

PANORAMAS ET SYNTHÈSES 11

**NOUVEAUX INVARIANTS
EN GÉOMÉTRIE ET EN TOPOLOGIE**

**Michèle Audin
John W. Morgan
Pierre Vogel**

avec une postface de
Daniel Bennequin

édité par François Dumas, Jean-Yves Le Dimet, Sylvie Paycha

Société Mathématique de France 2001
Publié avec le concours du Centre National de la Recherche Scientifique

M. Audin

Institut de Recherche Mathématique Avancée, Université Louis Pasteur et CNRS,
7 rue René Descartes, 67084 Strasbourg cedex, France.

E-mail : `Michele.Audin@math.u-strasbg.fr`

Url : `http://www-irma.u-strasbg.fr/~maudin`

J.W. Morgan

Department of Mathematics, Columbia University, New York, NY 10024 USA.

E-mail : `jm@math.columbia.edu`

Url : `www.math.columbia.edu/~jm`

P. Vogel

Université Paris 7, 2 place Jussieu, 75251 Paris Cedex 05.

E-mail : `vogel@math.jussieu.fr`

Url : `http://www.math.jussieu.fr/~vogel/`

D. Bennequin

Institut de mathématiques de Jussieu, Université de Paris 7, 4, place Jussieu,
75252 Paris cedex 05.

F. Dumas

LMP, Université Blaise Pascal, 63177 Aubière cedex.

E-mail : `francois.dumas@math.univ-bpclermont.fr`

J.-Y. Le Dimet

LMP, Université Blaise Pascal, 63177 Aubière cedex.

E-mail : `jy.ledimet@math.univ-bpclermont.fr`

S. Paycha

LMA, Université Blaise Pascal, 63177 Aubière cedex.

E-mail : `sylvie.paycha@math.univ-bpclermont.fr`

Url : `wwwlma.univ-bpclermont.fr/~paycha`

Classification mathématique par sujets (2000). — 5302, 8102, 53D, 53C23, 57M25, 57M27, 57R57, 58J.

Mots clefs. — Variétés symplectiques, courbes holomorphes, invariants de Seiberg-Witten, variétés de dimension 4, variétés de dimension 3, nœuds, invariants de type fini, théorie quantique des champs, dualité.

NOUVEAUX INVARIANTS EN GÉOMÉTRIE ET EN TOPOLOGIE

Michèle Audin, John W. Morgan, Pierre Vogel
avec une postface de Daniel Bennequin

édité par François Dumas, Jean-Yves Le Dimet, Sylvie Paycha

Résumé. — Ce volume traite des développements récents de trois types d'invariants géométriques :

- invariants symplectiques, dont les invariants de Gromov-Witten présentés par Michèle Audin,
- invariants de variétés de dimension 4 avec la théorie de Seiberg-Witten par John Morgan,
- invariants de type fini pour les variétés de dimension 3 décrits par Pierre Vogel.

Les liens entre ces trois classes d'invariants et la théorie des champs contemporaine sont abordés dans une postface de Daniel Bennequin.

Abstract (New Invariants in Geometry and Topology). — This volume offers a presentation of recent developments of three types of geometric invariants:

- symplectic invariants, including Gromov-Witten invariants, by Michèle Audin,
- invariants of four-manifolds and Seiberg-Witten theory, by John Morgan,
- finite type invariants for three-manifolds, by Pierre Vogel.

As a conclusion to this volume, Daniel Bennequin describes the links between these three types of invariants and contemporary quantum field theory.

TABLE DES MATIÈRES

Résumés des articles	vii
Abstracts	ix
Introduction	xi
 MICHÈLE AUDIN — <i>Invariants en géométrie symplectique via les courbes holomorphes</i>	1
1. Introduction : invariants, invariants mous, invariants grossiers.....	3
2. Les courbes pseudo-holomorphes, une introduction aux invariants de Gromov-Witten.....	8
3. Les espaces de modules d'applications stables.....	20
4. Propriétés et calculs d'invariants, enfin.....	41
5. Autres applications, autres invariants.....	52
6. Appendice : les espaces projectifs et leur famille.....	54
Références.....	56
 JOHN W. MORGAN — <i>Seiberg-Witten Invariants</i>	61
I. The Linear Algebra and Differential Geometry underlying the Seiberg-Witten Invariants.....	63
II. The Seiberg-Witten Equations and the Seiberg-Witten Moduli Space.....	73
III. Seiberg-Witten Invariants of Kähler Manifolds.....	82
IV. The Seiberg-Witten Invariants of Symplectic Four-manifolds.....	89
V. Other Results.....	92
References.....	97
 PIERRE VOGEL — <i>Invariants de type fini</i>	99
Introduction.....	99
1. Modules de diagrammes.....	100
2. L'intégrale de Kontsevich.....	104

3. Mouvements de Kirby.....	109
4. Les relations P_n	115
5. Propriétés de l'invariant de Le-Murakami-Ohtsuki.....	120
6. Invariants de type finis.....	121
Références.....	127

Postface

DANIEL BENNEQUIN — <i>Invariants contemporains</i>	131
I. Champs (semi-) classiques.....	131
II. Quantification des champs.....	134
III. Dualités des champs.....	138
IV. Monopôles.....	140
V. Supersymétrie. Théorie de Seiberg-Witten.....	142
VI. Dualité des supercordes.....	146
VII. Géométrie des invariants.....	150
Notes.....	152

RÉSUMÉS DES ARTICLES

<i>Invariants en géométrie symplectique via les courbes holomorphes</i> MICHÈLE AUDIN.....	1
---	---

Dans ce texte, je survole le problème des invariants des variétés symplectiques en m'attachant particulièrement à ceux qu'on obtient en « comptant » les courbes pseudo-holomorphes, les invariants de Gromov-Witten.

<i>Seiberg-Witten Invariants</i> JOHN W. MORGAN.....	61
---	----

Ces dernières années ont vu l'introduction de plusieurs invariants des variétés, invariants définis à l'aide des espaces de modules de certaines équations aux dérivées partielles. Il y a parmi ceux-ci les équations d'anti-autodualité et les invariants de Donaldson, les équations des courbes pseudo-holomorphes et les invariants de Gromov-Witten, et les équations de Seiberg-Witten et les invariants associés. Ces derniers sont le sujet de cet article. Nous commençons par développer les résultats fondamentaux qui sont nécessaires, puis nous définissons les invariants. Finalement, nous donnerons quelques applications aux variétés symplectiques et kählériennes.

<i>Invariants de type fini</i> PIERRE VOGEL.....	99
---	----

Il existe de nombreux invariants de Vassiliev (ou de type fini) des nœuds et des entrelacs. L'intégrale de Kontsevich est, en quelque sorte, l'invariant de type fini universel. Il prend ses valeurs dans un module de diagrammes $\mathcal{A}(\Gamma)$, Γ étant l'entrelacs considéré comme une courbe abstraite.

Il existe aussi des théories des invariants de type fini pour les variétés de dimension 3. Elles sont toutes plus ou moins équivalentes. La théorie de Kirby permet de décrire les variétés de dimension 3 à l'aide d'entrelacs. On peut alors extraire de l'intégrale de Kontsevich un invariant de variété de dimension 3. Lorsque la variété est une sphère d'homologie, cet invariant est l'invariant de type fini universel. Il prend ses valeurs dans le module de diagrammes $\mathcal{A}(\emptyset)$.

Invariants contemporains

DANIEL BENNEQUIN..... 131

Le texte se propose de présenter à des mathématiciens (ou à des étudiants en mathématiques) l'origine des dualités en théorie quantique des champs et en théorie des cordes (renormalisation, supersymétrie...), et d'expliquer comment ces dualités font apparaître de nouveaux invariants en topologie différentielle.

ABSTRACTS

<i>Invariants en géométrie symplectique via les courbes holomorphes</i> MICHÈLE AUDIN.....	1
---	---

In this survey, I discuss the problem of finding invariants for symplectic manifolds. I emphasize on the invariants that “count” holomorphic curves, namely on the Gromov-Witten invariants.

<i>Seiberg-Witten Invariants</i> JOHN W. MORGAN.....	61
---	----

Recent years have seen many new invariants of manifolds, invariants defined using moduli spaces of solutions to certain partial differential equations. These include the anti-self-dual equations and Donaldson invariants, the pseudo-holomorphic curve equations of Gromov and the resulting Gromov and Gromov-Witten invariants, quantum cohomology, various Chern-Simon’s invariants, Floer homology, and, the subject of this article, the Seiberg-Witten invariants. We cover the background necessary to define these invariants, give the definition, and then several applications to the topology of smooth four-manifolds including to symplectic four-manifolds and complex algebraic surfaces.

<i>Invariants de type fini</i> PIERRE VOGEL.....	99
---	----

There exist many Vassiliev (or finite type) invariants for knots or links. The Kontsevich integral is, in some sense, the universal finite type invariant. It takes values in a module of diagrams $\mathcal{A}(\Gamma)$, where Γ is the link considered as an abstract curve.

There is also finite type invariants theories for 3-dimensional manifolds. They are essentially equivalent. Using Kirby calculus, every 3-dimensional manifold may be described by links. So it is possible to extract from the Kontsevich integral an invariant of 3-dimensional manifold. If the manifold is a homology sphere, this invariant is the universal finite type invariant of the manifold. It takes values in the module of diagrams $\mathcal{A}(\emptyset)$.

Invariants contemporains

DANIEL BENNEQUIN..... 131

In this text we will present for mathematicians (or students in mathematics) the origin of dualities in Quantum Field Theory and String Theories (renormalization, supersymmetry, ...), and will explain to explain how these dualities are the source of new invariants in Differential Topology

INTRODUCTION

Les textes publiés dans ce volume correspondent à trois cours et un exposé présentés lors des journées « États de la Recherche » qui ont réuni une soixantaine de participants du 15 au 18 septembre 1999 à l'Université Blaise Pascal (Clermont-Ferrand 2)⁽¹⁾

Le projet d'organiser ces journées est né au sein du groupe de travail « Géométrie, Topologie, Analyse fonctionnelle » qui rassemble depuis quelques années des chercheurs des deux laboratoires de mathématiques de l'Université Blaise Pascal. Dans la diversité des centres d'intérêt (de la géométrie différentielle aux variétés de petite dimension, des algèbres d'opérateurs aux groupes quantiques) et dans les différences de problématiques (des origines physiques aux méthodes algébriques, homologiques ou de représentation), apparaissait un type de démarche mathématique présentant des points communs parmi lesquels le rôle central joué par la mise en évidence de certains invariants. Lorsque nous avons souhaité inscrire dans le développement de ce groupe de travail l'organisation de journées États de la Recherche, c'est assez naturellement autour de ce thème que nous les avons envisagées.

De plus, la période nous semblait propice. Sorti chargé d'histoire de sa période classique (essentiellement le XIX-ème siècle), le terme même d'invariant avait au XX-ème siècle connu bien des glissements, passant des objets mathématiques aux symétries de ces objets, puis voyant les symétries s'estomper à leur tour au profit de l'invariant en tant qu'objet. Il nous paraissait pertinent de présenter un « état des lieux » de certains développements récents de la notion d'invariant en partie impulsés par les idées de la physique contemporaine.

⁽¹⁾Ces journées ont été organisées conjointement par le Laboratoire de Mathématiques Pures (E.A. 986) et le Laboratoire de Mathématiques Appliquées (U.M.R. 6620) de l'Université Blaise Pascal, dans le cadre d'une école thématique du C.N.R.S. et de journées États de la Recherche de la S.M.F., avec le soutien financier du Ministère de l'Éducation Nationale de la Recherche et de la Technologie, du C.N.R.S., de l'Université Blaise Pascal, et de la région Auvergne. Que tous en soient ici remerciés.

Les trois chapitres suivants traitent d'invariants au sens classique où l'entendait Felix Klein dans le programme d'Erlangen, soit d'objets qui restent invariants sous l'action d'un groupe : le groupe des symplectomorphismes d'une variété symplectique dans le cas des invariants de Gromov-Witten, le groupe des difféomorphismes d'une variété de dimension 4 dans le cas des invariants de Seiberg-Witten, et le groupe des homéomorphismes d'une variété de dimension 3 dans le cas des invariants de type fini. Le premier chapitre (Michèle Audin) montre comment les invariants de Gromov-Witten dans le cadre des variétés symplectiques conduisent aux invariants de Seiberg-Witten en dimension 4. Le second (John Morgan) donne une construction à partir de théories de jauge de ces invariants de Seiberg-Witten. Finalement, l'invariant de Casson, qui peut aussi s'obtenir dans le cadre de la théorie de Seiberg-Witten, apparaît dans le troisième chapitre (Pierre Vogel) comme un invariant de type fini. Les liens qui unissent ces différents invariants et qui peuvent surprendre à première vue, s'expliquent assez naturellement par une brève incursion historique dans la théorie des champs. Ils sont pour la plupart de nature heuristique, établis par le biais de l'intégrale de Feynman, mais méritent malgré tout d'être notés car ils semblent porteurs de développements futurs autour de la notion d'invariant comme le suggère Daniel Bennequin dans la postface de cet ouvrage.

Les trois types d'invariants que l'on rencontre dans ce recueil ont en commun de pouvoir s'interpréter (au moins formellement) en termes de fonctions de corrélation. Entendons par fonction de corrélation une « intégrale de chemins » donnée par un objet formel du type

$$\langle O_1 \cdots O_k \rangle := \int_{\Gamma} \mathcal{D}\gamma e^{-\int \mathcal{L}(\gamma(x)) d\text{vol}(x)} O_1(\gamma) \cdots O_k(\gamma).$$

L'espace Γ sur lequel se fait l'intégration n'est pas un espace de « chemins » au sens propre du terme mais plutôt un espace de « champs » γ . Ces champs peuvent être des sections d'un fibré sur une variété sous-jacente M (par exemple une section d'un fibré spinoriel), des connexions sur un fibré principal basé sur M (dans une théorie de jauge, le champ est typiquement une connexion et le groupe de structure du fibré principal s'appelle le groupe de jauge) ou bien des applications d'une variété Σ vers une variété M (dans un modèle sigma par exemple). Dans ces trois cas, c'est la variété M sur laquelle on veut définir des invariants.

L'intégration sur l'espace des champs se fait à l'aide d'une mesure de volume formelle $\mathcal{D}\gamma$ sur Γ , l'intégrande étant formée d'« observables » $O_i : \Gamma \rightarrow \mathbb{C}$, $1 \leq i \leq k$. Le lagrangien $\mathcal{L}$ est une fonction sur l'espace des jets d'un certain ordre des éléments γ de Γ qui est intégrée par rapport à une mesure de volume ordinaire $d\text{vol}(x)$ (sur Σ dans le cas du modèle sigma, sur M dans les deux autres cas) pour donner l'action $\mathcal{A}(\gamma) := \int \mathcal{L}(\gamma(x)) d\text{vol}(x)$.

Le modèle est dit supersymétrique lorsque l'action est invariante par les transformations supersymétriques qui échangent bosons et fermions. Il est peut-être suffisant

ici – pour reprendre les termes de Michael Atiyah⁽²⁾ – de voir les théories de champs supersymétriques comme des constructions conduisant (formellement) à une géométrie différentielle sur certaines variétés de dimension infinie, ceci étant un point de vue naturel depuis les travaux de Witten faisant un usage géométrique de la supersymétrie⁽³⁾.

Certains lagrangiens ont pour particularité de donner lieu à des fonctions de corrélation indépendantes du choix de la métrique sur la variété sous-jacente, propriété qui caractérise ce qu'on appelle depuis une théorie topologique des champs, notion que Michael Atiyah a axiomatisée⁽⁴⁾, adaptant les axiomes proposés par Segal pour une théorie conforme des champs⁽⁵⁾. Cette interprétation physique des objets mathématiques que sont les invariants s'est petit à petit dégagée grâce aux travaux de physiciens, et en particulier à ceux d'Edward Witten.

Dans les années 80, Michael Atiyah proposa⁽⁶⁾ une théorie des champs non relativiste (*i.e.* la variable temps y joue un rôle différent de celui des variables d'espace) éclairant sous un jour nouveau des résultats intéressants pour la topologie de dimension 3 dûs à Andreas Floer⁽⁷⁾. À la suite de ces travaux, Michael Atiyah suggéra deux problèmes aux théoriciens des champs quantiques : donner une interprétation physique (certainement relativiste) de la théorie de Donaldson qui fournit des invariants de structures C^∞ en dimension 4 d'une part, chercher une définition intrinsèque du polynôme de Jones de la théorie des nœuds d'autre part qui donne lieu à des invariants de structures topologiques en dimension 3.

Edward Witten répondit⁽⁸⁾ par l'affirmative à la première question en proposant une interprétation des invariants de Donaldson comme fonctions de corrélation d'une théorie des champs $SU(2)$ (*i.e.* le groupe de jauge est $SU(2)$) supersymétrique tordue (ceci signifiant une entorse à la règle spin-statistique), la variété sous-jacente M étant de dimension 4. Ces fonctions de corrélation sont de la forme :

$$Z(k) = \langle O_1 \cdots O_k \rangle, \quad O_i := \int_{c_i} W_i,$$

W_i étant une forme différentielle de degré n_i définie en termes de champs de la théorie et c_i un cycle d'homologie de dimension n_i sur la variété. Bien que ces fonctions de corrélation soient des objets formels, Witten parvint à en extraire des formules explicites des invariants de Donaldson comme intégrales (ordinaires cette fois) sur l'espace

⁽²⁾« Topological Quantum Field Theories » in René Thom Festschrift (1989)

⁽³⁾« Supersymmetry and Morse Theory », Journ. Diff. Geom. 17 (1982)

⁽⁴⁾voir note 2

⁽⁵⁾« Two-dimensional conformal field theory and modular functors », IXth Internat. Conf. Math. Phys. Swansea, Adam Hilger 1989

⁽⁶⁾« New invariants of three- and four-manifolds » in « The mathematical heritage of Hermann Weyl », Proc. Symp. Pure Math., Vol 48, Amer. Math. Soc. 1988

⁽⁷⁾« Morse Theory for Fixed Points of Symplectic Diffeomorphisms », Bulletin A.M.S. n.16 (1987)

⁽⁸⁾« Topological Quantum Field Theory », Comm. Math. Phys. 1988

des modules des instantons. La théorie des invariants de Donaldson et leurs interprétations physiques devaient ensuite connaître des rebondissements grâce à des travaux de Nathan Seiberg et Edward Witten, développements sur lesquels nous reviendrons et qui font l'objet du chapitre 2 de cet ouvrage.

À la deuxième question posée par Atiyah concernant cette fois les invariants de nœuds, Witten répondit encore par l'affirmative⁽⁹⁾ en décrivant le polynôme de Jones comme fonction de corrélation, la variété sous-jacente étant cette fois de dimension 3 et le lagrangien dit de Chern-Simons étant donné par :

$$\mathcal{L}_{CS} = \frac{i}{4\pi} \text{Tr}(A \wedge dA + \frac{2}{3} A \wedge A \wedge A),$$

où $d + A$ est la forme locale d'une connexion sur un fibré vectoriel (de groupe de structure G quelconque) basé sur la variété sous-jacente. Encore une fois, il s'agit d'une théorie topologique dont les fonctions de corrélation sont liées à des invariants de nœuds traités dans le chapitre 3 de cet ouvrage.

Dans la discussion qui précède, nous avons passé sous silence un protagoniste de taille, la constante de couplage qui peut s'avérer indispensable pour relier fonctions de corrélations et invariants. Des constantes de couplage sont déjà présentes dans les deux théories mentionnées ci-dessus et pour bien faire, il faudrait écrire le lagrangien de Witten donnant lieu aux invariants de Donaldson sous la forme $\mathcal{L}/e^2$ et remplacer le lagrangien de Chern-Simons $\mathcal{L}_{CS}$ par $\frac{1}{h}\mathcal{L}_{CS}$, e et h étant les constantes de couplage. La méthode de la phase stationnaire redonne les solutions classiques en passant à la limite $e \rightarrow 0$ et $h \rightarrow 0$, celles-ci correspondant aux instantons dans le premier cas et aux connections plates dans le deuxième. Dans le premier cas, bien que les invariants construits à partir de ce lagrangien soient indépendants du choix de la constante de couplage e , c'est elle qui jouera un rôle prépondérant dans ce qui suit. Dans le cas du lagrangien de Chern-Simons, les coefficients du développement asymptotique de certaines fonctions de corrélation autour de $h = 0$ sont directement liés à l'intégrale de Kontsevich et donc aux invariants de Vassiliev.

L'idée de V.A. Vassiliev était au départ (1989) d'étudier les invariants de nœuds de la façon suivante : on considère l'espace de toutes les applications différentiables du cercle S^1 dans $\mathbb{R}^3$. L'espace des nœuds est le complémentaire dans cet espace de la partie non générique (ou ensemble des nœuds singuliers), appelée aussi « discriminant ». L'étude de la topologie du discriminant permet – au moyen de la dualité d'Alexander – d'obtenir des informations sur l'espace des nœuds. Par la suite, J. Birman et X-S. Lin ont donné une définition directe des invariants de Vassiliev en terme de nœuds singuliers et prouvé que la plupart des invariants connus sont des invariants de Vassiliev. Puis M. Kontsevich, inspiré par la théorie des graphes de Feynman, a construit une intégrale – appelée depuis intégrale de Kontsevich – qui constitue en

⁽⁹⁾« Quantum Field Theory and the Jones Polynomial », Comm. Math. Phys. 121 (1989)

quelque sorte l'invariant de Vassiliev universel. Ces travaux ont été améliorés par Dror Bar-Natan qui donne⁽¹⁰⁾ la définition heuristique suivante : « An invariant V is called (a Vassiliev invariant) of type m if its $(m + 1)$ st derivative vanishes ». En utilisant les idées évoquées ci-dessus et le « calcul de Kirby », T.T.Q. Le, J. Murakami et T. Ohtsuki ont construit⁽¹¹⁾ une intégrale de Kontsevich pour les variétés de dimension 3, permettant ainsi d'étendre à ces variétés la notion d'invariant de type fini ; ceci fait l'objet du chapitre 3 de ce volume, dû à Pierre Vogel.

Un des problèmes rencontrés en théorie des champs quantiques est de faire des prédictions raisonnables pour des théories de jauge non abéliennes (*i.e.* le groupe de structure G est non abélien) lorsque le couplage est fort, autrement dit, lorsqu'on ne peut raisonnablement pas faire des développements de Taylor en le paramètre de couplage (par exemple la constante e ci-dessus) parce que celui-ci est de taille non négligeable. C'est à partir de ses travaux⁽¹²⁾ avec Nathan Seiberg sur la théorie supersymétrique $N = 2$ en dimension 4 qu'Edward Witten a ensuite proposé⁽¹³⁾ une autre approche des invariants de Donaldson qui a donné lieu à la théorie appelée depuis « de Seiberg-Witten », présentée par John Morgan dans le deuxième chapitre de ce volume. Witten proposa d'étudier la théorie des champs « duale », qui contrairement à la théorie initiale, offre l'avantage d'être faiblement couplée (une grande constante de couplage e se transforme en une petite constante de couplage $1/e$) ce qui permet de la traiter perturbativement (*i.e.* par le biais de développements de Taylor en la constante de couplage). Les concepts de « dualité » de « couplage faible, fort » mettent en jeu les procédés de renormalisation que Daniel Bennequin décrit plus en détail dans la postface de ce volume.

L'histoire se prolonge ensuite avec les invariants de Gromov-Witten décrits par Michèle Audin dans le premier chapitre. Ils sont issus d'idées de Gromov sur les courbes pseudo-holomorphes, de Kontsevich sur des espaces de module de surfaces de Riemann⁽¹⁴⁾ et de Witten qui là encore pense en terme de fonctions de corrélation⁽¹⁵⁾. Le lagrangien $\mathcal{L}_\sigma$ est ici celui d'un sigma modèle supersymétrique – encore une fois il s'agit d'un modèle topologique – défini sur les applications $\phi : \Sigma \rightarrow M$ d'une surface de Riemann Σ vers une variété M . Les fonctions de corrélations correspondent à des

⁽¹⁰⁾ « On the Vassiliev knot invariants », Topology 34 (1995)

⁽¹¹⁾ « On a universal perturbative invariant of 3-manifold » Topology 37 (1998)

⁽¹²⁾ « Electric-magnetic duality, monopole condensation and confinement in $N = 2$ supersymmetric Yang-Mills theory », Nucl. Physics B 426 (1994)

⁽¹³⁾ « Monopoles and four-manifolds », Math. Research Letters 1 (1994)

⁽¹⁴⁾ « Gromov-Witten classes, Quantum cohomology, and enumerative geometry », Maxim Kontsevich et Yuri Manin, Comm. Math. Phys. (1994)

⁽¹⁵⁾ « Two-dimensional gravity and intersection theory on moduli space », Surveys in Differential Geometry (1991)

nombres d'intersection :

$$\begin{aligned} \langle O_{\alpha_1}(x_1) \cdots O_{\alpha_k}(x_k) \rangle &:= \int_{\text{Map}(\Sigma, M)} O_{\alpha_1}(x_1) \cdots O_{\alpha_n}(x_n) e^{-\mathcal{L}_\sigma(\phi)} \mathcal{D}\phi \\ &= \langle W_{\alpha_1} \smile \cdots \smile W_{\alpha_n}, [\mathcal{S}] \rangle \in \mathbb{Z} \end{aligned}$$

$x_1, \dots, x_k$ étant k points marqués sur Σ , $\mathcal{D}\phi$ une mesure de volume formelle sur l'espace des « chemins » $\text{Map}(\Sigma, M)$ et $O(x_i)$ des opérateurs à valeurs dans les distributions. Étant donnés $\alpha_1, \dots, \alpha_n \in H^*(M)$ on a posé $W_{\alpha_i} = (\text{év}_{x_i})^*(\alpha_i) \in H^*(\mathcal{S})$, où $\mathcal{S}$ est l'espace des modules des applications holomorphes $\phi : \Sigma \rightarrow M$, $[\mathcal{S}]$ est sa classe fondamentale et l'évaluation est définie⁽¹⁶⁾ par :

$$\begin{aligned} \text{év}_x : \mathcal{S} &\longrightarrow M \\ \phi &\longmapsto \phi(x). \end{aligned}$$

Dans la théorie de Gromov-Witten, lorsque la variété M est de Calabi-Yau, surgit aussi une dualité appelée symétrie miroir⁽¹⁷⁾ ; on peut, d'après les travaux de Givental⁽¹⁸⁾, associer à M une variété miroir $\widehat{M}$ dont les invariants se déduisent de ceux de M par échange des déformations complexes et kähleriennes. Ceci se lit du point de vue physique comme une dualité échangeant deux modèles sigma supersymétriques construits à partir de lagrangiens $\mathcal{L}$ et $\widehat{\mathcal{L}}$ sur des « chemins » $\phi : \Sigma \rightarrow M$ et $\widehat{\phi} : \Sigma \rightarrow \widehat{M}$. Enfin, lorsque M est une variété symplectique de dimension 4, comme l'explique John Morgan dans le Théorème IV.1.4, les invariants de Gromov-Witten sont liés aux invariants de Seiberg-Witten, ce qui nous ramène à notre point de départ.

Ainsi cette excursion succincte dans les développements de la théorie des champs des quinze dernières années nous a-t-elle naturellement conduits aux trois types d'invariants traités dans ce volume. Ces rencontres ne sont probablement pas fortuites comme le suggèrent par exemple des travaux de Dan Freed qui propose⁽¹⁹⁾ une interprétation de certains invariants que nous avons rencontrés comme des classes de cohomologie à l'aide d'intégrales de chemins généralisées (elles ne sont plus à valeurs réelles ou complexes mais dans des fibrés en droites). Sa proposition ne s'applique cependant rigoureusement qu'au cas très particulier où les intégrales de chemin se réduisent à des sommes discrètes.

F. Dumas, J.-Y. Le Dimet, S. Paycha

⁽¹⁶⁾ voir Chapitre 1 de ce volume

⁽¹⁷⁾ « Symétrie miroir », Claire Voisin, Panoramas et Synthèses, 2, Soc. Math. de France, 1996

⁽¹⁸⁾ « Homological geometry and mirror symmetry », I.C.M. Zürich (1994), Birkhäuser (1995)

⁽¹⁹⁾ « Characteristic numbers and generalized path integrals », in Geometry, Topology and Physics for Raoul Bott, (1994)

INVARIANTS EN GÉOMÉTRIE SYMPLECTIQUE VIA LES COURBES HOLOMORPHES

par

Michèle Audin

... sem falar que o número é de todas as coisas que há no mundo a menos exacta, diz-se quinhentos tijolos, diz-se quinhentos homens, e a diferença que há entre tijolo e homem é a diferença que se julga não haver entre quinhentos e quinhentos, quem isto não entender à primeira vez não merece que lho expliquem segunda^().*

Résumé. — Dans ce texte, je survole le problème des invariants des variétés symplectiques en m’attachant particulièrement à ceux qu’on obtient en « comptant » les courbes pseudo-holomorphes, les invariants de Gromov-Witten.

Abstract (Invariants in symplectic geometry via holomorphic curves). — In this survey, I discuss the problem of finding invariants for symplectic manifolds. I emphasize on the invariants that “count” holomorphic curves, namely on the Gromov-Witten invariants.

En géométrie et topologie aujourd’hui, les invariants affluent. Plus ou moins faciles à définir, ils sont en général difficiles à calculer.

Dans cet article, je vais présenter les invariants de Gromov-Witten. Même s’ils s’organisent de façon plus sophistiquée, ces invariants sont, en fin de compte, des *nombres*. Ici on compte des courbes pseudo-holomorphes. Par exemple, compter les droites du plan passant par deux points donnés, c’est calculer un invariant de Gromov-Witten. En général, il s’agit d’évaluer une classe de cohomologie — dans cet exemple, celle représentée par un couple de points du plan — sur la classe fondamentale d’un espace de courbes — dans cet exemple, l’espace des droites du plan.

Classification mathématique par sujets (2000). — 53C15, 14H15, 32G15, 53C23.

Mots clefs. — variétés symplectiques, variétés projectives complexes, courbes holomorphes, pseudo-holomorphes, espaces de modules des courbes, espaces de modules d’applications stables, invariants de Gromov-Witten.

(*)... sans compter que le nombre est de toutes les choses qui existent au monde la moins exacte, on dit cinq cents briques, on dit cinq cents hommes, et la différence qu’il y a entre une brique et un homme est la différence qu’on croit ne pas exister entre cinq cents et cinq cents, celui qui ne comprend pas cela du premier coup ne mérite pas qu’on le lui explique une seconde fois [70].

Le plan dont il est question dans cet exemple simple est le plan projectif complexe $\mathbf{P}^2(\mathbf{C})$, une des variétés symplectiques les plus simples qu'on puisse imaginer. Pour définir des nombres analogues dans une variété symplectique plus générale, il faudra d'abord avoir un analogue de la structure de variété complexe de $\mathbf{P}^2(\mathbf{C})$, pour pouvoir considérer des courbes « holomorphes » comme le sont les droites du plan. S'il y a toujours beaucoup de ces structures presque complexes, il n'y en a pas en général de meilleure que les autres. S'il y a aussi, localement, beaucoup de ces courbes pseudo-holomorphes, il n'est pas toujours vrai qu'elles forment de beaux espaces lisses et compacts comme celui des droites du plan. Il suffit d'avoir assisté à la convergence d'une famille de coniques lisses vers une conique dégénérée pour commencer à imaginer le problème.

L'utilisation des courbes pseudo-holomorphes initiée par Gromov dans [32], l'idée de Witten des invariants [82] et celle de Kontsevich des espaces d'applications stables [44], convenablement raffinées, permettent de se tirer de bien des mauvais pas et de définir, maintenant pour toutes les variétés symplectiques, des « invariants de Gromov-Witten » qui sont, en effet, des invariants du type de déformation de la variété symplectique.

Dans l'introduction, j'explique pourquoi on a besoin d'aller chercher des invariants si compliqués : il n'y a pas d'invariant local, pas d'invariant élémentaire global fin. J'explicitai ensuite les remarques ci-dessus : il n'est pas si facile de compter les courbes pseudo-holomorphes... mais le faire fournit effectivement des invariants pour les variétés symplectiques.

Ce texte. — La théorie est assez ardue dans sa réalisation. Il n'était donc pas question de donner des démonstrations complètes de résultats souvent très techniques. Pour ce survol, j'ai essayé de donner quelques idées, notamment des difficultés du sujet et de stratégies possibles pour les surmonter. J'ai donné des références nombreuses et, je l'espère, pas trop incomplètes. J'espère aussi que les nombreuses figures⁽¹⁾ présentes dans ce texte aideront à comprendre les principaux exemples et problèmes, permettant ainsi d'accéder plus facilement à la compréhension des articles originaux.

Remerciements. — Je remercie Sylvie Paycha pour sa lecture attentive d'une première version de ce texte, Daniel Bennequin, Philippe Eyssidieux, Pierre Vogel, Jean-Yves Welschinger et tous les participants des journées de Clermont-Ferrand pour leurs questions, remarques ou commentaires, les trois rapporteurs pour leur aide et surtout Pierre Baumann pour ses innombrables suggestions.

Note bibliographique. — Il y a de nombreux livres de géométrie symplectique sur le marché [50, 58, 3, 7] par exemple. Pour une introduction rapide, efficace et inspirée, je conseille vivement [1]. Une référence commode et assez complète pour

⁽¹⁾Je pense notamment aux figures 11, 12 et 13.

les propriétés de base des courbes holomorphes dans les variétés symplectiques et presque complexes est le livre [7], que je cite donc systématiquement, mais j'aime bien aussi [57]. J'ai aussi beaucoup utilisé (et même pillé) les articles ou ouvrages [47, 57, 4, 5], dont je remercie les auteurs pour leur aide involontaire.

Notations, dimensions. — Toutes les variétés différentielles dont il est question dans ce texte ont une dimension paire. Beaucoup d'entre elles sont complexes. C'est une source de confusion pour les dimensions. J'appelle *dimension* la dimension comme variété différentielle (réelle, donc). Quand j'utilise la dimension complexe, je l'appelle toujours *dimension complexe*, de sorte qu'il ne devrait pas y avoir d'ambiguïté. Par contre, toutes les *courbes* sont des courbes complexes — c'est-à-dire de dimension 2. On appelle « courbe rationnelle » une courbe de genre 0, c'est-à-dire une sphère ou un $\mathbf{P}^1(\mathbf{C})$.

1. Introduction : invariants, invariants mous, invariants grossiers

Une *variété symplectique* (W, ω) est une variété W munie d'une forme différentielle ω , de degré 2, qui est fermée et non-dégénérée, toutes propriétés qu'on résume en disant que ω est une *forme symplectique*.

Les premiers exemples de variétés symplectiques sont :

- d'abord $\mathbf{R}^{2n}$ avec une forme bilinéaire alternée non dégénérée, qui, dans une base convenable, peut toujours s'écrire $\omega_0 = \sum dx_i \wedge dy_i$
- et plus généralement les fibrés cotangents des variétés,
- toutes les surfaces orientables munies de formes volumes : toutes les 2-formes sur les surfaces sont fermées !
- toutes les variétés kählériennes et notamment toutes les variétés projectives complexes (j'utiliserai beaucoup les espaces projectifs complexes, voir le §6 pour ces exemples)
- et il y en a bien d'autres (kähleriennes ou non).

1.a. Invariants ? — Appelons *difféomorphisme symplectique*⁽²⁾ un difféomorphisme $\varphi : W_1 \rightarrow W_2$ qui préserve les formes symplectiques, c'est-à-dire, satisfait à $\varphi^* \omega_2 = \omega_1$.

On va s'intéresser ici à des invariants du type de difféomorphisme symplectique. Autant supposer la variété W fixée et considérer les formes symplectiques qu'elle admet, à difféomorphisme près. On considère donc l'ensemble $\mathcal{W}$ des formes symplectiques sur W . C'est un ouvert (condition de non-dégénérescence) de l'espace vectoriel⁽³⁾ des 2-formes fermées.

⁽²⁾C'est ce que la plupart des auteurs appellent un « symplectomorphisme ».

⁽³⁾muni, par exemple, de la topologie $\mathcal{C}^\infty$

Le groupe $\text{Diff}(W)$ des difféomorphismes de W opère sur $\mathcal{W}$. Un invariant à valeurs dans un espace topologique E est une application continue

$$\mathcal{W}/\text{Diff}(W) \longrightarrow E.$$

Par exemple l'application identique, avec $E = \mathcal{W}/\text{Diff}(W)$... mais ça ne fait pas avancer la science. L'idée est plutôt de prendre pour E un espace plus simple que $\mathcal{W}/\text{Diff}(W)$. Autre extrême, pas plus productif, toute application constante est un invariant.

Ici on utilisera des espaces discrets, notamment l'ensemble $\mathbf{Z}$ des nombres entiers. Un invariant à valeurs dans un espace discret (dit « invariant discret ») est constant sur les composantes connexes de $\mathcal{W}/\text{Diff}(W)$, on dit que c'est un invariant du type de déformation de la variété symplectique. Concrètement : deux formes ω_0 et ω sont dans la même classe de déformation s'il existe un difféomorphisme φ de W et un chemin ω_t ($t \in [0, 1]$) de formes symplectiques tels que $\varphi^*\omega = \omega_1$.

L'exemple trivial de l'invariant constant montre qu'à la difficulté de construire des invariants peut s'ajouter celle de montrer qu'ils ne sont pas trop grossiers, qu'ils peuvent effectivement distinguer deux formes symplectiques.

1.b. Pas d'invariant local. — Localement, toutes les variétés symplectiques d'une dimension donnée sont identiques : le théorème de Darboux affirme qu'elles possèdent, au voisinage de chaque point, des coordonnées locales dans lesquelles la forme symplectique s'écrit $\omega = \sum dx_i \wedge dy_i$.

Théorème 1.1 (Darboux). — Soit (W, ω) une variété symplectique de dimension $2n$ et soit x un point de W . Il existe un voisinage U de x et un difféomorphisme symplectique de $(U, \omega|_U)$ dans un ouvert de $(\mathbf{R}^{2n}, \omega_0)$.

Il n'y a donc pas d'invariant *local* du type de difféomorphisme symplectique d'une variété symplectique (W, ω) , comme l'est la courbure pour une variété riemannienne. Par exemple, si la courbure de la sphère empêche de dessiner des cartes « isométriques » de la Terre, il n'y a rien qui empêche de dessiner des cartes (locales) dans lesquelles les *aires* sont exactes, ces cartes sont dites « équivalentes ».

1.c. Structures presque complexes et invariants mous. — Pour une topologie classique, pensant en termes de groupes structurels, il n'y a pas de grande différence entre une 2-forme non dégénérée et une structure presque complexe.

– Dire qu'une variété est munie d'une 2-forme non dégénérée, c'est dire que le groupe structurel de son fibré tangent (qui est, *a priori*, $\text{GL}(2n; \mathbf{R})$) se réduit à $\text{Sp}(2n; \mathbf{R})$

– alors que dire qu'elle est presque complexe, c'est dire que ce groupe se réduit à $\text{GL}(n; \mathbf{C}) \subset \text{GL}(2n; \mathbf{R})$.

Or ces deux sous-groupes de $GL(2n; \mathbf{R})$ ont la même composante compacte, à savoir le groupe unitaire $U(n)$, qui est d'ailleurs leur intersection.

Un autre éclairage sur cette question est celui du géomètre kählérien. Une variété kählérienne est une variété analytique complexe sur laquelle est donnée une métrique hermitienne $(X, Y) \mapsto \langle X, Y \rangle_x$, que l'on peut décomposer en

$$\langle X, Y \rangle_x = g_x(X, Y) - i\omega_x(X, Y)$$

expression dans laquelle

- la partie réelle g est une métrique riemannienne sur la variété (réelle sous-jacente)
- et ω_x est, pour tout x , une forme bilinéaire (réelle) alternée : ω est une 2-forme différentielle.

Je laisse les lecteurs vérifier que $\omega(X, iY) = g(X, Y)$ et que la multiplication par i est une isométrie (de g comme de ω).

En géométrie kählérienne, on peut donc jouer à la fois avec une métrique riemannienne, une forme symplectique et une structure complexe. En géométrie symplectique, on ne dispose que de la forme symplectique ω . On introduit des *ersatz* de i en copiant ces propriétés. On appellera structure presque complexe *calibrée* par une forme ω sur la variété W une section J de $\text{End}(TW)$ telle que :

- On a $J_x^2 = -\text{Id}_{T_x W}$ en tout point x de W : c'est dire que J définit une structure de fibré vectoriel complexe sur TW (penser à la multiplication par i dans un espace vectoriel complexe).
- L'application $(X, Y) \mapsto \omega(X, JY)$ définit une métrique riemannienne sur W pour laquelle J est une isométrie.

Il y a beaucoup de structures presque complexes calibrées par une forme symplectique. Il n'y a pas en général de choix naturel d'une telle structure. Une des idées fécondes de Gromov dans [32] est justement de considérer l'ensemble de toutes ces structures. Il n'est pas difficile de démontrer la proposition suivante (c'est essentiellement la remarque, faite ci-dessus, que le groupe $\text{Sp}(2n; \mathbf{R})$ se rétracte sur $U(n)$, voir, par exemple, [2]).

Proposition 1.2. — *L'espace de toutes les structures presque complexes sur la variété W qui sont calibrées par la forme symplectique ω n'est pas vide et est contractile.*

Notons qu'en particulier, cet espace est connexe. Tous les invariants discrets de la variété presque complexe (W, J) (J calibrée par ω) se trouvent donc être, naturellement, des invariants⁽⁴⁾ de la variété symplectique (W, ω) . C'est le cas par exemple des classes de Chern du fibré vectoriel complexe (TW, J) . En particulier, une variété symplectique a des classes de Chern — ne dépendant pas de la structure presque complexe adaptée choisie.

⁽⁴⁾À condition toutefois qu'ils soient définis pour *toute* structure presque complexe. Voir la remarque 4.9.

Remarque 1.3. — Si ces invariants sont dits « mous⁽⁵⁾ », c’est parce qu’on n’utilise jamais, pour construire les structures presque complexes qui leur donnent naissance, le fait que la forme symplectique est *fermée*.

1.d. Un invariant bien grossier, le volume. — L’exemple des cartes de la Terre au § 1.b est un exemple de dimension 2. Dans cette dimension, il n’y a pas de différence entre une forme volume et une forme symplectique. En toute généralité, sur une variété symplectique (W, ω) de dimension $2n$, la forme $\omega^{\wedge n}$ est une forme volume (c’est précisément ce que dit la condition de non-dégénérescence) et donc les difféomorphismes symplectiques préservent le volume total de la variété.

On peut donc conclure que le volume total de la variété symplectique est un invariant de son type de difféomorphisme symplectique. Par exemple, il n’y a pas de difféomorphisme symplectique d’une grosse boule de $\mathbf{R}^{2n}$ dans une boule plus petite... mais cet invariant est bien grossier.

La « géométrie du volume » n’est pas très rigide. En voici deux illustrations.

– Un théorème classique de Moser [60] affirme que deux formes volume sur une variété sont difféomorphes si et seulement si elles donnent le même volume à cette variété.

– Si U et V sont deux ouverts connexes de $\mathbf{R}^n$ et si $\text{vol}(U) < \text{vol}(V)$, alors on peut plonger U dans V en préservant la forme volume.

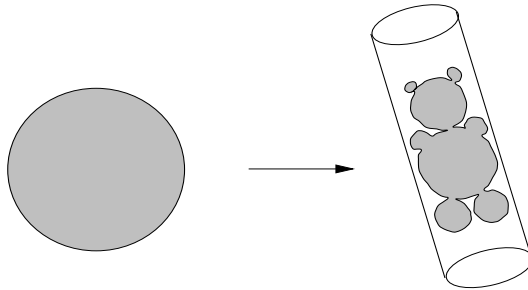


FIGURE 1. Une grosse boule dans un fin cylindre

La « largeur » de Gromov. — La situation en géométrie symplectique est beaucoup plus rigide. Un des premiers effets de l’introduction des courbes pseudo-holomorphes par Gromov (j’anticipe un peu sur le § suivant) est un résultat de géométrie symplectique complètement opposé : un exemple d’un ouvert de volume fini (une grosse

⁽⁵⁾Le qualificatif « mou » s’oppose à « dur », en anglais “soft” vs “hard”, la distinction est due à Gromov [33].

boule) qu'on ne peut pas plonger dans un autre, dont le volume est pourtant infini (un cylindre infini mais fin).

Théorème 1.4 (Gromov). — *S'il existe un plongement symplectique de la boule B_R^{2n+2} dans le cylindre $B_r^2 \times \mathbf{R}^{2n}$, alors le rayon R de la boule est au plus égal à celui, r , du cylindre.*

En fait, Gromov a défini un *invariant symplectique* plus fin que le volume, la *largeur*, qui est R pour la boule, r pour le cylindre et qui interdit le plongement quand $R > r$.

On peut comprendre ce théorème comme un « principe d'incertitude » en mécanique classique. C'est un point de vue que C. Viterbo défend dans un joli article grand public qui reste une introduction très recommandable à la topologie symplectique [78], point de vue repris dans un autre article grand public plus généraliste [6].

1.e. « Pseudo-holomorphic curves in symplectic manifolds ». — Classes de Chern, volume, on en était plus ou moins là en 1985, lorsque Gromov a montré, dans l'article [32] éponyme de ce paragraphe, qu'en étudiant les courbes pseudo-holomorphes qu'elles contiennent, on pouvait en apprendre beaucoup plus sur les variétés symplectiques⁽⁶⁾. Le théorème 1.4 n'est qu'une des manifestations de l'efficacité de ce nouvel outil.

En géométrie analytique complexe, on fait un grand usage des fonctions holomorphes. En général, sur une variété presque complexe (W, J) , il y a peu, voire pas, de fonctions holomorphes (même locales). Une fonction f sur un ouvert U de W est holomorphe si sa différentielle est, en tout point de U , une forme linéaire *complexe* :

$$\forall w \in W, \quad \forall X \in T_w W, \quad (df)_w(JX) = i(df)_w(X).$$

De façon heuristique, disons que s'il y avait beaucoup de fonctions holomorphes, on pourrait en faire des cartes locales analytiques, J serait *intégrable* et la variété serait analytique complexe.

La grande idée de Gromov, c'est que, par contre, une variété presque complexe contient toujours beaucoup de courbes holomorphes.

Si Σ est une surface de Riemann (sa structure complexe j est fixée), une application $u : \Sigma \rightarrow W$ est dite *J -holomorphe* (ou pseudo-holomorphe ou plus simplement holomorphe si on ne tient pas à préciser pour quelle structure presque complexe) si son application tangente est linéaire sur $\mathbf{C}$, c'est à dire si

$$T_z u \circ j(\zeta) = J \circ T_z u(\zeta), \quad \forall z \in \Sigma, \quad \forall \zeta \in T_z \Sigma.$$

C'est la même condition que pour une fonction, à cela près que la variété est devenue l'espace but.

⁽⁶⁾L'utilisation des courbes dans une variété pour comprendre la géométrie de cette variété fait penser au programme de Mori pour la classification des variétés algébriques, voir [14] (et sa bibliographie).

Espaces de courbes. — Ce sont les espaces de modules des courbes pseudo-holomorphes qui recèlent les *invariants de Gromov-Witten*, qui font l’objet de ce texte. Déjà présents dans [32], ces espaces de modules ont été tout de suite utilisés avec succès par McDuff pour classifier les variétés symplectiques de dimension 4 qui sont des fibrés en $\mathbf{P}^1(\mathbf{C})$ [55, 48]. Des formes précoces ont été définies et utilisées par Ruan [65, 66] pour distinguer des formes symplectiques sur certaines variétés symplectiques de dimension 6 (voir le § 2.e). Ils jouent la vedette en cohomologie quantique. Mentionnons enfin que les courbes pseudo-holomorphes ont aussi des applications en analyse complexe (voir par exemple [41]).

2. Les courbes pseudo-holomorphes, une introduction aux invariants de Gromov-Witten

Le but est ici de convaincre les lecteurs que les courbes pseudo-holomorphes peuvent être utiles en géométrie symplectique, en indiquant comment on les utilise pour interdire à la grosse boule de se plonger dans le fin cylindre. Ce schéma de la démonstration du théorème 1.4 met en effet en jeu les propriétés les plus importantes des espaces de courbes pseudo-holomorphes. Je l’utilise donc ici comme introduction aux idées les plus simples de la construction des invariants de Gromov-Witten — que je n’analyserai plus précisément qu’au § 3.

2.a. Existence locale. — J’ai mentionné le fait qu’il y avait beaucoup de courbes holomorphes dans une variété presque complexe. En coordonnées locales

$$u : z = x + iy \in \Sigma \longmapsto (u_1(x, y), \dots, u_{2n}(x, y)) \in W,$$

la condition d’holomorphicité s’écrit

$$\frac{\partial u_j}{\partial y} = \sum_{k=1}^{2n} J_{j,k}(u(x, y)) \frac{\partial u_k}{\partial x} \quad (1 \leq j \leq 2n).$$

C’est un système différentiel de $2n$ équations aux dérivées partielles en $2n$ fonctions inconnues.

Si la variété W est analytique réelle et si J est une section analytique réelle de $\text{End}(TW)$, le théorème de Cauchy-Kowalevskaya implique l’existence locale de solutions analytiques. Ceci reste vrai avec des hypothèses $\mathcal{C}^\infty$ parce que le système est quasi-linéaire (linéaire en les dérivées partielles) et elliptique. En effet, le système linéarisé n’est autre que le système des équations de Cauchy-Riemann, qui est elliptique. Voir aussi le § 2.d.

Les courbes pseudo-holomorphes jouissent de toutes les propriétés de régularité des solutions d’un système elliptique. Voir par exemple [73].

2.b. Application à la boule dans le cylindre. — Donnons ici une idée de la façon dont on peut utiliser les courbes pseudo-holomorphes pour démontrer le théorème 1.4 : les deux propriétés essentielles attendues des espaces de modules des courbes pseudo-holomorphes, à savoir lissité et compacité, sont en effet présentes dans cette démonstration.

On utilise une variété symplectique produit $W = S^2 \times V$, que je préfère⁽⁷⁾ écrire $\mathbf{P}^1(\mathbf{C}) \times V$, où l'on suppose que V est une variété symplectique « sans sphères », c'est-à-dire telle que $\pi_2(V) = 0$. Dans les applications, ce sera un tore. On démontre d'abord le théorème suivant.

Théorème 2.1 (Gromov [32]). — *Soit (V, σ) une variété symplectique compacte sans sphères et soit J une structure presque complexe sur $\mathbf{P}^1(\mathbf{C}) \times V$ calibrée par la forme symplectique $\omega \oplus \sigma$. Tout point (x, v) de $\mathbf{P}^1(\mathbf{C}) \times V$ est contenu dans exactement une J -sphère de la classe d'homologie de la fibre $\mathbf{P}^1(\mathbf{C}) \times \{v\}$ de v .*

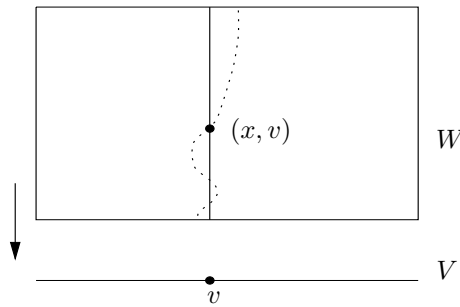


FIGURE 2

Voir la figure 2 pour une explication de cet énoncé : si la structure presque complexe J est le produit de structures presque complexes sur les facteurs, les fibres de la projection $\mathbf{P}^1 \times V \rightarrow V$ sont évidemment les J -sphères dont l'énoncé affirme l'existence et l'unicité... pour *toute* structure J .

Voici une idée de la démonstration. On considère l'ensemble $\mathcal{C}$ des couples (u, J) où J est une structure presque complexe calibrée et u une J -sphère de W dans la classe d'homologie de la fibre. On peut munir l'ensemble des couples (u, J) de classe $\mathcal{C}^k$ d'une structure de variété modelée sur un espace de Hölder ou de Sobolev (voir [35] et [12]), ceci définissant, quand k varie, une structure différentiable sur $\mathcal{C}$.

⁽⁷⁾À partir d'ici, je vais faire un grand usage des espaces projectifs avec leurs structures complexe, kählérienne, etc. usuelles. En cas de besoin, consulter le § 6.

Soit p la projection de $\mathcal{C}$ sur l'espace $\mathcal{J}$ des structures presque complexes calibrées. Le point clef est de montrer que p est surjective. Comme nous savons depuis la proposition 1.2 que $\mathcal{J}$ est connexe, il suffit de montrer que l'image de $\mathcal{C}$ est ouverte, fermée et non vide.

(1) On a déjà remarqué que $p(\mathcal{C})$ n'est pas vide : si J est une structure produit, elle est l'image de n'importe quel couple (u, J) où u représente une fibre.

(2) On montre que $p(\mathcal{C})$ est ouverte parce que p se linéarise en une application surjective. Il faut utiliser là un théorème d'annulation et un calcul d'indice (voir [32], [54]) basé sur le théorème de Riemann-Roch. Voir le § 2.d.

(3) Enfin, $p(\mathcal{C})$ est fermée grâce à un théorème de compacité, basé sur une propriété générale d'équicontinuité [32, 61]. Remarquons quand même que l'espace des applications pseudo-holomorphes $\mathbf{P}^1(\mathbf{C}) \rightarrow W$ n'est pas compact : le groupe « de reparamétrage » $\mathrm{PGL}(2; \mathbf{C})$, groupe des automorphismes de $\mathbf{P}^1(\mathbf{C})$, n'est pas compact ! On ne peut même pas espérer que l'espace des courbes J -holomorphes (non paramétrées) soit compact. Penser à une famille de coniques planes convergeant vers la réunion de deux droites (voir la figure 5). En fait c'est ce qui peut se produire de pire. Je reviendrai longuement sur cet exemple et sur le théorème de compacité aux §§ 3.b et 3.c. En tout cas ici, comme la classe d'homologie des fibres ne peut pas se décomposer, ce pire ne peut pas se produire et on a, effectivement, compacité. $\square$

Avec le même schéma de démonstration, on peut montrer, par exemple, un autre résultat impressionnant d'existence de courbes sur lequel je reviendrai au § 5.c.

Théorème 2.2 (Gromov [32]). — *Soit J une structure presque complexe calibrée par la forme de Kähler usuelle de $\mathbf{P}^2(\mathbf{C})$. Par deux points de $\mathbf{P}^2(\mathbf{C})$ passe une et une seule J -courbe de la classe d'homologie des droites projectives.* $\square$

Démonstration du théorème 1.4. — Armés du théorème 2.1, montrons donc qu'on ne peut pas plonger la grosse boule dans le fin cylindre. Supposons, au contraire, que ce soit possible. Soit

$$\varphi : B_R^{2n+2} \hookrightarrow B_r^2 \times \mathbf{R}^{2n}$$

un plongement symplectique. La stratégie est, d'abord, de remplacer $B_r^2 \times \mathbf{R}^{2n}$ par une variété de la forme $\mathbf{P}^1(\mathbf{C}) \times V$ pour pouvoir appliquer le théorème 2.1.

Considérons une boule fermée $\overline{B} = \overline{B}_{R'}^{2n+2}$ contenue dans B_R^{2n+2} . Comme $\varphi(\overline{B})$ est compacte, elle est contenue dans un produit

$$B_r^2 \times \left[-\frac{A}{2}, \frac{A}{2} \right]^{2n} \text{ pour } A > 0 \text{ assez grand.}$$

Remplaçons donc $\mathbf{R}^{2n}$ par $\left[-\frac{A}{2}, \frac{A}{2} \right]^{2n}$, ou, mieux, par le tore $T^{2n} = \mathbf{R}^{2n}/A\mathbf{Z}^{2n}$. La forme symplectique de $\mathbf{R}^{2n}$ étant invariante par translation, elle définit une forme symplectique σ sur T^{2n} .

On a donc un plongement symplectique de la boule ouverte $B_{R'}^{2n+2}$ dans $B_r^2 \times T^{2n}$. Le tore va être le V du théorème 2.1. Remplaçons maintenant le disque B_r^2 par $\mathbf{P}^1(\mathbf{C})$.

On vérifie directement que le disque B_r^2 est « symplectomorphe » au complémentaire d'un point dans la sphère de même volume, c'est-à-dire ici de rayon $r/2$. Munissons donc $\mathbf{P}^1(\mathbf{C})$ de la forme symplectique ω qui lui donne le volume πr^2 . On a maintenant un plongement symplectique de $B_{R'}^{2n+2}$ dans $\mathbf{P}^1(\mathbf{C}) \times T^{2n}$.

Prenons une boule de rayon R'' encore plus petit. Soit J_0 la structure complexe *standard* de $\mathbf{R}^{2n+2} = \mathbf{C}^{n+1}$. Comme l'application φ est un plongement symplectique, elle définit une structure complexe $\varphi_*(J_0)$ sur l'image $\varphi(\overline{B}_{R''})$, qui est calibrée par la restriction de la forme symplectique $\omega \oplus \sigma$ de $\mathbf{P}^1(\mathbf{C}) \times T^{2n}$. La proposition 1.2 ou une version relative permet de prolonger cette structure calibrée à tout $\mathbf{P}^1(\mathbf{C}) \times T^{2n}$. On applique maintenant le théorème 2.1 : il existe une courbe C holomorphe (pour cette structure bizarre) qui est homologue à $\mathbf{P}^1(\mathbf{C}) \times \{\star\}$ et passe par l'image $\varphi(0)$ du centre de la boule. On a

$$\int_C \omega \oplus \sigma = \langle \omega \oplus \sigma, [C] \rangle = \langle \omega \oplus \sigma, [\mathbf{P}^1(\mathbf{C}) \times \{\star\}] \rangle = \langle \omega, [\mathbf{P}^1(\mathbf{C})] \rangle = \pi r^2$$

par définition de ω . Pour l'intersection $C' = C \cap \varphi(\overline{B}_{R''})$, on a donc

$$\pi r^2 \geq \int_{C'} \omega \oplus \sigma = \int_{\varphi^{-1}(C')} \omega_0$$

puisque φ est un plongement symplectique.

Par construction-même, $\varphi^{-1}(C')$ est une courbe holomorphe pour la structure complexe standard de $\mathbf{R}^{2n+2}$ et son bord est contenu dans la sphère de rayon R'' . Maintenant, les courbes complexes d'une variété kählérienne sont minimales et les sous-variétés minimales satisfont à une propriété de monotonie, ici

$$\text{vol}(\overline{B}_{R''}^{2n+2} \cap \varphi^{-1}(C')) \geq \text{vol}(\overline{B}_{R''}^2) = \pi R''^2$$

(c'est, par exemple, le théorème 3.2.4 de [46]). On a donc $R'' < r$, ceci pour tout $R'' < R'$ et tout $R' < R$, donc $R \leq r$, la conclusion cherchée. $\square$

2.c. Invariants de Gromov-Witten, introduction. — On considère donc une variété symplectique (W, ω) et une structure presque complexe J calibrée par ω .

On va considérer des espaces de courbes J -holomorphes. Une courbe (paramétrée) $u : \Sigma \rightarrow W$ définit une classe d'homologie $A = u_*[\Sigma] \in H_2(W; \mathbf{Z})$. L'idée est de compter les courbes d'une classe d'homologie donnée satisfaisant certaines conditions d'incidence.

2.c.a. L'exemple des droites du plan. — Explicitons la différence entre « courbe » et « courbe paramétrée » sur un exemple simple. Considérons le cas où W est le plan projectif complexe (muni de sa structure complexe habituelle) et choisissons la classe d'homologie représentée par les droites projectives, c'est-à-dire le générateur L de $H_2(\mathbf{P}^2(\mathbf{C}); \mathbf{Z})$. Les courbes holomorphes qui représentent cette classe sont les

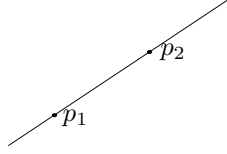


FIGURE 3. Par deux points...

courbes algébriques de degré 1 et donc, effectivement, les droites. Elles se laissent donc paramétrer par les applications « linéaires »

$$u : \mathbf{P}^1(\mathbf{C}) \longrightarrow \mathbf{P}^2(\mathbf{C}).$$

– L'espace de ces applications est de dimension complexe 5 : on donne u en donnant trois polynômes homogènes de degré 1 (ce qui fait six coefficients), à multiplication par un même scalaire près.

– L'espace des courbes non paramétrées est l'espace des droites de $\mathbf{P}^2(\mathbf{C})$, la grassmannienne $G_2(\mathbf{C}^3)$ (une autre version du plan projectif!), de dimension complexe 2.

La différence est le groupe de reparamétrage, le groupe des automorphismes de $\mathbf{P}^1(\mathbf{C})$ ici, à savoir $\mathrm{PGL}(2; \mathbf{C})$, qui est bien de dimension complexe 3, le compte est bon.

Toujours sur cet exemple simple, donnons aussi un exemple de condition d'incidence. Traduisons « par deux points passe une droite et une seule » en termes homologiques. On considère l'espace de modules $\mathcal{M}$, quotient de l'ensemble

$$\{(u, z_1, z_2) \mid z_1, z_2 \in \mathbf{P}^1(\mathbf{C}), z_1 \neq z_2, u : \mathbf{P}^1(\mathbf{C}) \rightarrow \mathbf{P}^2(\mathbf{C}), u_*[\mathbf{P}^1] = L\}$$

par $\mathrm{PGL}(2; \mathbf{C})$. C'est un fibré sur $G_2(\mathbf{C}^3)$ dont les fibres sont les couples de points distincts dans $\mathbf{P}^1(\mathbf{C})$. Donc la dimension complexe de $\mathcal{M}$ est 4.

Cet espace porte en lui une application d'évaluation

$$\begin{aligned} \mathcal{M} &\xrightarrow{\text{év}} \mathbf{P}^2(\mathbf{C}) \times \mathbf{P}^2(\mathbf{C}) \\ (u, z_1, z_2) &\longmapsto (u(z_1), u(z_2)). \end{aligned}$$

On remarquera que source et but ont la même dimension. L'assertion « par deux points... » est équivalente au fait que l'application év est de degré 1 : étant donné un point général⁽⁸⁾ (p_1, p_2) de $\mathbf{P}^2(\mathbf{C}) \times \mathbf{P}^2(\mathbf{C})$, son image inverse $\text{év}^{-1}(p_1, p_2)$ est constituée d'un unique élément (figure 3). En mentionnant le degré, j'ai passé sous silence le problème de compacité. Voir le § 2.d.

Retenons de cet exemple que « compter » les courbes satisfaisant certaines conditions d'incidence peut être, simplement, « calculer le degré » d'une application d'évaluation définie sur un espace de modules : l'idée est proche de celle utilisée par Donaldson [18] pour définir ses polynômes.

⁽⁸⁾ici, deux points p_1 et p_2 distincts dans $\mathbf{P}^2(\mathbf{C})$

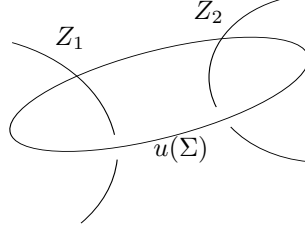


FIGURE 4

2.c.b. Espaces de modules et évaluation. — Revenons au cas général. On donne donc une classe $A \in H_2(W; \mathbf{Z})$ et des sous-variétés $Z_1, \dots, Z_m \subset W$ et on veut donner un sens au « nombre de courbes de genre g de la classe A coupant $Z_1, \dots, Z_m$ quelque part » (figure 4).

Bien sûr, ici, pour que l'invariant ne soit pas nul pour des raisons banales, il faut que la somme des codimensions des Z_i coïncide avec la dimension de l'espace des courbes de genre g de la classe A .

Comme dans l'exemple simple des droites de $\mathbf{P}^2(\mathbf{C})$, la stratégie est de définir un espace de modules $\overline{\mathcal{M}}_{g,m}^J(W, A)$ convenable, espace de courbes complexes de genre g avec m points marqués :

$$\overline{\mathcal{M}}_{g,m}^J(W, A) = \{(u, (z_1, \dots, z_m))\} / G,$$

G désignant bien sûr le groupe d'automorphismes de la situation.

Les points $z_1, \dots, z_m$ de Σ servent à construire les applications d'évaluation qui vont permettre d'imposer les conditions d'incidence :

$$\begin{aligned} \text{év}_m : \overline{\mathcal{M}}_{g,m}^J(W, A) &\longrightarrow W^m \\ (u, (z_1, \dots, z_m)) &\longmapsto (u(z_1), \dots, u(z_m)). \end{aligned}$$

Appelons $\xi_1, \dots, \xi_m$ les classes de cohomologie duales à $Z_1, \dots, Z_m$. Notre invariant devrait être

$$\Psi_{g,A}^J(\xi_1 \otimes \dots \otimes \xi_m) = \left\langle (\text{év}_m)^*(\xi_1 \otimes \dots \otimes \xi_m), [\overline{\mathcal{M}}_{g,m}^J(W, A)] \right\rangle \in \mathbf{Z}.$$

La difficulté est de construire les espaces de modules $\overline{\mathcal{M}}_{g,m}^J(W, A)$... et de démontrer que ceux-ci ont, sous certaines hypothèses, des classes fondamentales.

2.d. Quelques problèmes. — Ce sont, rapidement dit, la dépendance en J et les (graves) problèmes liés à l'existence d'une classe fondamentale : lissité, orientabilité, compacité, que je vais passer en revue dans ce paragraphe.

2.d.a. La dépendance en J . — On s'intéresse au nombre de courbes J -holomorphes de genre g dans la classe d'homologie $A \in H_2(W; \mathbf{Z})$, mais on aimerait que le résultat

ne dépende que de la forme symplectique et pas du choix de la structure presque complexe J que celle-ci calibre,

- soit parce que ce qu’on cherche, c’est un invariant de la forme symplectique,
- soit parce qu’il y a une structure complexe meilleure que les autres et que ce sont les courbes de celle-ci qu’on veut compter.

Un exemple naturel de cette dernière situation est celui de la géométrie énumérative des courbes planes ayant telle ou telle propriété d’incidence. Il s’agit ici de courbes holomorphes dans la variété (complexe!) $\mathbf{P}^2(\mathbf{C})$ et d’un problème très respectable (nombre de droites passant par deux points donnés, nombre de coniques passant par cinq points, *etc.*). Aucune géomètre algébriste ne se contentera d’un énoncé comme « Pour une structure presque complexe générique sur $\mathbf{P}^2(\mathbf{C})$, par deux points il passe une et une seule J -droite ». Ce qu’elle veut savoir, c’est combien de vraies droites pour la vraie structure complexe de $\mathbf{P}^2(\mathbf{C})$ il peut bien passer par ces deux points. Voir le § 3.d.a.

2.d.b. L’orientabilité. — Dans l’exemple de $\mathbf{P}^2(\mathbf{C})$ avec sa structure complexe usuelle et plus généralement d’une variété munie d’une structure complexe intégrable, l’espace de modules est, quand il est défini, complexe, donc le problème d’orientabilité ne se pose pas. Dans le cas non intégrable, il possède une structure presque complexe *stable*⁽⁹⁾ naturelle, qui définit son orientation (voir [54]).

2.d.c. La lissité. — Il y a d’abord un problème lié au fait que l’espace de modules est un *quotient*, mais ce problème n’est pas très grave, dès lors qu’on accepte des singularités du type *orbifold*, c’est-à-dire des quotients de variétés par des groupes finis. La seule retombée de cette tolérance est qu’alors la classe fondamentale est dans $H_*(\overline{\mathcal{M}}_{g,m}^J(W, A); \mathbf{Q})$ (pour avoir une dualité de Poincaré) et que les invariants obtenus sont *a priori* rationnels.

Il y a surtout un problème de régularité, pour que l’espace des applications holomorphes soit lisse. Si J est une structure presque complexe sur W , on peut considérer la partie anti- $\mathbf{C}$ -linéaire de du

$$\overline{\partial}_J u = \frac{1}{2}(du + J \circ du \circ j)$$

comme un élément de $\Omega^{0,1}(u^*TW)$ c’est-à-dire de $\mathcal{C}^\infty(\Lambda^{0,1}T^*\Sigma \otimes_J u^*TW)$.

Les courbes J -holomorphes sont les zéros (dans l’ensemble $\mathcal{X}(A)$ de toutes les applications représentant A) de $\overline{\partial}_J$ (c’est une autre façon d’écrire l’« équation de Cauchy-Riemann »). On peut considérer le « fibré vectoriel » $\mathcal{E} \rightarrow \mathcal{X}(A)$ dont la fibre $\mathcal{E}_u$ en $u \in \mathcal{X}(A)$ est $\Omega^{0,1}(u^*TW)$; $\overline{\partial}_J$ est alors une section de $\mathcal{E}$, et l’espace des courbes J -holomorphes de la classe A est l’ensemble des zéros de cette section. On peut alors

⁽⁹⁾C’est-à-dire que son fibré tangent possède une structure de fibré vectoriel complexe... une fois qu’on lui a ajouté un certain fibré trivial.

se demander si cette section est transverse à la section nulle, en d'autres termes si la composition D_u^J

$$\mathcal{C}^\infty(u^*TW) = T_u\mathcal{X} \xrightarrow{T_u(\bar{\partial}_J)} T_{(u,0)}\mathcal{E} = T_u\mathcal{X} \oplus \mathcal{E}_u \xrightarrow{\text{proj}} \mathcal{E}_u = \Omega^{0,1}(u^*TW)$$

est une application linéaire surjective pour toute solution u de $\bar{\partial}_Ju = 0$.

C'est un fait que l'application D_u^J est un opérateur de Fredholm (comme $\bar{\partial}$). Elle n'est pas en général linéaire sur $\mathbf{C}$ (elle l'est quand J est intégrable), mais sa partie antilinéaire est un opérateur d'ordre 0 (s'exprimant à l'aide du tenseur de Nijenhuis de J). On peut donc homotoper D_u^J , parmi les opérateurs de Fredholm, à sa partie $\mathbf{C}$ -linéaire, l'opérateur $\bar{\partial}$ de Dolbeault, qui définit une structure holomorphe sur u^*TW (voir le § 3.d.a). L'indice (réel) de D_u^J est alors donné par le théorème de Riemann-Roch (voir par exemple [27]) :

$$\text{Ind } D_u^J = 2\langle c_1, A \rangle + 2n(1 - g),$$

c'est la *dimension virtuelle* ou *attendue* de l'espace des courbes J -holomorphes de genre g dans la classe A (rappelons que $2n = \dim W$).

On dit que la structure presque complexe J est *régulière* pour la classe A si D_u^J est surjective pour toute courbe J -holomorphe $u : \Sigma \rightarrow W$ de la classe A . L'espace des courbes holomorphes est alors une variété et sa dimension est la dimension attendue.

Exemple 2.3. — Pour $W = \mathbf{P}^2(\mathbf{C})$, $A = L$ et $g = 0$, la formule donne une dimension virtuelle 10 pour l'espace des applications holomorphes de $\mathbf{P}^1(\mathbf{C})$ dans $\mathbf{P}^2(\mathbf{C})$ de la classe des droites. C'est cohérent avec l'évaluation faite ci-dessus au § 2.c.a : pour la structure complexe usuelle, la dimension complexe était 5.

Il n'est pas vrai en général que l'on ait une propriété de régularité assurant que l'espace des courbes J -holomorphes soit une variété de la dimension attendue. Illustrons ce problème par un exemple (ou plutôt un contre-exemple) simple.

Exemple 2.4. — Considérons l'espace $\tilde{\mathbf{P}}^2(\mathbf{C})$ obtenu en éclatant un point dans le plan projectif complexe $\mathbf{P}^2(\mathbf{C})$ et la classe E du diviseur exceptionnel (voir si nécessaire l'appendice (§ 6)). La formule de Riemann-Roch appliquée à la classe $2E$, à la structure complexe usuelle et aux courbes de genre nul ($\Sigma = \mathbf{P}^1(\mathbf{C})$) de cette classe donne une dimension virtuelle égale à 8. Mais les applications holomorphes de degré 2 de $\mathbf{P}^1(\mathbf{C})$ dans lui-même forment un espace de dimension (réelle) 10 : on donne deux polynômes homogènes de degré 2 soit six coefficients complexes, à multiplication par un même scalaire près, donc la dimension complexe est 5. Donc, pour toute structure complexe J assez proche de la structure standard, il y a une famille de dimension au moins 10 de courbes J -holomorphes rationnelles dans la classe $2E$. Le problème vient de l'existence de courbes qui sont des revêtements multiples (et ne se pose que pour des classes A de la forme mB pour $m \geq 2$). Voir aussi, pour cet exemple, les exemples 3.30 et le § 4.d.

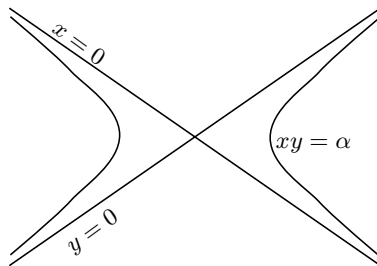


FIGURE 5. Conique dégénéralant

Le problème des revêtements multiples. — Considérons une classe d'homologie $A \in H_2(W; \mathbf{Z})$ d'une variété symplectique W de dimension $2n$ qui contient des courbes J -holomorphes rationnelles. Supposons la structure presque complexe calibrée J régulière pour la classe A , de sorte que la dimension de l'espace de ces courbes rationnelles (non paramétrées) est sa dimension virtuelle $2\langle c_1, A \rangle + 2n - 6$. Comme on a supposé que la classe A contient vraiment des courbes rationnelles, ce nombre est positif. Donc on a $\langle c_1, A \rangle \geq 3 - n$.

L'espace des courbes rationnelles J -holomorphes de la classe mA n'est pas vide non plus : en composant une courbe de la classe A avec une application de degré m de $\mathbf{P}^1(\mathbf{C})$ dans lui-même, on voit qu'il est de dimension au moins $4m + 2$ (l'espace des applications de degré m de $\mathbf{P}^1(\mathbf{C})$ dans lui-même est de dimension complexe $2m + 1$).

Or, la dimension virtuelle de cet espace est $2m\langle c_1, A \rangle + 2n - 6$, de sorte que, si l'on suppose $\langle c_1, A \rangle < 2$, on arrive à une catastrophe : pour m assez grand, cet espace, qui est de dimension au moins $4m + 2$, a une dimension virtuelle strictement inférieure à $4m + 2$. *Aucune* structure presque complexe ne peut être régulière pour la classe mA .

2.d.d. La compacité. — J'ai un peu triché en parlant du degré de l'application d'évaluation dans l'exemple des droites du plan. L'espace $\mathcal{M}$ que j'ai utilisé n'est, en effet, pas compact. C'est un fibré sur la grassmannienne, de fibre le complémentaire de la diagonale dans $\mathbf{P}^1(\mathbf{C}) \times \mathbf{P}^1(\mathbf{C})$: on a demandé que les deux points z_1 et z_2 soient distincts ! Le remède est très simple dans ce cas : on compactifie $\mathcal{M}$ brutalement en autorisant les deux points à coïncider. Le degré obtenu est bien 1 : l'image inverse $\text{év}^{-1}(p_1, p_2)$ ne rencontre pas la partie ajoutée pour compactifier.

Mais le problème peut être beaucoup plus sérieux. J'ai déjà évoqué le cas des coniques⁽¹⁰⁾ planes. Regardons le plus en détail. Chacun comprend que la suite de coniques lisses $xy = \alpha$ converge, quand α tend vers 0, vers la réunion des deux droites $xy = 0$ (figure 5). Explicitons le phénomène en termes d'applications. Soit, pour $\alpha \neq 0$,

⁽¹⁰⁾Exemple tout aussi inévitable que celui des droites passant par deux points *supra* et que celui de la cubique dégénéralant *infra*.

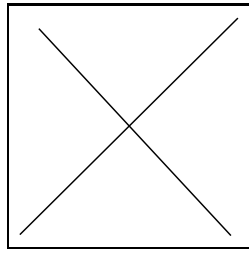
u_α l'application

$$\begin{aligned} u_\alpha : \mathbf{P}^1(\mathbf{C}) &\longmapsto \mathbf{P}^2(\mathbf{C}) \\ [x, y] &\longmapsto [x^2, \alpha y^2, xy] \end{aligned}$$

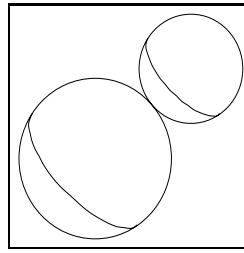
(en coordonnées homogènes). Le point $[0, 1]$ est envoyé sur $[0, 1, 0]$, point à l'infini de l'axe des y , et ceci pour tout $\alpha \neq 0$. Les autres points $[x, y]$ avec $x \neq 0$ sont envoyés sur les $[1, \alpha t^2, t]$. Quand α tend vers 0, ces points tendent vers ceux de l'axe des x .

Il n'y a *pas* d'application de $\mathbf{P}^1(\mathbf{C})$ dans $\mathbf{P}^2(\mathbf{C})$ qui paramètre la réunion de deux droites. Je reviendrai de façon plus précise sur ce phénomène au §3.c : c'est ce qu'on appelle la formation de « bulles ».

On voit que, pour compactifier raisonnablement l'espace des applications de degré 2 de $\mathbf{P}^1(\mathbf{C})$ dans $\mathbf{P}^2(\mathbf{C})$, il faut lui ajouter des applications définies sur la réunion de deux droites sécantes, ce que les topologues (réels) aiment dessiner comme sur la figure 6, justifiant ainsi la terminologie « bulle ». Plus généralement, des courbes,



(a) les droites de la géomètre



(b) et celles du topologue

FIGURE 6. Pourquoi « bulles » ? deux points de vue sur deux droites sécantes

disons rationnelles, d'une classe d'homologie A , peuvent très bien converger vers des courbes décomposées, quand la classe A elle-même se décompose en somme $A = B + C$ de classes contenant des J -courbes rationnelles.

Le problème des revêtements multiples (suite). — Aux problèmes de régularité soulevés par les revêtements multiples s'ajoutent des problèmes de compacité. Je viens d'évoquer le fait que, pour compactifier l'espace des courbes J -holomorphes d'une classe A décomposée en $A = B + C$, où les classes B et C contiennent des J -courbes, il faudrait rajouter à $\overline{\mathcal{M}}_{g,m}^J(W, A)$ des courbes des classes B et C .

Supposons que $A = B + mC$. La dimension virtuelle de l'espace des J -courbes rationnelles de la classe A est

$$2\langle c_1, B \rangle + 2m\langle c_1, C \rangle + 2n - 6.$$

Dans l'adhérence, il doit y avoir des courbes avec une composante de la classe B et une de la classe mC (et notamment des revêtements d'ordre m d'une courbe de la classe C). Donc, pour compactifier, il faut ajouter — entre autres choses — un espace

de dimension au moins égale à

$$[2\langle c_1, B \rangle + 2n - 6] + [2\langle c_1, C \rangle + 2n - 6 + 4m + 2].$$

En faisant croître m , on s'aperçoit que, si $\langle c_1, C \rangle < 2$, la dimension de cette composante du « bord » devient rapidement plus grande que la dimension de l'espace lui-même.

2.e. Oui, ce sont des invariants du type de déformation. — À titre de motivation pour la construction, je vais montrer, selon Ruan [66], un exemple dans lequel ces invariants permettent effectivement de distinguer deux formes symplectiques qui ont les mêmes invariants mous. Ruan considère deux variétés algébriques complexes de dimension complexe 3 construites par Gross [34] et dont ce dernier a démontré

- qu'elles sont difféomorphes comme variétés de dimension 6 et même comme variétés presque complexes
- mais qu'elles ne sont pas dans la même classe de déformation comme variétés algébriques.

Ruan distingue alors leurs classes de déformation comme variétés symplectiques grâce à un invariant de Gromov-Witten.

Les deux variétés sont construites comme hypersurfaces de variétés (toriques) de dimension 4

$$W_1 \subset \mathbf{P}^1(\mathbf{C}) \times \mathbf{P}^3(\mathbf{C}) = P_1 \text{ et } W_2 \subset \mathbf{P}(\mathcal{O}_{\mathbf{P}^1(\mathbf{C})}(-1) \oplus \mathcal{O}_{\mathbf{P}^1(\mathbf{C})}(2) \oplus \mathcal{O}_{\mathbf{P}^1(\mathbf{C})}(1)) = P_2,$$

de façon que leur première classe de Chern soit nulle⁽¹¹⁾ : on choisit une section assez générale s_i du fibré $\wedge^4 TP_i$ et on pose $W_i = s_i^{-1}(0)$. Les deux variétés W_1 et W_2 sont fibrées sur $\mathbf{P}^1(\mathbf{C})$ par construction. Il est facile de vérifier que les fibres sont des surfaces de degré 4 dans $\mathbf{P}^3$ (« surfaces $K3$ »).

Il est aussi très facile de vérifier que W_1 est simplement connexe et que sa cohomologie entière est sans torsion. Son H^2 est engendré par

- la classe f duale à la classe des fibres de la fibration
 - et la classe t qui provient de la classe d'Euler du fibré en droites $\mathcal{O}_{\mathbf{P}^1}(1) \rightarrow \mathbf{P}^1$.
- Notons (F, T) la base duale à (f, t) . La forme (cubique) d'intersection

$$H_2(W_1; \mathbf{Z}) \otimes H_2(W_1; \mathbf{Z}) \otimes H_2(W_1; \mathbf{Z}) \longrightarrow \mathbf{Z}$$

est alors la forme $2T^3 + 4T^2F$.

C'est un peu plus délicat, mais Ruan montre que W_2 a les mêmes propriétés : elle est simplement connexe, sa cohomologie est sans torsion, son H^2 est engendré par les classes f et t correspondantes et sa forme cubique est $2T^3 + 4T^2F$.

Un théorème de Wall [80] affirme que les variétés W_1 et W_2 sont difféomorphes comme variétés de dimension 6. Comme elles ont la même première classe de Chern, elles sont aussi difféomorphes comme variétés presque complexes (un théorème de

⁽¹¹⁾Les variétés W_1 et W_2 sont alors des « variétés de Calabi-Yau » de dimension 3.

Wu [83]). Pour distinguer leurs classes de déformations comme variétés algébriques, Gross utilise les cônes de Kähler⁽¹²⁾ (cône dans $H^2(W_i; \mathbf{R})$ engendré par les classes de cohomologie des formes de Kähler) qui sont

- le cône engendré par t et f pour W_1
- celui engendré par t et $t + f$ pour W_2 .

Pour distinguer leurs classes de déformation comme variétés symplectiques⁽¹³⁾ Ruan utilise, lui aussi, un argument basé sur cette différence entre les cônes de Kähler. Il calcule les invariants de Gromov-Witten pour la classe d'homologie $T - F$. Cette classe n'est représentée par aucune courbe holomorphe dans W_1 . En effet, pour $[\omega] = t + f$, on a

$$\langle [\omega], T - F \rangle = 0,$$

ce qui, en anticipant un peu, est contradictoire avec le lemme 3.6 : sur une courbe holomorphe, la forme symplectique est positive. Donc tous les invariants de Gromov-Witten associés à cette classe sont nuls. Il n'en est pas de même des invariants de W_2 pour cette même classe. On montre que, sur W_2 , il y a une (unique) courbe rationnelle dans la classe $T - F$. De plus, le fibré normal de cette courbe est $\mathcal{O}(-1) \oplus \mathcal{O}(-1)$, ce qui permet de montrer que la structure presque complexe J est régulière pour la classe $T - F$. L'espace de modules $\overline{\mathcal{M}}_{0,1}^J(W_2, T - F)$ est donc une variété de dimension 2 (un point sur une sphère) de sorte que, pour l'application d'évaluation

$$\begin{aligned} \overline{\mathcal{M}}_{0,1}^J(W_2, T - F) &\xrightarrow{\text{év}} W_2 \\ (u, z) &\longmapsto u(z) \end{aligned}$$

et pour $\xi \in H^2(W_2)$, on a

$$\Psi_{0,1}^{T-F}(\xi) = \langle \text{év}^* \xi, [\overline{\mathcal{M}}_{0,1}^J(W_2, T - F)] \rangle = \langle \xi, T - F \rangle$$

(nombre d'intersection de la classe duale à ξ avec la classe $T - F$, voir la proposition 4.2). Par exemple, pour $\xi = t$, on trouve $\Psi_{0,1}^{T-F}(t) = 1$, ce qui différencie les classes de déformation de W_1 et W_2 .

Remarque 2.5. — L'invariant utilisé ici pour distinguer les structures symplectiques sur $W = W_1 = W_2$ est donc une application

$$\begin{aligned} \mathcal{W} &\longrightarrow \text{Hom}(H^2(W; \mathbf{Z}), \mathbf{Z}) \\ \omega &\longmapsto (\xi \mapsto \Psi_{0,1}^{T-F}(\xi)). \end{aligned}$$

⁽¹²⁾et les résultats de [81]

⁽¹³⁾Remarquons qu'en identifiant W_1 et W_2 , on peut choisir des formes symplectiques dans la même classe de cohomologie, puisque les deux cônes de Kähler ont une intersection non triviale.

3. Les espaces de modules d'applications stables

On veut donc construire un espace de modules pour les applications J -holomorphes qui tienne compte des problèmes de dégénérescence et de compacité évoqués ci-dessus. La riche idée, c'est l'espace de modules d'applications stables de Kontsevich [44], à quoi nous amenaient les considérations précédentes. On remarquera qu'au cours de la partie précédente, on est passé insensiblement du « qu'est-ce qu'on compte ? » au « comment on le compte ? ».

Dans cette partie, je définis les espaces de modules d'applications stables de Kontsevich, avec quelques exemples simples. Je décris leur topologie, évoque leur compacité et m'interroge sur leur lissité.

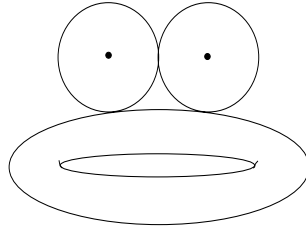


FIGURE 7

3.a. Espaces d'applications stables. — Soit W une variété symplectique munie d'une structure presque complexe J calibrée par sa forme symplectique.

On appelle *application stable* un paquet $(\Sigma, \vec{z}, u)$, dans lequel

- Σ est une courbe complexe de genre g . Elle est autorisée à avoir des points singuliers, mais seulement du type « points doubles ordinaires », c'est-à-dire des singularités représentées (au choix) comme sur la figure 6.

- la notation $\vec{z} = (z_1, \dots, z_m)$ désigne une suite ordonnée de m points distincts de Σ , qui ne doivent pas être des points singuliers de Σ et qu'on appelle des « points marqués »

- et $u : \Sigma \rightarrow W$ est une application J -holomorphe.

Les points singuliers et les points marqués sont regroupés sous le qualificatif commun de « spéciaux ». Le genre utilisé ici est le genre *arithmétique*, dont je rappelle une définition topologique. À la surface Σ , on associe un graphe Γ_Σ :

- les sommets correspondent aux composantes de Σ ,
- on relie deux sommets par une arête quand les composantes de Σ correspondantes se coupent.

Le genre *arithmétique* de Σ est, par définition, la somme des genres des composantes plus la dimension de $H_1(\Gamma_\Sigma)$. Par exemple, la figure 7 représente une courbe de genre arithmétique 2 (avec deux points marqués et trois points singuliers).

Il faut expliquer maintenant la notion de stabilité. Elle vient de la théorie de Deligne-Mumford des espaces de modules de courbes marquées stables.

Courbes marquées stables. — Une courbe marquée stable est un paquet $(\Sigma, \vec{z})$ comme ci-dessus, sauf qu'il n'y a pas d'application. La condition de stabilité sert à fabriquer l'espace $\overline{\mathcal{M}}_{g,m}$ des classes d'isomorphisme de courbes marquées. Un isomorphisme $(\Sigma, \vec{z}) \rightarrow (\Sigma', \vec{z}')$ est un isomorphisme $\phi : \Sigma \rightarrow \Sigma'$ qui satisfait à $\phi(z_i) = z'_i$ pour tout i . On dit que $(\Sigma, \vec{z})$ est *stable* si son groupe d'automorphisme est discret.

Exemples 3.1

- *En genre 0.* Les automorphismes de $\mathbf{P}^1(\mathbf{C})$ sont les homographies (éléments de $\mathrm{PGL}(2; \mathbf{C})$). Une homographie est déterminée par les images de trois points distincts. Les automorphismes de $\mathbf{P}^1(\mathbf{C})$ qui fixent deux points ou moins forment un sous-groupe non discret de $\mathrm{PGL}(2; \mathbf{C})$. Donc
 - la courbe marquée $(\mathbf{P}^1(\mathbf{C}), z_1, \dots, z_m)$ est stable si et seulement si $m \geq 3$,
 - l'espace $\overline{\mathcal{M}}_{0,3}$ est un point. On trouvera une description de $\overline{\mathcal{M}}_{0,4}$ au § 4.c.
- *En genre 1.* Une courbe lisse de genre 1 est, au choix près d'une origine, une courbe elliptique, c'est-à-dire un tore $\mathbf{C}/\Lambda$. En particulier, elle a un groupe d'automorphismes de dimension 1... et donc, si Σ est de genre 1, $(\Sigma, z_1, \dots, z_m)$ est stable si et seulement si $m \geq 1$.
- *En genre plus grand.* Les courbes de genre $g \geq 2$ ont un groupe d'automorphismes fini, ces courbes sont donc toujours stables.

Une façon courte d'exprimer la condition de stabilité est donc : $m + 2g \geq 3$.

Remarque 3.2. — J'utiliserai surtout $\overline{\mathcal{M}}_{0,m}$, la compactification de Deligne-Mumford-Knudsen de l'espace des classes d'isomorphisme de m points distincts ordonnés dans $\mathbf{P}^1(\mathbf{C})$. Voir [43, 42, 13, 36] pour des propriétés de ces espaces.

Applications stables. — Imposons à nos « applications » $(\Sigma, \vec{z}, u)$ une condition de stabilité analogue. Il faut commencer par préciser la notion d'isomorphisme. On dit que deux applications stables $(\Sigma, \vec{z}, u)$ et $(\Sigma', \vec{z}', u')$ sont *isomorphes* s'il existe un isomorphisme $\varphi : \Sigma \rightarrow \Sigma'$ tel que $\varphi(z_i) = z'_i$ pour tout i et que $u' \circ \varphi = u$.

On demande que les paquets n'aient pas d'automorphisme infinitésimal. C'est exactement demander que les composantes de Σ sur lesquelles u est constante soient des courbes *stables* au sens ci-dessus.

Plus concrètement, en pensant aux exemples 3.1,

- Si u est constante sur une composante de genre 0, cette composante doit porter au moins trois points spéciaux (singuliers ou marqués),
- de même, si u est constante sur une composante de genre 1, celle-ci doit porter un point spécial.

L'espace de modules des applications stables. — L'ensemble des classes d'isomorphisme des applications stables des courbes de genre g avec m points marqués représentant la classe A dans la variété presque complexe (W, J) sera noté $\overline{\mathcal{M}}_{g,m}^J(W, A)$.

Remarque 3.3. — Ce qu'on regarde ici n'est pas la courbe image dans W , mais ce n'est pas non plus tout à fait la courbe paramétrée puisqu'on quotiente par le groupe d'automorphismes.

Remarque 3.4. — On est obligé de considérer des courbes sources qui sont singulières, comme on l'a vu à propos des coniques au § 2.d. Les singularités de la courbe source sont des points doubles ordinaires, mais cela n'empêche pas la courbe image d'être éventuellement très singulière. On pensera aux exemples représentés sur la figure 8, les applications

- $t \mapsto (t^2, t^3)$, paramétrage d'une cubique cuspidale, la courbe source est une droite projective (lisse!),
- $t \mapsto (t^2, 0)$, paramétrage d'une conique dégénérée en une droite double qui n'est pas immersif (il n'est « injectif en aucun point »).

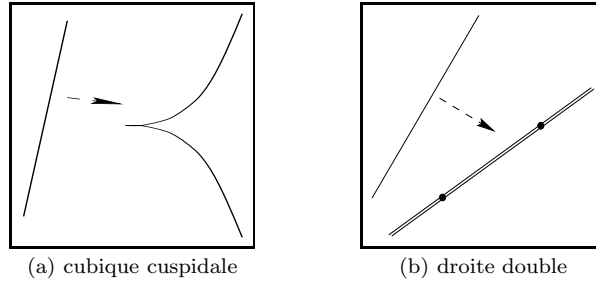


FIGURE 8. Applications singulières

Remarque 3.5. — La dimension attendue de $\overline{\mathcal{M}}_{g,m}^J(W, A)$ est $2\langle c_1, A \rangle + 2n(1 - g) + 2m - \dim G$ (où G est le groupe d'automorphismes d'une courbe lisse de genre g).

3.b. Exemples d'espaces d'applications stables. — Dans ce paragraphe, je donne quelques exemples simples, dans les cas où W est un point, une droite ou un plan projectif. Dans ces exemples, je ne mentionne pas explicitement la structure presque complexe dans la notation (écrivant, donc, $\overline{\mathcal{M}}$ pour $\overline{\mathcal{M}}^J$) puisque ces variétés ont une structure complexe préférable.

Le cas du point. — Quand W est un point, l'application u est constante et on retrouve l'espace $\overline{\mathcal{M}}_{g,m}$ des courbes marquées stables (rappelons que la condition de stabilité est $m + 2g \geq 3$).

Le cas de la classe nulle. — La classe d'homologie $A = 0$ ne peut être représentée que par des applications *constantes*. C'est une conséquence du lemme suivant.

Lemme 3.6. — *Soit ω une forme symplectique sur W et soit A une classe d'homologie de dimension 2. Si A peut être représentée par une application holomorphe pour une structure presque complexe calibrée par ω , alors on a $\langle \omega, A \rangle \geq 0$, avec égalité si et seulement si l'application holomorphe est constante.*

Démonstration. — Fixons la structure presque complexe J et considérons une application J -holomorphe $u : \Sigma \rightarrow W$. Sa différentielle est linéaire (sur $\mathbf{C}$). On a donc, pour tout $z \in \Sigma$ et tout $\alpha \in T_z \Sigma$,

$$(u^* \omega)_z(\alpha, i\alpha) = \omega_{u(z)}(T_z u(\alpha), J T_z u(\alpha)).$$

Comme ω calibre J , $X \mapsto \omega(X, JX)$ est une métrique riemannienne, donc

$$(u^* \omega)_z(\alpha, i\alpha) \geq 0.$$

Ainsi

$$\langle \omega, A \rangle = \langle \omega, u_*[\Sigma] \rangle = \langle u^* \omega, [\Sigma] \rangle \geq 0.$$

De plus, si u n'est pas constante, on peut trouver un ouvert $U \subset \Sigma$ et un champ de vecteurs non nul α sur U tels que $(u^* \omega)_z(\alpha, i\alpha) > 0$ pour $z \in U$. Donc $\langle \omega, A \rangle > 0$. $\square$

Si $A = 0$, on voit que l'application u est constante et donc l'espace de modules $\overline{\mathcal{M}}_{g,m}^J(W, 0)$ s'identifie à $\overline{\mathcal{M}}_{g,m} \times W$.

Applications de degré 1. — Un autre exemple très simple est celui de l'espace de modules $\overline{\mathcal{M}}_{0,0}(\mathbf{P}^1(\mathbf{C}), L)$. C'est l'espace des applications de degré 1

$$u : \mathbf{P}^1(\mathbf{C}) \longrightarrow \mathbf{P}^1(\mathbf{C}).$$

Aux automorphismes de $\mathbf{P}^1$ près, il n'y a qu'une seule telle application, l'application identique, de sorte que l'espace $\overline{\mathcal{M}}_{0,0}(\mathbf{P}^1(\mathbf{C}), L)$ est un point.

Remarque 3.7. — On peut considérer plus généralement la classe dL , toujours dans l'homologie de la droite projective complexe $\mathbf{P}^1(\mathbf{C})$. Cette fois, l'espace de modules $\overline{\mathcal{M}}_{0,0}(\mathbf{P}^1(\mathbf{C}), dL)$ est une compactification de l'espace de Hurwitz des polynômes de degré d . On peut aussi remplacer $\mathbf{P}^1(\mathbf{C})$ par $\mathbf{P}^n(\mathbf{C})$: il est clair que $\overline{\mathcal{M}}_{0,0}(\mathbf{P}^n(\mathbf{C}), L)$ est la grassmannienne $G(2, \mathbf{C}^{n+1})$ des droites de $\mathbf{P}^n(\mathbf{C})$ (comme on l'a déjà remarqué dans le cas $n = 2$ au § 2.c).

3.b.a. Coniques planes. — Considérons maintenant l'espace $\overline{\mathcal{M}}_{0,0}(\mathbf{P}^2(\mathbf{C}), 2L)$ des applications stables de degré 2 dans $\mathbf{P}^2(\mathbf{C})$. Il est constitué de quatre parties :

(1) L'espace des coniques lisses. Une conique propre est paramétrée par une application de degré 2

$$u : \mathbf{P}^1(\mathbf{C}) \longrightarrow \mathbf{P}^2(\mathbf{C}).$$

C'est un espace de dimension complexe 5 (trois polynômes homogènes de degré 2 à multiplication par un même scalaire près, modulo les automorphismes de $\mathbf{P}^1(\mathbf{C})$: $9 - 1 - 3 = 5$).

(2) L'espace des applications de degré 2 de $\mathbf{P}^1(\mathbf{C})$ dans $\mathbf{P}^2(\mathbf{C})$ dont l'image est une droite. Sa dimension complexe est 4 (la droite et les deux points de ramification de l'application : $2 + 2 = 4$).

(3) L'espace des applications $C \rightarrow \mathbf{P}^2(\mathbf{C})$, où C désigne la réunion de deux droites sécantes et l'image de u est aussi constituée de deux droites sécantes. La dimension complexe est encore 4 (deux droites dans $\mathbf{P}^2(\mathbf{C})$: $2 \times 2 = 4$).

(4) L'espace des applications $C \rightarrow \mathbf{P}^2(\mathbf{C})$, où C est (encore) la réunion de deux droites sécantes mais où cette fois l'image est formée d'une droite double. La dimension complexe de cet espace est 3 (une droite dans $\mathbf{P}^2(\mathbf{C})$ et un point sur cette droite : $2 + 1 = 3$).

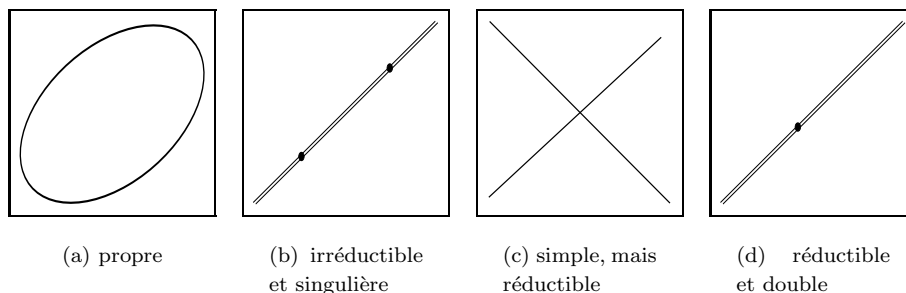


FIGURE 9. Coniques

3.b.b. Courbes rationnelles planes de degré d . — Une description des espaces de courbes de grand degré dans le plan serait assez compliquée (voir le cas des coniques ci-dessus). Mentionnons quand même le cas célèbre des courbes *rationnelles* de degré d .

Une courbe lisse de degré d n'est jamais rationnelle quand $d \geq 3$. Son genre est $(d-1)(d-2)/2$. Mais il y a quand même des courbes rationnelles de tout degré. Par exemple, une cubique plane singulière est toujours rationnelle (les cubiques lisses sont des courbes de genre 1, en accord avec la formule du genre), comme l'est la cubique cuspidale de la figure 8 : on abaisse le genre en ajoutant des singularités.

Naïvement et selon un décompte déjà utilisé plusieurs fois dans cet article, pour donner une application de degré d de $\mathbf{P}^1(\mathbf{C})$ dans $\mathbf{P}^2(\mathbf{C})$, il faut donner trois polynômes de degré d à un coefficient près, on marque m points et on quotiente par

$\mathrm{PGL}(2; \mathbf{C})$, le résultat de l'opération est une dimension complexe égale à $3d+2+m-3$. C'est d'ailleurs la dimension attendue, comme on le vérifiera sans mal⁽¹⁴⁾.

L'espace de modules et le but $(\mathbf{P}^2(\mathbf{C}))^m$ de l'évaluation ont même dimension si et seulement si $m = 3d - 1$. Le degré de l'application d'évaluation

$$\overline{\mathcal{M}}_{0,3d-1}(\mathbf{P}^2(\mathbf{C}), dL) \longrightarrow (\mathbf{P}^2(\mathbf{C}))^{3d-1}$$

est le nombre N_d de courbes rationnelles de degré d passant par $3d - 1$ points du plan (une droite par deux points, $N_1 = 1$, une conique par cinq points, $N_2 = 1 \dots$).

Ces nombres sont calculés par récurrence grâce à une formule, due à Kontsevich [44] (ici le théorème 4.8), qui est un des plus anciens et un des plus beaux succès de la théorie.

3.c. Topologie et compacité. — Il est dans la nature de l'espace $\overline{\mathcal{M}}_{g,m}^J(W, A)$ d'être relié à la fois à la variété W et à l'espace des courbes stables $\overline{\mathcal{M}}_{g,m}$ par, respectivement,

- l'évaluation des applications aux points marqués (à valeurs dans W^m),
- l'oubli de l'application.

Pour être utilisables, c'est bien le moins que ces applications soient continues : on se souvient sans doute que des calculs cohomologiques sont en vue (depuis le § 2.c). Pour l'instant, ce que nous avons défini n'est qu'un ensemble.

On se doute que cet ensemble peut être muni d'une structure plus riche. Sous certaines hypothèses, c'est une variété projective complexe. Même si la structure J n'est pas intégrable, c'est, toujours, un espace topologique : on peut le munir d'une variante de la topologie $\mathcal{C}^k$ (pour tout k).

Cette topologie est un peu compliquée, mais le jeu en vaut la chandelle :

– C'est pour cette topologie que s'exprime la propriété dite « de compacité » (voir [32] et [61]), ici le théorème 3.15.

– La définition de la topologie confirme et explicite la remarque faite ci-dessus qu'avec les classes d'isomorphisme d'applications stables, on est quelque part entre les applications et leurs images.

Topologie de $\overline{\mathcal{M}}_{g,m}^J(W, A)$. — Voici une description d'une base de voisinages de la classe de l'application stable $(\Sigma, \bar{z}, u)$. Fixons un nombre réel $\varepsilon > 0$, un voisinage U des points singuliers de Σ , un voisinage V de l'ensemble des points marqués et une métrique μ sur la partie lisse de Σ . Le voisinage que ces données définissent est formé des (classes d') applications stables $(\Sigma', \bar{z}', u')$ telles que :

- Il existe une application continue $\sigma : \Sigma' \rightarrow \Sigma$

⁽¹⁴⁾Voir le § 3.d.a pour une justification.

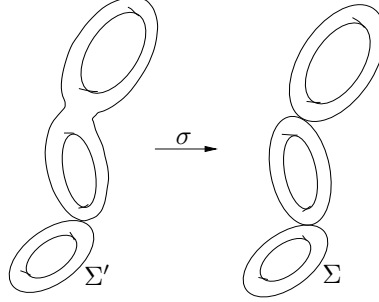


FIGURE 10

– telle que l'image inverse de chaque point double soit, soit un point double de Σ' , soit un anneau de module⁽¹⁵⁾ $\geq 1/\varepsilon$, ne contenant aucun point de $\tilde{z}'$ (voir la figure 10),

– et qui soit un difféomorphisme de classe $\mathcal{C}^k$ excepté au-dessus des points singuliers.

– Appelons J_Σ la structure complexe de Σ , considérée comme un automorphisme de $T\Sigma$ (la multiplication par i) et de même $J_{\Sigma'}$ celle de Σ' . La structure complexe $J_{\Sigma'}$ définit une structure complexe $\sigma_* J_{\Sigma'}$ sur le complémentaire de l'ensemble des points singuliers de Σ . On demande que $\|J_\Sigma - \sigma_* J_{\Sigma'}\|_{\mathcal{C}^k} < \varepsilon$.

– On demande bien sûr aussi que les points marqués de Σ' soient proches de ceux de Σ : $\sigma(z'_i) \in V$.

– On demande que u et u' soient proches l'une de l'autre : $\|u - u' \circ \sigma^{-1}\|_{\mathcal{C}^k} < \varepsilon$ en dehors de U .

– On demande, enfin, que $|\text{aire}(u) - \text{aire}(u')| < \varepsilon$.

Les normes utilisées pour exprimer ces conditions sont calculées grâce à la métrique μ fixée sur Σ et à la métrique définie par ω et J sur W .

Remarque 3.8. — Le cas où W est réduite à un point est celui de l'espace de modules des courbes (l'espace $\overline{\mathcal{M}}_{g,m}$). On a ainsi défini une topologie sur cet espace. Par exemple, la condition sur J_Σ et $J_{\Sigma'}$ exprime le fait que Σ et Σ' sont proches l'une de l'autre dans l'espace de modules $\overline{\mathcal{M}}_{g,0}$.

Exemples 3.9

(1) Après les droites et les coniques (§3.b), l'exemple le plus classique et le plus instructif est celui d'une famille de cubiques planes lisses convergeant vers une cubique rationnelle. On considère par exemple la cubique C_1 d'équation $y^2 = 4x^3 - x - 1$ et l'application $u_\alpha : C_1 \rightarrow \mathbf{P}^2(\mathbf{C})$ induite par $(x, y) \mapsto (\alpha^2 x, \alpha^3 y)$. L'image est la cubique

⁽¹⁵⁾Rappelons qu'un anneau complexe est conformément équivalent à $S^1 \times]0, L[$ pour un unique $L > 0$, son *module* : du point de vue conforme, être très long, c'est être très mince.

d'équation $y^2 = 4x^3 - \alpha^4 x - \alpha^6$, qui devient singulière et rationnelle pour $\alpha = 0$. Le phénomène est décrit, du point de vue des applications stables, sur la figure 11.

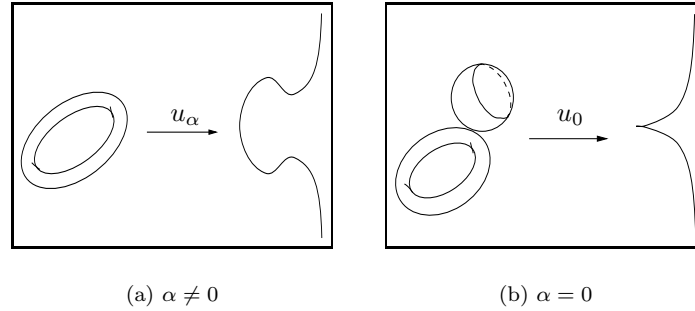


FIGURE 11

La courbe limite est constituée d'une composante de genre 1 avec une bulle. L'application stable limite

- est constante sur la composante de genre 1 (et l'envoie sur le point singulier)
- et paramètre la cubique cuspidale par la composante de genre 0.

Ce qu'on voit dans l'image, c'est la bulle.

(2) Voici un exemple que j'ai appris dans [25] et qui confirme le phénomène observé ci-dessus... et qui confirme aussi que la convergence peut être un peu compliquée.

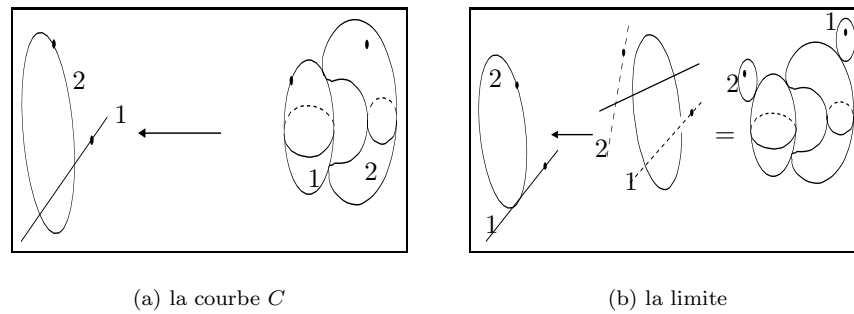


FIGURE 12

L'exemple image est très simple : il s'agit d'une cubique plane dégénérée C constituée d'une droite et d'une conique sécantes (figure 12). Il y a donc deux composantes rationnelles se coupant en deux points. Le genre arithmétique est 1. Il y a deux points doubles et on met un point marqué sur chacune des composantes.

On veut maintenant représenter par des applications stables, dans l'espace $\overline{\mathcal{M}}_{1,2}(\mathbf{P}^2(\mathbf{C}), 3L)$, le passage à une situation limite où la droite devient tangente à la conique.

On considère donc une famille $u_\alpha : C \rightarrow \mathbf{P}^2(\mathbf{C})$ où u_1 est l'inclusion et, où, quand α tend vers 0, la droite se rapproche d'une tangente à la conique.

Comme dans l'exemple ci-dessus, la courbe source est fixée : la figure est unique à automorphisme près puisque chaque composante est un $\mathbf{P}^1(\mathbf{C})$ avec trois points spéciaux. À la limite, à la source, il se crée deux bulles (les droites en pointillés sur la figure 12) qui vont paramétrer la courbe image limite, les autres composantes étant écrasées sur le point singulier.

(3) Dans l'exemple suivant, que m'a donné Jean-Claude Sikorav, on voit que les bulles peuvent très bien avoir elles-mêmes des points doubles.

Soit Σ_1 la cubique plane d'équation homogène $y^3 = x^3 - xz^2$ et soit $u_\alpha : \Sigma_1 \rightarrow \mathbf{P}^2$ l'application $[x, y, z] \mapsto [x, \alpha y, \alpha^2 z]$, dont l'image est la courbe Σ_α d'équation $\alpha y^3 = \alpha^4 x^3 - xz^2$. Faisons tendre α vers 0. La courbe limite $xz^2 = 0$ est formée

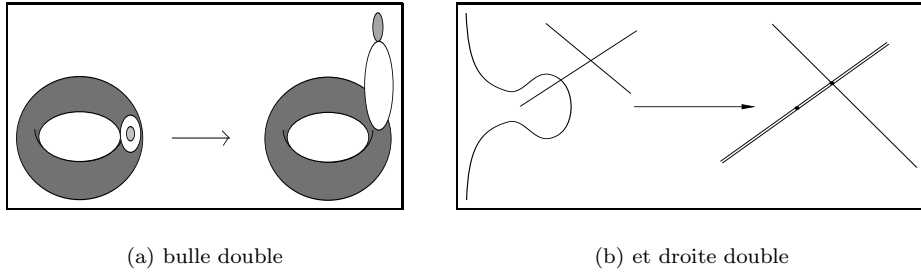


FIGURE 13

de deux droites dont l'une est double. L'application u_α converge vers l'application constante $u([x, y, z]) = [1, 0, 0]$ en dehors du point d'inflexion $[0, 0, 1]$. Donc une courbe rationnelle se forme, au point $[0, 0, 1]$, mais elle est réductible et formée de deux composantes dont l'une est un revêtement double. Une des composantes rationnelles est envoyée sur la droite double par une application de degré 2, avec donc deux points de ramification (comme sur la figure 9). La composante de genre 1 est écrasée sur un des points de ramification, l'autre composante rationnelle paramètre une droite passant par le deuxième point de ramification.

Remarque 3.10. — On trouvera de nombreux exemples de familles de cubiques planes convergeant vers des cubiques singulières dans [52]. Un bon exercice pourrait être de décrire ces exemples en termes de convergence dans un espace de modules d'applications stables.

Évaluation et contraction. — Décrivons maintenant les deux applications naturelles :
 – d'évaluation

$$\begin{aligned} \text{év}_m : \overline{\mathcal{M}}_{g,m}^J(W, A) &\longrightarrow W^m \\ (\Sigma, \vec{z}, u) &\longmapsto (u(z_1), \dots, u(z_m)) \end{aligned}$$

(pour $m \geq 1$),
 – et de contraction

$$\overline{\mathcal{M}}_{g,m}^J(W, A) \longrightarrow \overline{\mathcal{M}}_{g,m}$$

(pour $m + 2g \geq 3$).

La définition de l'évaluation est claire. La contraction est essentiellement l'application qui « oublie » u . Il y a toutefois là une petite difficulté technique que je vais expliquer. Si on essaie d'envoyer la classe d'isomorphisme de l'application stable $(\Sigma, \vec{z}, u)$ sur celle de la courbe $(\Sigma, \vec{z})$, on peut être gêné si $(\Sigma, \vec{z})$ a des composantes instables : celles-ci étaient autorisées dès que u n'y était pas constante. La solution est simple : on contracte ces composantes instables en des points, comme le montre la figure 14. C'est ce qui justifie le nom de « contraction » que porte cette application. On remarquera (toujours sur la figure 14) que cette contraction peut créer des points singuliers sur des composantes qui n'en avaient pas. La topologie que nous avons définie est assez

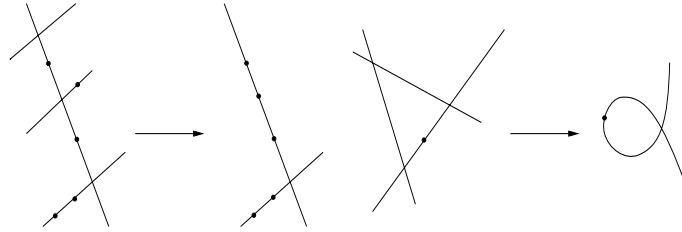


FIGURE 14. Contractions

bonne pour qu'on ait :

Proposition 3.11. — *Les deux applications d'évaluation aux points marqués*

$$\text{év}_m : \overline{\mathcal{M}}_{g,m}^J(W, A) \longrightarrow W^m$$

et de contraction

$$\overline{\mathcal{M}}_{g,m}^J(W, A) \longrightarrow \overline{\mathcal{M}}_{g,m}$$

sont continues.

□

Compacité. — Voici la propriété essentielle :

Proposition 3.12. — *L'espace $\overline{\mathcal{M}}_{g,m}^J(W, A)$ est compact.*

Idée de la démonstration. — Comme l'application de contraction est continue, le résultat est une conséquence de la compacité de la base $\overline{\mathcal{M}}_{g,m}$ (voir [43]) et de celle des fibres. Cette dernière est une conséquence du théorème de compacité de Gromov, que je vais rappeler maintenant. $\square$

Je donne deux versions du théorème de compacité de Gromov, pour lesquelles je renvoie à [32] et à [61]. D'autres démonstrations du théorème de Gromov se trouvent dans [84] ou [62]. À toutes celles et ceux qui douteraient que les énoncés proposés ici assurent la compacité d'un espace d'applications stables, je conseille la lecture de [25]. La relation entre la notion de stabilité et la séparation de l'espace de modules $\mathcal{Y}$ est notamment examinée.

Théorème 3.13. — *Soit W une variété riemannienne compacte munie d'une structure presque complexe J . Soit $u_n : \Sigma \rightarrow W$ une suite de courbes J -holomorphes d'aires bornées. Il existe une sous-suite convergeant vers une courbe J -holomorphe $u : \Sigma' \rightarrow W$ où Σ' est obtenue à partir de Σ en contractant quelques courbes fermées simples.*

Remarques 3.14

(1) Les structures presque complexes utilisées pourraient aussi dépendre de n (voir [32, 61]). Même avec cette modification, le résultat énoncé n'est qu'un cas particulier du théorème de Gromov, parce que j'ai fixé la courbe Σ .

(2) Le genre arithmétique de la courbe Σ' est égal au genre de Σ .

(3) Pour n assez grand, la classe d'homologie représentée par les u_n d'une sous-suite convergeant vers u est constante égale à celle représentée par u .

(4) Dans le cas qui nous intéresse, où W est une variété symplectique, J une structure presque complexe calibrée par la forme symplectique ω et où la métrique riemannienne est $\omega(\cdot, J\cdot)$, la condition d'aire bornée est automatique si les u_n représentent la même classe d'homologie A : cette aire est

$$\int_{\Sigma} u^* \omega = \langle u^* \omega, [\Sigma] \rangle = \langle \omega, u_* [\Sigma] \rangle = \langle \omega, A \rangle.$$

On utilise ici de façon décisive le fait que ω est *fermée*.

(5) Les bulles peuvent provenir de disques ou d'anneaux. Par exemple, la figure 13 montre deux bulles dont l'une provient d'un disque et l'autre d'un anneau.

La version « graphes » de ce théorème affirme qu'une limite de graphes est encore un graphe, à quelques bulles verticales près (voir la figure 15). La voici.

Théorème 3.15. — *Soient W une variété riemannienne compacte munie d'une structure presque complexe J et Σ une courbe complexe. Soit $u_n : \Sigma \rightarrow W$ une suite d'applications holomorphes d'aires bornées. Il existe une partie finie $\Gamma \subset \Sigma$ telle que :*

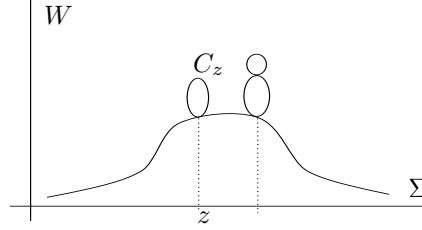


FIGURE 15

- une sous-suite de la suite des graphes $\tilde{u}_n : \Sigma \rightarrow \Sigma \times W$ converge vers le graphe $\tilde{u}$ d'une application J -holomorphe $u : \Sigma \rightarrow W$ en dehors de Γ .
- Si, pour $z \in \Gamma$, $\tilde{u}_n(z)$ ne converge pas vers $\tilde{u}(z)$, alors il y a une courbe rationnelle non triviale

$$\varphi_z : C_z \longrightarrow \{z\} \times W$$

passant par $u(z)$.

- Pour n assez grand, on a $u_n^*[\Sigma] = u_*[\Sigma] + \sum_{z \in \Gamma} (\varphi_z)_*[C_z] \in H_2(W; \mathbf{Z})$.

□

Remarque 3.16. — Supposons que W soit symplectique et que les courbes u_n représentent la même classe A . La conclusion homologique dit, notamment, que la classe A est décomposée en classe de la courbe limite plus les classes représentées par les bulles. Quand A ne peut pas se décomposer en somme $A = A_1 + A_2$ avec $\langle \omega, A_1 \rangle > 0$ et $\langle \omega, A_2 \rangle > 0$ (la classe A est alors dite « simple »), on a un vrai théorème de compacité. C'est ce qui se passe par exemple avec la classe des fibres de la projection $\mathbf{P}^1(\mathbf{C}) \times V \rightarrow V$ (dans le théorème 2.1) avec celle des droites dans le plan projectif... et qui ne se passe pas avec celle des coniques.

Le « bord » peut être trop gros. — Une autre raison qui complique la situation est le fait que l'espace de modules peut être considéré comme trop gros *aussi* parce qu'il est plus gros que la simple adhérence de l'espace des applications holomorphes stables $\Sigma \rightarrow W$ dont la source Σ est irréductible.

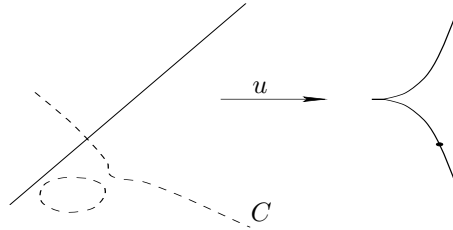


FIGURE 16

Le cas le plus simple où le problème se pose est celui des cubiques planes de genre 1. L'espace $\overline{\mathcal{M}}_{1,0}(\mathbf{P}^2(\mathbf{C}), 3L)$ est de dimension complexe 8. Il contient toutes les applications $u : C \cup \mathbf{P}^1(\mathbf{C}) \rightarrow \mathbf{P}^2(\mathbf{C})$, où C est une courbe de genre 1 coupant $\mathbf{P}^1(\mathbf{C})$ en un point, u est constante sur C et de degré 3 sur $\mathbf{P}^1(\mathbf{C})$. En d'autres termes, en utilisant les points marqués supplémentaires pour assembler les courbes, il contient un exemplaire de $\overline{\mathcal{M}}_{1,1} \times \overline{\mathcal{M}}_{0,1}(\mathbf{P}^2(\mathbf{C}), 3L)$... espace, lui, de dimension complexe 10.

Or une application $u : C \cup \mathbf{P}^1(\mathbf{C}) \rightarrow \mathbf{P}^2(\mathbf{C})$ n'est dans l'adhérence de l'ensemble des applications $\Sigma \rightarrow \mathbf{P}^2(\mathbf{C})$ (où Σ est irréductible) que si u a une singularité au point où les deux composantes se coupent : on l'a déjà dit, ajouter des singularités fait baisser le genre. La figure 16 montre un contre-exemple : la composante rationnelle $\mathbf{P}^1(\mathbf{C})$ est envoyée sur une cubique cuspidale par une application de degré 3, celle-ci envoie le point double sur un point régulier de la cubique et écrase la composante de genre 1.

Pour les questions énumératives, c'est un problème sérieux : on pourrait bien, en fin de compte (!) ne pas avoir compté ce que l'on pensait.

3.d. Lissité. — On a vu (exemple 2.4) qu'on ne peut pas espérer que $\overline{\mathcal{M}}_{g,m}^J(W, A)$ soit, en général, une variété (même en acceptant les orbifolds) de la dimension attendue. C'est quand même parfois le cas, comme je vais l'expliquer ci-dessous. Au § 3.e, j'expliquerai quelques stratégies possibles pour définir, en toute généralité, une classe fondamentale (virtuelle!) pour ces espaces.

Un énoncé de régularité. — À cause de l'exemple des revêtements multiples évoqué à deux reprises au § 2.d, il faut faire une hypothèse qui permette d'éviter le problème. On peut se restreindre, suivant [54], aux courbes *simples*, c'est-à-dire qui sont « injectives quelque part ». L'exemple le plus classique d'énoncé de régularité est la proposition 3.17 ci-dessous (voir [32, 57]).

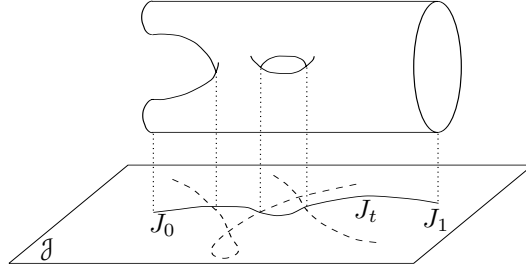


FIGURE 17

Proposition 3.17. — Soit (W, ω) une variété symplectique compacte de dimension $2n$ et soit $c_1 \in H^2(W; \mathbf{Z})$ sa première classe de Chern. Il existe une partie $\mathcal{J}_{\text{rég}}(A)$ dans l'espace $\mathcal{J}$ des structures presque complexes calibrées par ω , qui contient une intersection dénombrable d'ouverts denses et telle que

(1) L'espace $\mathcal{M}_g^A(J)$ des courbes J -holomorphes simples de la classe A est, pour $J \in \mathcal{J}_{\text{rég}}(A)$, une variété (lisse) orientée de dimension $2\langle c_1, A \rangle + 2n(1 - g)$.

(2) Si J_0 et $J_1 \in \mathcal{J}_{\text{rég}}(A)$, il existe une partie dense $\mathcal{J}_{\text{rég}}(J_0, J_1)$ dans l'espace des homotopies de J_0 à J_1 telle que, pour toute homotopie $(J_t)_{t \in [0,1]}$ dans cette partie, l'espace

$$\mathcal{M}_g^A((J_t)_{t \in [0,1]}) = \{(t, u) \mid u \in \mathcal{M}_g^A(J_t)\}$$

soit une variété orientée dont le bord orienté est $\mathcal{M}_g^A(J_1) - \mathcal{M}_g^A(J_0)$.

Cette proposition est illustrée sur la figure 17, où l'ouvert dense $\mathcal{J}_{\text{rég}}(A)$ est le complémentaire des hypersurfaces représentées en pointillés. Elle se démontre par les méthodes Fredholm évoquées au § 2.d.

3.d.a. Régularité de structures complexes usuelles. — Compter des courbes dans le plan est une activité humaine classique. Dans ce cas, j'ai déjà signalé qu'il était important de savoir si on peut

- effectivement compter ces courbes en calculant des invariants de Gromov-Witten
- utiliser les espaces de modules d'applications holomorphes pour la structure complexe usuelle pour calculer ces invariants.

La question est donc de savoir si la structure complexe usuelle du plan $\mathbf{P}^2(\mathbf{C})$ (et plus généralement d'une variété projective) est régulière pour telle ou telle classe.

Dans ce paragraphe, je vais mettre en évidence des conditions, sur une variété projective complexe W , qui garantissent que sa structure complexe est régulière pour les classes des courbes rationnelles et même que l'espace de modules de Kontsevich est, lui-même, une variété projective, peut-être pas lisse mais avec des singularités assez douces.

Le $\bar{\partial}$. — Commençons par quelques propriétés des structures complexes intégrables, inspirées par le § 3.5 de [57]. On a déjà signalé au § 2.d que, pour une structure complexe intégrable J , l'opérateur linéarisé D_u^J n'était autre que le $\bar{\partial}$ de Dolbeault. En particulier, D_u^J est surjective si et seulement si on a

$$\text{Im}(\bar{\partial} : \mathcal{C}^\infty(u^*TW) \longrightarrow \Omega^{0,1}(u^*TW)) = \Omega^{0,1}(u^*TW).$$

Comme Σ est une courbe, c'est équivalent à demander que $H_{\bar{\partial}}^{0,1}(u^*TW) = 0$.

Supposons maintenant que Σ soit rationnelle ($\Sigma = \mathbf{P}^1(\mathbf{C})$). Alors, grâce au théorème de Birkhoff⁽¹⁶⁾, nous savons que le fibré u^*TW , comme tous les fibrés sur $\mathbf{P}^1(\mathbf{C})$, se décompose en une somme de fibrés en droites. Pour tout fibré en droites $L \rightarrow \mathbf{P}^1(\mathbf{C})$, on a

$$H_{\bar{\partial}}^{0,1}(L) = \left(H_{\bar{\partial}}^{1,0}(L^*)\right)^* = \left(H^0(L^*(-2))\right)^*,$$

(un cas simple de la dualité de Serre, voir [31]) les différentes étoiles désignant les duals et le « -2 » venant du fibré cotangent ou canonique $\mathcal{O}(-2)$ de $\mathbf{P}^1(\mathbf{C})$. Ainsi

⁽¹⁶⁾Voir par exemple [63].

$\bar{\partial}$ est surjective si et seulement si pour chaque fibré en droites L apparaissant dans u^*TW , le fibré $L^*(-2)$ n'a pas de section holomorphe, c'est-à-dire si $c_1(L^*(-2)) < 0$, ou encore si $c_1(L) > -2$:

Proposition 3.18. — Soit $u : \mathbf{P}^1(\mathbf{C}) \rightarrow W$ une courbe J -holomorphe avec J intégrable. Pour que D_u^J soit surjective, il faut et il suffit que tous les fibrés en droites apparaissant dans la décomposition de Birkhoff de u^*TW aient une classe de Chern $c_1 \geq -1$.

Exemple 3.19. — Considérons l'espace projectif $\mathbf{P}^n(\mathbf{C})$ avec sa structure complexe naturelle J et la classe L des droites projectives. Soit $u : \mathbf{P}^1(\mathbf{C}) \rightarrow \mathbf{P}^n(\mathbf{C})$ une courbe holomorphe de la classe L (c'est à dire une droite). Alors

$$u^*T\mathbf{P}^n(\mathbf{C}) = \mathcal{O}(2) \oplus \mathcal{O}(1) \oplus \cdots \oplus \mathcal{O}(1),$$

D_u^J est surjective (ceci pour tout u) et J est régulière pour L .

Corollaire 3.20. — Supposons J intégrable et le fibré tangent TW engendré par ses sections holomorphes. Alors J est régulière pour toutes les classes contenant des courbes rationnelles.

En dimension 4, la situation est particulièrement simple.

Proposition 3.21. — Soit J une structure complexe (intégrable) sur une variété W de dimension 4. Soit $u : \mathbf{P}^1(\mathbf{C}) \rightarrow W$ une sphère J -holomorphe plongée. Alors $D_u = \bar{\partial}$ est surjective si et seulement si

$$\langle c_1(u^*TW), [\mathbf{P}^1(\mathbf{C})] \rangle \geq 1.$$

Démonstration. — On décompose le fibré $\mathcal{C}^\infty$ en $u^*TW = T\mathbf{P}^1(\mathbf{C}) \oplus N$. Soit $A = u_*[\mathbf{P}^1(\mathbf{C})]$. L'application u étant un plongement, on a

$$A \cdot A = \langle c_1(N), [\mathbf{P}^1(\mathbf{C})] \rangle \text{ et } \langle c_1(u^*TW), [\mathbf{P}^1(\mathbf{C})] \rangle = 2 + A \cdot A.$$

Considérons maintenant la décomposition analytique de u^*TW donnée par le théorème de Birkhoff

$$u^*TW = \mathcal{O}(m_1) \oplus \mathcal{O}(m_2).$$

On a $m_1 + m_2 = 2 + A \cdot A$. De plus, u^*TW contient $T\mathbf{P}^1(\mathbf{C})$ comme sous-fibré analytique. On a donc :

- soit que $T\mathbf{P}^1(\mathbf{C})$ est l'un des deux facteurs, disons $\mathcal{O}(m_1)$, alors $m_1 = 2$ et $m_2 = A \cdot A$, donc chaque m_i est supérieur ou égal à -1 ,
- soit qu'il s'envoie non trivialement dans chacun des deux, alors chacun des m_i est au moins égal à 2.

□

La formule d'adjonction (voir [31] pour le cas intégrable et [56] pour le cas d'une structure presque complexe) affirme que, si une classe $A \in H_2(W; \mathbf{Z})$ se représente par une courbe J -holomorphe plongée $u : C \rightarrow W$, alors toute application J -holomorphe $u' : C \rightarrow W$ représentant A est aussi un plongement. On déduit de la proposition 3.21 un critère simple et efficace :

Proposition 3.22. — *Soit J une structure complexe (intégrable) sur une variété W de dimension 4. Soit $A \in H_2(W; \mathbf{Z})$ une classe qui contient une sphère J -holomorphe plongée. Alors J est régulière pour A si et seulement si $A \cdot A \geq -1$.*

Exemple 3.23. — Il y a des contre-exemples très simples. En voici une famille. Il s'agit d'une des variétés différentielles $\mathbf{P}^1(\mathbf{C}) \times \mathbf{P}^1(\mathbf{C})$ ou $\tilde{\mathbf{P}}^2(\mathbf{C})$, vue comme espace total du fibré $\mathbf{P}(\mathcal{O}(-k) \oplus \mathbf{1}) \rightarrow \mathbf{P}^1(\mathbf{C})$, $k \geq 0$ (une « surface de Hirzebruch »)⁽¹⁷⁾. Elle possède une sphère holomorphe plongée d'auto-intersection $-k$ (la section nulle du fibré $\mathcal{O}(-k)$). Pour $k \neq 0, 1$, la structure complexe n'est pas régulière pour cette classe.

Variétés projectives convexes et lissité. — On va maintenant s'intéresser au « bord » de l'espace de modules $\overline{\mathcal{M}}_{0,0}(W, A)$, c'est-à-dire aux décompositions $A = A_1 + \cdots + A_k$ de la classe A en somme de classes représentables par des courbes rationnelles. On veut que les différentes strates de l'espace $\overline{\mathcal{M}}_{0,0}(W, A)$ soient lisses, c'est-à-dire que la structure J soit régulière pour toutes les classes A_i . On veut aussi éviter les problèmes dûs aux revêtements multiples évoqués au § 2.d et donc forcer une propriété de positivité des $\langle c_1, A_i \rangle$.

Définition 3.24. — La variété W est convexe pour la classe A , ou A -convexe si, pour toute décomposition

$$A = A_1 + \cdots + A_k$$

dans laquelle chaque A_j se représente par une courbe rationnelle, et pour toute application $u : \mathbf{P}^1(\mathbf{C}) \rightarrow W$ telle que $u_*[\mathbf{P}^1(\mathbf{C})] = A_j$ pour un certain j , on a

$$u^*TW = \bigoplus_{i=1}^n \mathcal{O}(m_i)$$

avec $m_i \geq 0$ pour tout i .

Il est clair que cette condition assure la régularité de la structure complexe pour chacune des classes A_j : on a vu dans la proposition 3.18 qu'il suffisait que tous les m_i soient ≥ -1 . Avoir les $m_i \geq 0$ garantit la positivité de la première classe de Chern, ce qu'affirme le lemme suivant.

⁽¹⁷⁾Du point de vue différentiel, il n'y a que deux classes d'isomorphisme de fibrés en $\mathbf{P}^1(\mathbf{C})$ sur $\mathbf{P}^1(\mathbf{C})$, parce que $\pi_1(\mathrm{PGL}(2; \mathbf{C})) \cong \mathbf{Z}/2$. La variété $\mathbf{P}(\mathcal{O}(-k) \oplus \mathbf{1})$ est donc diffeomorphe à $\mathbf{P}^1(\mathbf{C}) \times \mathbf{P}^1(\mathbf{C})$ quand k est pair et à $\tilde{\mathbf{P}}^2(\mathbf{C})$ quand k est impair.

Lemme 3.25. — *Si W est A -convexe, pour toute décomposition $A = A_1 + \cdots + A_k$ dans laquelle aucun des A_j n'est nul et tous sont représentables par des courbes rationnelles, on a $\langle c_1, A_j \rangle \geq 2$.*

Démonstration. — Par définition de la A -convexité, pour toute application holomorphe $u : \mathbf{P}^1(\mathbf{C}) \rightarrow W$ représentant une classe A_j , on a

$$u^*TW = \bigoplus_{i=1}^n \mathcal{O}(m_i) \text{ avec } m_i \geq 0 \text{ pour tout } i.$$

Mais, si A_j n'est pas nulle, u n'est pas constante et donc sa différentielle est injective (comme morphisme de faisceaux) et envoie un champ de vecteurs sur $\mathbf{P}^1(\mathbf{C})$ avec deux zéros sur une section non nulle de u^*TW avec au moins deux zéros. Donc au moins un des m_i est supérieur ou égal à 2. $\square$

Avec l'hypothèse de A -convexité, on peut démontrer :

Théorème 3.26. — *Si W est A -convexe, $\overline{\mathcal{M}}_{0,m}(W, A)$ est une variété projective normale et est localement le quotient d'une variété lisse par l'action d'un groupe fini. Sa dimension complexe est $\dim W + m - 3 + \langle c_1, A \rangle$.*

Il y a une très belle démonstration dans [26], dont voici un peu du parfum⁽¹⁸⁾. Les auteurs considèrent d'abord le cas où W est l'espace projectif $\mathbf{P}^n(\mathbf{C})$. Ils ajoutent des points marqués aux courbes de l'espace des applications stables, les points où ces courbes rencontrent les hyperplans de coordonnées (en supposant ces points distincts des points spéciaux déjà présents sur ces courbes). La figure 18 représente l'effet de cette opération sur une conique plane avec deux points marqués. Il faut ordonner les

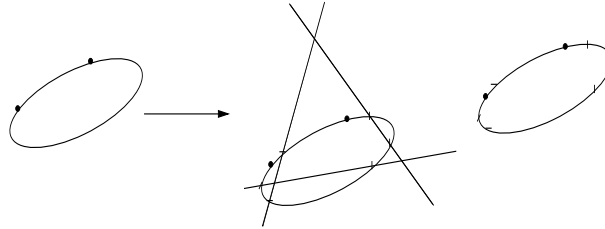


FIGURE 18

nouveaux points marqués, c'est ce qui va ensuite donner lieu aux quotients par des groupes finis présents dans l'énoncé. Les courbes avec tous ces points marqués sont ensuite considérées comme des courbes marquées stables abstraites.

⁽¹⁸⁾Voir plus généralement [15].

On remarquera que la donnée des points d'intersection de la courbe avec les hyperplans de coordonnées permet presque de reconstituer l'application stable, et précisément, permet de la reconstituer à l'action près du gros tore $(\mathbf{C}^*)^n$ sur $\mathbf{P}^n(\mathbf{C})$. On a donc établi une relation assez précise entre un gros ouvert de $\overline{\mathcal{M}}_{0,m}(\mathbf{P}^n(\mathbf{C}), A)$ et un sous-espace ouvert d'un espace de modules $\overline{\mathcal{M}}_{0,N}$ (je laisse aux volontaires le plaisir d'évaluer N).

En considérant différents choix d'hyperplans de coordonnées sur $\mathbf{P}^n(\mathbf{C})$, on a un recouvrement de $\overline{\mathcal{M}}_{0,m}(\mathbf{P}^n(\mathbf{C}), A)$ par de tels ouverts. On conclut en utilisant les propriétés de $\overline{\mathcal{M}}_{0,N}$. On passe enfin à une variété projective plus générale. $\square$

Remarque 3.27. — Les auteurs de [26] montrent aussi que $\overline{\mathcal{M}}_{0,m}(W, A)$ est un espace de modules au sens de la géométrie algébrique, c'est-à-dire qu'il représente un certain foncteur. Il y a aussi une « application stable universelle » (au moins au-dessus de l'espace des applications stables qui n'ont pas d'automorphisme du tout), c'est-à-dire un espace

$$\mathcal{U}_{0,m} \longrightarrow \overline{\mathcal{M}}_{0,m}(W, A)$$

qui ressemble beaucoup à $\overline{\mathcal{M}}_{0,m+1}(W, A)$ (la fibre au point $(\Sigma, \vec{z}, u)$ est la courbe Σ elle-même) et qui est muni d'une « application universelle » à valeurs dans W , l'évaluation au $(m+1)$ -ième point

$$(\Sigma, \vec{z}, u, z) \longmapsto u(z).$$

Définition 3.28. — On dit d'une variété projective W qu'elle est *convexe* si elle est convexe pour toutes les classes contenant des courbes rationnelles.

Remarque 3.29. — Cette propriété est équivalente à la suivante :

$$\forall u : \mathbf{P}^1(\mathbf{C}) \longrightarrow W, \quad H^1(\mathbf{P}^1(\mathbf{C}); u^*TW) = 0.$$

Elle doit être satisfaite par *toutes* les applications $u : \mathbf{P}^1(\mathbf{C}) \rightarrow W$. On en déduit que les degrés des facteurs de u^*TW sont ≥ 0 (si l'un d'eux était ≤ -1 , en composant avec une application de degré 2 de $\mathbf{P}^1(\mathbf{C})$ dans lui-même, on obtiendrait un fibré de degré < -1 ... et une contradiction).

Exemples 3.30

(1) Il y a de nombreux exemples de variétés projectives convexes, notamment toutes celles dont le fibré tangent est engendré par ses sections holomorphes globales. Tous les espaces homogènes G/P , où G est un groupe linéaire algébrique et P un sous-groupe parabolique, en particulier les espaces projectifs, les grassmanniennes et plus généralement les variétés de drapeaux, c'est-à-dire les quotients de $\mathrm{GL}(n; \mathbf{C})$ par les sous-groupes stabilisant un drapeau (complet ou non) $0 \subset P_1 \subset \dots \subset P_k = \mathbf{C}^n$.

(2) Il y a aussi beaucoup d'exemples de variétés non convexes. C'est le cas notamment de presque toutes les surfaces de Hirzebruch (voir l'exemple 3.23), les surfaces réglées ou espaces totaux des fibrés $\mathbf{P}(\mathcal{O}(-k) \oplus \mathbf{1}) \rightarrow \mathbf{P}^1(\mathbf{C})$. On a dit qu'une telle

surface contient une sphère plongée C d'auto-intersection $-k$, la section nulle du fibré $\mathcal{O}(-k)$. Pour un plongement $u : \mathbf{P}^1(\mathbf{C}) \rightarrow \mathbf{P}(\mathcal{O}(-k) \oplus \mathbf{1})$ d'image C , on a

$$\langle c_1(u^*TW), [\mathbf{P}^1(\mathbf{C})] \rangle = -k + 2.$$

Donc, si $u^*TW = \mathcal{O}(m_1) \oplus \mathcal{O}(m_2)$, $m_1 + m_2 = -k + 2$. Si $k \geq 2$, au moins un des m_i est ≤ -2 . Si $k = 1$, en considérant la composition de u avec une application de degré 2 de $\mathbf{P}^1$ dans lui-même, on arrive à la même conclusion. Donc la surface de Hirzebruch $\mathbf{P}(\mathcal{O}(-k) \oplus \mathbf{1})$ est convexe si et seulement si $k = 0$.

(3) Par exemple, pour $k = 1$, la surface de Hirzebruch est $\tilde{\mathbf{P}}^2(\mathbf{C})$, le plan éclaté en un point. Réexaminons donc l'exemple 2.4. La classe E du diviseur exceptionnel est représentée par une courbe rationnelle isolée. En composant avec une fraction rationnelle de degré 2, on obtient une courbe rationnelle de la classe $2E$. Comme on l'a déjà dit, les fractions rationnelles de degré 2 forment un espace de dimension complexe 5, ainsi

$$\dim_{\mathbf{C}} \overline{\mathcal{M}}_{0,3}(\tilde{\mathbf{P}}^2, 2E) \geq 5.$$

La dimension complexe attendue... et promise par le théorème 3.26 est 4, ce qui confirme la non-convexité de $\tilde{\mathbf{P}}^2(\mathbf{C})$.

3.e. Classes fondamentales virtuelles. — En général, l'opérateur D_u^J ne définit pas une section transverse à la section nulle et on n'a pas un espace de modules lisse de la dimension attendue.

Positivité, monotonie. — Pour éviter les nombreux problèmes évoqués ci-dessus à propos des revêtements multiples, on a commencé par définir les invariants de Gromov-Witten pour les variétés symplectiques *monotones*, celles pour lesquelles $c_1 = \lambda[\omega]$ pour un $\lambda > 0$, condition⁽¹⁹⁾ qui assure la positivité de $\langle c_1, A \rangle$ pour toutes les classes A contenant des courbes pseudo-holomorphes. Bien que le domaine d'application de cette théorie soit limité, c'est elle qui fait l'objet de l'exposé le plus lumineux, celui de McDuff et Salamon dans [57].

Perturbation, semi-positivité. — Les variétés symplectiques monotones étant somme toute assez rares, on a essayé d'étendre la construction à des variétés plus générales. En utilisant une méthode de perturbation due à Gromov (dans [32] encore), Ruan et Tian [67] ont élargi notablement le champ d'application de la théorie. Il s'agit de remplacer la condition de J -holomorphic

$$\frac{1}{2}(du + J \circ du \circ j) = 0$$

par la condition perturbée

$$\frac{1}{2}(du_z + J \circ du_z \circ j) = \nu(z, u(z))$$

⁽¹⁹⁾Pour une variété kählérienne, la condition est que la classe c_1 soit dans le cône des classes de Kähler. La variété est alors dite de *Fano*.

pour une certaine (petite) section anti- $\mathbf{C}$ -linéaire ν de $\text{Hom}(T\Sigma, TW) \rightarrow \Sigma \times W$. La ruse est que la courbe

$$\begin{aligned} \Sigma &\xrightarrow{\tilde{u}} \Sigma \times W \\ z &\longmapsto (z, u(z)) \end{aligned}$$

est alors J_ν -holomorphe pour une certaine structure presque complexe (proche de la structure produit). Remarquons que ces courbes $\tilde{u}$ sont des graphes et ne sont jamais, même à la limite, des revêtements multiples.

Pour faire fonctionner cette approche, on remplace la condition de positivité ci-dessus par une autre, plus faible mais néanmoins encore restrictive. On demande que pour toute classe d'homologie A telle que $\langle \omega, A \rangle > 0$ et $\langle c_1, A \rangle \geq 3 - n$, on ait $\langle c_1, A \rangle \geq 0$. On remarquera que c'est exactement la condition qui empêche le problème évoqué au § 2.d de se produire. On remarquera aussi que cette condition est automatiquement satisfaite par toutes les variétés symplectiques de dimension ≤ 6 (c'est-à-dire telles que $n \leq 3$).

Fibré excès. — Les constructions de classes fondamentales virtuelles sont basées sur l'idée du fibré « excès ». C'est une idée utilisée en topologie depuis quelques décennies (voir par exemple [64]). Supposons que X et Y soient deux sous-variétés d'une variété W se coupant le long d'une sous-variété lisse Z

$$\begin{array}{ccc} Z & \longrightarrow & X \\ \downarrow & & \downarrow \\ Y & \longrightarrow & W \end{array}$$

sans supposer que l'intersection soit transverse. On suppose toutefois que Z est lisse (ce que Quillen appelle dans [64] une intersection « propre »). On s'intéresse aux propriétés homologiques qu'aurait l'intersection transverse.

Sur la sous-variété Z , il y a un fibré vectoriel qui mesure le défaut de transversalité, le fibré excès $E = TW/TX + TY$ (qui serait nul si l'intersection était transversale). Alors l'intersection homologique de X et Y se représente par les zéros d'une section transverse de E .

Un cas particulier de cette situation est celui d'un fibré vectoriel $W \rightarrow X$ muni d'une section s . On appelle X la section nulle et Y l'image $s(X)$. On s'intéresse aux zéros de s , c'est-à-dire à l'intersection Z de X et de Y . Quand s est transverse à la section nulle, ses zéros représentent une classe d'homologie $[Z]$ duale à la classe d'Euler du fibré $W \rightarrow X$. Si Z est lisse sans que l'intersection soit transverse, on considère le fibré excès $E \rightarrow Z$. La classe duale à sa classe d'Euler est une classe $[Z]^{\text{vir}} \in H_m(Z; \mathbf{Z})$, où m est, précisément, la dimension $\dim X - \text{rg } W$ qu'aurait l'espace des zéros d'une section transverse... la dimension « virtuelle » de Z , en quelque sorte. On peut utiliser cette classe pour faire des calculs homologiques.

Le cas des espaces de modules d'applications stables est plus compliqué. Une des raisons est qu'on ne joue pas là avec des variétés et des fibrés de dimension ou de rang fini(e), ni même de rang constant.

Excès en géométrie algébrique. — Dans la construction algébro-géométrique de la classe fondamentale virtuelle, on se restreint au cas des variétés projectives lisses. L'obtention d'un cycle de la bonne dimension est réalisée grâce à une « théorie d'obstruction cotangente » (voir les articles de Behrend et Fantechi [11] et de Li et Tian [49]). Il faut noter que cette théorie, bien que très abstraite, permet effectivement de calculer des invariants de Gromov-Witten (voir [30, 74]).

Dans cette approche algébro-géométrique, la bonne notion d'espace est celle de « champ algébrique ». C'est une théorie, certes abstraite encore une fois, mais dans laquelle les quotients sont indolores, c'est pourquoi elle est d'usage courant dans la théorie des espaces de modules.

Excès simplifié. — Une famille d'exemples relevant de la géométrie algébrique et dans laquelle la théorie de l'excès s'applique simplement est celle des sous-variétés de variétés projectives convexes dont le fibré normal est « convexe » (une idée de Kontsevich [44] déjà exploitée par Givental [29]). J'en donnerai un exemple au § 4.d. Du point de vue pratique, ce cas, bien que très simple, est important : il contient notamment les variétés qui sont des intersections complètes dans des produits d'espaces projectifs.

Excès en géométrie symplectique. — Dans le cadre symplectique, où il n'y a pas de structure presque complexe meilleure que les autres, le rôle du champ algébrique est joué par une notion elle aussi assez sophistiquée, celle de « structure de Kuranishi ». Voir les travaux de Siebert [71] et de Fukaya et Ono [25].

L'approche de [71] utilise très directement la technique de l'excès. Quand on essaie d'appliquer cette technique aux sections Fredholm de fibrés banachiques sur des variétés banachiques (celles et ceux du § 2.d.c), tout se passe bien tant qu'on n'a pas besoin de changer la courbe source. Dès qu'on se met à considérer des bulles, les difficultés techniques sérieuses apparaissent. Siebert les surmonte, en fait, assez directement.

Il faut encore vérifier que le résultat obtenu ne dépend pas de la structure complexe utilisée pour l'obtenir. Voir, toujours, [71].

Relations. — Si l'approche symplectique est plus flexible, l'approche algébrique a l'avantage de ne jamais toucher à la structure complexe, ce qui est important pour les questions énumératives. Comme la plupart des variétés symplectiques utilisées dans la vie courante sont des variétés projectives complexes, les deux approches y coexistent. Il est donc nécessaire, aussi, de vérifier que les deux approches donnent les mêmes invariants. C'est bien le cas, comme l'a montré Siebert [72].

En pratique, on peut donc calculer les invariants de Gromov-Witten « symplectiques » (des variétés projectives) par la géométrie algébrique.

4. Propriétés et calculs d'invariants, enfin

Suivant Kontsevich et Manin [45], les invariants de Gromov-Witten devraient satisfaire un certain nombre de propriétés, que je vais examiner dans cette partie. Je donnerai notamment une idée de ce qu'est la célèbre règle de décomposition. Je montrerai aussi comment on peut l'utiliser pour calculer (récursivement) certains invariants de Gromov-Witten, les nombres N_d de courbes rationnelles de degré d passant par $3d-1$ points du plan. Je montrerai aussi un exemple relativement simple de calcul d'invariants à l'aide d'une classe fondamentale virtuelle, s'appliquant notamment au cas du plan éclaté en un point.

4.a. Propriétés. — On définit donc des invariants

$$\Psi_{g,m}^A \in \text{Hom}(H^*(W)^{\otimes m}; \mathbf{Q})$$

en intégrant sur la classe fondamentale (quelle qu'elle soit) :

$$\Psi_{g,m}^A(\xi_1 \otimes \cdots \otimes \xi_m) = \langle \text{év}_m^* \xi_1 \otimes \cdots \otimes \xi_m, [\overline{\mathcal{M}}_{g,m}^J(W, A)] \rangle.$$

Notations. — Je n'ai pas fait figurer le J dans la notation puisque le résultat n'en dépend pas. Le m est souvent inutile aussi : on voit bien sur quelle classe on calcule l'invariant. Quand le genre sera nul, je ne le ferai pas figurer non plus. Donc $\Psi^A(\xi^{\otimes m})$ signifiera, si $\xi \in H^*(W)$, la même chose que $\Psi_{0,m}^A(\xi^{\otimes m})$.

Voici une liste de propriétés auxquelles satisfont les invariants. Je donne des démonstrations avec classes fondamentales naïves ou des idées de démonstrations seulement.

Symétrie. — Il s'agit de l'invariance par le groupe symétrique permutant les facteurs dans W et les points marqués des courbes sources. Il est clair que cette propriété est satisfaite dans les cas où les approches directes ou naïves sont suffisantes. Elle se vérifie aussi facilement dans les approches sophistiquées évoquées au § 3.e puisque le groupe symétrique est fini.

Dimension. — On a $\Psi_{g,m}^A(\xi) = 0$ si le degré de la classe de cohomologie ξ ne coïncide pas avec la dimension virtuelle de $\overline{\mathcal{M}}_{g,m}^J(W, A)$.

Unité. — Appelons $1 \in H^0(W)$ l'unité du cup-produit.

Proposition 4.1. — On a $\Psi_{g,m}^A(\xi_1 \otimes \cdots \otimes \xi_{m-1} \otimes 1) = 0$ sauf si $(g, m) = (0, 3)$ et $A = 0$, cas où l'on a plus généralement

$$\Psi_{0,3}^0(\xi_1 \otimes \xi_2 \otimes \xi_3) = \langle \xi_1 \smile \xi_2 \smile \xi_3, [W] \rangle.$$

Démonstration. — Supposons d'abord $m \geq 3$ ou $g \neq 0$ et considérons le cas où $A = 0$. On a vu au §3.b que, pour la classe nulle, on a

$$\overline{\mathcal{M}}_{g,m}(W, 0) = \overline{\mathcal{M}}_{g,m} \times W.$$

On a dit aussi que l'application d'évaluation n'est autre que

$$\begin{aligned} \overline{\mathcal{M}}_{g,m} \times W &\longrightarrow W^m \\ (\Sigma, \vec{z}, x) &\longmapsto (x, x, \dots, x). \end{aligned}$$

Donc, si $p : \overline{\mathcal{M}}_{g,m} \times W \rightarrow W$ est la projection,

$$\begin{aligned} \langle (\text{é}v_m)^*(\xi_1 \otimes \dots \otimes \xi_m), [\overline{\mathcal{M}}_{g,m} \times W] \rangle &= \langle p^*\xi_1 \smile \dots \smile p^*\xi_m, [\overline{\mathcal{M}}_{g,m} \times W] \rangle \\ &= \begin{cases} 0 & \text{si } m > 3 \\ \langle \xi_1 \smile \xi_2 \smile \xi_3, [W] \rangle & \text{si } m = 3 \end{cases} \end{aligned}$$

puisque $\overline{\mathcal{M}}_{g,m}$ a une dimension strictement positive sous nos hypothèses.

Considérons maintenant l'unité $1 \in H^0(W)$ du cup-produit, c'est-à-dire la classe de cohomologie duale à la classe fondamentale, ou à l'inclusion $W \subset W$. Imposer qu'un des points marqués soit envoyé dans cette sous-variété, c'est ne rien imposer du tout. Formalisons cette remarque. Remarquons d'abord qu'on a, par définition

$$\Psi_{g,m}^A(\xi_1 \otimes \dots \otimes \xi_{m-1} \otimes 1) = \langle (\text{é}v_m)^*(\xi_1 \otimes \dots \otimes \xi_{m-1} \otimes 1), [\overline{\mathcal{M}}_{g,m}(W, A)] \rangle.$$

Si $g \neq 0$, ou $g = 0$ et $m \geq 4$, ou $A \neq 0$, il y a une application bien définie

$$\theta : \overline{\mathcal{M}}_{g,m}(W, A) \longrightarrow \overline{\mathcal{M}}_{g,m-1}(W, A),$$

qui oublie le dernier point marqué (et contracte les composantes devenues instables). Considérons le diagramme commutatif

$$\begin{array}{ccc} \overline{\mathcal{M}}_{g,m}(W, A) & \xrightarrow{\text{é}v_m} & W^m \\ \theta \downarrow & & \downarrow p \\ \overline{\mathcal{M}}_{g,m-1}(W, A) & \xrightarrow{\text{é}v_{m-1}} & W^{m-1} \end{array}$$

dans lequel la projection p oublie le dernier facteur. Comme $\xi_1 \otimes \dots \otimes \xi_{m-1} \otimes 1 = p^*(\xi_1 \otimes \dots \otimes \xi_{m-1})$, on a

$$\begin{aligned} \langle (\text{é}v_m)^*(\xi_1 \otimes \dots \otimes \xi_{m-1} \otimes 1), [\overline{\mathcal{M}}_{g,m}(W, A)] \rangle &= \langle \theta^*(\text{é}v_{m-1})^*(\xi_1 \otimes \dots \otimes \xi_{m-1}), [\overline{\mathcal{M}}_{g,m}(W, A)] \rangle \\ &= \langle (\text{é}v_{m-1})^*(\xi_1 \otimes \dots \otimes \xi_{m-1}), \theta_*[\overline{\mathcal{M}}_{g,m}(W, A)] \rangle. \end{aligned}$$

Comme $\dim \overline{\mathcal{M}}_{g,m-1}(W, A) < \dim \overline{\mathcal{M}}_{g,m}(W, A)$, la classe $\theta_*[\overline{\mathcal{M}}_{g,m}(W, A)]$ vit dans un groupe d'homologie trivial et donc elle est nulle. $\square$

Classes de degré 2. — Comme les classes d'homologie des courbes jouent un rôle assez primordial dans la construction des invariants de Gromov-Witten, il n'est pas étonnant que les classes de cohomologie de degré 2 jouent aussi un rôle particulier. C'est l'« axiome des diviseurs » de [45]. Voici la propriété précise. Pour la démonstration, je me placerai dans le cas d'une variété convexe.

Proposition 4.2. — *Soit ξ une classe de cohomologie de degré 2. Pour toute classe de cohomologie $\alpha \in H^*(W^{\otimes m})$, on a*

$$\Psi_{g,m+1}^A(\alpha \otimes \xi) = \langle \xi, A \rangle \Psi_{g,m}^A(\alpha).$$

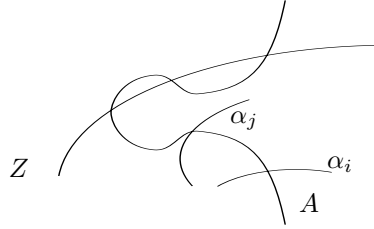


FIGURE 19

Soit Z un cycle de codimension 2 dual à la classe ξ . Les deux invariants qui apparaissent dans l'énoncé de la proposition sont

- D'une part $\Psi_{g,m}^A(\alpha)$ qui compte les courbes de genre g de la classe A avec m points marqués envoyés dans les cycles duaux aux α .
- D'autre part, l'invariant $\Psi_{g,m+1}^A(\alpha \otimes \xi)$ qui compte les mêmes courbes de genre g de la même classe A avec les mêmes conditions d'incidence, plus le fait qu'un point marqué supplémentaire est envoyé dans Z .

Dans les deux décomptes, les courbes sont exactement les mêmes. Il reste à choisir le $(m+1)$ -ième point marqué. On doit le prendre en un des points d'intersection de la courbe et de Z . Idéalement, il y a $\langle \xi, A \rangle$ tels points (figure 19), ce qui explique la formule, ce qui en donne aussi la démonstration dans le cas projectif complexe convexe, par exemple. $\square$

Exemple 4.3. — L'invariant de Gromov-Witten $\Psi_{0,3}^L(p^n \otimes p^n \otimes p)$ dans $\mathbf{P}^n(\mathbf{C})$ compte les droites passant par deux points généraux (classes d'homologie duales à p^n) et coupant un hyperplan assez général (classe d'homologie duale à p). Toute droite coupe un hyperplan général en un point donc la dernière condition n'ajoute rien. On a donc

$$\Psi_{0,3}^L(p^n \otimes p^n \otimes p) = \langle p, L \rangle \Psi_{0,2}^L(p^n \otimes p^n) \quad (= 1).$$

Règle de décomposition. — La propriété la plus difficile, à énoncer comme à démontrer, est la célèbre règle de décomposition. Je vais l'expliciter plus bas dans le cas du genre nul.

4.b. Potentiel de Gromov-Witten et cohomologie quantique. — On peut organiser formellement les invariants de Gromov-Witten de genre 0 en une somme infinie, le *potentiel de Gromov-Witten* :

$$\Psi(\xi) = \sum_{m \geq 3} \frac{1}{m!} \sum_A \Psi^A(\xi^{\otimes m}) q^A,$$

dont il n'est pas question de discuter la convergence ici. Dans cette formule

- ξ désigne une classe de cohomologie, $\xi \in H^*(W)$,
- la sommation porte sur toutes les classes $A \in H_2(X; \mathbf{Z})$,
- le symbole $\Psi^A(\xi^{\otimes m})$ est l'invariant de Gromov-Witten obtenu en évaluant la classe de cohomologie $\text{év}_m^* \xi^{\otimes m}$ sur la classe fondamentale de $\overline{\mathcal{M}}_{0,m}^J(W, A)$,
- q^A désigne la classe A vue multiplicativement dans l'anneau du groupe additif $H_2(W; \mathbf{Z})/\text{Torsion}$. On peut considérer $q = (q_1, \dots, q_r)$ comme une variable formelle, si $A_1, \dots, A_r$ désigne une base du groupe abélien libre $H_2(W; \mathbf{Z})/\text{Torsion}$ et si $A = a_1 A_1 + \dots + a_r A_r$, alors $q^A = q_1^{a_1} \dots q_r^{a_r}$.

Les dérivées troisièmes de Ψ permettent de définir, toujours modulo les questions de convergence, le *produit quantique* sur la cohomologie de W . On décrit la classe de cohomologie $\alpha \star_{\xi, q} \beta$ produit de α et β en l'évaluant sur une classe d'homologie c :

$$\begin{aligned} \langle (\alpha \star_{\xi, q} \beta), c \rangle &= (d^3 \Psi)_\xi(\alpha, \beta, \gamma) \\ &= \sum_{m \geq 0} \frac{1}{m!} \sum_A \Psi^A(\xi^{\otimes m} \otimes \alpha \otimes \beta \otimes \gamma) q^A, \end{aligned}$$

expression dans laquelle γ désigne la classe de cohomologie duale à c .

Les propriétés du produit quantique cachées dans les propriétés dont j'ai fait la liste au § 4.a sont

- sa commutativité (invariance par le groupe symétrique)
- le fait qu'il ait la même unité $1 \in H^0(W)$ que le cup-produit.

La règle de décomposition, quant à elle, exprime son associativité. Elle dit aussi que le potentiel de Gromov-Witten satisfait les équations dites WDVV, ou que la cohomologie munie du produit quantique est une variété de Frobenius (voir par exemple [16, 21, 37, 68, 53, 5, 69]).

4.c. La règle de décomposition en genre 0. — La version en genre 0 de la règle de décomposition concerne une sphère avec quatre points marqués qui dégénère sur un bouquet de deux sphères, chacune gardant deux des points marqués. Les démonstrations symplectiques sont basées sur un argument de recollement : il faut montrer que, près d'un bouquet de deux sphères J -holomorphes, il y a une sphère J -holomorphe. Je renvoie à [57] ou à [67] pour des explicitations de cet argument.

La démonstration dont je donne une idée ici vaut dans le cas projectif complexe. Elle fonctionne très bien pour les variétés convexes et vient de [26]. Bien que non symplectique, elle trouve sa place ici parce qu'elle est basée sur un argument géométrique finalement très simple qui en explique bien la raison d'être.

Fixons une base $(\partial_1, \dots, \partial_N)$ de $H^*(W)$. Appelons g la métrique définie sur $H^*(W)$ par le cup-produit. Dans la base (∂_i) ,

$$g_{i,j} = \langle \partial_i \smile \partial_j, [W] \rangle.$$

Notons $(g^{i,j})$ l'inverse de $(g_{i,j})$. On veut montrer :

Théorème 4.4. — *Pour toute classe ξ et tous indices i, j, r et s , on a*

$$\begin{aligned} \sum \frac{1}{m_1!m_2!} \Psi^{A_1}(\xi^{m_1} \partial_i \partial_j \partial_k) g^{k,n} \Psi^{A_2}(\xi^{m_2} \partial_r \partial_s \partial_n) q^{A_1+A_2} \\ = \sum \frac{1}{m_1!m_2!} \Psi^{A_1}(\xi^{m_1} \partial_r \partial_j \partial_k) g^{k,n} \Psi^{A_2}(\xi^{m_2} \partial_i \partial_s \partial_n) q^{A_1+A_2} \end{aligned}$$

(sommation sur m_1, m_2, A_1, A_2, k et n).

On veut décrire ce qui se passe dans l'homologie de $\overline{\mathcal{M}}_{0,m}(W, A)$ quand la classe A se décompose. Il faut donc considérer des courbes qui ont deux composantes, dont chacune contient une partie des points marqués. Voir la figure 21. On réduira finalement le calcul homologique au même problème dans l'espace de modules de « quatre points sur une droite », $\overline{\mathcal{M}}_{0,4}$.

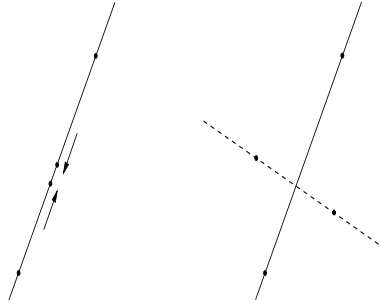


FIGURE 20

Quatre points sur une droite. — Regardons donc d'un peu plus près l'espace de Deligne-Mumford $\overline{\mathcal{M}}_{0,4}$. C'est une compactification de l'espace $\mathcal{M}_{0,4}$ des classes d'isomorphisme des ensembles de quatre points ordonnés sur $\mathbf{P}^1(\mathbf{C})$. Mais quatre points distincts de $\mathbf{P}^1(\mathbf{C})$ à isomorphisme près, c'est un point de $\mathbf{P}^1(\mathbf{C}) - \{\infty, 0, 1\}$, leur birapport : une unique homographie envoie z_1 sur ∞ , z_2 sur 0 et z_3 sur 1, elle envoie z_4 sur le birapport $[z_1, z_2, z_3, z_4]$. Ainsi on a

$$\mathcal{M}_{0,4} \cong \mathbf{P}^1 - \{\infty, 0, 1\}.$$

La compactification de Deligne-Mumford est un processus très impressionnant : on compactifie un espace de points distincts... en les gardant distincts. Quand deux points se rapprochent, ils créent une nouvelle composante rationnelle sur laquelle ils se retrouvent, toujours distincts. C'est ce que montre la figure 20.

Que peut bien être la compactification $\overline{\mathcal{M}}_{0,4}$ de $\mathbf{P}^1(\mathbf{C}) - \{\infty, 0, 1\}$? C'est $\mathbf{P}^1(\mathbf{C})$, bien sûr. On doit ajouter des courbes rationnelles réductibles constituées de deux droites sécantes contenant chacune deux des quatre points marqués. Il y a trois classes d'isomorphisme : à isomorphisme près, on ne voit que les numéros des points sur les composantes, lesquelles composantes sont d'ailleurs indistinguables. En utilisant le birapport, les trois points ajoutés $(14 | 23)$, $(24 | 13)$ et $(34 | 12)$ correspondent à ∞ , 0 et 1 respectivement.

Démonstration du théorème 4.4. — En identifiant les coefficients des puissances de q , on est amené à comprendre — dans l'homologie de $\overline{\mathcal{M}}_{0,m}(W, A)$ — ce qui se passe quand A se décompose en $A_1 + A_2$ (figure 21). Écrivons $\{1, \dots, m\} = E_1 \cup E_2$ (et supposons $\#E_i \geq 2$).

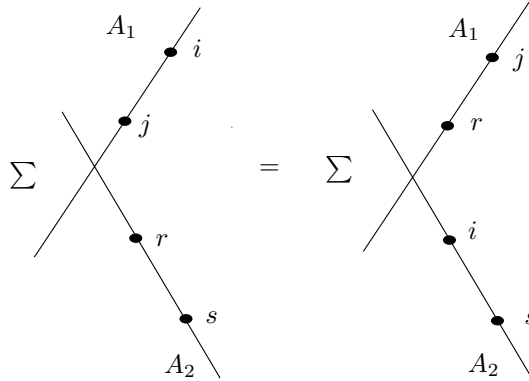


FIGURE 21

Fixons les décompositions de A et de $\{1, \dots, m\}$; on obtient un plongement

$$j : \overline{\mathcal{M}}_{0, E_1 \cup \{z\}}(W, A_1) \times_W \overline{\mathcal{M}}_{0, E_2 \cup \{z\}}(W, A_2) \longrightarrow \overline{\mathcal{M}}_{0, m}(W, A)$$

avec des notations que j'espère évidentes. Écrivons pour simplifier

$$M_i = \overline{\mathcal{M}}_{0, E_i \cup \{z\}}(W, A_i), \quad M = \overline{\mathcal{M}}_{0, m}(W, A) \quad \text{et} \quad D_{E_1, E_2}(A_1, A_2) = M_1 \times_W M_2.$$

Comme $\dim_{\mathbf{C}} M_i = n + \#E_i - 2 + \langle c_1, A_i \rangle$, la dimension complexe de $D_{E_1, E_2}(A_1, A_2)$ est $\dim_{\mathbf{C}} M - 1$, c'est un diviseur de notre espace de modules. Appelons i l'inclusion $D_{E_1, E_2}(A_1, A_2) \subset M_1 \times M_2$. Si un point marqué porte le numéro p , appelons e_p l'application d'évaluation en ce point.

Lemme 4.5. — Pour toute classe de cohomologie $\eta = \eta_1 \otimes \cdots \otimes \eta_m$ dans $H^*(W)^{\otimes m}$, on a

$$i_* j^* (\text{é}v_m)^*(\eta) = \sum_{k,n} g^{k,n} \left[\left(\prod_{p \in E_1} e_p^*(\eta_p) \right) e_z^*(\partial_k) \right] \times \left[\left(\prod_{q \in E_2} e_q^*(\eta_q) \right) e_z^*(\partial_n) \right].$$

La démonstration est un calcul direct, que je laisse pour plus tard. On en déduit l'égalité

$$\begin{aligned} \sum g^{k,n} \Psi^{A_1} \left(\left(\bigotimes_{p \in E_1} \eta_p \right) \otimes \partial_k \right) \Psi^{A_2} \left(\left(\bigotimes_{q \in E_2} \eta_q \right) \otimes \partial_n \right) \\ = \sum \langle \text{é}v_m^*(\eta), [D_{E_1, E_2}(A_1, A_2)] \rangle \end{aligned}$$

(les sommations portent sur les décompositions de $\{1, \dots, m\}$ en $E_1 \cup E_2$ avec $E_1 \supset \{a, b\}$ et $E_2 \supset \{c, d\}$). Posons

$$\begin{cases} \eta_1 = \cdots = \eta_{m-4} = \xi \\ \eta_{m-3} = \partial_i, \eta_{m-2} = \partial_j, \eta_{m-1} = \partial_r, \eta_m = \partial_s \\ a = m-3, b = m-2, c = m-1, d = m \end{cases}$$

dans le membre de gauche. Ce membre de gauche devient

$$\sum C(m_1, m_2) g^{k,n} \Psi^{A_1} (\xi^{m_1-2} \partial_i \partial_j \partial_k) \Psi^{A_2} (\xi^{m_2-2} \partial_r \partial_s \partial_n)$$

où $C(m_1, m_2)$ est le nombre de décompositions de $\{1, \dots, m\}$ en $E_1 \cup E_2$ avec $\#E_i = m_i$ et on fait la somme sur toutes les décompositions telles que $E_1 \supset \{a, b\}$ et $E_2 \supset \{c, d\}$. Finalement ce membre de gauche vaut

$$m! \sum_{\substack{m_1, m_2 \geq 0 \\ m_1 + m_2 = m-4}} \frac{1}{m_1! m_2!} g^{k,n} \Psi^{A_1} (\xi^{m_1} \partial_i \partial_j \partial_k) \Psi^{A_2} (\xi^{m_2} \partial_r \partial_s \partial_n).$$

Lemme 4.6. — Dans $H_* (\overline{\mathcal{M}}_{0,m}(W, A))$, si a, b, c et d sont quatre éléments distincts de $\{1, \dots, m\}$, on a

$$\sum_{\substack{a, b \in E_1 \\ c, d \in E_2}} [D_{E_1, E_2}(A_1, A_2)] = \sum_{\substack{c, b \in E_1 \\ a, d \in E_2}} [D_{E_1, E_2}(A_1, A_2)].$$

C'est précisément ce dont nous avons besoin pour finir la démonstration du théorème 4.4. $\square$

Démonstration du lemme 4.6. — C'est le cœur géométrique et la plus belle partie de la démonstration. L'idée est que $\overline{\mathcal{M}}_{0,4}$ n'est autre que $\mathbf{P}^1(\mathbf{C})$ (comme on l'a dit ci-dessus) et que deux points quelconques de $\mathbf{P}^1(\mathbf{C})$ représentent la même classe d'homologie.

Considérons la contraction composée

$$\overline{\mathcal{M}}_{0,m}(W, A) \longrightarrow \overline{\mathcal{M}}_{0,m} \longrightarrow \overline{\mathcal{M}}_{0,\{a,b,c,d\}}.$$

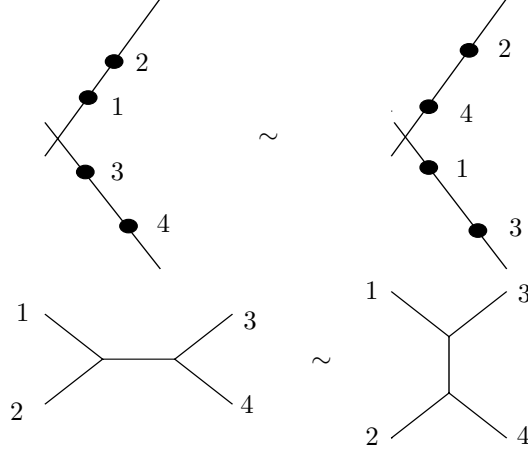


FIGURE 22

Le diviseur $\sum_{\substack{a,b \in E_1 \\ c,d \in E_2}} D_{E_1, E_2}(A_1, A_2)$ est l'image inverse du point $(a, b \mid c, d)$. Le diviseur

$$\sum_{\substack{c,b \in E_1 \\ a,d \in E_2}} D_{E_1, E_2}(A_1, A_2),$$

quant à lui, est l'image de $(c, b \mid a, d)$. Comme les deux points de $\mathbf{P}^1(\mathbf{C})$ sont équivalents, il en est de même des diviseurs dans $\overline{\mathcal{M}}_{0,m}(W, A)$. $\square$

Remarque 4.7. — Ce résultat s'exprime par les figures familières aux physiciens (figure 22). Voir [16].

Démonstration du lemme 4.5. — On commence par contempler le diagramme commutatif

$$\begin{array}{ccccc} M & \xleftarrow{j} & D_{E_1, E_2}(A_1, A_2) & \xrightarrow{i} & M_1 \times M_2 \\ \text{év}_m \downarrow & & \varepsilon \downarrow & & \downarrow \text{év}_{m_1+1} \times \text{év}_{m_2+1} \\ W^m & \xleftarrow{\text{proj}} & W^{m+1} & \xrightarrow{\Delta} & W^{m+2} \end{array}$$

dans lequel on voit qu'on a

$$\begin{aligned} i_* j^* (\text{év}_m)^* (\eta) &= i_* \varepsilon^* \text{proj}^* (\eta) \\ &= i_* \varepsilon^* (\eta \otimes D[W]) \\ &= (\text{év}_{m_1+1} \times \text{év}_{m_2+1})^* \Delta_* (\eta \otimes D[W]) \\ &= (\text{év}_{m_1+1} \times \text{év}_{m_2+1})^* (\eta \otimes D[\Delta]) \end{aligned}$$

(D désigne la dualité de Poincaré). On remarque maintenant que la classe duale à la diagonale dans $W \times W$ est

$$D[\Delta] = \sum_{k,n} g^{k,n} \partial_k \otimes \partial_n.$$

On a enfin

$$\begin{aligned} i_* j^* (\text{év}_m)^* (\eta) &= \sum_{k,n} g^{k,n} (\text{év}_{m_1+1} \times \text{év}_{m_2+1})^* (\eta \otimes \partial_k \otimes \partial_n) \\ &= \sum_{k,n} g^{k,n} \left[\left(\prod_{p \in E_1} e_p^*(\eta_p) \right) e_{z_0}^*(\partial_k) \right] \times \left[\left(\prod_{q \in E_2} e_q^*(\eta_q) \right) e_{z_0}^*(\partial_n) \right], \end{aligned}$$

la relation qu'on voulait démontrer. $\square$

Les courbes rationnelles de degré d . — La cohomologie de $\mathbf{P}^2(\mathbf{C})$ est l'anneau $\mathbf{Z}[p]/\langle p^3 \rangle$ où p désigne la classe duale à une droite.

Écrivons donc $\xi = \xi_0 + \xi_1 p + \xi_2 p^2$ pour les éléments de $H^*(\mathbf{P}^2)$, de sorte que le potentiel de Gromov-Witten

$$\Psi(\xi) = \sum_{m \geq 3} \frac{1}{m!} \sum_{d \geq 0} \Psi^{dL}(\xi^{\otimes m})$$

s'écrit

$$\Psi(\xi) = \frac{1}{2} (\xi_0^2 \xi_2 + \xi_0 \xi_1^2) + \sum_{d > 0} \sum_{m \geq 3} \frac{1}{m!} \Psi^{dL}(\xi^{\otimes m}).$$

En utilisant la proposition 4.2, comme p est une classe de degré 2, on peut réécrire cette somme comme

$$\sum_{m \geq 3} \sum_{d > 0} \Psi^{dL}((p^2)^{\otimes m}) \frac{\xi_2^m}{m!} \exp\langle \xi_1 p, dL \rangle.$$

On a déjà fait au §3.b.b le calcul de dimension qui montre que l'invariant $\Psi^{dL}((p^2)^{\otimes m})$ ne peut être non nul que si $m = 3d - 1$, auquel cas cet invariant est le nombre N_d de courbes rationnelles de degré d passant par $3d - 1$ points du plan en position générale. On peut donc réécrire le potentiel de Gromov-Witten dans le cas du plan sous la forme

$$\Psi(\xi) = \frac{1}{2} (\xi_0^2 \xi_2 + \xi_0 \xi_1^2) + \sum_{d \geq 1} N_d e^{d\xi_1} \frac{\xi_2^{3d-1}}{(3d-1)!}.$$

Dans cette dimension, la règle de décomposition (ou l'associativité du produit quantique, ou les équations WDVV) équivaut à l'équation

$$\frac{\partial^3 \Psi}{\partial \xi_2^3} = \left(\frac{\partial^3 \Psi}{\partial \xi_1^2 \partial \xi_2} \right)^2 - \frac{\partial^3 \Psi}{\partial \xi_1^3} \frac{\partial^3 \Psi}{\partial \xi_1 \partial \xi_2^2}.$$

On identifie les coefficients de $e^{d\xi_1} \xi_2^{3d-4}$ dans les deux membres et on obtient :

Théorème 4.8 (Kontsevich [44]). — Le nombre N_d de courbes rationnelles planes de degré d passant par $3d - 1$ points généraux est donné par la formule récursive

$$N_d + \sum_{\substack{d_1+d_2=d \\ d_1, d_2 \geq 1}} d_1^3 d_2 \binom{3d-4}{3d_1-1} N_{d_1} N_{d_2} = \sum_{\substack{d_1+d_2=d \\ d_1, d_2 \geq 1}} d_1^2 d_2^2 \binom{3d-4}{3d_1-2} N_{d_1} N_{d_2}$$

et la valeur initiale $N_1 = 1$. □

Cette relation permet théoriquement de calculer les N_d de proche en proche en commençant par $N_1 = 1$. On trouve $N_2 = 1$ (rien de neuf, une conique par cinq points en position générale), $N_3 = 12$, $N_4 = 620$ (ces deux nombres, moins populaires que N_1 et N_2 étaient néanmoins connus des spécialistes), $N_5 = 87\,304$ (à partir de là, on obtient des résultats qui n'étaient pas connus avant le théorème de Kontsevich), $N_6 = 26\,312\,976$ etc.

Remarque 4.9. — En utilisant des techniques proches de celle évoquée par la figure 17, on peut vérifier de façon assez simple (sans utiliser les techniques du § 3.e) que ce nombre N_d est le même pour toute structure presque complexe générique calibrée par la forme symplectique standard de $\mathbf{P}^2(\mathbf{C})$. Pour une démonstration et beaucoup de jolis résultats sur les courbes J -holomorphes planes et leurs singularités, voir [10].

4.d. Une classe fondamentale virtuelle mais simple. — On a déjà eu l'occasion de mentionner que les approches directes ne pouvaient pas fonctionner pour calculer les invariants Ψ^{2E} de $\tilde{\mathbf{P}}^2(\mathbf{C})$: la classe $2E$ n'est régulière pour aucune structure presque complexe (voir le § 2.4 et l'exemple 3.30). Dans ce cas pourtant, et plus généralement dans le cas des intersections complètes dans un produit d'espaces projectifs, il y a une version assez élémentaire de la classe fondamentale virtuelle (voir [44, 29]) qui permet le calcul de certains des invariants.

4.d.a. Fibrés vectoriels convexes. — Soit W une variété projective complexe *convexe* (voir la définition 3.28) et soit $V \rightarrow W$ un fibré vectoriel complexe. On dit que V est *convexe* s'il est engendré par ses sections holomorphes globales, c'est-à-dire si $H^1(W; V) = 0$.

Examinons maintenant l'ensemble Y des zéros d'une section holomorphe de V transverse à la section nulle : c'est une sous-variété de W dont le fibré normal est la restriction du fibré V à Y :

$$j : Y \subset W, \quad \nu_j = j^*V.$$

Il n'y a pas de raison pour que Y soit convexe. L'espace de modules $\overline{\mathcal{M}}_{0,m}(Y, A)$ peut donc être très mauvais. Voilà comment on peut faire fonctionner l'idée du fibré excès dans ce cas :

– L'espace de modules $\overline{\mathcal{M}}_{0,m}(W, j_*A)$ (j désigne l'inclusion $Y \subset W$), lui, est bon : la variété W est supposée convexe. On montre qu'il porte un fibré vectoriel $\mathcal{V}$ dont la fibre au point représenté par $u : C \rightarrow W$ est $H^0(C, u^*V)$. Pour ceci, on peut utiliser le

fait que $\overline{\mathcal{M}}_{0,m}(W, j_*A)$ est un espace de modules au sens de la géométrie algébrique, avec une application stable universelle (voir le § 3.d.a).

– Pour tout $\alpha \in H^*(W)^{\otimes m}$, on pose

$$\Psi^A((j^m)^*\alpha) = \langle (\text{év}_m)^*\alpha \smile e(\mathcal{V}), [\overline{\mathcal{M}}_{0,m}(W, j_*A)] \rangle.$$

Attention, on ne définit ainsi des invariants que sur les classes de cohomologie provenant de la cohomologie de la variété ambiante W . Cette réserve mise à part, c'est donc la classe duale à la classe d'Euler de $\mathcal{V}$ qui joue le rôle de classe fondamentale.

Remarques 4.10

(1) On peut vérifier que, si la variété Y est A -convexe, la définition donnée coïncide avec la définition précédente de $\Psi^A((j^m)^*\alpha)$.

(2) On vérifie aussi que le résultat ne dépend pas du choix du plongement de Y dans la variété convexe.

(3) Ces invariants satisfont la règle de décomposition 4.c, ce qu'on peut démontrer en utilisant la multiplicativité de la classe d'Euler.

Le plan éclaté en un point. — On va appliquer cette stratégie au cas où Y est la variété $\tilde{\mathbf{P}}^2(\mathbf{C})$ obtenue en éclatant $\mathbf{P}^2(\mathbf{C})$ en un point. Comme on l'a déjà signalé par exemple en § 3.d.a, si E est la classe du diviseur exceptionnel, Y n'est pas $2E$ -convexe.

On commence par plonger Y dans une variété convexe. C'est l'hypersurface de bidegré $(1, 1)$ dans $\mathbf{P}^1(\mathbf{C}) \times \mathbf{P}^2(\mathbf{C})$, en d'autres termes c'est le lieu des zéros d'une section du fibré vectoriel convexe $V = \mathcal{O}(1) \otimes \mathcal{O}(1)$:

$$Y = \{(\ell, d) \in \mathbf{P}^1(\mathbf{C}) \times \mathbf{P}^2(\mathbf{C}) \mid d \subset \ell \oplus \mathbf{C}\} = \{([a, b], [x, y, z]) \mid ay - bx = 0\}.$$

La cohomologie de Y est engendrée par des classes ambiantes, de sorte que la recette ci-dessus permet théoriquement de calculer tous les invariants de Gromov-Witten de Y . Ici nous allons nous intéresser à Ψ^{2E} .

La classe image j_*E est la classe F d'une fibre $\mathbf{P}^1(\mathbf{C}) \times \{\text{pt}\}$ dans $\mathbf{P}^1(\mathbf{C}) \times \mathbf{P}^2(\mathbf{C})$. On doit donc considérer l'espace de modules $\overline{\mathcal{M}}_{0,m}(\mathbf{P}^1(\mathbf{C}) \times \mathbf{P}^2(\mathbf{C}), 2F)$.

Comme les applications de la classe $2F$ ont une projection constante sur $\mathbf{P}^2(\mathbf{C})$, on a

$$\overline{\mathcal{M}}_{0,m}(\mathbf{P}^1(\mathbf{C}) \times \mathbf{P}^2(\mathbf{C}), 2F) \cong \overline{\mathcal{M}}_{0,m}(\mathbf{P}^1(\mathbf{C}), 2L) \times \mathbf{P}^2(\mathbf{C}).$$

Pour la même raison, le fibré $\mathcal{V}$ que nous considérons est de la forme

$$\mathcal{V} \cong \mathcal{V}' \otimes \mathcal{O}_{\mathbf{P}^2(\mathbf{C})}(1) \longrightarrow \overline{\mathcal{M}}_{0,m}(\mathbf{P}^1(\mathbf{C}), 2L) \times \mathbf{P}^2(\mathbf{C})$$

où $\mathcal{V}' \rightarrow \overline{\mathcal{M}}_{0,m}(\mathbf{P}^1(\mathbf{C}), 2L)$ est un fibré vectoriel qui a pour fibre — au point de l'espace de modules représenté par l'application $u' : C \rightarrow \mathbf{P}^1(\mathbf{C})$ — l'espace $H^0(C, u'^*\mathcal{O}(1))$. Le rang du fibré vectoriel complexe $\mathcal{V}'$ est 3 et

$$e(\mathcal{V}) = c_3(\mathcal{V}') \otimes 1 + c_2(\mathcal{V}') \otimes t + c_1(\mathcal{V}') \otimes t^2 \in H^6(\overline{\mathcal{M}}_{0,m}(\mathbf{P}^1(\mathbf{C}) \times \mathbf{P}^2(\mathbf{C}), 2F); \mathbf{Z})$$

(ici $t = c_1(\mathcal{O}_{\mathbf{P}^2(\mathbf{C})}(1))$ est le générateur de $H^2(\mathbf{P}^2(\mathbf{C}); \mathbf{Z})$).

Remarquons pour finir que $\mathcal{V}'$ a deux sections partout indépendantes, de sorte que $c_3(\mathcal{V}') = 0$, $c_2(\mathcal{V}') = 0$ et $e(\mathcal{V}) = c_1(\mathcal{V}') \otimes t^2$. Pour le vérifier, choisissons deux éléments indépendants s_1 et s_2 dans l'espace vectoriel $H^0(\mathbf{P}^1(\mathbf{C}), \mathcal{O}_{\mathbf{P}^1(\mathbf{C})}(1))$, qui est de dimension 2. Définissons deux sections $\tilde{s}_1, \tilde{s}_2$ de $\mathcal{V}'$ par

$$\tilde{s}_i(C, \bar{z}, u) = s_i \circ u.$$

Si $\lambda \tilde{s}_1 + \mu \tilde{s}_2$ s'annule en un point de $\overline{\mathcal{M}}_{0,m}(\mathbf{P}^1(\mathbf{C}), 2L)$ représenté par $(C, \bar{z}, u)$,

$$(\lambda s_1 + \mu s_2) \circ u \equiv 0.$$

L'image de u doit donc être contenue dans l'ensemble des zéros de $\lambda s_1 + \mu s_2$, qui est un point de $\mathbf{P}^1(\mathbf{C})$. Donc u doit être constante, une contradiction puisqu'elle est de degré 2. Les sections $\tilde{s}_1$ et $\tilde{s}_2$ sont donc bien partout indépendantes.

On en déduit un résultat d'annulation pour les invariants Ψ^{2E} :

Proposition 4.11. — *Si $\alpha \in H^*((\mathbf{P}^1(\mathbf{C}) \times \mathbf{P}^2(\mathbf{C}))^m)$ est divisible par t , alors*

$$\Psi^{2E}((j^m)^* \alpha) = 0.$$

Démonstration. — L'hypothèse sur la classe α est que, si on la décompose en

$$\alpha = \sum_i \alpha_1^i \otimes \cdots \otimes \alpha_m^i,$$

alors, pour tout i , un des α_j^i est de la forme $a \otimes bt$ pour un certain $a \in H^*(\mathbf{P}^1(\mathbf{C}))$, $b \in H^*(\mathbf{P}^2(\mathbf{C}))$. On veut donc vérifier que

$$\Psi^{2E}((j^m)^*(\alpha_1 \otimes \cdots \otimes \alpha_{m-1} \otimes (a \otimes bt))) = 0$$

pour tous les $\alpha_i \in H^*(\mathbf{P}^1(\mathbf{C}) \times \mathbf{P}^2(\mathbf{C}))$, $a \in H^*(\mathbf{P}^1(\mathbf{C}))$, $b \in H^*(\mathbf{P}^2(\mathbf{C}))$. Cet invariant a été défini comme

$$\langle \text{év}_m^*(\alpha_1 \otimes \cdots \otimes \alpha_{m-1} \otimes (a \otimes bt)) \smile (c_1(\mathcal{V}') \otimes t^2), [\overline{\mathcal{M}}_{0,m}(\mathbf{P}^1(\mathbf{C}), 2L) \times \mathbf{P}^2(\mathbf{C})] \rangle.$$

L'application

$$\text{év}_m : \overline{\mathcal{M}}_{0,m}(\mathbf{P}^1(\mathbf{C}) \times \mathbf{P}^2(\mathbf{C}), 2F) \longrightarrow (\mathbf{P}^1(\mathbf{C}) \times \mathbf{P}^2(\mathbf{C}))^m \cong (\mathbf{P}^1(\mathbf{C}))^m \times (\mathbf{P}^2(\mathbf{C}))^m$$

coïncide avec

$$\text{év}_m' \times \Delta : \overline{\mathcal{M}}_{0,m}(\mathbf{P}^1(\mathbf{C}), 2L) \times \mathbf{P}^2(\mathbf{C}) \longrightarrow (\mathbf{P}^1(\mathbf{C}))^m \times (\mathbf{P}^2(\mathbf{C}))^m.$$

Ainsi notre invariant est-il une somme de termes de la forme

$$\langle [\text{év}_m'^*(a' \otimes a) \otimes (b' \smile bt)] \smile (c_1(\mathcal{V}') \otimes t^2), [\overline{\mathcal{M}}_{0,m}(\mathbf{P}^1(\mathbf{C}), 2L)] \otimes [\mathbf{P}^2(\mathbf{C})] \rangle$$

qui sont tous nuls puisque $t^3 = 0$ dans $H^*(\mathbf{P}^2(\mathbf{C}))$. $\square$

5. Autres applications, autres invariants

Dans cette partie, je vais être très rapide et imprécise, renvoyant les lecteurs aux références citées.

5.a. Invariants équivariants, symétrie miroir. — J'ai déjà signalé au § 4.b que les invariants de Gromov-Witten et le produit quantique qu'ils définissent permettent de munir l'espace vectoriel complexe $H^*(W, \mathbf{C})$ d'une structure de variété de Frobenius. Les espaces de déploiements de singularités fournissent d'autres exemples de variétés de Frobenius. Une façon élégante et impressionnante [28] d'exprimer la « conjecture des miroirs » (voir [79]) est de postuler une équivalence entre ces deux types de variétés de Frobenius. Ce programme est réalisé par Givental [29] qui démontre cette conjecture en utilisant des invariants de Gromov-Witten équivariants.

5.b. Homologie de Floer, conjecture d'Arnold. — La conjecture d'Arnold met une borne inférieure au nombre de points fixes de certains difféomorphismes symplectiques. On sait depuis longtemps que les courbes holomorphes peuvent jouer un rôle dans une démonstration de cette conjecture, c'est ainsi que Floer l'avait démontrée pour les variétés symplectiques monotones⁽²⁰⁾ [24], fixant les bases de l'homologie qui porte son nom (voir aussi [38]).

Toujours dans [44], Kontsevich a suggéré que les nouveaux espaces de modules pourraient aussi être utilisés pour démontrer la conjecture sur une variété symplectique générale. C'est ce que font Liu et Tian dans [51], Fukaya et Ono dans [25].

5.c. Lien avec Seiberg-Witten. — Pour les variétés de dimension 4, les invariants de Gromov-Witten sont liés à ceux de Seiberg-Witten (voir le § IV du texte de John Morgan [59] dans ce volume), comme l'a montré Taubes (voir l'annonce générale de ces beaux résultats dans [75]).

Une conséquence étonnante : ce sont des invariants du type de *difféomorphisme* de la variété, contrairement à ce qui se passe en dimension 6 (voir le § 2.e).

Une autre conséquence : l'unicité de la structure symplectique sur $\mathbf{P}^2(\mathbf{C})$. Précisément, si ω_0 et ω sont des formes symplectiques sur $\mathbf{P}^2(\mathbf{C})$, il existe un réel $\lambda \neq 0$ et un difféomorphisme φ tel que $\varphi^*\omega = \lambda\omega_0$.

– En s'appuyant sur le théorème d'existence de J -droites (ici le théorème 2.2), on montre que, si pour une structure presque complexe J calibrée par une forme symplectique ω sur $\mathbf{P}^2(\mathbf{C})$, on peut faire passer une unique droite par deux points distincts, alors $(\mathbf{P}^2(\mathbf{C}), \omega)$ est difféomorphe à $\mathbf{P}^2(\mathbf{C})$ avec un multiple de la forme symplectique standard.

– Ce résultat de Gromov dans [32] a attendu dix ans le résultat complémentaire de Taubes dans [76]. Celui-ci montre la non-nullité d'un invariant de Seiberg-Witten, et l'égalité de cet invariant avec un invariant de Gromov-Witten. Ce dernier n'est donc pas nul, mais un invariant de Gromov-Witten non nul, c'est un résultat global d'existence (ici d'unicité aussi) de courbe holomorphe, ici de J -droites passant par deux points. D'où l'unicité de la structure symplectique de $\mathbf{P}^2(\mathbf{C})$.

⁽²⁰⁾ Voir le § 3.e.

5.d. Nouveaux invariants en dimension 4. — Les travaux de Donaldson [19, 20] ont ouvert de nouvelles perspectives. Il s'agit d'utiliser un analogue symplectique des sections holomorphes d'un fibré en droites complexes, par exemple pour construire des sous-variétés symplectiques d'une sous-variété symplectique donnée. Parmi les retombées de ces idées, citons notamment les invariants d'Auroux et Katzarkov [8].

Ces derniers utilisent un théorème d'Auroux [9], lui-même fondé sur les idées de [19], qui affirme que toute variété symplectique de dimension 4 est un revêtement de $\mathbf{P}^2(\mathbf{C})$ ramifié le long d'une surface symplectique Σ . En regardant cette surface critique comme un revêtement ramifié de degré d de $\mathbf{P}^1(\mathbf{C})$, les auteurs construisent des invariants des variétés symplectiques de dimension 4, qui sont

- une factorisation dans le groupe des tresses à d brins de la tresse Δ^2 (celle où on fait tourner tous les brins en même temps)
- et une représentation de monodromie $\pi_1(\mathbf{P}^2(\mathbf{C}) - \Sigma) \rightarrow \mathfrak{S}_n$ (ici n est le degré du revêtement $W \rightarrow \mathbf{P}^2(\mathbf{C})$).

Ce sont des invariants complets du type de déformation de la variété symplectique... les pessimistes en déduiront sans doute qu'ils sont inutilisables.

5.e. Homologie de contact. — Si V est une variété de contact — l'analogue, en dimension impaire, d'une variété symplectique — on peut considérer les courbes pseudo-holomorphes de la variété « symplectifiée » $W \cong V \times \mathbf{R}$. Les espaces de modules correspondants sont les principaux personnages d'une nouvelle théorie fascinante en cours d'élaboration par Eliashberg avec Givental, Hofer... mais dont on ne trouve pour l'instant de trace publiée que dans le très bref [22]. Voir toutefois [23].

5.f. Capacités. — La largeur de Gromov dont il a été question aux §§ 1.d et 2.b fait partie d'une famille d'invariants pour les variétés symplectiques non compactes. En concurrence avec les courbes holomorphes, les systèmes hamiltoniens fournissent des invariants de ce type, les « capacités », voir les travaux, notamment, d'Ekeland, Floer, Hofer, Viterbo, Zehnder. Voir les survols dans [77, 39].

6. Appendice : les espaces projectifs et leur famille

Dans cet appendice, je décris « la » forme de Kähler de l'espace projectif $\mathbf{P}^n(\mathbf{C})$ utilisée tout au long de cet article. Elle provient, bien sûr, de la forme de Kähler de $\mathbf{C}^{n+1}$.

La forme de Kähler de $\mathbf{C}^{n+1}$. — C'est simplement la forme symplectique standard $\omega = \sum dx_j \wedge dy_j$. Comme c'est aussi, au signe près, la partie imaginaire de la structure hermitienne canonique de $\mathbf{C}^{n+1}$, il est bien clair que c'est, en effet, une forme de Kähler (voir le § 1.c).

La forme de Kähler sur $\mathbf{P}^n(\mathbf{C})$. — On décrit la forme de Kähler de l'espace projectif en utilisant la définition de ce dernier comme quotient ; c'est un cas particulier du célèbre procédé de « réduction symplectique ». On considère donc l'espace projectif $\mathbf{P}^n(\mathbf{C})$ comme ce qu'il est vraiment, c'est-à-dire comme le quotient

$$\mathbf{P}^n(\mathbf{C}) = \mathbf{C}^{n+1} - \{0\} / z \sim \lambda z \text{ pour } \lambda \in \mathbf{C}^*,$$

ou encore, puisque toutes les droites complexes contiennent un cercle de vecteurs unitaires, comme le quotient de la sphère unité $S^{2n+1}/z \sim \lambda z$ pour $|\lambda| = 1$. On regarde l'inclusion et la projection :

$$\begin{array}{c} j : S^{2n+1} \subset \mathbf{C}^{n+1} \\ \downarrow p \\ \mathbf{P}^n(\mathbf{C}). \end{array}$$

La restriction $j^*\omega$ de la forme symplectique de $\mathbf{C}^{n+1}$ à la sphère a un noyau. Au point $z \in S^{2n+1}$, c'est simplement la droite réelle engendrée par iz . Le quotient tue précisément cette droite, de sorte qu'il y a une unique forme *symplectique*, disons σ , sur $\mathbf{P}^n(\mathbf{C})$ telle que

$$j^*\omega = p^*\sigma.$$

Sous-variétés. — Les variétés kählériennes sont nombreuses : les produits de variétés kählériennes, les sous-variétés complexes d'une variété kählérienne sont kählériennes, en particulier toutes les variétés projectives complexes sont kählériennes.

C'est le cas en particulier de l'espace $\tilde{\mathbf{P}}^2(\mathbf{C})$ obtenu en éclatant un point dans $\mathbf{P}^2(\mathbf{C})$, tout simplement parce que c'est une sous-variété complexe d'un produit d'espaces projectifs, comme on le voit en écrivant

$$\tilde{\mathbf{P}}^2(\mathbf{C}) = \{(\ell, d) \in \mathbf{P}^1(\mathbf{C}) \times \mathbf{P}^2(\mathbf{C}) \mid d \subset \ell \oplus \mathbf{C}\} = \{([a, b], [x, y, z]) \mid ay - bx = 0\}.$$

La projection sur le deuxième facteur est l'éclatement, le diviseur exceptionnel E est la fibre de $[0, 0, 1]$,

$$E = \{([a, b], [0, 0, 1]) \in \mathbf{P}^1(\mathbf{C}) \times \mathbf{P}^2(\mathbf{C})\}.$$

Le fibré normal de $\tilde{\mathbf{P}}^2(\mathbf{C})$ est le fibré en droites $\mathcal{O}(1) \otimes \mathcal{O}(1)$. En omettant de la notation les projections de $\mathbf{P}^1(\mathbf{C}) \times \mathbf{P}^2(\mathbf{C})$ sur les deux facteurs, on a donc

$$T\tilde{\mathbf{P}}^2(\mathbf{C}) = \oplus(\mathcal{O}(1) \otimes \mathcal{O}(1)) = T\mathbf{P}^1(\mathbf{C}) \oplus T\mathbf{P}^2(\mathbf{C}).$$

En appelant u, v , les images des générateurs des H^2 des facteurs dans $H^2(\tilde{\mathbf{P}}^2(\mathbf{C}))$, on voit que

$$\begin{aligned} -c_1(T\tilde{\mathbf{P}}^2(\mathbf{C})) &= u + 2v \\ -\langle c_1(T\tilde{\mathbf{P}}^2(\mathbf{C})), E \rangle &= \langle u + 2v, E \rangle = \langle u, [\mathbf{P}^1(\mathbf{C})] \rangle = 1, \end{aligned}$$

propriété qui a été utilisée plusieurs fois dans cet article.

Références

- [1] V. I. ARNOLD & A. B. GIVENTAL – « Symplectic geometry », *Dynamical systems, Encyclopædia of Math. Sci., Springer* (1985).
- [2] M. AUDIN – « Symplectic and almost complex manifolds », in [7], p. 41–74.
- [3] ———, *The topology of torus actions on symplectic manifolds*, Progress in Math., Birkhäuser, 1991.
- [4] ———, « Cohomologie quantique », *Séminaire Bourbaki, 1995-96, Astérisque* **241** (1997), p. 29–58.
- [5] ———, « An introduction to Frobenius manifolds, moduli spaces of stable maps and quantum cohomology », *preprint non publié* (1997).
- [6] M. AUDIN & P. IGLÉSIAS – « Optique, géométrie et topologie symplectique », *Images des mathématiques*, Supplément au *Courrier du C.N.R.S.* (1995).
- [7] M. AUDIN & J. LAFONTAINE (éds.) – *Holomorphic curves in symplectic geometry*, Progress in Math., Birkhäuser, 1994.
- [8] D. AUROUX & L. KATZARKOV – « Branched Coverings of $\mathbf{CP}^2$ and Invariants of Symplectic 4-manifolds », *Invent. Math.* **142** (2000), p. 631–673.
- [9] D. AUROUX – « Symplectic 4-manifolds as branched coverings of $\mathbf{CP}^2$ », *Invent. Math.* **139** (2000), p. 551–602.
- [10] J.-F. BARRAUD – « Courbes pseudo-holomorphes équisingulières en dimension 4 », *thèse, Université de Toulouse* (1998).
- [11] K. BEHREND & B. FANTECHI – « The intrinsic normal cone », *Invent. Math.* **128** (1997), p. 45–88.
- [12] D. BENNEQUIN – « Problèmes elliptiques, surfaces de Riemann et structures symplectiques », *Séminaire Bourbaki, Astérisque* **145-146** (1987), p. 111–136.
- [13] X. BUFF, J. FEHRENBACH, P. LOCHAK, L. SCHNEPS & P. VOGEL – *Espaces de modules des courbes, groupes modulaires et théorie des champs*, Panoramas et Synthèses, 7, Société Mathématique de France, 1999.
- [14] H. CLEMENS, J. KOLLÁR & S. MORI – « Higher-dimensional complex geometry », *Astérisque* **166** (1988).
- [15] M. CORNALBA – « A simple proof of the projectivity of Kontsevich’s space of maps », *Istit. Lombardo Accad. Sci. Lett. Rend. A* **129** (1995), p. 111–119.
- [16] P. DI FRANCESCO & C. ITZYKSON – « Quantum intersection rings », in [17].
- [17] R. DIJKGRAAF, C. FABER & G. VAN DER GEER (éds.) – *The moduli space of curves*, Progress in Math. 129, Birkhäuser, 1995.
- [18] S. DONALDSON – « The orientation of Yang-Mills moduli spaces and 4-manifold topology », *J. Differential Geom.* **26** (1987), p. 397–428.
- [19] ———, « Symplectic submanifolds and almost-complex geometry », *J. Differential Geom.* **44** (1996), no. 4, p. 666–705.
- [20] S. K. DONALDSON – « Lefschetz pencils on symplectic manifolds », *J. Differential Geom.* **53** (1999), no. 2, p. 205–236.
- [21] B. DUBROVIN – « Geometry of 2D topological field theories », *Lecture Notes in Math.* **1620** (1995).
- [22] Y. ELIASHBERG – « Invariants in contact topology », *Proceedings I.C.M., Vol. II (Berlin, 1998)*, Documenta Mathematica 1998, p. 327–338.
- [23] Y. ELIASHBERG, A. GIVENTAL & H. HOFER – « Introduction to Symplectic Field Theory », *preprint, math.SG/0010059* (2000).

- [24] A. FLOER – « Symplectic fixed points and holomorphic spheres », *Comm. Math. Phys.* **120** (1989), p. 575–611.
- [25] K. FUKAYA & K. ONO – « Arnold conjecture and Gromov-Witten invariant », *Topology* **38** (1999), p. 933–1048.
- [26] W. FULTON & R. PANDHARIPANDE – « Notes on stable maps and quantum cohomology », Algebraic geometry—Santa Cruz 1995, Amer. Math. Soc., Providence, RI, 1997, p. 45–96.
- [27] P. GAUDUCHON – « Connexions linéaires, classes de Chern, théorème de Riemann-Roch », in [7].
- [28] A. B. GIVENTAL – « Homological geometry and mirror symmetry », *I.C.M. Zürich (1994)*, Birkhäuser (1995), p. 473–480.
- [29] ———, « Equivariant Gromov-Witten invariants », *Internat. Math. Res. Notices* **13** (1996), p. 613–663.
- [30] T. GRABER & R. PANDHARIPANDE – « Localization of virtual classes », *preprint* (1997).
- [31] P. A. GRIFFITHS & J. HARRIS – *Principles of algebraic geometry*, Wiley, 1978.
- [32] M. GROMOV – « Pseudo-holomorphic curves in symplectic manifolds », *Invent. Math.* **82** (1985), p. 307–347.
- [33] ———, « Soft and hard symplectic geometry », *Proceedings of the International Congress of Mathematicians, Vol. 1, 2 (Berkeley, 1986)* (Providence, RI), Amer. Math. Soc., 1987, p. 81–98.
- [34] M. GROSS – « The deformation space of Calabi-Yau n -folds with canonical singularities can be obstructed », *Mirror symmetry, II*, Amer. Math. Soc., Providence, RI, 1997, p. 401–411.
- [35] R. HAMILTON – « The inverse function theorem of Nash and Moser », *Bull. Amer. Math. Soc.* **7** (1982), p. 65–222.
- [36] J. HARRIS & I. MORRISON – *Moduli of curves*, Springer, New York, 1998.
- [37] N. HITCHIN – « Frobenius manifolds (notes by D. Calderbank) », *Gauge theory and symplectic geometry, Montréal 1995* (J. Hurtubise & F. Lalonde, eds.), NATO ASI Series, Kluwer Academic Publishers, 1997, p. 69–112.
- [38] H. HOFER & D. SALAMON – « Floer homology and Novikov rings », in [40].
- [39] H. HOFER & E. ZEHNDER – « Symplectic invariants and Hamiltonian dynamics », in [40].
- [40] H. HOFER, C. H. TAUBES, A. WEINSTEIN & E. ZEHNDER (eds.) – *The Floer memorial volume*, Birkhäuser Verlag, Basel, 1995.
- [41] S. IVASHKOVICH & V. SHEVCHISHIN – « Structure of the moduli space in a neighborhood of a cusp-curve and meromorphic hulls », *Invent. Math.* **136** (1999), p. 571–602.
- [42] S. KEEL – « Intersection theory of moduli space of stable n -pointed curves of genus 0 », *Trans. Amer. Math. Soc.* **330** (1992), p. 545–574.
- [43] F. KNUDSEN – « Projectivity of the moduli space of stable curves II », *Math. Scand.* **52** (1983), p. 1225–1265.
- [44] M. KONTSEVICH – « Enumeration of rational curves via torus actions », in [17].
- [45] M. KONTSEVICH & Y. MANIN – « Gromov-Witten classes, quantum cohomology and enumerative geometry », *Commun. Math. Phys.* **164** (1994), p. 525–562.
- [46] J. LAFONTAINE – « Some relevant Riemannian geometry », in [7].
- [47] J. LAFONTAINE & M. AUDIN – « Applications of pseudo-holomorphic curves to symplectic topology », in [7], p. 1–14.

- [48] F. LALONDE & D. MCDUFF – « The classification of ruled symplectic 4-manifolds », *Math. Res. Letters* **3** (1996), p. 769–778.
- [49] J. LI & G. TIAN – « Virtual moduli cycles and Gromov-Witten invariants of algebraic varieties », *J. Amer. Math. Soc.* **11** (1998), p. 119–174.
- [50] P. LIBERMANN & C. M. MARLE – *Symplectic geometry and analytic mechanics*, Reidel, Boston, 1987.
- [51] G. LIU & G. TIAN – « Floer homology and Arnold conjecture », *J. Differential Geom.* **49** (1998), no. 1, p. 1–74.
- [52] V. LIZAN – Thèse, *Université de Toulouse* (1996).
- [53] Y. I. MANIN – *Frobenius manifolds, quantum cohomology, and moduli spaces*, Colloquium Publ., vol. 47, Amer. Math. Soc., 1999.
- [54] D. MCDUFF – « Examples of symplectic structures », *Invent. Math.* **89** (1987), p. 1–36.
- [55] ———, « The structure of rational and ruled symplectic 4-manifolds », *Journ. Amer. Math. Soc.* **3** (1990), p. 679–712.
- [56] ———, « The local behaviour of holomorphic curves in almost complex 4-manifolds », *Journal Diff. Geom.* **34** (1991), p. 143–164.
- [57] D. MCDUFF & D. SALAMON – *J-holomorphic curves and quantum cohomology*, Amer. Math. Soc., 1994, 2nd printing, 1995.
- [58] ———, *Introduction to symplectic topology*, The Clarendon Press Oxford University Press, New York, 1995, Oxford Science Publications.
- [59] J. MORGAN – « Seiberg-Witten invariants », *ce volume*.
- [60] J. MOSER – « On the volume elements on a manifold », *Trans. Amer. Math. Soc.* **120** (1965), p. 286–294.
- [61] P. PANSU – « Compactness », in [7].
- [62] T. H. PARKER & J. G. WOLFSON – « Pseudo-holomorphic maps and bubble trees », *J. Geom. Anal.* **3** (1993), p. 63–98.
- [63] A. PRESSLEY & G. SEGAL – *Loop groups*, The Clarendon Press Oxford University Press, New York, 1986, Oxford Science Publications.
- [64] D. QUILLEN – « Elementary proofs of some results of cobordism theory using Steenrod operations », *Advances in Math.* **7** (1971), p. 29–56.
- [65] Y. RUAN – « Symplectic topology on algebraic 3-folds », *Journal Diff. Geom.* **39** (1994), p. 215–227.
- [66] ———, « Topological sigma model and Donaldson type invariants in Gromov theory », *Duke Math. J.* **83** (1996), p. 461–500.
- [67] Y. RUAN & G. TIAN – « A mathematical theory of quantum cohomology », *Journal Diff. Geom.* **42** (1995), p. 259–367.
- [68] C. SABBABH – « Frobenius manifolds : isomonodromic deformations and infinitesimal period mappings », *Expo. Math.* **16** (1998), p. 1–57.
- [69] ———, *Déformations isomonodromiques et variétés de Frobenius*, à paraître, 2001.
- [70] J. SARAMAGO – *Memorial do Convento*⁽²¹⁾, Caminho, Lisboa, 1982.
- [71] B. SIEBERT – « Symplectic Gromov-Witten invariants », *New trends in algebraic geometry* (Warwick, 1996), Cambridge Univ. Press, 1999, p. 375–424.
- [72] ———, « Algebraic and symplectic Gromov-Witten invariants coincide », *Ann. Inst. Fourier (Grenoble)* **49** (1999), p. 1743–1795.

⁽²¹⁾Traduit en français sous le titre « Le Dieu manchot », éditions du Seuil.

- [73] J.-C. SIKORAV – « Some properties of holomorphic curves in almost complex manifolds », in [7].
- [74] H. SPIELBERG – « A formula for the Gromov-Witten invariants of toric varieties », *Thèse, Strasbourg* (1999).
- [75] C. H. TAUBES – « The Seiberg-Witten and Gromov invariants », *Math. Res. Lett.* **2** (1995), p. 221–238.
- [76] ———, « $sw \rightarrow gr$: from the Seiberg-Witten equations to pseudo-holomorphic curves », *J. Amer. Math. Soc.* **9** (1996), p. 845–918.
- [77] C. VITERBO – « Capacités symplectiques et applications [d'après Ekeland-Hofer, Gromov] », *Séminaire Bourbaki, 1988/89, Astérisque* (1989), no. 177-178.
- [78] ———, « Une introduction à la topologie symplectique », *Gazette des Mathématiciens* **54** (1992), p. 81–96.
- [79] C. VOISIN – *Symétrie miroir*, Panoramas et Synthèses, 2, Société Mathématique de France, 1996.
- [80] C. T. C. WALL – « Classification problems in differential topology of 6-manifolds », *Invent. Math.* **1** (1966), p. 335–374.
- [81] P. WILSON – « The Kähler cone on Calabi-Yau threefolds », *Invent. Math.* **28** (1988), p. 671–694.
- [82] E. WITTEN – « Two-dimensional gravity and intersection theory on moduli space », *Surveys in differential geometry* (Cambridge, MA, 1990), Lehigh Univ., Bethlehem, PA, 1991, p. 243–310.
- [83] W. T. WU – « Sur les classes caractéristiques des structures fibrées sphériques », *Actualités scientifiques et industrielles, Hermann* (1952).
- [84] R. YE – « Gromov's compactness theorem for pseudo holomorphic curves », *Trans. Amer. Math. Soc.* **342** (1994), p. 671–694.

SEIBERG-WITTEN INVARIANTS

by

John W. Morgan

Abstract. — Recent years have seen many new invariants of manifolds, invariants defined using moduli spaces of solutions to certain partial differential equations. These include the anti-self-dual equations and Donaldson invariants, the pseudoholomorphic curve equations of Gromov and the resulting Gromov and Gromov-Witten invariants, quantum cohomology, various Chern-Simon's invariants, Floer homology, and, the subject of this article, the Seiberg-Witten invariants. We cover the background necessary to define these invariants, give the definition, and then several applications to the topology of smooth four-manifolds including to symplectic four-manifolds and complex algebraic surfaces.

Résumé (Les invariants de Seiberg-Witten). — Ces dernières années ont vu l'introduction de plusieurs invariants des variétés, invariants définis à l'aide des espaces de modules de certaines équations aux dérivées partielles. Il y a parmi ceux-ci les équations d'anti-autodualité et les invariants de Donaldson, les équations des courbes pseudoholomorphes et les invariants de Gromov-Witten, et les équations de Seiberg-Witten et les invariants associés. Ces derniers sont le sujet de cet article. Nous commençons par développer les résultats fondamentaux qui sont nécessaires, puis nous définissons les invariants. Finalement, nous donnerons quelques applications aux variétés symplectiques et kählériennes.

The recent years have seen many new invariants introduced into geometry and topology. These invariants are defined using moduli spaces of solutions to certain elliptic partial differential equations (PDE). Under suitable genericity assumptions these moduli spaces will be finite dimensional manifolds of dimension given by the index of the elliptic operators which are the linearizations of these equations. One also needs a compactness or weak compactness property to make sure that, as one varies the parameters in the problem, the solutions cannot escape off to infinity. If this happens then one would not have an invariant. When this compactness property holds is not at all understood in general. Rather we now have several different and interesting

2000 Mathematics Subject Classification. — 57R57, 58J.

Key words and phrases. — Seiberg-Witten invariants, four-manifolds.

examples where it does, consequently where invariants result. These include the anti-self-dual (ASD) equations and Donaldson invariants, the pseudo-holomorphic curve equations of Gromov and the resulting Gromov and Gromov-Witten invariants, quantum cohomology, various Chern-Simons invariants, Floer homology, and, the subject of these lectures, the Seiberg-Witten invariants. The Seiberg-Witten equations exhibit the strongest form of compactness.

Amazingly enough, in all these contexts the invariants admit a quantum field theoretic definition – they are interpreted by the physicists as correlation coefficients in a certain renormalizable (and usually asymptotically free) field theory. This non-rigorous physics definition carries with it certain baggage – there is no mathematical rigorous foundations for quantum field theory. Nevertheless, new operators, equations and mathematical objects as well as relationships between them have been discovered through this mechanism. While these discoveries from quantum field theory (and also string theory) are not mathematical theorems, to date, they have always been mathematical truths. Many have been established rigorously using often very different techniques. The main interest is when these insights from physics are new and non-obvious from the mathematical point of view. A prime example of this is the Seiberg-Witten invariants (firmly established mathematically) and their relationship to the Donaldson invariants [2] (which is still conjectural).

Because the quantum field theories used to date have all been either asymptotically or infrared free and the correlation functions of interest topological in nature, these invariants can be computed using the moduli space of classical solutions to the equations of motion – from the point of view of physics this is where the moduli spaces of solutions to non-linear PDE's enters the picture. One can certainly imagine situations in which the quantum field theories do not limit in a nice way to the space of classical solutions of the equations of motion. In these cases, if they exist, it will be interesting to understand the mathematical formulation of the resulting quantum field theory invariants.

This article is an introduction to and a brief survey of some of the latest results of the subject of Seiberg-Witten invariants of smooth four-manifolds. The first section is a review of the necessary linear algebra and differential geometry (Clifford algebras, Spin and Spin^c -structures and Spin modules, and the Dirac operator). An excellent reference for this basic material is [14]. The second section introduces the Seiberg-Witten equations for a Spin^c -structure and shows how, beginning with the moduli space of solutions to these equations, to define a numerical invariant, the Seiberg-Witten (SW) invariant of the Spin^c -structure. Some of the very first, and still some of the nicest, applications of SW invariants are to Kähler manifolds. In the third section we identify the moduli space of solutions to the SW equations over a Kähler manifold with a purely holomorphic object. From this it is easy to explicitly compute the invariants for all algebraic surfaces. We give these results and deduce some topological

consequences such as the differentiable invariance of the Kodaira dimension and of the canonical class of the minimal model (up to sign). In the fourth section we turn to Taubes' generalization of the results in the Kähler case to the symplectic case where he identifies the SW invariants with the Gromov invariants counting the number of pseudo-holomorphic curves in a given homology class. Following Donaldson, we sketch a proof of Taubes' result in the simplest case. We also derive various topological corollaries of Taubes' result, such as the fact that $\mathbf{CP}^2$ has a unique symplectic structure up to symplectomorphism. Finally, in the last section we give a survey of other important results. We discuss the generalization of Taubes' results to symplectic manifolds with boundary due to Kronheimer-Mrowka and their finiteness result for symplectically fillable contact structures. We briefly discuss the Floer version of this theory for three-manifolds and then various gluing theorems that result, including the connected sum theorem and the blow-up formula for the SW invariants. We also state the result that a symplectic curve is genus minimizing in its homology class and indicate a proof in some cases. We end with the result of Fintushel-Stern constructing lots of four-manifolds homotopy equivalent to the $K3$ surface.

Many amazing results have been established for four-manifolds using these invariants and the related Donaldson polynomial invariants. General references for the state of the art in these subjects are [4] and [10]. For a general review of the subject of the topology of four-dimensional manifolds (and three-manifolds) the book [6] is a good reference. For results relating the Seiberg-Witten invariants to three-dimensional manifolds there is [12]. For results relating these invariants to symplectic structures and contact structures there are [15], [16], [17] and above all Taubes' revolutionary work in [25], [26] and [27]. The review in [23] examines the connections of these invariants to complex geometry.

In spite of all this amazing progress, many important questions remain to be answered, including a proof of the conjectured relation of these two sets of invariants both for four-manifolds and a better understanding of the relationship for three-manifolds. The topological significance of the three-manifold SW invariants is still a mystery. Whether or not there are other related, but more powerful, invariants of four-manifolds is also a wide open question. I hope that this introduction to the subject will help spread a sense of the beauty and depth of the subject and inspire others to tackle some of these questions.

I. The Linear Algebra and Differential Geometry underlying the Seiberg-Witten Invariants

I.1. Clifford Algebras. — The three-sphere, viewed as the group of quaternions of unit norm, acts by quaternion conjugation on the vector space of purely imaginary quaternions $\text{Im}(\mathbf{H})$ preserving the usual norm on this three-dimensional space. This

action thus determines a representation $S^3 \rightarrow \mathrm{SO}(3)$, realizing the spin double covering of $\mathrm{SO}(3)$. This implies easily that $\pi_1(\mathrm{SO}(n)) = \mathbf{Z}/2\mathbf{Z}$ for every $n \geq 3$; hence all these groups have connected double coverings. The double covering group of $\mathrm{SO}(n)$ is denoted $\mathrm{Spin}(n)$ and these double covering groups are called the *spin groups*. They cannot be realized as units in an $\mathbf{R}$ -division algebra since the only $\mathbf{R}$ -division algebras of finite dimension are $\mathbf{R}$, $\mathbf{C}$, and $\mathbf{H}$.

There is however a series of $\mathbf{R}$ -algebras, called Clifford algebras, the first three of which are $\mathbf{R}$, $\mathbf{C}$, and $\mathbf{H}$. One can obtain the spin double covers of $\mathrm{SO}(n)$, for every n , via a conjugation action of the subgroup of the units of the Clifford algebras on a certain vector subspace of the algebra.

I.1.1. Definition of the Clifford Algebras. — Let (V, Q) be a pair consisting of a finite dimensional real vector space V and a non-degenerate inner product Q on it. We define the Clifford algebra $\mathrm{Cl}(V, Q)$ as follows. It is a quotient algebra of the tensor algebra $T(V)$ determined by the relations $v \otimes v = -Q(v) \cdot 1$ for every $v \in V$. (Here 1 is the unit in $T(V)$.) Since the relations are homogeneous modulo two with respect to the grading of the tensor algebra, the $\mathbf{Z}$ grading on $T(V)$ descends to a decomposition $\mathrm{Cl}(V, Q) = \mathrm{Cl}_0(V, Q) \oplus \mathrm{Cl}_1(V, Q)$ with $\mathrm{Cl}_0(V, Q)$ being a subalgebra. It is easy to see that the subspace $V \subset T(V)$ descends to an embedding $V \subset \mathrm{Cl}(V)$ whose image generates the Clifford algebra as an algebra. Let $\{e_1, \dots, e_n\}$ be a basis for V with $Q(e_i) = \pm 1$ for all i and with $Q(e_i + e_j) = Q(e_i) + Q(e_j)$ for all $i \neq j$, or equivalently, $\langle e_i, e_j \rangle_Q = 0$ for $i \neq j$ where $\langle \cdot, \cdot \rangle_Q$ is the bilinear inner product associated with Q . It is also easy to see that the products $e_{i_1} \cdots e_{i_k}$ for $i_1 < \cdots < i_k$ generate $\mathrm{Cl}(V, Q)$ as a vector space and hence that the dimension of $\mathrm{Cl}(V, Q)$ over $\mathbf{R}$ is at most $2^{\dim V}$.

Examples. — Let $\mathbf{R}^n$ denote n -dimensional Euclidean space with its standard quadratic form. Of course, $\mathrm{Cl}(\mathbf{R}^0) = \mathbf{R}$. The algebra $\mathrm{Cl}(\mathbf{R}^1)$ is presented as

$$\mathbf{R}[x]/\{x^2 = -1\},$$

and hence is isomorphic as an algebra to $\mathbf{C}$. Thus, $\mathrm{Cl}(\mathbf{R}^1) \otimes_{\mathbf{R}} \mathbf{C}$ is isomorphic to $\mathbf{C} \oplus \mathbf{C}$. In the same manner,

$$\mathrm{Cl}(\mathbf{R}^2) = \mathbf{R}[x, y]/\{x^2 = y^2 = -1; xy = -yx\},$$

and hence is isomorphic to $\mathbf{H}$. Consequently, $\mathrm{Cl}(\mathbf{R}^2) \otimes_{\mathbf{R}} \mathbf{C}$ is isomorphic to $\mathbf{C}[2]$ the matrix algebra of complex 2×2 -matrices.

I.1.2. Their Structure. — The structure of these algebras is known completely and exhibits a mod 8 periodicity closely linked with Bott periodicity in real K -theory (see [14] for more details). We shall be concerned only with complex representations of these algebras and hence shall need only the structure of the complexified Clifford algebras. These are much easier to describe for they exhibit a mod 2 periodicity related to the periodicity of complex K -theory.

Lemma I.1.1. — *Let (V, Q) be a pair consisting of a finite dimensional real vector space and a positive definite quadratic form. Let n be the dimension of V . If $n = 2k$, then $\text{Cl}(V, Q) \otimes_{\mathbf{R}} \mathbf{C}$ is isomorphic to the algebra $\mathbf{C}[2^k]$ of complex matrices of size $2^k \times 2^k$. If $n = 2k + 1$, then $\text{Cl}(V, Q)$ is isomorphic to $\mathbf{C}[2^k] \oplus \mathbf{C}[2^k]$.*

Proof. — Consider a pair (W, Q) consisting of a finite dimensional vector space and a non-degenerate quadratic form and let $(W', Q') = (W, Q) \oplus \mathbf{R}^2$ (direct sum of quadratic spaces). Let $\{e_1, e_2\}$ be the standard orthonormal basis of $\mathbf{R}^2$. We define a map

$$\varphi: W \oplus \mathbf{R}^2 \longrightarrow (\text{Cl}(W, Q) \otimes \text{Cl}(\mathbf{R}^2)) \otimes_{\mathbf{R}} \mathbf{C}$$

by sending

$$\begin{aligned} w &\longmapsto i(w \otimes e_1 e_2); \text{ for all } w \in W, \\ e_i &\longmapsto 1 \otimes e_i; \text{ for } i = 1, 2. \end{aligned}$$

It is straightforward to see that for all $x \in W'$ we have $\varphi(x) \cdot \varphi(x)$ is the element $-Q'(x) \cdot 1$. By the defining property of the Clifford algebra, this implies that φ extends to an algebra homomorphism of $\text{Cl}(W', Q') \otimes_{\mathbf{R}} \mathbf{C}$. Clearly, this map is onto since it hits the $\mathbf{C}$ -algebra generating subspace $W' \subset \text{Cl}(W', Q')$. Given the upper bound on the dimension and the computations of $\text{Cl}(\mathbf{R}^1)$ and $\text{Cl}(\mathbf{R}^2)$ given above, one easily shows by induction that φ is an isomorphism. Since $\mathbf{C}[a] \otimes \mathbf{C}[b]$ is isomorphic to $\mathbf{C}[ab]$ the result follows by induction from the computations for $\mathbf{R}^1$ and $\mathbf{R}^2$. $\square$

Let (V, Q) be an oriented positive definite quadratic space over $\mathbf{R}$ and let $e_1, \dots, e_n$ be an orthonormal basis which is compatible with the orientation. We define the complex unit $\omega_{\mathbf{C}} \in \text{Cl}(V, Q) \otimes_{\mathbf{R}} \mathbf{C}$ by

$$\omega_{\mathbf{C}} = i^{\lfloor \frac{n+1}{2} \rfloor} e_1 \cdots e_n.$$

An easy computation shows that $\omega_{\mathbf{C}}$ depends only on the orientation and not the choice of orthonormal basis, and that $\omega_{\mathbf{C}}^2 = 1$. We let $\text{Cl}(V, Q)^{\pm}$ be the ± 1 eigenspaces for left multiplication by $\omega_{\mathbf{C}}$. When n is odd $\omega_{\mathbf{C}}$ is in the center of $\text{Cl}(V, Q)$ and hence its eigenspaces $(\text{Cl}(V, Q) \otimes_{\mathbf{R}} \mathbf{C})^{\pm}$ are subalgebras with the property that $x^+ \cdot x^- = 0$ for all $x^{\pm} \in \text{Cl}(V, Q)^{\pm}$. This, in fact, is a decomposition of the real algebra underlying the complex decomposition given in the last lemma in the case n is odd. Notice that $\text{Cl}_0(V, Q) \otimes_{\mathbf{R}} \mathbf{C}$ sits diagonally with respect to this decomposition in the sense that its projection to either factor is an isomorphism. This proves:

Lemma I.1.2. — *Let (V, Q) be a pair consisting of an oriented real vector space of dimension $2k + 1$ and a positive definite quadratic form. Then $\text{Cl}_0(V, Q) \otimes_{\mathbf{R}} \mathbf{C}$ is isomorphic to $\mathbf{C}[2^k]$. The algebra $\text{Cl}_0(V, Q)$ has, up to isomorphism, a unique irreducible, finite dimensional complex representation, denoted $S(V)$. The dimension of this representation is 2^k . Furthermore, $\text{Cl}(V, Q)$ has two non-isomorphic finite dimensional complex representations $S^{\pm}(V)$ each of which is of dimension 2^k coming*

from the irreducible representations of $\text{Cl}(V, Q)^\pm$. The restriction of either of these two representations to $\text{Cl}_0(V, Q)$ is irreducible and isomorphic to $S(V)$.

Proof. — All this is a consequence of the structure results above for the complexified Clifford algebras and Wedderburn's theorem, see Chapt. VII of [9]. $\square$

In the case when $n = 2k$ the results are slightly different. The following is straightforward to establish by induction.

Lemma I.1.3. — *Let (V, Q) be a pair consisting of an oriented finite dimensional real vector space and a positive definite quadratic form. Suppose that the dimension of V is even. There is a decomposition:*

$$\text{Cl}(V, Q) = \begin{pmatrix} \text{Cl}_0(V, Q)^+ & \text{Cl}_1(V, Q)^+ \\ \text{Cl}_1(V, Q)^- & \text{Cl}_0(V, Q)^- \end{pmatrix}.$$

In particular, $\text{Cl}_0(V, Q)^+ \otimes_{\mathbf{R}} \mathbf{C}$ is a matrix algebra over $\mathbf{C}$ of size 2^{k-1} . The unique (up to isomorphism) finite dimensional complex representation $S(V)$ of $\text{Cl}(V, Q)$ is of dimension 2^k . It decomposes over $\text{Cl}_0(V)$ into irreducible representations $S^\pm(V)$, each of dimension 2^{k-1} . Here $S^\pm(V)$ is the ± 1 -eigenspace for the action of $\omega_{\mathbf{C}}$. Reversing the orientation of V interchanges $S^\pm(V)$. Furthermore, Clifford multiplication identifies $(\text{Cl}_0(V, Q) \otimes_{\mathbf{R}} \mathbf{C})^+$ with the endomorphisms of $S^+(V)$ whereas Clifford multiplication by $(\text{Cl}_0(V, Q) \otimes_{\mathbf{R}} \mathbf{C})^-$ is trivial on $S^+(V)$.

In all cases these irreducible representations are called the *complex Clifford modules*.

Examples. — The dimensions that concern us most are $n = 3, 4$. We have $\text{Cl}(\mathbf{R}^3) = \mathbf{H} \oplus \mathbf{H}$ and the two complex Clifford modules $S^\pm(\mathbf{R}^3)$ are copies of $\mathbf{H}$ acted on by left multiplication by the $\pm$ -factor of $\text{Cl}(\mathbf{R}^3)$. Similarly, $\text{Cl}(\mathbf{R}^4)$ is isomorphic to $\mathbf{H}[2]$ and $\text{Cl}_0(\mathbf{R}^4)$ is the subalgebra of diagonal matrices, and hence is isomorphic to $\mathbf{H} \oplus \mathbf{H}$. Thus, the complex Clifford module $S(\mathbf{R}^4)$ is isomorphic to $\mathbf{H}^2$ and decomposes under $\text{Cl}_0(\mathbf{R}^4)$ into $\mathbf{H} \oplus \mathbf{H}$ with the action of $\text{Cl}_0(\mathbf{R}^4)$ on each factor being left quaternion multiplication. In all these examples the complex module structure is given by right multiplication by $\mathbf{C} \subset \mathbf{H}$.

I.1.3. Relation of the Clifford Algebra and the Exterior Algebra. — Let V be a finite dimensional real vector space and $\langle, \rangle$ a non-degenerate inner product on V with associated quadratic form Q . We have seen that $\text{Cl}(V, Q)$ and $\wedge^*(V)$ have the same dimension (namely, $2^{\dim V}$). In fact, the Clifford algebra can be identified with the exterior algebra (as vector spaces) in such a way that the multiplication in the Clifford algebra is a lower order perturbation of the usual exterior algebra multiplication. We define an action of V on $\wedge^*(V)$ by:

$$v \cdot (v_1 \wedge \cdots \wedge v_k) = v \wedge v_1 \wedge \cdots \wedge v_k + \sum_{i=1}^k (-1)^i \langle v, v_i \rangle_Q v_1 \wedge \cdots \wedge v_{i-1} \wedge v_{i+1} \wedge \cdots \wedge v_k,$$

and extending linearly. A direct computation shows that

$$(1) \quad v \cdot v \cdot (v_1 \wedge \cdots \wedge v_k) = -Q(v)(v_1 \wedge \cdots \wedge v_k),$$

and hence by linearity this is true for the action of v on all of $\wedge^* V$. This action of V of course determines a module structure for $\wedge^* V$ over the tensor algebra $T(V)$. Because of the defining relations of the Clifford algebra and Equation (1) it follows that this action factors to make $\wedge^*(V)$ a module over $\text{Cl}(V, Q)$. The real-linear map $\text{Cl}(V, Q) \rightarrow \wedge^*(V)$ defined by $x \mapsto x \cdot 1$ is clearly surjective. Since domain and range have the same dimension, it is a real-linear isomorphism. This identifies $\text{Cl}(V, Q)$ with $\wedge^*(V)$. It is easy to see that under this identification the multiplication of the Clifford algebra is given by

$$v \cdot (v_1 \wedge \cdots \wedge v_k) = v \wedge v_1 \wedge \cdots \wedge v_k + \sum_{i=1}^k (-1)^i \langle v, v_i \rangle_Q v_1 \wedge \cdots \wedge v_{i-1} \wedge v_{i+1} \wedge \cdots \wedge v_k.$$

From now on we use this identification implicitly.

I.2. The Spin Groups. — Let V be a finite dimensional real vector space and let Q be a positive definite quadratic form. We define $\text{Pin}(V, Q)$ as the subgroup of the multiplicative group of units of $\text{Cl}(V, Q)$ generated by $\{v \in V \mid Q(v) = 1\}$. We define $\text{Spin}(V, Q) = \text{Pin}(V, Q) \cap \text{Cl}_0(V, Q)$. It is easy to see that $\text{Spin}(V, Q)$ is a compact Lie group, see for example the proof of Theorem I.2.2 below. We denote $\text{Spin}(\mathbf{R}^n)$ by $\text{Spin}(n)$. When n is odd, we denote the representation of $\text{Spin}(n)$ obtained from either complex Clifford module $S^\pm(\mathbf{R}^n)$ of $\text{Cl}(\mathbf{R}^n)$ by Δ_n . (As we have seen these representations are isomorphic when restricted to $\text{Cl}_0(\mathbf{R}^n)$ and hence when restricted to $\text{Spin}(n)$.) When n is even the complex Clifford module $S(\mathbf{R}^n)$ decomposes as a representation of $\text{Spin}(n)$ into a direct sum of two irreducible representations under the decomposition into ± 1 -eigenspaces for $\omega_{\mathbf{C}}$. We denote these by $\Delta_n^\pm$. In all cases, these are irreducible unitary representations of $\text{Spin}(n)$.

Claim I.2.1. — *The conjugation action of $\text{Pin}(V, Q)$ on $\text{Cl}(V, Q)$ normalizes the subspace $V \subset \text{Cl}(V, Q)$ and acts on it by orthogonal transformations.*

Proof. — Let $v \in V$ satisfy $Q(v) = 1$ and let $w \in V$. If w is a multiple of v , then clearly $v w v^{-1} = w$. If w is orthogonal (under the bilinear form associated to Q) to v , then $v w v^{-1} = -w$. This shows that conjugation by v stabilizes $V \subset \text{Cl}(V, Q)$ and in fact is minus the reflection in v . Clearly, then the conjugation action of $\text{Pin}(V)$ induces an orthogonal action of $\text{Pin}(V, Q)$ on V . $\square$

Theorem I.2.2. — *$\text{Spin}(V, Q)$ is a subgroup of the group of units of norm one of $\text{Cl}_0(V, Q)$. The conjugation action of $\text{Spin}(V, Q)$ normalizes $V \subset \text{Cl}(V, Q)$ and gives an action of $\text{Spin}(V, Q)$ by orientation-preserving orthogonal transformations of V . This construction gives a non-trivial double covering homomorphism $\text{Spin}(V, Q) \rightarrow \text{SO}(V, Q)$.*

Proof. — Clearly, the restriction of the conjugation action to $\text{Spin}(V, Q)$ determines a homomorphism $\text{Spin}(V, Q) \rightarrow \text{SO}(V, Q)$. By construction, the image of this homomorphism contains all rotations in two-dimensional subspaces of V . Since these generate all of $\text{SO}(V, Q)$, it follows that this homomorphism is onto. The kernel of this homomorphism is the intersection of $\text{Spin}(V, Q)$ with the center of $\text{Cl}(V, Q)$. This intersection is ± 1 . Thus, we see that $\text{Spin}(V, Q)$ is a double covering of $\text{SO}(V, Q)$. It is easy to see that -1 is in the identity component of $\text{Spin}(V, Q)$ and hence $\text{Spin}(V, Q)$ is connected. In this manner we have constructed the non-trivial double covering of $\text{SO}(V, Q)$, which of course is its universal covering provided that the dimension of V is at least three. $\square$

Examples. — As we have seen $\text{Cl}_0(\mathbf{R}^3)$ is $\mathbf{H}$. The group of units of norm one of $\text{Cl}_0(\mathbf{R}^3)$ is thus the three-sphere. In this case $\text{Spin}(3)$ is the entire group of units of norm one in $\text{Cl}_0(\mathbf{R}^3)$. The restriction of either Clifford module $S^\pm(\mathbf{R}^3)$ to a representation of $\text{Spin}(3)$ yields the usual representation of S^3 on $\mathbf{H}$ by left multiplication, or equivalently the usual representation of $\text{SU}(2)$ on $\mathbf{C}^2$. Similarly, $\text{Cl}_0(\mathbf{R}^4) = \mathbf{H} \oplus \mathbf{H}$. In this case $\text{Spin}(4)$ is the product $S^3 \times S^3 = \text{SU}(2) \times \text{SU}(2)$ double covering $\text{SO}(4) = \text{SU}(2) \times_{\pm 1} \text{SU}(2)$. The Clifford module $\mathbf{H}^2$ decomposes under $\text{Spin}(4)$ as the direct sum of the defining two-dimensional representations of the two $\text{SU}(2)$ -factors.

I.2.1. The Groups $\text{Spin}^c(n)$. — There are closely related groups that will play a central role in the Seiberg-Witten equations. These are the groups $\text{Spin}^c(n)$.

Definition I.2.3. — Let V be a finite dimensional real vector space and let Q be a positive definite quadratic form on V . The group $\text{Spin}^c(V, Q)$ is the subgroup of $\text{Cl}_0(V, Q) \otimes_{\mathbf{R}} \mathbf{C}$ generated by $\text{Spin}(V, Q)$ and $S^1 \subset \mathbf{C}$. We denote by $\text{Spin}^c(n)$ the group $\text{Spin}^c(\mathbf{R}^n)$.

Clearly, the representations Δ_n , for n odd, and $\Delta_n^\pm$, for n even extend to irreducible unitary representations of $\text{Spin}^c(n)$. We denote by $\det: \text{Spin}^c(V, Q) \rightarrow S^1$ the projection modulo the normal subgroup $\text{Spin}(V, Q) \subset \text{Spin}^c(V, Q)$. The restriction of $\det$ to the center of $\text{Spin}^c(V, Q)$ yields a double covering $S^1 \rightarrow S^1$.

The following is clear.

Proposition I.2.4. — *The group $\text{Spin}^c(V, Q)$ is isomorphic to $\text{Spin}(V, Q) \times_{\pm 1} S^1$ and the center of $\text{Spin}^c(V, Q)$ is S^1 . The quotient of $\text{Spin}^c(V, Q)$ by its center is identified with $\text{SO}(V, Q)$. The Clifford modules of (V, Q) are representations of $\text{Spin}^c(V, Q)$.*

Examples. — $\text{Spin}^c(3) = \text{U}(2)$ and

$$\text{Spin}^c(4) = \{(A, B) \in \text{U}(2) \times \text{U}(2) \mid \det(A) = \det(B)\}.$$

I.3. Clifford Module Bundles. — Let $E \rightarrow M$ be an oriented, smooth, orthogonal vector bundle over a smooth manifold. Let n be the dimension of the fibers of this bundle. Then we automatically have the associated bundle of Clifford algebras $\text{Cl}(E) \rightarrow M$ and the subbundle $\text{Cl}_0(E)$. It is more subtle to decide whether the Clifford module bundles exist. In general they do not; there are topological obstructions to their existence. The standard way to construct them is as follows.

Let $P_E \rightarrow M$ be the $\text{SO}(n)$ -frame bundle of E (the bundle of oriented orthonormal frames). Suppose that we have a Spin-structure on E . That is to say there is a principal $\text{Spin}(n)$ -bundle, $\tilde{P} \rightarrow M$ and an isomorphism of principal $\text{SO}(n)$ bundles $\tilde{P}/\{\pm 1\} \rightarrow P_E$. If n is odd, we form the complex vector bundle $S(\tilde{P})$ associated to $\tilde{P}$ and the representation Δ_n . If n is even we form the bundles $S^\pm(\tilde{P})$ associated to $\tilde{P}$ and the representations $\Delta_n^\pm$. These bundles are unitary bundles of (complex) dimension $2^{\lfloor n/2 \rfloor}$. These bundles are bundles of modules over $\text{Cl}_0(E)$, which fiber-by-fiber are the irreducible Clifford modules over $\text{Cl}_0(E)$. In the case when n is even we set $S(\tilde{P})$ equal to $S^+(\tilde{P}) \oplus S^-(\tilde{P})$. Thus, if E has a Spin-structure the bundle of complex Clifford modules exists.

I.3.1. Spin-connections. — Now suppose that E has an orthogonal connection, with orthogonal covariant derivative ∇_E . There is a unique covariant derivative ∇ on $\text{Cl}(E)$ with the following properties:

- (1) The subbundle $E \subset \text{Cl}(E)$ is invariant under ∇ .
- (2) The restriction of ∇ to $E \subset \text{Cl}(E)$ is ∇_E .
- (3) For all x, y local sections of $\text{Cl}(E)$ and all $\tau \in TM$, we have $\nabla(x \cdot y)(\tau) = \nabla(x)(\tau) \cdot y + x \cdot \nabla(y)(\tau)$.

The covariant derivative ∇_E is induced by a connection on $P_E \rightarrow M$. We can view this connection as a family of horizontal subspaces in TP_E with certain properties. Lifting under the double covering mapping produces a $\text{Spin}(n)$ -connection on $\tilde{P}$. This connection induces a covariant derivative $\tilde{\nabla}$ on $S(\tilde{P})$. This is a unitary operator which is compatible with ∇ on $\text{Cl}_0(E)$ in the sense that for all local sections x of $\text{Cl}_0(E)$ and ψ of $S(\tilde{P})$ and all $\tau \in TM$ we have

$$\tilde{\nabla}(x \cdot \psi)(\tau) = \nabla(x)(\tau) \cdot \psi + x \cdot \tilde{\nabla}(\psi)(\tau).$$

I.3.2. The Obstruction to the Existence of a Spin-structure. — We have seen that, supposing that a Spin-structure exists on E , we have complex Clifford module bundles and connections satisfying all the natural properties that one would expect. But when does E have a Spin-structure? There is a well-known answer to this question. An $\text{SO}(n)$ -bundle $E \rightarrow M$ has a Spin-structure, *i.e.*, can be lifted to a $\text{Spin}(n)$ -bundle if and only if the second Stiefel-Whitney class $w_2(E) \in H^2(M; \mathbf{Z}/2\mathbf{Z})$ vanishes. This class is the obstruction to the existence of a double covering of the principal $\text{SO}(n)$ -bundle associated to E which restricts to each fiber to be a non-trivial double covering. We have the famous Wu formula, see [19] p.132 or [8] p.169, relating the second Stiefel-Whitney class $w_2(TM)$ of the tangent bundle of M with the self-intersection form: namely, $\langle w_2(TM), x \rangle = x \cdot x \pmod{2}$ for all $x \in H_2(M; \mathbf{Z}/2\mathbf{Z})$. Thus, the

frame bundle of a simply connected riemannian four-manifold has a Spin-structure if and only if the intersection form of the manifold is even. This shows that not all four-manifolds have Spin-structures.

I.3.3. Spin^c-liftings and Clifford Module Bundles. — By a Spin^c-lifting of an orthogonal bundle $E \rightarrow M$ we mean a Spin^c(n)-bundle $\tilde{P} \rightarrow M$ whose quotient modulo the center of Spin^c(n) is identified with P_E . The obstruction to such a lifting is $\delta(w_2(E)) \in H^3(M; \mathbf{Z})$, the integral Bockstein of the second Stiefel-Whitney class. The vanishing of this class is equivalent to the existence of a lifting of $w_2(E)$ to an integral class $c \in H^2(M; \mathbf{Z})$. We can see this correspondence directly. If $\tilde{P} \rightarrow M$ is a Spin^c-structure for E , then $c_1(\tilde{P})$ is an integral lift of $w_2(P)$. Conversely, suppose that we cover M by contractible open subsets U_i with contractible intersections. Fix trivializations of $E|_{U_i}$ and let $g_{ij}: U_i \cap U_j \rightarrow \text{SO}(n)$ be the resulting transition functions for E . For each ij , lift g_{ij} to a function $\tilde{g}_{ij}: U_i \cap U_j \rightarrow \text{Spin}(n)$. (This is possible since these intersections are assumed contractible.) Then for each $\{ijk\}$ we have

$$\tilde{g}_{jk} \circ \tilde{g}_{ij} = w_{ijk} \tilde{g}_{ik}$$

for some 2 Čech-cocycle w_{ijk} with values in $\{\pm 1\} \subset \text{Spin}(n)$. This is a Čech-cocycle representative for $w_2(E)$. Suppose that $\{w_{ijk}\}$ lifts to an integral 2 cochain $\{c_{ijk}\}$. With our conventions this means that for all ijk we have

$$\exp(\pi\sqrt{-1} c_{ijk}) = w_{ijk}.$$

From the exact sequence

$$0 \longrightarrow \mathbf{Z} \longrightarrow \mathbf{C} \xrightarrow{\exp} \mathbf{C}^* \longrightarrow 0,$$

and the fact that all cohomology of the sheaf of complex-valued C^∞ -functions vanishes, we see that there are smooth functions $\{\lambda_{ij}: U_i \cap U_j \rightarrow \mathbf{C}^*\}$ whose coboundary is $\{c_{ijk}\}$. This means that there are functions $\tilde{\lambda}_{ij}: U_i \cap U_j \rightarrow \mathbf{C}$ with $\exp(\tilde{\lambda}_{ij}) = \lambda_{ij}$ for all ij such that for all $\{ijk\}$ we have

$$\tilde{\lambda}_{jk} \tilde{\lambda}_{ij} = c_{ijk} \tilde{\lambda}_{ik}.$$

Let $\mu_{ij} = \exp(\frac{1}{2}\tilde{\lambda}_{ij})$. We have

$$\mu_{jk}\mu_{ij} = \exp(\pi\sqrt{-1} c_{ijk})\mu_{ik} = w_{ijk}\mu_{ik}.$$

Now we consider the products $\tilde{g}_{ij}\mu_{ij}: U_i \cap U_j \rightarrow \text{Spin}^c$. (Here, we are viewing $\mathbf{C}^*$ as the center of Spin^c.) The fact that the cochains $\{\tilde{g}_{ij}\}$ and μ_{ij} have coboundaries which lie in ± 1 and are equal, it follows that there products do indeed form a one cocycle for the group Spin^c and hence determine a Spin^c-bundle. This is a Spin^c-lifting of E . By a Spin^c-structure on a riemannian manifold we mean a Spin^c-lifting of the tangent bundle of the manifold.

As we have seen, the first Chern class c of the determinant line bundle of a Spin^c-lifting of E gives an integral lift of $w_2(E)$. Since the center of Spin^c is $\mathbf{C}^*$, one can

tensor a Spin^c -bundle by a line bundle. This does not affect the quotient $\text{SO}(n)$ -bundle. In fact all Spin^c -liftings of E arise by tensoring one given one by line bundles. Tensoring a Spin^c -bundle by a line bundle with first Chern class a changes the first Chern class of the determinant line bundle by $2a$. Thus, if there is no two-torsion in $H^2(M; \mathbf{Z})$, then the Spin^c -lifting is determined up to isomorphism by the first Chern class of its determinant line bundle. It is easy to see that as we tensor by a line bundle the spin bundles of the associated Spin^c -liftings change according to the equation: $S(\tilde{P} \otimes L) = S(\tilde{P}) \otimes L$.

The notion of a Spin^c -lifting of vector bundle $E \rightarrow M$ with an orthogonal structure $\langle \cdot, \cdot \rangle$ is a topological notion in the following sense. Suppose that $\langle \cdot, \cdot \rangle'$ is another orthogonal structure on E . Then there is a natural one-to-one correspondence between the Spin^c -liftings of the first orthogonal structure and those of the second. Thus, we can speak of a Spin^c -structure for E without having first imposed an orthogonal structure on E . Such a structure will determine a Spin^c -lifting of any orthogonal structure on E . As a special case of this, we can speak of a Spin^c -structure of a smooth manifold without having given that manifold a riemannian metric. For each metric we obtain a Spin^c -lifting of the orthogonal frame bundle associated with that metric.

The following lemma shows that in the case of four-manifolds Spin^c -structures always exist.

Lemma I.3.1. — *Let M be an oriented riemannian four-manifold. Then the frame bundle of TM admits a Spin^c -lifting. The set of all isomorphism classes of liftings is a principal homogeneous space for $H^2(M; \mathbf{Z})$.*

The standard reference for this result is [8], p.169. One can also find it on page 186 of [6]. It is such a pleasing algebraic topological argument that we sketch the proof here.

Proof. — We concern ourselves with the existence since the second result of the lemma is completely standard. As we have seen above, the problem is to show that $w_2(TM)$ has a lifting to an integral class. By the Universal Coefficient Theorem a mod 2 class has an integral lift if and only if its evaluation on all torsion integral classes vanishes. Thus, we must see that if $x \in H_2(M; \mathbf{Z})$ is a torsion class then $\langle w_2(TM), x \rangle = 0$. This simply means that $x \cdot x = 0 \pmod{2}$. But since x is an integral class $x \cdot x$ is an integer and since x is of finite order, a positive multiple of this integer is zero. The result follows. $\square$

Since the complex Clifford representations Δ_n and $\Delta_n^\pm$ of $\text{Spin}(n)$ extend to $\text{Spin}^c(n)$, if we have a Spin^c -lifting $\tilde{P}$ of P_E , then we can define the associated complex unitary vector bundles $S(\tilde{P})$ or $S^\pm(\tilde{P})$, which again are the bundles of complex Clifford modules over $\text{Cl}(E)$.

Suppose that we have a $\text{Spin}^c(n)$ -lifting $\tilde{P} \rightarrow M$ for E and that we have an $\text{SO}(n)$ -connection on E . The bundle $\tilde{P}$ is the double covering of the fibered product $P \times_M L$ where P is the principal $\text{SO}(n)$ -bundle associated to E and L is the principal $\text{U}(1)$ -bundle associated to $\det(\tilde{P})$. Thus, the given the connection on P together with a $\text{U}(1)$ -connection A on L will lift to the double covering to define a Spin^c -connection $\tilde{\nabla}_A$ on $\tilde{P}$ and hence a unitary covariant derivative on the bundle of complex Clifford modules $S(\tilde{P})$. Once again this connection is compatible with the connection on $\text{Cl}(E)$ determined by the connection on E .

I.4. The Dirac Operator. — Let M be an oriented riemannian manifold. Let $\tilde{P} \rightarrow M$ be a Spin^c -lifting of the tangent bundle of M . Let $S(\tilde{P}) \rightarrow M$ be the unitary bundle associated to $\tilde{P}$ and the complex Clifford module. Suppose that we have a $\text{U}(1)$ -connection A on the determinant line bundle L of $\tilde{P}$. We denote the curvature of this connection by $F_A \in \Omega^2(M; \mathbf{R})$. Let

$$\tilde{\nabla}_A: \Omega^0(M; S(\tilde{P})) \longrightarrow \Omega^0(M; T^*M \otimes S(\tilde{P})).$$

be the covariant derivative on $S(\tilde{P})$ determined by the Spin^c -connection on $\tilde{P}$ associated to the Levi-Civita connection on the tangent bundle of M and a connection A .

We also have the composition

$$T^*M \otimes S(\tilde{P}) \longrightarrow TM \otimes S(\tilde{P}) \longrightarrow S(\tilde{P}).$$

(The first map is the duality map given by the riemannian metric, and the second is Clifford multiplication.) Combining these two maps produces a formally self-adjoint first-order linear elliptic differential operator, the Dirac operator.

$$\not{D}_A: \Omega^0(M; S(\tilde{P})) \longrightarrow \Omega^0(M; S(\tilde{P})).$$

In the case when $\dim(M)$ is even and M is oriented, the Dirac operator sends sections of $S^+(\tilde{P})$ to sections of $S^-(\tilde{P})$ and vice-versa. When the dimension of M is 4, according to the Atiyah-Singer Theorem [1] the index of

$$\not{D}_A: \Omega^0(M; S^+(\tilde{P})) \longrightarrow \Omega^0(M; S^-(\tilde{P}))$$

is equal to

$$\frac{c_1(L)^2}{8} + \hat{A}(M) = \frac{c_1(L)^2 - \sigma(M)}{8},$$

where $\sigma(M)$ is the signature of M .

I.4.1. The Bochner Formula. — Another important fact about the Dirac operator is the Bochner formula, see p. 155 of [14], that for any section ψ of the spin bundle $S(\tilde{P})$ we have:

$$(2) \quad \not{D}_A \circ \not{D}_A(\psi) = \tilde{\nabla}_A^* \circ \tilde{\nabla}_A(\psi) + \frac{\kappa}{4}\psi + \frac{F_A}{2} \cdot \psi.$$

Here, $\widetilde{\nabla}_A^*$ is the adjoint of $\widetilde{\nabla}_A$, κ is the scalar curvature function of the manifold, and the multiplication in the last term is Clifford multiplication and uses the identification of the exterior algebra of T^*M with the Clifford algebra.

II. The Seiberg-Witten Equations and the Seiberg-Witten Moduli Space

II.1. The Equations. — We fix an oriented, compact, connected riemannian 4-manifold M and a Spin^c -structure $\widetilde{P} \rightarrow M$ for the tangent bundle. We denote by $L \rightarrow M$ the determinant line bundle of this Spin^c -structure. The Seiberg-Witten equations are equations on a pair (A, ψ) consisting of a $\text{U}(1)$ -connection on L and a section ψ of $S^+(\widetilde{P})$. The Seiberg-Witten (SW) equations are

$$(3) \quad \not{D}_A(\psi) = 0$$

$$(4) \quad F_A^+ = \psi \otimes \overline{\psi} - \frac{|\psi|^2}{2} \cdot \text{Id}.$$

We denote by $q(\psi)$ the quadratic function on the right-hand side of Equation (4). Equation (3) is straight-forward: it says that ψ is a harmonic plus spinor for the Dirac operator determined by the connection A on the determinant line bundle. Equation (4) requires some explanation. We have identified the differential forms with sections of the Clifford bundle $\text{Cl}(TM)$. Under this identification the even forms are identified with sections of $\text{Cl}_0(TM)$ and the forms

$$\mathbf{C}(1 + \omega_{\mathbf{C}}) \oplus \Omega_2^+(M; \mathbf{C})$$

are identified with the sections of $(\text{Cl}_0(TM) \otimes_{\mathbf{R}} \mathbf{C})^+$. According to Lemma I.1.3 this latter algebra is identified under Clifford multiplication with the sections of the endomorphism algebra of S^+ . The following lemma is a direct computation using the fact that $\text{Cl}_0^+(\mathbf{R}^4)$ is identified with the quaternions.

Lemma II.1.1. — *Under the identification of $\wedge^* T^*M$ with $\text{Cl}(TM)$ and under Clifford multiplication, $(\wedge_+^2 T^*M) \otimes i\mathbf{R}$ is identified with the space of self-adjoint trace-free endomorphisms of $S^+(\widetilde{P})$.*

This lemma tells how to interpret Equation (4): F_A^+ is a self-dual, purely imaginary two-form and hence is identified with a self-adjoint traceless endomorphism of $S^+(\widetilde{P})$. The term $\psi \otimes \overline{\psi} \in S^+(\widetilde{P}) \otimes \overline{S^+(\widetilde{P})}$ and the unitary structure on $S^+(\widetilde{P})$ gives an identification of $S^+(\widetilde{P}) \otimes \overline{S^+(\widetilde{P})}$ with $S^+(\widetilde{P}) \otimes (S^+(\widetilde{P}))^* = \text{End}(S^+(\widetilde{P}))$. Thus, the right-hand side of the second Seiberg-Witten equation is naturally an endomorphism of $S^+(\widetilde{P})$. In a local unitary trivialization for $S^+(\widetilde{P})$ in which the section ψ is given by the column vector

$$\begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix}$$

the endomorphism $\psi \otimes \bar{\psi}$ is given by

$$\begin{pmatrix} |\psi_1|^2 & \psi_1 \bar{\psi}_2 \\ \bar{\psi}_1 \psi_2 & |\psi_2|^2 \end{pmatrix}.$$

Clearly then, $\psi \otimes \bar{\psi}$ and $(|\psi|^2/2) \text{Id}$ are self-adjoint. It is also clear that they each have trace equal to $|\psi|^2$, so that their difference is traceless. In the local trivialization, we have

$$q(\psi) = \begin{pmatrix} \frac{|\psi_1|^2 - |\psi_2|^2}{2} & \psi_1 \bar{\psi}_2 \\ \bar{\psi}_1 \psi_2 & \frac{|\psi_2|^2 - |\psi_1|^2}{2} \end{pmatrix}.$$

Thus, the second Seiberg-Witten equation is to be viewed as an equation of traceless, self-adjoint endomorphisms of $S^+(\tilde{P})$. Using the inverse of Clifford multiplication, this equation can be viewed as an equation of self-dual, purely imaginary two-forms.

II.2. The Moduli Space. — The path we shall follow is to show that the solutions of these equations (modulo changes of gauge) form a reasonable moduli space. We then construct the Seiberg-Witten invariants of the Spin^c -structure on a manifold from the cohomology ring of this moduli space. In order to carry out this program we need to establish many technical results about the space of solutions of these equations.

II.2.1. The SW Operator. — We fix a compact, oriented smooth 4-manifold M with a riemannian metric g and a Spin^c -structure on M with determinant line bundle L . We view the Seiberg-Witten equations as defining an operator

$$SW_{\tilde{P}}: \mathcal{A}(L) \times \Gamma(S^+(\tilde{P})) \longrightarrow \Omega_+^2(M; i\mathbf{R}) \times \Gamma(S^-(\tilde{P})),$$

where $\mathcal{A}(L)$ is the (affine space) of $U(1)$ -connections on L and where for any vector bundle S , $\Gamma(S)$ denotes the vector space of sections of S . We also abbreviate the notation by setting

$$\mathcal{C}(\tilde{P}) = \mathcal{A}(L) \times \Gamma(S^+(\tilde{P})).$$

The leading order terms of $SW_{\tilde{P}}$ are first-order linear, but there are non-linear algebraic terms (terms involving no differentiations) which are quadratic in A and ψ . Thus, $SW_{\tilde{P}}$ is a non-linear first-order operator. It is not an elliptic operator because there is an infinite dimensional group of changes of gauge that act leaving the equation invariant. In particular, the space of solutions (if non-empty) will automatically be infinite dimensional because of the gauge symmetries.

II.2.2. Group of Changes of Gauge. — The group of changes of gauge that is relevant is the group of Spin^c -automorphisms $\tilde{P}$ which project to the identity on the frame bundle of the tangent bundle of M . These are exactly the maps of M to the center of $\text{Spin}^c(4)$, which is a circle. Thus, the group $\mathcal{G}$ of changes of gauge is $(S^1)^M$, viewed

as maps of M to the center of $\text{Spin}^c(4)$. The action of $\mathcal{G}$ on $\mathcal{C}(\tilde{P})$ is a right action and is given by

$$(A, \psi) \cdot \sigma = ((\sigma^2)^* A, \sigma^{-1} \cdot \psi).$$

The multiplication in the second factor is the usual multiplication of a section of a complex vector bundle by an element of S^1 viewed as a unit complex number. The map from the center of $\text{Spin}^c(4)$ to the determinant circle of $\text{Spin}^c(4)$ is the squaring map. Hence, σ^2 is the natural action of σ on the determinant line bundle. There is also an action of $\mathcal{G}$ on the value space of the Seiberg-Witten operator given by

$$(f, \psi^-) \cdot \sigma = (f, \sigma^{-1} \cdot \psi^-)$$

for a self-dual two-form f and a section ψ^- of $S^-(\tilde{P})$.

Lemma II.2.1. — *The Seiberg-Witten operator is equivariant with respect to these actions in the sense that*

$$SW_{\tilde{P}}((A, \psi) \cdot \sigma) = SW_{\tilde{P}}(A, \psi) \cdot \sigma.$$

In particular, the space of solutions to the Seiberg-Witten equations is invariant under the action of $\mathcal{G}$.

II.2.3. Definition of the Moduli Space

Definition II.2.2. — The Seiberg-Witten moduli space $\mathcal{M}(\tilde{P}, g)$ is the quotient space of gauge equivalence classes of solutions to the Seiberg-Witten equations associated to the riemannian manifold (M, g) and the Spin^c -structure $\tilde{P}$. When the metric g is clear from context we drop it from the notation and write $\mathcal{M}(\tilde{P})$ for the moduli space.

II.3. Properties of the Action by Changes of Gauge and the SW Operator

In order to study the properties of $\mathcal{M}(\tilde{P})$ we need to study in more detail the action $\mathcal{C}(\tilde{P}) \times \mathcal{G} \rightarrow \mathcal{C}(\tilde{P})$, and in particular, we wish to understand the possible stabilizers of points under this action.

Lemma II.3.1. — *If $(A, \psi)\sigma = (A, \psi)$, then either $\sigma = 1$ or $\psi = 0$ and σ is a constant map from M to S^1 .*

Proof. — The condition that $(\sigma^2)^* A = A$ tells us that $d\sigma = 0$, or equivalently that σ is a constant map. Now the equation $\sigma^{-1} \cdot \psi = \psi$ implies that either $\psi = 0$ or $\sigma = 1$. $\square$

Corollary II.3.2. — *If $(A, \psi) \in \mathcal{C}(\tilde{P})$ with $\psi \neq 0$, then the stabilizer of (A, ψ) in $\mathcal{G}$ is trivial. If $\psi = 0$, then its stabilizer is the circle of constant maps from M to S^1 .*

We call configurations $(A, 0)$ *singular* since their stabilizers are S^1 and all others *nonsingular*. The reader should be aware that the usual names for these types of configurations in the literature are ‘reducible’ and ‘irreducible’, respectively. The

reason for this terminology is by analogy with $SU(2)$ -gauge theory where the reducible connections are those whose stabilizer under the change of gauge action is S^1 . We will consistently use the terms ‘singular’ and ‘nonsingular’. We denote by $\mathcal{C}^{\text{ns}}(\tilde{P})$ the open subset of nonsingular configurations and by $\mathcal{C}^{\text{sing}}(\tilde{P})$ the closed subset of singular configurations.

Proposition II.3.3. — *For appropriate Sobolev completions of $\mathcal{G}$, $\mathcal{C}(\tilde{P})$ and $\mathcal{V} = \Omega_+^2(M; i\mathbf{R}) \times \Gamma(S^-(\tilde{P}))$ the following hold:*

- (1) $\mathcal{G}$ is a Hilbert Lie group.
- (2) $\mathcal{C}(\tilde{P})$ is a contractible Hilbert manifold.
- (3) The action of $\mathcal{G}$ on $\mathcal{C}(\tilde{P})$ is a smooth, proper action admitting local slices.
- (4) The quotient space $\mathcal{B}(\tilde{P})$ of this action is a stratified space with two strata. The closed subspace $\mathcal{B}^{\text{sing}}(\tilde{P})$ of gauge equivalence classes of singular configurations is a Hilbert manifold. The complement $\mathcal{B}^{\text{ns}}(\tilde{P})$ of nonsingular configurations is a smooth Hilbert manifold. The link of $\mathcal{B}^{\text{sing}}(\tilde{P})$ in $\mathcal{B}(\tilde{P})$ is the quotient of the complex Hilbert space $\Gamma(S^+(\tilde{P}))$ by the natural circle action.
- (5) There is a smooth Hilbert space bundle

$$\mathcal{C}^{\text{ns}}(\tilde{P}) \times_{\mathcal{G}} \mathcal{V} \longrightarrow \mathcal{B}^{\text{ns}}(\tilde{P}).$$

- (6) The operator $SW_{\tilde{P}}$ defines a smooth section of this Hilbert space bundle.

This is a fairly standard type of proposition given the lemmas above about the stabilizers of the action. The one thing that requires proof is the existence of local slices for the $\mathcal{G}$ -action. In fact, one proceeds as follows. Given $(A, \psi) \in \mathcal{C}(\tilde{P})$ there is a closed linear subspace $H \subset T_{(A, \psi)}\mathcal{C}(\tilde{P})$ invariant under the stabilizer in $\mathcal{G}$ of (A, ψ) . For any open neighborhood U of $0 \in H$, also invariant under the stabilizer, use the affine structure to embed U into $\mathcal{C}(\tilde{P})$ centered by (A, ψ) , and then, use the action to define a map

$$U \times_{\text{Stab}(A, \psi)} \mathcal{G} \longrightarrow \mathcal{C}(\tilde{P}).$$

The result is that, for sufficiently small U , this map is a diffeomorphism onto an open subset of $\mathcal{C}(\tilde{P})$. (In the special case of configurations of the form $(A_0, 0)$ we can take the linear subspace to be all configurations of the form (A, ψ) where $d^*(A - A_0) = 0$.) Clearly, U then becomes a local slice for the action. This proof of the existence of slices is by now a fairly standard argument, cf. [20, 4].

II.3.1. The Fredholm Property. — Now let us show that the operator $SW_{\tilde{P}}$ determines a Fredholm section of the Hilbert space bundle, and let us calculate the index of this section. To do this requires a study of the linearization of the gauge action and of $SW_{\tilde{P}}$ at a solution (A, ψ) of the SW equations.

The Lie algebra of $\mathcal{G}$ is $\Omega^0(M; i\mathbf{R})$ and the tangent space at any point to $\mathcal{C}(\tilde{P})$ is $\Omega^1(M; i\mathbf{R}) \oplus \Gamma(S^+(\tilde{P}))$. The linearization of the gauge action and the SW equations

at (A, ψ) gives the following sequence of operators:

$$\Omega^0(M; i\mathbf{R}) \longrightarrow \Omega^1(M; i\mathbf{R}) \oplus \Gamma(S^+(\tilde{P})) \longrightarrow \Omega_+^2(M; i\mathbf{R}) \oplus \Gamma(S^-(\tilde{P})).$$

The first map is given by

$$f \longmapsto (2df, -f \cdot \psi)$$

and the second map is given by

$$\begin{pmatrix} d_+ & -2B(\psi, \cdot) \\ \psi(\cdot) & \not\partial_A \end{pmatrix}.$$

Here, B is the bilinear form associated with $B(\psi, \psi) = q(\psi)$. It is easy to see that, provided that (A, ψ) is a solution to the SW equations, then this is a complex, and in fact an elliptic complex. To compute the index of the complex we can deform it through elliptic complexes to one which has the same first order terms but no zeroth order terms. (We do this by simply damping out the zeroth order terms.) Since this deformation is through elliptic complexes it does not change the index of the complex. The complex consisting of the purely first order terms is the sum of two complexes:

$$\Omega^0(M; i\mathbf{R}) \longrightarrow \Omega^1(M; i\mathbf{R}) \longrightarrow \Omega_+^2(M; i\mathbf{R})$$

with operators being d and d_+ and

$$0 \longrightarrow \Gamma(S^+(\tilde{P})) \longrightarrow \Gamma(S^-(\tilde{P}))$$

with operator $\not\partial_A$. The first complex has index $1 - b_1 + b_2^+$ and the second complex has real index $(\sigma(M) - c_1(L)^2)/4$. Putting these together we see that the index of the complex is $(2\chi(M) + 3\sigma(M) - c_1(L)^2)/4$. We set d equal to the negative of this number. We have shown:

Corollary II.3.4. — *The operator $SW_{\tilde{P}}$ defines a section of the Hilbert bundle in Proposition II.3.3 which is a Fredholm operator of index d . The zeros of this section are the open subset $\mathcal{M}^{\text{ns}}(\tilde{P}) \subset \mathcal{M}(\tilde{P})$ of gauge equivalence classes of nonsingular solutions to the Seiberg-Witten equations.*

II.4. Properties of the Moduli Space. — Given the Fredholm property and index computation above, one hopes to show that in the ‘generic situation’ the section $SW_{\tilde{P}}$ will actually be transverse to the zero section so that the moduli space of solutions to the SW equations modulo $\mathcal{G}$; *i.e.*, the preimage of 0 under the section SW , will in fact be a smooth manifold of dimension d . It requires some work to decide what generic situation should mean and to establish such a result.

There is, however, one completely general consequence of the Fredholm condition: The open subset $\mathcal{M}^{\text{ns}}(\tilde{P})$ of nonsingular solutions has a (finite dimensional) real analytic structure, and in fact, if $[A, \psi] \in \mathcal{M}(\tilde{P})$ is (the gauge equivalence class of) and nonsingular solution, then setting H^1 and H^2 equal to the finite dimensional vector spaces of the first and second degree cohomology of the elliptic complex associated

to (A, ψ) , there is an open neighborhood $U \subset H^1$ of 0 and a real analytic mapping $\varphi: U \rightarrow H^2$ whose differential vanishes at the origin, and a smooth embedding $i: U \subset \mathcal{B}^{\text{ns}}$ such $i(\varphi^{-1}(0))$ is a neighborhood of $[A, \psi]$ in $\mathcal{M}^{\text{ns}}(\tilde{P})$. This produces a real analytic structure on this neighborhood of $\mathcal{M}^{\text{ns}}(\tilde{P})$. As we pass from one such neighborhood to another the real analytic structures on the overlap are compatible, thus giving $\mathcal{M}^{\text{ns}}(\tilde{P})$ the structure of a (finite dimensional) real analytic variety. (See § 4.5 of [4] for more details in a closely related situation.) The formal dimension of this real analytic variety is d in the sense that $\dim(H^1) - \dim H^2 = d$. Thus, if the map φ were ‘generic’ the actual dimension of $\mathcal{M}^{\text{ns}}(\tilde{P})$ as a stratified set would be d .

Of course, there is no reason to expect this genericity in general. There are examples where at each point $H^2 \neq 0$ but φ is identically zero. In these cases one would view $\mathcal{M}^{\text{ns}}(\tilde{P})$ as a smooth manifold of higher dimension with an obstruction bundle over it. This moduli space can then be replaced by the zeros of a generic section of the obstruction bundle.

II.4.1. Avoiding Singular Solutions

Lemma II.4.1. — *Let M be a closed, oriented 4-manifold and let $\tilde{P}$ be a Spin^c -structure on M with determinant line bundle L . Suppose that $c_1(L)$ is not a torsion cohomology class. Then in the space of metrics there is a smooth submanifold of codimension $b_2^+(M)$ such that for any metric g not in this subset there are no singular solutions to the Seiberg-Witten equations for $\tilde{P}$ when M is given the riemannian structure g .*

Proof. — A singular solution is the same as an ASD connection A on L . (See p. 38 of [4] for a definition of ASD.) Since the curvature of a connection on L is closed, an ASD connection on L must have harmonic curvature. This would mean that the class of $c_1(L)$ has an ASD harmonic representative. The subset of $H^2(M; \mathbf{R})$ which have ASD harmonic representatives is a linear subspace of codimension $b_2^+(M)$. The condition on the Grassmannian of subspaces of codimension $b_2^+(M)$ that the subspace pass through the non-zero point $c_1(L)$ is codimension $b_2^+(M)$. By p. 148 in [4] the map from the space of metrics on M to the Grassmannian of subspaces of codimension $b_2^+(M)$ inside $H^2(M; \mathbf{R})$ given by associating to each metric the subspace of classes with ASD harmonic representatives is transverse to the submanifold consisting of all linear subspaces containing a given non-zero vector. The result now follows. $\square$

From now on for any manifold with $b_2^+(M) > 0$, when considering $\mathcal{M}(\tilde{P})$ we implicitly assume that we have chosen a metric out of the open set of the previous lemma so that there are no singular solutions to the SW equations.

II.4.2. The Perturbation. — Even assuming that we have arranged that there are no singular solutions to the Seiberg-Witten equations, it is not known whether simply varying the metric is sufficient to make the section $SW_{\tilde{P}}$ transverse to the zero-section and hence make the moduli space a smooth manifold of the correct dimension. Rather we consider a different sort of perturbation.

Given a self-dual real two-form $h \in \Omega_+^2(M; \mathbf{R})$ we define the perturbed Seiberg-Witten equations SW_h by:

$$\begin{aligned}\phi_A(\psi) &= 0 \\ F_A^+ - ih &= \psi \otimes \bar{\psi} - \frac{|\psi|^2}{2} \text{Id}.\end{aligned}$$

A standard application of Sard's theorem shows that:

Proposition II.4.2. — *Suppose that M is a closed, oriented riemannian manifold with $b_2^+(M) > 0$. Let $\tilde{P}$ be a Spin^c -structure for M . Then for a generic $h \in \Omega_+^2(M; \mathbf{R})$ every solution to the perturbed Seiberg-Witten equations SW_h is nonsingular and the section determined by SW_h is transverse to the zero section so that the moduli space $\mathcal{M}_h(\tilde{P})$ of gauge equivalence classes of solutions to SW_h is a smooth submanifold of $\mathcal{B}^{\text{ns}}(\tilde{P})$ of dimension d .*

II.4.3. Compactness of the Moduli Space. — So far everything that we have discussed is a special case of results that hold in much more general geometric contexts, e.g., the case of ASD connections on $\text{SU}(2)$ -bundles. But the next result, compactness, is special to this case.

Theorem II.4.3. — *The moduli space $\mathcal{M}(\tilde{P})$ is a compact space.*

Proof. — The basic point is that if $SW(A, \psi) = 0$ then there is an *a priori* pointwise bound on the norm of ψ . First, note that working with the Sobolev space of L_k^2 -configurations for any $k \geq 5$, we are considering objects which are C^2 . Hence, using this completion we see that the moduli space consists of C^2 -solutions.

Applying the Bochner formula (Equation (2)) we see that

$$0 = \nabla_A^* \nabla_A(\psi) + \frac{\kappa}{4} \psi + \frac{F_A}{2} \cdot \psi.$$

Since ψ is a section of $S^+(\tilde{P})$, Clifford multiplication of ψ by an ASD two-form vanishes and hence we can replace the last term in the above equation by $\frac{F_A^+}{2} \cdot \psi$. The SW equations tell us that this product is equal to $\frac{|\psi|^2}{4} \psi$. Taking pointwise inner product with ψ we see:

$$0 = \langle \nabla_A^* \nabla_A(\psi), \psi \rangle + \frac{\kappa}{4} |\psi|^2 + \frac{|\psi|^4}{4}.$$

A standard result shows that at a local maximum of $|\psi|^2$ the first term is greater than or equal to zero, and hence that at the maximum point x_0 for $|\psi|^2$ we have

$$\frac{|\psi(x_0)|^4}{4} \leq \frac{-\kappa(x_0)}{4} |\psi(x_0)|^2.$$

Thus, for all $x \in M$ we have

$$|\psi(x)|^2 \leq \max(\max_{y \in M}(-\kappa(y)), 0).$$

This is the *a priori* bound on $|\psi|^2$.

Given this bound it is a standard bootstrap argument to show that, up to change of gauge, (A, ψ) are bounded *a priori* in the Sobolev norm L_ℓ^2 for every ℓ . This means that (A, ψ) are bounded *a priori* in the C^k -norm for every k . Thus, given a sequence of solutions (A_n, ψ_n) , after passing to a subsequence and acting by a sequence of changes of gauge, we can arrange that the sequence converges in the C^∞ -topology to a limit (A, ψ) which clearly is a solution to the SW equations. This proves compactness. $\square$

Remark II.4.4

(1) The same argument shows that if h is C^1 then there is an *a priori* bound to the pointwise norm of $|\psi|$ for any solution (A, ψ) to SW_h . This implies that for any h which is C^1 the moduli space $\mathcal{M}_h(\tilde{P})$ is compact. Thus, for a generic such h , provided that $b_2^+(M) > 0$, the moduli space $\mathcal{M}_h(\tilde{P})$ is a compact smooth manifold of dimension d .

(2) It turns out that, for any $k \geq 3$, the moduli space of gauge equivalence classes of L_k^2 -solutions to the Seiberg-Witten equations is independent of k . Hence, the restriction to $k \geq 5$ in the above argument is not essential but is for technical simplicity only.

II.4.4. Orientability. — Let us assume that $\mathcal{M}_h(\tilde{P})$ is smooth. Then the top exterior power of its tangent bundle is identified with the restriction to the moduli space of the inverse of the real determinant line bundle of the elliptic complex associated to each point. Since these complexes deform through families of elliptic complexes to ones that are split as a direct sum of a constant complex

$$\Omega^0(M; i\mathbf{R}) \longrightarrow \Omega^1(M; i\mathbf{R}) \longrightarrow \Omega_+^2(M; i\mathbf{R})$$

with maps d and d^+ and a varying complex of Dirac operators

$$0 \longrightarrow \Gamma(S^+(\tilde{P})) \longrightarrow \Gamma(S^-(\tilde{P}))$$

we see that the determinant line bundle is equivalent to the tensor product of the determinant line of the first complex (which is a trivial line bundle over the base), and the real determinant line bundle of the family of Dirac operators. Since these latter operators are complex linear, the real determinant line bundle is trivialized by the complex structure. Thus, the determinant line bundle of the family of elliptic complexes is identified with the constant line which is the determinant line of

$$\Omega^0(M; i\mathbf{R}) \longrightarrow \Omega^1(M; i\mathbf{R}) \longrightarrow \Omega_+^2(M; i\mathbf{R}).$$

This determinant line is

$$H^0(M; \mathbf{R}) \otimes \wedge^{\text{top}} H^1(M; \mathbf{R})^{-1} \otimes \wedge^{\text{top}} H_+^2(M; \mathbf{R}).$$

In particular, we see that the moduli space is an orientable manifold, and an orientation of it is determined by choosing orientations for $H^1(M; \mathbf{R})$ and $H_+^2(M; \mathbf{R})$. From now on we assume that we have made such a choice of homology orientations for M .

II.5. Definition of the Seiberg-Witten Invariant. — Let M be a closed oriented smooth four-manifold with $b_2^+(M) > 0$ (and with homology orientation as discussed in the last section). Fix a Spin^c -structure $\tilde{P}$ on M . Then for any metric g on M and for a generic self-dual two-form the moduli space $\mathcal{M}_h(\tilde{P})$ associated to the perturbed equations SW_h is a compact smooth submanifold of $\mathcal{B}^{\text{ns}}(\tilde{P})$. We denote by d the dimension of this manifold. Fix a point $x \in M$, and let $\mathcal{G}^0$ be the subgroup of the group of gauge transformations which are the identity over x . The quotient $\mathcal{M}_h^0(\tilde{P})$ of the space of all solutions to SW_h by $\mathcal{G}^0$ is a circle bundle over $\mathcal{M}_h(\tilde{P})$. Let $\mu \in H^2(\mathcal{M}_h(\tilde{P}); \mathbf{Z})$ be the first Chern class of this circle bundle. Then we define the Seiberg-Witten invariant of $\tilde{P}$ to be

$$\int_{\mathcal{M}_h(\tilde{P})} \mu^{(d/2)},$$

provided that d is even, and to be zero if d is odd.

Theorem II.5.1. — *If $b_2^+(M) > 1$, then the Seiberg-Witten invariant as defined in the previous paragraph is independent of the choice of metric, perturbing two-form h and point $x \in M$. The value of the invariant depends only on the Spin^c -structure on M and hence defines a function*

$$SW_M: \{\text{Spin}^c\text{-structures on } M\} \longrightarrow \mathbf{Z}.$$

(Recall that the set of Spin^c -structures on M is a principal homogeneous space for $H^2(M; \mathbf{Z})$.)

Proof. — Given two choices of metric and perturbing two-form, we can find a one-parameter family of metrics and two-forms that connects the two given sets of choices. Associated to the family is a parametrized perturbed moduli space of all gauge equivalence classes of solutions $[A, \psi, t] \in \mathcal{B}^{\text{ns}}(\tilde{P}) \times [0, 1]$ with the property that (A, ψ) is a solution to the perturbed equations SW_{h_t} defined using the metric g_t . Since $b_2^+(M) > 1$, by Lemma II.4.1 we can assume that for every value of t , the perturbed equations SW_{h_t} for the metric g_t have no singular solutions. Thus, for a generic path of metrics and perturbations this parametrized moduli space will be a compact smooth oriented manifold of dimension $d + 1$ whose boundary is the difference of the two moduli spaces associated with the metrics and perturbations at $\{1\}$ and $\{0\}$. Fixing a point $x \in M$ we have a circle bundle over the parametrized moduli space which extends the circle bundles associated to x over the ends. This shows that the invariants defined using the metrics and perturbations at the two ends and the point x agree. As we vary the point x in M , the topological type of the circle bundle does not change and thus neither does its first Chern class. Hence, the definition of the invariant is also independent of the choice of $x \in M$. $\square$

II.5.1. The Case when $b_2^+(M) = 1$. — In the case when $b_2^+(M) = 1$ it is still the case that for a metric g and generic perturbation h that the moduli space $\mathcal{M}_h(\tilde{P})$ is a compact, oriented smooth manifold of dimension d . Thus, we can define, as before, an invariant $SW_h(\tilde{P}) \in \mathbf{Z}$. What is different in this case is that as we vary the metric g and the perturbation h in a one-parameter family, we will encounter intermediate values (g_t, h_t) for which there are singular solutions. Because of this the parametrized moduli space will have singularities over which the cohomology class μ will not extend. Thus, the value of $SW_h(\tilde{P})$ depends on (g, h) . It turns out that the value depends only on the g -self-dual harmonic form which is the image of h . In fact, there is a chamber structure in the second cohomology and one can view $SW_M(\tilde{P})$ as being a function defined on the union of the interior of the chambers and constant on each component (*i.e.*, each open chamber). There is also a wall-crossing formula which gives the change in the invariant as one crosses a wall and passes from one chamber to the next. In the case when $\pi_1(M) = \{1\}$ and $d \equiv 0 \pmod{2}$ and $d \geq 0$, the change as we cross a wall is always ± 1 . Thus, in this case the invariant has an arithmetic structure associated to a certain reflection in cohomology. It is a more complicated mathematical object than simply an invariant of the manifold, but still it is possible to use it to derive topological consequences for such manifolds.

III. Seiberg-Witten Invariants of Kähler Manifolds

As was indicated by Witten in his seminal paper [28] on applications of the Seiberg-Witten invariant to the topology of four-manifolds, for a Kähler manifold one can completely solve the SW equations and compute the SW invariants of these manifolds.

III.1. The Holomorphic Interpretation of the Objects in the Kähler Case

The method of solving the SW equations on a Kähler manifold is to reinterpret the objects and equations in this context in purely holomorphic terms – taking ‘differential’ out of the PDE and replacing the equations with purely algebraic ones.

III.1.1. The Spin^c -structure Associated with an Almost Complex Structure. — Let us begin this program by studying Spin^c -structures.

Lemma III.1.1. — *The homomorphism $U(n) \rightarrow \text{SO}(2n)$ has a unique lifting to a homomorphism $\rho: U(n) \rightarrow \text{Spin}^c(2n)$ whose composition with the determinant map from $\text{Spin}^c(2n) \rightarrow S^1$ is the usual determinant on $U(n)$.*

Proof. — Consider the natural map of $U(n)$ to $\text{SO}(2n) \times S^1$ given by the product of the natural inclusion and the determinant map. A direct computation of the effect of this map on fundamental groups shows that it lifts to a homomorphism $U(n) \rightarrow \text{Spin}^c(2n)$. This lifting homomorphism is unique. $\square$

Corollary III.1.2. — *Let X be a riemannian $2n$ -manifold with an almost complex structure $J: TX \rightarrow TX$ compatible with the metric. Then there is a natural Spin^c -structure $P_{X,J}$ for X whose determinant line bundle is the inverse of the canonical bundle of the almost complex structure on TX determined by J .*

Proof. — The almost complex structure J determines a reduction of the structure group of TX from $\text{SO}(2n)$ to $\text{U}(n)$. Applying the homomorphism ρ of the last lemma, this produces a Spin^c -structure on TX whose determinant line bundle is the determinant line bundle of the complex structure on TX . This is the inverse of the determinant line bundle of the cotangent bundle, which is, by definition, the canonical bundle of the almost complex structure. $\square$

III.1.2. The Spin Bundles and the Dirac Operator. — Let J be an almost complex structure on a manifold X , and let $P_{X,J}$ the Spin^c -structure associated to (X, J) by the last corollary. Consider $S = \wedge^{0,*}(T^*X \otimes \mathbf{C})$, differential forms of type $(0, *)$. There is a natural action of T^*X on S given by

$$w \cdot (w_1 \wedge \cdots \wedge w_k) = \sqrt{2} \left(\pi^{0,1}(w) \wedge w_1 \wedge \cdots \wedge w_k + \sum_{i=1}^k (-1)^i \langle \pi^{0,1}(w), w_i \rangle w_1 \wedge \cdots \wedge w_{i-1} \wedge w_{i+1} \wedge \cdots \wedge w_k \right),$$

where $\langle \cdot, \cdot \rangle$ is the hermitian inner product (complex linear in the second variable) associated to the metric. Direct computation shows that

$$w \cdot w \cdot (w_1 \wedge \cdots \wedge w_k) = -|w|^2 (w_1 \wedge \cdots \wedge w_k).$$

Thus, this action extends to an action of the bundle of Clifford algebras associated to T^*X . In fact this identifies with bundle of Clifford modules $S^+(P_{X,J})$ with $\wedge^{0,2*}(T^*X \otimes \mathbf{C})$ and $S^-(P_{X,J})$ with $\wedge^{0,2*+1}(T^*X \otimes \mathbf{C})$.

Now suppose that X is a Kähler manifold. Then at every point p there is a local holomorphic coordinate system in which the metric is standard to second order. Thus, at p the Levi-Civita connection in these coordinates is usual differentiation. From this and the above description of Clifford multiplication it is easy to see that:

Proposition III.1.3. — *Let X be a Kähler manifold of dimension n and let K_X be its canonical line bundle. Then the Spin^c -structure P_X on X determined by Corollary III.1.2 has determinant line bundle K_X^{-1} . This line bundle has the natural holomorphic connection A_X induced by the metric. The Dirac operator $\not{D}_{A_X}: \Gamma(S^+(P_X)) \rightarrow \Gamma(S^-(P_X))$ associated to the connection A_X on the determinant line bundle is*

$$\sqrt{2} \left(\bar{\partial} + \bar{\partial}^* \right): \Omega^{0,2*}(X; \mathbf{C}) \longrightarrow \Omega^{0,2*+1}(X; \mathbf{C}).$$

Any other Spin^c -structure on X is of the form $P_X \otimes L$ for a complex line bundle L . The spin bundles of this Spin^c -structure are

$$\begin{aligned} S^+(P_X \otimes L) &= \wedge^{0,2*}(T^*X \otimes \mathbf{C}) \otimes L \\ S^-(P_X \otimes L) &= \wedge^{0,2*+1}(T^*X \otimes \mathbf{C}) \otimes L. \end{aligned}$$

A $U(1)$ -connection A on the determinant line bundle $\mathcal{L} = K_X^{-1} \otimes L^{\otimes n}$ of $P_X \otimes L$ can be written uniquely as $A_X \otimes B^{\otimes n}$ for a $U(1)$ -connection B on L . The Dirac operator

$$\not{D}_A: \Gamma(S^+(P_X \otimes L)) \longrightarrow \Gamma(S^-(P_X \otimes L))$$

associated to the connection A on the determinant line bundle is

$$\sqrt{2} \left(\bar{\partial}_B + \bar{\partial}_B^* \right) : \Omega^{0,2*}(X; L) \longrightarrow \Omega^{0,2*+1}(X; L).$$

Here by $\bar{\partial}_B$ we mean the $(0,1)$ -part of the covariant derivative of B .

III.1.3. The SW Equations. — Let X be a Kähler surface, and let L be a complex line bundle over X . Then $P = P_X \otimes L$ is a Spin^c -structure on X with determinant line bundle $\mathcal{L} = K_X^{-1} \otimes L^2$. A section of $S^+(P)$ is an element $\psi = (\alpha, \beta)$ of $\Omega^0(X; L) \oplus \Omega^{0,2}(X; L)$. The Dirac equation becomes

$$(5) \quad \sqrt{2} \left(\bar{\partial}_B(\alpha) + \bar{\partial}_B^*(\beta) \right) = 0.$$

The SW curvature equation decomposes into two equations:

$$(6) \quad F_A^{0,2} = \frac{\bar{\alpha}\beta}{2}$$

$$(7) \quad \left(F_A^{1,1} \right)^+ = \frac{i}{4} (|\alpha|^2 - |\beta|^2) \omega,$$

where ω is the Kähler 2-form. Of course, since A_X is a holomorphic connection we see that

$$F_B^{0,2} = \frac{1}{2} F_A^{0,2} = \frac{\bar{\alpha}\beta}{4}.$$

III.2. Holomorphic Description of the Seiberg-Witten Moduli Space

Proposition III.2.1. — Let X be a compact connected Kähler surface and let $P = P_X \otimes L$ be a Spin^c -structure on X . The determinant line bundle of P is $K_X^{-1} \otimes L^{\otimes 2}$. Any solution to the SW equations (A, ψ) where $A = A_X \otimes B^{\otimes 2}$ is a $U(1)$ -connection on $K_X^{-1} \otimes L^{\otimes 2}$ and $\psi = (\alpha, \beta)$ is an element of $\Omega^0(X; L) \oplus \Omega^{0,2}(X; L)$ satisfies:

- (1) B determines a holomorphic structure on L .
- (2) α is a holomorphic section of L .
- (3) $\bar{\beta}$ is a holomorphic $(2,0)$ -form on X with values in L^{-1} .
- (4) If the degree of $\mathcal{L}$ (with respect to the Kähler form ω) is ≤ 0 , then $\beta = 0$. If the degree of $\mathcal{L}$ (with respect to the Kähler form ω) is ≥ 0 , then $\alpha = 0$.

Proof. — Applying $\bar{\partial}_B$ to the Dirac equation and taking L^2 -inner product with β , we see that

$$\langle \bar{\partial}_B \bar{\partial}_B(\alpha), \beta \rangle_{L^2} + \langle \bar{\partial}_B \bar{\partial}_B^*(\beta), \beta \rangle_{L^2} = 0.$$

Integrating by parts and using the $(0, 2)$ -curvature equation we see,

$$\left\langle \frac{\bar{\alpha}\beta}{4} \cdot \alpha, \beta \right\rangle_{L^2} + \|\bar{\partial}_B^*(\beta)\|_{L^2}^2 = 0.$$

This gives

$$\frac{1}{4} \|\bar{\alpha}\beta\|_{L^2}^2 + \|\bar{\partial}_B^*(\beta)\|_{L^2}^2 = 0,$$

and hence $F_B^{0,2} = (\bar{\alpha}\beta)/4 = 0$ and $\bar{\partial}_B^*(\beta) = 0$. It follows that B determines a holomorphic structure on L and that $\bar{\beta}$ is a holomorphic $(2, 0)$ form with values in L^{-1} . From the first equation we see that $\bar{\partial}\alpha = 0$, so that α is a holomorphic section of L . Thus, neither α nor β can vanish identically on a non-empty open subset unless that section vanishes identically everywhere. Since $\bar{\alpha}\beta = 0$, it follows that one of α and β is zero. Given this the last item is immediate from the $(1, 1)$ -curvature equation:

$$\deg(\mathcal{L}) = \int_X c_1(\mathcal{L}) \wedge \omega = \frac{-1}{8\pi} \int_X (|\alpha|^2 - |\beta|^2) \omega \wedge \omega. \quad \square$$

Suppose that we have $A = A_X \otimes B^{\otimes 2}$ and $\psi = (\alpha, \beta)$ satisfying all the above listed items. Then the Dirac equation is satisfied and the $(0, 2)$ -part of the SW curvature equation is satisfied. There remains the $(1, 1)$ -part of the SW curvature equation. It turns out that by varying the metric on L , one can find a holomorphic connection B' on L determining the same holomorphic structure as B so that the $(1, 1)$ -part of the curvature equation for $A' = A_X \otimes (B')^{\otimes 2}$ is satisfied. In fact, the solution is unique. Putting all this together we have the following, completely holomorphic description of the moduli space of solutions to the SW equations.

Theorem III.2.2. — *Let X be a compact Kähler surface with $b_2^+(X) > 1$, and let $P_X \otimes L$ be a Spin^c -structure with determinant $\mathcal{L} = K_X^{-1} \otimes L^{\otimes 2}$.*

(1) *Suppose that the degree of $\mathcal{L}$ is negative. Then the moduli space of solutions to the Seiberg-Witten equations is identified with the space of equivalence classes of pairs consisting of a holomorphic structure on L and a non-zero holomorphic section α of L . Two such pairs are equivalent if and only if there is a holomorphic isomorphism between the holomorphic structures carrying one section to a constant multiple of the other.*

(2) *Suppose that the degree of $\mathcal{L}$ is positive. Then the moduli space of solutions to the Seiberg-Witten equations is identified with the space of equivalence classes of pairs consisting of a holomorphic structure on L and a non-zero anti-holomorphic $(0, 2)$ -form with values in L^{-1} . Two such pairs are equivalent if and only if there is a holomorphic isomorphism between the holomorphic structures carrying one $(0, 2)$ -form to a constant multiple of the other.*

(3) Suppose that the degree of $\mathcal{L}$ is zero. Then all solutions to the SW equations have zero spinor field and the moduli space is identified with the moduli space of ASD connections on $\mathcal{L}$.

III.3. Computations in the Kähler Case. — As an immediate corollary of the previous theorem we have:

Corollary III.3.1. — Let X be a Kähler manifold with K_X of positive degree, and let P_X be the natural Spin^c -structure determined by the Kähler structure. Then $SW_X(P_X) = \pm 1$. Similarly, $SW_X(P_X \otimes K_X) = \pm 1$.

Proof. — We are considering the case when L is topologically trivial and when the determinant line bundle K_X^{-1} has negative degree. Thus, for any solution $(A, (\alpha, \beta))$ to the SW equations, we have that $\beta = 0$ and that B determines a holomorphic structure on L . This holomorphic structure has a non-zero holomorphic section α . Since L is topologically trivial, the existence of a non-zero holomorphic section means that L is holomorphically trivial and that α is a constant section. According to the previous theorem, this means that the SW moduli space is a single point. Using the same types of arguments as above it is an easy to show that in fact the linearized complex is exact at this solution so that this point is a smooth point of the moduli space. Hence, the SW invariant of P_X is ± 1 .

In the other case, the determinant line bundle has positive degree and hence $\alpha = 0$ whereas $\bar{\beta}$ is a holomorphic $(2,0)$ -form with values in K_X^{-1} . That is to say $\bar{\beta}$ is a holomorphic section of the trivial bundle. The same arguments as before show that there is a unique solution to the SW equations and that this solution is a smooth point of the moduli space, so that the SW invariant is ± 1 . $\square$

Remark III.3.2. — The Kähler structure on X induces natural orientations on $H^1(X; \mathbf{R})$ and $H_+^2(X; \mathbf{R})$ and hence a homology orientation for X . Using this homology orientation the SW invariant of P_X is 1.

III.3.1. Minimal Surfaces of General Type

Theorem III.3.3. — Let X be a minimal algebraic surface of general type and let P be a Spin^c -structure on X . Then:

- (1) $SW_X(P) = \pm 1$ if P is the Spin^c -structure P_X determined by the complex structure.
- (2) $SW_X(P) = \pm 1$ if $P = P_X \otimes K_X$.
- (3) $SW_X(P) = 0$ if P is not isomorphic to either P_X or $P_X \otimes K_X$.

Proof. — Let X be a minimal surface of general type. The canonical class K_X is numerically effective in the sense that for any holomorphic curve C in X we have $\langle K_X, C \rangle \geq 0$. Also, since a positive multiple of K_X is the hyperplane section of the canonical mapping (which is everywhere defined), we see that $K_X^2 > 0$. It follows from

the fact that K_X is numerically effective that its degree with respect to the Kähler class ω is ≥ 0 . By the Hodge index theorem (see p.80 of [7]), any class of type $(1, 1)$ orthogonal to ω is of non-positive square. This implies that the degree of K_X with respect to ω is positive. Thus, the first two computations following immediately from Corollary III.3.1. Now suppose that $P = P_X \otimes L$ is a Spin^c -structure with the formal dimension of the SW moduli space for P is non-negative. Let $\mathcal{L} = K_X^{-1} \otimes L^{\otimes 2}$ be the determinant line bundle for this Spin^c -structure. The dimension restriction means that $\mathcal{L}^2 \geq K_X^2 > 0$. In particular, by the Hodge index theorem $\mathcal{L}$ cannot be of degree zero with respect to ω .

Let us first consider the case when the degree of this determinant line bundle is negative. Then there is $0 \leq t$ such that $\omega \cdot (K_X + t\mathcal{L}) = 0$. Invoking the Hodge index theorem again we see that

$$0 \geq (K_X + t\mathcal{L})^2 = K_X^2 + 2tK_X \cdot \mathcal{L} + t^2\mathcal{L}^2.$$

It follows that $K_X \cdot \mathcal{L} \leq 0$. The minimum of the expression on the right-hand side obtains for $t = -(K_X \cdot \mathcal{L})/\mathcal{L}^2$ and the minimum value is:

$$K_X^2 - \frac{(K_X \cdot \mathcal{L})^2}{\mathcal{L}^2},$$

which of course is also ≤ 0 . Since L has a holomorphic section, we see that $K_X \cdot L \geq 0$ and hence $0 \geq K_X \cdot \mathcal{L} \geq -K_X^2$. Since $\mathcal{L}^2 \geq K_X^2 \geq -K_X \cdot \mathcal{L}$, and since $\mathcal{L}^2 > 0$, it follows that

$$K_X^2 - \frac{(K_X \cdot \mathcal{L})^2}{\mathcal{L}^2} \geq 0.$$

Thus,

$$K_X^2 - \frac{(K_X \cdot \mathcal{L})^2}{\mathcal{L}^2} = 0,$$

and hence $t = 1$, and $(K_X + \mathcal{L})^2 = 0$. By the Hodge index theorem, this implies that $L^2 = K_X + \mathcal{L}$ is a torsion class. This means that L is a line bundle whose first Chern class is torsion. Since L has a holomorphic section, it follows that L is trivial, i.e., $P = P_X$.

A similar argument works when $\mathcal{L}$ has positive degree. □

Similar arguments handle the case of elliptic or K3 surfaces and yield the following result:

Theorem III.3.4. — *Let X be a minimal surface, not rational or ruled. If X is a K3 surface, then $SW_X(P) = 0$ unless P is isomorphic to P_X in which case $SW_X(P) = \pm 1$. Suppose that X is an elliptic surface such that the first Chern class of K_X is not a torsion class. If P is a Spin^c -structure with determinant line bundle $\mathcal{L}$ and with $SW_X(P) \neq 0$, then $c_1(\mathcal{L})$ is a rational multiple of the fiber of the elliptic fibration. Furthermore, $c_1(\mathcal{L})$ lies on the interval $[-c_1(K_X), c_1(K_X)]$. All integral classes in this interval which are congruent modulo two to $c_1(K_X)$ have non-zero SW invariant.*

Proof (Sketch). — For a minimal elliptic surface X the first Chern class of the canonical class $c_1(K_X)$ is a rational multiple of the class of the generic fiber. Provided that the surface is not rational this multiple is ≥ 0 . Thus, K_X is numerically effective and $K_X^2 = 0$. The argument above applies in this case to show that if there is a SW solution for the Spin^c -structure $P_X \otimes L$ and if the formal dimension of this moduli space is non-negative, then the determinant line bundle $\mathcal{L}$ of $P_X \otimes L$ satisfies $\mathcal{L} \cdot K_X = 0$ and $\mathcal{L}^2 = 0$. This means that $\mathcal{L}$ is also a rational multiple of the generic fiber. $\square$

III.3.2. Non-minimal Surfaces. — The result on minimal algebraic surfaces contained in the last section extend via the blow-up formula (discussed in the next section) to non-minimal surfaces.

Theorem III.3.5. — *Let X' be an algebraic surface whose minimal model X is not rational nor ruled. Suppose that X' is obtained from X by blowing up k distinct points, and let $E_1, \dots, E_k$ be the exceptional curves in X' . Then X' is diffeomorphic to the connected sum of X and k copies of $\overline{\mathbb{CP}}^2$. We can decompose any Spin^c -structure on X' as a connection sum of a Spin^c -structure on X and k Spin^c -structures on $\overline{\mathbb{CP}}^2$. A Spin^c -structure on $\overline{\mathbb{CP}}^2$ is determined by an odd integer — the first Chern class of its determinant line bundle. We denote by P_r the Spin^c -structure on $\overline{\mathbb{CP}}^2$ associated with r . Then $SW_{X'}(P \#_{i=1}^k P_{r_i}) = SW_X(P)$ if every $r_i = \pm 1$ and is zero otherwise.*

III.4. Topological Consequences. — As consequences of the computations above we offer the following as typical topological consequences.

Corollary III.4.1. — *Let X and Y be algebraic surfaces. Suppose that $f: X \rightarrow Y$ is an orientation-preserving diffeomorphism. If X is not rational nor ruled, then the same is true for Y , and in fact X and Y have the same Kodaira dimension. In this case f sends the homology classes of the exceptional curves of X to those of Y , up to sign, and pulls back the canonical class of the minimal model of Y to that of the minimal model of X , again up to sign. In particular, if X is minimal then so is Y and f pulls back the canonical class of Y to plus or minus the canonical class of X .*

Corollary III.4.2. — *Two elliptic surfaces X and X' of positive Euler characteristic are diffeomorphic if and only if they are deformation equivalent in the sense that there is a connected family of elliptic surfaces containing both X and X' .*

Remark III.4.3. — By studying algebraic surfaces with $b_2^+ = 1$ one can also show that a rational or ruled surface is diffeomorphic only to other such surfaces. Thus, the Kodaira dimension of a complex surface is a differentiable invariant.

IV. The Seiberg-Witten Invariants of Symplectic Four-manifolds

In this section we discuss Taubes' work on the Seiberg-Witten invariants of symplectic four-manifolds. Taubes' has established generalizations of several of the results in the previous section for Kähler surfaces to more general symplectic four-manifolds.

IV.0.1. Calibrating Almost Complex Structures. — Given a symplectic $2n$ -manifold (X, ω) there is an almost complex structure $J: TX \rightarrow TX$ which calibrates the symplectic structure in the sense that $g(x, y) = \omega(x, Jy)$ is a riemannian metric for which J is an isometry and in which ω is a self-dual two-form. Furthermore, $\omega^n = (n!) d\text{vol}$.

Fixing such a calibration and induced metric there is the canonical Spin^c -structure $P_{X,J}$ associated to J . Its determinant line bundle is $K_{X,J}^{-1}$, the inverse of the canonical class of the almost complex structure. As we saw in the last section the spin bundles of this Spin^c -structure are

$$\begin{aligned} S^+(P_{X,J}) &= \wedge^{0,2*}(T^*X \otimes \mathbf{C}) \\ S^-(P_{X,J}) &= \wedge^{0,2*+1}(T^*X \otimes \mathbf{C}). \end{aligned}$$

More generally, for the Spin^c -structure $P_{X,J} \otimes L$ the spin bundles are given by

$$\begin{aligned} S^+(P_{X,J} \otimes L) &= \wedge^{0,2*}(T^*X \otimes \mathbf{C}) \otimes L \\ S^-(P_{X,J} \otimes L) &= \wedge^{0,2*+1}(T^*X \otimes \mathbf{C}) \otimes L. \end{aligned}$$

The Levi-Civita connection on X induces one on the bundle of two-forms of X and hence by projection induces a connection on K_X . We let A_X denote the inverse connection on K_X^{-1} . Just as in the Kähler case, it turns out that the Dirac operator $\not{D}_{A_X}: \Gamma(S^+(P_{X,J})) \rightarrow \Gamma(S^-(P_{X,J}))$ is given by

$$\sqrt{2} \left(\bar{\partial} + \bar{\partial}^* \right) : \Omega^{0,2*}(X; \mathbf{C}) \longrightarrow \Omega^{0,2*+1}(X; \mathbf{C}).$$

(Of course, it will not be the case that $\bar{\partial}^2 = 0$, since we do not have an honest complex structure.) Similarly, on the Spin^c -structure $P_{X,J} \otimes L$ with connection $A = A_X \otimes B^{\otimes 2}$ on the determinant line bundle $K_X^{-1} \otimes L^{\otimes 2}$, there is a related formula for the Dirac operator:

$$\not{D}_A = \sqrt{2} \left(\bar{\partial}_B + \bar{\partial}_B^* \right) : \Omega^{0,2*}(X; L) \rightarrow \Omega^{0,2*+1}(X; L).$$

Thus, in this case as well the SW equations become the system given in Equations (5), (6), and (7).

IV.1. Taubes' Perturbation. — Now we specialize to X a symplectic 4-manifold. One is not able to solve directly the SW equations in this context. Rather, Taubes found a perturbation of these equations which can be solved (at least in the limit as a parameter goes to ∞). The Taubes' perturbation affects only the curvature equation. It is:

$$F_A^+ = q(\psi) + F_{A_{X,J}}^+ - \frac{i}{4} \rho^2 \omega$$

where ρ is a real parameter, eventually to be taken arbitrarily large. With these changes the SW curvature equations for $A = A_{X,J} \otimes B^{\otimes 2}$ and $\psi = (\alpha, \beta)$ become

$$\begin{aligned} F_B^{0,2} &= \frac{\bar{\alpha}\beta}{4} \\ (F_B^{1,1})^+ &= \frac{i}{8}(|\alpha|^2 - |\beta|^2 - \rho^2)\omega. \end{aligned}$$

As an indication, let us solve these equations for the Spin^c -structure $P_{X,J}$. We give an argument due to Donaldson (see [3]) which simplified some of Taubes' earlier arguments.

Theorem IV.1.1. — *For all ρ sufficiently large, there is only one solution, up to gauge equivalence to the perturbed SW equations for the Spin^c -structure $P_{X,J}$. This solution has $\beta = 0$ and α a nonzero constant section.*

Proof. — Suppose that $A = A_{X,J} \otimes B^{\otimes 2}$ and $\psi = (\alpha, \beta)$ is a solution to the perturbed equations. The version of the Bochner formula valid in this context is

$$\nabla_B^* \nabla_B(f) = 2\bar{\partial}_B^* \bar{\partial}_B(f) + i(F_B^{1,1} \cdot \omega)f$$

for any function f . Furthermore, for any function f we have:

$$\bar{\partial} \bar{\partial}(f) = N \circ \partial(f) \quad \text{and hence} \quad \bar{\partial}_B \bar{\partial}_B(f) = N \circ \partial(f) + F_B^{0,2} \cdot f,$$

where $N: \Omega^{0,1}(X; \mathbb{C}) \rightarrow \Omega^{0,2}(X; \mathbb{C})$ is the so-called Nujenhuis tensor, which is the obstruction to the integrability of the almost complex structure. Thus,

$$\begin{aligned} & \int_X (|\nabla_B \alpha|^2) d\text{vol} \\ &= \int_X \left(\langle 2\bar{\partial}_B^* \bar{\partial}_B \alpha, \alpha \rangle - \frac{|\alpha|^2 - |\beta|^2 - \rho^2}{4} |\alpha|^2 \right) d\text{vol} \\ &= \int_X \left(-\langle 2\bar{\partial}_B^* \bar{\partial}_B(\beta), \alpha \rangle - \frac{|\alpha|^2 - |\beta|^2 - \rho^2}{4} |\alpha|^2 \right) d\text{vol} \\ &= \int_X \left(-\langle 2\beta, \bar{\partial}_B \bar{\partial}_B(\alpha) \rangle - \frac{|\alpha|^2 - |\beta|^2 - \rho^2}{4} |\alpha|^2 \right) d\text{vol} \\ &= \int_X -\langle 2\beta, N \circ \partial(\alpha) \rangle d\text{vol} - \int_X 2\langle \beta, F_B^{0,2} \cdot \alpha \rangle d\text{vol} + \int_X \frac{|\alpha|^2 |\beta|^2}{4} d\text{vol} \\ &\quad - \int_X \frac{|\alpha|^2 - \rho^2}{4} |\alpha|^2 d\text{vol} \\ &= \int_X -\langle 2\beta, N \circ \partial(\alpha) \rangle d\text{vol} - \int_X \frac{|\alpha|^2 |\beta|^2}{2} d\text{vol} + \int_X \frac{|\alpha|^2 |\beta|^2}{4} d\text{vol} \\ &\quad - \int_X \frac{(|\alpha|^2 - \rho^2)^2}{4} d\text{vol} - \rho^2 \int_X \frac{|\alpha|^2 - \rho^2}{4} d\text{vol} \end{aligned}$$

$$\begin{aligned}
&= \int_X -\langle 2\beta, N \circ \partial(\alpha) \rangle d\text{vol} - \int_X \frac{1}{4} |\alpha|^2 |\beta|^2 d\text{vol} - \int_X \frac{(|\alpha|^2 - \rho^2)^2}{4} d\text{vol} \\
&\quad - \rho^2 \int_X \frac{|\alpha|^2 - \rho^2}{4} d\text{vol}.
\end{aligned}$$

Since B is a connection on a trivial line bundle, we see that $\int_X F_B \wedge \omega = 0$. This translates through the SW equations into

$$\int_X (|\alpha|^2 - \rho^2) d\text{vol} = \int_X |\beta|^2 d\text{vol}.$$

Thus, we have

$$\int_X \left(|\nabla_B \alpha|^2 + \frac{(|\alpha|^2 - \rho^2)^2}{4} + \frac{1}{4} |\alpha|^2 |\beta|^2 + \rho^2 \frac{|\beta|^2}{4} \right) d\text{vol} = - \int_X \langle 2\beta, N \circ \partial(\alpha) \rangle d\text{vol}.$$

Since there is a constant $C > 0$ depending only on X and J such that

$$\left| - \int_X \langle 2\beta, N \circ \partial_B(\alpha) \rangle d\text{vol} \right| \leq C \|\beta\|_{L^2} \cdot \|\nabla_B(\alpha)\|_{L^2},$$

and since

$$C \|\beta\|_{L^2} \cdot \|\nabla_B(\alpha)\|_{L^2} \leq C (\|\beta\|_{L^2})^2 + \frac{1}{2} (\|\nabla_B(\alpha)\|_{L^2})^2,$$

we see that for $\rho \gg 1$, the above inequality implies $\nabla_B(\alpha) = 0$, $|\alpha|^2 - \rho^2 = 0$, and $\beta = 0$. It follows that $F_B^+ = 0$. Since B is a connection on a topologically trivial bundle, it follows that B is a flat connection. Of course, α is covariantly constant under B . This means that the flat connection B is in fact trivial and α is a constant section of norm ρ . All such are gauge equivalent, so in fact there is exactly one solution to the perturbed SW equations for L trivial and ρ sufficiently large. $\square$

As in the case of Kähler manifolds, one shows directly that the above solution is a smooth point in its moduli space. Thus, we have

Corollary IV.1.2. — *If X is a symplectic manifold with $b_2^+(X) > 1$, then $SW_X(P_{X,J}) = \pm 1$. Similarly, $SW_X(P_{X,J} \otimes K_X) = \pm 1$.*

Remark IV.1.3. — Notice that the same argument would work as long as $\deg L \leq 0$. Thus, any Spin^c -structure P whose determinant line bundle has degree (measured using ω) no larger than that of $P_{X,J}$ has trivial SW invariant unless $P = P_{X,J}$. By using a symmetry in the equations (or by directly considering the case when $\rho^2 \mapsto -\infty$) one can show that all Spin^c -structures on X with non-trivial SW invariants have determinant line bundles whose degrees (measured using ω) lie between $\pm$ the degree of K_X . At the extremes the only Spin^c -structures with non-zero SW invariants are $P_{X,J}$ and $P_{X,J} \otimes K_X$.

Taubes went on to show, by quite delicate and complicated estimates, a much more general result. As $\rho \mapsto \infty$ solutions $(B, (\alpha, \beta))$ to the perturbed SW equations for the Spin^c -structure $P = P_{X,J} \otimes L$ behave as follows: $\beta \mapsto 0$ and $\alpha^{-1}(0)$ converges to a pseudo-holomorphic curve in the homology class of $c_1(L)$. Thus, as $\rho \mapsto \infty$

the SW moduli space becomes identified with the Gromov moduli space of pseudo-holomorphic curves in the appropriate homology class. As a consequence, one has:

Theorem IV.1.4. — *Let X be a symplectic manifold with calibrating almost complex structure J . Suppose that $b_2^+(X) > 1$. Then the SW invariant of the Spin^c -structure $P = P_{X,J} \otimes L$ is identified with the Gromov invariant of (which counts the number of pseudo-holomorphic curves in) the homology class Poincaré dual to $c_1(L)$.*

There is also a variant of this theorem in the case of symplectic manifold with $b_2^+ = 1$. This identifies the Gromov invariant of $c_1(L)$ with the SW invariant of $P_{X,J} \otimes L$ computed in an appropriate chamber (the one determined by the perturbation for all $\rho \gg 1$).

As a consequence of this result Taubes showed:

Theorem IV.1.5. — *Let ω be a symplectic structure on $\mathbf{CP}^2$ compatible with the usual orientation. Then the Gromov invariant of the generator of $H_2(\mathbf{CP}^2; \mathbf{Z})$ is ± 1 . In particular, there is a pseudo-holomorphic S^2 embedded in $\mathbf{CP}^2$ representing the generator of H_2 .*

This gives a very nice generalization of Yau's uniqueness theorem [29] for Kähler structures on $\mathbf{CP}^2$.

Corollary IV.1.6 (Taubes [25]). — *Every symplectic structure on $\mathbf{CP}^2$ is symplectomorphic to the standard Kähler structure.*

Remark IV.1.7. — There is a quotient of the ball in $\mathbf{C}^2$ by a discrete subgroup of the complex hyperbolic group which is a Kähler surface with the same homology $\mathbf{CP}^2$. It is of course not symplectomorphic to $\mathbf{CP}^2$ since it is not simply connected. It is still an open question whether a symplectic manifold homotopy equivalent to $\mathbf{CP}^2$ is symplectomorphic to it.

Remark IV.1.8. — Using the identification of Gromov invariants and SW invariants and the blow-up formula (to be discussed in the next section), Taubes also showed that if X' is a symplectic manifold which is a connected sum of X and $\overline{\mathbf{CP}}^2$, then there is a symplectic S^2 representing the homology class of the exceptional curve. From this it is easy to see that X' is in fact a symplectic blow-up of X .

V. Other Results

In this section we survey various other results about the SW invariants.

V.1. Contact Structures. — A *contact structure* on an oriented 3-manifold is an oriented 2-plane field such that if α is a one-form defining the plane field then $\alpha \wedge d\alpha > 0$ at every point. There is the obvious notion of homotopy classes or deformation classes of contact structures on M . A contact structure on M^3 is said to be *symplectically fillable* if there is a compact symplectic four manifold (X, ω) and

a diffeomorphism $\varphi: M \rightarrow \partial X$ with the property that $\varphi^*(\omega)$ is positive on every two-plane in the contact structure. It is known that the homotopy classes of contact structures on a 3-manifold M are in one-to-one correspondence with $\mathbf{Z}$ cross the set of Spin^c -structures on M (compare Lemma 2.3 of [11]). Kronheimer-Mrowka [11] established a relative version of Taubes' theorem which applies to symplectic manifolds with contact boundaries. Using the basic compactness result for the space of solutions, they showed:

Theorem V.1.1 (Kronheimer-Mrowka [11]). — *Let M be an oriented 3-manifold. There are at most finitely many homotopy classes of contact structures on M that contain a symplectically fillable contact structure.*

V.2. Three-manifolds. — There is a variant of Seiberg-Witten theory for three-manifolds. It is related to Seiberg-Witten theory the way Floer theory is related to Donaldson's ASD theory. Let M be a closed oriented 3-manifold and let P be a Spin^c -structure on M . The SW equations are for a $U(1)$ -connection A on the determinant line bundle $\mathcal{L}$ of P and a section ψ of the bundle of spinors $S(P)$. The equations are:

$$\begin{aligned}\not{D}_A(\psi) &= 0 \\ F_A &= q(\psi) = \psi \otimes \overline{\psi} - \frac{1}{2}|\psi|^2 \text{Id}.\end{aligned}$$

In the second equation we have identified via Clifford multiplication the purely imaginary two-forms on M with the self-adjoint, traceless endomorphisms of $S(P)$. These equations are elliptic modulo the group of changes of gauge (which as before are all maps of M to S^1). Since the manifold is odd dimensional the formal dimension of the moduli space is zero, but quite often the moduli spaces are positive dimensional. As one varies the metric (and perturbations) the moduli spaces vary smoothly, except at the singular connections. A generic perturbation of these equations leads, exactly as in the usual Floer case to a chain complex whose homology is called the Seiberg-Witten Floer homology. (Care must be taken in dealing with the singular connections.)

V.2.1. Computations. — The only solution to the SW equations over S^3 are the trivial connection and the zero spinor field. The reason is that S^3 has a metric of positive curvature, and as in the 4-manifold case, applying the Bochner formula shows that in this case the equations have only singular solutions.

In the case when the three-manifold is the product of a surface Σ_g of genus $g > 1$ with S^1 , the only Spin^c -structures for which the SW equations have solutions are those whose determinant line bundle has first Chern class Poincaré dual to a multiple m times the fundamental class of S^1 where m is an even integer between $2 - 2g$ and $2g - 2$. In [24] it was shown that the moduli space of solutions P with $c_1(\det(P))$ Poincaré dual to $m[S^1]$, $m \neq 0$ is the $(2g - 2 - |m|)$ -fold symmetric product of Σ_g . In particular, in the two extremal cases, $m = 2 - 2g$ and $m = 2g - 2$ the moduli space is a single point.

If X is a three-manifold obtained by doing 0-surgery on a knot $K \subset S^3$, then the homology of X is the same as that of $S^2 \times S^1$. The determinant line bundles of Spin^c -structures on X are the Poincaré duals of even multiples of the fundamental class of the circle. Let $t \in H^2(X; \mathbf{Z})$ be the generator. In [18] it was shown that the Euler characteristics a_n of the moduli spaces for the Spin^c -structure associated to $2nt$, $n \in \mathbf{Z}$, are the coefficients a_n of the Alexander polynomial of the knot K (the version of the Alexander polynomial balanced about zero). It follows from these computations that the SW invariants of $X \times S^1$ are non-zero only for Spin^c -structures whose determinant line bundle is an even multiple of $t \in H^2(X \times S^1; \mathbf{Z})$, and that the SW invariant of $2nt$ is a_n . In particular, $X \times S^1$ can support a symplectic structure only if the Alexander polynomial of K is monic. The only general case when $X \times S^1$ is known to support a symplectic structure is when K is a fibered knot. In this case, as is well-known, the Alexander polynomial is monic.

V.3. The Connected Sum Theorem. — Since the only solution of the SW equations on S^3 is the trivial solution, if we have a four-manifold X which is a connected sum $A \# B$ and we choose a family of metrics on X in which the connected sum tube gets longer and longer, it turns out that the solutions converge to solutions on A and B . In this way we can associate to each solution on the connected sum a pair of solutions one on A and one on B . Conversely, for any pair of solutions – one on A and one on B , we can glue them together to form a solution on the connected sum, once the length of the connected sum tube is sufficiently long. There is an extra parameter in this process – the gluing parameter – which is the automorphism group of the trivial solution over S^3 , which is S^1 . As long as A and B have $b_2^+ \geq 1$, we can arrange that there are no singular solutions for A and for B . In this case we can then describe the moduli space for the connected sum as a circle bundle over the product of the moduli spaces for A and B . It is clear that the first Chern class μ used in computing the SW invariant pulls back from the product of the moduli spaces on A and B (and indeed from one of these factors) and thus, vanishes on the moduli space for the connected sum. This sketches a proof of the following connected sum theorem:

Theorem V.3.1. — *Let X be a connected sum $X = A \# B$ with $b_2^+(A)$ and $b_2^+(B)$ both at least one. Then $SW_X(P) = 0$ for all Spin^c -structures P on X .*

Putting this together with the earlier results for algebraic surfaces and symplectic manifolds, we see:

Corollary V.3.2. — *An algebraic surface or more generally a symplectic four-manifold can not be written as a connected sum $A \# B$ with $b_2^+(A)$ and $b_2^+(B)$ both greater than zero.*

V.4. The Blow-up Formula. — There is a variant of these results when $b_2^+(B) = 0$. Let us state it in the case $X \# \overline{\mathbf{CP}}^2$. Let e be the homology class of the exceptional

curve and let P_{ne} be the Spin^c -structure on $\overline{\mathbf{CP}}^2$ with determinant line bundle ne (here n is odd.)

Theorem V.4.1. — *Let P be a Spin^c -structure on X with $d(P) = 2k$. Then*

$$SW_{X\#\overline{\mathbf{CP}}^2}(P\#P_{ne}) = \pm SW_X(P)$$

for all n odd with $(n^2 - 1)/4 \leq d(P)$. All other Spin^c -structures on $X\#\overline{\mathbf{CP}}^2$ have zero SW-invariant.

The special case when X is of SW simple type in the sense that the only non-zero SW invariants occur for Spin^c -structures of formal dimension zero (eg symplectic manifolds), the formula simplifies: $SW_{X\#\overline{\mathbf{CP}}^2}(P\#P_{ne}) = 0$ unless $n = \pm 1$ and

$$SW_{X\#\overline{\mathbf{CP}}^2}(P\#P_{\pm e}) = SW_X(P).$$

V.5. Minimal Genus for Embedded Surfaces in Symplectic Manifolds

One especially nice consequence of the non-vanishing of SW-invariants for symplectic manifolds is a natural lower bound for the genus of riemann surfaces in the homology class of a symplectic curve in the manifold. Suppose that X is a symplectic manifold and that $C \subset X$ is a smooth connected symplectic submanifold. Let K_X be the canonical class. Then we have the adjunction formula:

$$K_X \cdot C + C \cdot C = 2g - 2.$$

Theorem V.5.1. — *Let X be a compact symplectic 4-manifold and let $C \subset X$ be a connected symplectic curve in X . If $C' \subset X$ is any smooth connected riemann surface homologous to C , then $g(C') \geq g(C)$.*

This theorem was originally proved in [13] and [21] in the case when $C \cdot C \geq 0$, and then extended to the general case more recently by [24]. Let us sketch the argument in the case $C \cdot C \geq 0$. By blowing-up along C we can assume that $C \cdot C = 0$, and hence that $K_X \cdot C = 2g - 2$. We take any curve C' in the same homology class as C and with the same genus as C . We cut out a tubular neighborhood of C' in X and double the resulting manifold with boundary. This produces a four-manifold which is split along $S^1 \times C'$. We take a metric on this manifold which has a submanifold isometric to $S^1 \times C' \times [0, T]$ for some sufficiently large T . We consider solutions to the SW equations on this manifold for Spin^c -structures which agree with P_X over each copy of $X - C'$. In particular, there is a unique solution on $S^1 \times C'$ for the restricted Spin^c -structure. This means that the moduli space of solutions is the product of the moduli space of solutions on the two sides. Since we know that the SW-invariant of each side is non-zero, it follows that the SW invariant of the resulting doubled manifold is also non-trivial. On the other hand, if we could choose a curve C'' in this homology class with lower genus than C , then we could C' to have the same genus, but to have a trivial handle contained in a ball. This would imply that the doubled

manifold was a connected sum of the form $Y \# S^2 \times S^2$ and hence by the connected sum theorem, it would have trivial SW invariants. This contradiction shows that C is genus minimizing in its homology class.

The general case when $C \cdot C < 0$ is treated by an argument of similar spirit but technically much more complicated. It uses the computations of [22] of the SW moduli spaces for S^1 -bundles over riemann surfaces.

V.6. Manifolds Homotopy Equivalent to the $K3$ -surface. — We end this survey with a result of Fintushel-Stern [5] which shows that there are many, many four-manifolds, even ones homotopy equivalent to a $K3$ -surface.

Let $X \rightarrow S^2$ be an elliptically fibered $K3$ -surface. Let $T^2 \subset X$ be a generic fiber of the fibration. Let $K \subset S^3$ be a knot. We form

$$X(K) = X - (T^2 \times D^2) \cup_{T^3} ((S^3 - K \times D^2) \times S^1),$$

where the union along the three-torus is taken in such a way that the homology of $X(K)$ agrees with that of X . Using a gluing theorem along the three-torus and the Meng-Taubes result about SW-invariants for $(S^3 - K) \times S^1$, Fintushel-Stern [5] showed the following:

Theorem V.6.1. — *Let $X \rightarrow S^2$ be an elliptically fibered $K3$ -surface. Let $K \subset S^3$ be a knot with Alexander polynomial $\sum_n a_n t^n$ (balanced so that $a_n = a_{-n}$). Then $X(K)$ is homotopy equivalent to X . Let $P \rightarrow X(K)$ be a Spin^c -structure. Then the SW invariant of P is zero unless the first Chern class of the determinant line bundle of P is an even multiple of the cohomology class Poincaré dual to a generic fiber. If the first Chern class of the determinant line bundle of P is $2k$ times the cohomology class Poincaré dual to a generic fiber then the SW-invariant is a_k .*

Corollary V.6.2. — *$X(K)$ as above is a symplectic manifold only if the Alexander polynomial of the knot K is monic.*

Corollary V.6.3. — *There are infinitely many differentiable distinct manifolds homotopy equivalent to the $K3$ -surface, infinitely many of which are not symplectic.*

Remark V.6.4. — This shows that the theory of four-manifolds homotopy equivalent to the $K3$ -surface (and in fact almost any algebraic surface containing an elliptic curve of square zero) is as complicated as the theory of Alexander polynomials of knots. Thus, there are lots and lots of four-manifolds some symplectic some not, with highly non-trivial SW-invariants. It seems natural to believe that the complete knot type (not just the Alexander polynomial of the knot) is reflected in the diffeomorphism type of $X(K)$. This leads naturally to the question, “Where are the four-manifold invariants that in this situation detect the other knot invariants of K such as the Jones polynomial, or is the Alexander polynomial of the knot a complete invariant of the diffeomorphism type of $X(K)$?” To answer questions like these requires new ideas

and/or new invariants. The physics of quantum field theory and string theory seem like a fertile ground on which to search for such new invariants.

References

- [1] M. ATIYAH & I. SINGER – Index of elliptic operators, III, *Ann. of Math.* **87** (1968), p. 546–604.
- [2] S. DONALDSON – Polynomial invariants for smooth 4-manifolds, *Topology* **29** (1990), p. 257–315.
- [3] ———, The Seiberg-Witten Equations and 4-manifold topology, *Bull. Amer. Math. Soc. (N.S.)* **33** (1996), p. 45–70.
- [4] S. DONALDSON & P. KRONHEIMER – *The geometry of four-manifolds*, Oxford University Press, 1990.
- [5] R. FINTUSHEL & R. STERN – Knots, Links, and 4-manifolds, *Invent. Math.* **134** (1998), p. 363–400.
- [6] R. GOMPF & A. STIPSITZ – *4-manifolds and Kirby calculus*, Graduate studies in Mathematics, vol. 20, American Mathematical Society, 1999.
- [7] P. GRIFFITHS & J. HARRIS – *Principles of Algebraic Geometry*, Pure and Applied Mathematics Series, John Wiley and Sons, New York, 1978.
- [8] F. HIRZEBRUCH & H. HOPF – Felder von Flächenelementen in 4-dimensionalen Mannigfaltigkeiten, *Math. Ann.* **136** (1958), p. 156–172.
- [9] N. JACOBSON – *Lectures in Abstract Algebra, Vol. II Linear Algebra*, Van Nostrand, New York, 1953.
- [10] P. KRONHEIMER – Embedded surfaces and gauge theory in three and four dimensions, *Surveys in differential geometry*, vol. III, International Press, 1998, p. 243–298.
- [11] P. KRONHEIMER & T. MROWKA – Monopoles and Contact Structures, *Invent. Math.* **130** (1997), p. 209–255.
- [12] ———, Scalar curvature and the Thurston norm, *Math. Res. Lett.* **4** (1997), p. 931–937.
- [13] ———, The genus of embedded surfaces in the projective plane, *Math. Res. Lett.* **1** (1994), p. 797–808.
- [14] B. LAWSON & M.-L. MICHELSON – *Spin Geometry*, Princeton Mathematical Series, vol. 38, Princeton University Press, 1989.
- [15] T.-J. LI & A.-K. LIU – The equivalence between SW and Gr in the case where $b^+ = 1$, *Internat. Math. Res. Notices* (1999), p. 335–345.
- [16] P. LISCA – Symplectic fillings and positive scalar curvature, *Geometry and Topology* **2** (1998), p. 509–525.
- [17] P. LISCA & G. MATIC – Tight contact structures and Seiberg-Witten invariants, *Invent. Math.* **129** (1998), p. 103–116.
- [18] G. MENG & C. TAUBES – SW=Milnor Torsion, *Math. Res. Lett.* **3** (1996), p. 661–674.
- [19] J. MILNOR & R. STASHEFF – *Characteristic Classes*, Ann. of Math. studies, vol. 76, Princeton University Press, 1974.
- [20] J. MORGAN – *The Seiberg-Witten Equations and applications to the topology of smooth four-manifolds*, Mathematical Notes, vol. 44, Princeton University Press, 1996.
- [21] J. MORGAN, Z. SZABO & C. TAUBES – A product formula for the Seiberg-Witten invariant and the generalized Thom conjecture, *J. Differential Geom.* **44** (1996), p. 706–789.

- [22] T. MROWKA, P. OZSVATH & B. YU – Seiberg-Witten monopoles on Seifert fibered spaces, *Comm in Analysis and Geometry* **5** (1997), p. 685–793.
- [23] S. NEMIROVSKII – Complex analysis and differential topology on complex surfaces, *Russian Math. Surveys* **54** (1999), p. 729–752.
- [24] P. OZSVATH & Z. SZABO – The symplectic Thom Conjecture, *Ann. of Math. (2)* **151** (2000), p. 93–124.
- [25] C. TAUBES – More constraints on symplectic forms from Seiberg-Witten invariants, *Math. Res. Lett.* **2** (1995), p. 9–13.
- [26] ———, $SW \Rightarrow Gr$: from the Seiberg-Witten equations to pseudo-holomorphic curves, *J. Amer. Math. Soc.* **9** (1996), p. 845–918.
- [27] ———, The Seiberg-Witten invariant and symplectic forms, *Math. Res. Lett.* **1** (1994), p. 809–822.
- [28] E. WITTEN – Monopoles and 4-manifolds, *Math. Res. Lett.* **1** (1994), p. 769–796.
- [29] S. YAU – Calabi’s conjecture and some new results in algebraic geometry, *Proc. Nat. Acad. Sci. U.S.A.* **74** (1977), p. 1798–1799.

INVARIANTS DE TYPE FINI

par

Pierre Vogel

Résumé. — Il existe de nombreux invariants de Vassiliev (ou de type fini) des nœuds et des entrelacs. L'intégrale de Kontsevich est, en quelque sorte, l'invariant de type fini universel. Il prend ses valeurs dans un module de diagrammes $\mathcal{A}(\Gamma)$, Γ étant l'entrelacs considéré comme une courbe abstraite.

Il existe aussi des théories des invariants de type fini pour les variétés de dimension 3. Elles sont toutes plus ou moins équivalentes. La théorie de Kirby permet de décrire les variétés de dimension 3 à l'aide d'entrelacs. On peut alors extraire de l'intégrale de Kontsevich un invariant de variété de dimension 3. Lorsque la variété est une sphère d'homologie, cet invariant est l'invariant de type fini universel. Il prend ses valeurs dans le module de diagrammes $\mathcal{A}(\emptyset)$.

Abstract (Finite type invariants). — There exist many Vassiliev (or finite type) invariants for knots or links. The Kontsevich integral is, in some sense, the universal finite type invariant. It takes values in a module of diagrams $\mathcal{A}(\Gamma)$, where Γ is the link considered as an abstract curve.

There is also finite type invariants theories for 3-dimensional manifolds. They are essentially equivalent. Using Kirby calculus, every 3-dimensional manifold may be described by links. So it is possible to extract from the Kontsevich integral an invariant of 3-dimensional manifold. If the manifold is a homology sphere, this invariant is the universal finite type invariant of the manifold. It takes values in the module of diagrams $\mathcal{A}(\emptyset)$.

Introduction

L'intégrale de Kontsevich associe à tout entrelacs $\alpha : \Gamma \rightarrow \mathbf{R}^3$ un élément $Z(\alpha)$ appartenant à un module $\widehat{\mathcal{A}(\Gamma)}$. Ce module est le complété d'un module gradué $\mathcal{A}(\Gamma)$ défini par générateurs et relations, les générateurs étant des diagrammes trivalents contenant Γ . Cet invariant est suffisamment général pour vérifier la propriété suivante : si I associe à un nœud ou un entrelacs son polynôme de Jones ou de Kauffman, ou

Classification mathématique par sujets (2000). — 57M27, 57M25.

Mots clefs. — Nœud, entrelacs, variété de dimension 3, théorie de Vassiliev, diagramme trivalent, invariants de type fini.

encore son polynôme HOMFLY, il existe pour toute courbe Γ une forme linéaire explicite f sur $\widehat{\mathcal{A}(\Gamma)}$ tel que I soit la composée de f avec l'intégrale de Kontsevich.

D'autre part, toute variété M compacte connexe orientée sans bord de dimension trois peut être définie comme chirurgie sur la sphère S^3 le long d'un entrelacs $\alpha : \Gamma \rightarrow S^3$. On peut alors définir pour tout entier n et toute courbe Γ , une application ι_n de $\widehat{\mathcal{A}(\Gamma)}$ dans $\widehat{\mathcal{A}(\emptyset)}$ qui ne dépend que de la signature de la forme d'enlacement de M et l'invariant de Le-Murakami-Ohtsuki (LMO) de M est simplement défini par la formule :

$$Z^{\text{LMO}}(M) = \sum_{n \geq 0} \iota_n(Z(\alpha))_n,$$

où $\iota_n(Z(\alpha))_n$ est la composante de degré n de $\iota_n(Z(\alpha))$.

Il existe plusieurs définitions des invariants de type fini des variétés de dimension trois, mais elles sont équivalentes dans le cas des sphères d'homologie entière. De plus l'invariant $Z^{\text{LMO}}(M)$ est dans ce cas l'invariant de type fini universel, de la même façon que l'intégrale de Kontsevich des nœuds est l'invariant de type fini (ou de Vassiliev) universel.

1. Modules de diagrammes

Soit Γ une courbe, c'est-à-dire une variété compacte de dimension 1. On appelle Γ -diagramme un triple (K, f, α) où :

- K est un graphe fini et f est un homéomorphisme de Γ sur un sous-graphe de K
- tous les points de $f(\partial\Gamma)$ sont des sommets univalents de K et tous les autres sommets de K sont trivalents
- α associe à chaque sommet trivalent x de K un ordre cyclique sur l'ensemble des trois arêtes de K issues de x .

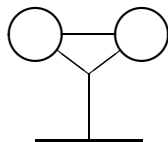
L'adhérence du complémentaire de Γ dans un Γ -diagramme K sera appelée la partie extérieure de K . C'est un diagramme uni-trivalent dont les sommets univalents sont situés sur la courbe Γ .

Pour simplifier, un Γ -diagramme (K, f, α) sera simplement noté K et l'application f considérée comme une inclusion. De plus les Γ -diagrammes seront toujours considérés à isomorphisme près, un isomorphisme de (K, f, α) sur (K', f', α') étant un isomorphisme de graphe φ de K sur K' qui respecte les ordres cycliques et tel que $\varphi \circ f$ soit de la forme $f' \circ \psi$ où ψ est un automorphisme de Γ isotope à l'identité.

Dans la pratique, un Γ -diagramme trivalent sera représenté par un graphe K contenant Γ et immergé dans le plan de telle façon que les ordres cycliques soient donnés par l'orientation standard du plan.

Voici un exemple de Γ -diagramme avec 6 sommets trivalents et deux sommets univalents, Γ étant formé de deux cercles et d'un intervalle. La partie extérieure a

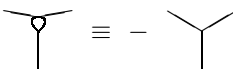
deux composantes connexes, un sommet trivalent et 5 sommets univalents.



Remarque. — Les graphes considérés ici sont tout-à-fait quelconques. Les arêtes ne sont pas orientées. Il peut exister plusieurs arêtes ayant des sommets donnés. Il se peut aussi que le graphe ait une composante homéomorphe à un cercle. En fait les seules contraintes considérées sont locales. Chaque sommet est univalent ou trivalent, c'est-à-dire qu'il part de ce sommet une ou trois arêtes.

1.1. Définition. — Soient k un corps de caractéristique 0 et Γ une courbe. On note $\mathcal{A}(\Gamma)$ le k -module défini par générateurs et relations de la façon suivante : les générateurs sont les Γ -diagrammes et les relations sont :

– la relation d'antisymétrie : si K' est obtenu à partir d'un Γ -diagramme trivalent K par changement d'un ordre cyclique, on a : $K \equiv -K'$.

(AS) 

– la relation IHX (ou relation de Jacobi) : si K , K' et K'' ne diffèrent qu'au voisinage d'une arête de la façon suivante :

$$K : \quad \text{I} \quad K' : \quad \text{H} \quad K'' : \quad \text{X}$$

on a :

(IHX)
$$K \equiv K' - K''$$

Dans la relation IHX, l'arête n'est pas nécessairement extérieure à Γ . Si ce n'est pas le cas, la relation prend la forme suivante :

(STU) 

Cette dernière relation est appelée relation STU dans la littérature.

Un Γ -diagramme a toujours un nombre pair de sommets trivalents. La moitié de ce nombre est appelé le degré du diagramme. Il est facile de vérifier que ce degré induit sur $\mathcal{A}(\Gamma)$ une structure de module gradué.

Il y a des variantes du module $\mathcal{A}(\Gamma)$ selon que l'on considère seulement certains types de Γ -diagrammes. Par exemple, si l'on considère uniquement des Γ -diagrammes K tels que chaque composante connexe de K rencontre Γ , on obtient un sous-module $\mathcal{A}_c(\Gamma)$ de $\mathcal{A}(\Gamma)$. Il est facile de vérifier que $\mathcal{A}_c(\Gamma)$ est de dimension finie en chaque degré. De plus, comme tout Γ -diagramme se décompose de façon unique comme union disjointe d'un $\emptyset$ -diagramme et d'un diagramme dont toute composante rencontre Γ , on obtient la décomposition suivante :

$$\mathcal{A}(\Gamma) = \mathcal{A}_c(\Gamma) \otimes \mathcal{A}(\emptyset).$$

Les modules gradués $\mathcal{A}$ et $\mathcal{A}_c$ ont un certain nombre de bonnes propriétés de fonctorialité par rapport à la courbe.

1.2. Proposition. — *Une inclusion i d'une courbe Γ dans une courbe Γ' induit des morphismes de modules gradués i_* de $\mathcal{A}(\Gamma)$ dans $\mathcal{A}(\Gamma')$ et de $\mathcal{A}_c(\Gamma)$ dans $\mathcal{A}_c(\Gamma')$. Ces morphismes ne dépendent que de la classe d'isotopie de l'inclusion i . Ainsi $\mathcal{A}$ et $\mathcal{A}_c$ sont des foncteurs de la catégorie des courbes et des classes d'isotopie de plongements dans la catégorie des modules gradués.*

1.3. Proposition. — *Soient Γ et Γ' deux courbes. Alors l'union disjointe induit des morphismes de modules gradués de $\mathcal{A}(\Gamma) \otimes \mathcal{A}(\Gamma')$ dans $\mathcal{A}(\Gamma \amalg \Gamma')$ et de $\mathcal{A}_c(\Gamma) \otimes \mathcal{A}_c(\Gamma')$ dans $\mathcal{A}_c(\Gamma \amalg \Gamma')$.*

1.4. Proposition. — *Soit f une application continue d'une courbe Γ dans une courbe Γ' qui envoie bord en bord. Alors f induit des morphismes de modules gradués f^* de $\mathcal{A}(\Gamma')$ dans $\mathcal{A}(\Gamma)$ et de $\mathcal{A}_c(\Gamma')$ dans $\mathcal{A}_c(\Gamma)$. De plus f^* ne dépend que de la classe d'homotopie de $f : (\Gamma, \partial\Gamma) \rightarrow (\Gamma', \partial\Gamma')$. De façon plus précise, $\mathcal{A}$ et $\mathcal{A}_c$ sont des foncteurs homotopiques contravariants de la catégorie des paires $(\Gamma, \partial\Gamma)$, où Γ est une courbe, dans la catégorie des modules gradués.*

Démonstrations. — Si i est une inclusion de Γ dans Γ' l'application i_* envoie un Γ -diagramme K en l'union de K et de Γ' , Γ étant identifiée à son image par i . Par construction l'application ainsi définie passe au quotient et définit une application toujours notée i_* de $\mathcal{A}(\Gamma)$ dans $\mathcal{A}(\Gamma')$ et de $\mathcal{A}_c(\Gamma)$ dans $\mathcal{A}_c(\Gamma')$ qui ne dépend que de la classe d'isotopie de i .

La proposition 1.3 est évidente mais la proposition suivante est plus sérieuse. Pour définir l'application f^* , nous allons procéder par étapes.

Soit K' un Γ' -diagramme. Notons $X(K') = \{u_i\}$ l'ensemble des sommets trivalents de K' situés sur Γ' . On dira que f est transverse à K' si f est différentiable sur un voisinage de $f^{-1}(X(K'))$ et a toutes ses valeurs critiques extérieures à $X(K')$. Nous allons tout d'abord définir $f^*(K')$ lorsque f est transverse à K' .

Soit H la partie extérieure de K' . C'est un graphe dont l'ensemble des sommets univalents sont les points $\{u_i\}$ et K' est l'union, au dessus de $X(K')$, de Γ' et de H . Notons S l'ensemble des fonctions s de $X(K')$ dans Γ telles que $f \circ s$ soit l'identité.

Si s est une fonction de S on peut définir le graphe K_s comme l'union de H et de Γ , les points u_i étant identifiés à leurs images par s dans Γ . Comme f est étale au voisinage de tout point de $s(X(K'))$, les ordres cycliques de K' définissent des ordres cycliques au voisinage de chaque sommet trivalent de K_s et K_s est un Γ -diagramme. On définit alors l'élément $f^*(K')$ comme la somme (dans $\mathcal{A}(\Gamma)$ ou $\mathcal{A}_c(\Gamma)$) des éléments K_s , $s \in S$.

Si f n'est pas transverse à K' , l'application f est homotope à une application différentiable g qui n'a aucun point critique au-dessus des points u_i . On pose alors : $f^*(K') = g^*(K')$.

Si f' est une autre application continue de $(\Gamma, \partial\Gamma)$ dans $(\Gamma', \partial\Gamma')$ homotope à f , et si g et g' sont des applications différentiables homotopes à f et f' et transverses à K' , g et g' sont homotopes par une homotopie différentiable h_t qui est transverse à K' sauf pour un nombre fini de valeurs de t . On peut aussi supposer que si h_t n'est pas transverse à K' , h_t n'a qu'un point critique qui est envoyé en $X(K')$ et que ce point critique est non dégénéré. On dira qu'une telle homotopie est transverse à K' .

Ainsi, pour montrer que $f^*(K')$ est bien définie et ne dépend que de la classe d'homotopie de f , il suffit de montrer que si h_t est une homotopie transverse à K' on a : $h_0^*(K') = h_1^*(K')$. Or $h_t^*(K')$ est définie pour tout t tel que h_t soit transverse à K' . De plus l'application $t \mapsto h_t^*(K')$ est localement constante là où elle est définie. Il reste à montrer que cette fonction n'a aucun saut, mais ceci est une conséquence immédiate de la relation AS. Ainsi $t \mapsto h_t^*(K')$ est constante et $f^*(K')$ est bien définie et ne dépend que de la classe d'homotopie de f .

Pour montrer que f^* est compatible avec les relations AS et IHX, il suffit de le faire lorsque f est différentiable et n'a aucune valeur critique située là où la géométrie des diagrammes change. Dans ce cas f^* envoie une relation AS en une somme de relations AS et une relation IHX en une somme de relations IHX. Le reste de la proposition 1.4 est alors facile à vérifier. $\square$

Corollaire. — Pour toute courbe Γ , $\mathcal{A}_c(\Gamma)$ est une coalgèbre cocommutative graduée.

Démonstration. — Soit f la première projection de $\Gamma \times \{0, 1\}$ vers Γ . Le morphisme f^* envoie $\mathcal{A}_c(\Gamma)$ en $\mathcal{A}_c(\Gamma \times \{0, 1\})$. Considérons le sous-module M de $\mathcal{A}_c(\Gamma \times \{0, 1\})$ engendré par tous les diagrammes ayant une composante qui rencontre à la fois $\Gamma \times \{0\}$ et $\Gamma \times \{1\}$. Il est facile de vérifier que le produit de $\mathcal{A}_c(\Gamma \times \{0\}) \otimes \mathcal{A}_c(\Gamma \times \{1\})$ dans $\mathcal{A}_c(\Gamma \times \{0, 1\})$ induit un isomorphisme ε de $\mathcal{A}_c(\Gamma) \otimes \mathcal{A}_c(\Gamma)$ sur $\mathcal{A}_c(\Gamma \times \{0, 1\})/M$. La diagonale Δ est alors l'application $\varepsilon^{-1} \circ f^*$. On vérifie aisément que $\mathcal{A}_c(\Gamma)$ munie de cette application Δ est une coalgèbre graduée cocommutative. $\square$

Remarque. — Ce corollaire n'est pas valable pour le module $\mathcal{A}(\Gamma)$. Par contre, on vérifie que $\mathcal{A}(\Gamma)$ est une coalgèbre cocommutative graduée sur l'algèbre commutative $\mathcal{A}(\emptyset)$.

1.5. Proposition. — Soit $n > 0$ un entier. Alors $\mathcal{P}_n = \mathcal{A}_c([0, 1] \times \{1, 2, \dots, n\})$ est une algèbre de Hopf cocommutative graduée connexe. Elle est commutative pour $n = 1$.

Démonstration. — La structure de coalgèbre est déjà définie. Le produit est défini grâce à l'application i de $[0, 1] \times \{0, 1\}$ dans $[0, 1]$ qui est croissante par rapport au paramètre de $[0, 1]$ et qui envoie $[0, 1] \times \{0\}$ en $[0, 1/2[$ et $[0, 1] \times \{1\}$ en $]1/2, 1]$:

$$\mathcal{A}_c([0, 1]) \otimes \mathcal{A}_c([0, 1]) \longrightarrow \mathcal{A}_c([0, 1] \times \{0, 1\}) \xrightarrow{i_*} \mathcal{A}_c([0, 1]).$$

L'antipode est plus compliquée à définir. On la construit par récurrence sur le degré et l'on vérifie que l'on a ainsi une structure d'algèbre de Hopf sur $\mathcal{P}_n$. Le fait que $\mathcal{P}_1$ (notée également $\mathcal{A}$ dans la littérature) soit commutative est montrée par exemple dans [BN]. $\square$

2. L'intégrale de Kontsevich

Initialement, l'intégrale de Kontsevich a été construite pour associer à tout nœud $\alpha : S^1 \rightarrow \mathbf{R}^3$ un élément $Z(\alpha)$ du complété d'un quotient de $\mathcal{A}_c(S^1)$. Elle a été ensuite étendue à tout entrelacs et, plus généralement, à tout tangle. Considérons tout d'abord le cas des tresses qui sont des tangles particuliers.

Le cas des tresses. — Soit T une tresse à n brins. On peut considérer cette tresse comme l'image d'un plongement α de $[0, 1] \times \{1, \dots, n\}$ dans $\mathbf{R}^2 \times [0, 1] = \mathbf{C} \times [0, 1]$ tel que, pour tout $i = 1, \dots, n$ et tout $t \in [0, 1]$, $\alpha(t, i)$ soit de la forme $(\beta_i(t), t)$. Les fonctions β_i sont différentiables de $[0, 1]$ dans $\mathbf{C}$. On note $\omega(t)$ la forme différentielle de degré 1 définie par :

$$\omega(t) = \sum_{1 \leq i < j \leq n} h_{ij} \frac{d\beta_i(t) - d\beta_j(t)}{\beta_i(t) - \beta_j(t)},$$

où h_{ij} est le diagramme obtenu en rajoutant une arête à $[0, 1] \times \{1, \dots, n\}$ attachée en les i -ième et j -ième brins et en prenant des ordres cycliques équivalents (*via* l'involution $(t, i) \leftrightarrow (t, j)$). La forme $\omega(t)$ est à coefficients dans l'algèbre $\mathcal{P}_n$ (définie sur le corps des complexes).

$$h_{ij} = \begin{array}{c} \left| \cdots \bigcirc \cdots \right| \\ \begin{array}{cccc} & & & \\ 1 & i & j & n \end{array} \end{array}$$

L'intégrale de Kontsevich de T est alors définie par la série suivante :

$$Z(T) = \alpha_* \sum_{p \geq 0} (2i\pi)^{-p} \int_{\Delta_p} \omega(t_1) \wedge \cdots \wedge \omega(t_p),$$

où Δ_p est le simplexe $0 \leq t_1 \leq \cdots \leq t_p \leq 1$. Cette intégrale appartient au module complété $\widehat{\mathcal{A}_c(T)}$ (défini sur $\mathbf{C}$), où T est considéré comme une courbe abstraite.

2.1. Proposition. — *Si deux tresses sont isotopes (à extrémités fixes), elles ont même intégrale de Kontsevich. De plus, si T_1 et T_2 sont deux tresses composables, on a un produit de $\mathcal{A}_c(T_1) \hat{\otimes} \mathcal{A}_c(T_2)$ dans $\mathcal{A}_c(T_1 T_2)$ défini par l'union des diagrammes et l'on a :*

$$Z(T_1 T_2) = Z(T_1)Z(T_2).$$

Esquisse de démonstration. — Si l'on considère une tresse comme un chemin dans l'espace de configuration X_n formé des n -uples de n points distincts dans $\mathbf{C}$, $\omega(t)$ apparaît comme l'image réciproque par le chemin d'une forme différentielle ω_n définie sur X_n . Cette forme différentielle, à valeurs dans l'algèbre $\mathcal{P}_n$ est en fait une connexion plate appelée connexion de Knizhnik-Zamolodchikov. Vu sous cet angle, l'intégrale $Z(T)$ apparaît comme la monodromie de cette connexion. D'autre part, on vérifie les deux égalités :

$$d\omega = 0, \quad \omega \wedge \omega = 0,$$

ce qui implique que la connexion est plate et l'intégrale ne dépend que de la classe d'homotopie (à extrémités fixes) du chemin dans l'espace de configuration X_n et $Z(T)$ est invariant par isotopie horizontale.

Le fait que Z soit compatible avec le produit est une simple application de la formule de Fubini. $\square$

Ces propriétés indiquent que Z est un foncteur entre deux catégories. La catégorie source est la catégorie des tresses $\mathcal{B}$. Les objets de $\mathcal{B}$ sont les parties finies de $\mathbf{C}$. Un morphisme dans $\mathcal{B}$ entre deux parties finies P et Q de $\mathbf{C}$ est une tresse, définie à isotopie horizontale près, contenue dans $\mathbf{C} \times [0, 1]$ et rencontrant $\mathbf{C} \times \{0\}$ en $P \times \{0\}$ et $\mathbf{C} \times \{1\}$ en $Q \times \{1\}$. La composition est le produit des tresses.

La catégorie but est plus compliquée à définir. Considérons la catégorie $\mathcal{D}$ dont les objets sont les ensembles finis et où un morphisme dans $\mathcal{D}$ d'un ensemble fini P vers un ensemble fini Q est un couple (Γ, u) , où Γ est un cobordisme de P vers Q , c'est-à-dire une courbe dont le bord est $P \amalg Q$, et u un élément de $\mathcal{A}(\Gamma)$. On considère de plus que deux couples (Γ, u) et (Γ', u') sont équivalents s'il existe un isomorphisme φ de Γ sur Γ' qui soit l'identité sur les bords et tel que : $u' = \varphi_*(u)$. La composition des cobordismes et l'union des diagrammes induisent la composition dans $\mathcal{D}$. De la même façon, on a les catégories $\mathcal{D}_c$, $\hat{\mathcal{D}}$ et $\hat{\mathcal{D}}_c$ obtenues en remplaçant $\mathcal{A}()$ par $\mathcal{A}_c()$ ou $\hat{\mathcal{A}}()$ ou encore $\hat{\mathcal{A}}_c()$.

On a donc :

2.2. Proposition. — *L'intégrale de Kontsevich est un foncteur de la catégorie des tresses $\mathcal{B}$ dans la catégorie $\hat{\mathcal{D}}_c$.*

La catégorie $\hat{\mathcal{D}}_c$ contient comme morphismes les algèbres $\mathcal{P}_n$. Elle contient aussi les groupes symétriques car une bijection d'un ensemble fini dans lui-même peut être vue comme un cobordisme Γ de X vers X et $(\Gamma, 1)$ est bien un automorphisme dans

la catégorie $\widehat{\mathcal{D}_c}$. En fait si Γ est un cobordisme d'un ensemble X vers un ensemble Y , on peut le considérer, *via* la correspondance $\Gamma \mapsto (\Gamma, 1)$, comme un morphisme de $\widehat{\mathcal{D}_c}$.

Le cas des tangles. — La catégorie des tresses est une sous-catégorie de la catégorie des tangles. Le foncteur Z n'a pas de prolongement à cette catégorie. Mais c'est le cas si l'on considère la catégorie des tangles munis d'un champ de vecteurs.

2.3. Définition. — On appelle tangle en bande ou b-tangle une sous-variété compacte Γ de dimension 1 de $\mathbf{C} \times [0, 1]$ munie d'un champ de vecteurs V possédant les propriétés suivantes :

- le bord de Γ est l'intersection $\Gamma \cap \mathbf{C} \times \{0, 1\}$,
- V est transverse à Γ , c'est-à-dire que $V(x)$ est, pour tout $x \in \Gamma$, un vecteur de $\mathbf{C} \times \mathbf{R}$ non tangent à la courbe Γ ,
- pour tout $x \in \partial\Gamma$, $V(x)$ est le vecteur $(i, 0)$.

Deux b-tangles (Γ, V) et (Γ', V') sont isotopes s'il existe une isotopie h_t de $\mathbf{C} \times [0, 1]$ dans lui-même qui soit l'identité sur $\mathbf{C} \times \{0, 1\}$, h_0 étant l'identité et $h = h_1$ un difféomorphisme tel que :

h envoie Γ en Γ' et le champ de vecteurs V' est homotope à $dh \circ V \circ h^{-1}$ par une homotopie qui est l'identité sur le bord de Γ' et qui est toujours transverse à Γ' .

Un entrelacs en bande est un b-tangle sans bord qui est donc entièrement contenu dans $\mathbf{C} \times]0, 1[$. Ce n'est autre qu'un entrelacs muni d'un champ de vecteurs transverse.

On représentera graphiquement un b-tangle par un diagramme dessiné dans une bande horizontale du plan, le bord supérieur de la bande correspondant à $\mathbf{R} \times \{0\}$ et le bord inférieur à $\mathbf{R} \times \{1\}$ et on munira le tangle correspondant du champ de vecteur $(i, 0)$.

Un b-tangle peut toujours être épaissi normalement au champ de vecteurs en des bandes orientées. Un tel épaississement sera appelé voisinage en bande du b-tangle. En fait la donnée du champ de vecteur est (à déformation près) équivalente à la donnée de la bande. D'où le nom de tangle en bande ou b-tangle. Si un b-tangle est représenté par un diagramme, la bande est un épaississement du diagramme dans le plan du dessin.

Les b-tangles forment une catégorie $\mathcal{T}$. Les objets de $\mathcal{T}$ sont les parties finies de $\mathbf{C}$. Un morphisme dans $\mathcal{T}$ d'une partie P vers une partie Q est une classe d'isotopie de b-tangles qui rencontrent $\mathbf{C} \times \{0, 1\}$ en $P \times \{0\}$ et $\mathbf{C} \times \{1\}$ en $Q \times \{1\}$. La composition $T \circ T'$ de deux b-tangles T et T' est la juxtaposition de T et de T' , en mettant T en $\mathbf{C} \times [0, 1/2]$ et T' en $\mathbf{C} \times [1/2, 1]$. On met le premier tangle au-dessus du deuxième.

En munissant une tresse du champ de vecteurs $(i, 0)$, on peut considérer toute tresse comme un b-tangle particulier et la catégorie des tresses $\mathcal{B}$ apparaît comme une sous-catégorie de la catégorie $\mathcal{T}$.

2.4. Proposition. — *Le foncteur Z admet un prolongement à la catégorie $\mathcal{T}$.*

Le prolongement n'est pas unique, mais il l'est si l'on veut que Z vérifie certaines conditions. Il est de plus unique si on le restreint aux entrelacs en bande et l'on a un invariant d'entrelacs en bande bien défini qui à chaque entrelacs E associe un élément $Z(E)$ de $\mathcal{A}(E)_c^\wedge$.

Cependant ce foncteur a une grave lacune : il est très difficile à calculer et n'a aucune propriété algébrique, mis à part le fait que c'est un foncteur.

La catégorie but de Z est la catégorie des diagrammes $\mathcal{D}_c^\wedge$. C'est une catégorie monoïdale qui possède une composition, mais aussi un produit tensoriel (l'union disjointe des diagrammes). Par contre, ce n'est nullement le cas de la catégorie des b-tangles $\mathcal{T}$ car il n'y a pas de façon canonique de définir l'union disjointe de deux tangles. Pour remédier à ce problème, on va modifier la catégorie des b-tangles pour en faire une catégorie monoïdale sur laquelle un foncteur monoïdal (comparable à Z) existe.

La catégorie des q-tangles. — Considérons le monoïde non associatif libre L engendré par un objet $\cdot$ appelé point. Tout élément X de L a un degré : le nombre d'occurrences du point dans X . Tout élément de L de degré > 1 s'écrit, de façon unique, comme le produit de deux éléments de L . De plus L possède une unique anti-involution $u \mapsto u^*$ qui envoie le point en lui-même. Les éléments de L seront appelés des q-objets.

Un q-objet est simplement un ensemble fini muni d'un parenthésage. Par exemple, il y a 5 q-objets de degré 4 :

$$u = (..)(..) \quad v = .(.(.)) \quad v^* = ((..).) \quad w = (.(.)). \quad w^* = .((..).)$$

À tout q-objet X on peut associer la partie $\hat{X}$ de $\mathbf{C}$ formé des entiers $1, 2, \dots, n$ correspondant aux n points de X . On définit alors la catégorie des q-tangles $\mathcal{T}'$ dont les objets sont les q-objets et les morphismes entre deux q-objets X et Y sont les morphismes dans la catégorie $\mathcal{T}$ de $\hat{X}$ vers $\hat{Y}$. En quelque sorte, la catégorie des q-tangles a les mêmes morphismes que la catégorie des b-tangles, mais ses objets sont les objets $\{1, 2, \dots, n\} \subset \mathbf{C}$ de $\mathcal{T}$ munis de parenthésages.

Cette catégorie est une catégorie monoïdale. Le produit tensoriel de deux q-objets est le produit de ces objets et le produit tensoriel de deux morphismes, c'est-à-dire de deux b-tangles est leur juxtaposition, le premier tangle étant mis entièrement à gauche du deuxième. Mais il y a une autre opération algébrique : le dédoublement.

Soient T un q-tangle, c'est-à-dire un morphisme dans $\mathcal{T}'$ d'un q-objet X vers un q-objet Y et Γ une composante connexe de T . Soient enfin $n \geq 0$ un entier et u un q-objet de degré n . Supposons tout d'abord que la composante Γ ne rencontre pas le bord de T . On peut remplacer Γ par n copies parallèles de Γ situées dans une bande voisinage de Γ transverse au champ de vecteurs. On obtient ainsi un nouveau q-tangle T' .

Si Γ joint un point de $\mathbf{C} \times \{0\}$ à un point de $\mathbf{C} \times \{1\}$, on peut procéder ainsi : les deux points du bord de Γ correspondent à un point x de X et un point y de Y . On peut alors remplacer x et y par u et obtenir des q -objets X' et Y' . On remplace ensuite Γ par n copies parallèles de Γ et l'on obtient ainsi un nouveau q -tangle T' .

Si Γ joint deux points de $\mathbf{C} \times \{0\}$ (ou de $\mathbf{C} \times \{1\}$), on peut faire de même, en remplaçant toutefois les deux points dans X (ou de Y), l'un par u et l'autre par u^* . Le q -tangle ainsi construit est appelé un dédoublement de T . Il dépend de la composante Γ , du nombre n et d'un q -objet de degré n .

La catégorie $\mathcal{T}'$ est engendrée, en tant que catégorie monoïdale, par un objet (le point) et les morphismes suivants :

$$X^+ = \begin{array}{c} \diagup \\ \diagdown \end{array}, \quad X^- = \begin{array}{c} \diagdown \\ \diagup \end{array}, \quad u = \bigcup, \quad v = \bigcap, \quad W = \begin{array}{c} \diagup \\ \diagdown \end{array} \parallel \begin{array}{c} \diagdown \\ \diagup \end{array},$$

le dernier morphisme étant le morphisme d'un objet $w_1(w_2w_3)$ vers $(w_1w_2)w_3$ représenté par un tangle trivial. De façon plus précise, on a une présentation par générateurs et relations de la catégorie monoïdale $\mathcal{T}'$, ce qui permet de construire des foncteurs de $\mathcal{T}'$ dans une catégorie monoïdale. En particulier, on obtient le résultat suivant de Drinfeld ([D]) :

2.5. Théorème. — *Il existe un foncteur Z' de catégories monoïdales de la catégorie $\mathcal{T}'$ dans la catégorie des diagrammes $\widehat{\mathcal{D}}_c$ tel que :*

- *pour tout q -tangle T de courbe sous-jacente Γ , $Z(T)$ est de la forme (Γ, u) , u étant un élément de $\mathcal{A}_c(\widehat{\Gamma})$ de terme constant 1 et de type groupe (c'est-à-dire $\Delta(u) = u \otimes u$)*
- *si T' est un dédoublement d'un q -tangle T , si Γ' et Γ sont leurs courbes sous-jacentes et si $Z(T)$ est de la forme (Γ, u) , on a $Z(T') = (\Gamma', f^*(u))$, f étant la projection canonique de Γ' sur Γ .*
- *si $X^\pm$ est un demi-twist positif ou négatif, on a $Z(X^\pm) = \exp(\pm h/2)\sigma$, σ étant la transposition de $\mathcal{S}_2$ et h l'élément suivant de $\mathcal{P}_2$:*

$$h = \begin{array}{c} \diagup \\ \diagdown \end{array} \parallel \begin{array}{c} \diagdown \\ \diagup \end{array} = - \begin{array}{c} \diagdown \\ \diagup \end{array} \parallel \begin{array}{c} \diagup \\ \diagdown \end{array}.$$

Le foncteur Z' n'est pas unique, mais il prend les mêmes valeurs que le foncteur Z sur les entrelacs en bande qui ne sont autre que les endomorphismes du vide dans les deux catégories $\mathcal{T}$ et $\mathcal{T}'$.

Il est caractérisé par les trois éléments :

$$\alpha = Z'(u), \quad \beta = Z'(v), \quad \Phi = Z'(w),$$

où w est l'associateur de $.(.)$ dans $(..)$. représenté par le b-tangle trivial.

$$u = \bigcup, \quad v = \bigcap, \quad w = \begin{array}{c} \diagup \\ \diagdown \end{array} \parallel \begin{array}{c} \diagdown \\ \diagup \end{array}.$$

L'élément Φ également appelé associateur appartient à l'algèbre $\mathcal{P}_3$. Quant aux éléments α et β , on peut les considérer dans l'algèbre $\mathcal{P}_1 = \widehat{\mathcal{A}}$, via les applications de u et v dans $[0, 1]$ qui envoient bord gauche en 0 et bord droit en 1.

L'associateur Φ n'est absolument pas quelconque. Il doit vérifier plusieurs propriétés. On ne connaît actuellement qu'un seul associateur explicitement. On l'obtient à l'aide d'une limite de l'intégrale de Kontsevich, en faisant tendre les points source et les points but vers des configurations limites correspondant aux parenthésages donnés. De plus, cet associateur correspond à un foncteur qui ne vérifie pas toutes les propriétés du théorème : la condition de dédoublement ne marche en effet que pour les tresses. Ceci dit il est montré dans [LM] que tous les associateurs sont conjugués par une transformation de type jauge, et qu'il existe un associateur à coefficients rationnels qui correspond à un foncteur Z' vérifiant toutes les hypothèses du théorème. On peut même supposer que Φ est symétrique, c'est-à-dire qu'il est invariant par la transformation $(t, i) \mapsto (1 - t, 3 - i) \in [0, 1] \times \{1, 2, 3\}$, et qu'il n'a de plus que des termes de degrés pairs.

Les éléments α et β par contre peuvent être choisis comme des éléments quelconques de type groupe, avec comme seules restrictions :

- α et β sont symétriques (invariant par la transformation $t \mapsto 1 - t$ de $[0, 1]$),
- $\alpha\beta(1 \otimes v)\Phi(u \otimes 1) = 1$, l'élément $(1 \otimes v)\Phi(u \otimes 1) \in \mathcal{A}_c((1 \otimes v)w(u \otimes 1))^\wedge$ étant considéré comme appartenant à $\mathcal{A}_c([0, 1])^\wedge = \widehat{\mathcal{A}}$.

Cette dernière formule est due à l'égalité suivante dans la catégorie des q-tangles : $1 = (1 \otimes v)w(u \otimes 1)$.

$$\left| \right. = \left| \right.$$

Cependant, si chaque composante connexe non fermée d'un q-tangle joint un point source à un point but, l'image de ce tangle ne dépend que de Φ et du produit $\alpha\beta$ et, par conséquent, que de Φ . En pratique on choisit : $\alpha = \beta$ et l'on note ν l'élément $\nu = \alpha\beta = ((1 \otimes v)\Phi(u \otimes 1))^{-1}$. On a donc dans ce cas :

$$\alpha = \beta = \nu^{1/2}.$$

L'élément ν n'est autre que l'image par Z (ou Z') du nœud trivial.

3. Mouvements de Kirby

Si E est un entrelacs en bande, on peut le considérer dans la sphère S^3 et faire la chirurgie de S^3 le long de E . Cette opération consiste à enlever pour chaque composante C de E un voisinage tubulaire de C et à recoller un exemplaire de $S^1 \times B^2$, en prenant soin que l'ancien parallèle de la composante (parallèle dans le sens de la

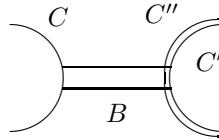
bande) devienne un méridien et le méridien un parallèle. La variété M_E ainsi obtenue est une variété connexe orientée compacte de dimension 3 qui ne dépend (à difféomorphisme orienté près) que de M et de la classe d'isotopie de l'entrelacs en bande E .

On sait que toute variété connexe orientée compacte de dimension 3 peut être obtenue de cette façon et que deux entrelacs en bande E et E' donnent des variétés difféomorphes (par un difféomorphisme direct) si et seulement si on peut passer de E à E' par une suite d'opérations élémentaires appelées mouvements de Kirby.

Le premier mouvement de Kirby consiste à rajouter à l'entrelacs un cercle trivial d'auto-enlacement ± 1 . Le second correspond à un glissement d'anse : si C et C' sont deux composantes de E , on peut remplacer C par une somme connexe de C avec C' et l'on obtient ainsi un nouvel entrelacs en bande E' .

Ainsi, pour obtenir un invariant de variétés de dimension 3, il suffit d'associer à chaque entrelacs en bande E un invariant $I(E)$ qui soit invariant par les deux mouvements de Kirby.

Le deuxième mouvement de Kirby (ou mouvement de type II) n'est pas très bien défini. Pour l'expliciter mieux, on choisit un intervalle C_0 de C et un intervalle C'_0 de C' , puis une bande B dans S^3 joignant C_0 à C'_0 , mais disjointe du reste de l'entrelacs. On supposera que cette bande est, au voisinage de C_0 et C'_0 , contenue dans les voisinages en bande de C et C' . On dédouble C' en rajoutant une copie parallèle C'' située dans le voisinage en bande de C' et rencontrant la bande B en un intervalle C''_0 . En enlevant la partie de la bande B située entre C' et C'' , on obtient une bande B' un peu plus petite que B . La somme connexe est alors l'union $C \cup C'' \cup \partial B'$ privée de l'intérieur de C_0 et C''_0 . On l'appellera la somme connexe de C et C' le long de la bande B .



Considérons l'invariant Z . Cet invariant peut se calculer avec l'intégrale de Kontsevich, mais également avec un foncteur de Drinfeld Z' . Cet invariant n'est pas invariant par rapport au second mouvement de Kirby.

On considère alors un nouvel invariant $\check{Z}$ qui se comporte mieux par rapport au second mouvement de Kirby. Soit E un entrelacs en bande. On peut multiplier $Z(E)$ par ν le long de chaque composante de E . On obtient ainsi un élément $\check{Z}(E)$ appartenant à $\mathcal{A}_c(E)$.

SOCIÉTÉ MATHÉMATIQUE DE FRANCE 2001

L'élément de X associée à un sommet du graphe est appelée sa couleur. Un tel graphe est donc appelé un diagramme X -colorié.

Soit $n \geq 0$ un entier et $[n]$ l'ensemble $\{1, 2, \dots, n\}$. Soit Γ une courbe orientée formée de n composantes, chacune isomorphe à un intervalle. Soit K un diagramme $[n]$ -colorié. Pour chaque $i \in [n]$, on peut attacher les sommets de K de couleur i sur la i -ième composante de Γ et mettre des ordres cycliques compatibles avec l'orientation de Γ . On obtient ainsi un Γ -diagramme. Mais cette construction dépend de la donnée pour chaque i d'un ordre total sur l'ensemble des sommets de K de couleur i . En prenant la moyenne de tous ces diagrammes, on obtient un élément $\varphi(K) \in \mathcal{A}(\Gamma)$. Dans le cas plus général où les composantes de Γ ne sont plus numérotées, on obtient une fonction φ de $\mathcal{B}(\pi_0(\Gamma))$ dans $\mathcal{A}(\Gamma)$. Avec une démonstration relativement semblable à la démonstration du théorème de Poincaré-Birkhoff-Witt, on montre le résultat suivant :

3.2. Proposition ([BN, LMO]). — *Soit Γ une courbe orientée sans composante fermée. Alors l'application $\varphi : \mathcal{B}(\pi_0(\Gamma)) \rightarrow \mathcal{A}(\Gamma)$ est un isomorphisme.*

Lorsque Γ a des composantes sans bord, la situation est différente. On peut toujours définir l'application $\varphi : \mathcal{B}(\pi_0(\Gamma)) \rightarrow \mathcal{A}(\Gamma)$, mais elle n'est plus injective.

Soient n un entier et $[n]^*$ l'ensemble $[n] \coprod \{*\}$. Soit $\mathcal{B}_0([n])$ le sous-module de $\mathcal{B}([n]^*)$ engendré par les graphes $[n]^*$ -coloriés n'ayant qu'un sommet de couleur $*$. Soit i un entier de $[n]$. On peut définir une application f_i de $\mathcal{B}_0([n])$ dans $\mathcal{B}([n])$ de la façon suivante :

Si K est un diagramme $[n]^*$ -colorié n'ayant qu'un sommet x de couleur $*$, on peut identifier x avec un point y de K situé près d'un sommet z de K de couleur i et mettre l'ordre cyclique en ce nouveau sommet en considérant d'abord l'arête de sommet x puis l'arête ne passant pas par z et enfin l'arête yz . En faisant la somme de tous ces graphes on obtient un élément $f_i(K) \in \mathcal{B}([n])$. Soit Γ une courbe orientée avec n composantes Γ_i , $i = 1, \dots, n$. On définit alors le module $\mathcal{C}(\Gamma)$ comme le module $\mathcal{B}([n])$ quotienté par les images des applications f_i pour chaque indice i tel que Γ_i soit une courbe fermée. On vérifie facilement le résultat suivant :

3.3. Proposition. — *Soit Γ une courbe orientée. Alors l'application φ induit un isomorphisme de $\mathcal{C}(\Gamma)$ sur $\mathcal{A}(\Gamma)$.*

Remarque. — On peut aussi définir des modules $B_c()$, $\widehat{\mathcal{B}}()$ et $\widehat{\mathcal{B}}_c()$ et des modules $\mathcal{C}_c()$, $\widehat{\mathcal{C}}()$ et $\widehat{\mathcal{C}}_c()$ et la proposition 3.3 se généralise immédiatement à ces nouvelles situations.

Remarques. — Si X est un ensemble, l'union disjointe des diagrammes induit sur $\mathcal{B}(X)$ une structure d'algèbre. Il y a de plus sur $\mathcal{B}(X)$ une structure de coalgèbre telle que les diagrammes X -coloriés connexes soient primitifs. En fait $\mathcal{B}(X)$ est une

algèbre de Hopf graduée. Si Γ est une courbe avec au moins une composante fermée, le module $\mathcal{C}(\Gamma)$ n'est pas une algèbre, c'est seulement une coalgèbre.

La situation est similaire pour les modules $\mathcal{A}(\Gamma)$. Ceux-ci sont des coalgèbres, mais ils n'ont pas toujours de structure d'algèbre. C'est le cas par exemple si Γ possède une composante fermée.

Cependant, même si $\mathcal{C}(\Gamma)$ et $\mathcal{A}(\Gamma)$ sont tous les deux des algèbres φ n'est pas nécessairement un morphisme d'algèbre. Ce n'est en particulier pas un morphisme d'algèbre dans le cas de l'intervalle. Mais φ est toujours un morphisme de coalgèbre.

Soit f une application d'une courbe Γ vers une courbe Γ' envoyant bord en bord. Cette application induit une application f^* de $\mathcal{A}(\Gamma')$ dans $\mathcal{A}(\Gamma)$. Elle induit aussi une application f^* de $\mathcal{C}(\Gamma')$ dans $\mathcal{C}(\Gamma)$. Cette application envoie un diagramme K en la somme de tous les diagrammes obtenus en relevant les couleurs de K . On vérifie que φ est compatible avec f^* . On a donc aussi une opération de dédoublement pour le foncteur $\mathcal{C}$.

Soit $n \geq 0$ un entier. Soit $\mathcal{A}(n)$ le quotient de l'algèbre $\mathcal{A}(\emptyset)^\wedge$ par l'idéal engendré par $c + 2n$, c représentant le graphe réduit à un cercle. Cette algèbre est isomorphe à la sous-algèbre $\mathcal{A}'(\emptyset)$ de $\mathcal{A}(\emptyset)^\wedge$ formée des diagrammes ne contenant aucun cercle isolé.

Soit Γ une courbe sans bord. Soit u un élément de $\mathcal{C}(\Gamma)$ représenté par un graphe $\pi_0(\Gamma)$ -colorié K . S'il existe un indice i pour lequel K a un nombre de sommet de couleur i différent de $2n$, on pose $\kappa_n(K) = 0$. Sinon on désigne par $\kappa_n(K)$ la somme de tous les diagrammes obtenus en connectant deux sommets univalents de même couleur. On vérifie que κ_n induit une application de $\mathcal{C}(\Gamma)$ dans le module $\mathcal{A}(\emptyset)$. En composant cette application par φ^{-1} d'un côté et par l'application quotient vers $\mathcal{A}(n)$ de l'autre, on obtient une application i_n de $\mathcal{A}_c(\Gamma)^\wedge$ dans $\mathcal{A}(n) \simeq \mathcal{A}'(\emptyset)$.

3.4. Théorème ([LMO]). — *Soit E un entrelacs en bande et E' l'entrelacs obtenu par un mouvement de Kirby de type II, alors les deux éléments $i_n(\check{Z}(E))$ et $i_n(\check{Z}(E'))$ coïncident en degré $\leq n$.*

La démonstration de ce théorème occupera l'essentiel de la partie 4.

3.5. Corollaire. — *Soient E un entrelacs en bande et M la matrice d'enlacement de E . Soit σ_+ (resp. σ_-) le nombre de valeurs propres strictement positives (resp. strictement négatives) de M . Soient U_+ et U_- les deux nœuds triviaux d'auto-enlacement 1 et -1 . Alors l'élément $\Omega_n(E)$, tronqué à l'ordre n de*

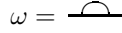
$$\frac{i_n(\check{Z}(E))}{i_n(\check{Z}(U_+))^{\sigma_+} i_n(\check{Z}(U_-))^{\sigma_-}}$$

est un invariant de la variété chirurgisée M_E noté $\Omega_n(M_E)$.

Démonstration. — Déterminons tout d'abord le terme constant de $i_n(\check{Z}(U_\pm))$. L'entrelacs en bande U_+ est représenté par le composé des trois q-tangles suivants :



On en déduit que $\check{Z}(U_+)$ est représenté par l'élément $\nu^2 \exp(\omega/2)$ de $\mathcal{A}_c([0, 1])^\wedge \simeq \mathcal{A}_c(S^1)^\wedge$, ω étant l'élément de $\mathcal{A}_c([0, 1])$ représenté par $[0, 1]$ auquel on a rajouté une arête.



On procède de même pour U_- et l'on voit que $\check{Z}(U_\pm)$ est représenté par $\nu^2 \exp(\pm\omega/2)$.

D'autre part, l'associateur Φ est un élément pair de type groupe de $\mathcal{P}_3$. C'est donc l'exponentielle d'un élément primitif pair u de $\mathcal{P}_3$. Cet élément vérifie $\Delta u = u \otimes 1 + 1 \otimes u$ et est représenté par une série $u = \sum u_n$ où chaque u_n est représenté par un $[0, 1] \times \{1, 2, 3\}$ -diagramme K_n de degré pair tel que la partie extérieure de K_n soit connexe non vide. On en déduit que Φ est représenté par une série de diagrammes K'_n ayant chacun au moins un sommet trivalent extérieur à la courbe $[0, 1] \times \{1, 2, 3\}$. Par définition de ν , on en déduit que ν est aussi représenté par une série de $[0, 1]$ -diagrammes K''_n ayant chacun au moins un sommet trivalent extérieur à $[0, 1]$.

Considérons maintenant l'idéal I de $\mathcal{A}_c([0, 1])^\wedge$ engendré par les diagrammes K ayant au moins un sommet extérieur à $[0, 1]$. On peut faire de même avec des graphes $[1]$ -coloriés et on obtient un idéal J de l'algèbre $\mathcal{B}_c([1])^\wedge$. L'isomorphisme φ de $\mathcal{B}_c([1])^\wedge$ sur $\mathcal{A}_c([0, 1])^\wedge$ passe au quotient et donne un isomorphisme φ' de $\mathcal{B}_c([1])^\wedge / J$ sur $\mathcal{A}_c([0, 1])^\wedge / I$. On vérifie alors la formule :

$$\varphi'^{-1}(\nu^2 \exp(\pm\omega/2)) = \varphi'^{-1}(\exp(\pm\omega/2)) = \exp(\pm\varphi'^{-1}(\omega)/2) = \exp(\pm\omega'/2),$$

où ω' est le graphe $[1]$ -colorié représenté par un intervalle. Notons g le terme de degré 0 de $i_n \circ \check{Z}$. Comme g est nul sur I , on a :

$$g(U_\pm) = g(\exp(\pm\omega/2)) = \frac{(\pm 1/2)^n}{n!} \kappa(\omega'^n),$$

où κ est l'opération fermeture : la somme de toutes les façons d'identifier les sommets univalents deux à deux.

Notons c le cercle dans $\mathcal{A}_c(\emptyset)^\wedge$. On a pour tout p :

$$\kappa(\omega'^p) = c\kappa(\omega'^{p-1}) + 2(p-1)\kappa(\omega'^{p-1}),$$

ce qui donne :

$$\kappa(\omega'^p) = (c + 2p - 2)(c + 2p - 4) \cdots (c + 2)c.$$

On a donc, modulo $c + 2n$:

$$\kappa(\omega'^n) \equiv (-2)(-4) \cdots (-2n) \equiv (-2)^n n!.$$

Cela donne en définitive :

$$g(U_{\pm}) = (\mp 1)^n,$$

et le terme de degré 0 de $i_n(\check{Z}(U_{\pm}))$ est égal à 1 ou à -1 .

Ainsi l'élément $\Omega_n(E)$ est bien défini et ne dépend que du tronqué à l'ordre n de $i_n(\check{Z}(E))$. Le théorème précédent montre que $\Omega_n(E)$ est inchangé par un mouvement de Kirby de type II et grâce à la présence du dénominateur, il est aussi inchangé par un mouvement de Kirby de type I. C'est donc un invariant de la variété M_E . $\square$

3.6. Théorème ([LMO]). — *Il existe un unique invariant Z^{LMO} qui associe à toute variété M orientée compacte connexe sans bord de dimension 3 un élément $Z^{\text{LMO}}(M)$ de $\mathcal{A}'(\emptyset)$ et qui possède les propriétés suivantes :*

- $Z^{\text{LMO}}(M)$ est un élément de type groupe de $\mathcal{A}'(\emptyset)$, c'est donc l'exponentielle d'un élément primitif de $\mathcal{A}'(\emptyset)$ représenté par une combinaison linéaire de diagrammes trivalents connexes ayant au moins un sommet
- pour toute variété M , et tout entier $n \geq 0$ on a :

$$\Omega_n(M) = \sum_{i \leq n} d(M)^{n-i} Z^{\text{LMO}}(M)_i,$$

$Z^{\text{LMO}}(M)_i$ désignant la composante de degré i de $Z^{\text{LMO}}(M)$ et $d(M)$ l'ordre du premier groupe d'homologie entière de M s'il est fini et 0 sinon.

L'invariant LMO Z^{LMO} est maintenant construit. La démonstration de ce dernier résultat sera donnée ultérieurement.

4. Les relations P_n

On a construit plus haut une application κ_n qui va, en particulier, de $\mathcal{B}([1])^{\widehat{}}$ dans $\mathcal{A}(\emptyset)^{\widehat{}}$. Cette application est nulle sur un diagramme dont le nombre de sommets univalents est différent de $2n$. Si un diagramme K a exactement $2n$ sommets univalents, $\kappa_n(K)$ est la somme de tous les diagrammes obtenus en connectant deux à deux les sommets univalents de K .

On peut généraliser cette construction à la situation suivante :

Soit X un ensemble fini. On pose $Y = X \coprod \{*\}$. On a alors une application κ_n de $\mathcal{B}(Y)^{\widehat{}}$ dans $\mathcal{B}(X)^{\widehat{}}$ consistant à effectuer l'opération précédente, mais uniquement pour les sommets de couleur $*$. Si K est un diagramme Y -colorié, $\kappa_n(K)$ est nul si le nombre de sommets de K de couleur $*$ est différent de $2n$. Sinon $\kappa_n(K)$ est la somme de tous les diagrammes obtenus en connectant deux à deux les sommets univalents de couleur $*$ de K .

Cette application est à valeurs dans $\mathcal{B}(X)^{\widehat{}}$. On appelle P_n son image. On appellera également P_n la relation modulo le sous-module P_n . On vérifie que l'application κ_n passe au quotient et on a aussi un sous-module P_n de $\mathcal{C}(\Gamma)^{\widehat{}}$ pour toute courbe Γ .

4.1. Proposition. — Soit Γ une courbe fermée. Soit $n \geq 0$ un entier. Alors tout élément de $\mathcal{C}(\Gamma)^\wedge$ est, modulo P_{n+1} , représenté par une combinaison linéaire de diagrammes ayant, pour chaque composante Γ_i de Γ , au plus $2n$ sommets de couleur i .

Démonstration. — On suppose que Γ a p composantes Γ_i . Les couleurs sont donc les entiers de $[p] = \{1, \dots, p\}$. Soit i une couleur. Soit X un élément de $\mathcal{C}(\Gamma)^\wedge$ représenté par une combinaison linéaire de diagrammes K ayant chacun au moins $2n+1$ sommets de couleurs i .

Notons v le diagramme formé d'un intervalle avec les deux sommets de couleur i . Supposons que X soit de la forme $v^k Y$, pour un certain entier $k \leq n$. Ceci est vrai pour $k = 0$. Pour simplifier l'exposition, on va supposer que Y est représenté par un diagramme (toujours noté Y). Ce diagramme Y a au moins $2n+1-2k$ sommets de couleur i . Soit u le diagramme formé d'un intervalle avec un sommet de couleur i et l'autre de couleur $*$. Choisissons $n+1-k$ sommets de Y de couleurs i et colorions-les par la couleur $*$. Ceci est possible car $n+1-k \leq 2n+1-2k$. On obtient ainsi un diagramme Z . On a :

$$\kappa_{n+1}(u^{n+k+1}Z) = \frac{(n+1+k)!}{(2k)!} (2k-1)(2k-3) \cdots 3 \cdot 1 \cdot v^k Y + v^{k+1} Y',$$

Y' ayant au moins $2n+1-2k-2$ sommets de couleur i avec, de plus, $Y' = 0$ si $k = n$. On peut ainsi, modulo P_{n+1} remplacer $v^k Y$ par un élément divisible par v^{k+1} . On obtient alors, de proche en proche :

$$X = v^0 X \equiv v^1 X_1 \equiv v^2 X_2 \equiv \cdots \equiv v^n X_n \equiv 0.$$

Ainsi tout diagramme ayant au moins $2n+1$ sommets d'une même couleur est équivalent à 0 modulo P_{n+1} , ce qui termine la démonstration. $\square$

La relation P_n peut aussi se définir sur les modules $\mathcal{A}(\Gamma)$ et elle est compatible avec l'isomorphisme φ . On en déduit :

4.2. Proposition. — Soit Γ une courbe fermée. Soit $n \geq 0$ un entier. Alors tout élément de $\mathcal{A}(\Gamma)^\wedge$ est, modulo P_{n+1} , représenté par une combinaison linéaire de diagrammes ayant au plus $2n$ sommets situés sur chaque composante de Γ .

Soit $n \geq 0$ un entier. On note j_n l'application quotient de $\mathcal{A}(n) = \mathcal{A}(\emptyset)^\wedge / (c+2n)$ dans $\mathcal{A}(\emptyset)^\wedge / (c+2n, P_{n+1})$, c désignant le diagramme représenté par un cercle.

4.3. Proposition. — L'application qui à un entrelacs en bande E associe l'élément $j_n \circ i_n \circ \check{Z}(E)$ est invariant par mouvement de Kirby de type II.

Démonstration. — On suppose que le mouvement de Kirby consiste à remplacer une composante C par sa somme connexe C'' avec une composante C' le long d'une bande B . D'après la proposition précédente, l'élément $\check{Z}(E)$ est représenté, modulo P_{n+1} , par une combinaison linéaire de diagrammes ayant au plus $2n$ sommets situés sur chaque composante de E . Mais si un de ces diagrammes a strictement moins de $2n$ sommets sur une composante de E , le diagramme correspondant par φ^{-1} a strictement

moins de $2n$ sommets d'une couleur donnée et son image par i_n est nulle. On peut donc supposer que tous ces diagrammes ont chacun exactement $2n$ sommets situés sur chaque composante de E et, quitte à déplacer les sommets situés sur E , on peut aussi supposer que ces sommets ne sont jamais situés sur $E \cap B$.

Considérons un de ces diagramme K . Soit C'_1 la partie de C' extérieure à B . Le dédoublement $f^*(K)$ est une somme de diagrammes H_p , H_p ayant p sommets situés sur C'_1 et $2n - p$ sommets situés sur la copie C''_1 parallèle à C'_1 . Si p est strictement inférieur à $2n$, $\kappa_n(H_p)$ est nul car il y strictement moins que $2n$ sommets sur C' . Il reste à considérer le cas $p = 2n$. Mais le diagramme H_{2n} n'est autre que le diagramme K . On en déduit que K et $f^*(K)$ ont, modulo P_{n+1} , même image par $\kappa_n \circ \varphi^{-1}$. Comme ceci a lieu pour chacun de ces diagrammes, les deux entrelacs E et E' ont même image par $j_n \circ i_n \circ \check{Z}$ ce qui montre la proposition. $\square$

Pour montrer le théorème 3.4, il reste à montrer le résultat suivant :

4.4. Proposition. — *L'application j_n est un isomorphisme en degré $\leq n$.*

Démonstration. — Le sous-module P_{n+1} de $\mathcal{A}(\emptyset) / (c+2n)$ est l'image de l'application $\kappa_{n+1} : \mathcal{B}([1]) \rightarrow \mathcal{A}(\emptyset) / (c+2n)$. Pour montrer la proposition, il suffit de montrer que si K est un diagramme $[1]$ -colorié ayant une image par κ_{n+1} de degré $\leq n$, cette image est nulle modulo $(c+2n)$.

Soit K un tel diagramme. Par hypothèse il a au plus $2n$ sommets trivalents. On peut supposer qu'il a exactement $2(n+1)$ sommets univalents car sinon son image par κ_{n+1} est nulle. Supposons qu'il existe deux sommets univalents de K qui sont joints par une arête. Alors K est le produit par u d'un diagramme K' , u étant un diagramme formé par un intervalle. On a alors :

$$\kappa_{n+1}(K) = \kappa_{n+1}(uK') = (c+2n)\kappa_n(K')$$

et $\kappa_{n+1}(K)$ est nul modulo $c+2n$.

Supposons maintenant qu'il existe deux sommets univalents de K joints par un chemin ne contenant qu'un seul sommet trivalent de K . Alors, par antisymétrie, l'élément K est nul dans $\mathcal{B}([1])$ ainsi que son image par κ_{n+1} .

Ainsi, si $\kappa_{n+1}(K)$ est non nul, aucune composante connexe de K n'est un arbre et les caractéristiques d'Euler de ces composantes sont négatives ou nulles. On en déduit que la caractéristique d'Euler χ de K est aussi négative ou nulle.

Soient p le nombre de sommets univalents de K , q le nombre de ses sommets trivalents et a le nombre de ses arêtes. En comptant le nombre d'arêtes orientées de K des deux façons possibles, on trouve la formule : $2a = p + 3q$, ce qui implique :

$$\chi = p + q - a = \frac{p - q}{2}$$

et q est supérieur ou égal à p .

Mais on a : $p = 2(n+1)$ et $q \leq 2n < p$, ce qui est contradictoire et ceci achève la démonstration de la proposition ainsi que celle du théorème 3.4. $\square$

4.5. Proposition. — Soient E un entrelacs en bande et $n \geq 0$ un entier. Soient $\delta(E)$ le déterminant de la matrice d'enlacement de E et p le nombre de composantes de E . Alors on a l'égalité suivante dans $\mathcal{A}'(\widehat{\emptyset})$:

$$i_{n+1}(\check{Z}(E)) = (-1)^p \delta(E) i_n(\check{Z}(E)).$$

Démonstration. — Pour définir la matrice d'enlacement de E , il faut numéroter et orienter les composantes de E . On notera Γ_i la i -ième composante de E . Pour tous i et j on notera a_{ij} l'enlacement de Γ_i avec Γ_j . Les nombres a_{ij} sont les coefficients de la matrice d'enlacement. Notons h_{ij} le E -diagramme obtenu en ajoutant à E une arête joignant Γ_i et Γ_j et en munissant ce diagramme des ordres cycliques correspondant à l'orientation de E . On vérifie alors la formule :

$$Z(E) = 1 + \sum_{ij} \frac{a_{ij}}{2} h_{ij} + U,$$

U étant une somme de termes de degrés > 1 . Comme ν n'a pas de terme de degré 1, on a :

$$\check{Z}(E) = 1 + \sum_{ij} \frac{a_{ij}}{2} h_{ij} + U',$$

où U' est une somme de termes de degrés > 1 .

Considérons l'isomorphisme $\varphi : \mathcal{C}_c(E) \widehat{\longrightarrow} \mathcal{A}_c(E) \widehat{\longrightarrow}$. On a :

$$\varphi^{-1}(\check{Z}(E)) = 1 + \sum_{ij} \frac{a_{ij}}{2} u_{ij} + V,$$

u_{ij} étant le diagramme formé par une arête dont les sommets sont de couleurs i et j , et V est une somme de termes de degrés > 1 . Mais $\varphi^{-1}(\check{Z}(E))$ est un élément de type groupe appartenant au module $\mathcal{C}_c(E) \widehat{\longrightarrow}$. Malheureusement ce module n'est pas une algèbre. Cependant on peut prendre un intervalle dans chaque composante de E de façon à avoir une courbe E' contenue dans E et relever $Z(E)$ en un élément de type groupe de $\mathcal{A}_c(E') \widehat{\longrightarrow}$. En appliquant φ^{-1} à cet élément, on voit que l'on a pu relever $\varphi^{-1}(\check{Z}(E))$ en un élément de type groupe de $\mathcal{C}_c(E') \widehat{\longrightarrow} = \mathcal{B}_c([p]) \widehat{\longrightarrow}$ qui est une algèbre de Hopf. Pour simplifier on notera toujours $\varphi^{-1}(\check{Z}(E))$ cet élément. Maintenant cet élément est l'exponentielle d'un élément primitif X de $\mathcal{B}_c([p])$:

$$X = \sum_{ij} \frac{a_{ij}}{2} u_{ij} + X',$$

où X' est une somme de diagrammes connexes ayant au moins un sommet trivalent.

On en déduit :

$$\varphi^{-1}(\check{Z}(E)) = \exp \left(\sum_{ij} \frac{a_{ij}}{2} u_{ij} \right) \sum_{k_1 \dots k_p} Y_{k_1 \dots k_p},$$

$Y_{k_1 \dots k_p}$ étant un diagramme avec au moins un sommet trivalent et, pour chaque $i \leq p$, k_i sommets univalents de couleur i .

Soient $k_1, k_2, \dots, k_p$ des entiers. On a une application $\kappa_{k_1 \dots k_p}$ de $B([p])^\wedge$ dans $\mathcal{A}(\emptyset)^\wedge / (c + 2n)$ qui est nulle en un diagramme pour lequel il existe une couleur i avec un nombre de sommets de cette couleur différent de $2k_i$ et qui envoie sinon un diagramme K en la somme de tous les diagrammes obtenu en connectant deux à deux les sommets de même couleur. L'application i_n n'est autre que $\kappa_{nn \dots n} \varphi^{-1}$.

Pour simplifier la démonstration on va se placer dans le cas où la matrice d'enlacement est diagonale. Voir [LMO] pour une démonstration complète. Les nombres a_{ij} sont donc nuls pour $i \neq j$, ce qui donne :

$$\varphi^{-1}(\check{Z}(E)) = \exp \left(\sum_i \frac{a_{ii}}{2} u_{ii} \right) \sum_{k_1 \dots k_p} Y_{k_1 \dots k_p}.$$

On a :

$$\kappa_{nn \dots n} \left(\exp \left(\sum_i \frac{a_{ii}}{2} u_{ii} \right) Y_{k_1 \dots k_p} \right) = \sum_{n_1 \dots n_p} \frac{\prod (a_{ii}/2)^{n_i}}{(\sum n_i)!} \kappa_{nn \dots n} \left(\prod u_{ii}^{n_i} Y_{k_1 \dots k_p} \right).$$

Cet élément est nul si les indices k_i ne sont pas tous pairs. On suppose donc que l'on a pour tout i : $k_i = 2q_i$. Mais alors la somme se réduit à un seul terme correspondant à $n_i = n - q_i$:

$$\kappa_{nn \dots n} \left(\exp \left(\sum_i \frac{a_{ii}}{2} u_{ii} \right) Y_{k_1 \dots k_p} \right) = \frac{\prod (a_{ii}/2)^{n-q_i}}{(\sum n - q_i)!} \kappa_{nn \dots n} \left(\prod u_{ii}^{n-q_i} Y_{k_1 \dots k_p} \right),$$

ce qui donne :

$$\kappa_{nn \dots n} \left(\exp \left(\sum_i \frac{a_{ii}}{2} u_{ii} \right) Y_{k_1 \dots k_p} \right) = \left(\prod (-a_{ii})^{n-q_i} \right) \kappa_{q_i \dots q_p} (Y_{k_1 \dots k_p}).$$

On procède de même avec l'entier $n+1$ et on obtient la formule :

$$\kappa_{n+1 \dots n+1} \left(\exp \left(\sum_i \frac{a_{ii}}{2} u_{ii} \right) Y_{k_1 \dots k_p} \right) = \left(\prod (-a_{ii}) \right) \kappa_{nn \dots n} \left(\exp \left(\sum_i \frac{a_{ii}}{2} u_{ii} \right) Y_{k_1 \dots k_p} \right).$$

On en déduit aisément la formule souhaitée. $\square$

Démonstration du théorème 3.6. — De la proposition précédente on déduit que, si M est obtenue par chirurgie sur un entrelacs en bande E , on a :

$$\Omega_{n+1}(M) = |\delta(E)| \Omega_n(M)$$

en degré $\leq n$. Mais $|\delta(E)|$ n'est autre que 0, si M n'est pas une sphère d'homologie rationnelle, et l'ordre de $H_1(M; \mathbf{Z})$ sinon. C'est donc le nombre $d(M)$ et l'on a pour tout $i \leq n$:

$$\Omega_{n+1}(M)_i = d(M) \Omega_n(M)_i.$$

On pose alors :

$$Z^{\text{LMO}}(M) = \sum_{n \geq 0} \Omega_n(M)_n$$

et l'on a pour tout n :

$$\Omega_n(M) = \sum_{i=0}^n d(M)^{n-i} \Omega_i(M)_i.$$

Le fait que $Z^{\text{LMO}}(M)$ soit un élément de type groupe résulte du fait que $\check{Z}(M)$ et $\varphi^{-1}(\check{Z}(M))$ sont de type groupe et d'un calcul explicite de $\kappa_{nn\dots n}(\varphi^{-1}(\check{Z}(M)))$. Voir [LMO] pour une démonstration complète. $\square$

5. Propriétés de l'invariant de Le-Murakami-Ohtsuki

5.1. Proposition. — Soient M et M' deux variétés compactes connexes orientées sans bord de dimension 3. Alors on a :

$$Z^{\text{LMO}}(M \# M')_n = \sum_{j+k=n} d(M)^k Z^{\text{LMO}}(M)_j d(M')^j Z^{\text{LMO}}(M')_k.$$

Démonstration. — On peut représenter M et M' comme chirurgies sur des entrelacs E et E' . On peut choisir ces entrelacs dans deux demi-espaces disjoints, l'un au-dessus de l'autre. On a alors :

$$\check{Z}(E \cup E') = \check{Z}(E) \check{Z}(E').$$

On en déduit pour tout n la formule :

$$\Omega_n(M \# M') = \Omega_n(M) \Omega_n(M'),$$

ce qui implique la formule souhaitée. $\square$

Si l'on restreint Z^{LMO} aux sphères d'homologie entière il devient multiplicatif. Mais ce n'est pas le cas pour l'invariant Z^{LMO} lui-même. On peut cependant modifier cet invariant en un invariant $\hat{Z}^{\text{LMO}}$ multiplicatif défini sur les sphères d'homologie rationnelle en posant, pour toute variété M tel que $H_1(M; \mathbf{Z})$ soit fini d'ordre d :

$$\hat{Z}^{\text{LMO}}(M) = \sum_n Z^{\text{LMO}}(M)_n d^{-n}.$$

5.2. Proposition. — Soient M une variété connexe compacte orientée sans bord de dimension 3 et $-M$ la même variété munie de l'orientation opposée. Soit b le premier nombre de Betti de M . Alors on a pour tout n :

$$Z^{\text{LMO}}(-M)_n = (-1)^{n(b+1)} Z^{\text{LMO}}(M)_n.$$

Démonstration. — Soit S l'endomorphisme qui multiplie un élément de degré n d'un module gradué par $(-1)^n$. On vérifie que si E' est l'image miroir d'un entrelacs en bande E on a : $Z(E') = S(Z(E))$. Comme ν est concentré en degrés pairs, on a aussi : $\check{Z}(E') = S(\check{Z}(E))$. Soient p le nombre de composantes de E et n un entier. Comme i_n diminue les degrés de pn , on a : $i_n(\check{Z}(E')) = (-1)^{pn} S(i_n(\check{Z}(E)))$. Soient M et $-M$ les variétés obtenues par chirurgie le long de E et E' . On a alors :

$$\Omega_n(-M) = (-1)^{n(p-\sigma_++\sigma_-)} S(\Omega_n(M)),$$

ce qui implique le résultat car on a : $p = b + \sigma_+ + \sigma_-$. $\square$

Extension au cas des variétés avec entrelacs. — Soient M une variété compacte orientée connexe sans bord de dimension 3 et E un entrelacs en bande contenu dans M . On peut construire la paire (M, E) par chirurgie de S^3 le long d'un entrelacs en bande E' disjoint d'un autre entrelacs en bande E . Soit $\check{Z}(E', E)$ l'invariant $Z(E \cup E')$ modifié par multiplication par ν le long de chaque composante de E' . On peut effectuer les constructions précédentes à partir de $\check{Z}(E', E)$, mais en ne modifiant que l'entrelacs E' . On obtient ainsi des invariants $\Omega_n(M, E)$ appartenant à $\mathcal{A}(\widehat{E})$ ainsi qu'un invariant LMO $Z^{\text{LMO}}(M, E)$ appartenant également à $\mathcal{A}(\widehat{E})$. De façon plus précise ces invariants appartiennent au sous-module $\mathcal{A}'(\widehat{E})$ de $\mathcal{A}(\widehat{E})$ engendré par les diagrammes ne contenant aucun cercle isolé extérieur à E .

Cet invariant généralise l'intégrale de Kontsevich, car si E est un entrelacs en bande on a :

$$Z(S^3, E) = Z(E).$$

6. Invariants de type finis

La théorie des invariants de Vassiliev, également appelés invariants de type fini, des entrelacs est basée sur certaines transformations élémentaires. La transformation la plus simple est la modification de croisements :

$$\begin{array}{c} \diagup \quad \diagdown \\ \diagdown \quad \diagup \end{array} \longleftrightarrow \begin{array}{c} \diagdown \quad \diagup \\ \diagup \quad \diagdown \end{array}$$

Ces deux entrelacs sont les désingularisés d'un entrelacs singulier ayant un point double. Les différences des deux singularisés des entrelacs singuliers ayant un point double engendrent un sous-module I_1 du module $\mathbf{Q}[\mathcal{E}]$ des combinaisons linéaires formelles d'entrelacs. Si l'on considère maintenant des entrelacs singuliers ayant n points doubles les sommes alternées des désingularisés de ces entrelacs engendrent un sous-module I_n de $\mathbf{Q}[\mathcal{E}]$. On obtient ainsi une filtration de $\mathbf{Q}[\mathcal{E}]$. La théorie de Vassiliev donne une description du gradué associé à cette filtration en terme de diagrammes trivalents.

On peut cependant considérer d'autres types de transformations élémentaires. Par exemple la transformation de type Δ :

$$(\Delta) \quad \begin{array}{c} \diagup \quad \diagdown \\ \diagdown \quad \diagup \\ \text{---} \end{array} \longleftrightarrow \begin{array}{c} \diagdown \quad \diagup \\ \diagup \quad \diagdown \\ \text{---} \end{array}$$

Cette transformation donne une autre filtration de $\mathbf{Q}[\mathcal{E}]$ mais celle-ci, à changement d'indices près, est la même que la filtration précédente.

On peut procéder de même avec les variétés de dimension 3. Cependant, les modifications élémentaires que l'on peut considérer sur les variétés donnent souvent une

filtration relativement triviale. C'est par exemple le cas lorsque l'on considère la transformation élémentaire $M_{E_+} \leftrightarrow M_{E_-}$, E_+ et E_- se déduisant l'un de l'autre par une modification de croisement. Cette transformation modifie la variété, mais surtout elle change son homologie. On construit ainsi une filtration du module $\mathbf{Q}[\mathcal{V}]$ engendré par toutes les variétés connexes compactes orientées de dimension 3. On montre que cette filtration ne donne que peu d'information.

La situation est très différente si l'on considère la transformation $M_E \leftrightarrow M_{E'}$, E et E' se déduisant par une transformation de type Δ .

La filtration de Gusarov–Habiro

6.1. Proposition. — *Si E' se déduit d'un entrelacs en bande E par une transformation de type Δ , les variétés chirurgisées M_E et $M_{E'}$ ont des homologies entières isomorphes.*

Démonstration. — Cela résulte du fait que les matrices d'enlacement de E et de E' sont les mêmes. $\square$

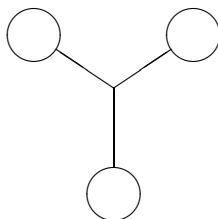
Soit $\mathcal{S}$ l'ensemble des classes d'isomorphisme de sphères d'homologie entière, c'est-à-dire de variétés connexes compactes orientées de dimension 3 ayant le type d'homologie entière de S^3 . La somme connexe induit sur $\mathcal{S}$ une structure de monoïde commutatif unitaire. On note $\mathbf{Q}[\mathcal{S}]$ le module des combinaisons linéaires formelles de sphères d'homologie entière. Ce module est une bialgèbre et les éléments de $\mathcal{S}$ sont de type groupe.

Considérons un diagramme trivalent K , une application α de K dans S^3 et un champ de vecteur V sur $\alpha(K)$ possédant les propriétés suivantes :

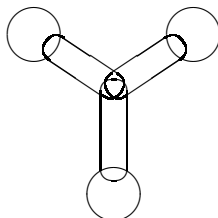
- α est un plongement différentiable sur toutes les arêtes de K ,
- pour tout sommet x de K les trois arêtes de $\alpha(K)$ issues de $\alpha(x)$ ont en ce point des tangentes coplanaires mais deux à deux non colinéaires,
- V est en tout sommet de $\alpha(K)$ non coplanaire au plan tangent au trois arêtes issues de ce sommet et en tout autre point de $\alpha(K)$ non tangent au graphe.

On appellera une telle donnée (K, α, V) un pseudo-diagramme en bande plongé dans S^3 . Un tel diagramme en bande admet un épaississement en une surface orientée transverse au champ de vecteur, ce qui induit un ordre cyclique sur les arêtes de K issues de chaque sommet de K . Un tel épaississement sera appelé épaississement en bande de (K, α, V) . On dira que (α, V) est un plongement en bande du diagramme K , si ces ordres cycliques sont les ordres cycliques donnés par la structure de diagramme de K . On dira dans ce cas que (K, α, V) est un diagramme en bande plongé.

Considérons le diagramme K suivant :



et supposons-le plongé dans S^3 par un plongement en bande. On supposera que les trois cercles de K sont orientés de façon standard. On peut épaissir chacune des trois arêtes de K qui ne sont pas des cercles en une bande orientée contenue dans un épaississement en bande de K . On obtient une courbe orientée formée de 6 cercles immergée dans la sphère :



On peut alors modifier cette courbe immergée en un entrelacs en bande $L(K)$, en prenant des croisements négatifs pour les 3 points doubles situés au bord du triangle central et des croisements positifs pour les 9 autres points doubles.

Près du sommet central de K , l'entrelacs $L(K)$ est un entrelac borroméen qui peut se dénouer par une transformation de type Δ . On appellera Y -diagramme un tel diagramme en bande plongé K . La construction précédente associe à tout Y -diagramme K un entrelacs en bande à 6 composantes $L(K)$. Si maintenant K est une union disjointe de Y -diagrammes, on peut effectuer la construction précédente sur chaque composante de K et l'on obtient un entrelacs en bande $L(K)$. Il est facile de vérifier que la matrice d'enlacement de $L(K)$ est toujours unimodulaire et la variété chirurgisée $M_{L(K)}$ est une sphère d'homologie entière.

Considérons un diagramme en bande plongé K dont chaque composante est un Y -diagramme. On pose :

$$\Sigma(K) = \sum_{K'} (-1)^{\#K'} M_{L(K')},$$

où K' est une union quelconque de composantes connexes de K , $\#K'$ le nombre de composantes de K' et $M_{L(K')}$ la variété obtenue par chirurgie le long de $L(K')$. Cette somme appartient à l'algèbre $\mathbf{Q}[S]$. On note $\mathcal{F}_n$ le sous-module de $\mathbf{Q}[S]$ engendré par les éléments $\Sigma(K)$, où K est un diagramme en bande plongé formé de n composantes qui sont chacune des Y -diagrammes. On montre aisément le résultat suivant :

6.2. Proposition. — *Les sous-modules $\mathcal{F}_n$ forment une filtration de la bialgèbre $\mathbf{Q}[\mathcal{S}]$, compatible avec la structure de bialgèbre, c'est-à-dire que l'on a pour tous p, q et n :*

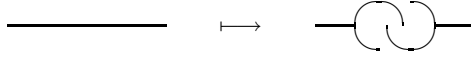
$$\mathcal{F}_p \mathcal{F}_q \subset \mathcal{F}_{p+q} \quad \Delta(\mathcal{F}_n) \subset \sum_{p+q=n} \mathcal{F}_p \otimes \mathcal{F}_q.$$

Remarque. — Cette filtration a été introduite par Gusarov et Habiro. Une autre filtration a été obtenue indépendamment par Garoufalidis et Ohtsuki. Mais celle-ci est équivalente à la filtration de Gusarov–Habiro.

De façon plus précise Gusarov et Habiro ont montré que pour tout n , $\mathcal{F}_{2n-1}$ est égal à $\mathcal{F}_{2n}$. Quant à la filtration $\mathcal{F}'_n$ de Garoufalidis et Ohtsuki, Garoufalidis, Levine et Ohtsuki ([GL1], [GO]) ont montré l'égalité $\mathcal{F}'_{3n-2} = \mathcal{F}'_{3n-1} = \mathcal{F}'_{3n}$ pour tout n . Enfin Gusarov ([Gu2]) a montré l'égalité $\mathcal{F}_{2n} = \mathcal{F}'_{3n}$.

La fonction de poids W . — L'intérêt de la construction de Habiro, est qu'il est facile de construire une application du module $\mathcal{A}'(\emptyset)$ engendré par les diagrammes trivalents n'ayant aucune composante circulaire dans le gradué associé à la filtration $\mathcal{F}_*$.

Considérons un tel diagramme trivalent K de degré n . On peut le plonger dans la sphère avec un plongement en bande et l'on obtient un diagramme en bande plongé K' . On effectue alors la transformation suivante : au voisinage du milieu de chaque arête de K' on modifie K' ainsi :



On obtient ainsi un nouveau diagramme en bande plongé K'' dont chaque composante connexe est un Y -diagramme. Comme K est supposé de degré n , K a $2n$ sommets et K'' a $2n$ composantes. L'élément $\Sigma(K'')$ est donc un élément de $\mathcal{F}_{2n}$. On a le résultat suivant :

6.3. Proposition. — *La construction ci-dessus induit une application bien définie W de $\mathcal{A}'(\emptyset)_n$ dans le quotient $\mathcal{F}_{2n}/\mathcal{F}_{2n+2}$. De plus W est surjective.*

Il faut montrer que $\Sigma(K'')$ est, modulo $\mathcal{F}_{2n+2}$, indépendant du plongement choisi. Il suffit pour cela de comparer $\Sigma(K''_1)$ et $\Sigma(K''_2)$ lorsque K_1 et K_2 sont obtenus l'un à partir de l'autre par une modification de croisement. Il faut ensuite montrer la compatibilité avec les relation AS et IHX. La relation AS est facile à vérifier. L'indépendance par rapport au plongement et la compatibilité avec la relation IHX sont plus délicates, mais il s'agit à chaque fois d'une vérification locale. Le fait que W est surjective est facile à vérifier.

6.4. Théorème. — *L'invariant LMO est l'invariant de type fini universel. C'est-à-dire qu'il induit pour tout $n > 0$ un isomorphisme de $\mathbf{Q}[\mathcal{S}]/\mathcal{F}_{2n+2}$ sur le module $\mathcal{A}'(\emptyset)$*

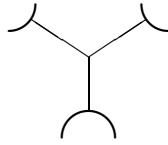
tronqué à l'ordre n . De plus les applications :

$$\mathcal{A}'(\varnothing)_n \xrightarrow{W} \mathcal{F}_{2n}/\mathcal{F}_{2n+2} \xrightarrow{Z^{\text{LMO}}} \mathcal{A}'(\varnothing)_n$$

sont des isomorphismes.

Esquisse de démonstration. — Montrons tout d'abord que l'invariant Ω_n est un invariant d'ordre au plus $2n$, c'est-à-dire qu'il s'annule sur $\Sigma(K)$ dès que K possède au moins $2n + 1$ composantes.

Soit donc K une union de $2n + 1$ Y -diagrammes. Considérons le diagramme en bande plongé K' obtenu en enlevant à K une composante K_0 . On peut dénouer $L(K_0)$ par une transformation de type Δ au voisinage du sommet central de K_0 . On obtient ainsi un nouvel entrelacs en bande $L_0(K_0)$ à 6 composantes qui est une union disjointe de trois entrelacs de Hopf. Le calcul de Kirby montre qu'une chirurgie le long de $L_0(K_0)$ ne modifie pas la variété, ce qui montre que les entrelacs en bande $L(K')$ et $L(K') \cup L_0(K_0)$ donnent la même variété par chirurgie. D'autre part, comme les matrices d'enlacement de $L(K_0)$ et $L_0(K_0)$ sont les mêmes, la différence $\varphi^{-1}(Z(L(K))) - \varphi^{-1}(Z(L(K') \cup L_0(K_0)))$ est une combinaison linéaire de diagrammes $L(K)$ -coloriés ayant chacun au moins un sommet trivalent et il en est de même de $\varphi^{-1}(\check{Z}(L(K))) - \varphi^{-1}(\check{Z}(L(K') \cup L_0(K_0)))$. Un calcul explicite montre que le premier terme non nul de $\varphi^{-1}(\check{Z}(L(K))) - \varphi^{-1}(\check{Z}(L(K') \cup L_0(K_0)))$ est en degré 2. Il est égal à $L(K')$ auquel on a rajouté trois arêtes près du sommet central de K :



Si maintenant K_0 est une union quelconque de composantes connexes de K , on peut effectuer une transformation de type Δ au voisinage de chacun des sommets centraux des composantes de K_0 et l'on obtient un entrelacs en bande plongé $L_0(K_0)$. Notons $L(K, K_0)$ l'union de $L(K - K_0)$ et de $L_0(K_0)$. Cet entrelacs en bande est obtenu à partir de $L(K)$ par des transformations de type Δ sur chaque composante de K_0 . Par itération du calcul précédent, on obtient que l'expression :

$$\sum_{K_0} (-1)^{\#K_0} \varphi^{-1}(\check{Z}(L(K, K_0)))$$

est une combinaison linéaire de diagrammes $L(K)$ -coloriés ayant chacun au moins $2n + 1$ sommets trivalents et la somme

$$\sum_{K_0} (-1)^{\#K_0} i_n(\check{Z}(L(K, K_0)))$$

est aussi une combinaison linéaire de diagrammes ayant au moins $2n + 1$ sommets trivalents. Cette somme commence donc en degré $n + 1$, ce qui implique :

$$\Omega_n(\Sigma(K)) = \sum_{K'} (-1)^{\#K'} \Omega_n(M_{L(K')}) = \sum_{K_0} (-1)^{\#K_0} \Omega_n(M_{L(K, K_0)}) = 0.$$

On en déduit que l'invariant LMO envoie chaque élément de $\mathcal{F}_{2n+2}$ en une série qui ne comporte que des termes de degrés $\geq n + 1$ et induit une application de $\mathbf{Q}[\mathcal{S}]/\mathcal{F}_{2n+2}$ dans le tronqué à l'ordre n de $\mathcal{A}'(\emptyset)$.

Pour terminer la démonstration il suffit de montrer que si K est un diagramme trivalent sans composante circulaire, sa classe dans $\mathcal{A}'(\emptyset)$ est égale à

$$\Omega_n(\Sigma(K)) = \sum_{K' \subset L(K)} i_n \circ \check{Z}(L(K) - K').$$

Mais, comme on l'a vu précédemment, on peut remplacer cette somme par :

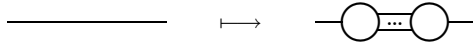
$$\Omega_n(\Sigma(K)) = \sum_{P \subset X} (-1)^{\#P} i_n \circ \check{Z}(L(K)_P),$$

où X est l'ensemble des sommets de K et $L(K)_P$ l'entrelacs en bande plongé $L(K)$ modifié par une transformation de type Δ près de chaque sommet de P .

Si K est de degré n il a $2n$ sommets et $3n$ arêtes. Par itération de la formule précédente on montre que la somme alternée $\sum (-1)^{\#P} \check{Z}(L(K)_P)$ est de la forme

$$\sum_{P \subset X} (-1)^{\#P} \check{Z}(L(K)_P) = U + V,$$

où U est le diagramme $\check{Z}(L_0(K))$ où l'on a attaché près de chaque sommet de K trois arêtes en trois points de $L_0(K)$ avec un sommet trivalent commun, et où V est un combinaison linéaire de diagrammes ayant au moins $2n + 1$ sommets trivalents situés en dehors de $L_0(K)$. Il reste à déterminer la contribution de U ayant exactement n sommets en chaque cercle de $L_0(K)$. Comme $L_0(K)$ est une union disjointe de $3n$ copies de l'entrelacs de Hopf, on voit que le seul terme de la somme alternée qui donne une image peut-être non nulle par i_n est le diagramme $\hat{K}$ multiplié par $(2n - 1)^{-3n}$, où $\hat{K}$ est le diagramme obtenu à partir de K en modifiant chaque arête de K de la façon suivante :



avec $2n - 1$ arêtes joignant les deux cercles. Un dernier calcul montre alors que $i_n(\hat{K})$ est égal à K , ce qui termine la démonstration. $\square$

L'invariant LMO est assez mystérieux. Si M est une variété de dimension 3 connexe compacte orientée sans bord, son invariant $Z^{\text{LMO}}(M)$ est un élément de type groupe de l'algèbre de Hopf $\mathcal{A}'(\emptyset)$; c'est donc l'exponentielle d'un élément primitif u . Cet élément u est une somme de diagrammes trivalents connexes ayant au moins un sommet.

La composante de u de degré 0 est nulle et sa composante de degré 1 est nécessairement un multiple du diagramme Θ suivant :

$$\Theta = \bigcirc \text{---} \bigcirc$$

Il existe donc un invariant c_1 à valeurs rationnelles tel que pour toute variété M compacte orientée connexe sans bord de dimension 3 le coefficient de degré 1 de $Z^{\text{LMO}}(M)$ est de la forme $c_1(M)\Theta$. Cet invariant est donné par le théorème suivant :

6.5. Théorème. — *L'invariant c_1 est relié à l'invariant de Casson–Walker–Lescop λ par la formule suivante :*

$$c_1(M) = (-1)^b \lambda(M)/2,$$

où b est le premier nombre de Betti de M .

Dans le cas où M est une sphère d'homologie entière, ce théorème se ramène à montrer que λ est nul sur le module $\mathcal{F}_3$ et à faire le calcul pour une variété particulière, par exemple la sphère de Poincaré. Ceci marche également pour les sphères d'homologie rationnelle car pour ces variétés, on montre que l'invariant L^{LMO} est aussi un invariant de type fini universel, du moins pour les variétés de dimension 3 dont le premier groupe d'homologie entière est un groupe fini donné et la forme d'enlacement est une forme donnée. La démonstration de ce fait ([H]) est plus délicate car la fonction W n'est plus disponible dans cette nouvelle situation.

Remarque. — Le module des primitifs de $\mathcal{A}'(\emptyset)$ est de dimension 1 en degré 1, 2 et 3. Outre l'invariant c_1 on a donc deux invariants c_2 et c_3 de degré 2 et 3 qui caractérisent l'invariant LMO en degré ≤ 3 . Ces invariants c_2 et c_3 sont à valeurs rationnelles, mais on n'a pas actuellement d'interprétation simple de ces invariants.

Remarque. — En grand degré on n'a pas d'interprétation de Z^{LMO} . Mais cet invariant est aussi riche que possible. En effet Habegger, Orr et Le ([HO, Le]) ont montré que si un élément u de degré $n > 0$ de $\mathcal{A}'(\emptyset)$ est représenté par un diagramme trivalent connexe, il existe une sphère d'homologie entière M dont l'invariant LMO est de la forme :

$$Z^{\text{LMO}}(M) = 1 + u + U,$$

U étant une somme de termes de degrés $> n$. Le résultat demeure si u est une combinaison linéaire à coefficients entiers de diagrammes trivalents connexes.

Références

- [BN] D. Bar-Natan – *On the Vassiliev knot invariants*, Topology **34** n°2 1995, 423–472.
- [D] V. Drinfeld – *On quasi-triangular quasi-Hopf algebras and a group connected with $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$* , Leningrad Math. J. **2** 1990, 829–860.

- [GL1] S. Garoufalidis and J. Levine – *On finite type invariants II*, Math. Annalen. **306** 1996, 691–718.
- [GL2] S. Garoufalidis and J. Levine – *Finite type invariants, the mapping class group and blinks*, J. Diff. Geom. **47** n°2 1997, 257–320.
- [GO] S. Garoufalidis and T. Ohtsuki – *On finite type invariants III : manifold weight systems*, Topology **37** n° 2 1998, 227–243.
- [Gu1] M. N. Gusarov – *On n -equivalence of knots and invariants of finite degrees*, in “Topology of Manifolds and Varieties”, 1996.
- [Gu2] M. N. Gusarov – *Finite type invariants and n -equivalence of 3-manifold*, CR. Acad. Sci Paris Sér I Math. **329** n° 6 1999, 517–522.
- [HO] N. Habegger and K. Orr – *Finite type Invariants, Realization and Vanishing*, J. Knot Theory Ram. **8** 1999, 1001-1007.
- [H] K. Habiro – *Claspers and finite type invariants of links*, Geom. Topol. **4** 2000, 1–83.
- [Ko] M. Kontsevich – *Vassiliev’s knots invariants*, Adv. Sov. Math. **16** n°2 1993, 137–150.
- [Le] T. T. Q. Le – *An invariant of integral homology 3-spheres which is universal for all finite types invariants*, AMS Translations series 2, **179** 1997, 75-100.
- [LM] T. T. Q. Le and J. Murakami – *The universal Vassiliev-Kontsevich invariant for framed oriented links*, Compositio Math. **102** 1996, 271–291.
- [LMO] T. T. Q. Le, J. Murakami and T. Ohtsuki – *On a universal perturbative invariant of 3-manifolds*, Topology **37** n°3 1998, 539–574.
- [Vo] P. Vogel – *Invariants de Vassiliev des nœuds*, Sémin. Bourbaki, 1992–93, Astérisque **216** 1993, 213–232.

POSTFACE

INVARIANTS CONTEMPORAINS

par

Daniel Bennequin

Résumé. — Le texte se propose de présenter à des mathématiciens (ou à des étudiants en mathématiques) l'origine des dualités en théorie quantique des champs et en théorie des cordes (renormalisation, supersymétrie...), et d'expliquer comment ces dualités font apparaître de nouveaux invariants en topologie différentielle.

Abstract (Contemporary invariants). — In this text we will present for mathematicians (or students in mathematics) the origin of dualities in Quantum Field Theory and String Theories (renormalization, supersymmetry, ...), and will explain to explain how these dualities are the source of new invariants in Differential Topology

I. Champs (semi-) classiques

Les trois théories d'invariants des chapitres précédents ont en commun de venir de la physique contemporaine, une physique nouvelle au point que nombre de physiciens se demandent où elle peut rejoindre les expériences. Si des mathématiciens lui contestent le nom de physique, ils ne peuvent pourtant plus l'ignorer comme source d'inspiration en mathématiques.

John Morgan a parlé d'invariants des variétés de dimension 4 d'après la récente théorie de champs introduite par Seiberg et Witten. Pierre Vogel a parlé d'invariants topologiques des nœuds et des variétés de dimension 3, dont l'origine est une théorie de champs appelée Chern-Simons, que Witten inventa pour comprendre l'invariant de

Classification mathématique par sujets (2000). — 53, 81, 00A30, 01A65, 53-02, 53C07, 53C80, 57R55, 57R57, 81T13, 81T16, 81T20, 81T30, 81T60, 81V05.

Mots clefs. — Théories de jauge, spineurs, théorie quantique des champs, groupe de renormalisation, dualité faible/fort, dualité électrique/magnétique, monopôle, supersymétrie, supercordes, M -théorie, théorie des invariants.

Le texte écrit est une version très étendue et détaillée de l'exposé oral. L'auteur tient à remercier beaucoup Sylvie Paycha pour tout ce qu'elle lui a apporté dans de nombreuses discussions et lettres au cours de l'élaboration de ce texte, et Marie-Paule Bressoulaly pour sa gentillesse, sa patience et sa compétence lors de la frappe d'un manuscrit plein de mots hirsutes.

Jones. Et la théorie des invariants symplectiques présentée par Michèle Audin s'appelle théorie de Gromov-Witten ; ses développements récents s'inspirent de la physique des théories de champs bidimensionnelles.

Le nom de Witten est symptomatique de cette nouvelle physique ; on lui doit l'invention systématique de théories de champs quantiques calculant ou justifiant des invariants topologiques. Bien que l'idée de telles théories soit due à des mathématiciens, M. Atiyah et G. Segal (1987), la plupart des exemples et le tableau d'ensemble reviennent à E. Witten. En revanche, le travail mathématique nécessaire pour mettre ces théories sous une forme rigoureuse à partir des indications de Witten, n'est pas une simple vérification ; il fait souvent appel à d'autres éléments de physique et de mathématique. Par exemple la géométrie et l'analyse non linéaire pour Kronheimer, Mrowka, Taubes à propos des invariants de Seiberg-Witten (*cf.* chapitre 2, §§ III, IV), ou les groupes quantiques pour Reshetikhin, Turaev à propos des invariants de Chern-Simons (qui sont des invariants de variétés de dimension 3 du type de ceux du chapitre 3), ou l'algèbre homologique et les équations différentielles pour Givental, Kontsevich, Manin à propos des invariants de Gromov-Witten (*cf.* chapitre 1, §§ 4, 5).

Notons une absente au tableau des théories inventoriées : la théorie de Yang-Mills. C'est la théorie des champs qui a été retenue pour décrire l'interaction des particules élémentaires, elle est à la base de la théorie des invariants de variétés différentielles de dimension 4 développée par Donaldson, et elle a motivé toute l'évolution récente y compris la découverte des invariants de Seiberg-Witten. Cela vaut la peine de s'attarder un peu sur elle. Référence (en plus de ce volume) : S.K. Donaldson and P.B. Kronheimer, *The geometry of four-manifolds*, Oxford University Press (1990).

Dans la théorie de Yang-Mills la variable est une connexion ∇ sur un fibré principal⁽¹⁾ P au-dessus d'une variété W de dimension 4. Le groupe structural de P est un groupe de Lie G ; c'est lui que la littérature physique nomme *groupe de jauge*. À ne pas confondre avec le *groupe d'invariance de jauge* $\mathcal{G}$ qui est le groupe des automorphismes de P (relevant l'identité de W dans W). En choisissant une trivialisations locale de P , on écrit $\nabla = d + A$, avec une 1-forme différentielle A à valeurs dans l'algèbre de Lie $\mathfrak{g}$ de G ; les éléments de $\mathcal{G}$ s'expriment par des fonctions f d'un ouvert de W dans le groupe G . Plus près du point de vue physique, les champs sont les composantes de A dans une base de $\mathfrak{g}$ car les choix de coordonnées font partie de l'appareil expérimental. Mais les mathématiciens, et beaucoup de physiciens, préfèrent penser que le champ est ∇ ; d'ailleurs les seules quantités sur lesquelles divers observateurs peuvent se mettre d'accord sont des fonctions de la classe de ∇ modulo $\mathcal{G}$, la *classe d'équivalence de jauge*. Un *changement de jauge* est une transformation de A en $f^{-1}Af + f^{-1}df$, on peut le voir, suivant les cas, comme l'effet de l'action d'un élément f de $\mathcal{G}$ ou comme l'effet d'un choix différent de trivialisations locale de P . La courbure F , ou F_A , de ∇ est la 2-forme définie par $F_A = dA + A \wedge A$ (*cf.* chapitre 2) ; à l'origine classique c'est F qui s'appelait le champ.

Pour décrire la dynamique classique de ∇ , choisissons une métrique riemannienne sur W et notons $*$ l'opérateur de Hodge-de Rham associé, échangeant les formes extérieures de degrés p et $4 - p$. Soit d_A la dérivée extérieure covariante, *i.e.* l'unique extension de ∇ en une dérivation (graduée) du module des formes différentielles à valeurs dans $\mathfrak{g}$ (sur l'algèbre extérieure des formes différentielles scalaires). L'identité de Bianchi, toujours satisfaite, est $d_A F_A = 0$, et l'équation de Yang-Mills s'écrit $d_A * F_A = 0$. Les solutions les plus remarquables sont les connexions satisfaisant à l'équation d'(anti)autodualité :

$$F_A = - * F_A$$

Ces solutions s'appellent des *instantons*, à cause de la forme en bosse particulière qu'elles prennent lorsque $W = S^4$. Le groupe $\mathcal{G}$ agit sur l'ensemble de ces solutions et l'ensemble quotient est la *variété de module des instantons*. Pour S^4 cette variété est décrite par la construction ADHM (Atiyah, Donaldson, Hitchin, Manin, 1978).

Dans la théorie de Donaldson le groupe de jauge G est SU_2 , les espaces de modules $\mathcal{M}_k$ dépendent d'un entier k classifiant le fibré P . Ces espaces sont sensibles au choix de la métrique, mais leur cohomologie (singulière) permet de construire certains polynômes sur $H_2(W)$ qui sont des invariants de difféomorphismes de la variété de base W (*cf.* chapitre 1 pour une exploitation analogue des modules de courbes pseudo-holomorphes)⁽²⁾.

De l'existence des instantons, Donaldson avait tiré des résultats révolutionnaires sur la topologie des variétés de dimension 4. C'était la première application remarquée de la physique des particules à la topologie différentielle ; mais, en regardant « mieux », on s'aperçoit que les premiers théorèmes de Donaldson (par exemple l'interdiction de certaines formes d'intersection sur $H_2(W)$) sont des approfondissements de corollaires du théorème de l'indice (par exemple le théorème de la signature de Rohlin). Or l'équation de Dirac, essentielle pour comprendre le théorème de l'indice, avait été inventée pour décrire les électrons (Dirac, 1928). Sans parler de la théorie de Hodge, directement attachée aux équations de Maxwell, qui sont les équations de Yang-Mills pour $G = U_1$, le cercle.

Mais comment donc apparaissent les équations de Seiberg-Witten ? Cette question va nous entraîner un peu loin dans une jungle d'idées et de techniques aventureuses où les concepts naissent, mais le voyage aidera à apprécier la nature des nouveaux invariants.

Comme l'explique J. Morgan (voir chapitre 2), (avec de petites variations de notations) la donnée d'une certaine structure Spin^c sur la variété W de dimension 4, détermine un fibré L sur W de fibre $\mathbb{C}$ et un fibré S^+ sur W de fibre $\mathbb{C}^2$; les équations de Seiberg-Witten portent sur une paire (A, ψ) d'une U_1 -connexion sur L et d'une section ψ de S^+ :

$$\mathcal{D}_A \psi = 0 \quad , \quad F_A^+ = \psi \otimes \overline{\psi} - \frac{1}{2} |\psi|^2 \text{Id},$$

où $\mathcal{D}_A$ est l'opérateur de Dirac et F_A^+ la composante auto-duale de la 2-forme de courbure (cf. chapitre 2). La section ψ de S^+ est l'exemple d'un *champ de spineurs*. La première équation est justement l'équation de Dirac⁽³⁾.

Ces équations auraient pu émerger en combinant des principes de théorie des champs, auto-dualité, (super)-symétrie (voir plus loin § V), mais l'histoire s'est faite autrement : en 1989 Witten a trouvé une théorie différente de Yang-Mills, appelée « $N = 2$ super-Yang-Mills twistée », dont la quantification décrit les polynômes de Donaldson. Au § V, nous verrons comment, en 1994, Seiberg et Witten ont découvert leurs équations dans une théorie de jauge « duale » de cette théorie, $N = 2$, *etc.* C'est d'ailleurs cette origine qui permet d'espérer qu'elles fourniront des invariants topologiques.

Des connexions comme ∇ ou A , des champs de spineurs comme ψ , ont été chargés de décrire la plupart des particules élémentaires connues, la faune des gluons, bosons intermédiaires, photons, leptons, quarks... Mais ces particules se manifestent avant tout dans le domaine quantique, et c'est un gros problème ou un vaste programme de passer des équations classiques à une dynamique quantique. Or la transition de Yang-Mills à Seiberg-Witten, qui comporte un exemple de « dualité contemporaine », ne se comprend qu'au cours de la quantification.

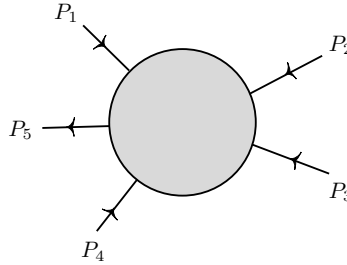
Dans les paragraphes suivants, j'essaie d'expliquer en quoi consiste cette dualité, ce qui ne peut se faire sans donner une idée de la *renormalisation*.

II. Quantification des champs

Pour certains auteurs la renormalisation est vraiment le cœur de la théorie quantique des champs, pour d'autres ce n'est qu'un outil technique. En tout cas il n'existe pas aujourd'hui d'autre voie pour proposer des nombres comparables avec l'expérience dans le domaine à la fois quantique et relativiste.

La confrontation des principes de la mécanique quantique et de la relativité a provoqué d'innombrables difficultés. Les physiciens l'avait deviné sur les relations d'incertitudes : si $\Delta p \Delta q \geq \hbar$ et $\Delta q \leq c \Delta t$, il faut $\Delta p \Delta t \geq \hbar c^{-1}$ ou encore $\Delta E \Delta t \geq \hbar$, si bien qu'il faut attendre quelque temps pour voir une mesure précise de l'impulsion, de l'énergie ou de la masse ($\sim 10^{-40}$ s pour 1g). Plutôt qu'une contradiction dans les principes, c'est une source de divergences ; là où on aimerait des nombres réels, il apparaît des intégrales divergentes qu'il faut régulariser, et puis des séries divergentes qu'il faut réorganiser, *etc.*

De même qu'en mécanique quantique la densité de probabilité de présence d'une particule est le carré du module d'une fonction d'onde, la théorie quantique des champs postule que la probabilité de trouver certaines particules $A_1, A_2, \dots$ avec des impulsions déterminées $p_1, p_2, \dots$ à l'entrée-sortie d'une collision peut s'obtenir à partir de nombres complexes notés $\langle A_1(p_1) A_2(p_2) \dots \rangle$, et appelés *amplitudes de probabilité*, ou juste *amplitudes*, ou quelquefois *fonctions de corrélation* en pensant à une ana-



logie avec la mécanique statistique. Ce sont les briques de l'édifice d'une théorie de champs quantique. Les lettres $A_1, A_2, \dots$ renvoient à des champs classiques, comme A, ψ du § I; dans le cadre quantique on les appelle encore des *champs*. Il est pratique de rassembler les amplitudes pour toutes les valeurs de $p_1, p_2, \dots$ possibles en faisant une (co)transformation de Fourier, comme celle qui fait passer de $\hat{f}(p)$ à $f(x)$; on obtient ainsi des distributions notées (abusivement) $\langle A_1(x_1)A_2(x_2)\dots \rangle$.

Mais les premiers à avoir surmonté les difficultés du problème (pour le photon et l'électron), en électromagnétisme, Feynman, Schwinger, Tomonaga... ont découvert un résultat inattendu : il n'y a pas de définition univoque des amplitudes, leur valeur est ambiguë. Au lieu de dépendre simplement des paramètres évidents (charges, masses...) et d'éventuelles conditions au bord (état du vide...) les amplitudes dépendent en plus de l'échelle d'observation. La théorie est même susceptible de traverser différentes phases, selon les valeurs de tous ces paramètres.

L'expérience suggère qu'à une certaine échelle les champs sont facilement découplés, sinon le mot d'interaction n'a pas de sens. Cependant le formalisme lagrangien et l'invariance de Lorentz forcent à regarder la théorie à toutes les échelles, les très grandes comme les très petites. Rien n'indique d'ailleurs a priori qu'un grossissement excessif ne réclame pas l'intervention de beaucoup d'autres champs. Les théories *renormalisables* sont celles qui s'adaptent à toutes les échelles. On a le droit d'admirer que la contrainte de renormalisabilité ait sélectionné des théories pour toutes les particules connues (plus quelques autres).

Une façon de présenter la quantification des champs s'est peu à peu imposée, celle de *l'intégrale fonctionnelle* (due à Feynman). Ce sera notre point de départ.

Au commencement est une théorie des champs classique, comme celle de Maxwell ou Yang-Mills, et les équations du mouvement sont les équations d'Euler-Lagrange d'une certaine fonction (des $A_i(x)$ et de leurs dérivées itérées en x), le *lagrangien* $\mathcal{L}(A_1, A_2, \dots)$ (par exemple pour Yang-Mills, $\mathcal{L}(A) = \text{Tr}(F_A \wedge *F_A)$.) Pour simplifier notons A l'ensemble des champs $A_1, A_2, \dots$. Feynman fait comme s'il existait une mesure de volume $\mathcal{D}A$ pour laquelle les amplitudes s'écriraient

$$\langle A_1(x_1)A_2(x_2)\dots \rangle = \frac{1}{Z} \int \mathcal{D}A \left(\exp \frac{i}{\hbar} \int_W \mathcal{L}(A) \right) A_1(x_1)A_2(x_2)\dots$$

où $Z = \int \mathcal{D}A \exp\left(\frac{i}{\hbar} \int_W \mathcal{L}(A)\right)$ s'appelle la *fonction de partition*. L'intégrale sur W de $\mathcal{L}(A)$ est l'*action classique*. Les intégrales suivant $\mathcal{D}A$ portent le nom d'*intégrales de Feynman*.

Cette écriture n'est qu'un guide formel, un programme à réaliser. Notons encore qu'il est nécessaire d'introduire aussi des fonctions F des champs A (et de leurs dérivées successives), qu'on considérera elles-même comme des champs (par exemple l'énergie, les moments...) et dont on prendra les moyennes, notées $\langle F_1(x_1)F_2(x_2) \cdots \rangle$.

Trouver un sens à de telles intégrales exige d'abord une « régularisation » : les deux principales recettes (pour les théories de jauge) sont la « troncature », élimination des A de trop grandes et trop petites impulsions, et la « régularisation dimensionnelle », déplacement en dimension non entière par prolongement analytique. Mais, dans un deuxième temps, il faut s'affranchir de ces régularisations arbitraires : des années 40 aux années 70 furent élaborés des procédés miraculeux (et convaincants) pour prendre la limite en adaptant arbitrairement un nombre fini de quantités, afin de définir la « partie finie » des amplitudes. Au début des années 70 l'approche de Wilson par le groupe de renormalisation a bien éclairci la situation. Passons par là.

Soit $\mathcal{L}$ un lagrangien de champ classique ; par exemple

$$\mathcal{L}(\varphi) = \frac{a}{2} |\nabla \varphi|^2 - \frac{m^2}{2} \varphi^2 - \frac{\lambda}{24} \varphi^4$$

pour la théorie dite φ^4 d'un champ scalaire (juste une fonction à valeur réelle sur l'espace-temps, modèle simplifié du magnétisme si l'on rend le temps imaginaire pur). Et soit U_0 la variété des paramètres dont $\mathcal{L}$ dépend ; dans l'exemple, U_0 est l'espace de coordonnées (a, m^2, λ) , a se nomme l'« intensité de champ », m la « masse » et λ la *constante de couplage* ; dans un $\mathcal{L}$ plus général, on trouve des coordonnées naturelles pour U_0 , ce sont souvent les coefficients de monômes de dérivées des champs, et on les appelle toutes « constantes de couplages » pour simplifier.

L'échelle d'observation des particules (que $A_1, A_2, \dots$ sont censés décrire) est donnée par l'inverse d'une distance ; elle est mesurée par un nombre Λ réel strictement positif, *i.e.* $\Lambda \in \mathbb{R}_+^\times$. Dans une théorie relativiste quantique, avec des unités où $c = \hbar = 1$, ce nombre Λ fixe aussi l'ordre de grandeur d'une fréquence, d'une énergie, d'une masse ou d'une impulsion.

La théorie achevée de $\mathcal{L}(A)$ doit prévoir des amplitudes $\langle A_1(x_1)A_2(x_2) \cdots A_n(x_n) \rangle$ bien définies à une échelle donnée, pour des valeurs données des constantes de couplages. Ce qui justifie la formulation de K.G. Wilson :

une théorie quantique relativiste est paramétrée par une variété V munie d'un isomorphisme π sur $U_0 \times \mathbb{R}_+^\times$; un point de V représente une collection cohérente d'amplitudes, à chaque point de V est associé un point $g \in U_0$ et un nombre $\Lambda \in \mathbb{R}_+^\times$.

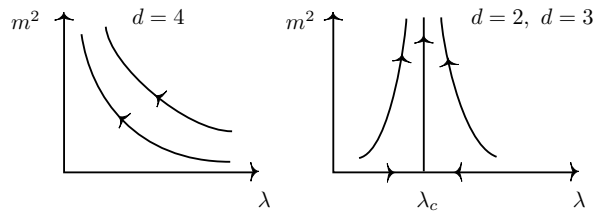
L'application π correspond aux *conditions de renormalisation*, interprétation des constantes de couplages physiques par certaines amplitudes bien choisies, par exemple la masse (fois i) comme pôle des fonctions à deux points $\langle \varphi(p_1)\varphi(p_2) \rangle$ en fonction de

l'impulsion $p = p_1 - p_2$, ou la constante λ de l'exemple φ^4 comme valeur d'une amplitude (convenablement normalisée) à quatre points $\langle \varphi(p_1)\varphi(p_2)\varphi(p_3)\varphi(p_4) \rangle$.

Pratiquement le point g est donné par des valeurs expérimentales ; par exemple, en électromagnétisme, il contient la masse et la charge de l'électron. Mais l'expérience montre que les « mêmes » particules à des échelles différentes donnent des points différents de U_0 . Aussi V doit-elle être munie d'un feuilletage ρ de dimension 1 ; ce feuilletage dit l'équivalence des amplitudes (l'évolution d'une même théorie suivant Λ), et π ne le projette pas, en général, sur le feuilletage trivial des parallèles à $\mathbb{R}_+^\times$. Par contre un choix d'unité d'échelle (le dernier choix restant) est a priori indifférent ; ce qui entraîne que les feuilles de ρ se projettent sur les trajectoires d'un champ de vecteurs (autonome) β sur U_0 .

Le champ β , ainsi que son action (simple) sur les amplitudes (Gell-Mann et Low), porte le nom de *flot de renormalisation* ou *groupe de renormalisation*. C'est une sorte de *groupe de Galois* de la théorie, comme le concevaient (en 1953) les initiateurs Stueckelberg et Perelmann⁽⁴⁾.

Dans l'exemple φ^4 la dynamique de β se lit dans le plan de coordonnées $\left(\frac{m^2}{a}, \frac{\lambda}{a}\right)$; elle dépend beaucoup de la dimension d de l'espace-temps. Par exemple :



La dynamique de β sur U_0 a des conséquences importantes pour la dynamique des champs dans l'espace-temps. *A priori* β pourrait engendrer un flot compliqué, cycles limites, attracteurs étranges, *etc*, mais une conjecture qui tient bon prétend que β est le gradient d'une fonction. D'ailleurs ce sont surtout les points d'équilibre qui ont retenu l'attention. Les points de U_0 qui sont attracteurs quand $\Lambda \rightarrow \infty$ sont dits *stables dans l'ultra-violet* (ou UV-stable), ceux qui sont attracteurs quand $\Lambda \rightarrow 0$ sont dits *stables dans l'infrarouge* (ou IR-stable) (ainsi 0 et ∞ sont les deux racines de l'arc-en-ciel). Lorsqu'un point correspondant à une théorie découplée, sans interaction, est UV-stable, la théorie est dite *asymptotiquement libre*. D'après 't Hooft, c'est le cas de la théorie de Yang-Mills non abélienne qui décrit les interactions fortes (QCD), interactions entre quarks, au contraire de l'électrodynamique quantique (QED) où l'absence d'interaction est plutôt IR-stable (le propre des théories de jauge abéliennes). Donc, aux grands moments de transfert (c'est-à-dire Λ grand), le couplage entre quarks et fermions tend vers zéro ; la théorie perturbative peut faire des prédictions. Le problème est à grande échelle (c'est-à-dire Λ petit). Là, logiquement, la théorie des perturbations ne peut plus dire grand chose.

De plus, il faut être honnête, pratiquement rien du schéma de renormalisation, U, R, β , n'a été mis en place en dehors de calculs perturbatifs ; c'est-à-dire que les physiciens ont dû choisir des paramètres, considérés comme petits, par exemple $1/137 = e^2/\hbar c$ ou un couplage (renormalisé) g (ou plusieurs) et procéder uniquement ordre par ordre en les puissances de g : ce n'est que chaque terme du développement des amplitudes qui va peu à peu se renormaliser, ρ et β ne seront définis que pas à pas, perturbativement.

(Quelques références :

K.G. Wilson, *The renormalization group and critical phenomena*, Reviews of Modern Physics, vol. 55 n° 3 (1983).

K.G. Wilson, *Renormalization of a Scalar Field Theory in Strong Coupling*, Physical Review D, vol. 6 n° 2 (1972), 419–426.

J. Polchinski, *Renormalization and effective lagrangians*, Nuclear Physics B 231 ; (1984) 261–295.

G. 't Hooft, M. Veltman, *Diagrammar*, CERN (1973) et G. 't Hooft, *The renormalization group in QFT*, (1988), in *Under the spell of the gauge principle*, World Scientific (1994).)

Il faut maintenant ajouter, pour cette partie, le tome 1 du livre édité par P. Deligne, P. Etingof et al., *Quantum Fields and Strings : A course for mathematicians*, AMS, IAS, 1999.

III. Dualités des champs

À l'échelle humaine les quarks ne se manifestent que par paires ou par triplets ; on n'arrive pas à les séparer les uns des autres. Par contre, aux très fines échelles de certains accélérateurs, les quarks se montrent un par un, ou plutôt comme deux ou trois toutes petites billes dans un petit sac. La plupart des physiciens sont persuadés que ce « confinement » est un effet non perturbatif en théorie de Yang-Mills. Les quarks et les gluons qui les font interagir sont décrits par le couplage des équations de Dirac avec celles de Yang-Mills pour le groupe SU_3 . Depuis les années 70, avec les calculs sur réseaux (Wilson) et la découverte des effets topologiques (Wilson, 't Hooft, Polyakov), l'intérêt des « renormalisateurs » s'est de plus en plus porté sur les « effets non perturbatifs » en couplage fort. Faisant appel à la topologie du groupe de jauge, l'explication (conjecturée) du confinement passe par la « dualité ».

Il est déjà temps d'indiquer ce qui s'appelle *dualité* dans une théorie quantique des champs ; d'abord ce qu'est une « dualité faible/fort » : on dit qu'on a une telle dualité lorsque, après renormalisation, deux familles $U_0, \hat{U}_0$ de lagrangiens classiques sur des champs a priori différents $A, \hat{A}$ décrivent une même famille d'amplitudes, mais en échangeant couplage fort sur A avec couplage faible sur $\hat{A}$ et réciproquement⁽⁵⁾. (Pour une parfaite dualité, on considère la collection complète des amplitudes, mais il peut arriver que la collection se réduise à une fonction de partition.)

Ainsi, très schématiquement, deux théories de champs classiques sont duales lorsque du point de vue quantique elles sont équivalentes (ou partiellement équivalentes).

Un prototype de dualité faible/fort se trouve en mécanique statistique avec un modèle discret : le modèle d'Ising en dimension 2.⁽⁶⁾

Comme le dit J. Polchinski (*String duality*, Rev. of Modern Physics vol. 68, (1996)), la dualité du fort et du faible est analogue à la transformation de Fourier où une fonction qui s'étale dans l'espace des positions devient très concentrée dans l'espace des moments. La mécanique quantique toute entière repose sur cette dualité, elle l'énonce sous la forme d'une dualité onde/corpuscule $\Delta p \Delta q \geq \hbar$. L'analogie est plus précise encore si l'on regarde la représentation d'une gaussienne : du point de vue des positions, « dans la base des masses de Dirac δq », la gaussienne s'écrit $G(q) = \frac{1}{\sqrt{2\pi\sigma}} \exp(-q^2/2\sigma)$, et du point de vue des moments, après transformation de Fourier « dans la base des ondes pures $\exp(ipq)$ », on a $\hat{G}(p) = \exp(-p^2\sigma/2)$. On voit que σ est remplacé par σ^{-1} .

La transformation de Fourier échange les masses de Dirac des points avec celles des hyperplans (distributions de simples couches) ; on retrouve ainsi la « dualité projective », surtout développée au XIX^e siècle (Poncelet, Brianchon, Gergonne) mais prenant ses racines dans les travaux de Desargues et Pascal. Sous la forme de Fourier elle échange l'infiniment grand et l'infiniment petit. Et là entre les deux se tient l'homme de René Descartes et de Blaise Pascal, « un néant à l'égard de l'infini, un tout à l'égard du néant, un milieu entre rien et tout ; infiniment éloigné de comprendre les extrêmes. La fin des choses et leurs principes lui sont invinciblement cachés dans un secret impénétrable... ». Mais la physique de Pascal est résolument plus moderne que celles de Descartes, puisqu'elle « substitue aux évidences de la Raison, les vertiges et l'effroi des Abîmes » (cf. J.P. Fanton d'Andon, *l'horreur du vide*, 1978). La quête des physiciens, construisant un gigantesque appareil mathématique dont on n'aperçoit ici qu'un tout petit bout, à certains donne de l'assurance à d'autres de l'effroi.

Et pourtant « manque d'avoir contemplé ces infinis, ils continuent de se porter téméairement à la recherche de la nature comme s'ils avaient quelque proportion avec elle ». (Pensées de Pascal, fragment n° 185 *Disproportion de l'homme* dans l'édition de Le Guern, édition Folio. Voir aussi fragment n° 397, *Infini rien*)

Derrière certaines transformations de Fourier, la formule de Poisson et la transformation modulaire des fonctions Θ , 't Hooft a montré qu'on peut discerner une dualité faible/fort entre l'énergie électrique et l'énergie magnétique de certains modèles de particules (en « phase de Coulomb », cf. « *Under the spell* »... p.508).

Justement, la dualité entre électricité et magnétisme est à l'origine de l'invention des équations de Seiberg-Witten.

Dans le vide les équations de Maxwell $d * F = 0$ et $dF = 0$ (tautologique puisque $F = dA$), s'écrivent

$$\begin{aligned} \nabla \times B &= \frac{\partial E}{\partial t}, & \nabla \times E &= -\frac{\partial B}{\partial t}, \\ \nabla \cdot B &= 0, & \nabla \cdot E &= 0, \end{aligned}$$

en décomposant la courbure F en un champ électrique E et un champ magnétique B . Elles sont invariantes sous l'action de l'opérateur $*$ qui se traduit par les transformations $E \mapsto B$, $B \mapsto -E$. Pour que cette symétrie $*$ reste valable en présence de charges et de courants, on doit ajouter non seulement des charges et courants électriques mais aussi des charges et courants magnétiques; c'est ce que proposa Dirac en 1931 avec l'introduction de nouvelles particules, les *monopôles magnétiques*. Dirac montra comment les principes quantiques forcent la relation $eb \in 2\pi\mathbb{Z}$ entre une charge électrique ponctuelle e et une charge magnétique ponctuelle b , expliquant ainsi pourquoi toutes les charges électriques connues sont des multiples entiers d'une même charge élémentaire (ici, avec Dirac et la tradition, on a choisi les unités de Gauss, où $4\pi = \hbar c$). Comme e est le couplage de l'électron et du photon, b serait le couplage du monopôle magnétique et du « photon magnétique ». Remarquons que b est de l'ordre de $1/e$; on aurait donc une dualité qui échange les couplages faibles et forts. Cependant, à un niveau plus profond (tenant compte des différences de nature entre les potentiels, électriques et magnétiques), de nombreuses difficultés s'élèvent contre cette dualité en électromagnétisme ordinaire; voir E. Witten « *Duality, space time and quantum mechanics* », Physics Today, Mai 1997.

Dans les théories de jauge non abéliennes on retrouve de tels monopôles, magnétiquement chargés, mais avec moins d'obstacles pour une dualité, comme l'ont suggéré C. Montonen et D. Olive en 1977. Nous allons voir comment au § suivant. Référence pour les §§ III à VI, Tome 2 du livre édité par P. Deligne, P. Etingof et al., *Quantum Fields and Strings : A course for mathematicians*, AMS, IAS, 1999.

IV. Monopôles

Dans une théorie de champ de jauge la constante de couplage g apparaît souvent dans la formule $\nabla = d + gA^{(7)}$; elle peut s'interpréter comme une charge électrique élémentaire. Les *monopôles non abéliens*, découverts indépendamment par 't Hooft et Polyakov en 1974, offrent un exemple de particules qui se manifestent très peu (par des termes $\exp(-g^{-1})$ exponentiellement petit en g) dans la théorie perturbative, en « couplage faible », mais qui dominent les amplitudes de la théorie quantique en « couplage fort », cf. § II. Ils possèdent une charge magnétique (de l'ordre de g^{-1}), qui joue le rôle dual de la charge électrique. C'est un bon début pour une dualité faible/fort et électrique/magnétique.

Les monopôles de 't Hooft et Polyakov sont des solutions particulières des équations de Yang-Mills-Higgs, version classique du « modèle standard » pour les particules en « interaction faible ». Pour ces interactions faibles (principales responsables des effets

radioactifs), la théorie standard n'est pas purement celle de Yang et Mills : il y a bien un groupe de jauge G (ici $SU_2 \times U_1$ ou $SU_3 \times SU_2 \times U_1$ si l'on veut tenir compte des quarks) et une connexion ∇ , mais il y a aussi au moins un champ φ , appelé *champ de Higgs*, section du fibré adjoint en algèbre de Lie $\mathfrak{g}$. Le lagrangien de Yang-Mills-Higgs est

$$\mathcal{L}(A, \varphi) = \text{Tr}(F_A \wedge *F_A) + |d_A \varphi|^2 + \lambda(|\varphi|^2 - a^2)^2.$$

Mais les monopôles les plus étudiés (nous ne parlerons que de ceux-là) sont des solutions (statiques) des équations de Bogomolny (en dimension 4 et en temps imaginaire pur) :

$$F_A = *d_A \varphi.$$

Ce sont les équations dérivées de $\mathcal{L}$ dans la limite $\lambda \rightarrow \infty$, introduite par Prasad et Sommerfield (1975). Pour que ces monopôles soient des solutions de Yang-Mills-Higgs dans $\mathbb{R}^4$ d'énergie finie, on impose qu'à l'infini dans $\mathbb{R}^3$ (à temps fixé) on ait dans chaque direction, $|\varphi| = a$; si bien que la configuration à l'infini du champ de Higgs φ fournit un invariant topologique, une classe caractéristique. Par exemple pour le groupe de jauge $G = SU_2$, φ définit une application de la sphère à l'infini de $\mathbb{R}^3$ dans la sphère de rayon a de SO_3 ; son degré est un nombre entier k qui peut s'interpréter comme une charge magnétique, $b = 2\pi k g^{-1}$. Autrement dit, les plans perpendiculaires aux directions de φ définissent, en dehors d'un compact de $\mathbb{R}^3$, un fibré en plans, de groupe structural SO_2 ; le nombre k est sa classe d'Euler. On dit alors que φ a *spontanément brisé* le groupe de jauge SU_2 sur le sous-groupe $H = SO_2$. (Attention : à proprement parler aucune symétrie n'est brisée, la symétrie de jauge $\mathcal{G}$ est intacte). Suivant B. Julia et A. Zee (1974) et M.K. Prasad et C.M. Sommerfield (1975) il existe aussi des *dyons*, devinés par J. Schwinger, possédant à la fois une charge magnétique et une charge électrique; alors la condition de quantification s'exprime par $e_n b_m - e_m b_n \in 2\pi\mathbb{Z}$, pour deux particules n et m de charges électriques e_n, e_m et magnétiques b_n, b_m . (J. Schwinger, D. Zwanziger 1968). (N. Manton, M. Atiyah et N. Hitchin, 1984, ont montré que l'interaction de monopôles « lents » produit des dyons.)

Polyakov et 't Hooft (1975, 1976) évaluèrent l'effet des instantons et des monopôles dans la théorie des perturbations⁽⁸⁾ : si g est la constante de couplage de jauge, les contributions aux amplitudes calculées par intégrales de Feynmann viennent avec un facteur de l'ordre de $\exp(-8\pi^2/g^2)$. 't Hooft et Polyakov remarquèrent aussi l'importance d'un facteur $\exp(i\theta)$ devant les amplitudes. Cet angle θ n'apparaît que dans la combinaison complexe $(\theta/2\pi) + i(4\pi/g^2)$; il vient de la possibilité d'ajouter un terme $\frac{g^2\theta}{32\pi^2} \text{Tr}(F_A^2)$ au lagrangien de Yang-Mills-Higgs. Ce terme est de nature topologique et ne modifie pas la théorie classique⁽⁹⁾.

Dans une théorie de jauge en dimension 4, pour $G = SU_n$ avec des champs de Higgs, mais sans brisure de symétrie, après un fixage de jauge (c'est à dire une trivialisatation unitaire des fibrés), 't Hooft montre que la théorie renferme des champs

vectoriels avec ou sans masse, des champs scalaires et des SU_2 -monopôles. En couplage fort, on espère que la théorie se réexprime, dualement, en séries de $1/g^2$ à partir de monopôles. Les physiciens attendent un comportement statistique analogue à celui des supraconducteurs, mais avec les rôles de l'électricité et du magnétisme échangés : plus précisément ils suspectent un comportement en troupeau de moutons des paires de monopôles. Or il est établi que les champs magnétiques, qui ne sont pas piégés dans les supraconducteurs, en sont rejetés ; c'est ce qui s'appelle « l'effet Meissner ». Dès lors un effet Meissner dual exclurait, ou confinerait les courants électriques. C'est le scénario de confinement des quarks (*cf.* début du § III) que 't Hooft proposa.

't Hooft détailla même la description d'un portrait de plusieurs phases pour la chromodynamique quantique renormalisée ; au moins trois phases : le mode du confinement, dual du mode de Higgs et une « phase de Coulomb » où l'électricité et le magnétisme sont symétriques, qui permet des forces de longue portée. Cette description, pour le groupe de jauge SU_n , fait jouer un rôle prépondérant au centre $\mathbb{Z}/n\mathbb{Z}$ de SU_n ⁽¹⁰⁾.

Dans un des cas les plus simples de la théorie de Yang-Mills-Higgs, $G = SO_3$, $H = U_1$ (modèle de Georgi-Glashow), Montonen, et Olive on fait une proposition précise pour tester la dualité : une formule que Bogomolny, Prasad, Sommerfield (BPS) avaient établie pour la masse des dyons dans la limite classique : $M = C\sqrt{e^2 + b^2}$, reste vraie après quantification et renormalisation : $M = u\sqrt{e^2 + b^2}$ où u est la moyenne de $\text{Tr}(\varphi^2)$, manifestant ainsi une symétrie de rotation pour l'électricité et le magnétisme (*cf.* « *Magnetic monopoles as gauge particles* » Physics Letters 72 (1977).)

Cette dualité exacte faible/fort, électrique/magnétique n'a pas pu être établie telle quelle, mais elle a pu être établie dans des théories de jauge avec suffisamment de « supersymétries ». Les dualités donnent une action complète de $SL_2(\mathbb{Z})$ sur $\tau = (\theta/2\pi) + i(4\pi/g^2)$ (ce qui aura son importance pour Seiberg-Witten). Indication en 1978 de Witten-Olive, argument de Seiberg en 1988, conclusion en 1994 de Ashoke Sen (Physics Letter B 329, 217–221).

Dans la littérature physique, *état BPS* signifie une particule dont la masse est proportionnelle à la charge, ce qui résulte en général d'une supersymétrie.

La nécessité de faire appel à la supersymétrie nous « oblige » à en parler à présent.

V. Supersymétrie. Théorie de Seiberg-Witten

La définition d'une particule commence avec la considération de l'espace lorentzien plat. Là le groupe de Poincaré (translations, rotations, contractions...) peut agir sur la collection des amplitudes d'une théorie de champs quantique. Pour une bonne part de l'étude, l'action de l'algèbre de Lie du groupe de Poincaré suffit. Les morceaux irréductibles donnent lieu aux « particules ». Or la *supersymétrie* est la seule extension de l'action de l'algèbre de Lie dans le cadre des algèbres de Lie graduées (dites *super-algèbres*) qui soit permise par les contraintes (unitarité...) (Théorème de

Haag, Lopuszanski, Sohnius, 1975). (Référence : J. Wess, J. Bagger, *Supersymmetry and Supergravity*, Princeton 1992.) La supersymétrie s'intègre en un « super-groupe de Poincaré ». On peut présenter géométriquement les choses en disant que l'espace-temps usuel V est étendu par un produit de N facteurs (à une infinité de dimensions) $S \otimes \Lambda^- \oplus V \otimes \Lambda_0^+$, où S est l'espace d'une représentation spinorielle, Λ l'algèbre extérieure sur $\mathbb{R}^\infty$, $\Lambda^+ \oplus \Lambda^-$ la décomposition en parties paire et impaire, $\Lambda^+ = \mathbb{R} \oplus \Lambda_0^+$; alors le nombre entier N se nomme *nombre de supersymétries*, la super-algèbre et son super-groupe donnent une algèbre et un groupe de transformations de l'espace étendu, le *super-espace*. Les théories de jauge s'étendent en théories de « super-Yang-Mills-Higgs », mais ces théories supersymétriques se révèlent plus rigides que les ordinaires; en particulier la géométrie analytique complexe entre en scène naturellement : les interactions des « superchamps » Φ_i sont gouvernées par deux types de termes dans le lagrangien, $\mathcal{K}(\Phi, \bar{\Phi})$ et $\mathcal{F}(\Phi)$; les Φ_i (qui sont des multiplets de vecteurs, fermions et bosons) se comportent comme des coordonnées holomorphes d'un « super-espace interne », le terme $\mathcal{F}$, appelé *super-potentiel*, en est une fonction holomorphe, alors que le terme $\mathcal{K}$ est la fonction génératrice d'une métrique kählerienne $g_{ij} = \partial^2 \mathcal{K} / \partial \Phi_i \partial \bar{\Phi}_j$. (cf. Wess et Bagger, *loc. cit.* ou Green, Schwarz et Witten *Superstring theory*, volume 2).

La première raison de faire confiance à la supersymétrie n'est pas d'origine expérimentale mais d'origine esthétique : des lagrangiens phénoménologiques longs et compliqués se ramassent en termes simples dans la langue supersymétrique. Une deuxième raison est le succès en mathématiques : après la superalgèbre et l'application aux théorèmes d'indice, la théorie de Seiberg-Witten en est une preuve. Un parallèle est tentant avec la géométrie symplectique : ce n'était pas pour coller aux expériences nouvelles que Lagrange donnait aux équations de la mécanique la forme hamiltonienne, c'était par souci de symétrie. Assez récemment la géométrie symplectique a montré sa puissance en géométrie et en topologie différentielle, de Poincaré et Birkhoff à Arnold et Gromov. Voir le chapitre écrit par Michèle Audin.

Les supersymétries garantissent des simplifications radicales de la renormalisation; c'est ce que N. Seiberg a su merveilleusement bien exploiter dans les quinze dernières années. Un point essentiel est la possibilité de prolongements analytiques (peut-être ambigus, c'est-à-dire sur des revêtements ramifiés de $\mathbb{C}$) des amplitudes renormalisées, en fonction des constantes de couplages. C'est ce que Seiberg a appelé la « puissance de l'holomorphic ».

Après que Seiberg ait intégré la théorie $N = 4$ en dimension 4, Seiberg et Witten ont presque entièrement résolu la théorie $N = 2$. Dans les théories de jauge supersymétriques $N = 2$ ou 4, on trouve encore des dyons; si le groupe de jauge G est brisé en un tore $H = U_1^r$, les dyons forment un réseau $\mathbb{Z}^{2r}$, par exemple; si $G = SU_2$, chaque dyon est représenté par un point dans $\mathbb{C}$, soit $Z = a n_e + a_D n_b$, $(n_e, n_b) \in \mathbb{Z}^2$, avec $a = \sqrt{u}$, $u = \langle \text{Tr}(\varphi^2) \rangle$. Les nombres complexes a et a_D apparaissent comme couplage

et couplage dual. Si $\mathcal{F}$ est le superpotentiel effectif, il vient

$$a_D = \frac{\partial \mathcal{F}}{\partial a}, \quad \tau = \frac{\theta}{2\pi} + i \frac{4\pi}{g^2} = \frac{\partial^2 \mathcal{F}}{\partial a^2}.$$

Le nombre u caractérise l'état du vide ; si $u \neq 0$ la symétrie SU_2 est (spontanément) brisée en U_1 , si $u = 0$ la symétrie n'est pas brisée (cf. § IV). Et si Λ est l'échelle pour renormaliser (cf. § II), le groupe de renormalisation β s'exprime par la dépendance de toutes les quantités, a, a_D, τ , etc. en fonction de u et Λ .

Pour $N = 4$ la renormalisation est « triviale », mais $N = 2$ réclame une renormalisation et $\mathcal{F}$ est une série de a et $\ln(a)$. On regarde a et a_D comme des fonctions de u et λ ; Seiberg et Witten démontrent que, lorsque a tend vers ∞ , on a

$$a_D \sim \frac{2ia}{\pi} \ln\left(\frac{a}{\Lambda}\right) + i \frac{a}{\pi}.$$

Ainsi la monodromie à l'infini (en u) de a_D est non triviale, ce qui force une singularité dans le plan de u , autre que $+\infty$. Un autre argument, prenant en compte la positivité de $\text{Im}(da_D d\bar{a})$, force un groupe de monodromie non abélien et au moins deux singularités en plus de ∞ sont nécessaires. (voir N. Seiberg et E. Witten, *Electric magnetic duality, monopole condensation and confinement in $N = 2$ supersymmetric Yang-Mills theory*, Nuclear Physics B 426 (1994), 19–52 et *Monopoles, duality and chiral symmetry breaking in $N = 2$ supersymmetric QCD*, Nuclear Physics B 431 (1994), 484–550.).

Tout d'abord, cela permet un Ansatz de solution complète, que personne ne conteste : a et a_D seraient les périodes de la forme différentielle $(x - u) dx / 2\pi y$ sur la courbe elliptique « universelle »

$$y^2 = (x^2 - \Lambda^4)(x - u).$$

(L'Ansatz pose en particulier que les deux singularités en u ajoutées par la quantification, sont $\pm\Lambda^2$; ceci se révèle une bonne normalisation pour rendre compte de la limite $\Lambda \rightarrow \infty$).

Et plus généralement, pour un groupe de jauge G de rang r , la théorie de champs renormalisée est décrite par une courbe complexe Σ de genre r variant au-dessus d'un tore de dimension r , (nommé « branche de Coulomb »). Plus précisément le groupe de renormalisation fait évoluer les couplages, interprétés comme des périodes d'une forme holomorphe, en fonction de Λ et des paramètres de brisure de symétrie, (cf. § IV) suivant un système dynamique algébriquement complètement intégrable associé à G par Hitchin. (Voir par exemple : Donagi et Witten, 1995).

Ensuite, une limite convenable de la théorie $N = 2$ supersymétrique près d'une des singularités $u = \pm\Lambda^2$ va donner ce qui s'appelle aussi « théorie de Seiberg-Witten » : une théorie de champs très particulière exploitée par Witten pour inspecter la topologie des variétés de dimension 4. En effet l'interprétation physique d'une singularité dans le plan des u est l'irruption de quelques particules sans-masse supplémentaires. Par exemple dans la théorie classique, en $u = 0$, ce sont les gluons. Ici, dans la théorie

quantique, pour $u = \pm\Lambda^2$, l'hypothèse la plus naturelle, retenue par Seiberg-Witten, est qu'un monopôle magnétique perde sa masse et domine la théorie en basse énergie. Près de ce point, (dans le langage équivalent d'une théorie associée $N = 1$, cf. Wess et Bagger *loc. cit.*) le monopôle est formé de deux champs de spineurs, $\psi, \tilde{\psi}$ et d'un photon dual A , le superpotentiel $\mathcal{F} = A\psi\tilde{\psi} + mU(A)$ (U polynomiale) est deviné par Seiberg-Witten (formule 5. 17 du premier article) et donne l'équation de Seiberg-Witten, pour une U_1 -connexion A sur un fibré en droite L et ψ un champ de spineurs, section de $S^+ \otimes L$:

$$\mathcal{D}_A \psi = 0, \quad F_A^+ = [\psi\tilde{\psi}]^+$$

(F_A^+ est la partie autoduale de la courbure et $[\psi\tilde{\psi}]^+$ la composante dans Λ_+^2 pour la décomposition $S^+ \otimes S^+ = \Lambda^0 + \Lambda_+^2$ cf. § I) (Références : E. Witten, *Monopoles and 4-manifolds*, Math. Res. Letters 1 (1994) 761–796 et le chapitre de J. Morgan dans ce livre.)⁽¹¹⁾

Lorsque certaines amplitudes $\langle A_1(x_1)A_2(x_2)\cdots \rangle$ d'une théorie de champs quantique sont des invariants topologiques de l'espace-temps sous-jacent ou de structures ajoutées dessus, comme les fibrés, on dit qu'on a affaire à une *théorie topologique des champs*. La pertinence de ces théories s'est imposée dans les années 80. Rétrospectivement on peut faire remonter leur origine à Dirac, 1931, ou même à Gauss et Ampère, 1821. Les principes des théories topologiques des champs, rejoignant et rassemblant une foule de résultats dûs à Donaldson, Floer, Jones et Gromov, furent exposés par Witten dans « *Topological Quantum fields theory* » et dans « *Quantum field theory and the Jones Polynomial* » en 1988 (voir le chapitre écrit par Pierre Vogel). Dans ces articles on trouve l'interprétation en terme d'amplitudes des invariants des nœuds et des variétés de dimension 3 et 4.

À l'origine du travail de Witten est une question de M. Atiyah et G. Segal. En effet, comme Atiyah venait de trouver une origine en théorie quantique de la théorie de Floer pour les sphères d'homologie de dimension 3, il demanda à Witten d'en faire autant pour la théorie de Donaldson. C'est alors (1988) que Witten inventa sa version « tordue » de Yang-Mills $N = 2$ supersymétrique dont les polynômes de Donaldson sont des amplitudes (la torsion signifie une entorse à la règle spin-statistique). Cette théorie, comme Yang-Mills ordinaire pour SU_2 ou SU_3 , est asymptotiquement libre, faiblement couplée en UV et fortement couplée en IR (cf. § II pour les définitions). Mais puisque les invariants de Donaldson sont indépendants de la métrique, on peut les calculer aussi bien en IR qu'en UV. Le problème était de contrôler le couplage fort de la nouvelle théorie. De premiers travaux de Seiberg, portant sur la théorie $N = 1$ en IR, ont permis à Witten d'obtenir les théorèmes voulus pour les surfaces complexes kählériennes avec $H^{2,0} \neq 0$ (cf. *Supersymmetric Yang-Mills Theory on a Four-manifold*, J. Math. Phys. 35 (1994)). À présent Witten peut faire appel à la théorie duale, celle des monopôles, qui est abélienne et facile à contrôler en IR, et définit des invariants de variétés différentiables de dimension 4. Ces nouveaux invariants

sont remarquablement efficaces. Reste (toujours en 2001) à démontrer que ce sont les mêmes que ceux de Donaldson ; cela confirmerait une partie de la « dualité » devinée. Les plus optimistes (dont Seiberg et Witten) pensent qu'il existe une déformation de ces théories « super » jusqu'à la théorie Yang-Mills-Higgs ordinaire, il en résulterait la dualité attendue pour les quarks et les gluons.

VI. Dualité des supercordes

Arrivé à ce point le lecteur sera sans doute déçu d'apprendre qu'il n'a pas encore vu le vrai pays des dualités. En effet, les spécialistes sont de plus en plus persuadés que le fin mot de l'histoire se trouve caché dans les propriétés non perturbatives des *théories de supercordes* (Voir par exemple, E. Witten, Physics Today, Mai 1997.) Une autre dualité que celles de Seiberg-Witten ou Montonen-Olive-Sen, celle des « symétries miroirs » est importante au chapitre I de Michèle Audin sur Gromov-Witten, où elle provient également de l'étude des supercordes ; d'où l'intérêt double d'un dernier détour. C'est la dernière excursion.

La théorie des supercordes est actuellement la seule théorie imaginée qui semble réussir à unifier toutes les forces fondamentales, en particulier elle a l'ambition de réconcilier la mécanique quantique et la gravitation. Depuis longtemps les principes de la relativité générale et ceux de la mécanique quantique souffrent d'une incompatibilité grave ; par exemple l'existence de « rayons de Schwarzschild » entraîne que, pour explorer des régions de plus en plus petites, il faut utiliser des objets de moins en moins lourds alors que les inégalités d'Heisenberg exigent le contraire. En dessous de la longueur de Planck ($\sim 10^{-32}$ cm) les deux théories paraissent irréconciliables. (Notons qu'à l'échelle usuelle des particules élémentaires, avec $\hbar = c = 1$, une telle longueur représente une gigantesque énergie, de l'ordre de 10^{-5} g, le poids de certaines bactéries). La théorie de la gravitation d'Einstein ne résiste pas aux petites échelles, elle est non renormalisable (*cf.* § II).

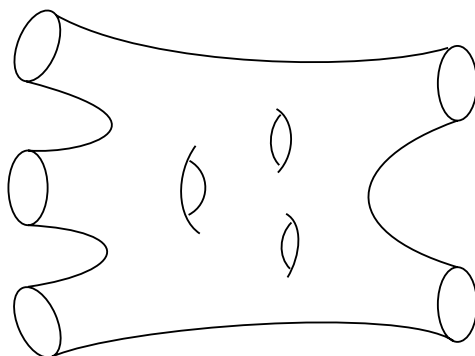
L'idée des cordes est d'attribuer aux particules très énergétiques, la forme d'une ficelle d'une longueur de l'ordre de 10^{-32} cm justement, et de « quantifier deux fois » les mouvements et les accidents de cette ficelle relativiste dans l'espace-temps de Minkowski pour retrouver, sans contradictions, toute la physique connue. D'où vient cette idée ? Quel est son rapport avec la gravitation ? Pour le comprendre, reportons-nous à la plus naïve des questions qui se pose en théorie quantique des champs : comment un champ devient-il une particule ? En effet, dans les chambres à bulles, ou parmi les fils tendus à la sortie des accélérateurs, toutes les particules élémentaires ont des trajectoires. Dans quelle approximation passe-t-on des amplitudes des connexions et des sections de fibrés spin à des projectiles ? La réponse est que la relation se fait en deux temps, en « deux quantifications » : une première quantification mène de l'action du point matériel (sans masse) dans l'espace de Minkowski à l'équation de Klein-Gordon $d * d\varphi = 0$; c'est l'équation de Schrödinger de la particule. La

deuxième quantification fait passer à une assemblée de champs φ , pour tenir compte des créations et annihilations notamment⁽¹²⁾.

Afin de rendre compte des interactions, il est possible, par exemple, d'introduire un potentiel φ^4 , ce qui donne la théorie de champ scalaire mentionnée au § II. Pour les particules chargées, les photons ou les super-points, les interactions peuvent s'expliquer en couplant les équations de Dirac, de Maxwell et d'Einstein. De même les quarks, les bosons W et Z , les gluons sont soumis au couplage de Dirac et de Yang-Mills. L'exigence de la renormalisation sélectionne sévèrement le choix des interactions.

La généralisation de ces quantifications aux cordes et supercordes relativistes n'est pas encore complètement comprise, mais les éléments recueillis depuis plus de 25 ans ont convaincu la plupart des théoriciens des particules de son intérêt.

En premier lieu, la vision des interactions est changée : dans une limite où $\hbar$ serait négligeable, mais pas la longueur typique des cordes, mesurée par la constante α' , inverse de la *tension de la corde*, on aurait affaire à des objets possédant une dimension d'espace et une de temps. Les trajectoires sont remplacées par des cylindres. Et, pour décrire les *interactions* de cordes il suffit de remplacer le cylindre par des surfaces dont le bord a plus de 2 composantes.



Le nombre n de trous compte le nombre de cordes à l'entrée et à la sortie de l'accident. On montre que le genre k de la surface (son nombre d'anses) mesure l'ordre des *corrections quantiques* dans les amplitudes de champs de cordes⁽¹³⁾.

La deuxième bonne nouvelle apportée par les cordes est la présence naturelle de particules véhiculant des interactions gravitationnelles : les *gravitons*. La première quantification des théories de cordes est une théorie de champ bidimensionnelle qui incorpore, au premier ordre en α' , la théorie de la gravitation d'Einstein classique à coté de la théorie Yang-Mills classique dans l'espace ambiant. D'ailleurs la constante α' vaut $\hbar \frac{G}{c^3}$, où G est la constante de gravitation universelle de Newton. Ainsi la théorie des cordes fixe toutes les unités de mesure de la physique, ce qui en ferait une « théorie de tout » (« theory of everything »). Mais aux ordres plus élevés en α' des corrections arrivent qui pourraient bien régulariser la gravitation aux grandes énergies

et guider sa quantification. Notons que les théories de cordes conservent un sens, au moins perturbatif, si l'on remplace l'espace temps plat de Minkowski par des variétés lorentziennes munies de métriques d'Einstein, *i.e.* solutions des équations d'Einstein classiques ; cela provient miraculeusement des équations du groupe de renormalisation (*cf.* Green, Schwarz, Witten et Polyakov, références ci-dessous).

Une autre « prédiction » des cordes est la *supersymétrie* : sans elle, c'est-à-dire sans passer aux supercordes, des particules aberrantes violant les principes de la relativité (*i.e.* se propageant plus vite que les photons) s'imposeraient subrepticement.

Enfin, et surtout, déjà la première quantification des cordes et supercordes est une théorie quantique de champ très exigeante : si l'on veut y maintenir, au niveau quantique, les symétries de la corde, ou supercorde classique (à savoir le groupe de (super) Poincaré ambiant (*cf.* § V), les difféomorphismes de la surface source, les dilatations locales de sa métrique (transformations de Weyl), les supersymétries de supersurface), il n'y a que 5 possibilités, toutes en dimension 10 d'espace-temps. Celles-ci portent les noms suivants : Type I, Type IIA, Type IIB, Hétérotique SO_{32} , Hétérotique $E_8 \times E_8$ ⁽¹⁴⁾.

Références : M.B.Green, J.H.Schwarz et E.Witten, *Superstring theory*, Vol. I et II Cambridge 1987, et A.M.Polyakov, *Gauge fields and strings* 1987.

Pendant longtemps (de 1975 à 1995 environ) on n'a connu de la théorie des cordes et supercordes deux fois quantifiée que des bribes perturbatives. Comme si, pour la théorie de Yang-Mills quantique, on ne savait calculer que les termes en g , g^2 ou g^3 des amplitudes. Mais depuis cinq ans des informations sont parues sur le « couplage fort » (*cf.* E.Witten, *Reflections on the fate of space-time*, Physics today, Avril 96 et J.Polchinski, *String duality*, *loc. cit.*). Les nouvelles connaissances s'appuient toutes sur d'extraordinaires dualités, plus spectaculaires encore qu'en théorie quantique des champs car elles échangent même les topologies et les dimensions des espaces-temps recevant les cordes.

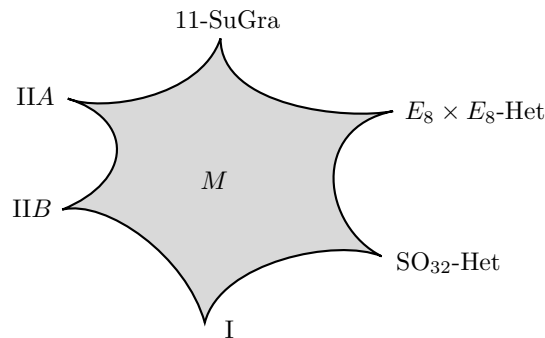
Les premières dualités des théories de cordes furent repérées vers 1990, dans les théories « hétérotiques » (A.Font, L.Ibanez, D.Lüst, F.Quevedo). Elles s'appellent *T-dualités*, *T* comme « target ». Elles changent la géométrie de l'espace ambiant et sont faible/fort pour certains couplages de jauge⁽¹⁵⁾.

En rapport avec le chapitre 1 sur Gromov-Witten, il faut mentionner que l'origine des « symétries miroirs » se trouve dans ces dualités⁽¹⁶⁾.

Ensuite, après les *T*-dualités, parut la *S*-dualité (1993, J.H.Schwarz et A.Sen, *Duality symmetries of 4D heterotic strings*, Physics Letters B 312, 105–114), échangeant les couplages faibles et forts de la corde ainsi que les charges électriques et magnétiques des champs de jauge, (dans les théories hétérotiques compactifiées sur 22+6 dimensions). Cette découverte a provoqué un changement radical du point de vue sur les cordes, ouvrant sur la théorie non perturbative. Elle fut vite suivie de la mise en

évidence de dualités faible/fort pour tous les types de cordes (type I, type II *etc.*), appelées *U*-dualités (*U* comme « unity ») (Townsend, Witten, 1995, *cf.* E. Witten, *String theory dynamics in various dimensions*, Nuclear Physics B 443 (1995) 85–126)⁽¹⁷⁾.

Il semble à présent que les dualités réussissent à permuter entre elles toutes les formes de supercordes connues, plus une nouvelle théorie mystérieuse, baptisée *M*-théorie. Il en sort l'espoir de définir une unique théorie renormalisée d'un genre nouveau, dont les théories de supercordes, et la *M*-théorie, seraient des limites variées. Par un glissement de sens bien compréhensible, la théorie *M* a fini par vouloir dire l'unique théorie non perturbative cherchée. Selon M.B. Green (1999), les dualités entre supercordes (et *M*-théorie) réussissent même à échanger des « limites classiques » avec des « effets quantiques » dans une autre théorie⁽¹⁸⁾.



Dans les limites de couplages forts, de nouveaux « objets étendus », généralisant en théorie des cordes et supercordes, les monopôles et les instantons des §§ III, IV, V, surviennent et dominent : il y a des *trous noirs*, des *membranes* ou *d-branes* (*i.e.* des sous-variétés de dimension $d + 1$ quelconque), des *D-branes*, *D* comme Dirichlet, qui sont des sous-variétés sur lesquelles les extrémités des cordes peuvent s'appuyer (voir Polchinski, *loc. cit.* et aussi J. Polchinski, *Combinatorics of boundaries in string theory*, Physical Review D, R 6041–6045, 1994).

C'est en *M*-théorie que la « bonne interprétation » des monopôles de Seiberg-Witten paraît résider, E. Witten « *Solutions of four-dimensional field theories via M-theory* » Nuclear Physics B 500 (1997) 3–42 : (pour ceux qui n'ont pas encore le vertige) les états BPS donnant lieu aux monopôles seraient des « approximations effectives en basse énergie de solitons en *M*-théorie, (des 2-branes), localisés sur des variétés de dimension 3, accrochées à des *D*-branes de dimension 6, dans un espace-temps courbe à 11 dimensions de la forme $X \times \mathbb{R}^7$, où X est une surface projective complexe ». Ces variétés de dimension 3 sont produits d'une droite $\mathbb{R}$ dans un $\mathbb{R}^4$ fixé de $\mathbb{R}^7$ avec un disque holomorphe D dans X , dont le bord appartient à une surface de Riemann Σ fixée. C'est cette surface Σ , « bien sûr », qui est à la base du système intégrable de Hitchin intervenant dans la théorie de Seiberg-Witten.

VII. Géométrie des invariants

Le sens classique du mot « invariant » dans le contexte d'une « théorie des invariants » repose sur la combinaison de trois termes ; par exemple :

- l'excentricité est un invariant de l'ellipse pour les déplacements,
- le discriminant est un invariant de l'équation du second degré pour les homographies,
- la signature est un invariant de la variété par cobordisme.

De façon générale, la quantité (ou qualité) Z est un invariant de l'objet X , par la transformation T lorsque $T(X) = X'$ entraîne $Z(X') = Z(X)$. Le plus souvent, X est un objet d'une certaine « catégorie », T un morphisme et Z un foncteur, mais il peut arriver que X soit une catégorie, T un foncteur, Z une transformation naturelle, *etc.* (À la rubrique « invariant » le dictionnaire mentionnera n'importe quoi qui reste fixe quand autre chose change, mais ça ne fait pas une théorie pour autant). La définition s'étend facilement aux « covariants », *etc.*

Dans les exemples plus anciens, les transformations T considérées forment un groupe, mais dans les trois chapitres que réunit ce livre-ci, T est bien une flèche dans une catégorie. Pour Michèle Audin, principalement X est une variété symplectique (W, ω) (ou une structure presque-complexe sur W), T une transformation symplectique, Z une suite de nombres ; pour Pierre Vogel, X est une variété de dimension 3 (ou un nœud), T un homéomorphisme, Z une fonction (dépendant aussi d'une algèbre de Lie ou d'une autre structure) ; pour John Morgan, X est une variété de dimension 4, T un difféomorphisme, Z une fonction des structures spin_c sur X . Notons qu'à chacune des théories présentées correspond aussi un « groupe d'invariance », celui des transformations symplectiques de W , celui des homéomorphismes de X ou celui des difféomorphismes de X , et qu'on s'intéresse aussi à des invariants de sous-objets particuliers de X relatifs à ces groupes. Il y a chaque fois ces deux aspects passif et actif, d'une théorie de relativité dont parle H. Weyl (*Classical groups*).

Au début de l'histoire des invariants, disons du XVII^{ème} au XIX^{ème} siècle, l'accent était mis sur les objets X ; c'est-à-eux qu'on semblait vouloir accéder à travers les invariants.

Puis l'accent s'est déplacé sur les transformations T , avant tout sur les groupes qu'elles forment. Le livre d'Hermann Weyl sur les groupes classiques le raconte bien : la théorie des représentations linéaires des groupes a remplacé la classique « théorie des invariants ». Les principes de Klein (le célèbre « programme d'Erlangen ») l'ont proclamé : le groupe doit primer sur l'espace géométrique. Même s'il reprend ses droits, comme dans la théorie des espaces symétriques de Cartan (ou des immeubles de Tits) (*cf.* l'exposé de J. Tits au congrès international de 1974 à Vancouver) l'objet X est dérivé, ou un outil. De même le développement des catégories, l'apparition de nouveaux morphismes comme les cobordismes de Pontrjagin et Thom (*cf.* l'exposé de H. Hopf au congrès international de 1958 à Edimbourg), ont insisté sur l'intérêt de mettre en tête le côté « transformations ».

Mais la particularité des invariants contemporains, le signe de l'inflexion actuelle, me semble être la mise en avant de Z ; l'invariant lui-même a priorité sur l'objet et la symétrie. C'est avant tout la « forme de l'invariant » qui compte. Ce qui rappelle la transition du quantique au classique : en premier arrive un « spectre », une collection de nombres ou d'amplitudes, qui se mesurent, et ensuite, on leur attache des représentations géométriques, des modèles qui donnent un sens. Par exemple, Z est une fonction de partition, une fonction génératrice, mais elle traduit diverses théories duales $U, \widehat{U}, V, \widehat{V}, \dots$, plusieurs intégrales enveloppant divers champs (comme disait Gilles Châtelet), sur des variétés de topologies différentes, presque indifférentes. Si des objets X, X' viennent au secours de Z , les théorèmes à établir les concernent sous forme d'une correspondance entre X et X' , souvent surprenante. Ensuite on baptise « symétrie » un analogue de T : la théorie des groupes quantiques est typique ; elle qui sert en retour à construire des invariants de nœuds, est apparue d'abord à la place des symétries pour garantir des relations de commutations, des invariances de spectres d'opérateurs, dans la théorie des systèmes complètement intégrables en mécanique quantique et en mécanique statistique (Faddeev, Reshetikhin, Takhtadjan). Le nouveau point de vue est source de « nouvelles symétries ».

Dans les travaux de Grothendieck, on pouvait sentir l'ébranlement de l'ordre X, T, Z ; une structure algébrique au départ, des constructions d'algèbre homologique, d'infini-catégories appelaient tout un réseau de relations d'invariance. Le nouvel aspect est spécialement marqué dans les travaux de Kontsevich (*cf.* le chapitre écrit par Pierre Vogel). Il n'y a qu'à voir comment, partant de classes d'homologie d'algèbre, il aboutit aux invariants de variétés 3 (*Feynman diagrams and low dimensional topology*, 1994), ou comment il propose d'expliquer le phénomène des symétries miroirs, la dualité entre invariants de Gromov-Witten des variétés symplectiques et les faisceaux sur des variétés algébriques, à partir d'algèbres homotopiques A_∞ de Stasheff (*Homological Algebra and Mirror Symmetry*, 1994). La présence partout dans ces travaux des *modules de courbes*, est le signe d'une métamorphose des invariants.

Faible ou fort, électrique ou magnétique, Infini ou rien ?

Est-ce que la méditation physique et mathématique s'accorde au plus humain ? « Entre le rêve et la vie..., la nuit et le jour, l'invisible et le visible », le poète cherchait sa voie. Dans *Aurélia*, Gérard de Nerval écrivait :

La dame que je suivais, développant sa taille élancée dans un mouvement qui faisait miroiter les plis de sa robe en taffetas changeant, entoura gracieusement de son bras nu une longue tige de rose trémière, puis elle se mit à grandir dans un clair rayon de lumière, de telle sorte que peu à peu le jardin prenait sa forme, et les parterres et les arbres devenaient les rosaces et les festons de ses vêtements ; tandis que sa figure et ses bras imprimaient leurs contours aux nuages pourprés du ciel. Je la perdais de vue, à mesure qu'elle se transfigurait, car elle semblait s'évanouir dans sa propre grandeur « Oh ! ne fuis pas !, m'écriai-je, ... car la nature meurt avec toi ! »

Notes

⁽¹⁾Un fibré principal $P \xrightarrow{\pi} M$ de groupe structural G est l'application quotient d'une action libre de G sur P à droite; dans une trivialisation locale au-dessus de $U \subset M$, $P|_U$ s'identifie à $U \times G$. Une connexion principale ∇ sur P est un sous-fibré du fibré tangent à P , invariant par G , qui permet de relever à P les chemins dans M ; en trivialisation locale, ∇ est donnée par une 1-forme différentielle A à valeurs dans l'algèbre de Lie $\mathfrak{g}$ de G ; étant donné un point $u_0 \in U$, un élément g_0 de G et un chemin $\gamma : [0, 1] \rightarrow U$, tel que $\gamma(0) = u_0$, l'unique chemin $(\gamma, g) : [0, 1] \rightarrow U \times G$ relevant γ , tel que $g(0) = g_0$, est obtenu en résolvant l'équation $\frac{dg}{dt}g^{-1} = -A\left(\frac{d\gamma}{dt}\right)$.

⁽²⁾Sur une variété W de la forme $M^3 \times \mathbb{R}$ (avec une métrique produit et le fibré P trivial), toute connexion est équivalente, modulo changement de jauge, à une connexion $\nabla = d + A$, telle que A n'ait pas de terme en dt ($t \in \mathbb{R}$). Si bien que ∇ peut être vue comme une trajectoire de 1-formes différentielles sur W à valeurs dans $\mathfrak{g}$. L'équation d'antidualité s'interprète alors comme l'équation différentielle associée au gradient de la fonctionnelle de Chern-Simons

$$CS(A) = \int_M \text{Tr} \left(A \wedge dA + \frac{2}{3} A \wedge A \wedge A \right).$$

Pour récupérer les invariants topologiques les plus intéressants, il ne suffit pas de considérer la théorie classique, la théorie quantique est nécessaire. (Notons que la collection complète des invariants définis par Reshetikhin et Turaev demande de faire varier le groupe de jauge et ses représentations.)

Si l'on considère le cas particulier des variétés M de la forme $\Sigma^2 \times \mathbb{R}$ (donc $W = \Sigma \times \mathbb{R}^2$) où Σ est une surface de Riemann, avec la métrique produit sur M (et W), on tombe sur les équations de Cauchy-Riemann et la théorie de Gromov-Witten n'est pas loin. En effet, notons s, t les coordonnées sur $\mathbb{R}^2$ et u, v les coordonnées sur Σ . Soit A une 1-forme sans terme en dt , une nouvelle trivialisation de P enlève la composante en ds ; alors A s'écrit $\alpha du + \beta dv$ et l'équation d'antidualité équivaut au système suivant :

$$\frac{\partial \alpha}{\partial t} = \frac{\partial \beta}{\partial s}, \quad \frac{\partial \beta}{\partial t} = -\frac{\partial \alpha}{\partial s}, \quad \frac{\partial \alpha}{\partial v} - \frac{\partial \beta}{\partial u} = \alpha \wedge \beta - \beta \wedge \alpha,$$

c'est-à-dire les équations de Cauchy-Riemann pour $\alpha + i\beta$ en fonction de $t + is$, jointes aux équations de courbure nulle de A . Si $G = U_1$ on a affaire aux fonctions holomorphes de deux variables $t + is, u + iv$ (ou $u - iv$), pour G compact simple la théorie de Wess-Zumino-Witten sur Σ intervient.

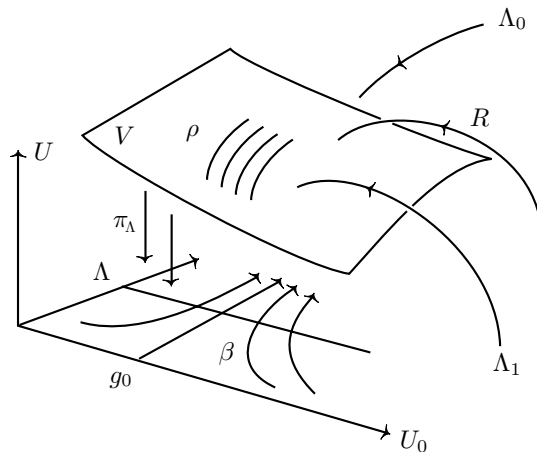
(La théorie initiale de Gromov, développée en particulier par Floer, avant que Witten ne la place en perspective physique, se présentait comme la délinéarisation des équations de Cauchy-Riemann).

⁽³⁾Les équations de Seiberg-Witten portent sur deux champs, A et ψ : si (e_1, e_2, e_3, e_4) est un repère de l'espace en dimension 4 et (φ_1, φ_2) une trivialisation de S^+ , $A = \alpha_1 e_1 + \dots + \alpha_4 e_4$, $\psi = (B_1 \varphi_1 + B_2 \varphi_2)$, avec $B_1 = \beta_1 + i\beta_2$, $B_2 = \beta_3 + i\beta_4 \in \mathbb{C}$. Posons $A_1 = \alpha_1 + i\alpha_2$ et $A_2 = \alpha_3 + i\alpha_4$ et introduisons les opérateurs $P_1 = i\nabla_1 + \nabla_2$, $P_2 = i\nabla_3 + \nabla_4$, où ∇_i est la dérivation covariante dans la direction e_i .

Les équations de Seiberg-Witten s'écrivent :

$$\begin{bmatrix} \overline{P}_2 & -\overline{P}_1 \\ P_1 & P_2 \end{bmatrix} \begin{bmatrix} B_2 \\ B_1 \end{bmatrix} = \begin{bmatrix} -A_2 & A_1 \\ -\overline{A}_1 & \overline{A}_2 \end{bmatrix} \begin{bmatrix} B_2 \\ B_1 \end{bmatrix}, \quad \begin{bmatrix} \overline{P}_2 & -\overline{P}_1 \\ P_1 & P_2 \end{bmatrix} \begin{bmatrix} A_2 \\ A_1 \end{bmatrix} = \begin{bmatrix} B_2 & \overline{B}_1 \\ B_1 & -\overline{B}_2 \end{bmatrix} \begin{bmatrix} \overline{B}_1 \\ \overline{B}_2 \end{bmatrix}.$$

⁽⁴⁾Donnons une idée de la construction de V , ρ , et π , ou plutôt racontons l'histoire comme elle devrait se passer (ce qui est assez loin d'être prouvé). Pour simplifier je vais partir de la régularisation tronquant les grandes impulsions. On commence par plonger U_0 dans un espace U assez grand de lagrangiens (en général U est de dimension infinie), en supposant que chaque point de U donne une valeur finie de l'intégrale fonctionnelle tronquée au-delà d'une certaine fréquence. En outre, pour tout couple (Λ_0, Λ_1) dans $\mathbb{R}_+^\times \times \mathbb{R}_+^\times$ tel que $\Lambda_1 < \Lambda_0$ on se donne une fibration $\pi_{\Lambda_1, \Lambda_0}$ de U sur U_0 qui impose la valeur de certaines amplitudes tronquées par Λ_0 à l'échelle physique d'observation Λ_1 . (Dans l'exemple φ^4 , les valeurs de a, m^2, λ sont données par le développement en impulsions de l'intégrale du lagrangien jusqu'aux monômes de degré ≤ 4 en φ .) On suppose en plus que, partant de $\mathcal{L}$ dans U , tronquant l'intégrale de Feymann au-dessus de Λ_0 et calculant la moyenne sur les Λ dans l'intervalle $[\Lambda_1, \Lambda_0]$, on obtient les mêmes amplitudes qu'en tronquant simplement à Λ_0 un autre lagrangien $\mathcal{L}' = R_{\Lambda_1, \Lambda_0}(\mathcal{L})$, quitte à faire un changement de variables $x' = f(x)$ et normaliser autrement les champs $A'_i = Z_i A_i$. Ce $\mathcal{L}'$ est le « potentiel effectif » selon Wilson.



Le (semi-) groupe de renormalisation est défini, en action multiplicative, sur $U \times]0, 1[$ par $R_\lambda(\mathcal{L}, \Lambda) = (R_{\lambda\Lambda, \Lambda}(\mathcal{L}), \lambda\Lambda)$. On a $R_{\Lambda_2, \Lambda_1} \circ R_{\Lambda_1, \Lambda_0} = R_{\Lambda_2, \Lambda_0}$ donc $R_{\lambda\mu} = R_\lambda \circ R_\mu$, $\lambda < 1$, $\mu < 1$. La variété $U_{\Lambda_1, \Lambda_0} = R_{\Lambda_1, \Lambda_0}(U_0)$ est censée couper transversalement les fibres de $\pi_{\Lambda_1, \Lambda_0}$. Tout cela est espéré pour que la théorie soit renormalisable mais c'est surtout le point suivant qui constitue un « théorème de renormalisation » : lorsque Λ_0 tend vers $+\infty$ la variété U_{Λ_1, Λ_0} converge vers une sous-variété U_{Λ_1} de U et la restriction de $\pi_{\Lambda_1, \Lambda_0}$ à U_{Λ_1, Λ_0} converge vers une application π_{Λ_1} de U_{Λ_1} sur U_0 . Pour tout couple $\Lambda_1 < \Lambda_2$ on a $R_{\Lambda_1, \Lambda_2}(U_{\Lambda_2}) = U_{\Lambda_1}$. Les $U_\Lambda \times \{\Lambda\}$ forment une sous-variété V de $U \times \mathbb{R}_+^\times$ que la collection des π_Λ envoie dans $U_0 \times \mathbb{R}_+^\times$; le semi-groupe R induit un flot sur V ; c'est l'invariance d'échelle qui assure que ce flot provient d'un champ de vecteurs β sur U_0 . On peut voir V et son feuilletage ρ comme la « vraie théorie » (ou l'ensemble des vraies théories), l'identification à $U_0 \times \mathbb{R}_+^\times$ et le champ β sont des souvenirs du lagrangien classique. L'autre relation est $V = \lim_{\lambda \rightarrow 0} R_\lambda(U_0 \times \mathbb{R}_+^*)$ mais pour $w = (u, \Lambda_0) \in W$ le point u_λ de U_0 tel que $u = R_{\Lambda, \lambda\Lambda_0}(u_\lambda)$ n'a pas de limite dans U_0 quand $\Lambda \rightarrow \infty$, $\lambda\Lambda_0 \rightarrow \Lambda$.

⁽⁵⁾Rappelons (cf. § II) que les théories renormalisées, calculant les « vraies » amplitudes sont décrites par un feuilletage ρ sur une variété V : il peut donc arriver que deux projections $\pi : V \rightarrow U_0 \times \mathbb{R}_+^*$ et $\hat{\pi} : V \rightarrow \hat{U}_0 \times \mathbb{R}_+^*$ engendrent des dynamiques β et $\hat{\beta}$ très différentes sur U_0 et $\hat{U}_0$, une source pour β peut devenir un puits pour $\hat{\beta}$ et réciproquement.

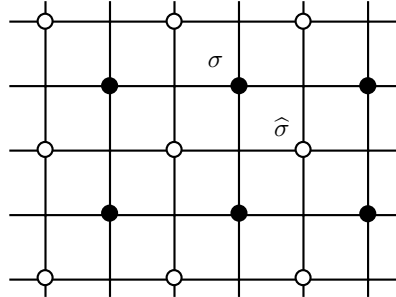
⁽⁶⁾Ce qui remplace le champ A (ou ψ) est une application σ de $\mathbb{Z}^2$ dans $\{+1, -1\}$ appelée configuration de spin. Les amplitudes sont remplacées par les corrélations de spin : pour $x_1, x_2, \dots, x_n$ dans $\mathbb{Z}^2$,

$$\langle \sigma(x_1) \cdots \sigma(x_n) \rangle = \frac{1}{Z} \sum_{\sigma} e^{-\beta H(\sigma)} \sigma(x_1) \cdots \sigma(x_n)$$

où $Z = \sum_{\sigma} e^{-\beta H(\sigma)}$, est la fonction de partition. Les sommes portent sur toutes les configurations possibles σ , le coefficient β est le produit d'une constante universelle par l'inverse de la température T , il joue le rôle du couplage, et $H(\sigma)$ est l'énergie de la configuration σ ($H(\sigma)$ est une combinaison linéaire des produits de deux spins voisins). Pour que les sommes aient un sens, il faut remplacer $\mathbb{Z}^2$ par une partie finie ; ce qui est une troncature pour régulariser.

Le système des spins à la température T sur un réseau plan carré est dual du système de spins sur le réseau dual à une température $\hat{T}$ de l'ordre de $1/T$. (Plus exactement il existe une constante J telle que $\text{sh}(J\hat{T}^{-1}) \text{sh}(JT^{-1}) = 1$). Le réseau dual s'obtient en mettant un sommet au centre de chaque face :

Kadanoff et Ceva ont défini des fonctions à valeurs réelles sur le réseau dual, une fonction $\mu_\sigma(x)$ pour chaque configuration σ , appelées « désordres », ou « mesures du désordre » ; on peut les interpréter comme des champs duaux $\hat{\sigma}$ des σ .



Il existe un procédé, nommé « renormalisation par blocs de spins », comprenant troncature et moyennisation (cf. § II) qui permet de prendre la limite des fonctions de corrélations lorsque, de façon coordonnée, $|T - \hat{T}|$ tend vers 0, la taille des boîtes finies tend vers l'infini et la maille du réseau tend vers 0. Si $T = \hat{T}$, T s'appelle *température critique*; la limite s'appelle *limite critique*. Le réseau devient infini en même temps que sa maille rétrécit et les corrélations renormalisées sont celles d'une théorie de champs quantique en temps imaginaire, dite « euclidienne ». Il est possible d'identifier cette limite à la théorie euclidienne bidimensionnelle d'un champ scalaire φ de lagrangien $\frac{1}{2}(\nabla\varphi)^2 - \frac{m^2}{2}\varphi^2 - \frac{\lambda}{24}\varphi^4$ (cf. § II). Le λ joue le rôle de T^{-1} , une dualité échange $\lambda \rightarrow \infty$ et $\lambda \rightarrow 0$.

⁽⁷⁾Cette convention interprète g comme un coefficient de dilatation dans l'algèbre de Lie $\mathfrak{g}$, elle donne aussi $F_A = dA + gA \wedge A$. Mais il est aussi fréquent de rencontrer g sous la forme d'un facteur g^{-2} devant $\mathcal{L}(A)$. L'équivalence entre les deux points de vue est discutée par exemple dans le livre *Quantum field theory* d'Itzykson et Zuber. À vrai dire, dans notre monde actuel, à mi-chemin entre la parfaite harmonie et le chaos bestial, la théorie de jauge n'est pas « unifiée » et il est nécessaire d'introduire plusieurs constantes de couplages de jauge, $g_1, g_2, \dots$ une par facteur simple de G , une pour le facteur abélien U_1 , etc. Par exemple, la constante des interactions faibles est bien plus petite que celle des interactions fortes.

⁽⁸⁾Lorsqu'une fonction de potentiel classique possède plusieurs minima, certaines solutions de la théorie classique associée (par les équations de Newton) sont localisées près des puits, mais il leur correspond des états asymptotiques de la théorie quantique entre lesquels la probabilité de transition (calculée par l'équation de Schrödinger) n'est pas nulle. C'est ce qui s'appelle « l'effet tunnel » en mécanique quantique. On constate alors que ces probabilités de « sauter au-dessus des barrières de potentiels » ont un ordre de grandeur de la taille d'exponentielles $\exp(S)$, où S est l'action de solutions classiques prises en temps imaginaire pur. Comme les instantons, les monopôles qui sont utiles en théorie quantique des champs sont des solutions de la théorie classique euclidienne (c'est-à-dire en temps imaginaire), leur effet est une modification des amplitudes par l'exponentielle de leur action.

⁽⁹⁾Notons, selon la discussion de Polchinski (« *String duality* » *loc. cit.* p. 1248), que le paramètre α qui sert à développer les amplitudes en séries $\sum a_n \alpha^n$ est $\hbar g^2$; or, dans une théorie renormalisable et « Borel resommable », l'erreur due à la resommation est de l'ordre de $\exp(-1/\alpha)$; elle est donc comparable aux « effets topologiques ». Des conséquences observables expérimentalement de ces effets ont malgré tout pu être établies, elles sont discutées par exemple dans 't Hooft « *Under the spell...* » *loc. cit.*

⁽¹⁰⁾Plus généralement on peut tenter l'analyse des théories de Yang-Mills avec tout groupe de jauge compact G et des champs de Higgs ramenant spontanément la symétrie à un sous-groupe H . Goddard, Nuyts et Olive (« *Gauge theory and magnetic charge* », Nuclear Physics B 125 (1977)) ont démontré que la « charge magnétique » des monopôles de 't Hooft et Polyakov est naturellement une représentation d'un groupe de Lie compact $H^\vee$ qu'on peut construire explicitement à partir de H , le « groupe dual ». Ils montrent $(H^\vee)^\vee = H$. Par exemple $U_1^\vee = U_1$, alors la charge est un entier. Et ils suggèrent l'existence d'une vraie dualité faible/fort échangeant H et $H^\vee$. Les nombres quantiques de Noether (électrique) et topologiques (magnétiques) s'échangeraient. Les monopôles pour H formeraient une théorie de jauge pour $H^\vee$.

Il est très remarquable, mais aussi rarement signalé, que le groupe dual $H^\vee$ est exactement celui que R.P. Langlands avait défini à la fin des années 60 pour les besoins de l'arithmétique et de la théorie des groupes. Par exemple $SU_n^\vee = PSU_n = SU_n/(\mathbb{Z}/n\mathbb{Z})$. La dualité induit la transposition des matrices de Cartan, ainsi les séries A, D, E, F, G sont auto-duales mais B et C s'échangent. Le groupe dual de Langlands intervient dans un vaste ensemble de conjectures reliant l'analyse harmonique, les formes automorphes et la théorie des nombres, pour donner des lois de réciprocité nouvelles. (Langlands parle de groupes réductifs complexes ; ce sont les complexifications des groupes compacts. Il travaille aussi avec des adèles et pas seulement sur $\mathbb{R}$ ou $\mathbb{C}$, ce qui fait une « corde arithmétique ». cf. « *Problems in the theory of automorphic forms* » in Lectures in Modern Analysis and Applications III, Lect. Notes in Math 170 (1970)). Il serait extraordinaire qu'une même théorie mathématique à venir, éclaire à la fois la vraie nature des quarks et la correspondance de Langlands (dont un cas particulier traité par A. Wiles a entraîné le théorème de Fermat).

⁽¹¹⁾L'origine supersymétrique des équations de Seiberg-Witten explique les symétries apparentes du système de la note (3).

⁽¹²⁾Pour une particule chargée soumise à une force de Lorentz, l'équation de Dirac sur un champ ψ remplace l'équation de Schrödinger ; la théorie quantique des champs la traite comme une équation de « champ classique ». Pour le champ électromagnétique, on a l'air de sauter la première quantification, mais ce n'est qu'une impression passagère, car le « champ quantique » associé est considéré aussitôt comme une assemblée d'oscillateurs (avec polarisation *etc.*) : la particule réintroduite est le *photon*. La généralisation supersymétrique, c'est-à-dire la quantification d'un point matériel

dans le super-espace de Minkowski (*cf.* § V), donne les équations de Maxwell et de Dirac découplées, à coté de la linéarisation des équations d'Einstein.

⁽¹³⁾En théorie des cordes les champs $A(x)$ deviennent des expressions $A(x(z, \bar{z}))$ où z désigne une coordonnée d'un point sur une surface de Riemann. Les amplitudes sont des moyennes sur des espaces de modules de structures conformes de surfaces de Riemann avec des points marqués.

⁽¹⁴⁾La théorie de type I est la seule à comporter des cordes *ouvertes* (*i.e.* à extrémités libres) en même temps que les cordes *fermées*; elle possède un groupe de jauge SO_{32} et une supersymétrie. Les types II possèdent $N = 2$ supersymétries d'espace-temps ambiant et aucun groupe de jauge; le type II A est symétrique par réflexion de l'espace ambiant alors que le type II B ne l'est pas (conditions de chiralités). Les théories hétérotiques sont des théories de cordes fermées où les ondes se propageant d'un coté sur le cercle (par exemple à droite) sont de type bosonique alors que les ondes se propageant de l'autre coté sont fermioniques; un groupe de jauge est autorisé, mais il ne peut être que SO_{32} ou $E_8 \times E_8$. À vrai dire les cordes bosoniques doivent se propager dans un espace-temps de dimension 26, donc la corde hétérotique, en un certain sens, se meut dans un espace-temps de dimension $26 + 10 = 36$.

⁽¹⁵⁾Une conséquence de cette dualité T est l'existence d'un plus petit diamètre possible pour les dimensions compactifiées dans l'espace ambiant, *cf.* E. Witten, Physics Today, Avril 96 *loc. cit.* Le mécanisme est le suivant : une particule sans masse, ou une corde, se propageant le long d'un cercle de rayon R voit son moment quantifié $p = nR^{-1}$, avec $n \in \mathbb{N}$, et son énergie de même $E_n = |n|R^{-1}$. Mais une corde peut aussi s'enrouler m fois autour du cercle, elle possède alors une énergie $|m|R(\alpha')^{-1}$. Il y a une dualité, « généralisant la dualité du modèle d'Ising » (vue au § III ci-dessus), qui échange les deux spectres en échangeant R et $\alpha'R^{-1}$. Montrons, suivant Witten, comment ceci explique la longueur de Planck comme limite aux explorations possibles de l'espace : d'après Heisenberg, à l'échelle d'un moment p , on explore une distance $x \sim \hbar p^{-1}$. Il semblerait donc qu'avec des p de plus en plus grands, on puisse explorer x de plus en plus petit. Mais en théorie des cordes, ce « microscope de Heisenberg » ne fonctionne plus : si l'énergie devient trop grande, la corde s'étend et au lieu d'explorer de petites distances on assiste à la propagation d'une grande corde. Grossièrement, tout se passe comme si le principe d'incertitude avait deux termes

$$\Delta x \geq \frac{\hbar}{\Delta p} + \alpha' \frac{\Delta p}{\hbar},$$

le premier terme étant celui des relations d'incertitude quantique habituelle, le second terme reflétant une nouvelle incertitude ou bizarrerie (fuzziness) due à la théorie des cordes. Avec les deux termes ensemble, il apparaît un *minimum absolu de l'incertitude sur les longueurs*, d'ordre $\sqrt{\alpha'} \sim 10^{-32}$ cm dans toute expérience.

⁽¹⁶⁾Comme les seules théories de supercordes satisfaisantes sont à valeurs dans une variété W de dimension 10, une possibilité pour rattacher ces théories au monde apparent (et retomber sur terre) est que W soit le produit de $\mathbb{R}^4$ (ou d'un espace-temps einsteinien de dimension 4) avec une variété de tout petit diamètre de dimension 6. Cette variété a tendance à être une *variété de Calabi-Yau*, c'est-à-dire une variété analytique complexe X de dimension 3 sur $\mathbb{C}$ munie d'une métrique kählérienne dont la forme volume est la norme d'une 3-forme holomorphe. Dans la limite où α' tend vers 0, la théorie des supercordes à valeurs dans X est une théorie de champ à 2 dimensions avec une invariance (super) conforme, qui détermine les particules visibles dans $\mathbb{R}^4$. Une dualité (faible/fort) est conjecturée, elle est supposée échanger la théorie à valeurs dans X avec une théorie du même genre à valeurs dans une autre variété de Calabi-Yau. Cette dualité a lieu avec le type II ; elle échange le type II A avec le type II B. La variété Y s'appelle *variété miroir* de X ; on a $h^{1,0}(Y) = h^{1,1}(X)$ et inversement $h^{1,1}(Y) = h^{1,0}(X)$. Les intégrales de Feymann se localisent, pour II A du côté X , sur les courbes holomorphes dans X alors qu'elles se localisent directement dans Y pour le modèle dual. Les invariants de Gromov-Witten de X se calculent ainsi à partir de la théorie des variations des structures complexes de Y . Les prédictions numériques, commencées à partir des années 80 (Lerche, Vafa, Warner...), comptant des courbes holomorphes dans X , se vérifient peu à peu (cf. par exemple M. Kontsevich, exposé 801 au Séminaire Bourbaki. *Mirror symmetry in dimension 3*, Juin 95.) Les espaces X et Y donneraient des résultats très différents en physique classique ou quantique usuelle, mais ils sont équivalents du point de vue des supercordes. « En théorie des cordes il n'y a pas réellement d'espace-temps classique, mais seulement une théorie de champ bidimensionnelle » (cf. M. Kontsevich *loc. cit.*)

⁽¹⁷⁾Ces dualités s'organisent pour former des sous-groupes discrets des groupes de Lie de symétrie globales des théories classiques de champs ou de cordes sous-jacentes (supergravité, super Yang-Mills, etc.) Par exemple les premières T -dualités forment un groupe $O(6, 22; \mathbb{Z})$ et les premières S -dualités un sous-groupe $SL_2(\mathbb{Z})$ d'un $SL_2(\mathbb{R})$, analogue à celui de Montonen-Olive-Sen. Pour une application des dualités à la théorie des formes automorphes, voir l'exposé Bourbaki 891 de M. Kontsevich, *Product formulas for modular forms on $O(2, n)$* [after R. Borcherds], Juin 95.

⁽¹⁸⁾Il existe en dimension 11 d'espace-temps une exceptionnelle théorie supersymétrique de la gravitation, une théorie de *supergravité* (cf. Wess et Bagger, *loc. cit.*), celle de dimension maximale ; elle fut découverte en 78 par E. Cremmer, B. Julia et J. Scherk. En compactifiant cette supergravité à 11 dimensions sur un cercle, on sait retrouver une théorie de supergravité à 10 dimensions, qui se trouve être une limite de « basses énergies » ($\alpha' \sim 0$) de la théorie de supercorde II A. La nouvelle théorie M (conjecturée), (M comme matrice, membranes, merveilles, etc.), se tiendrait dans un espace-temps de dimension 11, elle aurait aux « basses énergies », comme limite

la supergravité de Cremmer, Julia, Scherk et redonnerait la théorie II A par « compactification d'une dimension sur un cercle ».

Le couplage fort d'une théorie de type I est équivalent au couplage faible de l'hétérotique SO_{32} . Les couplages forts des théories II A, ainsi que hétérotique $E_8 \times E_8$ sont liés à la théorie M en couplage faible. La théorie II B voit ses couplages forts et faibles échangés, avec une symétrie $SL_2(\mathbb{Z})$ complète, *etc. cf.* Polchinski *loc. cit.* et Witten *loc. cit.*

PANORAMAS ET SYNTHÈSES

2001

11. M. AUDIN, J.W. MORGAN, P. VOGEL, D. BENNEQUIN – *Nouveaux invariants en Géométrie et en Topologie* (édité par F. DUMAS, J.-Y. LE DIMET, S. PAYCHA)

2000

10. C. ANÉ, S. BLACHÈRE, D. CHAFAÏ, P. FOUGÈRES, I. GENTIL, F. MALRIEU, C. ROBERTO, G. SCHEFFER – *Sur les inégalités de Sobolev logarithmiques* (avec une préface de D. BAKRY et M. LEDOUX)
9. W. CRAIG – *Problèmes de petits diviseurs dans les équations aux dérivées partielles*

1999

8. D. CERVEAU, É. GHYS, N. SIBONY, J.-C. YOCCOZ (avec la collaboration de M. FLEXOR) – *Dynamique et géométrie complexes*
7. X. BUFF, J. FEHRENBACH, P. LOCHAK, L. SCHNEPS, P. VOGEL – *Espace de modules des courbes, groupes modulaires et théorie des champs*

1998

6. D. ALPAY – *Algorithme de Schur, espaces à noyau reproduisant et théorie des systèmes*

1997

5. C. KASSEL, M. ROSSO, V. TURAEV – *Quantum groups and knot invariants*

1996

4. M. ZINSMEISTER – *Formalisme thermodynamique et systèmes dynamiques holomorphes*
3. J. BERTIN, J.-P. DEMAILLY, L. ILLUSIE, C. PETERS – *Introduction à la théorie de Hodge*
2. C. VOISIN – *Symétrie miroir*

1995

1. S. TABACHNIKOV – *Billiards*

1994

0. J.-P. RAMIS – *Séries divergentes et théories asymptotiques*

