 2^{-N}),$$

et le choix de la partie entière de $\log|x - y|/\log 2$ comme valeur de N donne l'estimation cherchée.

L'ensemble des fonctions localement quasilipschitziennes sur un ouvert de \mathbf{R}^n étant stable par multiplication par les fonctions à support compact et par changement de variables, on peut parler de champ de vecteurs localement quasilipschitzien sur une variété différentiable. On a alors le théorème de Cauchy-(quasi)Lipschitz suivant:

Proposition 4. *Soit v un champ de vecteurs quasilipschitzien sur une variété compacte M . Pour tout $x_0 \in M$, il existe une unique courbe $\gamma \in C^1(\mathbf{R}, M)$ solution de l'équation différentielle*

$$\frac{d}{dt}\gamma = v(\gamma), \quad \gamma(0) = x_0.$$

De plus, il existe $k > 0$ tel que le flot (ϕ^t) ainsi défini vérifie

$$\forall t \in \mathbf{R}, \phi^t \in C^{e^{-k|t|}}(M, M).$$

Démonstration. Il suffit de raisonner localement, donc dans \mathbf{R}^n . Si, pour $j = 1, 2$, γ_j est une courbe intégrale de v sur un intervalle $[0, T]$ avec $\gamma_j(0) = x_j$, on a

$$\frac{d}{dt}|\gamma_1 - \gamma_2|^2 = 2(\gamma_1 - \gamma_2, v(\gamma_1) - v(\gamma_2)).$$

Posant $f(t) = |\gamma_1(t) - \gamma_2(t)|^2$, le fait que v soit quasilipschitzien entraîne alors

$$\frac{d}{dt}f \leq Cf \left(1 + \log_+ \frac{1}{f}\right).$$

On en déduit, pour tout $t \in [0, T]$, et si $f(0) \leq 1$,

$$f(t) \leq f(0)e^{-ct},$$

ce qui prouve l'unicité et la régularité du flot. L'existence découle classiquement d'une approximation de v par des champs réguliers et du théorème d'Ascoli.

On notera que le résultat ci-dessus est inchangé pour un champ $v = v(t, x)$ dépendant continûment du paramètre t .

2.2. On se propose maintenant d'étudier des solutions singulières de l'équation (E) dont la donnée de Cauchy vérifie $\text{rot}u^0 \in L^\infty(M)$. Compte tenu de l'équation de transport (T), on s'attend à ce que, pour tout temps t , $\omega(t)$ soit dans $L^\infty(M)$, donc que le champ des vitesses soit quasilipschitzien. Les équations (E) et (T) doivent alors être réinterprétées au sens des distributions, de la façon suivante,

$$(E') \quad \partial_t u + \text{div}(u \otimes \tilde{u}) = -\nabla p, \quad \text{div}u = 0,$$

$$(T') \quad \partial_t \omega + \text{div}(u\omega) = 0,$$

la condition $\text{div}u = 0$ assurant l'équivalence des deux formulations lorsque u est une solution régulière.

Théorème 5. (Yudovich [Y], 1963). Soit u^0 un champ de vecteurs tel que $\text{rot}u^0 \in L^\infty$, $\text{div}u^0 = 0$. Alors il existe une unique solution u de l'équation (E') vérifiant $u \in C^0(\mathbf{R} \times M, TM)$, $\text{rot}u \in L^\infty(\mathbf{R} \times M)$. De plus, le champ u est quasilipschitzien par rapport à x , le flot ϕ^t de u à partir de $t = 0$ conserve le volume, et l'on a, pour tout réel t , pour presque tout $x \in M$,

$$\omega(t, \phi^t(x)) = \text{rot}u^0(x).$$

Démonstration. L'idée centrale dans ce résultat apparaît dans la démonstration suivante de l'unicité. En reprenant la démonstration donnée au §1.1, on constate que, puisque ∇v n'est plus L^∞ , on ne peut plus conclure aussi simplement par le lemme de Gronwall. En revanche, puisque les opérateurs pseudodifférentiels d'ordre 0 sont bornés sur L^p pour tout $p \in]1, \infty[$, on a $\nabla u(t) \in L^p(M)$, et l'inégalité de Hölder donne alors

$$\begin{aligned} \frac{d}{dt} \|u(t) - v(t)\|_{L^2}^2 &\leq 2\|\nabla v(t)\|_{L^p} \|u(t) - v(t)\|_{L^{2p'}}^2 \\ &\leq 2M^{2/p} \|\nabla v(t)\|_{L^p} \|u(t) - v(t)\|_{L^2}^{2/p'}, \end{aligned}$$

où p' désigne l'exposant conjugué de p et $M = \sup_{t \in [0, T]} [u(t) - v(t)]_0$. La belle idée de Yudovitch est alors d'étudier cette inéquation différentielle lorsque p tend vers l'infini, en s'appuyant sur le résultat suivant, corollaire du théorème d'interpolation de Marcinkiewicz (*cf.* par exemple Stein [St]) et de la théorie de Calderon–Zygmund (*cf.* par exemple [CM]):

Lemme 6. Soient M une variété compacte, A un opérateur pseudodifférentiel d'ordre 0 sur M . Alors il existe une constante C telle que, pour tout $p < \infty$, on ait l'estimation

$$\|Af\|_{L^p} \leq Cp\|f\|_{L^\infty}.$$

En utilisant le lemme 6 et en posant $g(t) = \|u(t) - v(t)\|_{L^2}^2$, on obtient

$$\frac{d}{dt}g \leq CpM^{2/p}g^{1-1/p}.$$

En intégrant cette inéquation et en tenant compte du fait que $g(0) = 0$, il vient, pour tout $t \geq 0$,

$$g(t) \leq M^2(Ct)^p.$$

Cette inégalité ayant lieu pour tout $p < \infty$, on en déduit que $g(t) = 0$ pour tout $t < 1/C$ et finalement $g = 0$, ce qui démontre l'unicité. Pour démontrer l'existence, on régularise la donnée de Cauchy u^0 et, en utilisant le théorème 1, on obtient une suite (u_n) de solutions régulières de (E) ; l'équation (T) entraîne que $\text{rot}u_n$ est borné dans L^∞ , donc $u_n(t)$ est borné dans C_*^1 uniformément par rapport à t . Par ailleurs, l'argument utilisé ci-dessus pour l'unicité montre que la suite (u_n) est de Cauchy dans $C^0([0, T], L^2)$ pour tout $T > 0$. Par interpolation, on en déduit que u_n converge uniformément vers une solution u de (E') ; les autres propriétés énoncées s'obtiennent alors aisément par passage à la limite.

3. DÉMONSTRATION DU THÉORÈME A.

3.1. Position du problème.

Soit u^0 un champ de vitesses sur M dont le tourbillon est

$$\omega^0 = a1_{D^0} + b(1 - 1_{D^0}),$$

où D^0 est un ouvert régulier de M , et où a et b sont des nombres réels vérifiant

$$a \text{vol}(D^0) + b(\text{vol}(M) - \text{vol}(D^0)) = 0.$$

Pour simplifier, nous supposons de plus que la frontière de D_0 est connexe. Désignons par u la solution de Yudovitch associée à cette donnée de Cauchy, par ϕ^t le flot de u à partir de $t = 0$. Alors le tourbillon de $u(t)$ est donné par

$$\omega(t) = a1_{D^t} + b(1 - 1_{D^t}),$$

où $D^t = \phi^t(D^0)$. Choisissons un champ régulier X^0 non nul sur un voisinage compact K^0 de ∂D^0 , tangent à cette courbe, et désignons par $\gamma^0 : \mathbf{S}_1 \rightarrow \partial D^0$ un difféomorphisme tel que

$$\frac{d}{ds}\gamma^0 = X^0(\gamma^0).$$

Pour étudier la régularité de $\gamma^t = \phi^t(\gamma^0)$, on remarque que, formellement, si le champ $X^t = \phi_*^t(X^0)$ était suffisamment régulier, on aurait

$$\frac{d}{ds}\gamma^t = X^t(\gamma^t), \quad \nabla_{\frac{d}{ds}}^k \frac{d}{ds}\gamma^t = \nabla_{X^t}^k X^t(\gamma^t).$$

Une information de régularité sur X^t est donc immédiatement transformée en une information analogue sur γ^t . On rend rigoureux ce raisonnement en régularisant les données et en considérant la suite (u_n) de solutions régulières correspondantes. En remarquant que $\gamma_n^t = \phi_n^t(\gamma^0)$ converge uniformément vers γ^t , on est ramené à estimer $X_n^t = \phi_{n*}^t(X^0)$ et ses dérivées covariantes par rapport à lui-même, par exemple en norme C^α , pour $\alpha \in]0, 1[$ fixé. Nous allons nous restreindre ici au cas de $[X_n^t]_\alpha$, ce qui entraînera que la courbe est $C^{1+\alpha}$. Les dérivées d'ordre supérieur se traitent ensuite plus facilement. L'information cruciale sur la suite des données de Cauchy régularisées (ω_n^0) est alors que $L_{X^0}\omega_n^0$ est bornée dans l'espace L^∞ , et que ω_n^0 est bornée dans C^α sur le complémentaire $(K^0)^c$ de K^0 . Outre une estimation de $[X_n^t]_\alpha$, on va en déduire une estimation de la norme Lipschitz de $u_n(t)$, qui lui est intimement liée, et on prouvera donc que *le champ $u(t)$ est en fait lipschitzien pour tout temps.*

Si K est un compact de M et X est un champ de vecteurs sur M , on pose

$$I(K, X) = \min_{x \in K} |X(x)|.$$

On obtient alors le résultat souhaité en appliquant à la suite de solutions régularisées le théorème suivant:

THÉORÈME 7. (Chemin [C4]). *Pour tout $\alpha \in]0, 1[$, il existe une constante C_α telle que, pour toute solution régulière u de l'équation d'Euler, pour tout compact K^0 de M , pour tout champ de vecteurs régulier X^0 ne s'annulant pas sur K^0 , on ait, pour tout $t \geq 0$, l'estimation suivante, où u^0 désigne la donnée de Cauchy de u et ω^0 son tourbillon :*

$$[u(t)]_1 \leq C_\alpha M(u^0) \log N_\alpha(K^0, X^0, \omega^0) e^{C_\alpha M(u^0)t},$$

$$\|X^t\|_\alpha \leq C_\alpha \left(\|X^0\|_\alpha + [\operatorname{div} X^0]_0 + \frac{[L_{X^0} \omega^0]_0}{[\omega^0]_0} \right) N_\alpha(K^0, X^0, \omega^0)^{2 \exp C_\alpha M(u^0)t}$$

$$\text{avec } M(u^0) = [\omega^0]_0 + \|u^0\|_{L^2},$$

$$N_\alpha(K^0, X^0, \omega^0) = 2 + \frac{\|X^0\|_\alpha + [\operatorname{div} X^0]_0}{I(K^0, X^0)} + \frac{[L_{X^0} \omega^0]_0}{I(K^0, X^0)[\omega^0]_0} + \frac{\|\omega^0\|_{\alpha, (K^0)^c}}{[\omega^0]_0}.$$

La démonstration du théorème 7 fait l'objet des deux sections suivantes.

3.2. L'estimation fondamentale.

Le point crucial dans la démonstration du théorème 7 est une inégalité logarithmique sur le laplacien en dimension 2:

Proposition 8. *Pour tout $\alpha \in]0, 1[$, il existe une constante $C_\alpha > 0$ telle que, pour tous champs de vecteurs u et X de classe C^∞ sur M , vérifiant $\operatorname{div} u = 0$, pour tout compact K de M sur lequel X ne s'annule pas, on ait l'inégalité*

$$[\nabla u]_0 \leq C([\omega]_0 + \|u\|_{L^2}) \log \left(2 + \frac{\|X\|_\alpha + [\operatorname{div} X]_0}{I(K, X)} + \frac{[L_X \omega]_0}{I(K, X)[\omega]_0} + \frac{\|\omega\|_{\alpha, K^c}}{[\omega]_0} \right),$$

où l'on a posé $\omega = \operatorname{rot} u$.

Démonstration. Elle repose sur les trois lemmes suivants.

Lemme 9. *Soit $\alpha \in]0, 1[$. On a les estimations suivantes entre normes hölderiennes et la norme L^∞ :*

$$(i) \quad [f]_0 \leq C_\alpha \|f\|_0 \log \left(2 + \frac{\|f\|_\alpha}{\|f\|_0} \right)$$

$$(ii) \quad [a \cdot b]_0 \leq C_{\alpha, p} [a]_0 \|b\|_0 \log \left(2 + \frac{\|a\|_\alpha}{[a]_0} + \frac{\|a \cdot b\|_\alpha}{[a]_0 \|b\|_0} \right),$$

$$\text{avec } a = (a_1, \dots, a_p), \quad b = (b_1, \dots, b_p), \quad a \cdot b = a_1 b_1 + \dots + a_p b_p$$

$$(iii) \quad \|L_Z B f\|_\alpha \leq C_{\alpha, \alpha', B} ([L_Z f]_0 + (\|Z\|_{\alpha'} + [\operatorname{div} Z]_0)[f]_0),$$

pour tout $\alpha' > \alpha$, tout opérateur pseudodifférentiel B d'ordre -1 et tout champ de vecteurs Z .

Il suffit bien sûr de montrer les trois inégalités ci-dessus dans \mathbf{R}^n . Elles découlent alors de la caractérisation des espaces de Hölder par les décompositions dyadiques et du calcul paradifférentiel de Bony [B1]; l'inégalité (iii) est une adaptation d'un lemme de commutateur de [B1]. Pour démontrer (i), on écrit, pour tout N ,

$$\begin{aligned} [f]_0 &\leq \sum_{p \leq N} [\Delta_p f]_0 + \sum_{p > N} [\Delta_p f]_0, \\ &\leq CN \|f\|_0 + C_\alpha 2^{-N\alpha} \|f\|_\alpha, \end{aligned}$$

et on conclut en optimisant le choix de N comme dans la preuve du lemme 3. Enfin, la preuve de (ii) combine ce type d'interpolation logarithmique avec le paraproduit de Bony. Notons en passant que, si l'on applique (i) à $f = Ag$, avec A pseudodifférentiel d'ordre 0, on obtient, puisque A est borné dans C_*^0 ,

$$\begin{aligned} [Af]_0 &\leq C_\alpha \|f\|_0 \log \left(2 + \frac{\|f\|_\alpha}{\|f\|_0} \right), \\ &\leq C_\alpha [f]_0 \log \left(2 + \frac{\|f\|_\alpha}{[f]_0} \right), \end{aligned}$$

ce qui est une forme précisée du lemme 2. (avec $A = \nabla B$.) Voici une variante de ce type d'estimation, où la régularité hölderienne est cette fois remplacée par une hypothèse de localisation:

Lemme 10. Soient A un opérateur pseudodifférentiel d'ordre 0. Il existe $C > 0$ tel que, si F et G sont deux fermés disjoints dans M , et si f est une fonction essentiellement bornée supportée par F , alors

$$\sup_{x \in G} |Af(x)| \leq C [f]_0 \left(1 + \log_+ \frac{1}{d(F, G)} \right).$$

Là encore il suffit de raisonner en coordonnées locales; le lemme découle alors immédiatement de la formule

$$Af(x) = \int K(x, y) f(y) dy$$

où K est un noyau à support compact vérifiant $K(x, y) \leq C|x - y|^{-n}$. Cette approche donne d'ailleurs une autre démonstration du lemme 2.

Enfin, la dimension 2 est utilisée dans la remarque élémentaire suivante:

Lemme 11. Soient E un plan vectoriel réel, Y un vecteur non nul de E et J un endomorphisme de E tel que $J^2 = -1$. Alors l'application

$$\begin{aligned} \text{End}(E) &\longrightarrow \mathbf{R}^2 \times E \\ T &\mapsto (\text{tr}(T), \text{tr}(JT), T(Y)) \end{aligned}$$

est un isomorphisme d'espaces vectoriels.

Indiquons maintenant brièvement comment on aboutit à la proposition 8. Grâce au lemme 10, l'estimation de ∇u ne pose pas de problème dans un domaine L suffisamment éloigné de K ; en effet, écrivons $\omega = \omega_1 + \omega_2$, où ω_1 est supporté dans un voisinage K' de K disjoint de L , et où ω_2 est supporté dans K^c . D'après le lemme 10, la contribution de ω_1 à ∇u sur L est estimée par $[\omega]_0 \log(2 + \theta^{-1})$, où $\theta = d(L, K')$, tandis que la contribution de ω_2 est estimée à l'aide du lemme 9, (i), par $[\omega]_0 \log(2 + \theta^{-1} + \|\omega\|_{\alpha, K^c} / [\omega]_0)$, si l'on suppose de plus que $d(K, K'^c)$ est également de l'ordre de θ . Le paramètre θ sera ajusté plus loin.

Il reste à estimer ∇u sur le voisinage $K'' = L^c$ de K . Pour cela, on utilise la conséquence suivante du lemme 11 : il existe $C > 0$ tel que, pour toute section T de $\text{End}TM$, pour tout champ de vecteurs Y sur M tel que $|Y(x)| = 1$ sur K'' , on a, pour tout $x \in K''$,

$$|T(x)| \leq C(|\text{tr}(T(x))| + |\text{tr}(JT(x))| + |T(x)Y(x)|).$$

Appliquant cette inégalité à $T = \nabla u$, on obtient

$$(3) \quad |\nabla u(x)| \leq C(|\omega(x)| + |\nabla_Y u(x)|).$$

On choisit alors $Y = \chi_\theta X / |X|$, où χ_θ est une fonction C^∞ valant 1 près de K'' et supportée dans un voisinage K''' d'ordre θ de K'' , de sorte que l'on ait, pour tout $\beta \in]0, 1[$,

$$\|\chi_\theta\|_\beta \leq C\theta^{-\beta}.$$

Le choix de θ intervient ici pour garantir que l'on puisse minorer $|X|$ sur K''' ; en choisissant $\theta = (I(K, X) / A \|X\|_\alpha)^{1/\alpha}$ où A est une constante assez grande, on a facilement

$$I(K''', X) \geq \frac{1}{2} I(K, X), \quad \|Y\|_\alpha \leq C \frac{\|X\|_\alpha}{I(K, X)}, \quad \|\nabla_Y u\|_\alpha \leq C \frac{\|X\|_\alpha \|\nabla_X u\|_\alpha}{I(K, X)^2}.$$

Le lemme 9, (ii), conduit alors à

$$[\nabla_Y u]_0 \leq C([\omega]_0 + \|u\|_{L^2}) \log \left(2 + \frac{\|X\|_\alpha}{I(K, X)} + \frac{\|\nabla_X u\|_\alpha}{I(K, X)([\omega]_0 + \|u\|_{L^2})} \right),$$

où l'on a de plus utilisé que $\|u\|_0 \leq C([\omega]_0 + \|u\|_{L^2})$. En revenant à l'inégalité (3) et en appliquant le lemme 9, (iii), pour estimer $\|\nabla_X u\|_\alpha$, on obtient la proposition 8.

3.3. Propagation de la régularité.

La fin de la preuve du théorème 7 s'appuie sur le lemme suivant de propagation de la régularité höldérienne dans une équation de transport:

Lemme 12. *Soit $\alpha \in]0, 1[$. Il existe une constante C telle que, pour tout $u \in C^\infty(\mathbf{R}_+ \times M, TM)$ et pour toutes fonctions f, g sur $\mathbf{R}_+ \times M$ vérifiant $\partial_t f + L_u f = g$, on ait l'estimation*

$$\|f(t)\|_\alpha \leq \|f(0)\|_\alpha e^{\int_0^t [u(s)]_1 ds} + \int_0^t \|g(s)\|_\alpha e^{\int_s^t [u(\sigma)]_1 d\sigma} ds.$$

L'estimation ci-dessus s'obtient aisément à l'aide de la méthode des caractéristiques.

Reprenons les notations du paragraphe 3.1; de l'équation de transport

$$\partial_t L_X \omega + L_u L_X \omega = 0,$$

on déduit d'abord que $[L_X \omega]_0 = [L_X^0 \omega^0]_0$. Par ailleurs, puisque ϕ^t conserve le volume,

$$[\operatorname{div} X]_0 = [\operatorname{div} X^0]_0.$$

On applique alors le lemme 12 à l'équation de transport

$$\partial_t X + \nabla_u X = \nabla_X u,$$

en tenant compte de l'estimation

$$\|\nabla_X u\|_\alpha \leq C_\alpha ([L_X \omega]_0 + [u]_1 (\|X\|_\alpha + [\operatorname{div} X]_0)),$$

qui est une variante du lemme 9,(iii) ; on obtient, par le lemme de Gronwall,

$$\|X^t\|_\alpha \leq C_\alpha \left(\|X^0\|_\alpha + [\operatorname{div} X^0]_0 + \frac{[L_{X^0} \omega^0]_0}{[\omega^0]_0} \right) e^{2 \int_0^t [u(s)]_1 ds}.$$

De plus, on montre directement que

$$I(K^0, X^0) \leq I(K^t, X^t) e^{\int_0^t [u(s)]_1 ds}, \quad \|\omega(t)\|_{\alpha, (K^t)^c} \leq \|\omega^0\|_{\alpha, (K^0)^c} e^{\int_0^t [u(s)]_1 ds}.$$

En injectant ces trois estimations dans l'inégalité fournie par la proposition 8, on obtient finalement

$$[u(t)]_1 \leq C_\alpha ([\omega^0]_0 + \|u^0\|_{L^2}) \left(\log N_\alpha(K^0, X^0, \omega^0) + \int_0^t [u(s)]_1 ds \right),$$

où l'on a posé

$$N_\alpha(K^0, X^0, \omega^0) = 2 + \frac{\|X^0\|_\alpha + [\operatorname{div} X^0]_0}{I(K^0, X^0)} + \frac{[L_{X^0} \omega^0]_0}{I(K^0, X^0)[\omega^0]_0} + \frac{\|\omega^0\|_{\alpha, (K^0)^c}}{[\omega^0]_0}.$$

On en déduit le théorème 7.

4. DÉMONSTRATION DU THÉORÈME B.

4.1. Elimination de la pression.

Lemme 13. Soient $T > 0$ et $u \in L^2([0, T] \times M, TM)$. Les conditions suivantes sont équivalentes:

(i) u est solution du système d'Euler

$$(E') \quad \partial_t u + \operatorname{div}(u \otimes \tilde{u}) = -\nabla p, \quad \operatorname{div} u = 0.$$

(ii) u est solution de l'équation

$$(E'') \quad \partial_t u + A \left(u \otimes \tilde{u} - \frac{1}{2} g(u, u) I \right) = 0, \quad \operatorname{div} u = 0,$$

où $A = J \nabla \Delta^{-1} \operatorname{rot} \operatorname{div} + H \operatorname{div}$.

Considérons l'opérateur pseudodifférentiel $P = J\nabla\Delta^{-1}\text{rot} + H$, de sorte que $A = P\text{div}$. D'après la décomposition de Hodge des sections-distributions de TM (ou champs distributions), P est le projecteur sur l'espace des champs à divergence nulle parallèlement à l'espace des champs de gradients. Le système d'Euler (E') est donc équivalent à

$$\partial_t u + P\text{div}(u \otimes \tilde{u}) = 0, \quad \text{div} u = 0.$$

Pour toute distribution scalaire f , on a $\text{div}(fI) = \nabla f$; on peut donc ajouter à l'argument de $P\text{div}$ dans la première équation ci-dessus n'importe quel tenseur du type fI ; on choisit $f = -\frac{1}{2}g(u, u)$, et on obtient le résultat.

Remarques.

Le choix très arbitraire de f que nous venons de faire va trouver sa justification dans la section 4.3 ci-dessous.

L'équation (E'') montre en particulier que toute solution u du système d'Euler qui est de carré intégrable sur $]0, T[\times M$ est continue sur $[0, T]$ à valeurs dans les champs distributions sur M . On pourra donc parler, pour *tout* $t \in [0, T]$, du champ distribution $u(t)$ et de la distribution $\omega(t)$.

4.2. Un théorème de stabilité.

Rappelons qu'une famille \mathcal{F} de fonctions τ -intégrables sur M est dite uniformément τ -intégrable si, pour tout $\varepsilon > 0$, il existe $\delta > 0$ tel que, pour toute partie borélienne E de M vérifiant $\int_E \tau \leq \delta$, on ait

$$\forall f \in \mathcal{F}, \int_E |f| \tau \leq \varepsilon.$$

Le critère de Dunford-Pettis (cf. par exemple [Gr]) assure que la famille \mathcal{F} est alors relativement compacte dans $L^1(M, \tau)$ pour la topologie faible $\sigma(L^1, L^\infty)$. Le théorème B est en fait conséquence du résultat général suivant.

THÉORÈME 14. (Delort [D1,2]). Soient $T > 0$ et (u_n) une suite de solutions de carré intégrable sur $]0, T[\times M$ du système d'Euler conservatif (E'). On suppose que, pour tout $t \in [0, T]$, $u_n(t)$ est de carré intégrable, que son tourbillon $\omega_n(t)$ est une mesure de Radon sur M , et qu'il existe une constante C telle que

$$(4) \quad \forall t \in [0, T], \int_M |u_n(t)|^2 \tau + \int_M |\omega_n(t)| \leq C.$$

En outre, on suppose que, pour tout $t \in [0, T]$, il existe une suite $\sigma_n(t)$ de mesures de Radon positives (resp. négatives) sur M telle que $(\omega_n(t) - \sigma_n(t))$ soit uniformément τ -intégrable sur M pour $n \in \mathbf{N}, t \in [0, T]$. Alors toute valeur d'adhérence de la suite (u_n) pour la topologie faible de $L^2([0, T] \times M, TM)$ est solution du système d'Euler (E').

La démonstration du théorème 14 fait l'objet de la section suivante. Dans l'immédiat, indiquons comment il entraîne le théorème B.

Soit u^0 un champ L^2 à divergence nulle dont le tourbillon ω^0 est une mesure de Radon à partie singulière σ^0 positive (resp. négative). Posons $\omega_n^0 = e^{\Delta/n} \omega^0$ et $u_n^0 = B\omega_n^0 + H(u^0)$. Il est clair que u_n^0 converge vers u^0 en moyenne quadratique. Soit (u_n) la suite de solutions C^∞ de (E) sur $[0, \infty[\times M$, dont la suite des données de Cauchy est (u_n^0) . Alors la conservation de l'énergie cinétique et l'équation de transport (T) entraînent les estimations (4). Par ailleurs, soit σ_n la solution de l'équation de transport (T) avec $u = u_n$ et $\sigma_n(0) = e^{\Delta/n} \sigma^0$. On a donc $\sigma_n \geq 0$ (resp. ≤ 0). Remarquons que $\text{rot} u_n^0 - \sigma_n^0$ n'est autre que $e^{\Delta/n} f^0$, où f^0 est la dérivée de Radon-Nikodym de ω^0 par rapport à τ . Il en résulte que la suite $(\omega_n^0 - \sigma_n^0)$ est uniformément τ -intégrable, et, en utilisant à nouveau l'équation de transport (T) et le fait que le flot de u_n conserve le volume, on en déduit que la famille $(\omega_n(t) - \sigma_n(t))_{n,t}$ est uniformément τ -intégrable. Puisque, d'après (4), la suite (u_n) est bornée dans $L^2([0, T] \times M)$, elle admet une valeur d'adhérence faible u qui est donc solution de (E'). De plus, l'équation (E'') entraîne que la suite (u_n) est équicontinue sur $[0, T]$ à valeurs dans les champs distributions sur M , donc $u(0)$ est la limite de u_n^0 , c'est-à-dire u^0 . De même, pour tout $t \in [0, T]$, $u(t)$ est valeur d'adhérence de $u_n(t)$ donc de carré intégrable, et son tourbillon est une mesure de Radon à partie singulière positive (resp. négative), ce qui achève la démonstration du théorème B.

4.3. Démonstration du théorème 14.

En utilisant l'équicontinuité en temps découlant de l'équation (E''), on constate qu'il suffit de montrer que si, pour tout $t \in [0, T]$, $u_n(t)$ converge faiblement vers $u(t)$ dans $L^2(M, TM)$, alors u est solution de (E''). Compte tenu de l'équation (E'') et du théorème de convergence dominée, le théorème 14 est donc conséquence du résultat suivant, indépendant de la variable t :

Lemme 15. Soit (u_n) une suite bornée de champs de vecteurs de carrés intégrables à divergence nulle, dont la suite des tourbillons (ω_n) est bornée dans l'espace des mesures de Radon. On suppose qu'il existe une suite (σ_n) de mesures de Radon positives (resp. négatives) sur M telle que la suite $(\omega_n - \sigma_n)$ soit uniformément τ -intégrable. Alors, si u_n converge vers u faiblement, $u_n \otimes \tilde{u}_n - g(u_n, u_n)I/2$ converge vers $u \otimes \tilde{u} - g(u, u)I/2$ au sens des tenseurs distributions.

Bien sûr, nous allons montrer le lemme 15 en utilisant la relation de Biot et Savart

$$u_n = B\omega_n + H(u_n).$$

Puisque la suite $H(u_n)$ est bornée dans l'espace de dimension finie \mathcal{H} , elle est relativement compacte, et l'on est ramené au cas où $u_n = B\omega_n$. Le lemme 15 est alors conséquence d'un résultat très général de "compensation microlocale" par positivité, que nous allons énoncer en dimension quelconque.

Pour cela, il est utile d'introduire un peu de vocabulaire. Soit E un espace vectoriel réel de dimension d , et soit f une fonction continue sur E , homogène de degré $-d$. On désigne par ρ le champ radial sur $E \setminus \{0\}$, donné en coordonnées linéaires (ξ_1, \dots, ξ_d) par $\rho = \sum \xi_j \partial_{\xi_j}$. Soit λ une mesure de Lebesgue sur E , et soit φ une fonction de classe C^1 sur E , à support compact, et valant 1 près de l'origine. Alors la quantité $\int_E f L_\rho \varphi d\lambda$ ne dépend pas de φ . En effet, si ψ est une fonction ayant les mêmes propriétés que φ , la fonction de classe C^1 $\varphi - \psi$ est supportée dans un compact de $E \setminus \{0\}$, et une intégration par parties donne

$$\int_E f L_\rho(\varphi - \psi) d\lambda = \int_E (\varphi - \psi)(-L_\rho - d)f d\lambda,$$

qui est nul d'après la relation d'Euler. Suivant Wodzicki (*cf.* par exemple l'exposé de Kassel [K] dans ce séminaire), désignons par $\text{res}(f, \lambda)$ cette quantité. Remarquons qu'elle possède une interprétation métrique très simple : si Q est un produit scalaire sur E dont λ est l'élément de volume, en approchant la fonction indicatrice de la boule unité de Q par de telles fonctions φ , on obtient que $\text{res}(f, \lambda)$ est l'intégrale de f sur la sphère unité de Q pour la mesure de Lebesgue.

Supposons maintenant que M soit une variété de dimension d et que f soit une fonction continue sur $T^*M \setminus M$, homogène de degré $-d$. D'après ce qui précède,

l'expression en coordonnées locales

$$\text{res}(f) = |dx^1 \wedge \dots \wedge dx^d| \text{res} \left(f(x, \cdot), |d\xi_1 \wedge \dots \wedge d\xi_d| \right)$$

définit sur M une densité (c'est-à-dire une d -forme tordue, au sens de [VAR]) indépendante du choix des coordonnées ; convenons d'appeler résidu de f cette densité. Le résultat clé de Delort s'énonce alors de la façon suivante :

Lemme 16. *Soit M une variété compacte de dimension d munie d'une densité > 0 identifiant distributions et distributions-densités. Soit (ω_n) une suite bornée de mesures de Radon sur M , également bornée dans l'espace de Sobolev $H^{-d/2}(M)$, et convergeant faiblement vers ω . On suppose qu'il existe une suite (σ_n) de mesures de Radon positives et une suite (f_n) uniformément intégrable sur M telles que*

$$\omega_n = \sigma_n + f_n.$$

Alors, pour tout opérateur pseudodifférentiel scalaire A sur M , d'ordre $-d$, et dont le symbole principal a un résidu nul, on a

$$(A\omega_n, \omega_n)_{H^{d/2}, H^{-d/2}} \rightarrow (A\omega, \omega)_{H^{d/2}, H^{-d/2}} .$$

Démonstration. On suppose par exemple $\sigma_n \geq 0$. En utilisant une partition de l'unité, on se ramène à prouver un résultat analogue sur \mathbf{R}^d , ω_n étant supportée dans un compact fixe.

Première étape. Si ω est une mesure de Radon à support compact dans \mathbf{R}^d appartenant de plus à $H^{-d/2}$, alors ω est diffuse, c'est-à-dire que, pour tout point x_0 , $\omega(\{x_0\}) = 0$.

En effet, si φ est une fonction C^∞ à support compact près de l'origine, le théorème de convergence dominée nous indique que

$$\omega(\{x_0\}) = \lim_{r \rightarrow 0} \int \varphi\left(\frac{x - x_0}{r}\right) \omega(dx).$$

Comme la famille $(\varphi((x - x_0)/r))$ est bornée dans $H^{d/2}$ et converge faiblement vers 0, on en déduit l'assertion.

Deuxième étape. Si de plus $\omega = \sigma + f$, où $\sigma \geq 0$ et $f \in L^1$, alors, pour toute fonction $g \in L^1$, $\sigma + g$ est diffuse.

En effet, $\sigma + g = \omega + g - f$.

Troisième étape. Le noyau de l'opérateur A est de la forme $k(x, x - y)$, où $k = k(x, z)$ est une fonction de classe C^∞ sur $\mathbf{R}^d \times \mathbf{R}^d \setminus \{0\}$, et l'hypothèse de résidu nul signifie exactement que $k(x, z)$ est bornée près du lieu singulier $z = 0$ (cf. par exemple [CM]). De plus, si φ est une fonction C^∞ à support compact valant 1 près de l'origine, on a, si f et g appartiennent à $H^{-d/2}$,

$$(Af, g)_{H^{d/2}, H^{-d/2}} = \lim_{r \rightarrow 0} \int k(x, x - y) \left(1 - \varphi\left(\frac{x - y}{r}\right)\right) f(y) \bar{g}(x) dx dy.$$

Dès lors, si ω est comme à la première étape, on a

$$(A\omega, \omega)_{H^{d/2}, H^{-d/2}} = \int k(x, x - y) \omega(dy) \bar{\omega}(dx),$$

d'après le théorème de convergence dominée, puisque, en vertu du théorème de Fubini, la mesure $\omega \otimes \bar{\omega}$ ne charge pas la diagonale.

Quatrième étape. On utilise maintenant le lemme élémentaire suivant de théorie de la mesure:

Lemme 17. Soient X un espace métrique compact, (μ_n) une suite de mesures de Radon sur X convergeant vaguement vers μ , et (ν_n) une suite de mesures de Radon positives convergeant vaguement vers ν , telle que $|\mu_n| \leq \nu_n$. Si f est une fonction borélienne bornée sur X , continue en dehors d'un fermé de mesure nulle pour ν , alors

$$\int_X f d\mu_n \rightarrow \int_X f d\mu.$$

Il suffit d'appliquer ce lemme à $X = K \times K$ où K est un compact de \mathbf{R}^d , $\mu_n = \omega_n \otimes \bar{\omega}_n$, $\mu = \omega \otimes \bar{\omega}$, $\nu_n = \sigma_n + |f_n|$. Comme (f_n) est uniformément intégrable, on peut supposer que f_n converge faiblement vers $f \in L^1$ et que $|f_n|$ converge faiblement vers $g \in L^1$. Alors σ_n converge vaguement vers la mesure positive $\sigma = \omega - f$ et on peut prendre $\nu = (\sigma + g) \otimes (\sigma + g)$. Il reste à vérifier que ν ne charge pas la diagonale, ce qui est une conséquence immédiate de la seconde étape et du théorème de Fubini.

La démonstration du lemme 16 est maintenant aisée. Soient α une 1-forme et X un champ de vecteurs réguliers sur M . Alors

$$\langle u_n \otimes \tilde{u}_n - \frac{1}{2}g(u_n, u_n), \alpha \otimes X \rangle_{\mathcal{D}', \mathcal{D}} = (A\omega_n, \omega_n)_{H^1, H^{-1}},$$

où $A = \langle \alpha, B \rangle \langle \tilde{B}, X \rangle - \frac{1}{2} \langle \tilde{B}, X \rangle B$. Son symbole principal est donc, sur la sphère unité de g dans T^*M ,

$$a(x, \xi) = g(*\xi, \alpha)g(*\xi, \tilde{X}) - \frac{1}{2}g(\alpha, \tilde{X}).$$

Son résidu est donc

$$\tau \int_{S^1} \left(g(\xi, \alpha)g(\xi, \tilde{X}) - \frac{1}{2}g(\alpha, \tilde{X}) \right) d\sigma(\xi) = 0,$$

comme le montre un calcul élémentaire. Ceci achève la démonstration.

Remarque.

La démonstration ci-dessus s'étend au cas d'une suite (u_n) de solutions de l'équation de Navier-Stokes

$$\partial_t u_n + \operatorname{div}(u_n \otimes \tilde{u}_n) - \nu_n \Delta u_n = -\nabla p_n, \quad \operatorname{div} u_n = 0,$$

où (ν_n) est une suite de nombres > 0 convergeant vers zéro, modélisant la viscosité du fluide. Les solutions mises en évidence par Delort peuvent donc être vues comme le modèle limite de nappes de tourbillon dans un fluide peu visqueux.

BIBLIOGRAPHIE

- [A1] S. ALINHAC - *Remarques sur l'instabilité du problème des poches de tourbillon*, J. of Functional Analysis **98** (1991), 361-379.
- [A2] S. ALINHAC - *Un phénomène de concentration évanescence pour des flots non-stationnaires incompressibles en dimension 2*, Commun. Math. Phys. **127** (1990), 585-596.
- [Ar] V. ARNOLD - *Sur la géométrie différentielle des groupes de Lie de dimension infinie et ses applications à l'hydrodynamique des fluides parfaits*, Ann. Inst. Fourier **16** (1966), 319-361.
- [BF] C. BARDOS, U. FRISCH, C. SULEM, P.-L. SULEM - *Finite time analyticity for the two and three-dimensional Kelvin-Helmholtz instability*, Commun. Math. Phys. **80** (1981), 485-516.

- [Ba] G.K. BATCHELOR - *An Introduction to Fluid Mechanics*, Cambridge University Press, 1970.
- [Be] A.L. BERTOZZI - *Existence, uniqueness, and a necessary and sufficient condition for blow up of smooth solutions to the contour dynamics equation*, prépublication, Université de Princeton, Mai 1992.
- [BC] A.L. BERTOZZI, P. CONSTANTIN - *Global regularity for vortex patches*, prépublication, Université de Chicago, Mars 1992.
- [Bi] G. BIRKHOFF - *Helmholtz and Taylor instability*, in *Hydrodynamic instability*, Proc. Symp. Appl. Math. vol.13, AMS, 1962, 55-76.
- [B1] J.-M. BONY - *Calcul symbolique et propagation des singularités pour les équations aux dérivées partielles non linéaires*, Ann. Sc. Ecole Normale Sup. **14** (1981), 209-246.
- [B2] J.-M. BONY - *Analyse microlocale et singularités non linéaires*, in *Nonlinear Hyperbolic Problems*, Lecture Note in Mathematics n^o 1402, Springer, 1989.
- [Br] Y. BRENIER - *The least action principle and the related concept of generalized flows for incompressible perfect fluids*, J. Amer. Math. Soc. **2** (1989), 225-255.
- [CO] R. CAFLISH, O. ORELLANA - *Longtime existence for a slightly perturbed vortex sheet*, Comm. Pure Appl. Math. **39** (1986), 807-838.
- [C1] J.-Y. CHEMIN - *Autour du problème des vortex patches*, Séminaire Equations aux dérivées partielles, 1989-1990, Ecole Polytechnique.
- [C2] J.-Y. CHEMIN - *Régularité des trajectoires d'un fluide parfait incompressible remplissant l'espace*, prépublication, Ecole Polytechnique 1990, à paraître au J. Math. Pures et Appliquées.
- [C3] J.-Y. CHEMIN - *Sur le mouvement des particules d'un fluide parfait incompressible bidimensionnel*, Inventiones Math. **103** (1991), 599-629.
- [C4] J.-Y. CHEMIN - *Persistance de structures géométriques dans les fluides incompressibles bidimensionnels*, prépublication, Ecole Polytechnique, Avril 1991.
- [C5] J.-Y. CHEMIN - *Evolution d'une singularité ponctuelle dans des équations strictement hyperboliques non linéaires*, Amer. J. Math. **112** (1990), 805-860..
- [CL] R. COIFMAN, P.-L. LIONS, Y. MEYER, S. SEMMES - *Compensated compactness and Hardy spaces*, prépublication, Université Paris-Dauphine, n^o 9123, 1991.

- [CM] R. COIFMAN, Y. MEYER - *Au-delà des opérateurs pseudo-différentiels*, Astérisque **57** Société Mathématique de France, 1978.
- [CT] P. CONSTANTIN, E. TITI - *On the evolution of nearly circular vortex patches*, Commun. Math. Phys. **119** (1988), 177-198.
- [D1] J.-M. DELORT - *Existence de nappes de tourbillon pour l'équation d'Euler sur le plan*, Séminaire Equations aux dérivées partielles, 1990-1991, Ecole Polytechnique.
- [D2] J.-M. DELORT - *Existence de nappes de tourbillon en dimension deux*, J. Amer. Math. Soc. **4** (1991), 553-586.
- [DM1] R.J. DIPERNA, A. MAJDA - *Oscillations and concentrations in weak solutions of the incompressible fluid equations*, Commun. Math. Phys. **108** (1987), 667-689.
- [DM2] R.J. DIPERNA, A. MAJDA - *Concentrations in regularizations for 2-D incompressible flow*, Comm. Pure Appl. Math. **40** (1987), 301-345.
- [DM3] R.J. DIPERNA, A. MAJDA - *Reduced Hausdorff dimension and concentration-cancellation for 2-D incompressible flow*, J. Amer. Math. Soc. **1** (1988), 59-95.
- [D] D.G. DRITSCHEL - *Contour dynamics and contour surgery : numerical algorithms for extended, high-resolution modeling of vortex dynamics in two-dimensional, inviscid, incompressible flows*, Computer Physics Reports **10** (1989), 77-146.
- [DI] D.G. DRITSCHEL, M.E. McINTIRE - *Does contour dynamics go singular?*, Physics of Fluids A **2** (1990), 748-753.
- [DR] J. DUCHON, R. ROBERT - *Global Vortex Sheet Solutions of Euler Equations in the Plane*, Journal of Differential Equations **73** (1988), 215-224.
- [EB] D. EBIN, J.E. MARSDEN - *Groups of diffeomorphisms and the motion of an incompressible fluid*, Ann. of Math. **40** (1970), 102-163.
- [E] L. EULER - *Principes généraux du mouvement des fluides*, Mémoires de l'Académie des Sciences de Berlin **11** (1755), 274-315, 316-361, Opera Omnia, series Secunda, livre 12.
- [EM] L.C. EVANS, S. MÜLLER - *Hardy spaces and the two-dimensional Euler equations with nonnegative vorticity*, prépublication.
- [GT] C. GREENGARD, E. THOMANN - *On DiPerna-Majda concentration sets for two-dimensional incompressible flow*, Comm. Pure Appl. Math. **41**

- (1988), 295-303.
- [Gr] A. GROTHENDIECK - *Espaces vectoriels topologiques*, S.Paulo, 1958.
- [H] H.L.F. von HELMHOLTZ - *Über discontinuierlich Flüssigkeitsbewegungen*, Monastberichte der konigl. Akad. Wissenschaften zu Berlin (1868), 215-228.
- [K] C. KASSEL - *Le résidu non commutatif (d'après M. Wodzicki)*, Séminaire Bourbaki, 1988-89, exposé n° 708, Astérisque **177-178**, Société Mathématique de France.
- [Ka] T. KATO - *On classical solutions of the two-dimensional non-stationary Euler equation*, Arch. Rat. Mech. Anal. **27** (1968), 188-200.
- [Ke] Lord KELVIN (Sir W. THOMPSON) - *Mathematical and physical papers*, **4** Cambridge University Press, 1910.
- [Kr] R. KRASNY - *Desingularization of periodic vortex sheet roll-up*, J. Comput. Phys. **65** (1986), 292-313.
- [L] G. LEBEAU - *Interaction des singularités pour les équations aux dérivées partielles non linéaires (d'après J.-M. Bony et al.)*, Séminaire Bourbaki, 1984-85, exposé n° 642, Astérisque **133-134**, Société Mathématique de France.
- [Li] L.LICHTENSTEIN - *Über einige Existenzprobleme der Hydrodynamik homogener unzusammendrückbarer, reibungsloser Flüssigkeiten und die Helmholtz'schen Wirbelsätze*, Mathemat. Zeit. **23** (1925), 89-154, 310-316, **26** (1927), 196-323, 387-415, 725, **32** (1930), 608.
- [Lo] P.-L. LIONS - *Remarks on incompressible models of Fluid Mechanics*, pré-publication, Université de Paris-Dauphine, n° 9206, 1992.
- [M1] A. MAJDA - *Vorticity and the mathematical theory of incompressible fluid flow*, Comm. Pure Appl. Math. **39** (1986), supplément, 187-220.
- [M2] A. MAJDA - *Vortex dynamics : Numerical Analysis, scientific computing, and mathematical theory*, Proceedings of the First International Congress for Industrial and Applied Mathematics, S.I.A.M.Publ., 1988, 153-182.
- [M3] A. MAJDA - *The interaction of nonlinear analysis and modern applied mathematics*, Proc. I.C.M., Kyoto, 1990.
- [R] G. de RHAM - *Variétés différentiables*, Hermann, Actualités Sci. Indust., Paris, 1960.
- [SB] P. SAFFMAN, G. BAKER - *Vortex interactions*, Ann. Rev. Fluid Mech. **11** (1979), 95-122.

- [S] P. SERFATI - *Vortex patches bidimensionnels*, manuscrit, Février 1990, et thèse Université de Paris VI, septembre 1992.
- [Sh] A. SHNIRELMAN - *On the geometry of the group of diffeomorphisms and the dynamics of an ideal incompressible fluid*, Math. USSR Sbornik **56** (1987), 79-105.
- [St] E. STEIN - *Singular Integrals and Differentiability Properties of Functions*, Princeton University Press, 1970.
- [T] Sir G. TAYLOR - *The instability of liquid surfaces when accelerated in a direction perpendicular to their planes*, Proc. Roy. Soc. London, Ser.A, vol. **201** (1950), 192-196.
- [VAR] N. BOURBAKI - *Variétés différentielles et analytiques, fascicule de résultats, paragraphes 8 à 15*, Hermann, Paris, 1971.
- [W] W. WOLIBNER - *Un théorème sur l'existence du mouvement plan d'un fluide parfait, homogène, incompressible, pendant un temps infiniment long*, Math. Z. **37** (1933), 698-726.
- [Y] V. YUDOVITCH - *Niestatsionarnye tchetenia idealnoï niecjimaiemoï jidkosti*, Journal vtychisl. mat. i mat. fis. **3** (1963), 1032-1066.
- [Z] N. ZABUSKY, M. HUGHES, K. ROBERTS - *Contour dynamics for the Euler equations in two dimensions*, J. Comput. Phys. **30** (1979), 96-106.

Patrick GÉRARD

Université de Paris-Sud
Département de Mathématiques
Bâtiment 425
F-91405 ORSAY Cedex

Astérisque

LEONID PASTUR

Eigenvalue distribution of random operators and matrices

Astérisque, tome 206 (1992), Séminaire Bourbaki,
exp. n° 758, p. 445-461

<http://www.numdam.org/item?id=SB_1991-1992__34__445_0>

© Société mathématique de France, 1992, tous droits réservés.

L'accès aux archives de la collection « Astérisque » (<http://smf4.emath.fr/Publications/Asterisque/>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

EIGENVALUE DISTRIBUTION OF RANDOM OPERATORS AND MATRICES

by Leonid PASTUR

1. INTRODUCTION

The study of the eigenvalue distribution of various classes of selfadjoint operators is one of the oldest and most popular branches of the spectral theory. Recall, for instance, the well-known problem of H. Weyl on the high energy asymptotic behaviour of the counting function

$$(1) \quad \mathcal{N}_\Lambda(\lambda) = \#\{\lambda_i \leq \lambda\}$$

of eigenvalues λ_i of a selfadjoint boundary value problem for an elliptic operator A_Λ in a compact domain $\Lambda \subset \mathbb{R}^d$. According to H. Weyl, in the simplest but important case of $A_\Lambda = -\Delta$,

$$(2) \quad \mathcal{N}_\Lambda(\lambda) = c_d |\Lambda| \lambda^{d/2} + o(\lambda^{d/2}), \quad \lambda \rightarrow \infty,$$

where $|\Lambda|$ is the volume of Λ and c_d depends only on the dimensionality of the space.

The study of the subsequent terms on this and similar asymptotic formulae for various cases has produced a wide variety of beautiful and important results and has revealed many deep interconnections of the spectral theory with geometry, topology, ergodic theory, various analytic and asymptotic methods, etc.

Let me remind now that one of the main Weyl's motivations for studying this problem was justification of Rayleigh and Jeans derivation of the formula

for the spectral distribution of the black body radiation and of Debye derivation of the formula for the specific heat of a crystal. These problems that played an important role in the development of quantum mechanics, can be formulated as problems of constructing of the thermodynamics of the ideal Bose gas. One of the main mathematical physics concepts in statistical mechanics is the concept of the thermodynamic limit, designed in order to study the bulk properties of macroscopically large systems. From this point of view it is rather natural to consider some sequence Λ_k of compact domains expanding into the whole \mathbb{R}^d as $k \rightarrow \infty$, to prove the existence for each $\lambda \in \mathbb{R}$ of the limit

$$(3) \quad \lim_{k \rightarrow \infty} N_{\Lambda_k}(\lambda) \equiv N(\lambda),$$

where

$$(4) \quad N_{\Lambda}(\lambda) = |\Lambda|^{-1} \mathcal{N}_{\Lambda}(\lambda),$$

to check the independence of this limit of a sequence $\{\Lambda_k\}$ for a sufficiently broad family of sequences and after that to study the nondecreasing function (3) for various ranges of λ (in particular, for $\lambda \rightarrow \infty$) and other parameters.

In the case of the Laplacian and a sufficiently regular sequence of domains this problem can be reduced to the problem of finding the high energy asymptotic of $N_{\Lambda}(\lambda)$ for a fixed Λ . This is why H. Weyl was rather interested in the proof of the independence of the leading term in (2) of the shape of a domain and considered this result as one of his main achievements.

The problem of studying $N_{\Lambda}(\lambda)$, known as the integrated density of states (IDS), can also be formulated in the discrete case, i.e. for a sequence of $n \times n$ matrices with $n \rightarrow \infty$. Here the analogue of (4) is

$$(5) \quad N_n(\lambda) = \#\{\lambda_i \leq \lambda\} n^{-1}$$

In the case when a matrix is the restriction to a finite set $\Lambda \subset \mathbb{Z}^d$ of the Toeplitz operator the problem was considered by Grenander and Szegő [1].

In this paper I am going to discuss three classes of random differential and matrix operators for which the above formulated problem can be studied rather completely. The first class includes differential and finite-difference operators

with random coefficients. The second class includes random matrices with independent and identically distributed entries. Thus, the main difference between these two classes is that operators of the first class have nonzero entries only for a finite and Λ -independent number of diagonals adjacent to the principal one, while for operators (matrices) of the second class all entries have, roughly speaking, the same order of magnitude. This difference turns out to be rather serious and results in different forms of problems, answers and techniques. Therefore it is somewhat surprising that there exists operators that are in a certain sense interpolating between these two classes. This is our third class.

2. DIFFERENTIAL AND FINITE-DIFFERENCE OPERATORS WITH RANDOM COEFFICIENTS

It is clear that if we are going to prove the existence of the limit (3), we should impose certain conditions on the coefficients of the respective operators. Indeed, consider, for instance, the Schrödinger operator H_Λ defined in Λ by the operation

$$(6) \quad -\Delta + q(x)$$

and some (say Dirichlet) boundary conditions on $\partial\Lambda$. Then it is easy to see that if $q(x) \rightarrow \infty$ as $|x| \rightarrow \infty$, then $N(\lambda) \equiv 0$ and if $q(x) \rightarrow 0$ as $|x| \rightarrow \infty$ then $N(\lambda) = N_0(\lambda)$, where

$$(7) \quad N_0(\lambda) = c_d \lambda^{d/2}$$

is, according to (2), the integrated density of states of $-\Delta$. Thus, to obtain a nontrivial result (neither 0 nor $N_0(\lambda)$) we should consider a nonzero potential $q(x)$ that does not grow and does not decay at infinity. Moreover, it is clear that $q(x)$ should behave rather “regularly” at infinity not to produce too irregular oscillations of $N_\Lambda(\lambda)$ for large Λ .

The simplest nontrivial case is that of a periodic potential. The respective Schrödinger operator describes the electron motion in an ideal solid. A rather general class of potentials for which there exists the limit (3) is given by realizations (sample functions) of metrically transitive (ergodic) random fields in \mathbb{R}^d which model disordered solids. This means that we consider not a single

operator H_Λ with some particular $q(x)$ but rather a family $H_\Lambda(\omega)$ of operators corresponding to potentials of the form

$$(8) \quad q(x, \omega) = Q(T_x \omega),$$

where $Q(\omega)$, $\omega \in \Omega$, is a measurable function on a probability space Ω in which a group T_x , $x \in \mathbb{R}^d$, of measure-preserving and metrically transitive automorphisms acts. A simple example of (8) in the discrete case, i.e. for the finite-difference analogue of (6)

$$(9) \quad -\Delta_{\text{disc}} + q(x), \quad x \in \mathbb{Z}^d,$$

is a family of independent identically distributed (i.i.d.) random variables. This operator is a discretized model for electron motion in completely disordered solids, as amorphous substances, disordered alloys, etc.

The existence of the limit (3) with probability 1 and its nonrandomness for the Schrödinger operator with potentials having the form (8) and bounded below was given by myself in early seventies. Later this problem was considered by several authors and the results which are rather close to optimal were obtained (see, e.g., books [2,3] and references therein). For example, in the case of the operator (9) the IDS exists for any metrically transitive field (8) on \mathbb{Z}^d which is finite with probability 1.

Let us discuss now some properties of the IDS, restricting ourselves mainly to the case of the Schrödinger operators (6) and (9) (for the respective proofs and references see, e.g., books [2,3]).

(i) **Smoothness.** In the discrete case (i.e. for (9)) and in the continuous one-dimensional case (i.e. for (6) with $d = 1$) the IDS is always continuous. This can be proved on the basis of simple ergodic arguments. A more refined technique based on the notion of the Lyapunov exponent yields the log-Hölder property of the IDS:

$$|N(\lambda_1) - N(\lambda_2)| \leq \text{const} |\log |\lambda_1 - \lambda_2||^{-1}.$$

There are counter examples showing that this estimate is optimal for the whole class of metrically transitive potentials. However, if we consider the discrete operator (9) with i.i.d. random potential whose distribution function is

$F(dq) = f(q)dq$, $\sup_{q \in \mathbb{R}} f(q) \leq f_0 < \infty$, then, according to Wegner

$$N(d\lambda) = \rho(\lambda)d\lambda, \quad \sup_{\lambda \in \mathbb{R}} \rho(\lambda) \leq f_0.$$

(Here and below I denote by the same symbol $N(\cdot)$ the nondecreasing function and the measure that are related in the obvious way.)

Further, assuming some smoothness of $f(q)$, one can deduce much stronger smoothness of $\rho(\lambda)$, up to its C^∞ or real analyticity. On the other hand, it is known that if for $d = 1$ an i.i.d. random potential takes two values, say 0 and q_0 , then if q_0 is large enough the IDS has a singular continuous component [2,3].

Smoothness of the IDS, being of considerable interest in itself, plays an important role in the proofs of the Anderson localization, i.e. of the presence of a point component in the spectrum of operators (6) and (9).

It should be emphasized that most of the known results on the smoothness of the IDS and other spectral properties of the random operator are proved for the discrete case (9). The continuous case is much more technically difficult and much less studied so far.

(ii) **Explicit formulae.** The IDS is found explicitly for the 1-dimensional continuous case with the Markov random potential taking two values and for the multidimensional discrete case with the Cauchy-distributed or certain quasi-periodic potential. In the latter cases the IDS is an analytic function in some strip $|\operatorname{Im} \lambda| \leq \text{const}$. These formulae provide a considerable amount of quantitative information on the behaviour of the IDS on various parts of the spectrum.

(iii) **Asymptotic behaviour.**

(a) *High energies.* Let the potential $q(x)$ in (6) be a metrically transitive field on \mathbb{R}^d such that

$$\mathbb{E}\{|q(0)|^{p+1}\} < \infty,$$

where p is the smallest even number greater than $d/2$. Then

$$(10) \quad N(\lambda) = N_0(\lambda)(1 + o(1)), \quad \lambda \rightarrow \infty,$$

where $N_0(\lambda)$ is the IDS for $-\Delta$ and is given by (7). This is an analogue of the Weyl asymptotic (2) containing the leading term only. The naive perturbation

theory yields for the subsequent term in the one-dimensional case

$$(11) \quad N(\lambda) = N_0(\lambda) - (4\pi\lambda)^{-1} \int_0^\infty B(x) \cos 2\sqrt{\lambda}x \, dx.$$

Here we assume that $\mathbb{E}\{q(x)\} = 0$ and set $\mathbb{E}\{q(x)q(0)\} = B(x)$. This formula can be justified in many interesting cases. But the proof of the multidimensional analogue of (11) is still absent.

However, the problem of its justification seems rather important, because it is the simplest nontrivial case of the semi-classical asymptotics for the random Schrödinger operator. In the last decade physicists have found a lot of beautiful results by using various versions of the perturbation theory and the semi-classical approximation. These results are known as the weak localization theory. Their mathematical meaning is completely unexplored.

(b) *Low energies.* This is an asymptotic region that has no analogues in the conventional spectral theory but turns out to be rather rich and interesting for random operators. Let me mention here several typical results.

According to the quantum mechanical ideology, the low energy part of the spectrum should depend strongly on the particular potential. Nevertheless there are several rather well defined types of asymptotic behaviour of the IDS at low energies.

The simplest one corresponds to the unbounded potential. Consider, for instance, a potential whose probability distribution $F(dq)$ is such that

$$(12) \quad \lim_{q \rightarrow -\infty} \frac{\log F(q)}{\log F(q+a)} = 1, \quad \forall a \in \mathbb{R}.$$

Then for the operators (6) and (9) ,

$$(13) \quad \log N(\lambda) = \log F(\lambda)(1 + o(1)), \quad \lambda \rightarrow -\infty.$$

In the discrete case, i.e. for (9), the proof of (13) is fairly simple and uses in fact elementary variational arguments. In the continuous case, i.e. for (6), the proof of (13) is more complicated, although it also can be carried out in the variational terms. It was given by myself for the Gaussian random potential and for the so-called Poisson potential

$$(14) \quad q(x) = \sum_j u(x - x_j)$$

where $u(x)$ is a nonpositive function with a compact support and $\{x_j\}$ are the Poisson random points in \mathbb{R}^d .

It is believed that the asymptotic relation (13) is valid in the continuous case for a rather wide class of random potentials satisfying some sufficiently weak continuity condition.

Another type of asymptotic behaviour takes place for a random potential bounded from below, in particular for (14) with nonnegative $u(x)$. Here the spectrum is \mathbb{R}^+ and

$$(15) \quad \log N(\lambda) = -c_d n \lambda^{-d/2} (1 + o(1)), \quad \lambda \rightarrow 0_+,$$

where c_d depends only on d and n is the concentration of the Poisson points.

This asymptotic was suggested by I. Lifshitz and was proved by several authors including myself. The proof is based on the Wiener integral technique and an important ingredient of the proof are the deep results by Donsker and Varadhan on the large deviations for the Wiener process, or, in other words, on the infinite dimensional Laplace method.

It is natural to study the subsequent terms in the asymptotic formulae (13) and (15), in particular, the preexponential factor. This factor was found rigorously only in some one-dimensional cases.

Let me mention one more type of the asymptotic behaviour of the IDS. Consider the elliptic operator

$$(16) \quad - \sum_{i,j=1}^d \frac{\partial}{\partial x_i} \left(a_{ij}(x) \frac{\partial}{\partial x_j} \right),$$

where $a_{ij}(x)$ are random fields of the form (8). The spectrum of this operator is also \mathbb{R}^+ and we are interested in the behaviour of $N(\lambda)$ for $\lambda \rightarrow 0_+$. The change of the variables $\lambda = \varepsilon^2$, $x_i = \xi_i \varepsilon^{-1}$ and $\varepsilon^d N(\varepsilon^2) = \nu(\varepsilon)$ reduces our problem to the study of (16) in which $a_{ij}(x)$ are replaced by the fastly oscillating coefficients $a_{ij}(\xi/\varepsilon)$. This problem is the objective of the homogenization theory which has been developed in the last decades.

A rather general result which is suitable for the spectral theory has been proved by S. Kozlov. By using this result, one can prove that for the random

operator (16) (cf. (7) and (10))

$$N(\lambda) = c(a)\lambda^{d/2}(1 + o(1)), \quad \lambda \rightarrow 0_+,$$

where $c(a)$ is a rather complicated functional of a_{ij} 's. Its explicit form is known in the 1-dimensional case and some 2-dimensional examples.

3. RANDOM MATRICES

Let us consider the symmetric $n \times n$ random matrices $V^{(n)}$ with the entries

$$(17) \quad V_{ij}^{(n)} = n^{-1/2}W_{ij}, \quad i, j = 1, \dots, n$$

where W_{ij} are independent (except the symmetry condition $W_{ij} = W_{ji}$) random variables such that

$$(18) \quad \mathbb{E}\{W_{ij}\} = 0, \quad \mathbb{E}\{W_{ij}W_{kl}\} = v^2(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}).$$

The simplest and best known case of such W_{ij} 's corresponds to the Gaussian ones where their joint distribution can be written as

$$(19) \quad Z_n^{-1} \exp(-n \operatorname{Tr} W^2/4a^2) dW, \quad dW = \prod_1^n dW_{ii} \prod_{i < j} dW_{ij},$$

where Z_n is the normalization constant. This matrix ensemble was introduced by Wigner to describe some properties of the spectra of the energy levels of heavy nuclei.

There are many beautiful results for this matrix ensemble (see e.g. books [4,5] and references therein). In particular, if $\lambda_1, \dots, \lambda_n$ are eigenvalues of the respective random matrix and $N(\lambda)$ is specified by (5), then with probability 1,

$$(20) \quad \lim_{n \rightarrow \infty} N_n(\lambda) \equiv N(\lambda) = \int_{-\infty}^{\lambda} \rho(\mu) d\mu,$$

where

$$(21) \quad \rho(\lambda) = \frac{1}{2\pi v^2} \begin{cases} \sqrt{4v^2 - \lambda^2}, & |\lambda| \leq 2v \\ 0, & \text{otherwise.} \end{cases}$$

This is the well known Wigner or semicircle law. Wigner and several other authors have extended these results to more general cases. The optimal result proved by myself [6] is that if we replace in (20) convergence with probability 1 by convergence in probability, then (20) and (21) are valid under conditions that W_{ij} 's for $i < j$ are identically distributed and satisfy (18).

A somewhat more general form of this condition which is an analogue of the Lindeberg condition well known in the probability theory turns out to be necessary and sufficient [5,6]. Thus, the semicircle law (20) and (21) is as universal a form of the limiting eigenvalue distribution of random matrices with independent entries as the normal distribution is universal for the normalized sums of independent variables.

The latter results were obtained by constructing certain recursion relations in n for $n^{-1} \text{Tr}(V^{(n)} - z)^{-1}$, $\text{Im } z \neq 0$ and by studying their asymptotic behaviour for $n \rightarrow \infty$. This technique allow us to consider the more general matrix ensemble

$$(22) \quad H_{ij}^{(n)} = h_i \delta_{ij} + n^{-1/2} W_{ij}, \quad i, j = 1, \dots, n$$

with the same W_{ij} 's and h_i 's admitting the existence of the limit

$$(23) \quad \lim_{n \rightarrow \infty} \#\{h_i \leq \lambda\} n^{-1} \equiv N_0(\lambda)$$

for each λ which is a continuity point of $N_0(\lambda)$. This limit may be called the unperturbed integrated density of states. Consider the Stieltjes transform

$$r_o(z) = \int (\mu - z)^{-1} N_0(d\mu), \quad \text{Im } z \neq 0$$

of $N_0(d\mu)$. According to [6], if

$$r_n(z) = \int (\mu - z)^{-1} N_n(d\mu),$$

then the limit $r(z) = \lim_{n \rightarrow \infty} r_n(z)$ exists in probability for all $\text{Im } z \neq 0$ and can be found as a unique solution of the functional equation

$$(24) \quad r(z) = r_o(z + v^2 r(z))$$

in the class of functions, which are analytic for all $\text{Im } z \neq 0$ and such that $\text{Im } r(z) \cdot \text{Im } z > 0$, $\text{Im } z \neq 0$.

The IDS that corresponds to (24) is known as the deformed semicircle law. By using Eq. (24) one can prove the existence of the density of states, its boundedness, the location of the support of $\rho(\lambda)$, its asymptotic behaviour near end points of the support etc.

It is noteworthy that the above results for the ensemble (22) can be regarded as a limiting case of more general result obtained by Marchenko and myself (see [6]). Indeed, consider the following ensemble of $n \times n$ matrices

$$(25) \quad H^{(n)} = h^{(n)} + \sum_{i=1}^m \tau_i(\cdot, q_i) q_i,$$

where $h^{(n)}$ is an arbitrary matrix having the limiting IDS $N_0(d\lambda)$, τ_i and q_i are independent identically distributed random variables and unit vectors in \mathbb{R}^n respectively and the distribution of the latter is “close enough” to the uniform distribution. Then, if $n \rightarrow \infty$, $m \rightarrow \infty$, $m/n \rightarrow c < \infty$, there exists the limiting IDS $N(d\lambda)$ of the ensemble (25) and the Stieltjes transforms $r(z)$ and $r_o(z)$ of $N(d\lambda)$ and $N_0(d\lambda)$ are related as

$$(26) \quad r(z) = r_o \left(z + c \int_{-\infty}^{\infty} \tau(1 + \tau r(z))^{-1} \sigma(d\tau) \right),$$

where $\sigma(d\tau)$ is the distribution of τ_i . It is easy to see that if we redenote in (26) $z + c \mathbb{E}\{\tau\}$ by z and after that perform the limiting transition $c \rightarrow \infty$, $\mathbb{E}\{\tau^2\} \rightarrow 0$, $c \mathbb{E}\{\tau^2\} \rightarrow v^2$, then we arrive at (24).

One can regard (24) and (26) as a complete solution of the eigenvalue problem for the ensembles (22), (25) and (27).

As I have mentioned above, the initial physical motivation for these problems was provided by nuclear physics. In the last decades the random matrix ensembles have been used in many other branches of quantum physics. In particular, the problem of the study of the quantum kicked rotator that is an archetype model in quantum chaology is related to the study of the eigenvalues distribution of random matrices

$$(27) \quad V_{ij}^{(n,b)} = (2b_n + 1)^{-1/2} \phi((i - j)/b_n) W_{ij}.$$

Here $b_n \rightarrow \infty$ as $n \rightarrow \infty$ and $\phi(t) = \phi(-t) \in \mathbb{R}$ is a piecewise continuous and bounded function with a compact support and

$$(28) \quad \int \phi^2(t) dt = 1.$$

In particular, if $\phi(t) \equiv 1$ and $2b_n + 1 = n$, we obtain (17) and if $\phi(t)$ is the indicator of the interval $(-1, 1)$, then (27), (28) defines the so-called band matrices that have nonzero i.i.d. entries only inside the “band” of a width $2b_n + 1$ around the principal diagonal of $V^{(n,b)}$.

The IDS of this matrix ensemble have been found recently by several groups of authors [8-10]. The most general result was proved in [10]. It asserts that the IDS of (27) is given by the semicircle law (20), (21) if $\lim_{n \rightarrow \infty} b_n n^{-1}$ equals 0 or $1/2$. If, however, $0 < \lim_{n \rightarrow \infty} b_n n^{-1} < 1/2$, then the limiting eigenvalue distribution exists but it is not the semicircle law. Its Stieltjes transform can be found as a unique solution of some non-linear integral equation. The situation can be “corrected” in a sense that the semicircle law can still be obtained if we consider a certain periodic function $\phi(t)$ in (27).

The proof of these results is based on a certain new approach. It consists in deriving an infinite system of linear equations for the moments of the diagonal matrix elements of the resolvent $(V^{(n,b)} - z)^{-1}$ and in asymptotic solution of this system for b_n , $n \rightarrow \infty$. This method proves to be rather efficient and general and allows us to consider a wide variety of related problems (see review [11]). In particular it can be proved that the deformed semicircle law (24) is valid for any not necessarily diagonal “unperturbed” matrix $h^{(n)} = \{h_{ij}\}_{i,j=1}^n$ in (22) for which the IDS exists, i.e. if μ_1, \dots, μ_n are the eigenvalues of $h^{(n)}$, then we only need to assume that the limit (5) exists.

In the next section we consider other problems that can be solved by the same approach.

I have discussed matrix ensembles that share with the Gaussian ensemble (19) the property of statistical independence of all functionally independent entries. All these ensembles have the same IDS described by the semicircle law (21). There is another generalization of (19) which is defined by the probability

distribution of the form

$$(29) \quad Z_n^{-1} \exp(-n \operatorname{Tr} V(W)) dW,$$

where $V(t)$ is a real-valued function that grows at infinity faster than $a \log |t| \forall a > 0$. The Gaussian ensemble (19) corresponds obviously to $V(t) = t^2/4a^2$. Polynomials of an even degree $p > 2$ appear in quantum field theory, statistical mechanics of random surfaces, combinatorics, etc. (see [7] and references therein).

The IDS of these ensembles differs from the semicircle law. Generically its support consists of $p/2$ intervals with square root zeros of the density of states at each endpoint. However, by varying $V(t)$ (e.g. the coefficients in a polynomial $V(t)$), one can obtain a variety of degenerated cases with smaller number of intervals and other behaviour of $\rho(\lambda)$ at their end points. For instance, if $V(t) = |t|^\alpha, \alpha > 1$, then

$$(30) \quad \rho(\lambda) = v_\alpha(\lambda B_\alpha^{-1}) B_\alpha^{-1},$$

where

$$v_\alpha(t) = v_\alpha(-t) = \frac{2\alpha}{\pi} \begin{cases} \int_{|t|}^1 \tau^{\alpha-1} (\tau^2 - t^2)^{-1/2} d\tau, & |t| \leq 1 \\ 0, & |t| > 1 \end{cases}$$

$$(31) \quad B_\alpha = \left(2\alpha\pi^{-1} \int_0^1 t^\alpha (1 - t^2)^{-1/2} dt \right)^{1/\alpha}.$$

There is a beautiful approach to the study of the ensembles (29). It is based on the possibility of expressing the expectation of any unitary invariant of a random matrix belonging to the ensemble (29) via polynomials orthogonal on \mathbb{R} with respect to the weight $\exp(-nV(\lambda))$ [4,7]. Physicists have found a lot of very interesting properties of the eigenvalue distribution of the ensemble (29) by using this approach. However, rigorous proofs are not numerous here (see however paper [15] in which the rigorous derivation of the form of the density of states for some class of even polynomials in (29) is given). One of the reasons is that rigorous study should be based on precise asymptotic formulae for the respective

orthogonal polynomials. The best recent result is for the case of $V(t) = |t|^\alpha$ only [13]. The corresponding asymptotic formulae allow us to obtain rigorously the form (30), (31) of the density of states of the ensemble (29) with the same $V(t)$ [14]. However, these formulae are not precise enough to calculate other quite important characteristics of random matrices. Besides, almost nothing is known rigorously in case of a nonmonomial $V(t)$, especially for a nonconvex one, e.g. for $V(t) = at^2 + bt^4$ with $a < 0$ and $b > 0$. The corresponding sufficiently precise asymptotic formulae for the orthogonal polynomials, which should be in many respect analogous to semiclassical formulae of quantum mechanics, would be of great use and importance in spectral theory.

In the last three years a new wave of activity in this field has been initiated by the progress in the study of models of two-dimensional gravity and the string theory [12]. This progress being translated into the random matrix theory language consists in establishment of the form of the behaviour of the mean prelimit density of states $\rho_n(\lambda) = \mathbb{E}\{n^{-1} \sum_1^n \delta(\lambda - \lambda_i)\}$ of the ensemble (29) with the special polynomials $V(\lambda)$ for $n \rightarrow \infty$ and λ tending simultaneously to an end point of the support of the limiting density of states $\rho(\lambda) = \lim_{n \rightarrow \infty} \rho_n(\lambda)$. These studies revealed many beautiful connections of the random matrix theory with the theory of integrable systems, the spectral theory of Jacobi matrices, the theory of orthogonal polynomials and have raised a lot of problems that are of great interest for both mathematical and theoretical physics.

4. "INTERPOLATING" FAMILIES OF RANDOM OPERATORS

The two classes of random operators that were considered in preceding sections have different origin and have been studied independently and by rather different techniques. Therefore it seems rather natural to look for interconnections between these classes. In the present section we are going to consider three families of random operators belonging to the first class. However, these families depend on certain parameters (we call them the interaction radius, the dimensionality of space and the number of components, respectively) in such a way that the limits of the IDS of the corresponding random operators for infinite

values of these parameters coincide with the deformed semicircle law which is typical for the second class.

Our first family consists of the operators $H^{(R)} = h + V^{(R)}$ acting in $l^2(\mathbb{Z}^d)$ and defined by the matrices

$$(32) \quad H^{(R)}(x, y) = h(x - y) + R^{-d/2} \phi((x - y)R^{-1})W(x, y), \quad x, y \in \mathbb{Z}^d,$$

where $h(-x) = \overline{h(x)}$, $h(x) \in l^1(\mathbb{Z}^d)$, $R < \infty$, $\phi(t)$, $t \in \mathbb{R}^d$, is a piecewise continuous function with compact support and $W(x, y)$ are independent (except the symmetry condition $W(x, y) = W(y, x)$) and identically distributed random variables such that (cf. (18))

$$(33) \quad \mathbb{E}\{W(x, y)\} = 0, \quad \mathbb{E}\{W(x, y)W(s, t)\} = v^2(\delta_{xs}\delta_{yt} + \delta_{xt}\delta_{ys}).$$

The matrix (32) defines the d -dimensional finite difference operator. In particular, if $h(x) = 0$, $|x| \neq 1$, $R = 1$ and $\phi(t) = 0$ for $|t| > 1$, then this operator is of the second order. However, it differs from the discrete Schrödinger operator (6) in that it has random off-diagonal entries.

The random operator $H^{(R)}$, defined by (32), is a metrically transitive operator in the terminology of book [3]. In particular, it admits the nonrandom IDS $N^{(R)}(\lambda)$ defined by (3) and (4) in which $H_\Lambda^{(R)}$ is a restriction of $H^{(R)}$ to a finite “box” $\Lambda \in \mathbb{Z}^d$. We have proved recently [16] that there exists the limit $\lim_{R \rightarrow \infty} N^{(R)}(\lambda)$ and that its Stieltjes transform satisfies equation (24), where $r_o(z)$ is the same transform of the IDS of the Toeplitz operator $h(x - y)$ and can be easily expressed via the symbol of this operator [1,3]. Thus, $\lim_{R \rightarrow \infty} N^{(R)}(\lambda)$ coincides with the deformed semicircle law. In particular, if the nonrandom part of (32) is absent (i.e. $h(x) \equiv 0$), then this limit coincides with the semicircle law (20), (21).

Our second family $H^{(d)}$ of random operators acts also in $l^2(\mathbb{Z}^d)$ and is defined by the random matrix

$$(34) \quad H^{(d)}(x, y) = h_d(x - y) + d^{-1/2} \delta(|x - y| - 1)W(x, y),$$

where

$$(35) \quad h_d(x) = d^{-1/2} \sum_{i=1}^d h_1(x_i - y_i) \prod_{k \neq i} \delta(x_k - y_k),$$

$h_1(x) \in l_1(\mathbb{Z})$ and the random variables $W(x, y)$ are the same as in (33). This operator also belongs to the class of metrically transitive operators, in particular it admits the IDS $N^{(d)}(\lambda)$. Now, sending d to infinity, we again obtain the limiting eigenvalue distribution coinciding with the deformed semicircle law. The form (35) of the unperturbed part of (34) is more special than that in (32). This special form is needed to guarantee the existence of the $d = \infty$ limit for the unperturbed IDS which in this case turns out to be of the Gaussian form.

The third family of operators $H^{(n)}$ acts in $l^2(\mathbb{Z}^d) \otimes \mathbb{C}_n$ and is defined by the random “block” matrix

$$(36) \quad H^{(n)}(\alpha, x; \beta, y) = h(x - y)\delta_{\alpha\beta} + n^{-1/2}\delta(x - y)W_{\alpha\beta}(x),$$

where $\alpha, \beta = 1, \dots, n$, $x, y \in \mathbb{Z}^d$, $h(x)$ is the same as in (32) and $W_{\alpha\beta}(x)$ are independent (except the symmetry condition $W_{\alpha\beta}(x) = W_{\beta\alpha}(x)$) identically distributed random variables such that (cf. (18) and (32))

$$\mathbb{E}\{W_{\alpha\beta}(x)\} = 0, \quad \mathbb{E}\{W_{\alpha\beta}(x)W_{\gamma\delta}(y)\} = v^2\delta(x - y)(\delta_{\alpha\gamma}\delta_{\beta\delta} + \delta_{\alpha\delta}\delta_{\beta\gamma}).$$

The IDS $N^{(n)}(\lambda)$ of $H^{(n)}$ is defined as the limit of $(n|\Lambda|)^{-1}\mathcal{N}_{\Lambda, n}(\lambda)$, where $\mathcal{N}_{\Lambda, n}(\lambda)$ is the counting function of the eigenvalues of the restriction $H_{\Lambda}^{(n)}$ of (36) to the finite domain $\Lambda \in \mathbb{Z}^d$.

Here we send to infinity the order n of the blocks. The resulting IDS is again the deformed semicircle law.

Operators $H^{(n)}$ were introduced by Wegner [17], who obtained the latter result at the theoretical physics level of rigour by using the perturbation theory with respect to the random part of (36).

Thus we see once again that the deformed semicircle law is a rather general and universal form of the limiting eigenvalue distribution.

The results formulated in this section were proved [16] by using the general approach that was briefly outlined in the previous section. The main technical result here is the statement that in the limit $R, d, n \rightarrow \infty$ the moments $\mathbb{E}\{\prod_{i=1}^l G(x_i, y_i; z_i)\}$ take the form $\prod_{i=1}^l \mathbb{E}\{G(x_i, y_i; z_i)\}$ for any fixed $l \geq 1$, $x_i, y_i \in \mathbb{Z}^d$ and $|\text{Im } z_i| \geq \eta, i = 1, \dots, l$ with sufficiently large η . Here $G(x, y; z) =$

$(H^{(a)} - z)^{-1}(x, y)$ for $a = R, d$ and $G(x, y; z) = n^{-1} \sum_{\alpha=1}^n (H^{(n)} - z)^{-1}(\alpha, x; \alpha, y)$ for $a = n$.

We prove this result by constructing a certain infinite system of equations for this family of moments and by proving that this system admits a unique solution that can be written in the factorized form in the limits $R, d, n = \infty$. The function $\mathcal{G}(x - y; z) = \lim_{R, d, n \rightarrow \infty} \mathbb{E}\{G(x, y; z)\}$ that determinates this form can be found as a unique solution of a certain nonlinear integral equation. Its solubility condition yields (24).

The reader familiar with statistical mechanics will find a close analogy of the results described above with the limit of the infinite interaction radius or the limit of the infinite dimensionality of space which are widely used in statistical mechanics giving a rigorous form of the mean field approximation. Thus, our limiting results can be regarded as analogues of the respective results in statistical mechanics.

REFERENCES

- [1] U. GRENANDER, G. SZEGÖ, *Toeplitz form and their applications*, Univ. Calif. Press, Berkeley, 1958.
- [2] R. CARMONA, J. LACROIX, *Spectral theory of the random Schrödinger operator*, Birkhäuser, Basel, 1990.
- [3] L.A. PASTUR, A.L. FIGOTIN, *Spectra of Random and Almost Periodic Operators*, Grundlehren der Math. Wiss., Springer, Heidelberg, 1992.
- [4] M. L. MEHTA, *Random matrices*, Academic Press, N. Y., 1967.
- [5] V. GIRKO, *Spectral theory of random matrices*, Nauka, Moscow, 1988.
- [6] L.A. PASTUR, Spectra of Random Self-Adjoint Operators, *Russ. Math. Surveys*, **28**, 1 (1973), 1-67.
- [7] D. BESSIS, C. ITZYKSON, J. ZUBER, Quantum field theory techniques, *Adv. App. Math.*, **1**, 2 (1980), 109-157.
- [8] M. KUS, M. LEWENSTEIN, F. HAAKE, Density of eigenvalues pf random band matrices, *Phys. Review*, **A44** (1991), 2800-2808.
- [9] L. BOGACHEV, S. MOLCHANOV, L. PASTUR, *Teor. Matem. Fyz.*, **90** (1991), 31-42 (in Russian).

- [10] A. KHORUNZHY, S. MOLCHANOV, L. PASTUR, *Teor. Matem. Fyz.*, **92**, 2, (1992) (in Russian).
- [11] A. KHORUNZHY, B. KHORUZHENKO, L. PASTUR, M. SHCHERBINA, The Large- n Limit in Statistical Mechanics and Spectral Theory of Disordered Systems. In “*Phase Transitions and Critical Phenomena*”, eds. C.Domb and J.L.Lebowitz, **15**, 73, Academic Press, London, 1992.
- [12] E. BREZIN, V. KAZAKOV, Exactly solvable field theories of closed strings, *Phys. Lett.*, **B236**, 2 (1990), 144-150; M. DOUGLAS, S. SHENKER, *Nucl. Phys.*, **B335** (1990), 635; D. GROSS, A. MIGDAL, Nonperturbative solution of the Ising model on a random surface, *Phys. Rev. Lett.*, **64**, 7 (1990), 717-720.
- [13] D. S. LUBINSKY, *Strong asymptotics for extremal errors and polynomials associated with Erdős-type weights*, Pitman Research Notes in Math., **202**, Longmans, Harlow, 1989.
- [14] L. PASTUR (in press).
- [15] L. PASTUR, V. TKACHENKO (in preparation).
- [16] A. KHORUNZHY, L. PASTUR (in preparation).
- [17] F. WEGNER, Disordered system with n orbitals per site: $n = \infty$ limit, *Phys. Rev.*, **B19** (1979), 783-792.

Leonid PASTUR

*Math. Division, Institute for Low Temperature Physics
Ukrainian Academy of Sciences
310164, Kharkov, Ukraine*

Astérisque

HAROLD ROSENBERG

**Some recent developments in the theory of properly
embedded minimal surfaces in \mathbb{R}^3**

Astérisque, tome 206 (1992), Séminaire Bourbaki,
exp. n° 759, p. 463-535

http://www.numdam.org/item?id=SB_1991-1992__34__463_0

© Société mathématique de France, 1992, tous droits réservés.

L'accès aux archives de la collection « Astérisque » (<http://smf4.emath.fr/Publications/Asterisque/>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

**SOME RECENT DEVELOPMENTS IN THE THEORY OF
PROPERLY EMBEDDED MINIMAL SURFACES IN \mathbb{R}^3**

by **Harold ROSENBERG**

In the past decade there has been considerable progress in our understanding of minimal surfaces in three dimensional manifolds. In this seminar I would like to discuss a small part of the work that has been done concerning properly embedded minimal surfaces (which I will refer to as m -surfaces) in \mathbb{R}^3 .

Until 1982, the only examples of such surfaces we knew were periodic minimal surfaces and the catenoid and plane, and they came to us from the last century : the helicoid, Scherk's surfaces, Riemann's surface, Schwarz's surfaces, etc. An m -surface is periodic if it is invariant by a non trivial discrete group of isometries acting freely on \mathbb{R}^3 . Our surfaces are always assumed connected unless stated otherwise. We denote by $C(M)$ the total curvature of M : $C(M) = \int_M K$, K the gaussian curvature of M .

In 1982, C. Costa wrote down the formulae for a complete minimal surface, modelled on a 3-punctured torus, of $C(M) = -12\pi$, which he believed was embedded [Cost.-1,2]. D. Hoffman and W. Meeks *looked at* the surface on a computer and with the aid of the symmetries they detected, they proved the Costa surface is embedded (James Hoffman did the graphics). Subsequently families of finite total curvature m -surfaces have been constructed [H.-M.-2], figures 1 and 2.

All the examples we know today, of m -surfaces in \mathbb{R}^3 , are periodic or of finite total curvature. One of the important open problems is to decide if there are other examples.

We will discuss some of the main results concerning m -surfaces of finite total curvature : the theorems of R. Schoen [Sch.-1] and Lopez-Ros [Lo.-Ros]; each theorem is a characterization of the catenoid among m -surfaces of finite total curvature. Schoen's theorem assumes exactly two ends and the Lopez-Ros theorem assumes genus zero.

We discuss the curvature estimates of stable minimal surfaces, initiated by Heinz for graphs and in general by R. Schoen. We show how the curvature estimates are used to construct stable limits of least area surfaces, and we give applications.

We discuss the annular end theorem and the strong halfspace theorem of Hoffman-Meeks. This latter result says that two properly immersed disjoint minimal surfaces in \mathbb{R}^3 are planes; this is very useful.

We discuss the work of Meeks and myself on the finite total curvature conjecture : an m -surface in \mathbb{R}^3 of finite topology and at least two ends is of finite total curvature. A corollary of our work is that such a surface is of finite conformal type.

We discuss the work of Meeks and myself on periodic minimal surfaces. The main result is that finite topology of the quotient surface implies finite total curvature of this quotient surface. If this is so then the (quotient) surface is parametrized by meromorphic data on a compact Riemann surface (a Weierstrass type representation).

This theorem yields topological and geometrical obstructions for the existence of such surfaces. For example, the number of ends of such surfaces is always at least two (except for the plane). If the surface is doubly periodic and orientable (in the quotient) then the number of ends is at least four. These are topological obstructions, we will discuss geometrical obstructions in section VII. For example, if all the ends are not parallel (as in Scherk's doubly periodic surface) then the group G is commensurable. This means there are two independent elements of G of the same length.

We prove the plane and the helicoid are the only simply connected m -surfaces in \mathbb{R}^3 with an infinite symmetry group.

We discuss the sum of minimal surfaces and some applications.

There is no known topological obstruction to realizing a complete, orientable, non compact surface as an m -surface in \mathbb{R}^3 .

Finally we discuss some problems, conjectures, and related results.

I have decided not to discuss the construction of the beautiful examples of Costa, Karcher, Hoffman and Meeks. Their influence on this subject has been enormous, and as H. Karcher says : “What a magnificent picture of a conformal map.” I would like to thank David Hoffman and Hermann Karcher for their work and inspiration. Some of the computer graphics were done at the Geometry, Analysis, Graphics Laboratory at the University of Massachusetts at Amherst by Jim Hoffman, Ed Thayer and Fusheng Wei. The remaining computer graphics were done by Hermann Karcher and Konrad Polthier working with SFB256 at Bonn. I thank you all. I received a great deal of help with the material preparation of this manuscript by Hermann Karcher and Katrin Wendland. I thank you both.

The paper is organized as follows.

1. How the classical examples are constructed.
2. The Weierstrass representation and the geometry of the ends of a finite total curvature minimal surface in \mathbb{R}^3 .
 - 2.1 Osserman’s parametrization of finite total curvature surfaces
 - 2.2 The geometry of finite total curvature ends
3. The characterizations of the catenoid by R. Schoen and Lopez–Ros.
 - 3.1 The theorem of R. Schoen
 - 3.2 The theorem of Lopez-Ros
 - 3.3 and 3.4 The maximum principle at infinity
 - 3.5 The monotonicity formula
4. Curvature estimates for stable minimal surfaces.
 - 4.1 The Barbosa-Do Carmo stability criteria
 - 4.2 An idea of the proof of Heinz’s theorem
5. Compactness of least area families and construction of complementary finite total curvature surfaces.

6. The annular end theorem and the strong halfspace theorem of Hoffman-Meeks.

6.1 The annular end theorem

6.2, 6.3 and 6.4 The finite conformal type theorem and its corollaries

6.5 The strong halfspace theorem

7. Doubly periodic minimal surfaces.

7.1, 7.2 and 7.3 The finite total curvature theorem for doubly periodic minimal surfaces

7.4 The total curvature formula

7.5, 7.6 and 7.7 Global topological and geometrical properties

7.8, 7.9, 7.10 and 7.11 The sum of minimal surfaces and applications

8. Singly periodic minimal surfaces.

8.1 The finite total curvature theorem for singly periodic minimal surfaces

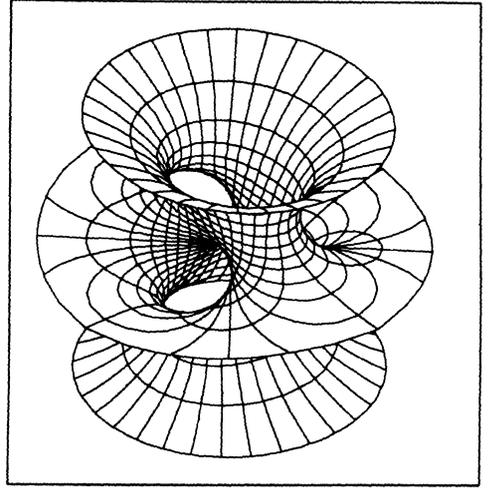
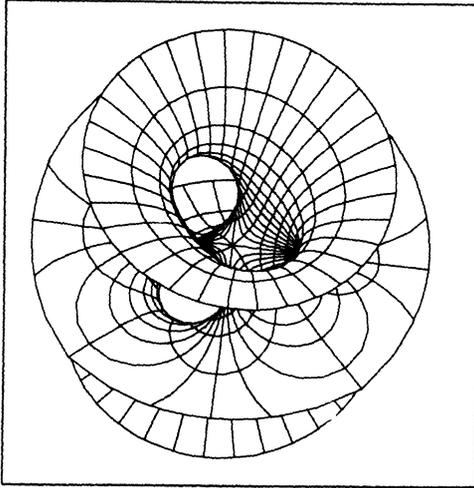
8.2, 8.3 and 8.4 The generalized Weierstrass representation

8.5 The geometry of finite total curvature ends

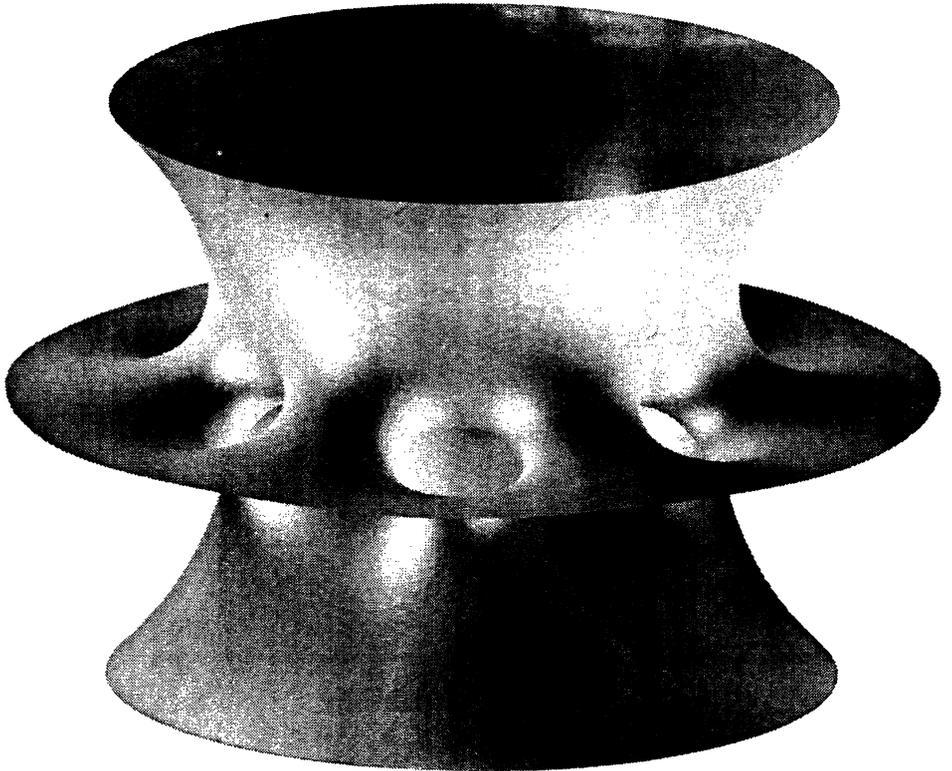
8.6 The winding number of an end

8.7 - 8.12 Applications of the finite total curvature theorem

9. Some problems, conjectures, and related results.



1



2

1. HOW THE CLASSICAL EXAMPLES ARE CONSTRUCTED

Let Γ be a polygonal Jordan curve in \mathbb{R}^3 and M_0 a compact minimal surface with $\partial M_0 = \Gamma$. The Schwarz reflection principle [Oss.-1] allows us to extend M_0 across each edge of Γ by rotating M_0 by π about each edge. Continuing the reflections along each edge that develops one obtains a complete minimal surface M , which will have singularities in general and self intersections; i.e. M will be immersed.

If the angle at which two edges of Γ meet is irrational the M will turn infinitely often about this vertex. So if one wants embedded examples then Γ should be chosen to have vertex angles of the form $2\pi/n$, and M_0 should be chosen embedded. For example the polygons of figures 3 and 4 provide properly embedded examples which are invariant by 3-independent translations (trily periodic examples).

Riemann, Schwarz and Weierstrass found minimal surfaces M_0 with $\partial M_0 = \Gamma$ by explicitly solving the Riemann mapping problem and the Weierstrass representation (this is well explained in Darboux [Darb.]). Today we find M_0 by other techniques. Douglas and Rado proved that any rectifiable Jordan curve Γ in \mathbb{R}^n bounds a least area minimal disc M_0 , [Doug.],[Rado-1,2]. Subsequently, R. Osserman proved M_0 had no geometric branch points (i.e. a least area disc with boundary Γ is immersed [Oss.-1]). Finally using geometric measure theory, Reifenberg proved there is always an embedded minimal surface M_0 with $\partial M_0 = \Gamma$, [Reif.].

Now if one chooses Γ and M_0 well, the complete surface M obtained by the reflections of M_0 in all edges (that develop) will be an embedded triply periodic surface. The quotient of M by a group G generated by 3-independent translations will be a compact minimal surface of finite genus embedded in the flat 3-torus \mathbb{R}^3/G . The geometry and topology of these surfaces has been studied by W. Meeks [M.-3], and H. Karcher [K.-3],[K.-4].

Many of the other known examples of infinite total curvature (doubly and singly periodic examples) are constructed by taking Γ to be a non compact polygon and M_0 a complete embedded minimal surface with boundary Γ . Again one does all possible reflections of M_0 across the edges of Γ (and its iterates) to construct a complete minimal surface M . So how does one

find M_0 when Γ is infinite? There is a general theory which attacks this problem (the Jenkins–Serrin theorem [J.-S.] and the conjugate Plateau construction [K.-1]) but rather than discuss this. I will describe how one can obtain some examples directly.

First, let us construct Scherk's (first) surface by solving a compact Plateau problem and taking limits. Consider the polygon $\Gamma(n)$ of figure 5-a. For each integer n , choose $\Gamma(n)$, as in figure 5-a, so that

- $\Gamma(n)$ projects to a square in the horizontal plane, and
- the top edges are at height n , and the bottom edges at height $-n$.

Now let $\Sigma(n)$ be the least area disc with boundary $\Gamma(n)$. It is not hard to prove, that $\Sigma(n)$ is a graph over the square in the horizontal plane to which $\Gamma(n)$ projects. (More generally, Rado has proved that if a Jordan curve Γ projects to a convex planar curve C , then any minimal surface bounded by Γ is a graph over the planar domain bounded by C [M.-2]).

Now $\Sigma(n)$ inherits the symmetries of $\Gamma(n)$ so there is a point p_n of $\Sigma(n)$ at vertical height zero where the tangent plane of $\Sigma(n)$ is horizontal.

Now as $n \rightarrow \infty$, the surfaces $\Sigma(n)$ all pass through the same point $p_n = p$. Then the functions defining the graphs $\Sigma(n)$ converge to a function f , defined on the interior of the square, with boundary values $+\infty$ on two opposite sides of the square and $-\infty$ on the two other sides. The graph of f is a minimal surface with boundary the four vertical lines over the vertices of the square, figure 5-b.

Now do Schwarz reflection of the graph of f about the four vertical lines, and about all the vertical lines one obtains. This yields Scherk's minimal surface M . M projects to the infinite array of squares in the horizontal plane, which form the (black squares say) of an infinite checkerboard pattern, figure 5-c.

Now one can form quotients of M by independent horizontal translations to obtain properly embedded, finite topology (and finite total curvature) surfaces in flat manifolds $\mathbb{T}^2 \times \mathbb{R}$, \mathbb{T}^2 a flat 2-torus.

The simplest way to do this yields a projective plane punctured in two points. We now describe some of these examples.

Let P be the square to which $\Gamma(n)$ projects and let v_1, v_2 be the vectors

determined by the sides of P . Let $G(v_1 + v_2, v_1 - v_2)$ denote the group generated by the translations $v_1 + v_2, v_1 - v_2$. Then $G(v_1 + v_2, v_1 - v_2)$ leaves M invariant and the quotient is topologically a projective plane minus two points, of total curvature -2π , figure 6-a.

A fundamental domain for $G(2v_1, 2v_2)$ is two contiguous copies of P (figure 6-b) and the quotient of M by this group is conformally diffeomorphic to a 4-punctured sphere and is of total curvature -4π in $T^2 \times \mathbb{R}$.

One can realize S^2 minus any even number of points this way : let $G = G(2nv_1, v_2)$, a fundamental domain consists of $2n$ copies of P (figure 6-c). To obtain a torus minus four points, let $G = G(2(v_1 + v_2), 2(v_1 - v_2))$. A fundamental domain is four copies of P (figure 6-d). The total curvature is -8π and there are four ends.

One obtains the Klein bottle from the group $G(v_1 - v_2, 2(v_1 + v_2))$. A fundamental domain is given in figure 6-e. There are two ends and the total curvature is -4π . Placing n copies of P diagonally and letting $G = G(v_1 - v_2, n(v_1 + v_2))$ we obtain the connected sum of n projective planes minus two points. By taking appropriate oriented two-sheeted covers of the nonorientable examples just described one obtains every possible orientable surface minus four points.

Notice that in all these examples, the ends are asymptotic to flat cylinders, which happen to be vertical. Also the top ends are not parallel to the bottom ends here. There are examples with all the ends parallel and non vertical [M.-R.1]. H. Karcher has constructed, an easy to visualize example of a torus minus four points in $\mathbb{T} \times \mathbb{R}$, with all the ends parallel; we call this a Karcher saddle, [K.-2].

In Karcher's example one has a rectangle P and a minimal graph over the part of P bounded by L_1, L_2, C_1 and C_2 (figure 7-a). The function is 0 on C_1, C_2 and $+\infty$ on L_1, L_2 . Moreover, the graph is vertical along $C_1 \cup C_2$. This implies C_1, C_2 are planar lines of curvature and the graph can be extended by reflection in the plane of P . This new surface has four vertical lines as boundary (figure 7-b).

To obtain a torus minus four points, place two copies of P diagonally and quotient by the group $G(2v_1, 2v_2)$, figure 7-c.

In all of the above examples, the geometry and topology of the quotient

surfaces are related by the formula :

$$C(M) = 2\pi\mathcal{X}(M) .$$

This is a special case of the result :

THEOREM 1.1 [M.-R.1]. — *Let $M \subset T \times \mathbb{R}$ be a properly embedded minimal surface of finite topology. Then M has finite total curvature and*

$$C(M) = 2\pi\mathcal{X}(M) .$$

In our construction of Scherk's surface we started with a square P over which we took limits of minimal graphs. If we started with a rhombus P the same construction works; the graphs with boundary the polygons $\Gamma(n)$ would still have their vertical point p_n at height zero. So Scherk's surfaces exist over checkerboard patterns defined by rhombi. However, had we started with a parallelogram P with sides of unequal length, the points p_n will always *drift off* to infinity and the limiting surface will be two disjoint vertical strips (figure 8).

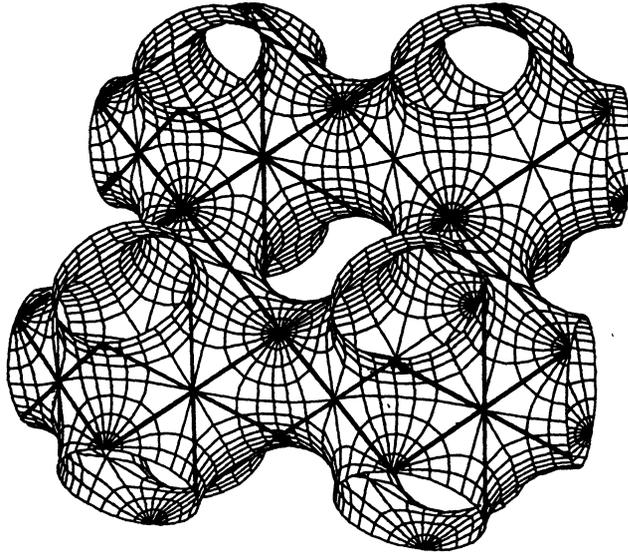
This a special case of our result :

THEOREM 1.2 [M.-R.-1]. — *Let M be a properly embedded minimal surface in $\mathbb{T} \times \mathbb{R}$ of finite topology. If the ends of M are not parallel then $\mathbb{T} \times \mathbb{R}$ has a commensurable lattice and the ends of M are vertical.*

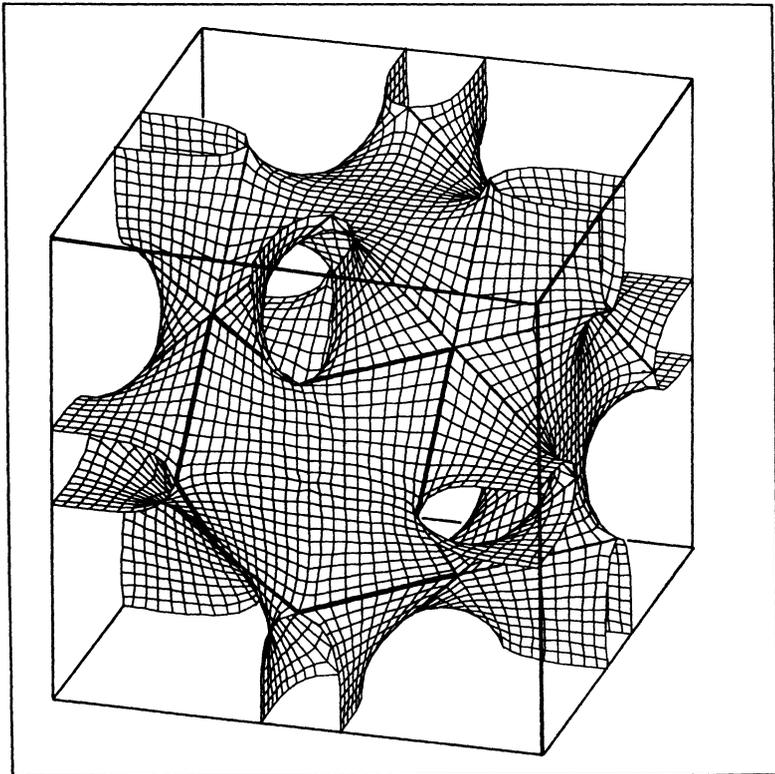
By commensurable lattice we mean $\mathbb{T} \times \mathbb{R} = \mathbb{R}^3/G$, and G has two linearly independent vectors of equal length.

There is a theorem of Jenkins and Serrin which yields Scherk's surface over a rhombus (hence the complete surface by reflection in the vertical lines over the vertices). We state a special case of their result.

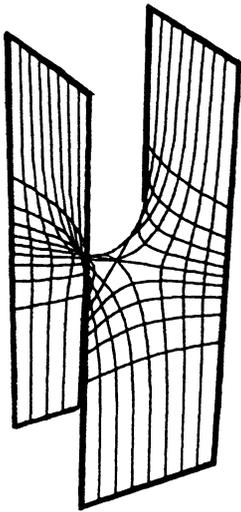
THEOREM 1.3 [J.-S.]. — *Let C be a polygonal Jordan curve in the plane with an even number of sides. Let P be the compact planar domain bounded by C and let φ be the data on C which is $+\infty$ and $-\infty$ on adjacent sides of C . A necessary and sufficient condition that φ extend to a (finite valued) minimal graph over P , is the sum of the lengths of the edges of C where φ is $+\infty$, equals the sum of the lengths of the edges of C where φ is $-\infty$.*



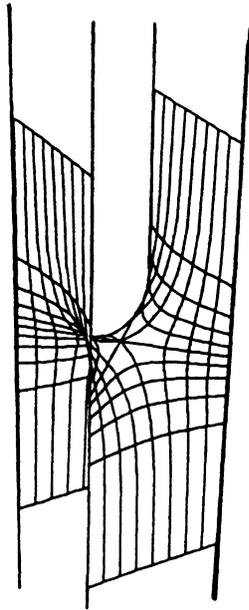
3



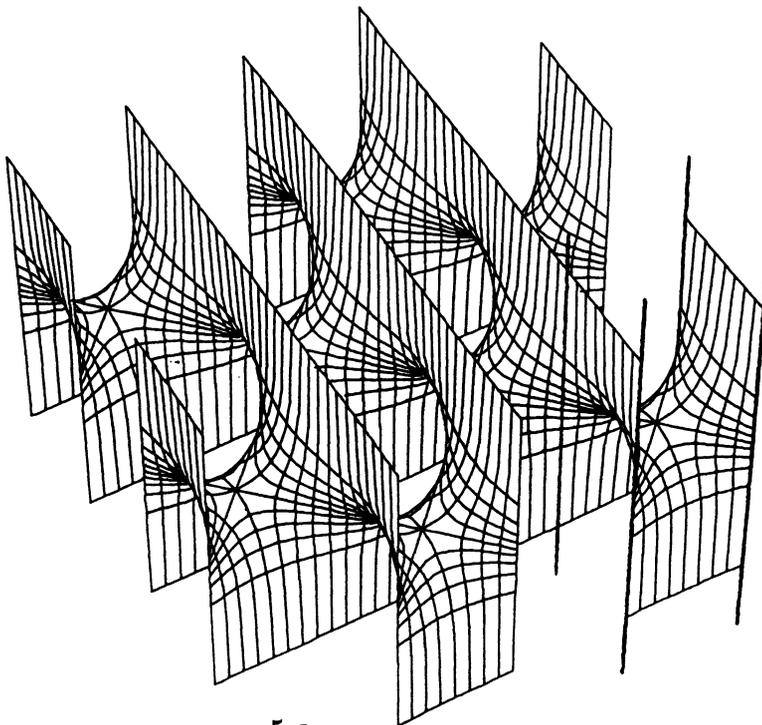
4



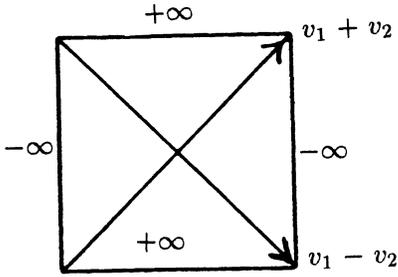
5.a



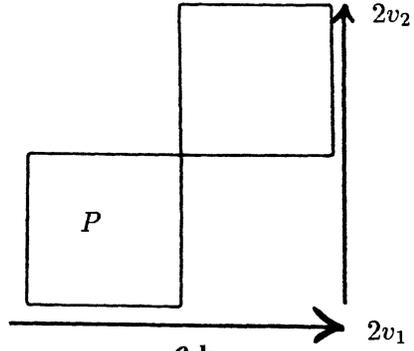
5.b



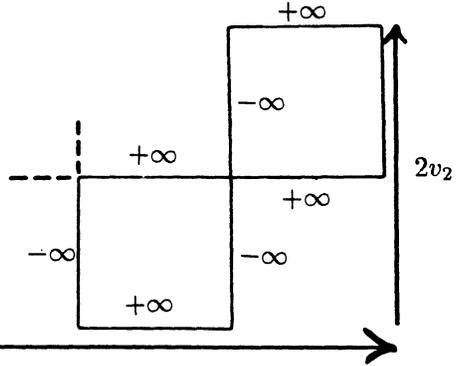
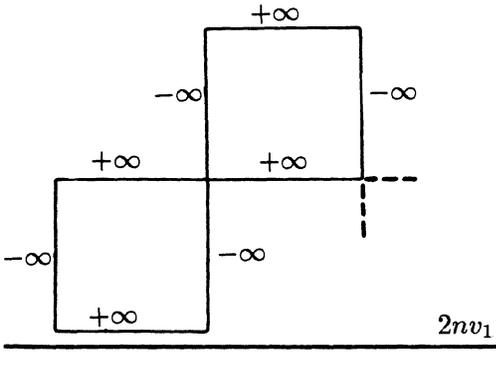
5.c



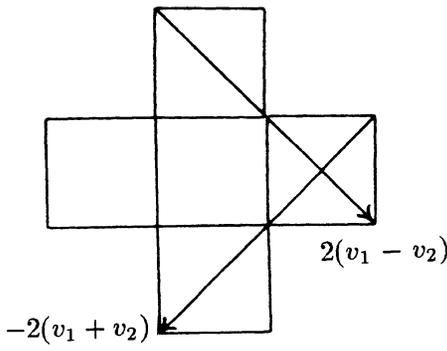
6.a



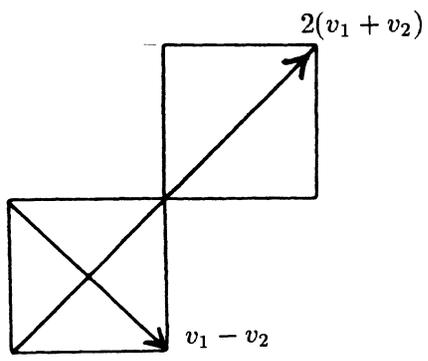
6.b



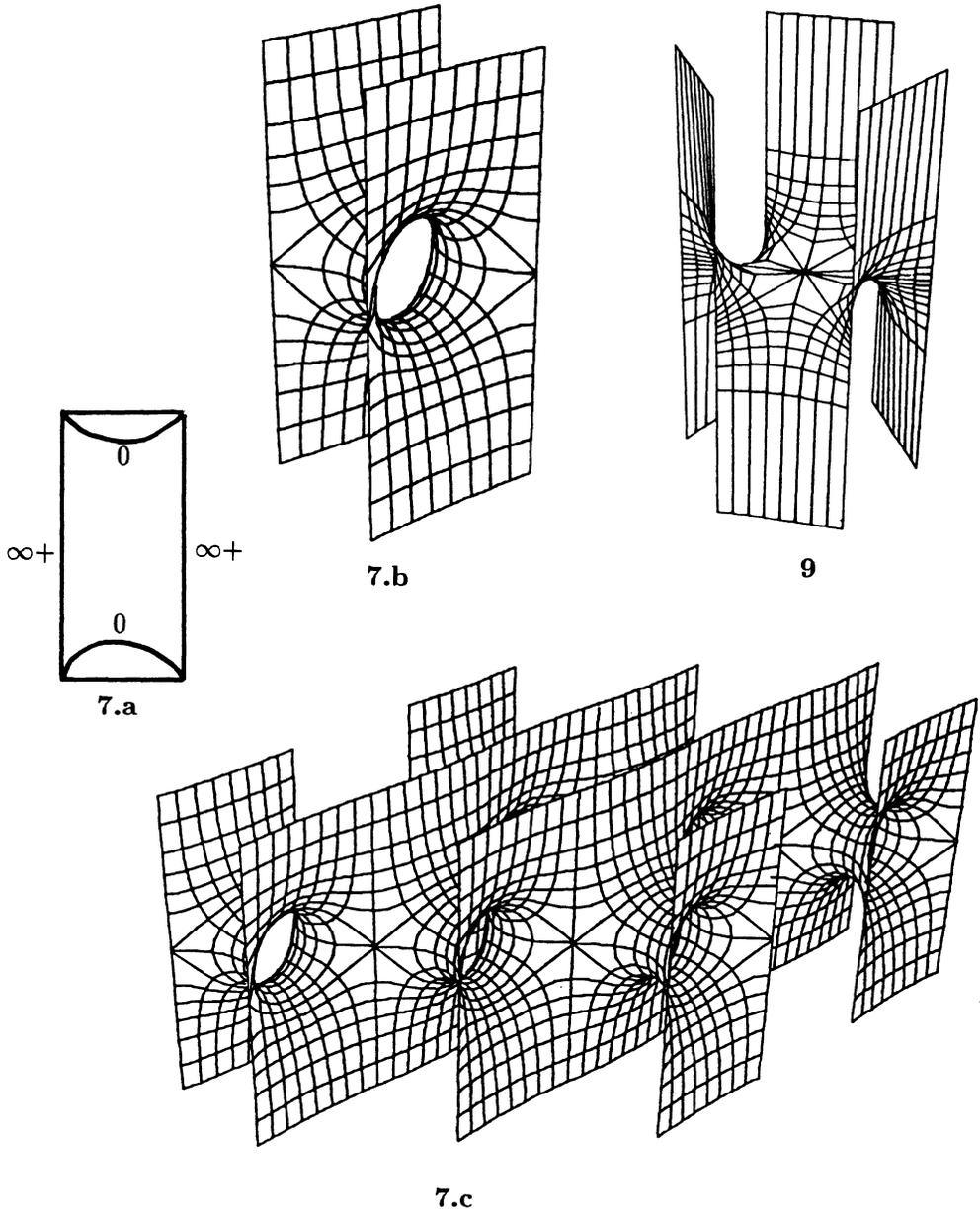
6.d



6.e



6.f



When φ does extend to a minimal graph, the graph is bounded by the vertical lines over the vertices of C . For example, the Jenkins-Serrin graph over a regular hexagon is shown in figure 9.

Clearly the theorem of Jenkins-Serrin implies Scherk's surface exists over a parallelogram precisely when it is a rhombus.

There are generalizations of Jenkins-Serrin theorem to non compact domains which have proved useful to construct complete surfaces [R.-S.E.], [K.-1].

2. THE WEIERSTRASS REPRESENTATION AND THE GEOMETRY OF THE ENDS OF A FINITE TOTAL CURVATURE MINIMAL SURFACE IN \mathbb{R}^3

Consider a Riemann surface M and a conformal map $\phi : M \rightarrow \mathbb{C}^3$ satisfying $\sum_{i=1}^3 \phi_i^2 = 0$, $\phi = (\phi_1, \phi_2, \phi_3)$. Then $X(z) = Re \int_{z_0}^z \phi$, is a minimal surface in \mathbb{R}^3 . It is not hard to see (and can be found in [Oss.-1], for example) that every surface in \mathbb{R}^3 , whose mean curvature vanishes, is locally of this form. In order for the surface in \mathbb{R}^3 to be modelled on M , one needs the integral of ϕ to be independent of the path on M between z_0 and z ; this is called the period condition : for every cycle γ on M ,

$$Re \int_{\gamma} \phi(z) dz = 0 .$$

Also, in order for M to be immersed in \mathbb{R}^3 , one requires $\sum_{i=1}^3 |\phi_i(z)| \neq 0$ for $z \in M$.

We summarize this in the definition : a minimal surface, in \mathbb{R}^3 , modelled on the Riemann surface M , is a conformal map $\phi : M \rightarrow \mathbb{C}^3$ satisfying :

- $\phi_1^2 + \phi_2^2 + \phi_3^2 = 0$ on M and $|\phi_1|^2 + |\phi_2|^2 + |\phi_3|^2$ never vanishes,
- the period condition : $Re \int_{\gamma} \phi(z) dz = 0$ for all cycles γ on M .

In case the period condition is not satisfied, one considers the minimal surface modelled on the universal conformal covering space of M (i.e. \mathbb{C} or

the open unit disc). It is common, even when X is not an embedding, to speak of $X(M)$ as the minimal surface modelled on M .

The three coordinate functions of ϕ and the one equation : $\phi_1^2 + \phi_2^2 + \phi_3^2 = 1$, can be reduced to two conditions. One classical way to do this is called the Weierstrass representation. Assuming $\phi_1 - i\phi_2$ is not identically zero (this corresponds to M a plane), let

$$g = \frac{\phi_3}{\phi_1 - i\phi_2} \quad \text{and} \quad \omega(z) = (\phi_1 - i\phi_2)dz .$$

Then g is a meromorphic function on M , ω a holomorphic one form, and M is obtained from (g, ω) by :

$$(W) \quad X(z) = \operatorname{Re} \int_{z_0}^z \left(\frac{\omega}{2}(1 - g^2), \frac{i\omega}{2}(1 + g^2), g\omega \right) .$$

This is called the Weierstrass representation of M . Notice, the poles of g are the zeros of ω and a pole of order k of g corresponds to a zero of order $2k$ of ω .

It is easy to see that a pair (g, ω) as above, i.e. g is meromorphic on M , ω holomorphic on M , satisfying the zero-pole condition, defines a minimal surface by using the formula (W). Naturally one needs the period condition for the surface to be modelled on M .

The meromorphic map g has an important geometrical meaning : it is the Gauss map of M ; more precisely, it is the composition of the usual Gauss map of $X(M)$ with stereographic projection of the unit sphere (centered at the origin) to the equatorial plane, from the north pole.

The geometric invariants of M are expressed in terms of (g, ω) . The induced metric on M is given by :

$$ds = \frac{|\omega|}{2}(1 + |g|^2) ,$$

and the curvature of M :

$$K = - \left[\frac{4|g'|}{|\omega|(1 + |g|^2)^2} \right]^2 .$$

2.1. Osserman's parametrization of finite total curvature surfaces

A Riemann surface M is said to be of finite conformal type if there is a compact Riemann surface \overline{M} such that M is conformally equivalent to \overline{M} punctured in a finite number of points.

The importance of finite total curvature for a complete minimal surface in \mathbb{R}^3 is made clear by the following theorem of Osserman.

THEOREM 2.2 [Oss.-1]. — *Let M be a complete immersed minimal surface in \mathbb{R}^3 and $|C(M)| = \int_M |K| dA < \infty$. Then M is of finite conformal type and M can be parametrized by meromorphic data on a compact Riemann surface. More precisely, if \overline{M} denotes the conformal compactification of M (so M is conformally $\overline{M} - \{p_1, \dots, p_n\}$) then the Weierstrass representation (g, ω) of M extends meromorphically to \overline{M} .*

Thus, in some sense, the theory of finite total curvature minimal surfaces in \mathbb{R}^3 is a problem in Riemann surface theory. But, we are very far from an understanding of this subject. How does one see M in terms of (g, ω) ? When is M embedded? Which M exist?

It is interesting to understand what Osserman's theorem has to do with minimal surfaces. In fact, an important part of this theorem is independent of minimality. A complete Riemannian two manifold of finite total curvature is of finite conformal type. This is a theorem of Huber [Hub.]; a modern proof can be found in [Wh.]. The hard part of Huber's theorem is the conformal type since the Cohn-Vossen inequality ($C(M) \leq 2\pi\chi(M)$ for complete 2-manifolds of non positive curvature) implies the topological type is finite when $C(M)$ is finite. Assuming this, it is not hard to extend (g, ω) to the punctures. An end A of M is conformally a punctured disc: $A = \mathbb{D}^* = \{0 < |z| \leq 1\}$. The Gauss map g extends meromorphically to the origin since the total curvature of A is the area of the Gaussian image of A , counted with multiplicity. If the puncture were an essential singularity then g would take on almost every value infinitely often and the spherical area would be infinite. Now rotate M so g is finite at the puncture. Since the metric on M is $\frac{|\omega|}{2}(1 + |g|^2)$, and the metric is complete, one has $\int_\gamma |\omega| = \infty$, for every divergent path γ on A ; i.e. γ converges to the origin viewed in \mathbb{D}^* .

Then one proves ω has a pole at the origin by a function theory argument (cf; [Oss.-1]).

2.2. The geometry of finite total curvature ends

Now it is not difficult to analyse an end A of finite total curvature. Parametrize A conformally by \mathbb{D}^* and let (g, ω) be the Weierstrass representation of A . After a rotation of A in \mathbb{R}^3 we can suppose $g(0) = 0$, so $g(z) = z^k$ after a conformal reparametrization of a subend of A . The metric complete at 0 tells us that ω must have a pole at 0 : $\omega(z) = \left(\frac{c}{z^\ell} + \mathcal{O}(|z|^{-\ell-1})\right) dz$ in a neighbourhood of 0.

A direct calculation, using the Weierstrass representation (W), yields :

$$E) \quad 2[x_1(z) - ix_2(z)] = \int_{z_0}^z \omega - \overline{\int_{z_0}^z g^2 \omega}.$$

We know $g^2 \omega$ has a milder pole than ω at 0 so $x_1 - ix_2$ has a pole of order ℓ at 0. Consider the image of the circle $re^{i\theta}$, $0 \leq \theta \leq 2\pi$, $r > 0$, r small, by the map $x_1 - ix_2$. The image curve turns $\ell - 1$ times about the x_3 -axis. Since the curve must close (i.e. A is an annulus) we have $\ell > 1$; in fact, the coefficient of $\frac{1}{z}$ in ω must be 0, since x_1 and x_2 are single valued on A .

If the end A is embedded, then the image curve : $(x_1 - ix_2)(re^{i\theta})$, $0 \leq \theta \leq 2\pi$ turns once around the x_3 -axis, hence $\ell = 2$ and

$$\omega = \left(\frac{c}{z^2} + h(z)\right) dz, \quad h \text{ holomorphic}$$

in a neighbourhood of 0.

Now $x_3 = \operatorname{Re} \int g \omega$, and $g(z) = z^k$ near 0, so $|x_3|$ is bounded means $k \geq 2$; these are the planar ends. Notice that if $k = 1$, c is real since x_3 is well defined on the end.

Integrating ϕ_3 we obtain the development of x_3 (the constants are the integration constants) :

$$x_3(z) = \begin{cases} c \ln|z| + c_0 + \mathcal{O}(|z|^2), & \text{if } k = 1, \\ d_0 + \mathcal{O}(|z|), & \text{if } k \neq 1. \end{cases}$$

From the equation E we obtain :

$$|z| = \frac{|c|}{2|x|} + \mathcal{O}(|x|^{-2}) .$$

Substitute this in $x_3(z)$ to obtain :

$$x_3 = a \ln|x| + b + \frac{c_1 x_1 + c_2 x_2}{|x|^2} + \mathcal{O}(|x|^{-2}) ,$$

where the coefficients are real constants ($a = -c$).

Hence an embedded finite total curvature end is asymptotic to a planar or a catenoid end.

3. THE CHARACTERIZATIONS OF THE CATENOID BY R. SCHOEN AND LOPEZ-ROS

In the class of finite total curvature minimal surfaces in \mathbb{R}^3 we have two fundamental theorems; each a characterization of the catenoid.

THEOREM 3.1 [Sch.-1]. — *Let M be a complete immersed finite total curvature minimal surface in \mathbb{R}^3 with two ends, each embedded. Then M is a catenoid.*

Remark : We will see that M is embedded; this follows immediately from the monotonicity formula.

THEOREM 3.2 ([Lo.-Ros] and [P.-Ros]). — *Let M be an m -surface in \mathbb{R}^3 of finite total curvature and genus zero. Then M is a plane or a catenoid.*

We make a short digression. The theorem of R. Schoen is rather surprising. Why can't one add a handle to a catenoid (figure 10) ?

Related to this, we have the unsolved conjecture of W. Meeks : let C_1 and C_2 be convex curves in parallel planes and let M be a compact connected minimal surface with $\partial M = C_1 \cup C_2$. Then M has genus zero.

Convexity is certainly necessary here as the following example shows. Let M_1, M_2 be two pieces of catenoids placed as in figure 11-a.

The boundaries of M_1, M_2 are in parallel planes. Now join the top and bottom boundary circles by narrow bridges (figure 11-b). By the Bridge principle [Cour.], [Smale], there is a minimal surface M , which is close to M_1 and M_2 (near M_1, M_2) and fills the bridges.

Now we shall begin the proofs of theorems 3.1, 3.2. A useful tool is the maximum principle at infinity. The usual maximum principle implies that the distance between two disjoint properly immersed minimal surfaces in \mathbb{R}^3 cannot be realized at points $p_1 \in \text{int}(M_1), p_2 \in \text{int}(M_2)$, unless M_1 and M_2 are parallel planes. Now what happens when M_1 and M_2 are asymptotic at infinity?

THEOREM 3.3 (Maximum Principle at Infinity [L.-R.],[M.-R.-2]). — *Let M_1 and M_2 be disjoint properly immersed minimal surfaces with compact boundary in a complete flat three manifold. If $\partial M_1 = \partial M_2 = \phi$ then M_1 and M_2 are flat. Otherwise*

$$\text{dist}(M_1, M_2) = \min\{\text{dist}(M_1, \partial M_2), \text{dist}(M_2, \partial M_1)\} .$$

In fact, the case $\partial M_1 = \partial M_2 = \phi$ is the strong halfspace theorem of Hoffman–Meeks. This case does not arise in the proofs of the theorems of Schoen and Lopez–Ros and we discuss the strong halfspace theorem in VI. What we need (and prove) here is the special case of the maximum principle at infinity, first proved by Langevin and Rosenberg.

THEOREM 3.4 [L.-R.]. — *Let M_1 and M_2 be disjoint finite total curvature embedded minimal surfaces in \mathbb{R}^3 with compact boundaries. Then $\text{dist}(M_1, M_2) > 0$.*

The proof of this theorem uses the notion of flux on a minimal surface M .

Let α be an oriented cycle on M and denote the complex structure operator of M by J (J is rotation by $\pi/2$ in each tangent space of M). The flux of α is

$$\text{Flux}(\alpha) = \int_{\alpha} J(\alpha') ,$$

where α' is the unit tangent vector to $\alpha : J(\alpha')$ is a conormal field to M along α ; a unit normal to α , tangent to M .

Since the coordinate functions are harmonic on M , a direct application of the divergence theorem shows that $Flux(\alpha)$ only depends on the homology class of α ; so flux should be thought of as an \mathbb{R}^3 -valued function on the homology of M .

Now, we established in II that an embedded finite total curvature end of a minimal surface in \mathbb{R}^3 , has a limiting normal vector (which we suppose vertical here) and the end can be written as a graph $u(x)$, for $|x|$ large :

$$u(x) = a \ell n|x| + b + \frac{c_1 x_1 + c_2 x_2}{|x|^2} + \mathcal{O}(|x|^{-2}) .$$

We say the end is a catenoid type end if $a \neq 0$ (a is the logarithmic growth rate of the end) and the end is planar if $a = 0$. In the first case the end is geometrically asymptotic to a catenoid and to a horizontal plane when $a = 0$.

The development of u can be differentiated term by term, so the outward pointing conormal vector to the end, along the curve $C_R = \{(x, y, u(x, y))/x^2 + y^2 = R^2\}$ is easily calculated to be

$$\nu = \frac{1}{R}(x, y, a) + \mathcal{O}(|R|^{-2}) .$$

Hence $Flux(C_R) = (0, 0, 2\pi a) + \mathcal{O}(|R|^{-1})$.

Since the flux only depends on the homology class of C_R , we have $Flux(C_R) = (0, 0, 2\pi a)$. In particular, the $Flux$ vector only depends on the logarithmic growth rate of the end and is parallel to the limiting normal vector.

Now we can prove theorem 3.4. Assume there are two embedded disjoint minimal surfaces M_1, M_2 of finite total curvature, compact boundaries, and $\text{dist}(M_1, M_2) = 0$. We will see this leads to a contradiction.

Since finite total curvature of M_1, M_2 , implies finite topological type, each M_i has a finite number of ends, each end topologically an annulus. Since $\text{dist}(M_1, M_2) = 0$ and M_1, M_2 have compact boundaries, there must be an end E_1 of M_1 and E_2 of M_2 such that $\text{dist}(E_1, E_2) = 0$. After a

rotation in \mathbb{R}^3 , we can assume E_1 and E_2 are asymptotic to the same horizontal catenoid end of growth rate a (if $a = 0$ its a horizontal plane).

Since $E_1 \cap E_2 = \phi$, we can assume E_1 lies above E_2 (we can take E_1 and E_2 to be graphs). After a small vertical downward translation E'_1 of E_1 , $\partial E'_1$ still lies above E_2 , but outside of a large ball, E'_1 lies below E_2 . Hence $E'_1 \cap E_2$ is a compact nonempty one dimensional analytic subset of both E'_1 and E_2 .

We now show that $E'_1 \cap E_2$ is a simple closed curve γ , homotopically non trivial on E'_1 and E_2 , and E'_1 is transverse to E_2 along γ . Since E'_1 is a graph, the projection $\pi : \mathbb{R}^3 \rightarrow \mathbb{R}^2$ of $E'_1 \cap E_2$ is a compact nonempty one-dimensional analytic variety of \mathbb{R}^2 . Let D be a disc in \mathbb{R}^2 so that E'_1 is a graph over $\mathbb{R}^2 - D$. If $\pi(E'_1 \cap E_2)$ is not a connected, homotopically non trivial simple closed curve in $\mathbb{R}^2 - D$, then $\mathbb{R}^2 - \pi(E'_1 \cap E_2)$ contains a compact component disjoint from D . This is impossible since the lifts of this component to E_2 and E'_1 correspond to different solutions to the minimal surface equation with the same boundary values (impossible by the usual maximum principle). Hence E'_1 intersects E_2 transversely in a single curve γ that is homotopically non trivial on E'_1 and E_2 . Let \tilde{E}_1 and \tilde{E}_2 denote the ends of E'_1 , E_2 respectively, with boundary γ .

The surfaces \tilde{E}_1 and \tilde{E}_2 represent distinct solutions to the minimal surface equation (they are graphs) over the unbounded region Δ of \mathbb{R}^2 with boundary $\pi(\gamma)$, and they have the same boundary values along $\pi(\gamma)$. Since \tilde{E}_1 and \tilde{E}_2 are asymptotic to translates of a fixed vertical catenoid, they have the same logarithmic growth rate.

Let X_1 and X_2 denote the gradient of the third coordinate functions of \tilde{E}_1 and \tilde{E}_2 , respectively. If ν_1 , ν_2 denote the conormal (upward pointing) vectors to \tilde{E}_1 and \tilde{E}_2 along γ , then

$$Flux(\tilde{E}_1) = (0, 0, \int_{\gamma} X_1 \cdot \nu_1) = (0, 0, \int_{\gamma} X_2 \cdot \nu_2) .$$

However, along γ , \tilde{E}_1 lies below \tilde{E}_2 , so $X_1 \cdot \nu_1 < X_2 \cdot \nu_2$ at each point of γ , and this contradicts $\int_{\gamma} X_1 \cdot \nu_1 = \int_{\gamma} X_2 \cdot \nu_2$.

A basic result in minimal surface theory is the monotonicity formula; a proof may be found in [G.-T.].

THEOREM 3.5 (Monotonicity Formula). — *Let M be a properly immersed minimal surface in \mathbb{R}^3 , $x \in M$, and $D_R(x) =$ the euclidean ball of \mathbb{R}^3 , of radius R , centered at x . Let k be the number of sheets of M passing through x and $\Sigma(R) = M \cap D_R(x)$. Then $\frac{|\Sigma(R)|}{\pi R^2 k}$ is a monotone increasing function of R , which tends to one as $R \rightarrow 0$. Here $|\Sigma(R)|$ denotes the area of $\Sigma(R)$. In fact, each sheet of $\Sigma(R)$, passing through x , has area growing at least as fast as πR^2 , the area of the flat disc through x of radius R .*

COROLLARY. — *Let M be a complete finite total curvature surface in \mathbb{R}^3 with exactly two embedded ends. Then M is embedded.*

Proof of Corollary : Each end E of M can be written as the graph of a function u over \mathbb{R}^2 — a compact disc. A simple calculation shows the area growth of E is Euclidean, i.e. $\frac{|E \cap D_R|}{\pi R^2} \rightarrow 1$ as $R \rightarrow \infty$. Since M has exactly two ends, each embedded, we conclude $f(R) = \frac{|M \cap D_R|}{2\pi R^2} \rightarrow 1$ as $R \rightarrow \infty$. If M had a point of self intersection, the monotonicity formula implies $f(R) = 1$ for all R hence M is the union of two flat planes.

Proof of Theorem 3.1 : Let E_1 and E_2 be the ends of M and let γ_1 and γ_2 be smooth Jordan curves on E_1 and E_2 respectively, each homotopically non trivial on its end. Let Σ be the compact submanifold of M bounded by $\gamma_1 \cup \gamma_2$. Since γ_1 is homologous to γ_2 in Σ , the flux of γ_1 equals the flux of γ_2 . Hence, if the flux of γ_1 is different from 0, then the limiting normal vectors to E_1 and E_2 are parallel and if a_1 and a_2 are the logarithmic growth rates of E_1 and E_2 respectively, then $a_1 = -a_2$. Also the flux formula implies that if a_1 is zero then so is a_2 . Thus both ends are simultaneously catenoid type ends or planar type ends. Moreover, since we know M is embedded, the ends are always parallel, i.e. their limiting normal vectors are parallel.

After a rotation of M we can assume the ends are horizontal.

We observe first that neither end of M is planar. For if E_1 were planar then we could find a horizontal plane P disjoint from M . Then move P towards M by parallel translation. There would be a first point of contact with the horizontal plane (at a finite point of M , or at infinity) and this contradicts the maximum principle at infinity, or the usual maximum

principle.

So we can assume E_1 is a catenoid type end (above E_2) with growth rate $a_1 \neq 0$ and $a_1 = -a_2$.

We will now prove there is a horizontal plane P which is a plane of symmetry of M . This allows us to say that the catenoids to which E_1 and E_2 are asymptotic, have the same vertical axis. One finds P by applying the Alexandrov reflection technique and the maximum principle at infinity.

Let $P(t)$ be the horizontal plane $x_3 = t$. For each t , let M_t^+ be the part of M , on and above $P(t)$ and M_t^- the part of M on and below $P(t)$. Let M_t^* be the symmetry of M_t^+ by $P(t)$:

$$M_t^* = \{(x, 2t - x_3)/(x, x_3) \in M_t^+\} .$$

A surface S has locally bounded slope if the tangent plane to every interior point of S never contains a vertical line. Finally we say a subset A is above a subset B , written $A \geq B$, if for every $x \in \mathbb{R}^2$ for which $p^{-1}(x) \cap A \neq \emptyset$ and $p^{-1}(x) \cap B \neq \emptyset$, we have all points of $p^{-1}(x) \cap A$ lying above all points of $p^{-1}(x) \cap B$. Here $p: \mathbb{R}^3 \rightarrow \mathbb{R}^2$ is the projection to the horizontal.

Now for t large, M_t^+ is a graph of locally bounded slope over (a part of) $P(t)$; M_t^+ is part of E_1 for t large. M_t^* is then a catenoid type end of growth rate $-a_1$; the same as E_2 . Thus for t sufficiently large, M_t^* is above M_t^- .

Now consider decreasing t , to s say, and the surface M_s^* . We claim that if for each τ , $s \leq \tau \leq t$, M is never vertical along $P(\tau)$, then M_s^* is above M_s^- . Otherwise, there would be a first interior point of contact of some M_τ^* with M_τ^- and the usual maximum principle yields $M_\tau^* = M_\tau^-$ so M is vertical along $P(\tau) \cap M$. Here we have used the maximum principle at infinity to say the end of M_τ^* is a strictly positive distance from E_2 .

Since it is not possible that M_s^* is above M_s^- for all s , there is a largest τ such that M is vertical at some point p of $M \cap P(\tau)$. M_τ^* is above M_τ^- hence M_τ^* and M_τ^- have a common boundary, they are tangent at p , and one is locally on one side of the other at p . Thus $M_\tau^* = M_\tau^-$ by the boundary maximum principle and M is invariant by reflection in $P(\tau)$.

Now consider the development of E_2 as a graph :

$$u_2(x) = a \ell n|x| + b + \frac{c_1 x_1}{|x|^2} + \frac{c_2 x_2}{|x|^2} + \mathcal{O}(|x|^{-2}) .$$

We wish to translate M so that the axis of the catenoid passes through the origin. Let $x_1 = y_1 + \alpha_1$, $x_2 = y_2 + \alpha_2$. Then u_2 in terms of y is given by :

$$u_2(y) = a \ell n|y| + b + \frac{\tilde{c}_1 y_1}{|y|^2} + \frac{\tilde{c}_2 y_2}{|y|^2} + \mathcal{O}(|y|^{-2}) ,$$

where $\tilde{c}_i = c_i + a\alpha_i$, $i = 1, 2$. Thus letting $\alpha_i = -\frac{c_i}{a}$, $i = 1, 2$, and calling y by x again, we have :

$$u_2(x) = a \ell n|x| + b + \mathcal{O}(|x|^{-2}) .$$

Assuming the horizontal plane $x_3 = 0$ is the plane of symmetry of M (which we can suppose after a vertical translation of M), we have $u_2(x) = -u_1(x)$ and the development of $u_1(x)$ is :

$$u_1(x) = -a \ell n|x| - b + \mathcal{O}(|x|^{-2}) .$$

We will now prove M is a surface of revolution about the x_3 -axis, thus a catenoid. Since the expression of u_2 is invariant by rotation about the origin in the $x = (x_1, x_2)$ plane, it suffices to prove the plane $x_1 = 0$ is a plane of symmetry of M .

Denote by $P(t)$ the planes $x_1 = t$ (now we shall think of the x_1 axis as vertical) and for K a (large) constant let $M \cap \{|x_3| = K\} = B_1 \cup B_2 = B$, where $B_2 = M \cap \{x_3 = K\}$, $B_1 = M \cap \{x_3 = -K\}$.

Here is the idea of what we shall do next. Fix $t > 0$. For K large, B_1 and B_2 are close (C^1 -close) to circles in the planes $|x_3| = K$, centered at the x_3 axis. So B_t^+ is a graph of bounded slope over $P(0)$ and $B_t^* \geq B_t^-$, cf. figure 12.

Let $\Sigma = \Sigma(K)$ be the compact part of M bounded by B . By doing the Alexandrov reflection technique with the horizontal planes $P(s)$, coming down to $P(t)$, from above $\Sigma(K)$, one proves that Σ_t^+ is a graph of bounded slope over $P(0)$ and $\Sigma_t^* \geq \Sigma_t^-$. By construction, this will continue to hold for any larger value of K . So $M_t^* \geq M_t^-$. Since this is true for any $t > 0$, we have $M_0^* \geq M_0^-$. Now do the same argument from below, i.e. start with $-t$ and come up, from below with horizontal planes to conclude $M_0^- \leq M_0^+$. hence M is invariant by symmetry in $P(0)$.

We now make the above discussion precise. First we prove B_t^+ is a graph of bounded slope and $B_t^* \geq B_t^-$. Clearly it suffices to do this for B_2 (B_1 is the same argument). We have

$$u_2(x) = a \ln|x| + b + \mathcal{O}(|x|^{-2}),$$

hence

$$\frac{\partial u_2}{\partial x_1} = \frac{ax_1}{|x|^2} + \mathcal{O}(|x|^{-3}).$$

Consequently for $x_1 \geq t$ and $|x|$ sufficiently large (depending on t) we have $\frac{\partial u_2}{\partial x_1} > 0$. The normal vector η to B_2 , in the plane $x_3 = K$ is

$$\eta = \left(\frac{ax_1}{|x|^2} + \mathcal{O}(|x|^{-3}), \frac{ax_2}{|x|^2} + \mathcal{O}(|x|^{-3}), 0 \right).$$

The first coordinate of η is non zero for $|x|$ large, and $x_1 \geq t$, so B_t^+ is a graph of bounded slope.

Now on B_2 we have $u_2 = K$, so

$$\ln|x| + \mathcal{O}(|x|^{-2}) = \frac{1}{a}(K - b)$$

hence $|x|e^{\mathcal{O}(|x|^{-2})} = R$ for a large R . Since $e^{\mathcal{O}(|x|^{-2})} = 1 + \mathcal{O}(|x|^{-2})$, it follows that $|x| = R + \mathcal{O}(|x|^{-1})$, so B_2 is close to a circle for $|x|$ large. Let C denote the circle of radius R in $x_3 = K$, centered at the origin. Clearly

$$\text{dist}(C_t^*, C_{t/2}^-) \geq \varepsilon(t),$$

where $\varepsilon(t) > 0$, depends only on t . Hence if K is sufficiently large, $B_{2,t}^* \geq B_{2,t/2}^-$. Since $B_{2,t/2}^+$ is a graph over $P(0)$, it follows that $B_{2,t}^* \geq B_{2,t}^-$.

To complete the proof of R. Schoen's theorem, it remains to show $\Sigma_t^* \geq \Sigma_t^-$, where $\Sigma = \Sigma(K)$ is the compact submanifold of M bounded by B .

Let T be the maximum value of x_1 on Σ ; it is realised on B since x_1 is harmonic.

Define $J = \{s \in [t, T] / \Sigma_s^+ \text{ is a graph of locally bounded slope over } P(s) \text{ and } \Sigma_s^* \geq \Sigma_s^-\}$. We prove the theorem by showing J is a non empty open and closed subset of $[t, T]$, so $t \in J$.

Let $p \in B$ be a point where $x_3 = T$. Then $T \in J$ so J is non empty. Since a compact minimal surface is in the convex hull of its boundary (the maximum principle applied to the coordinate functions) Σ is in the slab between the planes $x_3 = K$ and $x_3 = -K$. By the boundary maximum principle, Σ is never tangent to the planes $|x_3| = K$; i.e; Σ is transverse to $|x_3| = K$ along B . So for s near T , $s < T$, Σ_s^+ is a graph over $P(s)$ of locally bounded slope. Notice also, that if $s \in J$ and $s < s_1 \leq T$, then $s_1 \in J$.

First we show J is closed. Suppose $(s, T) \subset J$. If Σ_s^+ is not a graph then there is s_1 , $s < s_1 \leq T$ and $x \in P(o)$ such that (x, s) and (x, s_1) are on the same vertical. We choose s_1 so there are no other points of Σ on the vertical between (x, s) and (x, s_1) . We know that (x, s) is an interior point of Σ since B_t^+ is a graph and Σ touches $|x_3| = K$ only along B .

Now Σ is a graph in a neighborhood of (x, s_1) and not vertical in the neighborhood. So the vertical lines to this neighborhood, descend to fill a neighborhood of (x, s) . It follows that Σ is below $P(s)$ in a neighborhood of (x, s) , otherwise there would be an (\bar{x}, \bar{s}) on Σ , near (x, s) , with $\bar{s} > s$. But then (\bar{x}, \bar{s}) would have another point of Σ , above it, on the vertical, so Σ_s^+ would not be a graph. Clearly Σ below $P(s)$ at (x, s) implies $\Sigma = P(s)$ by the maximum principle, which is a contradiction. Thus J is closed.

Next we show J is open. Let $(x, s) \in \Sigma$ and let assume $[s, T] \subset J$.

Let $D \subset \Sigma$ be a disc containing (x, s) . Notice that Σ is not vertical at (x, s) , for if this were so, consider the discs D_s^* and D_s^- . Then D_s^* is locally on one side of D_s^- in a neighborhood of (x, s) , so $D_s^* = D_s^-$ by the boundary maximum principle and $P(s)$ is a plane of symmetry of Σ . This is impossible since B is not symmetric in $P(s)$.

Thus Σ is a graph in a neighbourhood U of $P(s)$ and not vertical in U . It remains to show $\Sigma_\tau^* \geq \Sigma_\tau^-$ for τ near s . Since $\Sigma \cap U$ is a graph, we have $\Sigma_\tau^* - V \geq \Sigma_\tau^-$ for V a neighborhood of $P(s)$, $V \subset U$, and τ near s . Also $\Sigma_\tau^+ - V$ is compact and its image under reflection in $P(s)$ is disjoint from Σ_s^- , so by continuity, for τ near s , we have $\Sigma_\tau^* - V \geq \Sigma_\tau^-$. This means $\Sigma_\tau^* \geq \Sigma_\tau^-$ for τ near s .

The last argument using Alexandrov reflection actually proves more : under certain circumstances, a minimal surface inherits the symmetries of its boundary. More precisely, R. Schoen proved :

THEOREM 3.6. — *Suppose $\Omega \subset \mathbb{R}^n$ is a compact domain whose boundary is mean convex. Let $B^{n-1} \subset \mathbb{R}^{n+1}$ be a compact embedded manifold (not necessarily connected) satisfying : $B \subset (\partial\Omega) \times \mathbb{R}$, B_0^+ is a graph of locally bounded slope over $P(0) = \mathbb{R}^n \times (0)$ and $B_0^+ \geq B_0^-$. Let M be an embedded minimal hypersurface with $\partial M = B$ and all interior points in $\Omega \times \mathbb{R}$. Then M_0^+ is a graph over $P(0)$ of locally bounded slope and $M_0^+ \geq M_0^-$.*

In fact, R. Schoen proved this theorem only assuming M immersed, but one has to work a little more than we did to obtain this generality. The hypothesis, the interior points of M are in $\Omega \times \mathbb{R}$ is not serious, since if an interior point p of M is in $(\partial\Omega) \times \mathbb{R}$ then, by the maximum principle, the connected component of p in M is entirely contained in $(\partial\Omega) \times \mathbb{R}$, so one can disregard these components of M .

It is easy to construct examples of boundaries B whose symmetries do not pass to a minimal M with $\partial M = B$. For example consider two copies of a dumbbell curve in parallel planes as in figure 13.

The (asymmetric) minimal M can be obtained by applying the bridge principle to a catenoid bounded by the two circles on the right, and the two discs bounded by the two circles on the left.

We now begin the proof of the Lopez–Ros theorem : the plane and the catenoid are the only properly embedded minimal surfaces with finite total curvature and genus zero.

Here is the idea. Suppose M is a embedded finite total curvature surface in \mathbb{R}^3 . We know M has a finite number of ends which we can assume horizontal (i.e. their limiting normals are vertical) after a rotation of M in \mathbb{R}^3 . Each end is asymptotic to a horizontal plane or to a catenoid.

Lopez and Ros deform M , among minimal surfaces by deforming the Weierstrass data. If (g, ω) is the Weierstrass data of M they consider the data $(\lambda g, \frac{\omega}{\lambda})$, where λ is a positive real number. One checks that this data defines an immersion X_λ of M . In fact this holds whenever the flux of M is vertical. In our case this is so, since M has genus zero so all the flux is at the ends. The planar ends have zero flux and the catenoid type ends have vertical flux since their limiting normals are vertical.

Next one observes that each catenoid end of M is a (vertical) catenoid end of X_λ with the same logarithmic growth rate; it can move up or down during the deformation. The planar ends stay planar ends at the same height.

By the maximum principle at infinity, the distance between the ends of X_λ is strictly positive, as λ goes from 1 to infinity. Therefore each X_λ is an embedded surface.

If M has a point p where the normal vector is vertical, then there is a neighborhood D of p and a $\lambda > 1$ such that $X_\lambda(D)$ is not embedded (assuming M is not a plane). One sees this by proving X_λ converges to an Enneper surface (which is not embedded) near p , as $\lambda \rightarrow \infty$. Thus M has no points p where the normal is vertical.

If M has a planar end A , then a similar analysis proves $X_\lambda(D)$ is not embedded, for λ large and D a subend of A (assuming M not flat).

Thus all the ends of M are catenoid type and the Gauss map has no zeros or poles on M . The conformal compactification of M is the sphere S and the foliation by the level curves of x_3 is non singular on M and has a singularity of positive index at each puncture. Hence there are exactly two ends in M .

We could now refer to R. Schoen's theorem to conclude M is a catenoid but it is easy to prove this directly. Since the Gauss map has all its zeros and poles at the ends, g has degree one. After a conformal reparametrisation of M one can assume $g(z) = z$. A little residue theory then proves $\omega(z) = c \frac{dz}{z^2}$, $c \in \mathbb{R}$, hence M is a catenoid [Oss.-1].

Now we enter into the details of this argument.

For each cycle γ on M , one can calculate the flux of γ by :

$$\int_{\gamma} (\phi_1, \phi_2, \phi_3) = i \text{ Flux}(\gamma) .$$

From this formula, it follows easily that the following three conditions are equivalent to M having vertical flux :

- 1) the forms ϕ_1 and ϕ_2 are exact,
- 2) the forms ω and $g^2\omega$ are exact,
- 3) for each $\lambda > 0$, the immersion X_λ is well defined on M .

The formula for the metric and curvature show that each $X_\lambda(M) = M_\lambda$ is a complete minimal surface of the same total curvature as M .

Now suppose $p \in M$ is a point where the normal to M is vertical, say $(0, 0, -1)$. Parametrize a neighborhood of p conformally by $\{0 \leq |z| < \varepsilon\}$, with

$$g(z) = z^k, \quad \omega = (a + zh(z))dz,$$

where $a \in \mathbb{C}^*$ and h is holomorphic in $D(t) = \{|z| < \varepsilon\}$. Introduce the conformal coordinate $\xi = \lambda^{1/k}z$ on $D(\lambda^{1/k}\varepsilon)$. Then X_λ is parametrized by

$$g_\lambda(\xi) = \xi^k, \quad \omega_\lambda = \frac{1}{\lambda^{1+1/k}} \left(a + \frac{\xi}{\lambda^{1/k}} h\left(\frac{\xi}{\lambda^{1/k}}\right) \right) d\xi.$$

Now dilate X_λ by $\lambda^{1+1/k}$ to obtain \tilde{X}_λ . As $\lambda \rightarrow \infty$, \tilde{X}_λ converges, on compact subsets of \mathbb{C} , to the minimal surface $X_\infty : \mathbb{C} \rightarrow \mathbb{R}^3$, with Weierstrass data :

$$g_\infty(\xi) = \xi^k, \quad \omega_\infty = ad\xi.$$

This is a complete surface with a non embedded end; there are transversal self intersections. Hence X_λ , for λ large, has self intersections.

If the normal at p is $(0, 0, 1)$, then turn M upside down.

Now suppose A is a planar end of M (and M is not a plane), and let the limiting normal vector to A be $(0, 0, -1)$. Parametrize a subend of A by the Weierstrass data in $D(\varepsilon)$:

$$g(z) = z^k, \quad \omega = \left(\frac{a}{z^2} + h(z) \right) dz,$$

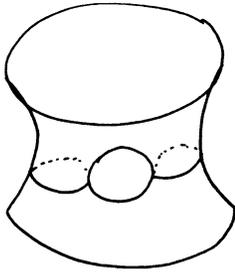
where $a \in \mathbb{C}^*$ and h holomorphic in $D(\varepsilon)$.

We have a parametrization of the end of M_λ in $D(\lambda^{1/k}\varepsilon)$, $\xi = \lambda^{1/k}z$,

$$g_\lambda(\xi) = \xi^k, \quad \omega_\lambda = \frac{1}{\lambda^{1-1/k}} \left(\frac{a}{\xi^2} + \frac{1}{\lambda^{2/k}} h\left(\frac{\xi}{\lambda^{1/k}}\right) \right) d\xi, \quad k > 1.$$

After a homothety by $\lambda^{1-1/k}$, we obtain a new minimal surface \widehat{M}_λ . When $\lambda \rightarrow \infty$, \widehat{M}_λ converges uniformly on compact subsets of \mathbb{C}^* to $X_\infty : \mathbb{C}^* \rightarrow \mathbb{R}^3$ defined by

$$g_\infty(\xi) = \xi^k, \quad \omega_\infty = \frac{a}{\xi^2} d\xi, \quad \xi \in \mathbb{C}^*.$$



10

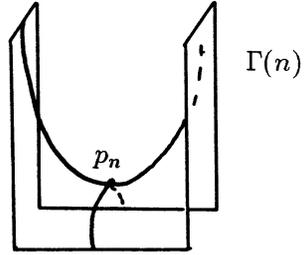


Figure 8

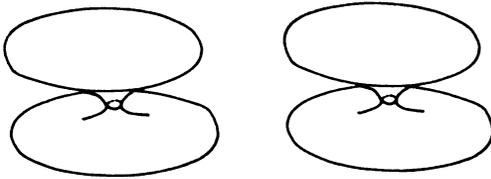


Figure 11.a

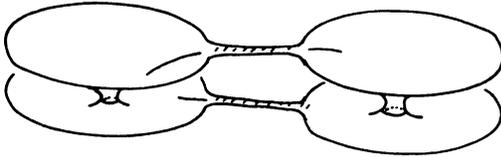
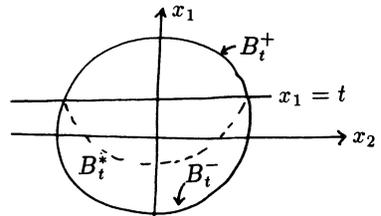


Figure 11.b



12

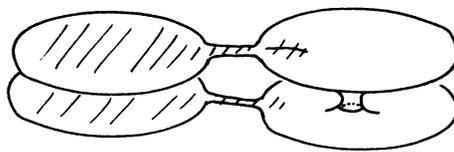


Figure 13

If $k = 1$, this is a catenoid, and if $k > 1$, the surface has a non embedded end at infinity. So for λ large, M_λ is not embedded, since A is a planar end.

It remains to prove each X_λ is an embedding. Let $J = \{\lambda/X_\lambda \text{ is injective}\}$. If $\lambda_0 \in J$ then the distance between two fixed ends of $X_{\lambda_0}(M)$ is strictly positive, by the maximum principle at infinity. Clearly this distance is a continuous function of λ (it may be infinite). Hence for λ near λ_0 , X_λ is also an embedding and J is open.

Suppose $\lambda_n \in J$, $\lambda_n \rightarrow \lambda$ as $n \rightarrow \infty$. If X_λ is not injective then there are points $x, y \in M$, $x \neq y$, with $X_\lambda(x) = X_\lambda(y)$. The intersection of M_λ at $X_\lambda(x)$ and $X_\lambda(y)$ cannot be one dimensional so a neighborhood of x and a neighborhood of y have the same image by X_λ , by the usual maximum principle (we used $\lambda_n \in J$ here). Hence $X_\lambda : M \rightarrow M_\lambda$ is a finite covering of the (embedded) minimal surface M_λ . Again, by the maximum principle at infinity, there is an ε -tubular neighborhood U of M_λ that is embedded. The ends of M_{λ_n} vary continuously so for n large, $M_{\lambda_n} \subset U$. But then the orthogonal projection of M_{λ_n} to M_λ is a diffeomorphism so X_λ is also a diffeomorphism. This contradiction shows J is closed and completes the proof of the Lopez-Ros theorem.

4. CURVATURE ESTIMATES FOR STABLE MINIMAL SURFACES

In 1952, E. Heinz proved that if M is a minimal graph over the disc D_R of radius R ($D_R = \{x^2 + y^2 \leq R^2\}$) and if K_0 is the Gaussian curvature of M at the origin, then [Heinz] :

$$|K_0| \leq \frac{4\pi^3}{3R^2} .$$

This result was generalized by E. Hopf, Finn and Osserman, [E.Hopf-1],[F.-Oss.], to parametric minimal surfaces whose Gauss map misses an open set.

The most general theorem was obtained by R. Schoen [Sch.-2] : there is a universal constant $C > 0$ such that if M is a stable minimal (immersed

and complete) surface in a flat three manifold then

$$|K(p)| \leq \frac{C}{d(p)^2},$$

where $p \in M$, and $d(p)$ is the intrinsic distance of p to ∂M . Stable means that every compact domain D of M minimizes area up to second order, among normal variations of D leaving the boundary fixed; we will make this precise shortly.

This theorem of Schoen is a very important tool for the study of minimal surfaces in three-manifolds. Notice that this implies the only complete immersed stable minimal surfaces with no boundary in flat 3-manifolds are totally geodesic. So, for example, in \mathbb{R}^3 they are planes. This result was also proved by Do Carmo and Peng [Do C.-P.].

Why is Schoen's result a generalization of Heinz's theorem, i.e. why is a minimal graph stable? In general a foliation by minimal hypersurfaces implies each leaf is stable. For the unit vector field n to the foliation is divergence free in the ambient space. Let D be a compact domain in a leaf and \tilde{D} a chain with $\partial\tilde{D} = \partial D$; so that $D \cup \tilde{D}$ is a cycle, homologous to zero. The divergence theorem implies the flux of n across D equals the flux of n across \tilde{D} ; i.e.

$$area(D) = \int_D \langle n, n \rangle = \int_{\tilde{D}} \langle n, n_{\tilde{D}} \rangle \leq area(\tilde{D}),$$

where $n_{\tilde{D}}$ is the unit normal vector field to \tilde{D} . Hence D is area minimizing in its homology class.

Now the vertical translation of a graph foliates a solid cylinder and the above argument shows that for every $R' < R$, the part of M over $D_{R'}$ minimizes area up to second order. Letting $R' \rightarrow R$ we see this is also true for M .

4.1. The Barbosa-Do Carmo stability criteria

There is an important criteria for stability of a domain on a minimal surface in \mathbb{R}^3 due to Barbosa and Do Carmo [B.-Do C.] which implies that graphs are stable. Their theorem says that an immersed minimal surface in

\mathbb{R}^3 is stable if the area of the spherical image (by the Gauss map) is less than 2π .

I would like to make a few comments on their theorem. Let D be a compact domain on the minimal surface M , n a unit normal vector field to M and f a piecewise smooth function on D which vanishes on ∂D . The vector field $Y = fn$ on D , induces a normal variation of D and the second derivative of area of this variation is :

$$- \int_D f(\Delta f - 2Kf)$$

where Δ is the intrinsic Laplacian of M . The operator $L = \Delta - 2K$ is the stability (or Jacobi) operator of M . M stable means the above integral is strictly positive for all compact domains D and non constant f on D , vanishing on ∂D . Hence if one can find a non constant f , $f = 0$ on ∂D , in the kernel of L (such f are called Jacobi fields), D is not stable.

Now suppose D is a domain on which the Gauss map g is a branched covering onto $g(D)$. Then Schwarz proved that if the first eigenvalue λ_1 of the spherical Laplacian Δ_s on $g(D)$ is less than two, D is not stable. Here is the proof. Let u be a function on $g(D)$, u positive in interior $g(D)$, zero on $\partial g(D)$ and $\Delta_s u + \lambda_1 u = 0$. Define $f = u \circ g$. Since $g(\partial D) = \partial(gD)$, f vanishes on ∂D and is positive in interior D . The second variation defined by f is :

$$- \int_D f \Delta_s f + 2f^2 = (\lambda_1 - 2) \int_D f^2 < 0 .$$

Hence D is not stable (the above integrals are taken on the complement of the branch points of g).

Now here is the idea of the proof of the Barbosa-Do Carmo stability criteria. If D is not stable then one can find a domain $\tilde{D} \subset D$ and a function u on \tilde{D} , $u > 0$ on $\text{int } \tilde{D}$, $u = 0$ on $\partial \tilde{D}$ and $\Delta u - 2Ku = 0$.

One then averages u via the Gauss map to obtain a function f on $g(\tilde{D})$ satisfying

$$\int_{g(\tilde{D})} |\text{grad } f|^2 \leq 2 \int_{g(\tilde{D})} f^2 .$$

This inequality implies $\lambda_1(g(\tilde{D})) \leq 2$.

However, among all spherical domains having a fixed area, the spherical cap minimizes the first eigenvalue of the Laplacian. But a spherical cap in an open hemisphere has $\lambda_1 > 2$ (the coordinate functions of \mathbb{R}^3 satisfy $\Delta_s + 2 = 0$ and they are positive on an open hemisphere, zero on its boundary) so this is a contradiction.

4.2. An idea of the proof of Heinz’s theorem

Let us suppose M is the graph of a function f whose gradient vanishes at the origin. This gradient hypothesis makes the proof simpler. The idea is to compare M to a Scherk graph.

We can assume f defines a minimal graph on D_R , $f(0, 0) = 0$, and $|\nabla f(0, 0)| = 0$. Rotate the graph of f so the x -axis is a principal direction, curving upwards.

Let N be a Scherk graph defined over a square of side length 2, centered at the origin, with boundary values $+\infty$ on the vertical sides of the square and $-\infty$ on the horizontal sides. Assume also N passes through the origin; clearly N is horizontal at the origin.

Let \tilde{K}_0 be the Gauss curvature of N at the origin. A homothety of N by $C > 0$, from the origin, transforms N to a minimal graph N_C defined over a square $\diamond(C)$ containing D_C . Since curvature is multiplied by $\frac{1}{C^2}$ under this homothety, the curvature K_C of N_C at the origin, satisfies

$$|K_C| \leq \frac{|\tilde{K}_0|}{C^2} .$$

Notice that N_C is horizontal at the origin and one of the principal curvatures of N_C is along the x -axis and points upward.

Choose $C > 0$ so that the principal curvature of N_C , along the x -axis at the origin, equals the corresponding principal curvature of M at the origin. Then $K_0 = K_C$ at this point.

Now if $R \leq C$ then $D_R \subset D_C$ and

$$|K_0| = |K_C| \leq \frac{|\tilde{K}_0|}{C^2} \leq \frac{|\tilde{K}_0|}{R^2} .$$

If $\diamond(C) \subset D_R$ then consider $M \cap N_C$. Both surfaces are tangent at \mathcal{O} so they are equal or $M \cap N_C$ is a one dimensional analytic curve, singular

at \mathcal{O} , and with at least six branches passing through the singularity. Since N_C is asymptotic to infinity on the boundary of the square $\diamond(C)$, except at the four vertices, there must be a compact component of $M \cap N_C$ strictly contained in the vertical region over the interior of $\diamond(C)$ (at most one branch of $M \cap N_C$ can go to a fixed vertex of $\diamond(C)$ since M is a graph). Then there is a Jordan curve α in $\diamond(C)$ along which M and N_C agree. Since they are both graphs over the interior of α and one has unicity of such minimal graphs by the usual maximum principle, we have $M = N_C$; a contradiction. Thus $\diamond(C)$ is not contained in D_R , hence $C > R/\sqrt{2}$ and

$$|K_0| = |K_C| \leq \frac{|\tilde{K}_0|}{C^2} < \frac{2|\tilde{K}_0|}{R^2},$$

and a Heinz type estimate is established.

5. COMPACTNESS OF LEAST AREA FAMILIES AND CONSTRUCTION OF COMPLEMENTARY FINITE TOTAL CURVATURE SURFACES

A technique used often to study a complete minimal surface M in a flat 3-manifold N is to construct finite total curvature minimal surfaces Σ , with $\partial\Sigma$ compact and non empty, Σ non compact, such that $\partial\Sigma \subset M$ and $\text{int}(\Sigma) \cap M = \emptyset$. Such surfaces Σ trap M in small regions of N which makes the geometry of M understandable. We will see several examples of this technique.

First I would like to explain how Σ can be obtained. Let Ω be a complete region of N , whose boundary is a good barrier for solving the least-area Plateau problem (this theory was developed by Meeks and Yau [M.-Y.]). This means $\partial\Omega = C$ is a 2-dimensional variety, smooth except along an analytic one dimensional variety, such that

- C is mean convex at the smooth points, i.e., the mean curvature vector at such points, points into Ω (the zero vector points into Ω), and
- at a non smooth point of C , the angle between the smooth faces of C , at the point, is less than or equal to π (measured in Ω).

Then Meeks and Yau proved that any smooth embedded 1-cycle Γ in $\partial\Omega$, that is null homologous in Ω is the boundary of a compact least area surface Σ_Γ in Ω , and Σ_Γ is smooth and embedded. The idea is to solve the Plateau problem in N by taking a limit of embedded surfaces with boundary Γ whose areas converge to the infimum of all possible areas. Then one checks that such a minimizing sequence can be constructed to stay in Ω . The mean convexity (and angle condition) implies that surfaces leaving Ω will increase area when crossing $\partial\Omega$. Then one works (considerably) to extract a subsequence that converges to a smooth embedded surface.

One can also use geometric measure theory to obtain Σ_Γ [Simon]. Again, we are assuming $C = \partial\Omega$ is a good barrier and $\Gamma \subset \partial\Omega$ a smooth one cycle (i.e. a collection of disjoint smooth Jordan curves). If Γ bounds an oriented 2-chain in Ω then Γ bounds a smooth embedded orientable surface Σ_Γ in Ω which minimizes area among all orientable 2-chains in Ω with boundary Γ . If Γ is a Z_2 -boundary in Ω then Γ bounds a smooth embedded least area surface in the same relative Z_2 -homology class. If Γ bounds an orientable (immersed) surface of genus n in Ω , then Σ_Γ can be chosen of genus at most n and of least area in its homotopy class.

Now we will discuss how the least area compact minimal surfaces Σ_Γ can converge to finite total curvature, non compact, minimal surfaces Σ .

We assume M orientable, A an end of M , $A \subset \partial\Omega$, and Γ a smooth Jordan curve on A , not homologous to zero in Ω . Let $A_1 \subset A_2 \subset \dots$ be an increasing sequence of compact submanifolds of A , which exhausts A , and $\partial A_i = \Gamma \cup \Gamma_i$. By our previous discussion of how one can solve the Plateau problem in Ω using geometric measure theory, we know there exists a least area smooth embedded surface Σ_i in Ω such that $\partial\Sigma_i = \Gamma \cup \Gamma_i$ and Σ_i is Z_2 -homologous to A_i rel ∂A_i . Since A_i is orientable and $\Sigma_i \cup A_i$ is Z_2 -homologous to zero, Σ_i is also orientable. Since Γ is not homologous to zero in Ω , Σ_i can be chosen connected.

Now we will show a subsequence of the Σ_i converges to a stable embedded minimal surface Σ with $\partial\Sigma = \Gamma$.

Observe that there are uniform local area bounds for the family Σ_i . For if $B \subset \Omega$ is a ball of radius r , ∂B transverse to Σ_i , then $\partial B \cap \Sigma_i$ is a 1-cycle on ∂B that bounds (mod 2) a 2-chain on ∂B of area at most $2\pi r^2$. Since Σ_i

minimizes area bounded by ∂A_i (in the Z_2 -homology class), we conclude $B \cap \Sigma_i$ has area at most $2\pi r^2$. Similarly if B is a ball centered at a point of $\partial \Sigma_i$, then the area of $\Sigma_i \cap B$ is at most the area of ∂B .

Now let $B(r) \subset \Omega$. By the curvature estimates of R. Schoen, after choosing a possibly smaller r , each component of $\Sigma_i \cap B(r)$ that intersects $B(r/2)$ can be expressed as a graph, of small gradient, over a plane P_i in $B(r)$, passing through the center of the ball, and P_i does not depend on the component. By the uniform area estimates, $\Sigma_i \cap B(r/2)$ contains a bounded number of components independent of i and hence there a a bounded number of associated graphs. Suppose for the moment that for every i , $\Sigma_i \cap B(r/2)$ contains one component. Choose a subsequence of the P_i to converge to a plane P through the center of the ball. Then the standard compactness theorem for minimal graphs implies a subsequence of the graphs $\Sigma_i \cap B(r/2)$ converge to a minimal graph over $P \cap B(r/2)$. When $\Sigma_i \cap B(r/2)$ has more than one component, we do the above argument to each component and obtain a (uniformly bounded) finite number of graphs over $P \cap B(r/2)$, to which the subsequence of $\Sigma_i \cap B(r/2)$ converges.

Now Ω has a countable basis of balls B_n where for every n and subsequence Σ_{i_λ} of Σ_i , the $\Sigma_{i_\lambda} \cap B_n$ have a convergent subsequence in B_n . Suppose the subsequence $\Sigma_{i_\lambda} \cap B_1$ converges in B_1 . Then the associated sequence of graphs in $B_2 \cap \Sigma_{i_\lambda}$ has a subsequence converging in $B_2 \cup B_1$. Continue in this manner from B_i to B_{i+1} and take a diagonal subsequence. This yields a subsequence of Σ_i that converges to a smooth minimal surface Σ , with $\partial \Sigma = \Gamma$. It is not hard to see that Σ is embedded and stable (since it's a limit of least area embedded surfaces). Also the boundary regularity theorem of Hardt and Simons implies Σ is smooth along Γ [H.-S.]. Finally, the theorem of Doris Fisher Colbrie yields that Σ has finite total curvature [F.C.].

In particular, this technique yields :

LEMMA 5.1. — *Let M be a properly embedded minimal surface in \mathbb{R}^3 with more than one end. Then there is an end of a catenoid or of a plane in the complement of M .*

Proof: Let Γ be a smooth Jordan curve on M that separates M into two non compact components, one of which we denote by A . M separates \mathbb{R}^3 into two connected components and Γ cannot be homologous to zero in both components; let Ω be a component such that Γ is not homologous to zero in Ω .

By our previous discussion, there is a finite total curvature embedded minimal surface Σ_Γ in Ω with $\partial\Sigma_\Gamma = \Gamma$. More precisely, $C = \partial\Omega = M$ is a minimal surface hence a good barrier for solving the Plateau problem. Let $A_i \subset A_{i+1}$ be an exhaustion of A with $\partial A_i = \Gamma \cup \Gamma_i$. Let Σ_i be an embedded minimal surface in Ω , Z_2 -homologous to A_i , with $\partial\Sigma_i = \partial A_i$. As before, a subsequence of Σ_i converges to Σ_Γ .

Now it may be that $\Sigma_\Gamma \subset M$ (if it touches M at one interior point, then since it's on one side of M at this point, it is contained in M). In this case, at least one end of M is of finite total curvature, so asymptotic to a planar end or catenoid end B . By the maximum principle at infinity, the distance between the ends of M is strictly positive. So B can be translated into Ω to be disjoint from M . Similarly, if $\text{int}\Sigma_\Gamma \subset \text{int}\Omega$, then the ends of Σ_Γ are a strictly positive distance from M so the conclusion of the lemma is clear.

There is a slight refinement of this lemma which is useful.

LEMMA 5.2. — *Let B be a ball in \mathbb{R}^3 and A_1, A_2 properly embedded minimal surfaces, non compact with $\partial A_1, \partial A_2$ smooth Jordan curves such that $B \cap (A_1 \cup A_2) = \partial A_1 \cup \partial A_2$ and $A_1 \cap A_2 = \emptyset$. Let Δ be the annulus on ∂B , bounded by $\partial A_1 \cup \partial A_2$ and let Ω be the connected component of $\mathbb{R}^3 - (A_1 \cup A_2 \cup \Delta)$ disjoint from B . Then there is an end of a plane or a catenoid in the interior of Ω . Moreover, ∂A_1 is the boundary of a smooth embedded surface Σ in Ω and outside of a larger ball \tilde{B} containing B , Σ is a finite total curvature minimal surface that separates ends of A_1 and A_2 , i.e. any path from A_1 to A_2 in $\mathbb{R}^3 - \tilde{B}$, meets Σ .*

Proof: Let $\Gamma = \partial A_1$ and consider Ω , with $\partial\Omega = A_1 \cup \Delta \cup A_2$. If $\partial\Omega$ were a good barrier for solving the Plateau problem then the construction of $\Sigma = \Sigma_\Gamma$ proceeds exactly as in the previous lemma. However ∂B is not mean convex with respect to Ω . One changes the Riemannian metric of \mathbb{R}^3

in a neighborhood of Δ in Ω so that $\partial\Omega$ is a good barrier in the new metric (cf. [M.-Y.] for the details). Then one proceeds as before. (This lemma remains true even if A_1 and A_2 are properly immersed; [M.-R.-2].)

6. THE ANNULAR END THEOREM AND THE STRONG HALFSpace THEOREM OF HOFFMAN-MEEKS

We can now give an idea of the proof of the following important result.

THEOREM 6.1 [H.-M.-3]. — *Let M be a properly embedded minimal surface in \mathbb{R}^3 , then M can have at most two annular ends of infinite total curvature.*

Sketch of Proof: Let A_1, A_2 and A_3 be distinct annular ends of M . It is not hard to find a ball B such that $B \cap (A_1 \cup A_2 \cup A_3) = \partial A_1 \cup \partial A_2 \cup \partial A_3$. Using the previous lemma, one traps one of the ends, A_1 say, between standard ends E_1, E_2 (each is a catenoid or planar end).

Now one proves that A_1 has finite total curvature. This is the difficult part of the proof. One proves the tangent plane to A_1 is never vertical outside of some compact set (then the Gaussian image of this subend is in a hemisphere hence has area less than 2π , so by the stability theorem of Barbosa-Do Carmo, A_1 is stable so of finite total curvature). To prove the tangent plane of A_1 is eventually never vertical one constructs foliations of the region between E_1, E_2 by minimal annuli whose boundaries are on $E_1 \cup E_2$. Then one studies the contact of A_1 with the foliation. The only contact points are of saddle type (by the usual maximum principle) and the topology of A_1 being simple one is able to show A_1 is eventually transverse to the foliation which implies there are no vertical tangent planes. Hence at most two annular ends of M can have infinite total curvature.

Now what about the two remaining annular ends, can they have infinite total curvature? This is unknown and it is one of the most important problems in this subject today. Meeks and I have proved :

THEOREM 6.2 [M.-R.-3]. — *Let M be a properly embedded minimal*

surface in \mathbb{R}^3 , with more than one end. If A is an annular end of M then (after a rotation of M in \mathbb{R}^3), either A is asymptotic to a horizontal plane (hence has finite total curvature) or x_3/A is a proper harmonic function. In particular, every such A is conformally the punctured disc $D^* = \{z \in \mathbb{C} / 0 < |z| \leq 1\}$.

COROLLARY 6.3. — *If M is a properly embedded minimal surface of finite topology and more than one end, then M has finite conformal type.*

COROLLARY 6.4. — *If M is a properly embedded minimal annulus then after a rotation of M , M intersects every horizontal plane in a simple closed curve.*

COROLLARY 6.5. — *An m -surface in \mathbb{R}^3 with a helicoidal type end has exactly one end.*

The strong halfspace theorem

There are complete immersed non planar minimal surfaces in a halfspace of \mathbb{R}^3 . Jorge and Xavier constructed such examples in a slab [J.-Xav.]. It is not known if such examples exist in a ball.

However, if the immersion is proper, Hoffman and Meeks proved this is not possible. They prove more :

THEOREM 6.6. — (the strong halfspace theorem [H.-M.-4]) *If M_1 and M_2 are disjoint properly immersed minimal surfaces in \mathbb{R}^3 then they are parallel planes.*

Proof : Assume first that M_2 is a plane (the (x, y) plane say) and M_1 is in the upper halfspace. After a vertical translation we can assume $\text{dist}(M_1, M_2) = 0$.

Let D_t be the disc of radius t in M_2 centered at the origin. Since M_1 is properly immersed, there is a $t > 0$ such that $\text{dist}(D_1, M_1) > t$. Choose $t < 1/4$. Let γ be the vertical upward translation of ∂D_1 , a distance t . By our choice of t and D_1 , $\gamma \cup \partial D_1$ is the boundary of a stable catenoid C_1 . For each $t > 1$, $\gamma \cup \partial D_t$ is the boundary of a stable catenoid C_t and C_{t_1} is above C_{t_2} when $1 \leq t_2 \leq t_1$. As $t \rightarrow \infty$, the C_t converge to the horizontal plane at height t , less the disc E in this plane bounded by γ ; figure 14.

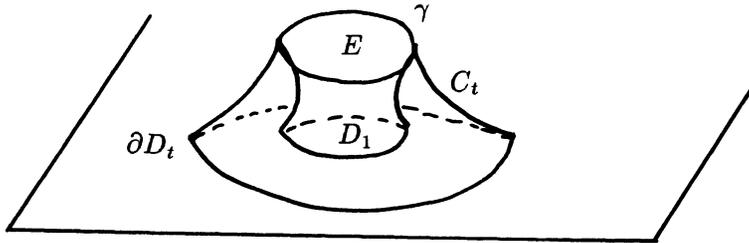


Figure 14

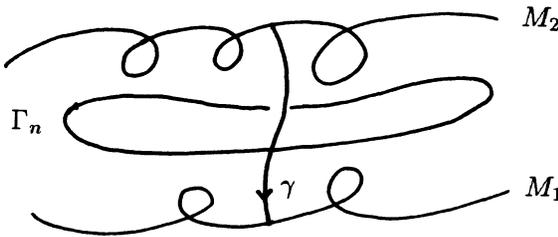


Figure 15

Clearly $E \cup C_t \cup D_t$ bounds a compact topological ball and the limit of these balls as $t \rightarrow \infty$ is the slab between M_2 and the horizontal plane at height t .

Now M_2 is properly immersed in \mathbb{R}^3 and $\text{dist}(M_2, M_1) = 0$ so there is a smallest t such that $C_t \cap M_2 \neq \emptyset$. But then C_t is on one side of M_2 at this point of first contact so $C_t = M_2$ by the maximum principle. This proves the strong halfspace theorem in the special case that M_2 is a plane.

Now suppose M_1 and M_2 are disjoint and properly immersed. We will find a plane between M_1 and M_2 so by what we have just proved M_1 and M_2 are planes too.

Let Ω be the connected region of \mathbb{R}^3 whose boundary is contained in

$M_1 \cup M_2$ and the boundary contains points of both M_1 and M_2 . Notice that $\partial\Omega$ is a good barrier for solving the Plateau problem.

Let γ be an arc in Ω joining a point of M_1 to a point of M_2 and let Γ_n be Jordan curves in Ω such that the linking number of Γ_n and γ is one and Γ_n is in the complement of the ball of radius n centered at a fixed point of γ ; figure 15.

Let Σ_n be a least area smooth immersed minimal surface in Ω with $\partial\Sigma_n = \Gamma_n$. As in V, a subsequence of the Σ_n converge to a complete stable minimal surface $\Sigma \subset \Omega$. Σ is non empty since each Σ_n intersects γ by our linking number restriction.

By R. Schoens theorem Σ is a plane. Clearly if Σ ever touched M_1 or M_2 then they would be planes too. This completes the proof of the strong halfspace theorem.

7. DOUBLY PERIODIC MINIMAL SURFACES

We call a minimal surface in \mathbb{R}^3 periodic if it is connected and invariant by a non trivial discrete group G of isometries that acts freely on \mathbb{R}^3 . In fact we study the quotient minimal surface in \mathbb{R}^3/G . In fact, all connected properly embedded minimal surfaces M in \mathbb{R}^3/G arise this way, since, by the strong halfspace theorem the lift of M to \mathbb{R}^3 is a connected minimal surface invariant by G (assuming M not planar). Notice that this implies that $\pi_1(M) \rightarrow \pi_1(\mathbb{R}^3/G)$ is surjective under our hypothesis on M .

Our main result relates the topology of M to its total curvature $C(M)$.

THEOREM 7.1 (the finite total curvature theorem, [M.-R.-4]). — *Let M be a properly embedded minimal surface in a non simply connected complete flat 3-manifold. Then M has finite topology if and only if $C(M)$ is finite. When $C(M)$ is finite, we have the formula*

$$C(M) = 2\pi(\mathcal{X}(M) - W(M)) ,$$

where $W(M)$ is the total winding number of the ends of M (we define this later). When $N = T^2 \times \mathbb{R}$, $W(M) = 0$.

Notice that one needs to assume N not simply connected; the helicoid in \mathbb{R}^3 has infinite total curvature and finite topology.

I would like to discuss the proof of this theorem (at least for doubly periodic surfaces) and give some applications. A complete flat 3-manifold is finitely covered by T^3 , $T^2 \times \mathbb{R}$ or \mathbb{R}^3/S_θ , S_θ a screw motion around the x_3 -axis, followed by rotation by θ about this axis. So our theorem concerns $T^2 \times \mathbb{R}$ and \mathbb{R}^3/S_θ (doubly and singly periodic surfaces).

Now let G be generated by two independent translations so that $\mathbb{R}^3/G = T \times \mathbb{R}$, T a flat 2-torus. Let $x_3 : T \times \mathbb{R} \rightarrow \mathbb{R}$ denote the *third* coordinate function, $T_t = T \times \{t\}$ the level set of x_3 at height t . We let $D^* = \{0 < |z| \leq 1\}$ be the punctured disc in \mathbb{C} .

LEMMA 7.2. — *Let A be an annulus diffeomorphic to D^* and $X : A \rightarrow T \times \mathbb{R}$ a proper minimal immersion of A . Then A contains a proper subannulus A' which can be conformally parametrized by D^* . In this parametrization $x_3/A'(z) = c \ln|z|$ where c is constant.*

Proof: Let $X_3 = x_3 \circ X : A \rightarrow \mathbb{R}$; X_3 is a proper map. Since A has one end, X_3 is bounded from above or below but not both, so assume X_3 is bounded below. After translating $X(A)$ vertically downward, we can make the boundary of the annulus have negative x_3 -coordinate and X_3 has 0 as a regular value. Hence $\Delta = X_3^{-1}(-\infty, 0]$ is a compact smooth submanifold of A . Δ contains exactly one component containing ∂A and the other components have x_3 -coordinate zero. The maximum principle for the harmonic function X_3 implies Δ is connected and by elementary topology Δ is an annulus, and $A' = X_3^{-1}[0, \infty)$ is a proper subannulus of A .

The function X_3/A' is a proper nonnegative harmonic function with zero boundary values. It is an easy exercise in elementary complex analysis to prove that A' can be conformally parametrized by D^* and $X_3 = c \ln|z|$ for some constant c .

Now let M be a properly immersed minimal surface in $T \times \mathbb{R}$, of finite topology. By the above lemma, each annular end of M is conformally D^* so M has finite conformal type. We want to know M is of finite total curvature

when embedded so we may as well assume M is orientable (by passing to a two sheeted covering). Then the Gauss map $g : M \rightarrow S^2$ can be defined and is conformal; two liftings of a point of M to \mathbb{R}^3 differ by a translation that leaves the oriented unit normal vector field to the lifted surface, invariant.

Our technique to prove M has finite total curvature is to prove the punctures of the annular ends of M are removable singularities of the Gauss map g . Since the total curvature is the area of the spherical image of M by g , this suffices. In general one shows the puncture is not an essential singularity by trapping an end A in a region of space which controls the values of g . If g misses to many values near the puncture then the singularity is removable. Now we can do this.

THEOREM 7.3. — *Let $X : A \rightarrow T \times \mathbb{R}$ be a proper minimal embedding of D^* . Then A has finite total curvature.*

Proof : By the previous lemma 7.2, we can suppose $A = D^*$ and $X_3(z) = c \ln|z|$; we shall identify A with $X(A)$. We take $C < 0$ so that $X_3 \geq 0$ on A . Let $C_t = A \cap T_t$; each C_t is a simple closed curve. The proof divides into two cases : C_0 a generator of $\pi_1(T_0)$ or not. We shall consider the first case here and we refer the reader to [M.-R.-1] for the second case.

C_0 generates a cyclic subgroup G in $\pi_1(T \times \mathbb{R})$. Let $p : \widetilde{T \times \mathbb{R}} \rightarrow T \times \mathbb{R}$ be the Riemannian covering space such that $p_*\pi_1(\widetilde{T \times \mathbb{R}}) = G$. $\widetilde{T \times \mathbb{R}}$ is isometric to $(S^1 \times \mathbb{R}) \times \mathbb{R}$ and the generator of $\pi_1(T \times \mathbb{R})/G$ acts naturally on $H = p^{-1}(x_3^{-1}[0, \infty))$ as a translation. For notational convenience, let A also denote a lifting of A to H . Since ∂A is compact and $S^1 \times \mathbb{R} = \partial H$ is non compact, we can choose a closed geodesic α in ∂H such that $\alpha \cap \partial A = \emptyset$. Choose a covering transformation σ such that α is contained in the interior of the compact annulus Δ with boundary ∂A and $\sigma(\partial A)$; figure 16.

Let $\Omega \subset \widetilde{T \times \mathbb{R}}$ be the component of $H - (A \cup \sigma A)$ whose boundary contains $A \cup \sigma A$, and Ω_t the points of Ω at height at most t . Notice that Ω_t is a good barrier for the Plateau problem; its boundary consists of four minimal surfaces meeting at angles less than or equal to π .

Let α_t be a Jordan curve in the interior of the smooth annulus of $\partial\Omega_t$ at height t , such that α_t is homotopic to α . Let Σ_t be a least area embedded

smooth surface with $\partial\Sigma_t = \alpha \cup \alpha_t$, $\Sigma_t \subset \Omega_t$, figure 16. First observe that Σ_t is orientable : this will follow by showing Σ_t separates Ω_t . If not then there is a simple closed curve δ in Ω_t which intersects Σ_t transversely in one point. But $\pi_1(\Omega_t)$ is generated by $\pi_1(\partial A)$ hence δ is homotopic to a multiple of ∂A and ∂A has zero intersection number with Σ_t . Since the Z_2 -intersection numbers are well defined in homotopy classes, this is impossible and Σ_t is orientable.

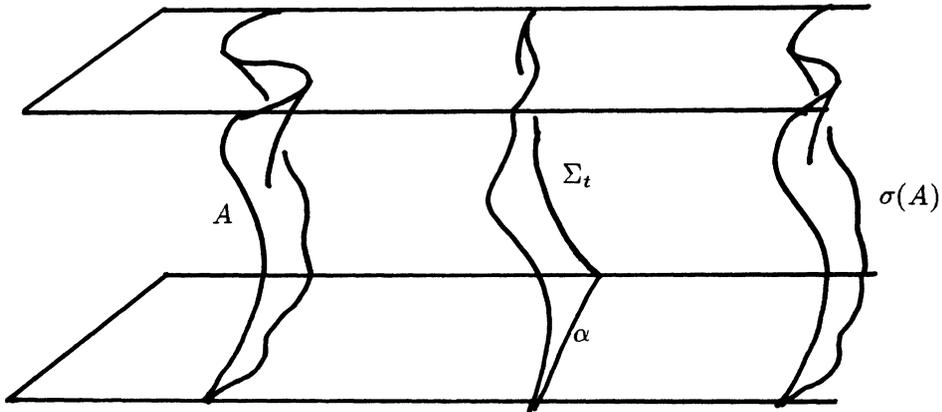


Figure 16

By 5, a subsequence of Σ_t converge to a smooth embedded stable surface Σ , $\partial\Sigma = \alpha$, $\Sigma \subset W$. By the usual maximum principle $\text{int}\Sigma \subset \text{int}W$.

We now prove Σ is part of a plane. Since α is the quotient of a straight line in \mathbb{R}^3 , we can extend Σ by Schwarz reflection R_α to a properly embedded minimal surface $\tilde{\Sigma} \subset T \times \mathbb{R}$. Note that since $\sigma(\Omega) \cap \Omega = \sigma(A)$ and $\Sigma \cap \partial\Omega = \alpha$, $\sigma\Sigma \cap \Sigma = \phi$. Let $R_{\sigma\alpha}$ be Schwarz reflection about $\sigma\alpha$ (i.e. rotation by π about $\sigma\alpha$). We claim that $\tilde{\Sigma}$ and $\Sigma' = \sigma\Sigma \cup R_{\sigma\alpha}(\sigma\Sigma)$ are two properly embedded disjoint minimal surfaces. Note that $R_{\sigma\alpha} \circ \sigma = R_{\tilde{\alpha}}$, and $R_{\tilde{\alpha}} \cdot R_{\sigma\alpha} \cdot \sigma = id$, where $\tilde{\alpha}$ is the geodesic on ∂H halfway between α and $\sigma\alpha$. Hence if $\tilde{\Sigma} \cap \Sigma' \neq \phi$, then $R_\alpha\Sigma \cap R_{\sigma\alpha}(\sigma\Sigma) \neq \phi$. Composing the last inequality with $R_{\tilde{\alpha}}$ yields $R_{\tilde{\alpha}}R_\alpha(\Sigma) \cap \Sigma \neq \phi$. But $R_{\tilde{\alpha}}R_\alpha = \sigma$ so

$\sigma \Sigma \cap \Sigma \neq \phi$, a contradiction.

Now that $\tilde{\Sigma}$ and Σ' are disjoint the strong halfspace theorem says their lifts to \mathbb{R}^3 are planes, hence Σ and $\sigma\Sigma$ are parallel flat annuli in H .

Let $P(\theta)$ be a flat annulus which contains α and which makes an angle θ with the horizontal plane ∂H . Choose θ sufficiently small so that $P(\theta)$ intersects the region bounded by Σ and $\sigma\Sigma$ in a compact set and $P(\theta)$ intersects A transversally in a smooth curve. This is possible since A intersects ∂H transversally in a single curve.

Consider the foliation of $T \times \mathbb{R}$ by planes parallel to $P(\theta)$ (flat annuli in fact). Notice that each leaf intersects $P(\theta)$ in a compact set. This foliation is defined by the level sets of a linear function whose restriction to A is a proper harmonic function. Hence this harmonic function has no critical points on A above $P(\theta)$. In particular, the normal to $P(\theta)$ is never attained as a normal vector to the part of A above $P(\theta)$.

Since θ can vary in an interval, the Gauss map on A misses a curve of values, hence the puncture is not an essential singularity and A has finite total curvature.

Now we can analyse the geometry of the ends of an m -surface of $T \times \mathbb{R}$ of finite topological type; we will see they converge geometrically to flat annuli. Before proving this we analyse immersed finite total curvature surfaces in $T \times \mathbb{R}$.

THEOREM 7.4. — *Let M be a properly immersed minimal surface in $T \times \mathbb{R}$, of finite total curvature. Let A_1, \dots, A_ℓ be the ends of M with vertical limiting normal vectors and let n_i be the branching order of the Gauss map at the end A_i . Then $C(M) = 2\pi(\mathcal{X}(M) - \sum_{\lambda=1}^{\ell} n_\lambda)$. In particular, if M has no horizontal ends, then $C(M) = 2\pi\mathcal{X}(M)$.*

Proof: When M is nonorientable and we pass to the oriented two sheeted cover of M , then all the terms in above formula multiply by two. This is obvious for $C(M)$ and $\mathcal{X}(M)$; each end of M lifts to two ends in the cover. Hence we can assume M is orientable.

Let $M_t = M \cap (T \times [-t, t])$. By Gauss-Bonnet, we have :

$$C(M) = \lim_{t \rightarrow \infty} \int_{M_t} K = \lim_{t \rightarrow \infty} (2\pi \mathcal{X}(M_t) - \int_{\partial M_t} \kappa_g).$$

For large values of t , $\mathcal{X}(M_t) = \mathcal{X}(M)$ so we must calculate $\int_{\partial M_t} \kappa_g$.

First consider a component C_t of ∂M_t that is on an end E having a nonvertical limiting normal vector v . We shall prove $\int_{C_t} \kappa_g \rightarrow 0$ as $t \rightarrow \infty$. Let \vec{a} be a horizontal unit vector orthogonal to v ; since v is not vertical, there are exactly two such vectors. Choose the orientation of \vec{a} so that C'_t converges to \vec{a} as $t \rightarrow \infty$ (C'_t is oriented by M_t and ' denotes derivative with respect to arc length).

Let $d\vec{a}$ be the closed one form defined by orthogonal projection on \vec{a} (the line parallel to \vec{a}). We have $\int_{C_{t_1}} d\vec{a} = \int_{C_{t_2}} d\vec{a}$ since $C_{t_1} - C_{t_2}$ bounds on E . As $t \rightarrow \infty$, $C'_t \rightarrow \vec{a}$ hence $\int_{C_t} ds$ converges to $\int_{C_t} d\vec{a}$. In particular, the lengths of the C_t are uniformly bounded.

Let X be the conormal vector field along C_t , i.e. X is tangent to M , $\langle X, C'_t \rangle = 0$, $|X| = 1$ and X points into M_t . Let \vec{a}^\perp be the unit normal vector to \vec{a} , tangent to T_t and whose direction is C''_t , when $C''_t \neq 0$.

We have

$$\begin{aligned} \kappa_g(C_t) &= \langle C''_t, X \rangle = |C''_t| \cos(\sphericalangle(C''_t, X)) \\ &= \kappa \cos(\sphericalangle(\vec{a}^\perp, X) + \varepsilon), \end{aligned}$$

where κ is the curvature of C_t , viewed as a planar horizontal curve and $\varepsilon \rightarrow 0$ as $t \rightarrow \infty$.

Now compute κ by thinking of C_t as a planar section of M . Let P be the plane at $C_t(s)$ generated by the normal n to M at $C_t(s)$ and $C'_t(s)$. Let $\kappa_n(s)$ be the normal curvature, i.e. the curvature of the curve $P \cap M$ at $C_t(s)$.

We have $\kappa_n(s) = \kappa \cos \psi$ where ψ is the angle between $C''_t(s)$ and n . Since the limiting normal is not vertical, $\cos \psi$ is bounded away from zero. Hence if $\kappa_n \rightarrow 0$ as $t \rightarrow \infty$ then so does κ and κ_g .

Since M is minimal, the principal curvatures κ_1, κ_2 of M are equal in modulus. The normal curvatures are between κ_1 and κ_2 so it suffices to

prove

$$K = \kappa_1 \kappa_2 \longrightarrow 0 \quad \text{as } t \longrightarrow \infty .$$

In a conformal parametrization of the end by D^* the induced metric is $ds = \lambda|dz|$ where $\lambda = \frac{|f(z)|}{2}(1 + |g(z)|^2)$, f a non vanishing holomorphic function on D^* , g the Gauss map. The curvature $K = \frac{-\Delta \ell n \lambda}{2\lambda}$. Since $\ell n|f|$ is harmonic in D^* and g extends holomorphically to 0, we have $\Delta \ell n \lambda = \Delta \ell n(1 + |g|^2)$ is bounded in a neighborhood of 0. The metric ds is complete at 0 so $\lambda \rightarrow \infty$ as $|z| \rightarrow 0$, and this proves $K \rightarrow 0$, hence $\int_{C_t} \kappa_g ds \rightarrow 0$ as $t \rightarrow \infty$.

Now consider an end A of M with a vertical limiting normal vector v . By lemma 7.2, we can conformally parametrize A by D^* so that $x_3 = C \ell n|z|$. The Gauss map has a zero or pole at 0 of order n , the branching order of A .

We can assume $g(z) = z^n +$ higher order terms near 0. As z goes once around the circle $|z| = r$ counterclockwise, the normal vector to A along C_t goes n times around the vertical vector v , always turning in the same sense when r is small. Hence the normal vector to the curve C_t in T_t , turns monotonically counterclockwise, n times around the origin.

Let κ be the planar curvature of C_t (in T_t). By the last paragraph, $\kappa > 0$ and $\int_{C_t} \kappa ds = 2\pi n$. Now $\kappa_g = \kappa \cos \phi$, ϕ the angle between the conormal to C_t in M and the horizontal. We have $\phi \rightarrow 0$ as $t \rightarrow \infty$ hence $\int_{C_t} \kappa_g ds \rightarrow 2\pi n$. This completes the proof of theorem 7.4.

Remark : We deduce from the above argument, that if the end A is embedded, it can not have a vertical limiting normal vector v . For if v is vertical, the curves C_t have positive curvature κ and $\int_{C_t} \kappa ds = 2\pi n$. Clearly this means C_t is a convex curve, null homotopic on T_t . Hence A lifts to a finite total curvature embedded end in \mathbb{R}^3 which must be a catenoid (asymptotically). Clearly this can not be embedded in $T \times \mathbb{R}$.

THEOREM 7.5. — *Let A be a properly embedded minimal annular end in $T \times \mathbb{R}$. Then A is asymptotic to a flat cylinder. Moreover two distinct annular ends of an m -surface in $T \times \mathbb{R}$, converge to distinct flat cylinders.*

Proof: We know we can assume A is parametrized by D^* and $X_3 = c \ell n|z|$

with $c < 0$. By the analysis in the proof of theorem 7.4, we may assume that for every small $\varepsilon > 0$, there exists a $T > 0$ such that for $t > T$, each curve $C_t = A \cap T_t$ is contained in the interior of an ε -tubular neighborhood B_t of a geodesic; $B_t \subset T_t$. Fix $\varepsilon > 0$, and assume, after possibly translating A downward, that C_t has this property for $t \geq 0$. Let α_t and β_t be the boundary curves of B_t . For small t , it is clear there exists a unique flat annulus $F_t \subset T \times \mathbb{R} - A$ with boundary $\alpha_0 \cup \alpha_t$. We shall check that such an F_t exists for all t and varies continuously with t . Clearly the set of t for which F_t exists is open, since F_t and $A \cap (T \times [0, t])$ are compact. Also, $\partial F_t \cap \partial A = \phi$ so the maximum principle implies the limit of such F_t is also an example.

In the same manner, we define flat annuli E_t with boundary $\beta_0 \cup \beta_t$, disjoint from A . A subsequence of E_t converges to a flat annulus E , with $\partial E = \alpha_0$, $E \cap A = \phi$ (by the maximum principle). Similarly F_t has a subsequence converging to a flat annulus F , $\partial F = \beta_0$, $F \cap A = \phi$ and $E \cap F = \phi$. Hence E and F are parallel at a distance ε and A is between E and F . Now do the same argument at a height such that the C_t are within $\varepsilon/2$ of a geodesic on T_t . Letting $\varepsilon \rightarrow 0$ we get the desired limit flat annulus.

By the maximum principle at infinity, it follows that distinct ends converge to distinct flat annuli.

7.6. Global topological and geometrical properties

Recall that $T \times \mathbb{R} = \mathbb{R}^3/G$ has a commensurable lattice if G contains two linearly independent vectors of equal length.

THEOREM 7.6. — *Let M be an m -surface of $T \times \mathbb{R}$ of finite topological type (M not flat). Then :*

1. *If M is orientable, then M separates $T \times \mathbb{R}$. In this case, the number of top ends, as well as the number of bottom ends of M is even. In particular M has at least four ends.*

2. *If M is nonorientable, then the number of top ends, as well as the number of bottom ends, is odd. In particular, whether M is orientable or not, the number of ends is even.*

3. The top ends of M are parallel to the bottom ends of M if and only if the subgroup of $H_1(T \times \mathbb{R})$ generated by the loops on the ends of M is cyclic. If the ends are parallel then the number of top ends equals the number of bottom ends. In particular, by 1, if M is orientable and has parallel ends, then the number of ends is a multiple of four.

4. If the ends of M are not parallel, then they are vertical and $T \times \mathbb{R}$ has a commensurable lattice.

Proof : Assume M orientable and let \widehat{M} be the connected lifting of M to \mathbb{R}^3 . Let G be the translation group defining M . if $\sigma \in G$ then $\sigma\widehat{M} = \widehat{M}$ by the strong halfspace theorem. \widehat{M} separates \mathbb{R}^3 into two connected components A and B and σ conserves orientation so $\sigma A = A$. Hence M bounds A/G in $T \times \mathbb{R}$ and M separates $T \times \mathbb{R}$.

For t large, we know that $M \cap T_t$ consists of a finite number of pairwise disjoint simple closed curves C_1, \dots, C_n and each C_i is approximately a geodesic. Here n is the number of top ends. Similarly, $M \cap T_t = D_1 \cup \dots \cup D_m$ for $t < 0$, $|t|$ large, each D_j an almost geodesic and m equals the number of bottom ends. Since M separates $T \times \mathbb{R}$ into two components, A and B say, each C_i has two sides on T_t , one in A , the other in B . Hence both n and m are even. This proves 1. We leave 2 to the reader, or refer to [M.-R.-1].

Let P be a flat annulus parallel to the limiting top ends and Q a flat annulus representing the limit of the bottom ends. Let \vec{a} and \vec{b} denote the limiting directions of C_i and D_j respectively, $|\vec{a}| = |\vec{b}| = 1$. Let X denote the conormal vector field to ∂M_t ; X is tangent to M , $|X| = 1$, $X \perp \partial M_t$, and X points upward. Let \vec{v} be the upward unit vector field tangent to P and normal to \vec{a} . Similarly let \vec{w} be the unit field tangent to Q , normal to \vec{b} and pointing upward.

The flux of \vec{v} across a curve C_j is $\int_{C_j} \langle \vec{v}, X \rangle ds$. As $t \rightarrow \infty$, X converges to \vec{v} , C_j converges to a geodesic A_j . Hence the flux of \vec{v} across ∂M_t for t large, is

$$\sum_{j=1}^n \int_{|A_j|} \langle \vec{v}, \vec{v} \rangle ds = n|A_1| .$$

Similarly, the flux of \vec{w} across D_j is $\int_{D_j} \langle \vec{w}, \vec{w} \rangle |B_j|$, where B_j is the

limiting geodesic of D_j .

Since \vec{v} is the gradient of a coordinate function, which is harmonic on M , the flux of \vec{v} across $\partial(M \cap T \times [-t, t])$ is zero. hence $n|A_1| = \langle \vec{v}, \vec{w} \rangle m|B_1|$. In particular $n|A_1| \leq m|B_1|$ and equality holds if and only if $\vec{v} = \vec{w}$. Now turn M upside down to conclude $m|B_1| \leq \langle \vec{v}, \vec{w} \rangle n|A_1|$. Hence $\vec{v} = \vec{w}$ and $n|A_1| = m|B_1|$.

If \vec{v} is not vertical then there is a unique horizontal direction \vec{a} normal to \vec{v} . Hence $\vec{a} = \vec{b}$ when \vec{v} is not vertical and the top ends are parallel to the bottom ends.

If the ends are all parallel, then the subgroup of $H_1(T \times \mathbb{R})$ generated by the ends is the cyclic subgroup generated by A_1 . If the ends generate a cyclic subgroup with generator A then $\vec{a} = \vec{b}$ and $\vec{v} = \vec{w}$ so the ends are parallel.

If the ends are not parallel then they are vertical and $n|A_1| = m|B_1|$. Since \vec{a} and \vec{b} are independent, the vectors $n|A_1|\vec{a}$ and $m|B_1|\vec{b}$ are independent and of equal length. So the lattice is commensurable. This completes the proof of theorem 7.6.

Now it is not hard to give necessary conditions for a given doubly periodic minimal surface to have nonparallel ends which forces the ambient space to have a commensurable lattice. We leave the proof to the reader or refer to [M.-R.-1].

THEOREM 7.7. — *Let $M \subset T \times \mathbb{R}$ be a non flat m -surface of finite topology. Then the ends are not parallel if 1, 2 or 3 holds :*

1. M is orientable and the number of ends is not a multiple of four
2. M is a planar domain
3. $\chi(M)$ is odd.

7.8. The sum of finite total curvature minimal surfaces (minimal herissons).

Let M_1, M_2 be finite total curvature complete non planar minimal surfaces in \mathbb{R}^3 with Gauss maps g_1, g_2 . Let $p_1, \dots, p_n, q_1, \dots, q_m$ be the punctures of M_1, M_2 respectively and $\overline{M}_1, \overline{M}_2$ the compactified Riemann

surfaces. If one fixes a unit vector $z \in S^2$, one can add (in \mathbb{R}^3) all points in \mathbb{R}^3 having z as normal. As z varies in S^2 this yields a complete (branched) minimal surface or a point. More precisely, Rosenberg and Toubiana have proved :

THEOREM 7.8 [R.-T.-2]. — *The set*

$$M_1 + M_2 = \left\{ \sum_{x \in g_1^{-1}(z)} x + \sum_{y \in g_2^{-1}(z)} y / z \in S^2 - W \right\}$$

is a complete minimal surface in \mathbb{R}^3 (or a point) of total curvature -4π . Here W is some subset of $g_1\{p_1, \dots, p_n\} \cup g_2\{q_1, \dots, q_m\}$.

The normal vector to $M_1 + M_2$ at the point $\sum_{x \in g_1^{-1}(z)} x + \sum_{y \in g_2^{-1}(z)} y$ is z .

Thus $M_1 + M_2$ is naturally parametrized by $S^2 - W$; denote this parametrization by \hat{g} . If $M_1 + M_2$ is not a point, then \hat{g} is a conformal injection, which explains the total curvature -4π of $M_1 + M_2$.

The (possible) branch points of $M_1 + M_2$ are geometric branch points, however the Weierstrass data of $M_1 + M_2$ is meromorphic at these branch points; the $\hat{\omega}$ of $M_1 + M_2$ vanishes at the branch points. Notice these points are quite distinct from the branch points of the Gauss map; in general, \hat{g} is injective where $\hat{\omega}$ vanishes.

The sum operation is very useful for detecting symmetries in a surface M . For example Rosenberg and Toubiana have proved :

THEOREM 7.9 [R.-T.-2]. — *Let M be a complete finite total curvature minimal surface in \mathbb{R}^3 . if all the ends of M are asymptotic to planes (planar ends) then $M + M$ is a point.*

The idea of the proof is simple. At a planar end of M , the points having a fixed normal direction (near the limiting normal) are distributed in space so as to have the same barycenter (like the roots of unity). So a planar end puncture becomes a regular point in $M + M$.

Since $T \times \mathbb{R}$ is an abelian group under addition and the Gauss map is invariant under translation, the sum $M_1 + M_2$ is also defined in $T \times \mathbb{R}$ and has total curvature -4π or 0.

Meeks and Rosenberg have proved.

THEOREM 7.10 [M.-R.-1]. — *Let M be a finite total curvature complete immersed minimal surface in $T \times \mathbb{R}$. Then*

1. *If the ends of M converge to parallel flat annuli, then $M + M$ is a point.*

2. *If M is embedded and the ends of M are not parallel, then $M + M$ is a Scherk surface.*

Finally, we apply this theorem to obtain :

THEOREM 7.11. — *Suppose M is an m -surface of finite topology in $T \times \mathbb{R}$ and $T \times \mathbb{R}$ has an incommensurable lattice. Then*

1. *$M + M$ is a point.*

2. *If M has genus one and four parallel ends (e.g. a Karcher saddle), then after a translation of M (so that a zero of Gaussian curvature occurs at the origin), the order two points in the group $(\mathbb{R}^2/G) \times \mathbb{R}$ are the zeros of the Gaussian curvature of M . In this case M separates $T \times \mathbb{R}$ into two isometric components.*

8. SINGLY PERIODIC MINIMAL SURFACES

We have a wealth of beautiful examples of singly periodic m -surfaces. The helicoid is the easiest to grasp : take a horizontal line ℓ , passing through the x_3 -axis; then rotate ℓ with constant velocity while rising vertically with constant velocity. This surface \widetilde{M} is invariant under screw motions S_θ and for a fixed θ , $M = \widetilde{M}/S_\theta$ is conformally a two-punctured sphere of finite total curvature -2θ . In \mathbb{R}^3/S_θ , M has two annular ends, each a helicoid end. Notice that M no longer has a well defined Gauss map. The Gauss map g of \widetilde{M} induces a multivalued meromorphic map on M where distinct determinations of its values differ by λ^m where $\lambda = e^{i\theta}$: if p and q are points of \mathbb{R}^3 with $S_\theta^m(p) = q$, $p, q \in \widetilde{M}$, then the normal vectors to \widetilde{M} at p and q differ by rotation about the x_3 -axis by $m\theta$, hence their stereographic projections to the horizontal complex plane differ by multiplication by λ^m .

From the point of view of the Weierstrass representation of \widetilde{M} in \mathbb{R}^3 , \widetilde{M} is the conjugate surface of the catenoid. On $\mathbb{C}^* = \mathbb{C} - (0)$, the data : $g(z) = z$, $\omega_\tau(z) = e^{i\tau} \frac{dz}{z^2}$, defines a complete minimal surface \widetilde{M}_τ for each real τ ; the catenoid is $\tau = 0$ and the helicoid $\tau = \pi/2$. The surfaces \widetilde{M}_τ are not embedded for $0 < \tau < \pi/2$, however each \widetilde{M}_τ has two annular ends and they are embedded. Each end is a *helicoid-catenoid* type end. The intersection of M_τ with a large vertical cylinder of radius R , centered at the x_3 -axis, consists of two helices (like a barber pole). As $R \rightarrow \infty$, the helices rise on the cylinder like $\ell n R$; so they look like helicoids and catenoids (actually one helice rises and the other descends). We will see later that when \widetilde{M} is an embedded singly periodic surface, there are no annular ends of this type. The number of ends will be even and half of them would rise as $R \rightarrow \infty$, while the other half would descend. So \widetilde{M} couldn't be embedded. Hoffman and Wei have shown that one can add a handle to the helicoid in a periodic manner [H.-Wei.], figures 17-a,b.

Another singly periodic example is the conjugate surface of Scherks doubly periodic m -surface. In terms of Weierstrass data in $T \times \mathbb{R}$, $g(z) = z$ and $\omega(z) = \frac{dz}{z^4-1}$, parametrizes Scherk's 1'st surface by the sphere punctured at the four roots of unity. The reader can easily check that x_1 and x_2 are multi-valued, and x_3 single valued. The data for the conjugate surface (Scherks' second surface) is $g(z) = z$, $\omega = \frac{i dz}{z^4-1}$, also modelled on S^2 minus the fourth roots of unity. Now x_3 has a period and x_1, x_2 are single valued; the surface is invariant by a vertical translation T . There are four annular ends that are the quotient, by T , of vertical ends, asymptotic to planes. Ends of this type are called Scherk type ends, figure 18-a.

Karcher has shown that one can construct singly periodic m -surfaces of this nature with $2n$ Scherk type ends, for any $n \geq 2$. Moreover, he is able to deform these surfaces to singly periodic m -surfaces, invariant by screw motions S_θ , so that the Scherk type ends become helicoid type ends [K.-2], figure 18-b . He does this with the generalized Weierstrass representation we develop in this chapter.

We discuss one more example, the Riemann example. This surface is invariant by a translation T (not a vertical translation) and the horizontal

sections $x_3 = \text{constant}$ are circles and lines. The ends in \mathbb{R}^3 are asymptotic to horizontal planes (located at the heights whose sections are lines), figure 19. The simplest orientable quotient of this is a two punctured torus of total curvature -8π . The Riemann examples form a one parameter family and the conjugate surface of a Riemann example is also a Riemann example (this is a good exercise). We refer the reader to [H.-M.-5] for an excellent discussion of these surfaces. Callahan, Hoffman and Meeks have generalized the Riemann examples [C-H-M], figure 20-a. Also Hoffman and Wei have shown that one can add a handle to Riemann's surface (one handle between every other pair of planar ends) to obtain a singly periodic m -surface, a surface of genus one with three punctures [H.-Wei], figure 20-b.

8.1. The finite total curvature theorem

Our theorem 7.1 states that an M surface in \mathbb{R}^3/S_θ is of finite topology if and only if it is of finite total curvature. I will briefly outline the structure of the proof, and refer the courageous reader to [M.-R.-4] for the details.

Let $M \subset \mathbb{R}^3/S_\theta$ be an m -surface of finite topology. The problem is to show the ends (topologically annular) are of finite total curvature. This is done by trapping an end of M between *standard ends*; i.e. two ends of finite total curvature whose geometry one understands. Then, using foliations by stable minimal annuli of the region between the standard ends (that trapped the end of M) one proves the end of M is stable hence of finite total curvature.

The first part of the proof requires an understanding of the finite total curvature annular ends. In II, we explained the geometry of embedded finite total curvature ends in \mathbb{R}^3 using the Weierstrass representation and the fact that the Weierstrass data (g, ω) extends meromorphically to the puncture. An annular end A in $\mathbb{R}^3/S_\theta = N$ has a multi-valued Gauss map so the first thing we need to know is the existence of a limit tangent plane of A at infinity. Assuming A has finite total curvature Huber's theorem tells us A can be conformally parametrized by D^* . Now we have the Picard-type theorem (whose proof uses elementary complex analysis).

THEOREM 8.2 [M.-R.-4]. — *Let g be a multi-valued meromorphic map*

on D^* : $g = \widehat{g}(\exp^{-1})$, with $\widehat{g}(z + 2\pi i) = \lambda \widehat{g}(z)$ for $z \in \{x + iy/x \leq 0\}$, and $|\lambda| = 1$. If the area of the image of g (i.e. the restriction of g to D^* slit along a radial line), counted with multiplicity, is finite, then g extends continuously to the origin.

8.3 The generalized Weierstrass representation

Now we use this result to obtain a Weierstrass representation for A (meromorphic on D) as follows. We can assume $\lambda \neq 1$ since this is the usual Weierstrass representation. Then the limiting value of g is 0 or ∞ since it is fixed by multiplication by λ ; so assume $g(0) = 0$. Let $\theta = 2\pi a$ with $0 < a < 1$. Clearly the map $z^{1-a}g(z)$ is bounded in a neighborhood of 0, so $g(z) = z^{a-1}h(z)$ with h holomorphic in a neighborhood of 0. Hence $\frac{dg}{g}$ is a well defined meromorphic one form on D^* that extends meromorphically to 0. One obtains the multi-valued g from this form by $g = \exp(\int \frac{dg}{g})$.

Notice that the third coordinate function x_3 is defined, up to a constant on N so dx_3 is well defined on N . Let $\eta = dx_3 + i(*dx_3)$. It is easy to see that η is meromorphic on D^* and extends meromorphically to 0.

We then can take as Weierstrass data on A the pair of one forms $(\frac{dg}{g}, \eta)$, which extend meromorphically to the puncture. In general, we have :

THEOREM 8.4 [M.-R.-4]. — *let M be a complete finite total curvature minimal surface in \mathbb{R}^3/S_θ . There exists a conformal compactification \overline{M} of M , and meromorphic one forms $(\frac{dg}{g}, \eta)$ on \overline{M} , such that M is parametrized by*

$$X(z) = Re \int (g + \frac{1}{g}, ig - \frac{i}{g}, 2)\eta$$

where $g = \exp(\int \frac{dg}{g})$.

8.5. The geometry of finite total curvature ends

Now using this parametrization we describe the asymptotic geometry of embedded annular ends.

THEOREM 8.6 [M.-R.-4]. — *A properly embedded minimal annulus in*

\mathbb{R}^3/S_θ , of finite total curvature, is asymptotic to a plane, a flat vertical annulus (a Scherk type end) or to a helicoid–catenoid type end (with horizontal limit tangent plane). If the end A is part of an m -surface of finite total curvature then A can not be a helicoid–catenoid type end. If $\theta \neq 0$ and A is asymptotic to a plane, then the plane is horizontal. If θ is irrational, then A is not a Scherk type end.

8.6. The winding number of an end

Using this theorem we can calculate the flux and total curvature of m -surfaces M of finite total curvature. For the latter one proceeds as follows. Let $\gamma \subset \mathbb{R}^3/S_\theta$ be the quotient of the x_3 -axis and let T_R be a tubular neighborhood of radius R of γ . For R large, $M_R = M \cap T_R$ is bounded by k Jordan curves C_1, \dots, C_k on ∂T_R , pairwise disjoint, and each C_i converges to a vertical line (Scherk type end) or to a horizontal circle (planar end) or to a helice on ∂T . Now one calculates the total curvature of M_R using Gauss–Bonnet and let $R \rightarrow \infty$. The boundary term is what we call the winding number of the end.

More generally, let A be a properly immersed annular end in \mathbb{R}^3/S_θ . We know a subend of A is disjoint from γ so we assume $A \cap \gamma = \emptyset$. Then ∂A is homotopic to a cycle on ∂T_R of the form $n\alpha + m\beta$ where α is a horizontal circle on ∂T_R and β is the quotient of the right handed helicoidal arc that joins a point p to $S_\theta(p)$ and projects to an embedded cycle on ∂T_R . The winding number of A is defined to be $\frac{1}{2\pi}|2\pi n + m\theta|$. It's easy to see that this doesn't depend on R for R large, and in the case of standard ends it is the limit of the total geodesic curvature of the C_1, \dots, C_k .

When M is a complete minimal surface of finite total curvature in \mathbb{R}^3/S_θ , the winding number of M is defined to be the sum of the winding numbers of it's ends. When M is embedded, this is k times the winding number of one end, k the number of ends.

Now the formula of 7.1 should be clear to the reader :

$$C(M) = 2\pi(\mathcal{X}(M) - W(M)) .$$

When the ends are Scherk type ends this is $C(M) = 2\pi\mathcal{X}(M)$. When they are k -planar ends, it is $C(M) = 2\pi(\mathcal{X}(M) - k)$.

Applications of the finite total curvature theorem

We have seen in VII, that a non planar orientable m -surface in a flat 3-manifold separates the space (this followed easily from the strong half-space theorem). This fact, together with 8.1 and our knowledge of the geometry of finite total curvature ends 8.6, yields a topological obstruction for the existence of certain m -surfaces :

THEOREM 8.7 [M.-R.-4]. — *Let M be an orientable non planar m -surface, of finite topology, in a non simply connected flat 3-manifold. Then the number of ends of M is even.*

Erik Toubiana has proved that an m -surface in \mathbb{R}^3/T , T a translation, that has finite, non zero, total curvature and the topology of a two punctured sphere (i.e. an annulus) is a helicoid [T]. Using 8.1 we generalize this result to \mathbb{R}^3/S_θ :

THEOREM 8.8 [M.-R.-4]. — *Let M be an m -surface in \mathbb{R}^3/S_θ , topologically an annulus, and not flat. Then M is a helicoid.*

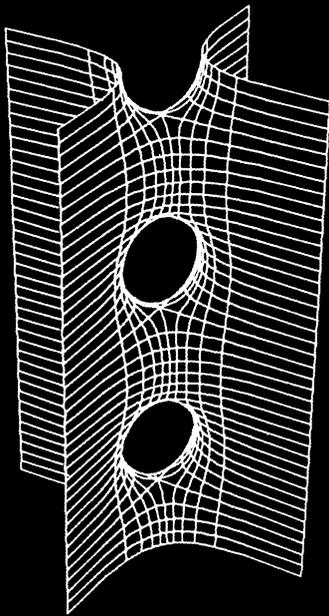
Perez and Ros have generalized the Toubiana theorem to genus zero :

THEOREM 8.9 [P.-Ros]. — *The helicoid is the only genus zero m -surface in \mathbb{R}^3/T with a finite number of helicoidal type ends.*

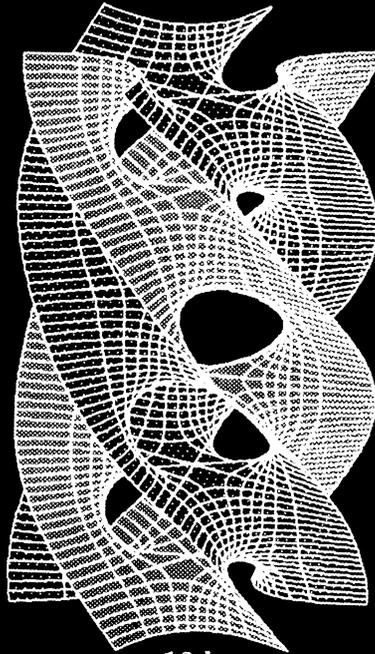
Their technique of proof uses the Lopez-Ros deformation described in III, and theorem 8.1. They also prove (with this technique) that there are no genus one m -surfaces in \mathbb{R}^3/S_θ , $\theta \neq 0$, with a finite number of planar ends. In other words : one can not screw the Riemann example. Notice that the Karcher deformations of Scherk's singly periodic surface shows that one can screw Scherk's surface [K.-2], figure 18-b.

Theorem 8.8 yields a unicity theorem for the helicoid in \mathbb{R}^3 :

THEOREM 8.10 [M.-R.-4]. — *The plane and the helicoid are the only simply connected m -surfaces in \mathbb{R}^3 with an infinite symmetry group.*



18.a



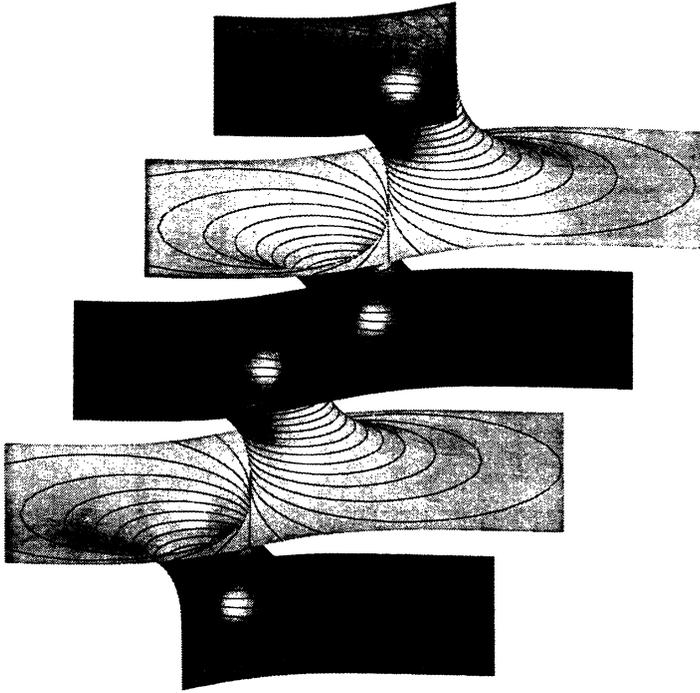
18.b



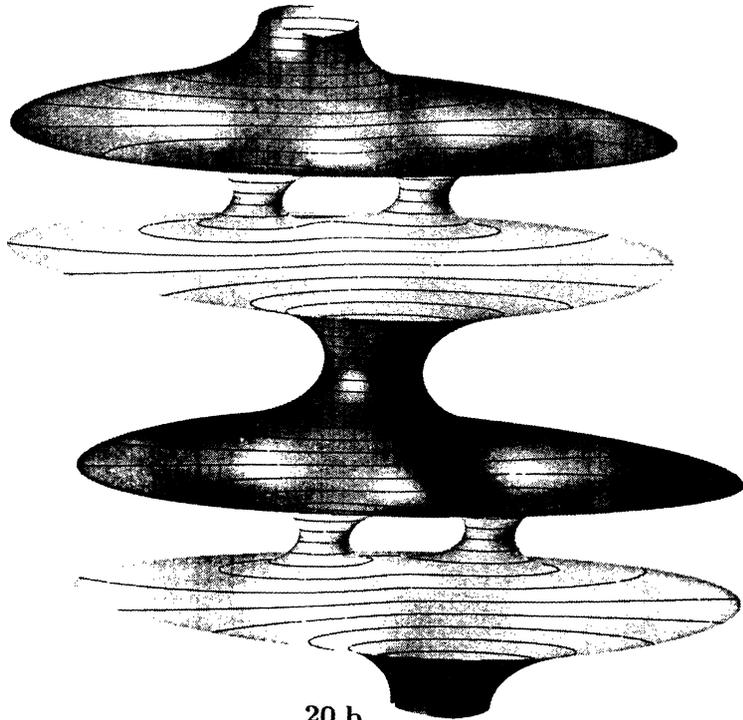
17.a



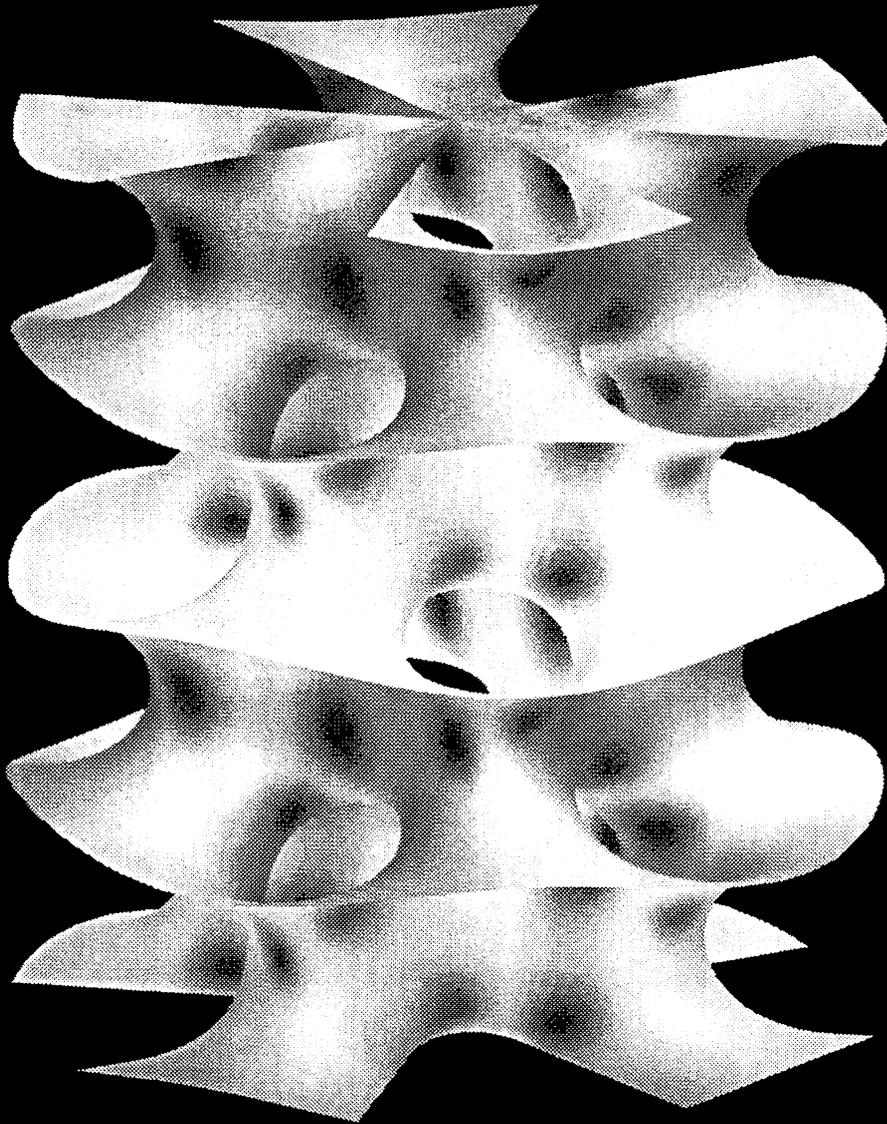
17.b



19



20.b



20.a

Another application allows us to classify the sum surface :

THEOREM 8.11 [M.-R.-4]. — *Let $M \subset \mathbb{R}^3/T$ be an m surface of finite topology, T a non trivial translation. If M has a helicoid end, then $M + M$ is a helicoid. if M has a planar end then $M + M$ is a point. if M has four Scherk type ends, then $M + M$ is a Scherk surface.*

Callahan, Hoffman and Meeks have proved a good structure theorem for singly periodic m -surfaces with more than one end.

THEOREM 8.12 [C.-H.-M.]. — *Let $M \subset \mathbb{R}^3$ be an m -surface with infinite symmetry group and more than one end. Then either M is a catenoid or M has the following properties :*

- 1) $\text{Sym}(M)$ contains an infinite cyclic subgroup S of finite index, generated by a screw motion S_θ .
- 2) M/S has finite topology precisely when M/S has finite total curvature.
- 3) there exists a plane whose intersection with M consists of a finite number of simple closed curves.

As a corollary of their theorem, they prove that a doubly periodic m -surface in \mathbb{R}^3 has one end.

9. SOME PROBLEMS, CONJECTURES AND RELATED RESULTS

Perhaps there are many m -surfaces in \mathbb{R}^3 of finite topology and infinite total curvature. For the moment, the only known example is the helicoid. Are there any others? Mark Soret has proved there are no others near the helicoid (graphs over the helicoid in an ε -tubular neighborhood of the helicoid [M.-S.]).

A less general question is to decide if the helicoid and plane are the only simply connected m -surfaces in \mathbb{R}^3 . We know this to be the case if the surface has an infinite symmetry group; theorem 8.10.

Maybe every infinite total curvature m -surface in \mathbb{R}^3 has an infinite

symmetry group (I doubt it) in which case the answer would be affirmative. All of the examples of infinite total curvature m -surfaces we know today are constructed using symmetries. There is no good reason (as far as I am concerned) to believe there are no others. It is likely that one can add exactly one handle (maybe more) to the helicoid to create an m -surface of infinite total curvature and non periodic.

There is an important difference when an m -surface in \mathbb{R}^3 has more than one end. We have seen in V, that this enables us to find planar or catenoid ends in the complement of M . Then Hoffman and Meeks proved that at most two annular ends of such M can have infinite total curvature (6.1). This leads Hoffman and Meeks to conjecture :

The finite total curvature conjecture [H.-M.-3] :

An annular end of an m -surface in \mathbb{R}^3 , with at least two ends has finite total curvature.

This is related to the Nitsche conjecture : a minimal surface that meets every horizontal plane in a Jordan curve is a catenoid. Nitsche proved this assuming the Jordan curves are star shaped [N].

Now Meeks and I have proved (6.2) that an annular end of an m -surface in \mathbb{R}^3 with at least two ends, is either of finite total curvature or contains a subend which meets every horizontal plane, in the upper halfspace of \mathbb{R}^3 , in a Jordan curve (after a Euclidean motion of the surface).

Hence the finite total curvature conjecture is a consequence of an affirmative answer to the following conjecture of Meeks and me :

The generalized Nitsche conjecture [M.-R.-3] :

Let A be a minimal annular end such that $A \cap \{x_3 = c \geq 0\}$ is a Jordan curve for every $c \geq 0$. Then A has finite total curvature.

Notice that this question concerns one holomorphic function g in the punctured disc and the problem is whether the origin is an essential singularity. Since A (or a subend of A) can be conformally parametrized by D^* with $x_3 = K \ln|z|$ the Weierstrass data of A is of the form $(g, \frac{1}{zg})$. It seems difficult to relate the singularity of g at the origin with the property

that A is embedded. Erik Toubiana and I have constructed examples of immersed annuli meeting every horizontal plane transversally and of infinite total curvature (g has an essential singularity [R.-T.-1]). We have even constructed such immersions in a slab of \mathbb{R}^3 . From time to time, I find myself working on the Nitsche conjecture using techniques from complex analysis (concerning essential singularities), but not using A embedded. Fortunately, this doesn't happen to me very often.

An affirmative answer to the generalized Nitsche conjecture would imply that finite topology, m -surfaces with more than one end can be parametrized by meromorphic data on a compact Riemann surface. All of the examples of properly embedded m -surfaces in \mathbb{R}^3 , that we presently know, do have this property : all of the infinite total curvature examples we know are periodic and have quotients of finite topology. We saw in VIII that the generalized Weierstrass representation is meromorphic on a compact Riemann surface.

What are the m -surfaces in \mathbb{R}^3 with exactly one end, topologically an annulus. For the moment, we know of only the plane and the helicoid, but as I said earlier, it is likely one can add a handle to a helicoid. Perhaps one can realise all compact surfaces, of arbitrary genus, with one puncture. Let us call an annular end algebraic if it is conformally a punctured disk and $\frac{dg}{g}$ and η extend meromorphically to the puncture. Is every finite topology m -surface in \mathbb{R}^3 algebraic? Is a properly embedded minimal annular end algebraic? Can one at least decide if it is conformally a punctured disk? I can prove that a minimally immersed annulus whose total curvature grows polynomially (*i.e.* $\int_{D_r} |K| \leq cr^n$, where D_r is a geodesic disk of radius r) is conformally a punctured disk. This growth condition should imply algebraic. An interesting related problem is to study minimal surfaces whose intersection with every plane $x_3 = \text{constant}$, is one properly embedded real line. Is such a surface conformally \mathbb{C} ? Is it a plane or a helicoid?

What are the genus zero m -surfaces in \mathbb{R}^3 ? The only examples we know are the plane, the helicoid, and the singly periodic Riemann examples. Meeks has conjectured that if the surface is also periodic then it is one of the these three examples [M.-1].

What are the genus zero m -surfaces in the other flat 3-manifolds than \mathbb{R}^3 ? In $T^2 \times \mathbb{R}$ we believe the only such examples lift to a Scherk surface in \mathbb{R}^3 (notice Scherk has infinite genus in \mathbb{R}^3). This was proved by Meeks and me when such a surface has 4 ends [M.-R.-1] and Wei extended this to 6 ends [Wei].

In \mathbb{R}^3/S_θ , $\theta = 0$ theorem 8.9 of Perez-Ros says the only genus zero finite topology example with helicoid type ends is the helicoid. What are all the genus zero examples in \mathbb{R}^3/S_θ ? Notice the Riemann example has genus zero in \mathbb{R}^3 and genus one in \mathbb{R}^3/T .

A (too) general question is to classify the genus g finite topology m -surfaces in $T^2 \times \mathbb{R}$ or \mathbb{R}^3/S_θ . For $g = 0$ or 1, I believe the problem is presently within our grasp. Certainly the same problem in \mathbb{R}^3 is beyond our means for the moment. Until recently, the only doubly periodic examples we knew were coverings of the Scherk surface or the Karcher saddles, together with their families constructed by Meeks and me [M.-R.-1]. Then F. Wei found a very beautiful example (using conjugate Plateau techniques or Weierstrass representation) to construct a genus two doubly periodic example with two top ends and two bottom ends, all parallel, different from the other known examples. Wei's surface had no lines as in the Scherk's surface and Karcher saddle [Wei], figure 21-a. Using Wei's idea, Karcher was able to add a handle to Scherk's surface (so the new surface has the same end behavior as the Scherk surface and is of genus one; personal communication), figure 21-b. Rabah Souam has proved that neither Wei's nor Karcher's surface could exist if one tried to keep the four vertical lines on the surface (thesis; Paris VII).

D. Hoffman conjectures that if M is a finite total curvature m -surface in \mathbb{R}^3 then the number of ends of M is less than or equal to the genus of M plus two. He believes that to add an end to an embedded minimal surface of finite total curvature in \mathbb{R}^3 , one must increase the genus (contrary to the Riemann example).

Another interesting subject to pursue is the relationship between the intrinsic isometries of an m -surface M (i.e. its symmetry group) and the ambient isometries leaving M invariant (its isometry group). When M is an m -surface in \mathbb{R}^3 , Meeks conjectures that every symmetry of M extends

to an isometry of \mathbb{R}^3 . Meeks and I have proved this for doubly periodic m -surfaces; in fact we proved more (rigidity) : let $f_1 : M \rightarrow \mathbb{R}^3$ be a doubly periodic m -surface and suppose $f_2 : M \rightarrow \mathbb{R}^3$ is another isometric minimal immersion of M , then there is an isometry ϕ of \mathbb{R}^3 such that $\phi f_2 = f_1$, [M.-R.-1]. Choi, Meeks and White have proved that an m -surface in \mathbb{R}^3 with more than one end is rigid [C.-M.-W.].

Singly periodic m -surfaces in \mathbb{R}^3 are not rigid (the helicoid) however it is true that their symmetry group equals their isometry group when M/S_θ has finite topology [M.-1].

Meeks has also conjectured that a non simply connected m -surface M in \mathbb{R}^3 is rigid (maybe the helicoid is the only non rigid m -surface in \mathbb{R}^3 ?) : any other isometric proper minimal immersion of M is congruent to M [M.-1].

Perhaps the notion of rigidity should be restricted to isometric minimal embeddings of M (not immersions). Then the helicoid is probably rigid.

Meeks has extended the finite total curvature theorem 7.1 to finite genus doubly periodic surfaces. He proved an m -surface in $T^2 \times \mathbb{R}$ of finite genus has finite total curvature [M.-1]. Does this remain true in \mathbb{R}^3/S_θ ?

In what generality does the maximum principle at infinity remain valid? Can one remove the hypothesis $\partial M_1, \partial M_2$ compact? The minimum distance between M_1 and M_2 (assumed disjoint) should not be realizable at *interior points at infinity*.

In the same spirit, Antonio Ros asked me the following question : suppose M is an m -surface in \mathbb{R}^3 ; can an end of M be an accumulation point of other ends of M ? More precisely, can there be a divergent sequence x_n on an end A of M and a sequence $y_n \in M - A$ such that $\text{dist}(x_n, y_n) \rightarrow 0$, as $n \rightarrow \infty$?

A problem that arises when studying m -surfaces in \mathbb{R}^3 of finite total curvature is the following : do all the catenoid ends have the same axis? This is unknown even for three catenoid ends.

There has been important work done by Celso Costa on the problem of classifying m -surfaces in \mathbb{R}^3 of finite total curvature : he classified those of total curvature -12π [Cost.-3]. His proof uses very difficult calculations in elliptic function theory. It would be very interesting to understand this

from another point of view.

There has been much important and beautiful recent work done on minimal surfaces that I have not discussed. I consider my most important omission the theorem of Frohman and Meeks that two one-ended m -surfaces in \mathbb{R}^3 of the same genus are ambiently isotopic [F.-M.].

There is also the very beautiful work of Fujimoto on values of the Gauss map [Fuj.-1,2]; but it is not clear to me this has anything to do with the surface being embedded or not.

Meeks and White have studied the space of minimal submanifolds of \mathbb{R}^3 bounded by two convex Jordan curves C_1 and C_2 . When C_1 and C_2 are in parallel planes they proved there are 0, 1 or 2 minimal annuli with boundary $C_1 \cup C_2$ [M.-Wh.].

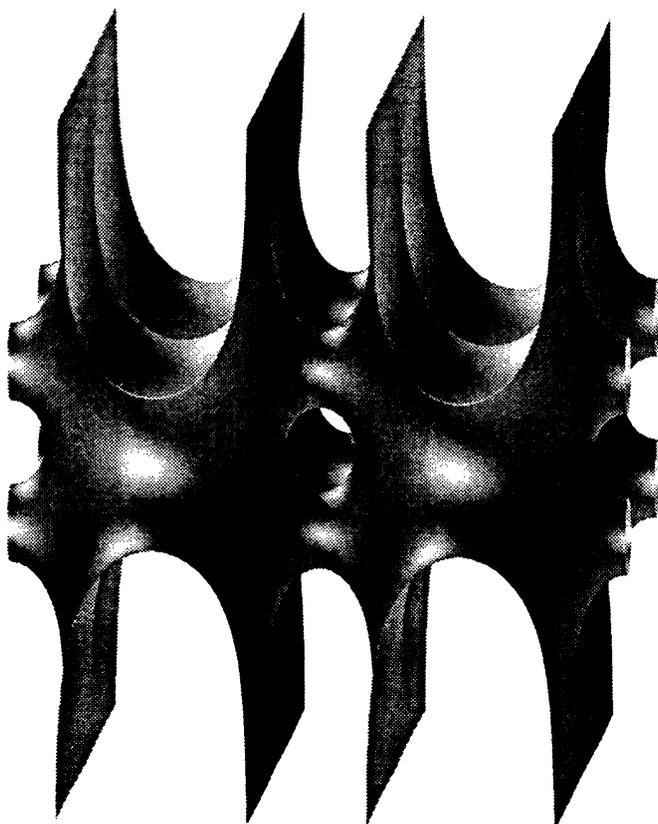
Finally, let me mention the problem of how, and when, can one desingularize a minimal variety : given two m -surfaces M_1, M_2 in \mathbb{R}^3 , when is there an m -surface M that is close to $M_1 \cup M_2$ outside of a neighborhood of $M_1 \cup M_2$? In many examples, the desingularization M looks like a string of handles along $M_1 \cup M_2$. Here are some examples. Scherks singly periodic surface is the desingularization of two orthogonal planes. Karcher's singly periodic generalization of this Scherk surface is the desingularization of n planes meeting along an axis; figure 18-a.

A helicoid and its rotation about its axis, meet along the axis. Karcher's examples desingularize this (and in general, n helicoids meeting along their axis) desingularize this by a string of handles along the axis; figure 18-a.

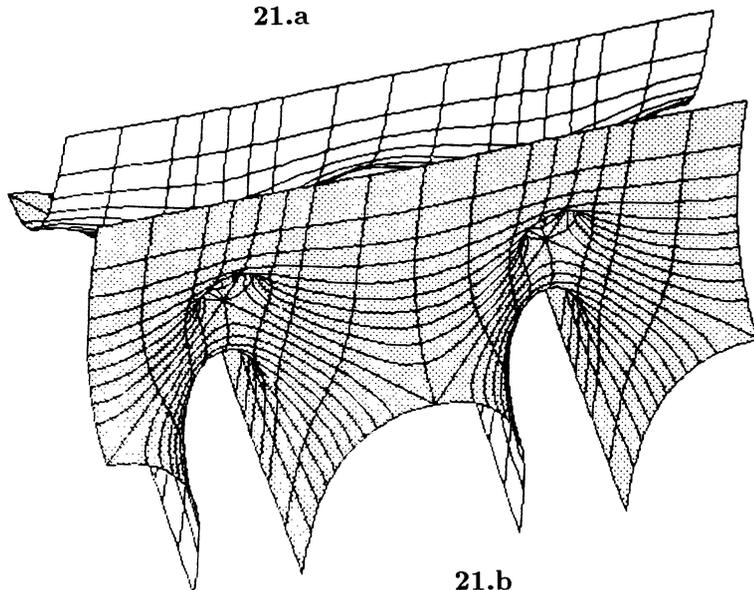
Costa's finite total curvature m -surface, with 3-ends, can be thought of as the desingularization of the vertical catenoid and the horizontal plane passing through the waist circle; figure 1. The higher genus examples of Hoffman and Meeks with 3-ends are a better illustration of this (figure 2) : one places a string of handles around the circle of intersection of the catenoid and the horizontal plane.

When $M_1 \cap M_2$ is a Jordan curve C , a necessary condition for desingularization appears to be : $\int_C n_1 \cdot n_2 = 0$, where n_1 is the normal to C in M_1 and n_2 the normal to M_2 along C .

How to make sense of this is not at all clear. How can one do *minimal surgery* on $M_1 \cup M_2$?



21.a



21.b

BIBLIOGRAPHY

- [B.Do C.] J.L. Barbosa and M. Do Carmo. On the size of a stable minimal surface in \mathbb{R}^3 . *American Journal of Mathematics* 98(2) : 515-528, 1976.
- [C.-H.-M] M. Callahan, D. Hoffman, and W. H. Meeks III. The structure of singly-periodic minimal surfaces. *Inventiones Math.* 99 : 455-481, 1990.
- [Cost.-1] C. Costa. *Imersões minimas en \mathbb{R}^3 de gênero un e curvatura total finita*. PhD thesis, IMPA, Rio de Janeiro, Brazil, 1982.
- [Cost.-2] C. Costa. Example of a complete minimal immersion in \mathbb{R}^3 of genus one and three embedded ends. *Bull. Soc. Bras. Mat.* 15 : 47-54, 1984.
- [Cost.-3] C. Costa. Uniqueness of minimal surfaces embedded in \mathbb{R}^3 with total curvature -12π . *Journal of Differential Geometry* 30(3) : 597-618, 1989.
- [Cour.] R. Courant. *Dirichlet's Principle, Conformal Mapping and Minimal Surfaces*. Interscience Publishers, Inc., New York, 1950.
- [Darb.] G. Darboux. *Leçons sur la théorie générale des surfaces et les applications géométriques du calcul infinitésimal*. Gauthier-Villars, Paris, 1st part, 2nd edition, 1914.
- [Do C.-P.] M. Do Carmo and C.K. Peng. Stable minimal surfaces in \mathbb{R}^3 are planes. *Bulletin of the AMS* 1 : 903-906, 1979.
- [Doug.] J. Douglas, Solution of the problem of Plateau, *Trans. AMS* 33 : 263-321, 1931.
- [F.-Oss.] R. Finn and R. Osserman. On the Gauss curvature of non-parametric minimal surfaces, *J. Anal. Math.* 12 : 351-364, 1964.
- [F.C.] D. Fischer-Colbrie. On complete minimal surfaces with finite Morse index in 3-manifolds. *Inventiones Math.* 82 : 121-132, 1985.
- [Fr.-M.] C. Frohman and W.H. Meeks III. The topological uniqueness of complete one-ended minimal surfaces and Heegard surfaces in \mathbb{R}^3 , preprint.
- [Fuj.-1] H. Fujimoto. On the number of exceptional values of the Gauss maps of minimal surfaces. *Journal of the Math. Society of Japan* 40(2) : 235-247, 1988.
- [Fuj.-2] H. Fujimoto. Modified defect relations for the Gauss map of minimal surfaces. *Journal of Differential Geometry* 29 : 245-262, 1989.
- [G.-T.] D. Gilbarg and N.S. Trudinger. *Elliptic partial differential equations of*

- second order*. Springer-Verlag, New York, 2nd edition, 1983.
- [H.-S.] R. Hardt and L. Simon. Boundary regularity and embedded minimal solutions for the oriented Plateau problem. *Annals of Math.* 110 : 439-486, 1979.
- [Heinz] E. Heinz. Über die Lösungen der Minimalflächengleichung. *Nachr. Akad. Wiss. Göttingen Math. Phys.* K1, II (1952) 51-56.
- [H.-M.-1] D. Hoffman and W.H. Meeks III. A complete embedded minimal surface in \mathbb{R}^3 with genus one and three ends. *Journal of Differential Geometry* 21 : 109-127, 1985.
- [H.-M.-2] D. Hoffman and W.H. Meeks III. Properties of properly embedded minimal surfaces of finite total curvature. *Bulletin of the AMS* 17(2) : 296-300, 1987.
- [H.-M.-3] D. Hoffman and W.H. Meeks III. The asymptotic behavior of properly embedded minimal surfaces of finite topology. *Journal of AMS* 2(4) : 667-681, 1989
- [H.-M.-4] D. Hoffman and W.H. Meeks III. The strong halfspace theorem for minimal surfaces. *Inventiones Math.* 101 : 373-377, 1990.
- [H.-M.-5] D. Hoffman and W.H. Meeks III. Minimal surfaces based on the catenoid. *Amer. Math. Monthly, Special Geometry Issue* 97(8) : 702-730, 1990.
- [H.-Wei] D. Hoffman and F. Wei. Adding handles to the helicoid, preprint.
- [E.H.] E. Hopf. On an inequality for minimal surfaces $z = f(x, y)$, *J. Rat. Mech. Anal.* 2 : 519-522, 1953.
- [Hub.] A. Huber. On subharmonic functions and differential geometry in the large. *Commentari Mathematici Helvetici* 32 : 181-206, 1957.
- [J.-S.] H. Jenkins, J. Serrin. Variational problems of minimal surface type II, *Arch. Rat. Mech. Analysis* 21 : 321-342, 1966.
- [J.-Xav.] L. Jorge, F. Xavier. A complete minimal surface in a slab of \mathbb{R}^3 , *Annals of Maths*, 1980, 203-206.
- [K.-1] H. Karcher. Construction of minimal surfaces. *Surveys in Geometry*, pages 1-96, 1989. University of Tokyo, 1989, and Lecture Notes No.12, SFB256, Bonn, 1989.
- [K.-2] H. Karcher. Embedded minimal surfaces derived from Scherk's examples. *Manuscripta Math.* 62 : 83-114, 1988.
- [K.-3] H. Karcher. The triply periodic minimal surfaces of Alan Schoen and

- their constant mean curvature companions. *Manuscripta Math.* 64 : 291-357, 1989.
- [K.-4] H. Karcher. Construction of higher genus embedded minimal surfaces. *Geom. and Top. of Sub. III* World Sc. 174-191, 1990.
- [L.-R.] R. Langevin and H. Rosenberg. A maximum principle at infinity for minimal surfaces and applications. *Duke Math. Journal* 57 : 819-828, 1988.
- [Lo.-Ros] F.J. Lopez and A. Ros. On embedded complete minimal surfaces of genus zero. *Journal of Differential Geometry* 33(1) : 293-300, 1991.
- [M.-1] W.H. Meeks III. The geometry, topology and existence of periodic minimal surfaces, preprint.
- [M.-2] W.H. Meeks III. *Lectures on Plateau's Problem*. Instituto de Matematica Pura e Aplicada (IMPA), Rio de Janeiro, Brazil, 1978.
- [M.-3] W.H. Meeks III. The theory of triply-periodic minimal surfaces. *Indiana University Math. Journal* 39(3) : 877-936, 1990.
- [M.-R.-1] W.H. Meeks III and H. Rosenberg. The global theory of doubly periodic minimal surfaces. *Inventiones Math.* 97 : 351-379, 1989.
- [M.-R.-2] W.H. Meeks III and H. Rosenberg. The maximum principle at infinity for minimal surfaces in flat three-manifolds. *Commentari Mathematici Helvetici* 65 : 255-270, 1990.
- [M.-R.-3] W.H. Meeks III and H. Rosenberg. The geometry and conformal structure of properly embedded minimal surfaces of finite topology in \mathbb{R}^3 , to appear in *Invent. Math.*
- [M.-R.-4] W.H. Meeks III and H. Rosenberg. The geometry of periodic minimal surfaces, to appear in *Comment. Math. Helv.*
- [M.-Wh.] W.H. Meeks III and B. White. Minimal surfaces bounded by convex curves in parallel planes. *Commentari Mathematici Helvetici* 66 : 263-278, 1991.
- [M.-Y.] W.H. Meeks and S.T. Yau. The existence of embedded minimal surfaces and the problem of uniqueness. *Math. Z.* 179 : 151-168, 1982.
- [N.] J.C.C. Nitsche. A characterization of the catenoid. *Journal of Math. Mech.* 11 : 293-302, 1962.
- [Oss.-1] R. Osserman. Global properties of minimal surfaces in E^3 and E^n . *Annals of Math.* 80(2) : 340-364, 1964.

- [Oss.-2] R. Osserman. On the Gauss curvature of minimal surfaces. *Trans. AMS* 96 : 115-128, 1960.
- [P.-Ros] J. Pérez and A. Ros. Some uniqueness and nonexistence theorems for embedded minimal surfaces, preprint.
- [Rado-1] T. Rado. The problem of the least area and the problem of Plateau. *Math. Z.* 32 : 763-796, 1930.
- [Rado-2] T. Rado. On the problem of Plateau. *Ergebnisse der Mathematik und ihrer Grenzgebiete*, Springer-Verlag, Berlin 1933.
- [Reif.] R. Reifenberg. Solution for the Plateau problem for m -dimensional surfaces of varying topological type. *Acta Math.* 104 : 1-92, 1960.
- [R.-T.-1] H. Rosenberg and E. Toubiana. A cylindrical type complete minimal surface in a slab of \mathbb{R}^3 . *Bull. Sc. Math. III*, pages 241-245, 1987.
- [R.-T.-2] H. Rosenberg and E. Toubiana. Complete minimal surfaces and minimal herissons. *Journal of Differential Geometry* 28 : 115-132, 1988.
- [R.-S.E.] Sa Earp and H. Rosenberg. The Dirichlet problem for the minimal surface equation on unbounded planar domains. *Journal de Mathématiques Pures et Appliquées* 68 : 163-183, 1989.
- [Sch.-1] R. Schoen. Uniqueness, symmetry, and embeddedness of minimal surfaces. *Journal of Differential Geometry* 18 : 791-809, 1983.
- [Sch.-2] R. Schoen. *Estimates for Stable Minimal Surfaces in Three Dimensional Manifolds*, volume 103 of *Annals of Math. Studies*. Princeton University Press, 1983.
- [Simon] L. Simon. Lectures on geometric measure theory. In *Proceedings of the Center for Mathematical Analysis*, volume 3, Canberra, Australia, 1983. Australian National University.
- [Smale] N. Smale. A bridge principle for minimal and constant mean curvature submanifolds of \mathbb{R}^n . *Invent. Math.* 90 : 505-549, 1987.
- [M.S.] M. Soret. Deformations de surfaces minimales. *Thèse Univ. Paris VII*, 1992.
- [Souam] R. Souam. Stabilité et unicité des surfaces minimales. *Thèse Univ. Paris VII*, 1992.
- [T.] E. Toubiana. On the uniqueness of the helicoid. *Ann. Inst. Four.* 38 : 121-132, 1988.
- [Wei] F. Wei. Some existence and uniqueness theorems for doubly periodic

minimal surfaces, to appear in *Invent. Math.*

[Wh.] B. White. Complete surfaces of finite total curvature. *Journ. Diff. Geom.*
26 : 315-326, 1987.

Harold ROSENBERG

Université de Paris VII

U.F.R. de Mathématiques

URA 212 du C.N.R.S.

Tour 45-55 - 5ème étage

2, place Jussieu

F-75251 PARIS CEDEX 05