Ralph Meyer, Thomas Schick, and Yuri Tschinkel (Ed.)

Trends in Mathematics

erschienen in der Reihe „Mathematisches Institut. Seminare“ der
Universitätsdrucke des Universitätsverlags Göttingen 2008

Ralph Meyer,
Thomas Schick,
and Yuri Tschinkel (Ed.)

Trends in Mathematics

Mathematisches Institut
Georg-August-Universität
Göttingen

Courant-Colloquium, October 2007

Universitätsverlag Göttingen
2008

Bibliographische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliographie; detaillierte bibliographische Daten sind im Internet über <http://dnb.ddb.de> abrufbar.

Address of the Editors
Ralf Meyer, rameyer@uni-math.gwdg.de@uni-math.gwdg.de

Thomas Schick, schick@uni-math.gwdg.de

Yuri Tschinkel, yuri@uni-math.gwdg.de

Mathematisches Institut, Bunsenstr. 3-5, 37073 Göttingen.
URL: http://www.uni-math.gwdg.de/tschinkel

Typesetting and Layout: Stefan Wiedmann
Cover Image: Hyperboloid of revolution of one sheet;
collection of the Mathematics Institute Göttingen (photographer Margo Bargheer)

http://univerlag.uni-goettingen.de
ISBN: 978-3-940344-56-4

CONTENTS

ABSTRACTS

Euclidean propagators in noncommutative quantum field theory
DOROTHEA BAHNS 1

It is shown that the 2-point and the 4-point function of bosonic fields on the noncommutative Minkowski space are distributions which are boundary values of analytic functions. Contrary to what one might expect, this connection to analytic functions does not provide a connection to the popular Euclidean Feynman rules of noncommutative field theory, and thereby explains why renormalization in the framework of those latter rules crudely differs from renormalization in the Minkowskian regime.

Geometric construction of cluster algebras and cluster categories
KARIN BAUR 13

In this note we explain how to obtain cluster algebras from triangulations of (punctured) discs following the approach of **[13]**. Furthermore, we give a description of m-cluster categories via diagonals (arcs) in (punctured) polygons and of m-cluster categories via powers of translation quivers as given in joint work with R. Marsh (**[1]**, **[3]**).

Dispersive Equations and Hyperbolic Orbits
HANS CHRISTIANSON 31

We consider several related problems in linear PDE on manifolds, both when the underlying manifold is compact and when it is non-compact. Our general assumption is that the classical flow has a single periodic hyperbolic (unstable) orbit. The standard techniques from linear PDE relying on dispersive effects and non-trapping assumptions do not directly apply in this situation, but the unstable nature of the periodic orbit allows us to prove slightly weaker results.

Forms in many variables
RAINER DIETMANN 47

We give a survey about two new applications of the Hardy-Littlewood circle method from analytic number theory to arithmetic problems. These problems are finding a polynomial bound for the smallest integer solution of a quadratic Diophantine equation, and showing that the intersection of a system of rational cubic forms in sufficiently many variables admits a rational linear space of zeros of given dimension.

This note is an informal report on the joint paper **[4]** of the author with Jeffrey Giansiracusa, which grew out of the attempt to understand the topology of the moduli stack of stable curves. The main result is the construction of a map from the moduli stack to a certain infinite loop space, which is surjective on homology in a certain range. This shows the existence of many torsion classes in the homology of $\overline{\mathfrak{M}}_{g,n}$. We give a geometric description of some of the new torsion classes. Also, we give a new proof of an (old) theorem computing the second homology of the moduli stack.

In this survey, we describe three tropical enumerative problems and the corresponding moduli spaces of tropical curves. They have the structure of weighted polyhedral complexes. We observe similarities in the definitions of the weights, aiming at a better understanding of the tropical structure of the moduli spaces.

We give a leisurely account of the relationship between volume and L^2-Betti numbers on closed, aspherical manifolds based on the results in **[4]** – albeit with a different point of view. This paper grew out of a talk presented at the first colloquium of the Courant Center in Göttingen in October 2007.

We give an overview over current results on the global structure of affine Deligne-Lusztig varieties associated to a hyperspecial maximal compact subgroup. In particular, we discuss a formula for their dimensions and the set of connected components of the closed affine Deligne-Lusztig varieties.

Unlike Lie algebras which one-to-one correspond to simply connected Lie groups, Lie algebroids (integrable or not) one-to-one correspond to a sort of étale stacky groupoids (W-groupoids). Following Sullivan's spacial realization of a differential algebra, we construct a canonical integrating Lie 2-groupoid for every Lie algebroid. Finally we discuss how to lift Lie algebroid morphisms to W-groupoid morphisms (Lie II). Examples of Poisson manifolds and symplectic stacky groupoids are provided. This paper contains essentially some ideas of proofs and examples, for a complete treatment please refer to **[29]** which also proves some connectedness result.

INTRODUCTION

This volume contains the Proceedings of the Courant-Colloquium *Göttingen Trends in Mathematics* which took place at the Mathematics Institute of the University of Göttingen in October 12-14, 2007.

Yuri Tschinkel

15.07.2008

Mathematisches Institut, Courant-Colloquium Trends in Mathematics
(R. MEYER, T. SCHICK, Y. TSCHINKEL, eds.), p. 1–12
Georg-August Universität Göttingen, 2008

EUCLIDEAN PROPAGATORS IN NONCOMMUTATIVE QUANTUM FIELD THEORY

Dorothea Bahns

Mathematisches Institut, Universität Göttingen, Bunsenstr. 3-5, 37073 Göttingen
E-mail : bahns@uni-math.gwdg.de

Abstract. It is shown that the 2-point and the 4-point function of bosonic fields on the noncommutative Minkowski space are distributions which are boundary values of analytic functions. Contrary to what one might expect, this connection to analytic functions does not provide a connection to the popular Euclidean Feynman rules of noncommutative field theory, and thereby explains why renormalization in the framework of those latter rules crudely differs from renormalization in the Minkowskian regime.

1. Introduction

A quantum field theoretic model is to a large part determined by the choice of a partial differential operator. For physical reasons, this operator has to be *hyperbolic*, and one of its fundamental solutions, the so-called Feynman propagator, is the building block in any calculation of physically relevant quantities. Nonetheless, ever since proposed by Symanzik in 1966 [**9**] based on ideas of Schwinger, the so-called Euclidean framework has played a very important role. In this framework, the building block is the so-called Schwinger function, a fundamental solution of an *elliptic* partial differential operator. The Euclidean formalism not only simplifies calculations, but seems to be indispensable in constructive quantum field theory. The remarkable theorem of Osterwalder and Schrader gives sufficient conditions for the possibility to recover the original hyperbolic (physically meaningful) field

2000 ***Mathematics Subject Classification.*** 81T75,46F20.

theory from a Euclidean framework, and therefore justifies the Euclidean framework in ordinary quantum field theory. I will recall below how the Schwinger function of the Euclidean framework of scalar field theory is derived by analytic continuation from the hyperbolic theory and how it relates to the Feynman propagator.

As one possibility to incorporate gravitational aspects into quantum field theory, for some years now, much research has been done on quantum fields on noncommutative spaces, the most popular of which is the noncommutative "Moyal space" whose coordinates are subject to commutation relations of the Heisenberg type [**4**]. Already in that early paper, a possible setting for hyperbolic perturbative quantum field theory was proposed, where the field algebra is endowed with a noncommutative product, the twisted (convolution) product. Notwithstanding, the vast majority of publications on field theory on noncommutative spaces ("noncommutative field theory") has been and still is formulated within a Euclidean setting. This setting was not derived from a hyperbolic noncommutative theory but from the Euclidean framework of ordinary field theory by replacing all products with twisted ones. I shall refer to this approach as the traditional noncommutative Euclidean framework. Despite some attempts, it has not been possible to relate this traditional noncommutative Euclidean setting to some hyperbolic noncommutative theory – in fact, there is evidence that it might be impossible to do so, unless the time variable commutes with all space variables.

After some years, it became clear that even the simplest quantum field theoretic models, the massive scalar models, have very peculiar properties when formulated within the traditional Euclidean noncommutative framework.(1) Most notably, the so-called ultraviolet–infrared mixing problem noted in [**7**] severly limits the type of models that can be defined at all [**5, 6**].

In contrast to these results, I have shown [**1**] that in a hyperbolic setting, the ultraviolet-infrared mixing effect is not present at least in the most prominent example graph that exibits ultraviolet–infrared mixing in the traditional Euclidean realm. I shall present this result in a longer and more technical article shortly. However, the calculations and the combinatorial aspects being quite complicated, I have not yet been able to find a general proof of the conjecture that the ultraviolet-infrared mixing problem may be absent in this hyperbolic noncommutative setting.

---

(1) In fact, as is common in the literature, I will only consider this simplest example of field theories here. Already such massive scalar models have enough structure to enable us to study the principles of noncommutative field theory, while massless theories and gauge theories are notoriously difficult.

For this reason, I thought it desirable to find a Euclidean framework that was actually derived from a hyperbolic noncommutative setting. This being achieved, it is to be hoped that a Schwinger functional can be found which should greatly simplify the combinatorial aspects of perturbation theory, and that the full Euclidean machinery might indeed make it possible to investigate renormalizability and the possible absence of the ultraviolet-infrared mixing problem in general.

In this note, I will show that one can indeed derive a noncommutative Euclidean framework from a hyperbolic theory on the Moyal space, and that this framework is *not* the traditional one that is investigated in the literature. In contrast to this traditional framework, the new Euclidean framework can moreover be related to a setting involving Feynman propagators via an analytic continuation similar to the one of ordinary quantum field theory. To start, I will recall in the next section how the Schwinger function is derived in ordinary massive scalar quantum field theory and how it is related to the Feynman propagator. In the third section, I will then derive a Euclidean 4-point function from a noncommutative hyperbolic Wightman function and comment on how to proceed for arbitrarily high order. I will show that the Euclidean framework thus derived differs from the traditional noncommutative approach. Moreover, the relation to Feynman propagators is clarified. In an outlook I will briefly comment on further possible research that ensues from these new results.

2. Euclidean methods in quantum field theory

The hyperbolic partial differential operator of massive scalar field theory is the massive Klein–Gordon operator $P := \frac{\partial^2}{\partial {x_0}^2} - \Delta_{\mathbf{x}} + m^2$ on $\mathbb{R}^4$ where $\Delta_{\mathbf{x}}$ denotes the Laplace operator on $\mathbb{R}^3$, $\mathbf{x} \in \mathbb{R}^3$, and $m > 0$ is a real parameter, called the field's mass. As mentioned in the introduction, all the relevant quantities of a scalar field theoretic model can be calculated from a fundamental solution of this operator. Recall here that a distribution $E \in \mathscr{D}'(\mathbb{R}^n)$ is a fundamental solution (or Green's function) of a partial linear differential operator $P(\partial)$ on $\mathbb{R}^n$ provided that in the sense of distributions, $P(\partial)E = -\delta$ with δ denoting the δ-distribution.

Our starting point here, however, is the 2-point-function $\Delta_+ \in \mathscr{S}'(\mathbb{R}^4)$, a tempered distribution which is a solution (not a fundamental solution) of the Klein–Gordon equation, $P\Delta_+ = 0$ in the sense of distributions. For $x = (x_0, \mathbf{x}) \in \mathbb{R}^4$, $x_0 \in \mathbb{R}$,

$\mathbf{x} \in \mathbb{R}^3$, it is given explicitly by

$$\Delta_+(x) = \frac{1}{(2\pi)^3} \int \frac{\mathrm{d}^3\mathbf{k}}{2\omega_\mathbf{k}} \mathrm{e}^{-\mathrm{i}\omega_\mathbf{k} x_0 + \mathrm{i}\mathbf{kx}}, \qquad \text{where } \omega_\mathbf{k} = \sqrt{\mathbf{k}^2 + m^2},$$

an expression which in fact makes sense as an oscillatory integral, see [**8**, Sec IX.10] for details. Here and in what follows, boldface letters denote elements of $\mathbb{R}^3$ and an expression such as $\mathbf{kx}$ is shorthand for the canonical scalar product of $\mathbf{k}$ and $\mathbf{x}$.

It is well-known that Δ_+ is the boundary value (in the sense of distributions) of an analytic function. To see this, let us first fix some notation. Let $a \in \mathbb{R}^n$ with $|a| = 1$, let $\theta \in (0, \pi/2)$, and let ay denote the canonical scalar product in $\mathbb{R}^n$. The set $\Gamma_{a,\theta} = \{y \in \mathbb{R}^n \mid ya > |y| \cos\theta\} \subset \mathbb{R}^n$ is called the cone about a with opening angle θ. Let with $\Gamma^*_{a,\theta}$ denote the dual cone, $\Gamma^*_{a,\theta} = \Gamma_{a,\frac{\pi}{2}-\theta}$. For temperered distributions whose support is contained in the closure of a cone, the following general assertion holds:

Theorem 1 ([8], Thm IX.16). *Let u be a tempered distribution with support in the closure of a cone $\Gamma_{a,\theta}$, $a \in \mathbb{R}^n$, $0 < \theta < \pi/2$. Then its Fourier transform $\tilde{u}$ is the boundary value (in the sense of $\mathscr{S}'$) of a function f which is analytic in the tube $\mathbb{R}^n - \mathrm{i}\Gamma^*_{a,\theta} = \{z \in \mathbb{C}^n | -\mathrm{im}\, z \in \Gamma^*_{a,\theta}\} \subset \mathbb{C}^n$.*

Observe that for $\tilde{u}$ to be the boundary value of f in the sense of $\mathscr{S}'$ means that for any $\eta \in \Gamma^*_{a,\theta}$ we have for $t \searrow 0$ in $\mathbb{R}$ from above,

$$\int f(x - \mathrm{i}t\eta)\, g(x)\, \mathrm{d}x \;\rightarrow\; \tilde{u}(g) \qquad \forall\, g \in \mathscr{S}(\mathbb{R}^4).$$

The Fourier transform $\tilde{\Delta}_+$ of the 2-point function,

$$\tilde{\Delta}_+(p_0, \mathbf{p}) = \frac{1}{2\omega_\mathbf{p}} \delta(p_0 - \omega_\mathbf{p}), \tag{2.1}$$

is a tempered distribution whose support (the positive mass shell) is contained in the closure of the cone $V = \Gamma_{(1,\mathbf{0}),\pi/4}$ (the forward light cone). Applied to $u := \tilde{\Delta}_+$, Theorem 1 thus guarantees that $\tilde{u} = \Delta_+$ is the boundary value of a function f which is analytic in $\mathbb{R}^4 - \mathrm{i}V$ (observe that $V^* = V$). Explicitly, for $x = (x_0, \mathbf{x}) \in \mathbb{R}^4$ and $\eta = (x_4, \mathbf{0}) \in V$ (hence $x_4 > 0$), we have in this case

$$f(x - \mathrm{i}\eta) = \frac{1}{(2\pi)^3} \int \frac{\mathrm{d}^3\mathbf{k}}{2\omega_\mathbf{k}} \mathrm{e}^{+\mathrm{i}\mathbf{kx} - \omega_\mathbf{k}(x_4 + \mathrm{i}x_0)}. \tag{2.2}$$

We now define a function s via $s(\mathbf{x}, x_4 + \mathrm{i}x_0) := f(x - \mathrm{i}\eta)$ for x and η as above. Making use of the identity

$$\frac{1}{2\omega_{\mathbf{k}}} \mathrm{e}^{-\omega_{\mathbf{k}} x_4} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathrm{d}k_4 \, \frac{\mathrm{e}^{\mathrm{i}k_4 x_4}}{k^2 + m^2} \qquad \text{for } x_4 > 0 \tag{2.3}$$

where $k = (\mathbf{k}, k_4) \in \mathbb{R}^4$, $k^2 = \mathbf{k}^2 + k_4^2$, and setting $x_0 = 0$ in (2.2), we then find that

$$s(x) = \frac{1}{(2\pi)^4} \int \mathrm{d}^4 k \, \frac{\mathrm{e}^{\mathrm{i}kx}}{k^2 + m^2} \tag{2.4}$$

where $x = (\mathbf{x}, x_4) \in \mathbb{R}^4$, $x_4 > 0$. One now considers a distribution $S \in \mathscr{S}'(\mathbb{R}^4)$, the so-called *Schwinger function*, whose formal integral kernel is given by the Fourier transform of the smooth function on $\mathbb{R}^4$,

$$\tilde{S}(k) = \frac{1}{k^2 + m^2}.$$

By definition, when restricted to the upper half space $x_4 > 0$, $S(\mathbf{x}, x_4)$ is (pointwise) equal to the function s given in (2.4). Observe also that S is the unique fundamental solution of the *elliptic* partial differential operator $\Delta - m^2$ with Δ the Laplace operator on $\mathbb{R}^4$.

As mentioned in the introduction, the building block in hyperbolic perturbation theory is the Feynman propagator Δ_F, a fundamental solution for the Klein–Gordon operator $P = \frac{\partial^2}{\partial {x_0}^2} - \Delta_{\mathbf{x}} + m^2$. Without going into details, let me mention that, remarkably, the Fourier transform $\tilde{S}$ of the Schwinger function is the analytic continuation of the Fourier transform $\tilde{\Delta}_F$ of the Feynman propagator (up to a sign), where, formally, for w given by $w(\mathbf{p}, p_4 - \mathrm{i}p_0) := \tilde{\Delta}_F(p_0 + \mathrm{i}p_4, \mathbf{p})$ we have $\tilde{S}(\mathbf{p}, p_4) = -w(\mathbf{p}, p_4)$.

3. Analytic continuation in the noncommutative case

It would be beyond the scope of this note to explain the possible perturbative setups for massive scalar fields on the noncommutative Moyal space with hyperbolic signature (see [**2**] for a comparison). Only two features of such noncommutative (hyperbolic) field theories matter here. The first is the fact that our starting point still is the Klein–Gordon operator and the 2-point-function discussed in the previous section. The second important feature – and this feature is shared by the traditional noncommutative Euclidean formalism – is the fact that one has to consider not only products but also twisted products of distributions.

To fix notation, we note here that for two Schwartz functions $f, g \in \mathscr{S}(\mathbb{R}^4)$ this twisted product (Moyal product) is

$$f * g\,(x) = \int d^4k \int d^4p\; \tilde{f}(k)\, \tilde{g}(p)\, e^{-i(p+k)x}\, e^{-\frac{i}{2}p\theta k} \tag{3.1}$$

for $x \in \mathbb{R}^4$, where $\tilde{f}$ and $\tilde{g}$ denote the Fourier transforms of f and g, respectively, and where θ is a *nondegenerate* antisymmetric 4×4-matrix. Observe that in a Euclidean theory, a product such as kx stands for the canonical scalar product, whereas in a hyperbolic setting, it denotes a Lorentz product, $kx = k_0x_0 - \mathbf{kx}$. The oscillating factor $e^{-\frac{i}{2}p\theta k}$ is also called the twisting.

3.1. Tensor product of 2-point functions. Since the 2-point function remains unchanged in noncommutative field theory, we have to consider higher order correlation functions in order to see a difference between field theory on Moyal space and ordinary field theory. Again, it would be beyond the scope of this note to explain the whole setup. It will be sufficient to consider the particular example of the so-called 4-point function of massive scalar field theory. In ordinary field theory, this is a distribution given by 2-fold tensor products of 2-point functions,

$$\Delta_+^{(2)}(x,y) = \frac{1}{(2\pi)^6}\int \frac{d^3\mathbf{k}}{2\omega_{\mathbf{k}}}\frac{d^3\mathbf{p}}{2\omega_{\mathbf{p}}}\, e^{-i(\omega_{\mathbf{k}}x_0+\omega_{\mathbf{p}}y_0)+i(\mathbf{kx}+\mathbf{py})}$$

By standard arguments from microlocal analysis involving the wavefront set of distributions, it can be shown that even the pullback of this tensor product with respect to the diagonal map, that is, the product in the sense of Hörmander, is a well-defined distribution $\in \mathscr{S}'(\mathbb{R}^4)$ (see for instance [**8**, Chap IX.10]). In order to avoid issues regarding renormalization later, in this note, however, only tensor products of distributions will be considered.

It is not difficult to see that $\Delta_+^{(2)}$ is again the boundary value of an analytic function:

Lemma 2. *The tempered distribution $\Delta_+^{(2)}$ is the boundary value of a function f_2 which is analytic in $\mathbb{R}^4\times\mathbb{R}^4 - iV\times V$. Explicitly, for*

$$z = (x_0, \mathbf{x}, y_0, \mathbf{y}) \in \mathbb{R}^4\times\mathbb{R}^4 \qquad \textit{and} \qquad \eta = (x_4, \mathbf{0}, y_4, \mathbf{0}) \in V\times V$$

(hence x_4 and $y_4 > 0$), we have

$$f_2(z - i\eta) = \frac{1}{(2\pi)^6}\int\frac{d^3\mathbf{k}}{2\omega_{\mathbf{k}}}\int\frac{d^3\mathbf{p}}{2\omega_{\mathbf{p}}}\, e^{-\omega_{\mathbf{k}}(x_4+ix_0)-\omega_{\mathbf{p}}(y_4+iy_0)+i\mathbf{kx}+i\mathbf{py}}$$

and for the function s_2 defined by $s_2(\mathbf{x}, x_4 + \mathrm{i}x_0, \mathbf{y}, y_4 + \mathrm{i}y_0) = f_2(z - \mathrm{i}\eta)$ with η and z as above, we find for $(x, y) = (\mathbf{x}, x_4, \mathbf{y}, y_4) \in \mathbb{R}^4 \times \mathbb{R}^4$, x_4 and $y_4 > 0$,

$$s_2(x,y) = \frac{1}{(2\pi)^8} \int \mathrm{d}^4 k \int \mathrm{d}^4 p \, \frac{1}{k^2+m^2} \, \frac{1}{p^2+m^2} \, \mathrm{e}^{+\mathrm{i}kx+\mathrm{i}py}, \tag{3.2}$$

where $p^2 = \mathbf{p}^2 + p_4^2$, likewise for k^2.

Proof. The first claim is a direct consequence of Theorem 1 applied with respect to x and y separately, and the second claim follows again from the identity (2.3). □

As in the previous section, one now defines a distribution S_2 as the Fourier transform (in $\mathbb{R}^4 \times \mathbb{R}^4$) of the smooth function

$$\tilde{S}(k)\,\tilde{S}(p) = \frac{1}{k^2+m^2} \frac{1}{p^2+m^2}$$

which, when restricted to $\mathbb{R}^3 \times \mathbb{R}_{>0} \times \mathbb{R}^3 \times \mathbb{R}_{>0}$, is equal to s_2. The reader who is familiar with quantum field theory will of course recognize that when one considers the pullback of $\Delta_+^{(2)}$ with respect to the diagonal map, formally, one finds $x = y$ in (3.2) and that in this case S_2 becomes the Fourier transform (in $\mathbb{R}^4$) of the convolution $\tilde{S} \times \tilde{S}(k) = \int \mathrm{d}^4 p \, \frac{1}{(k-p)^2+m^2} \, \frac{1}{p^2+m^2}$.

It is well-known that the same procedure can be applied more generally. Each contribution to the (hyperbolic) $2n$-point function (or Wightman function) is an n-fold tensor product of 2-point functions (n-point functions for odd n vanish). In order to find the corresponding higher order Schwinger function, one considers the analytic continuation according to Theorem 1 in each of the n variables and proceeds in the same manner as explained for the 4-point function above.

3.2. Twisted product of 2-point functions. In [**3**], it was shown how $2n$-point functions are calculated in hyperbolic massive scalar field theory on the noncommutative Moyal space (n-point functions for n odd still vanish). As it turns out, the first deviation from ordinary field theory shows up in the 4-point function, where one of the contributions is a twisted tensor product of two 2-point functions,

$$\Delta_+^{(\star 2)}(x,y) = \int \frac{\mathrm{d}^3\mathbf{k}}{2\omega_{\mathbf{k}}} \frac{\mathrm{d}^3\mathbf{p}}{2\omega_{\mathbf{p}}} \, \mathrm{e}^{-\mathrm{i}(\omega_{\mathbf{k}} x_0 + \omega_{\mathbf{p}} y_0) + \mathrm{i}(\mathbf{k}\mathbf{x} + \mathbf{p}\mathbf{x})} \, \mathrm{e}^{-\mathrm{i}\tilde{p}\theta\tilde{k}} \tag{3.3}$$

where $\tilde{k} = (\omega_{\mathbf{k}}, \mathbf{k})$, and $\tilde{p} = (\omega_{\mathbf{p}}, \mathbf{p})$, such that their Lorentz square is $k \cdot k = (m^2 + \mathbf{k}^2) - \mathbf{k}^2 = m^2$ and their first component is positive. In the terminology of physics, this means that the "momenta" k and p in the oscillating factor are *on-shell.* This will turn out to be very important later on. It is important to note that, while our starting point is the twisted product (3.1), the vectors in the twisting are on-shell as a consequence of the support properties of $\tilde{\Delta}_+(k_0,\mathbf{k}) = \frac{1}{\omega_\mathbf{k}}\,\delta(k_0-\omega_\mathbf{k})$. Note also that the factor 2 in the oscillating factor compared to the ordinary twisting in (3.1) is correct.

Once more, we now apply Theorem 1.

Lemma 3. *The tempered distribution $\Delta_+^{(\star 2)}$ is the boundary value of a function f_2^θ which is analytic in $\mathbb{R}^4\times\mathbb{R}^4 - \mathrm{i}V\times V$. Explicitly, for $z=(x_0,\mathbf{x},y_0,\mathbf{y})\in\mathbb{R}^4\times\mathbb{R}^4$ and $\eta=(x_4,\mathbf{0},y_4,\mathbf{0})\in V\times V$ (hence x_4 and $y_4>0$), we have*

$$f_2^\theta(z-\mathrm{i}\eta) = \frac{1}{(2\pi)^6}\int\frac{\mathrm{d}^3\mathbf{k}}{2\omega_\mathbf{k}}\int\frac{\mathrm{d}^3\mathbf{p}}{2\omega_\mathbf{p}}\,\mathrm{e}^{-\omega_\mathbf{k}(x_4+\mathrm{i}x_0)-\omega_\mathbf{p}(y_4+\mathrm{i}y_0)+\mathrm{i}\mathbf{kx}+\mathrm{i}\mathbf{py}}\,\mathrm{e}^{-\mathrm{i}\tilde{p}\theta\tilde{k}}$$

where $\tilde{k}=(\omega_\mathbf{k},\mathbf{k})$, $\tilde{p}=(\omega_\mathbf{p},\mathbf{p})$. For the function s_2^θ defined by

$$s_2^\theta(\mathbf{x},x_4+\mathrm{i}x_0,\mathbf{y},y_4+\mathrm{i}y_0) = f_2^\theta(z-\mathrm{i}\eta)$$

with η and z as above, we then find for $(x,y)=(\mathbf{x},x_4,\mathbf{y},y_4)\in\mathbb{R}^4\times\mathbb{R}^4$ with $x_4,y_4>0$,

$$s_2^\theta(x,y) = \frac{1}{(2\pi)^8}\int\mathrm{d}^4k\int\mathrm{d}^4p\,\frac{1}{k^2+m^2}\,\frac{1}{p^2+m^2}\,\mathrm{e}^{+\mathrm{i}kx}\,\mathrm{e}^{+\mathrm{i}py}\,\mathrm{e}^{-\mathrm{i}\tilde{p}\theta\tilde{k}} \tag{3.4}$$

where $p^2=\mathbf{p}^2+p_4^2$, likewise for k^2, and with $\tilde{p}$ and $\tilde{k}$ as above.

Proof. Since the Fourier transform of $\Delta_+^{(\star 2)}$ is still a tempered distribution with support contained in the closure of $V\times V$, the first claim follows from Theorem 1. The second claim again follows from the identity (2.3) – which, as should be noted, does not affect the twisting factor. □

Observe that s_2^θ and s_2 from Lemma 2 differ only by the oscillating factor $\mathrm{e}^{-\mathrm{i}\tilde{p}\theta\tilde{k}}$. As before, we now define a distribution S_2^θ as the Fourier transform (in $\mathbb{R}^4\times\mathbb{R}^4$) of the smooth function

$$\tilde{S}_2^\theta(k,p) = \frac{1}{k^2+m^2}\,\frac{1}{p^2+m^2}\,\mathrm{e}^{-\mathrm{i}\tilde{p}\theta\tilde{k}}. \tag{3.5}$$

Again, the reader who is familiar with the field will recognize that in the case of coinciding points, instead of $\tilde{S}_2^\theta(k,p)$ one considers the Fourier transform of the twisted convolution

$$\int\mathrm{d}^4p\,\frac{1}{(k-p)^2+m^2}\,\frac{1}{p^2+m^2}\,\mathrm{e}^{-\mathrm{i}\tilde{p}\,\theta\,\widetilde{k-p}}$$

where $\widetilde{k-p}=(\omega_{\mathbf{k}-\mathbf{p}},k-p)$.

It is very important to note that all the momenta which appear in the oscillating factors in all the above expressions are on-shell, being of the form $\tilde{p} = (\omega_{\mathbf{p}}, \mathbf{p})$, likewise for k or $p-k$. The oscillating factor therefore distinguishes the components of $(\mathbf{p}, p_4)$ and is, in particular, independent of the fourth component p_4. The reason for this lies in the fact that the Fourier transform of the 2-point function forces the momenta in the oscillating factor to be on-shell, and this is not changed by the analytic continuation.

This turns out to be crucial in the following assertion:

Remark 4. Since the oscillating factor in (3.5) is independent of one of the components of k and $p \in \mathbb{R}^4$, it is obvious that $\tilde{S}_2^{\theta}$ is the analytic continuation a product of Feynman propagators with an on-shell twisting: For

$$w_2^{\theta}(\mathbf{k}, k_4 - \mathrm{i}k_0, \mathbf{p}, p_4 - \mathrm{i}p_0) := \tilde{\Delta}_F(k_0 + \mathrm{i}k_4, \mathbf{k})\, \tilde{\Delta}_F(p_0 + \mathrm{i}p_4, \mathbf{p})\, \mathrm{e}^{-\mathrm{i}\tilde{p}\theta\tilde{k}}$$

we find $\tilde{S}_2^{\theta}(\mathbf{k}, k_4, \mathbf{p}, p_4) = -w^{\theta}(\mathbf{k}, k_4, \mathbf{p}, p_4)$.

All this remains true when one calculates the higher order Schwinger functions from the $2n$-point functions that were calculated in [**3**]. Again, performing the analytic continuation in the n variables separately, one finds on-shell twistings (though they become more and more complicated), and the analytic continuation of the corresponding Fourier transform of Feynman propagators can be performed as in Remark 4 while leaving the twistings unchanged.

This is the essential difference to the traditional noncommutative Euclidean framework employed in the literature. In this latter framework, starting point are Schwinger functions, and of course, when twisted products appear, by (3.1) the oscillating factors depend on all four components of a momentum vector $k = (\mathbf{k}, k_4)$. For instance, instead of $\tilde{s}_2^{\theta}$ as in (3.5), one finds the following expression

$$\tilde{e}_2^{\theta}(k, p) = \frac{1}{k^2 + m^2}\,\frac{1}{p^2 + m^2}\,\mathrm{e}^{-\mathrm{i}p\theta k} \tag{3.6}$$

where $k = (\mathbf{k}, k_4)$ and $p = (\mathbf{p}, p_4)$ and θ is a nondegenerate 4×4-matrix. So far, it was not possible to relate this framework to a hyperbolic one, the main difficulty being the dependence of the oscillating factor on k_4. Naively copying the procedure sketched on page 5 and in Remark 4 leads to exponentially increasing terms which render the integrals ill-defined. So far, the only way out found seems to be to make the oscillating factor independent of one of the components in an *ad hoc* way, by requiring θ to be a matrix of rank 2 (“spacelike noncommutativity”).

Remark 4 shows that such measures are unnecessary when the new noncommutative Euclidean framework derived from the hyperbolic n-point functions is employed.

4. Outlook

It will be shown elsewhere that at least in the most prominent example graphs, the ultraviolet-infrared mixing problem is absent in this new Euclidean framework. However, as can be easily seen already in the example S_2^θ discussed above, the higher order Schwinger functions are not symmetric with respect to reflections in the origin. This may jeopardize the possibility to set up a complete consistent perturbative framework using a Schwinger functional and further research must be done in that direction.

Other than that, the results presented here open many interesting possibilities for future research. For one thing, one should try to generalize the Osterwader Schrader Theorem in this setting. Also, it would be most interesting to study whether the ultraviolet-infrared mixing problem appears in this setting at all. And last but not least, a thorough understanding of the new Euclidean setup should enable us to learn more about hyperbolic noncommutative models – which in themselves have proved to be quite difficult to treat. It is certainly to be hoped that from a Euclidean perturbative setup to be developped from the ideas presented here, general proofs of renormalizability of hyperbolic noncommutative field theory will at last be possible.

References

[1] D. Bahns – "Talks: European Science Foundation Workshop Noncommutative Quantum Field Theory", Erwin Schrödinger Institute, Nov 26-29, 2007, and 4th Vienna Central European Seminar on Particle Physics and Quantum Field Theory, Nov 30 - Dec 02, 2007.

[2] ______, "Perturbative methods on the noncommutative minkowski space", Ph.D. Thesis, 2004.

[3] ______, "Local counterterms on the noncommutative Minkowski space", Rigorous quantum field theory, Progr. Math., vol. 251, Birkhäuser, Basel, 2007, p. 11–26.

[4] S. Doplicher, K. Fredenhagen & J. E. Roberts – "The quantum structure of spacetime at the Planck scale and quantum fields", *Comm. Math. Phys.* **172** (1995), no. 1, p. 187–220.

[5] H. Grosse & R. Wulkenhaar – "Renormalisation of φ^4-theory on noncommutative $\mathbb{R}^4$ in the matrix base", *Comm. Math. Phys.* **256** (2005), no. 2, p. 305–374.

[6] R. GURAU, V. RIVASSEAU & A. TANASA – "A translation-invariant renormalizable non-commutative scalar model", arXiv:0802.0791 [math-ph].

[7] S. MINWALLA, M. VAN RAAMSDONK & N. SEIBERG – "Noncommutative perturbative dynamics", *J. High Energy Phys.* (2000), no. 2, p. Paper 20, 31.

[8] M. REED & B. SIMON – *Methods of modern mathematical physics. II. Fourier analysis, self-adjointness*, Academic Press [Harcourt Brace Jovanovich Publishers], New York, 1975.

[9] K. SYMANZIK – "Euclidean quantum field theory. I. Equations for a scalar model", *J. Mathematical Phys.* **7** (1966), p. 510–525.

Mathematisches Institut, Courant-Colloquium Trends in Mathematics
(R. Meyer, T. Schick, Y. Tschinkel, eds.), p. 13–30
Georg-August Universität Göttingen, 2008

GEOMETRIC CONSTRUCTION OF CLUSTER ALGEBRAS AND CLUSTER CATEGORIES

Karin Baur

Department of Mathematics, ETH Zürich Rämistrasse 101, 8092 Zürich, Switzerland
E-mail : baur@math.ethz.ch

***Abstract*.** In this note we explain how to obtain cluster algebras from triangulations of (punctured) discs following the approach of **[13]**. Furthermore, we give a description of m-cluster categories via diagonals (arcs) in (punctured) polygons and of m-cluster categories via powers of translation quivers as given in joint work with R. Marsh (**[1]**, **[3]**).

1. Introduction

This article is an expanded version of a talk presented at the Courant-Colloquium "Göttingen trends in Mathematics" in October 2007. It is a survey on two approaches to cluster algebras and (m-)cluster categories via geometric constructions.

Cluster algebras where introduced in 2001 by Fomin and Zelevinsky, cf. [**14**]. They arose from the study of two related problems.

Problem 1 (Canonical basis). Understand the *canonical basis* (Lusztig), or *crystal basis* (Kashiwara) of quantized enveloping algebras associated to a semisimple complex Lie algebra. It is expected that the positive part of the quantized enveloping algebra has a (quantum) cluster algebra structure, with the so-called cluster monomials forming part of the dual canonical basis.

This picture motivated the definition of *cluster variables.*

2000 ***Mathematics Subject Classification*.** 16G20, 16G70, 18E30.

Problem 2 (Total positivity). An invertible matrix with real entries is called *totally positive* if all its minors are positive. This notion has been extended to all reductive groups by Lusztig [**28**]. To check total positivity of an upper uni-triangular matrix, only a certain collection of the non-zero minors needs to be checked (disregarding the minors which are zero because of the uni-triangular from). The minimal sets of such all have the same cardinality. When one of them is removed, it can often be replaced by a unique alternative minor. The two minors are connected through a certain relation.

This exchange (*mutation for minors*) motivated the definition of cluster mutation.

The subject of cluster algebra is a very young and dynamic one. In the past few years, connections to various other fields arose. We briefly mention a few of them here.

– Poisson geometry (integrable systems), Teichmüller spaces (local coordinate systems), cf. Gekhtman-Shapiro-Vainshtein [**22, 23**] and Fock-Goncharov [**12**];

– Y-systems in thermodynamic Bethe Ansatz (families of rational functions defined by recurrences which were introduced by Zamolodchikov [**35**]), cf. [**14**];

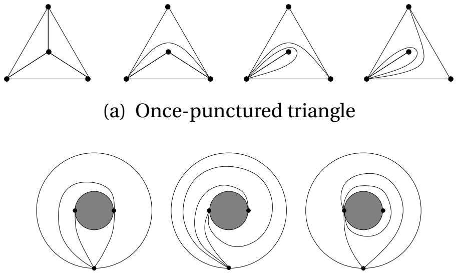
– Stasheff polytopes, associahedra, Chapoton-Fomin-Zelevinsky [**8**];

– ad-nilpotent ideals of Borel subalgebras in Lie algebras, Panyushev [**29**];

– Preprojective algebra models, Geiss-Leclerc-Schröer, [**20**], [**21**];

– Representation theory, tilting theory, cf. [**6**].

In this article, we will first recall triangulations of surfaces with marked points and associate certain integral valued matrices to them. Then we will give a brief introduction to cluster algebras (Section 3). In Section 4 we show how to associate cluster algebras to triangulations of (punctured) discs. Then we explain what cluster categories and m-cluster categories are (Section 5) and give a combinatorial model to describe m-cluster categories via arcs in a polygon in Section 6, cf. Theorems 6.3, 6.4 as given in our joint work with R. Marsh ([**1**], [**3**]). In addition, we obtain a descriptions of the m-cluster categories using the notion of the power of a translation quiver (Theorem 6.5). At the end we describe connections to other work, pose several questions and show new directions in this young and dynamic field (Section 7).



(a) Once-punctured triangle

(b) Annulus

FIGURE 1. Examples of triangulations

2. Triangulated surfaces

Now we recall the triangulation of surfaces following the approach of Fomin, Shapiro and Thurston [**13**]. Let S be a connected oriented Riemann surface with boundary. Fix a finite set M of *marked points* on S. Marked points in the interior of S are called *punctures*.

We consider triangulations of S whose vertices are at the marked points in M and whose edges are pairwise non-intersecting curves, so-called *arcs* connecting marked points. The most important example for us is the case where S is a disc with marked points on the boundary and with at most one puncture. We will later restrict to that case but for the moment we explain the general picture.

It is convenient to exclude cases where there are no such triangulations (or only one such). We always assume that M is non-empty and that each boundary component has at least one marked point. And we disallow the cases (S, M) with one boundary component, $|M| = 1$ with ≤ 1 puncture and $|M| \in \{2, 3\}$ with no puncture.

In case S is a (punctured) disc we will also call it a *(punctured) polygon*. E.g. if (S, M) has three marked points on the boundary and a puncture, we will say that S is a once-punctured triangle.

Note that the pair (S, M) is defined (up to homeomorphism) by the genus of S, by the numbers of boundary components, of marked points on each boundary component and of punctures. Two examples of such triangulations are given in Figure 1.

Definition. A curve in S (up to isotopy relative M) is an *arc* γ *in* (S,M) if

(i) the endpoints of γ are marked points in M;

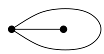
(ii) γ does not intersect itself (but its endpoints might coincide);

(iii) relative interior of γ is disjoint from M and from the boundary of S;

(iv) γ does not cut out an unpunctured monogon or digon.

The set of all arcs in (S,M) is usually infinite as we can already see in the case of the annulus of Figure 1(b). One can show that it is finite if and only if (S,M) is a disk with at most one puncture, i.e. if (S,M) is the object of our interest.

Two arcs are said to be *compatible* if they do not intersect in the interior of S. An *ideal triangulation* is a maximal collection T of pairwise compatible arcs. The arcs of T cut S into the so-called *ideal triangles.* These triangles may be self-folded, e.g. along the horizontal arc in the picture below:



An easy count shows that the once-punctured triangle has ten ideal triangulations, the four of Figure 1, with the rotations of the last three (by 120° and 240°).

In fact we can say more: the number of arcs in an ideal triangulation is an invariant of (S,M), we call it the *rank* of (S,M). There is a formula for it, cf. [**11**]: if g is the genus of S, b the number of boundary components, p the number of punctures, c the number of marked points on the boundary, then the rank of (S,M) is

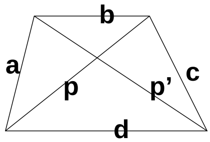
$$6g+3b+3p+c-6$$

The rank of the once punctured triangle of Figure 1(a) is thus three as expected.

For small rank, [**13**, Example 2.12] gives a list of all possible choices of (S,M). The word "type" appearing in the list refers to the Dynkin type of to the corresponding cluster algebra as will be explained later:

Rank 1	unpunctured square (type A_1)
Rank 2	unpunctured pentagon (type A_2)
	once-punctured digon (type $A_1 \times A_1$)
	annulus with one marked point on each boundary component
Rank 3	unpunctured hexagon (type A_3)
	once-punctured triangle (type $A_3 = D_3$)
	annulus with one marked point on one boundary component,
	two on the other once-punctured torus.

If T is an ideal triangulation of (S,M) and p an arc of T as in the picture below, we can replace p by an arc p' through a so-called flip or Whitehead move:



Here we allow that some of the sides $\{a,b,c,d\}$ coincide. A consequence of a result of Hatcher ([**25**]) is that for any two ideal triangulations T and T' there exists a sequence of flips leading from T to T'.

We next want to associate a matrix to an ideal triangulation of (S,M). This works as follows. Let T be an ideal triangulation of (S,M), label the arcs of T by $1,2,\ldots,n$. Then define $B(T)$ to be the following $n\times n$-square matrix

$$B(T)=\sum_{\Delta} B^{\Delta}$$

where the $n\times n$-matrices B^{Δ} are defined for each triangle $\triangle$ of T by

$$b_{ij}^{\Delta} = \begin{cases} 1 & \text{if } \Delta \text{ has sides } i \text{ and } j \text{ where } j \text{ is a clockwise neighbour of } i; \\ -1 & \text{if } \Delta \text{ has sides } i \text{ and } j \text{ where } i \text{ is a clockwise neighbour of } j; \\ 0 & \text{otherwise.} \end{cases}$$

The matrix $B(T)$ is skew-symmetric with entries $0,\pm1,\pm2$.

Remark. In order to simplify the definition of b_{ij}^{Δ} we have cheated a little bit. Whenever the triangle Δ is self-folded along an arc i, then in the right hand side of the definition of the entry b_{ij}^{Δ}, the arc i has to be replaced by its enclosing loop $l(i)$, cf. Figure 2.

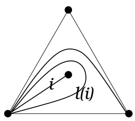
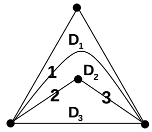


FIGURE 2. Enclosing loop $l(i)$ of the arc i

Example 2.1. (1) We compute $B(T)$ for the triangulated punctured triangle.

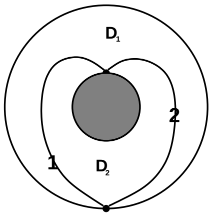


It is:

$$\underbrace{B^{D_1}}_{=0}+B^{D_2}+B^{D_3} = \begin{pmatrix} 0 & 1 & -1 \\ -1 & 0 & 1 \\ 1 & -1 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 & -1 \\ -1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

(2) Take an annulus with one marked point on each boundary and the triangulation T as in the picture. Then $B(T)$ is

$$B^{D_1}+B^{D_2} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 2 \\ -2 & 0 \end{pmatrix}$$



3. Cluster algebras

In this section we present a very short introduction to cluster algebras, following Fomin-Zelevinsky [**14**]. A cluster algebra $\mathscr{A} = \mathscr{A}(\underline{x}, B)$ is a subring of the ring $\mathbb{F} = \mathbb{Q}(u_1, \dots, u_m)$, associated to a *seed* $(\underline{x}, B)$ defined in the following way.

(i) A seed is a pair $(\underline{x}, B)$ consisting of a *cluster* $\underline{x} = (x_1, \dots, x_m)$ where $\underline{x}$ is a free generating set of $\mathbb{F}$ over $\mathbb{Q}$ and $B = (b_{xy})_{xy}$ is a sign-symmetric $m \times m$ matrix with integer coefficients, i.e. $b_{xy} \in \mathbb{Z}$ for all $1 \le x, y \le m$ and if $b_{xy} > 0$ then $b_{yx} < 0$.

(ii) A seed $(\underline{x}, B)$ can be mutated to another seed $(\underline{x}', B')$: *mutation at* $z \in \underline{x}$ is the map μ_z: $(\underline{x}, B) \mapsto (\underline{x}', B')$: $\underline{x}' = \underline{x} - z \cup z'$ where z' is defined via the *exchange relation*

$$zz' = \prod_{\substack{x\in\underline{x} \\ b_{xz}>0}} x^{b_{xz}} + \prod_{\substack{x\in\underline{x} \\ b_{xz}<0}} x^{-b_{xz}}$$

and B' is defined similarly via *matrix mutation*:

$$b'_{xy} = \begin{cases} -b_{xy} & \text{if } x = z \text{ or } y = z \\ b_{xy} + \frac{1}{2}(|b_{xz}| \cdot b_{zy} + b_x \cdot |b_{zy}|) & \text{otherwise.} \end{cases}$$

Note: μ is involutive, i.e. $\mu_{z'}(\mu_z((\underline{x}, B))) = (\underline{x}, B)$.

Two seeds $(\underline{x}, B)$ and $(\underline{x}', B')$ are said to be *mutation-equivalent* if one can be obtained from the other through a sequence of mutations. The *cluster variables* are defined to be the union of all clusters of a mutation-equivalence class (of a given seed). These appear in overlapping sets. Finally, the corresponding *cluster algebra*

$\mathscr{A} = \mathscr{A}(\underline{x}, B)$ is the subring of $\mathbb{F}$ generated by all the cluster variables. (Here we are defining cluster algebras with trivial coefficients.) A cluster algebra is said to be *of finite type* if there exists only a finite number of cluster variables.

One can show that up to isomorphism of cluster algebras $\mathscr{A}(\underline{x}, B)$ does not depend on the initial choice of a free generating set $\underline{x}$.

Example 3.1. (Type A_2). We start with the pair $\underline{x} = (x_1, x_2)$, B= $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$. In a first step we mutate x_1. from $x_1 x_1' = 1 + x_2$ we obtain $x_1' = \frac{1+x_2}{x_1}$. The next mutation is at x_2 (mutation at x_1' would lead us back to x_1), we have $x_2' = \frac{x_1+1}{x_2}$. And then $x_1'' = \frac{x_1+x_2+1}{x_1 x_2}$; $x_2'' = x_1$, $x_1''' = x_2$.
In particular, we obtain five cluster variables in this example.

Some of the main results on cluster algebras are summarized here:

– *Laurent phenomenon:* $\mathscr{A}(\underline{x}, B)$ sits inside $\mathbb{Z}[x_1^{\pm}, \dots, x_m^{\pm}]$, i.e. every element of the cluster algebra is an integer Laurent polynomial in the variables of $\underline{x}$ (cf. [**15**]);

– Classification of finite type cluster algebras by roots systems, [**16**] (cluster algebras of finite type can be classified by Dynkin diagrams);

– Realizations of algebras of regular functions on double Bruhat cells in terms of cluster algebras ([**4**]).

Examples of cluster algebras are: Coordinate rings of SL_2, SL_3 ([**17**]); Plücker coordinates on $Gr_{2,n+3}$ ([**32**], [**22**]).

Cluster algebras and quivers. We will now explain how to associate a quiver to a seed of a cluster algebra.

Recall that a *quiver* $\Gamma = (\Gamma_0, \Gamma_1)$ is an oriented graph with vertices Γ_0 and arrows Γ_1 between them. E.g.

$$1 \xrightarrow{\alpha} 2 \xrightarrow{\beta} 3$$

with $\Gamma_0 = \{1, 2, 3\}$ and $\Gamma_1 = \{\alpha, \beta\}$;

Any skew-symmetric $m \times m$-matrix B determines a quiver $\Gamma(B)$ with m vertices. One labels the columns of B by $\{1, 2, \dots, m\}$ and sets $\Gamma_0 = \{1, 2, \dots, m\}$. Then one draws b_{xy} arrows from x to y if $b_{xy} > 0$ (for $x, y \in \Gamma_0$).

Such a quiver has no loops and for any two vertices $i \neq j$ of $\Gamma(B)$, there are only arrows in one direction between them.

So in particular, if the matrix B of a seed $(\underline{x}, B)$ is skew-symmetric, it determines a quiver in this way.

Example 3.2. The matrix B= $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ from Example 3.1 above gives the quiver:

$$1 \longrightarrow 2$$

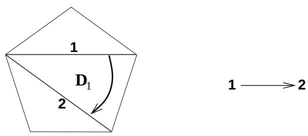
Clearly, this process is reversible: a quiver whose arrows only go in one direction between any given pair $i \neq j$ of vertices and without loops gives rise to a skew-symmetric matrix which we will denote by $B(\Gamma)$.

4. From triangulations to cluster algebras

From now on we assume that (S,M) is a disc with at most one puncture. We want to show how a triangulation T of (S,M) determines a cluster algebra. Label the arcs of T by $1,2,\dots,n$.

Then we define a cluster $\underline{x}_T = (x_1,\dots,x_n)$ by sending $i \mapsto x_i$ and choose as a matrix the the skew-symmetric matrix associated $B(T)$ associated to T in Section 2. This clearly produces a seed $(\underline{x}_T, B(T))$. Thus to the triangulation T of the disc (S,M) we have associated the seed $(\underline{x}_T, B(T))$ and hence obtain a cluster algbra $\mathscr{A} = \mathscr{A}(\underline{x}_T, B(T))$.

Example 4.1. Consider an unpunctured pentagon as below. In the triangulation, we label the arcs 1 and 2. They form a triangle D_1 together with a boundary arc and 2 is the clockwise neighbour of 1.



Then the seed we obtain is $((x_1,x_2), B(T))$ with $B(T) = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ as in Example 3.1 above.

5. Cluster categories

Cluster categories are quotients of derived categories of module categories. They were introduced by Buan-Marsh-Reineke-Reiten-Todorov ([**6**]).

Independently, Caldero-Chapoton-Schiffler have introduced the cluster categories (in type A_n) in 2005 ([**7**]) using a graphical description. Later, Schiffler extended this to type D_n in ([**31**]).

The aim behind the definition of cluster categories was to model cluster algebras using the representation theory of quivers. This was motivated by the observation that the cluster variables of a cluster algebra of finite type are parametrized by the almost positive roots of the corresponding root system.

Cluster categories have led to new developments in the theory of the (dual of the) canonical bases, they provide insight into cluster algebras. They have also developed into a field of their own. E.g. they have led to the definition of cluster-tilting theory.

Let us describe the construction of cluster categories, following [**6**].

We start with a quiver Q whose underlying graph is a simply-laced Dynkin diagram (i.e. of type ADE). Denote by $D^b(kQ)$ the bounded derived category of finite dimensional kQ-modules (we assume that the field k is algebraically closed). Note that the shape of the quiver of $D^b(kQ)$ is $Q\times\mathbb{Z}$ with certain connecting arrows. By *quiver of* $D^b(kQ)$ we mean the Auslander-Reiten quiver of $D^b(kQ)$, i.e. the quiver whose vertices are the isomorphism classes of indecomposable modules and whose arrows come from irreducible maps between them.

This quiver has two well-known graph automorphisms: τ ("Auslander-Reiten translate") which sends each vertex to its neighbour to the left. And [1] (the "shift") which sends a vertex in a copy of the module category of kQ to the corresponding vertex in the next copy of the module category.

The *cluster category*, $\mathscr{C}$, is now defined as the orbit category of $D^b(kQ)$ under a canonical automorphism:

$$\mathscr{C} := \mathscr{C}(Q) := D^b(kQ)/\tau^{-1}\circ[1]$$

One can show that this is independent of the chosen orientation of Q. More generally, Keller ([**27**]) has introduced the m-cluster category, $\mathscr{C}^m$ as follows:

$$\mathscr{C}^m := \mathscr{C}^m(Q) := D^b(kQ)/\tau^{-1}\circ[m]$$

Keller has shown in [**27**] that $\mathscr{C}^m$ is triangulated and a Calabi-Yau category of dimension $m+1$. Furthermore, $\mathscr{C}^m$ is Krull-Schmidt ([**6**]). The m-cluster category has attracted a lot of interest over the last few years. In particular, it has been studied by Keller-Reiten, Thomas, Wralsen, Zhu, B-Marsh, Assem-Brüstle-Schiffler-Todorov, Amiot, Wralsen, etc.

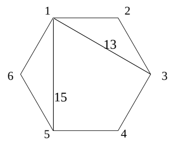
Our goal for this note is to describe $\mathscr{C}^m$ using diagonals of a polygon (type A_n) and arcs in a punctured polygon (type D_n).

6. From arcs via quivers to cluster categories

Let us first recall the notion of a stable translation quiver due to Riedtmann [**30**].

Definition. A *stable translation quiver* is a pair (Γ, τ) where $\Gamma = (\Gamma_0, \Gamma_1)$ is a quiver (locally finite, without loops) and $\tau : \Gamma_0 \to \Gamma_0$ is a bijective map such that the number of arrows from x to y equals the number of arrows from τy to x for all $x,\ y \in \Gamma_0$. The map τ is called the *translation* of (Γ, τ).

Now we are ready to define a quiver Γ from a hexagon (see figure below) as follows:



Γ_0: The vertices are the diagonals $(i\,j)$ of the hexagon $(1 \le i < j-1 \le 7)$.
Γ_1: The arrows are of the form $(i\,j) \to (i, j+1)$, $(i\,j) \to (i+1, j)$ provided the target is also a diagonal in the hexagon $(i, j \in \mathbb{Z}_6)$.
Set τ: $(i\,j) \mapsto (i-1, j-1)$ to be anti-clockwise rotation about the center.

The quiver obtained this way from the hexagon is the following:

$$\begin{tikzcd}[column sep=small]
 & & 15 \arrow[dr] \arrow[rr, dash, dashed] & & 26 \arrow[dr] \arrow[rr, dash, dashed] & & 13 \arrow[dr] & & \\
 & 14 \arrow[ur] \arrow[dr] \arrow[rr, dash, dashed] & & 25 \arrow[ur] \arrow[dr] \arrow[rr, dash, dashed] & & 36 \arrow[ur] \arrow[dr] \arrow[rr, dash, dashed] & & 14 \arrow[dr] & \\
13 \arrow[ur] \arrow[rr, dash, dashed] & & 24 \arrow[ur] \arrow[rr, dash, dashed] & & 35 \arrow[ur] \arrow[rr, dash, dashed] & & 46 \arrow[ur] \arrow[rr, dash, dashed] & & 15
\end{tikzcd}$$

It clearly is an example of a stable translation quiver.

Note that such a quiver can be defined for any polygon. Denote the quiver arising in that way by $\Gamma(n,1)$ if $n+2$ is the number of vertices of the polygon. (The use of n instead of $n+2$ in the notation of the quiver $\Gamma(n,1)$ and the extra entry 1 are used to make this compatible with the more general setting involving m-diagonals described below). Caldero, Chapton and Schiffler have shown that the cluster category can be obtained via diagonals in a polygon:

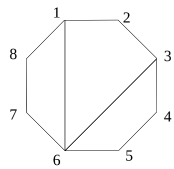
Theorem 6.1 ([**7**])**.** *The quiver of the cluster category $\mathscr{C} = \mathscr{C}(A_{n-1})$ is isomorphic to the quiver $\Gamma(n,1)$ obtained from an $(n+2)$-gon.*

As before in the case of the bounded derived category, the *quiver of* $\mathscr{C}$ is an abbreviation for the Auslander-Reiten quiver of $\mathscr{C}$. It has as vertices the indecomposable objects of $\mathscr{C}$, and as arrows are the irreducible maps between them.

To be able to model m-cluster categories we now generalize the notion of diagonal and introduce the so-called m-diagonals. We start with a polygon Π with $nm+2$ vertices ($n, m \in \mathbb{N}$), labeled by $1, 2, \ldots, nm+2$.

Definition. An *m-diagonal* is a diagonal (ij) dividing Π into an $mj+2$-gon and an $m(n-j)+2$-gon ($1 \leq j \leq \frac{n-1}{2}$).

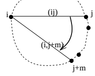
Example 6.2. To illustrate this, let Π be an octagon, $n = 3$, $m = 2$. In that case, $1 \leq j \leq 1$, so any 2-diagonal has to divide Π into a quadrilateral and a hexagon.



Observe that each maximal set of non-crossing 2-diagonals contains two elements. They are $\{(16), (36)\}$, $\{(16), (25)\}$, $\{(16), (14)\}$ and rotated version of these.

Recall that the number of arcs in a triangulation (see Section 2) is an invariant of a disc (S, M), called the rank of (S, M). In the same way, the maximal number of non-crossing m-diagonals is an invariant of the polygon. It is equal to $n-1$ (for the $nm+2$-gon Π).

Using m-diagonals we can now define a translation quiver $\Gamma(n, m) = (\Gamma, \tau_m)$:
Γ_1: $(ij) \to (ij')$ if (ij), $B_{jj'}$ and (ij') span an $m+2$-gon ($B_{jj'}$ is the boundary j to j', going clockwise).
Γ_0: The vertices are the m-diagonals (ij) in Π (with $1 \leq i < j-m$).
Γ_1: The arrows are of the form $(ij) \to (i, j+m)$, $(ij) \to (i+m, j)$ provided the target is still inside the polygon. In other words: (ij), $(i, j+m)$ and the boundary arc j to $j+m$ (resp. (ij), $(i+m, j)$ and the boundary arc from i to $i+m$) form an $m+2$-gon as in the picture:



Furthermore, let τ_m be anti-clockwise rotation (about center) through the angle $m\frac{2\pi}{nm+2}$.

Remark. It is clear that $\Gamma(n,m)$ is a stable translation quiver. In case $m = 1$, we recover the usual diagonals.

The quiver $\Gamma(3,2)$ for the octagon from the previous example is thus:

$$\begin{tikzcd}
 & 16 \arrow[rr, dash, dashed] \arrow[dr] & & 38 \arrow[rr, dash, dashed] \arrow[dr] & & 25 \arrow[rr, dash, dashed] \arrow[dr] & & 47 \arrow[rr, dash, dashed] \arrow[dr] & & 16 \\
14 \arrow[rr, dash, dashed] \arrow[ur] & & 36 \arrow[rr, dash, dashed] \arrow[ur] & & 58 \arrow[rr, dash, dashed] \arrow[ur] & & 27 \arrow[rr, dash, dashed] \arrow[ur] & & 14 \arrow[ur] &
\end{tikzcd}$$

Then one can show that the A-type m-cluster category can be obtained using m-diagonals in a polygon:

Theorem 6.3 ([1]). *The quiver of the m-cluster category $\mathscr{C}^m = \mathscr{C}^m(A_{n-1})$ is isomorphic to the quiver $(\Gamma(m,n),\tau_m)$ obtained from m-diagonals in an $nm+2$-gon.*

The proof of our result uses Happel's description of the Auslander-Reiten-quiver of $D^b(kQ)$ where Q is of Dynkin type A_{n-1} and combinatorial analysis of $\Gamma(n,m)$. For details we refer to [**1**, Section 5].

The description in type D. We have a similar description of the m-cluster categories of D-type. Instead of working with a polygon (or unpunctured disc) we now have to use a punctured polygon. Let Π be a punctured $nm-m+1$-gon. Instead of using the term diagonal, we now speak of *arcs* in Π. An arc going from i to j, homotopic equivalent to the boundary B_{ij} from i to j (going clockwise) is denoted by (ij). By (ii) we denote an arc homotopic equivalent to the boundary B_{ii} with endpoints in i. And $(i0)$ is an arc homotopic equivalent to the arc between i and the puncture 0. We will say that an n-gon is *degenerate* if it has n sides and $n-1$ vertices.

For details and examples we refer to [**3**, Section 3].

Definition. An *m-arc* of Π is an arc (ij) such that

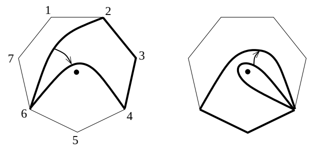
(i) (ij) and B_{ij} (the boundary from i to j, going clockwise) form an $km+2$-gon for some k,

(ii) (ij) and B_{ji} (the boundary from j to i, going clockwise) form an $lm+2$-gon for some l.

Furthermore, (ii) and $(i0)$ are called *m-arcs* if (ii) and B_{ii} form a degenerate $km+2$-gon for some k.

Then we can define a translation quiver $\Gamma = \Gamma_{\odot}(n,m)$ as follows:
Γ_0: The vertices are the m-arcs of Π
Γ_1: The arrows are the so-called m-moves between vertices:

We say that $(ij) \to (ik)$ is an *m-move* if (ij), B_{jk} and (ik) span a (degenerate) $m+2$-gon. In the figure below there are two examples of 2-moves.



τ_m: rotation anti-clockwise (about center), through an angle of $\frac{m}{nm-m+1}$.

Clearly, the pair $(\Gamma_{\odot}(n,m), \tau_m)$ is also a stable translation quiver. We can now formulate the statement.

Theorem 6.4 (Theorem 3.6 of [3]). *The quiver of the m-cluster category $\mathscr{C}^m = \mathscr{C}^m(D_n)$ is isomorphic to the quiver $(\Gamma_{\odot}(m,n), \tau_m)$ obtained from m-arcs in an $nm - m + 1$-gon.*

The m-th power of a translation quiver. We will now describe another way to obtain m-cluster categories directly from the diagonals or arcs in a (punctured) polygon. Let (Γ,τ) a translation quiver as before. Then we define the *m-th power of* Γ, Γ^m, to be the quiver whose vertices are the vertices of Γ (i.e. $\Gamma_0^m = \Gamma_0$). The arrows in Γ^m are paths of length m, going in an unique direction. (To be precise, we ask that such a path is *sectional,* i.e. that in a path $x_0 \to x_1 \to \cdots \to x_{m-1} \to x_m$ of length m we have $\tau x_{i+1} \neq x_{i-1}$ whenever τx_{i+1} is defined.) And the translation τ^m of Γ^m is obtained by repeating the original translation m times.

Definition. The quiver (Γ^m, τ^m) as defined above is called the *m-th power of* (Γ,τ).

With this we are ready to formulate the result:

Theorem 6.5 ([1]). $\Gamma(n,m)$ *is a connected component of*

$$(\Gamma(nm,1))^m = (\Gamma(\textit{cluster category}))^m.$$

Remark. Observe that (Γ^m, τ^m) is again a stable translation quiver. However, even if (Γ,τ) is connected, the m-th power is not connected in general!

Example 6.6. To illustrate this consider the quiver $\Gamma(6,1)$ of an octagon.

$$\begin{tikzcd}[column sep=small]
 & & & & 17 \arrow[rr, dash, dashed] \arrow[dr] & & 28 \arrow[rr, dash, dashed] \arrow[dr] & & 13 \arrow[dr] & & & & \\
 & & & 16 \arrow[rr, dash, dashed] \arrow[ur] \arrow[dr] & & 27 \arrow[rr, dash, dashed] \arrow[ur] \arrow[dr] & & 38 \arrow[rr, dash, dashed] \arrow[ur] \arrow[dr] & & 14 \arrow[dr] & & & \\
 & & 15 \arrow[rr, dash, dashed] \arrow[ur] \arrow[dr] & & 26 \arrow[rr, dash, dashed] \arrow[ur] \arrow[dr] & & 37 \arrow[rr, dash, dashed] \arrow[ur] \arrow[dr] & & 48 \arrow[rr, dash, dashed] \arrow[ur] \arrow[dr] & & 15 \arrow[dr] & & \\
 & 14 \arrow[rr, dash, dashed] \arrow[ur] \arrow[dr] & & 25 \arrow[rr, dash, dashed] \arrow[ur] \arrow[dr] & & 36 \arrow[rr, dash, dashed] \arrow[ur] \arrow[dr] & & 47 \arrow[rr, dash, dashed] \arrow[ur] \arrow[dr] & & 58 \arrow[rr, dash, dashed] \arrow[ur] \arrow[dr] & & 16 \arrow[dr] & \\
13 \arrow[rr, dash, dashed] \arrow[ur] & & 24 \arrow[rr, dash, dashed] \arrow[ur] & & 35 \arrow[rr, dash, dashed] \arrow[ur] & & 46 \arrow[rr, dash, dashed] \arrow[ur] & & 57 \arrow[rr, dash, dashed] \arrow[ur] & & 68 \arrow[rr, dash, dashed] \arrow[ur] & & 17
\end{tikzcd}$$

Its second power has three components: one component is $\Gamma(2,2)$. The two other components are both quivers of quotients $D^b(\mathrm{A}_3)/[1]$ of $D^b(A_3)$ by the shift. In particular, we have thus given a geometric construction of a quotient of $D^b(A_3)$ which is not an m-cluster category!

The three components of the second power of $\Gamma(6,1)$ are:

$$\begin{tikzcd}[column sep=small]
 & 16 \arrow[rr, dash, dashed] \arrow[dr] & & 38 \arrow[rr, dash, dashed] \arrow[dr] & & 25 \arrow[rr, dash, dashed] \arrow[dr] & & 47 \arrow[rr, dash, dashed] \arrow[dr] & & 16 \\
14 \arrow[rr, dash, dashed] \arrow[ur] & & 36 \arrow[rr, dash, dashed] \arrow[ur] & & 58 \arrow[rr, dash, dashed] \arrow[ur] & & 27 \arrow[rr, dash, dashed] \arrow[ur] & & 14 \arrow[ur] &
\end{tikzcd}$$

$$\begin{tikzcd}[column sep=tiny]
 & & 17 \arrow[rr, dash, dashed] \arrow[dr] & & 13 \arrow[dr] & & \\
 & 15 \arrow[rr, dash, dashed] \arrow[ur] \arrow[dr] & & 37 \arrow[rr, dash, dashed] \arrow[ur] \arrow[dr] & & 15 \arrow[dr] & \\
13 \arrow[rr, dash, dashed] \arrow[ur] & & 35 \arrow[rr, dash, dashed] \arrow[ur] & & 57 \arrow[rr, dash, dashed] \arrow[ur] & & 17
\end{tikzcd}
\qquad
\begin{tikzcd}[column sep=tiny]
 & & 28 \arrow[rr, dash, dashed] \arrow[dr] & & 24 \arrow[dr] & & \\
 & 26 \arrow[rr, dash, dashed] \arrow[ur] \arrow[dr] & & 48 \arrow[rr, dash, dashed] \arrow[ur] \arrow[dr] & & 26 \arrow[dr] & \\
24 \arrow[rr, dash, dashed] \arrow[ur] & & 46 \arrow[rr, dash, dashed] \arrow[ur] & & 68 \arrow[rr, dash, dashed] \arrow[ur] & & 24
\end{tikzcd}$$

Remark. We have a corresponding result for type D, see Theorem 5.1 of [**3**]. In addition, in type D we give an explicit description of *all* connected components appearing in the m-th power of $\Gamma_{\odot}(nm-m+1,1)$

7. Connections and future directions

To finish we want to provide a short outlook and describe some open problems and possible future directions.

– In recent work with Robert Marsh ([**2**]) we provide a link between cluster algebra combinatorics and perfect matchings (for vertices and edges of a triangulation). This uses work of Conway-Coxeter ([**9**], [**10**]), and of Broline-Crowe-Isaacs ([**5**]) on frieze patterns of positive integers.

– Y-systems can be defined in general for pairs (G, H) of Dynkin types. Zamolodchikovs periodicity conjecture for general Y-systems have been proved for $G = A_1$ and $H = A_n$ by Frenkel-Szenes ([**19**]), by Gliozzi-Tateo ([**24**]) and for $G = A_1$, H any Dynkin type by Fomin-Zelevinsky ([**18**]) using cluster algebras theory. More recently, the cases $G = A_k$, $H = A_n$ have been solved ([**33**], [**34**], independently). In this context, there are various open questions. First of all: can periodicity be proved for G of arbitrary of Dynkin type using the theory of cluster algebras? This is not even known for $G = A_k$. Second: what would be a good counterpart on the side of Y-systems to the geometric model for m-cluster categories? And thirdly: In current work with Marsh we have discovered classes of infinite periodic systems ($G = A_1$, $H = A_\infty$). Does this have a translation to the setting of Y-systems?

– The approach to model cluster algebras with discs (S, M) works for types A and D ([**13**]) and for types B, C under certain modifications ([**8**]).
Open: what can be said about the exceptional types, in particular, is there a way to model type E using a disc with marked points?

– Jorgensen ([**26**]) has obtained m-cluster categories as quotient categories of cluster categories via deletion of rows (τ-orbits). They inherit a triangulated structure. This process can be viewed as a reverse to our construction using the m-th power of a quiver. *Question*: how can we explain the triangulated structure of m-cluster categories via the m-th power of a quiver?

References

[1] K. BAUR & R. J. MARSH – "A geometric construction of m-cluster categories", to appear in Trans. AMS.

[2] ______, "Ptolemy relations for punctured discs", preprint, arXiv:0711.1443v1 [math.CO].

[3] ______, "A geometric description of the m-cluster categories of type D_n.", (2007).

[4] A. BERENSTEIN, S. FOMIN & A. ZELEVINSKY – "Cluster algebras. III: Upper bounds and double Bruhat cells.", *Duke Math. J.* **126** (2005), no. 1, p. 1–52.

[5] D. BROLINE, D. CROWE & I. ISAACS – "The geometry of frieze patterns.", *Geom. Dedicata* **3** (1974), p. 171–176.

[6] A. B. BUAN, R. MARSH, M. REINEKE, I. REITEN & G. TODOROV – "Tilting theory and cluster combinatorics.", *Adv. Math.* **204** (2006), no. 2, p. 572–618.

[7] P. CALDERO, F. CHAPOTON & R. SCHIFFLER – "Quivers with relations arising from clusters (A_n case).", *Trans. Am. Math. Soc.* **358** (2006), no. 3, p. 1347–1364.

[8] F. CHAPOTON, S. FOMIN & A. ZELEVINSKY – "Polytopal realizations of generalized associahedra.", *Can. Math. Bull.* **45** (2002), no. 4, p. 537–566.

[9] J. CONWAY & H. COXETER – "Triangulated polygons and frieze patterns.", *Math. Gaz.* **57** (1973), p. 87–94.

[10] ———, "Triangulated polygons and frieze patterns.", *Math. Gaz.* **57** (1973), p. 175–183.

[11] F. V. FOCK & A. B. GONCHAROV – "Dual Teichmüller and lamination spaces.", Papadopoulos, Athanase (ed.), Handbook of Teichmüller theory. Volume I. Zürich: European Mathematical Society (EMS). IRMA Lectures in Mathematics and Theoretical Physics 11, 647-684 (2007)., 2007.

[12] V. FOCK & A. GONCHAROV – "Moduli spaces of local systems and higher Teichmüller theory.", (2006).

[13] S. FOMIN, D. SHAPIRO & D. THURSTON – "Cluster algebras and triangulated discs. Part i: Cluster complexes", Preprint arxiv:math.RA/0608367v3, August 2006.

[14] S. FOMIN & A. ZELEVINSKY – "Cluster algebras. I: Foundations.", *J. Am. Math. Soc.* **15** (2002), no. 2, p. 497–529.

[15] ———, "The Laurent phenomenon.", *Adv. Appl. Math.* **28** (2002), no. 2, p. 119–144.

[16] ———, "Cluster algebras. II: Finite type classification.", *Invent. Math.* **154** (2003), no. 1, p. 63–121.

[17] ———, "Cluster algebras: Notes for the CDM-03 conference.", Mazur, Barry (ed.) et al., Current developments in mathematics, 2003. Somerville, MA: International Press. 1-34 (2003)., 2003.

[18] ———, "Y-systems and generalized associahedra.", (2003).

[19] E. FRENKEL & A. SZENES – "Thermodynamic Bethe ansatz and dilogarithm identities. I.", *Math. Res. Lett.* **2** (1995), no. 6, p. 677–693.

[20] C. GEISS, B. LECLERC & J. SCHRÖER – "Semicanonical bases and preprojective algebras.", (2005).

[21] ———, "Semicanonical bases and preprojective algebras. II: A multiplication formula.", *Compos. Math.* **143** (2007), no. 5, p. 1313–1334.

[22] M. GEKHTMAN, M. SHAPIRO & A. VAINSHTEIN – "Cluster algebras and Poisson geometry.", *Mosc. Math. J.* **3** (2003), no. 3, p. 899–934.

[23] M. GEKHTMAN, M. SHAPIRO & A. VAINSHTEIN – "Cluster algebras and Weil-Petersson forms.", *Duke Math. J.* **127** (2005), no. 2, p. 291–311.

[24] F. GLIOZZI & R. TATEO – "Thermodynamic Bethe ansatz and three-fold triangulations.", *Int. J. Mod. Phys. A* **11** (1996), no. 22, p. 4051–4064.

[25] A. E. HATCHER – "On triangulations of surfaces.", *Topology Appl.* **40** (1991), no. 2, p. 189–194.

[26] P. JORGENSEN – "Quotients of cluster categories", Preprint, arXiv:0705.1117v1.

[27] B. KELLER – "On triangulated orbit categories.", *Doc. Math., J. DMV* **10** (2005), p. 551–581.

[28] G. LUSZTIG – "Total positivity in reductive groups.", Brylinski, Jean-Luc (ed.) et al., Lie theory and geometry: in honor of Bertram Kostant on the occasion of his 65th birthday. Invited papers, some originated at a symposium held at MIT, Cambridge, MA, USA in May 1993. Boston, MA: Birkhäuser. Prog. Math. 123, 531-568 (1994)., 1994.

[29] D. I. PANYUSHEV – "ad-nilpotent ideals of a Borel subalgebra: generators and duality.", *J. Algebra* **274** (2004), no. 2, p. 822–846.

[30] C. RIEDTMANN – "Algebren, Darstellungsköcher, Überlagerungen und zurück.", *Comment. Math. Helv.* **55** (1980), p. 199–224.

[31] R. SCHIFFLER – "A geometric model for cluster categories of type D_n.", *J. Algebr. Comb.* **27** (2008), no. 1, p. 1–21.

[32] J. S. SCOTT – "Grassmannians and cluster algebras.", *Proc. Lond. Math. Soc., III. Ser.* **92** (2006), no. 2, p. 345–380.

[33] A. SZENES – "Periodicity of y-systems and flat connections", Preprint arxiv:math.RT/0606377.

[34] A. Y. VOLKOV – "On zamolodchikov's periodicity conjecture", Preprint arxiv:hep-th/0606094.

[35] A. ZAMOLODCHIKOV – "On the thermodynamic Bethe ansatz equation in the sinh-Gordon model.", *J. Phys. A, Math. Gen.* **39** (2006), no. 41, p. 12863–12887.

Mathematisches Institut, Courant-Colloquium Trends in Mathematics
(R. MEYER, T. SCHICK, Y. TSCHINKEL, eds.), p. 31–46
Georg-August Universität Göttingen, 2008

DISPERSIVE EQUATIONS AND HYPERBOLIC ORBITS

Hans Christianson

Department of Mathematics, Massachusetts Institute of Technology, 77 Mass. Ave., Cambridge, MA 02139, USA • *E-mail :* hans@math.mit.edu

Abstract. We consider several related problems in linear PDE on manifolds, both when the underlying manifold is compact and when it is non-compact. Our general assumption is that the classical flow has a single periodic hyperbolic (unstable) orbit. The standard techniques from linear PDE relying on dispersive effects and non-trapping assumptions do not directly apply in this situation, but the unstable nature of the periodic orbit allows us to prove slightly weaker results.

1. Introduction and Applications in PDE

In this note we report on some recent results in the field of Partial Differential Equations (PDE) on manifolds. The most fundamental problems in this area stem from trying to understand the effects of classical geometry on the solutions to PDE. We consider the effects of a single hyperbolic "trapped" ray on solutions to a number of problems on both compact and noncompact Riemannian manifolds.

1.1. Eigenfunction Concentration. Let (M, g) be a smooth, compact, Riemannian manifold with or without boundary, and let $-\Delta_g$ be the (non-negative) Laplace-Beltrami operator acting on functions (with Dirichlet boundary conditions if $\partial M \neq \emptyset$). We consider the eigenvalue problem:

$$\left\{ \begin{array}{c} (-\Delta_g - \lambda^2) u = 0 \\ \int_M |u|^2 dx = 1, \text{ and } u|_{\partial M} = 0 \text{ if } \partial M \neq \emptyset \end{array} \right. \tag{1.1}$$

2000 ***Mathematics Subject Classification.*** 58J99, 35B99.

A very natural question to ask is “what do the solutions *look like*?” The idea is that the underlying *classical* geometry (of Hamiltonian dynamics on the cotangent bundle) affects solutions to PDE on the manifold. For eigenfunctions, the presence of periodic geodesics may cause concentration or “scarring” phenomena. There are 3 cases we describe here: ergodic classical dynamics, elliptic periodic geodesics, and hyperbolic periodic geodesics. The first two cases are well-studied in the literature (see references below), and the case of an isolated hyperbolic geodesic is the subject of these notes.

The ergodic case was studied first by Zelditch [**20**] and Colin de Verdière [**17**], and has been generalized to many different situations (Figure 1 is an example of ergodic dynamics). The idea is that since ergodic flow is roughly “evenly-mixing”, then eigenfunctions should somehow become evenly distributed as the eigenvalue or energy increases. To be more precise, let $U \subset M$ be an open set. Then it is known that

$$\int_U |u|^2 dx \simeq \frac{\text{Area } U}{\text{Area } M}, \quad \lambda \to \infty$$

along a *density* 1 subsequence of eigenvalues λ.

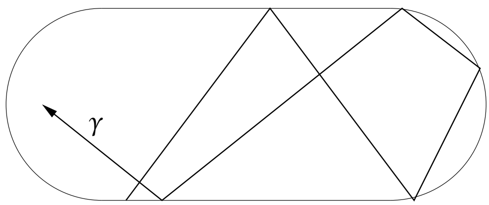


FIGURE 1. The Bunimovich stadium is known to have ergodic billiard dynamics.

In the case of an elliptic (stable) periodic geodesic (see Figure 2), we expect the stability of the classical dynamics to lead to well-concentrated eigenfunctions. What we can prove is that there are well-localized quasimodes (approximate eigenfunctions). In fact, the techniques used to prove Theorem 2 below provide a new proof of the following well-known Theorem.

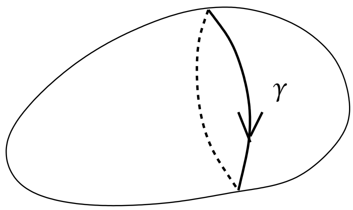


FIGURE 2. A compact manifold without boundary and stable geodesic.

Theorem 1 (Theorem 6, [6]). *Suppose that $\gamma \subset M$ is an elliptic periodic geodesic (making only transversal reflections with ∂M if $\gamma \cap \partial M = \emptyset$). Let $U \supset \gamma$ be any open neighbourhood. Then there exist functions $\varphi_j \in L^2(X)$ and values $\lambda_j \in [0,\infty)$ satisfying*

$$\begin{cases} (-\Delta_g - \lambda_j^2)\varphi_j = \mathcal{O}(\lambda_j^{-\infty}) \\ \int_M |\varphi_j|^2 dx = 1, \text{ and } \varphi_j|_{\partial M} = 0 \text{ if } \partial M \neq \emptyset, \end{cases}$$

and

$$\int_{M \setminus U} |\varphi_j|^2 dx = \mathcal{O}(\lambda_j^{-\infty})$$

as $\lambda_j \to \infty$.

In the case of an unstable geodesic (see Figures 3 and 4), the situation changes. If the classical flow is not also ergodic, there is no reason to expect eigenfunctions will not concentrate on the orbit. However, the unstable nature of the geodesic suggests if there is concentration, it will be very weak. We have the following Theorem.

Theorem 2 (Main Theorems [7, 6]). *Suppose that $\gamma \subset M$ is an unstable closed geodesic, making only transversal reflections with ∂M if $\gamma \cap \partial M \neq \emptyset$, and $U \supset \gamma$ is a sufficiently small neighbourhood. Then there exists a constant $C > 0$ such that if u satisfies* (1.1),

$$\int_{M \setminus U} |u|^2 dx \geqslant \frac{1}{C \log |\lambda|}, \quad |\lambda| \to \infty.$$

Remark 1. The conclusion of Theorem 2 applies also to semi-hyperbolic geodesics; that is, those which are unstable in a minimum of one direction in phase space.

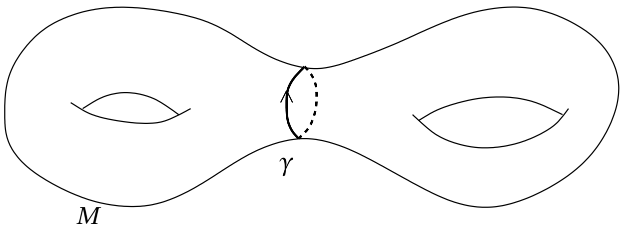


FIGURE 3. A compact manifold M without boundary and unstable closed geodesic γ.

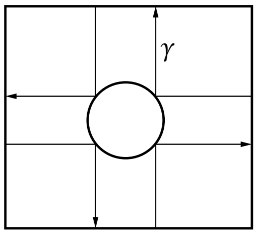


FIGURE 4. A compact manifold M with boundary and unstable closed geodesic γ.

The estimates in Theorem 2 are the same as those obtained in the case the classical flow is completely integrable. Thus we conclude the completely integrable case is the "worst" in some sense.

Theorem 2 generalizes results of Colin de Verdière-Parisse [**18**] for a surface of revolution, which is of course completely integrable. In [**18**] they show in addition that this estimate is sharp in this case. Theorem 2 also generalizes work of Burq-Zworski [**2**] for real hyperbolic geodesics.

1.2. Damped Wave Equation. The estimates used to prove Theorem 2 have an immediate application to the study of solutions to the damped wave equation on compact manifolds. Let $a(x) \in \mathscr{C}^\infty(M)$, $a \geqslant 0$, and consider the damped wave equation

$$\begin{cases} (\partial_t^2 - \Delta_g + 2a(x)\partial_t)u = 0 \\ u(x,0) = 0, \ \ u_t(x,0) = f(x) \end{cases}$$

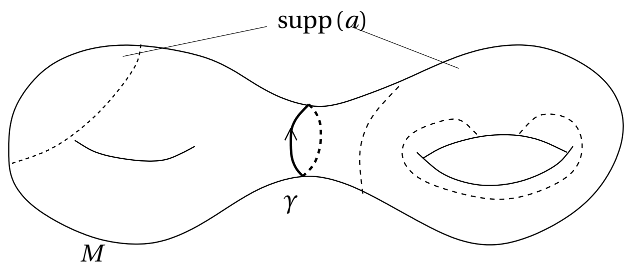


FIGURE 5. A compact manifold without boundary, unstable geodesic γ, and a controls M geometrically outside a neighbourhood of γ.

We define the energy $E(t)$ to be

$$E(t) := \|\partial_t u\|^2_{L^2(M)} + \|\nabla u\|^2_{L^2(M)}.$$

Integrations by parts shows if $a \equiv 0$, then

$$E(t) \equiv \|f\|^2_{L^2(M)},$$

while if $a > 0$ somewhere,

$$E'(t) < 0.$$

We say a controls M geometrically if every geodesic meets the support of a within some finite fixed time. We have the classical result of Rauch-Taylor [**16**]: if $a \geqslant \delta > 0$ everywhere or we have geometric control, then

$$E(t) \leqslant Ce^{-t/C}\|f\|^2_{L^2(M)}.$$

If $\gamma \subset U$ is a hyperbolic periodic geodesic, we may only have geometric control *outside of U* (see Figure 5).

Theorem 3 (Theorem 5 [7]). *If a controls M geometrically outside U, then for any* $\varepsilon > 0$ *there exists* $C > 0$ *such that*

$$E(t) \leqslant Ce^{-t/C}\|f\|^2_{H^\varepsilon(M)}.$$

Remark 2. This generalizes the work of Rauch-Taylor mentioned above in the global control case, as well as Lebeau [**10**] for a surface of revolution with incomplete geometric control.

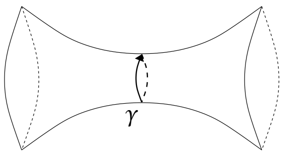


FIGURE 6. A piece of the catenoid.

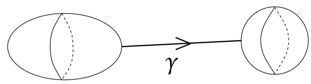


FIGURE 7. $\mathbb{R}^n$ with two convex bodies removed.

1.3. Schrödinger Equation. Now assume (M, g) is a non-compact, Riemannian manifold with or without a compact boundary and $M = \mathbb{R}^n$ (or finitely many copies) outside a compact set.

Let $r \simeq \operatorname{dist}_g(x, x_0)$ be a "radial" variable, and denote by $e^{it\Delta_g} u_0$ the solution to the initial value Schödinger problem:

$$\begin{cases} (i\partial_t + \Delta_g)u(x,t) = 0 \\ u(x,0) = u_0, \end{cases}$$

for u_0 in an appropriate Hilbert space. The operator $e^{it\Delta_g}$ is unitary from $H^s(M) \to H^s(M)$ for any fixed t, but if we also integrate in time we gain some regularity. More precisely, in $\mathbb{R}^n$ (or more general non-trapping, asymptotically Euclidean geometries), one proves the following "local smoothing estimate":

$$\int_0^T \left\| \langle r \rangle^{-1/2-\varepsilon} e^{it\Delta_g} u_0 \right\|^2_{H^{1/2}} dt \leqslant C \| u_0 \|^2_{L^2}.$$

Next suppose $\gamma \subset M$ is a closed hyperbolic geodesic with only transversal reflections with ∂M, and M is non-trapping otherwise (see Figures 6 and 7 for some examples).

Theorem 4 (Theorem 1 [5]). *Suppose that M, g, r, and γ satisfy the assumptions above. Then for any $\varepsilon > 0$ there exists $C > 0$ such that*

$$\int_0^T \left\| \langle r \rangle^{-1/2-\varepsilon} e^{it\Delta_g} u_0 \right\|^2_{H^{1/2-\varepsilon}} dt \leqslant C \| u_0 \|^2_{L^2}.$$

Remark 3. Local smoothing has been studied by many authors. In [**8**], Doi shows one has the sharp $H^{1/2}$ smoothing effect if and only if M is asymptotically Euclidean and non-trapping.

In [**1**], Burq proves Theorem 4 in the case of $\mathbb{R}^n$ with several convex bodies removed with some assumptions.

There are many applications coming from local smoothing type estimates, including Strichartz estimates and the study of nonlinear Schrödinger equations.

Remark 4. Theorem 4 has recently been extended in [**4**] to the case of a thin hyperbolic fractal trapped set using estimates of Nonnenmacher-Zworski in [**15**].

1.4. Wave Equation. We make the same assumptions here as in §1.3 with the addition that we assume $\dim M = n \geqslant 3$ is odd. We study solutions to the linear wave equation:

$$\begin{cases} (-D_t^2 - \Delta_g) u(x,t) = 0, \ (x,t) \in M \times [0,\infty) \\ u(x,0) = u_0 \in H^1(M), \\ D_t u(x,0) = u_1 \in L^2(M). \end{cases} \tag{1.2}$$

For $\psi \in \mathscr{C}_c^\infty(M)$, we define the local energy:

$$E(t) = \int_M \left|\psi \partial_t u\right|^2 + \left|\psi \nabla u\right|^2 dx.$$

Theorem 5 (Theorem 2 [5]). *For each $\psi \in \mathscr{C}_c^\infty(M)$, $\varepsilon > 0$, and each pair*

$$\begin{aligned} u_0 &\in \mathscr{C}_c^\infty(M) \cap H^1(M), \text{ and} \\ u_1 &\in \mathscr{C}_c^\infty(M) \cap L^2(M), \end{aligned}$$

there is a constant $C > 0$ such that

$$E(t) \leqslant Ce^{-t^{1/2}/C} \left(\| u_0 \|^2_{H^{1+\varepsilon}(M)} + \| u_1 \|^2_{H^\varepsilon(M)} \right).$$

Remark 5. We think of Theorem 5 as an analogue of the *Sharp Huygen's Principle.*

This generalizes similar results of Morawetz [**13**], Morawetz-Phillips [**9**], Morawetz-Ralston-Strauss [**14**], and Vodev [**19**], among many others.

There are immediate applications to the study of solutions to nonlinear wave equations (see, for example, [**12, 11**]).

Remark 6. Theorem 5 has now been generalized. In [**3**], we prove a weak polynomial rate of decay in the case of a thin fractal hyperbolic trapped set as studied in [**15**].

1.5. A Common Theme. All of the results discussed employ some common ideas. That is, each problem can be reduced via some trick to proving a high-energy resolvent estimate. In order to prove the high-energy resolvent estimates we need, we use a semiclassical rescaling, microlocal analysis near the periodic orbit, and microlocal "gluing" techniques to paste this together with known non-trapping estimates.

For the non-concentration result we make a semiclassical rescaling, $\lambda = \sqrt{z}/h$, and consider the equation

$$(-h^2\Delta - 1)u = E(h)u, \;\; E(h) = \mathcal{O}(h).$$

We prove a cutoff estimate:

$$\operatorname{supp}\chi \sim \gamma \implies \|\chi u\| \leqslant \frac{\sqrt{\log(1/h)}}{h}\| - h^2\Delta\chi u\|,$$

and a commutator argument finishes the proof.

In order to prove energy decay for the damped wave equation, *formally* we cut off in time, take the Fourier transform in t, and consider the equation

$$(-\tau^2 + i2\tau a(x) - \Delta)\hat{u} = f.$$

Rescaling $\tau = \sqrt{z}/h$ yields:

$$h^{-2}P(h) = h^{-2}(-h^2\Delta + i2\sqrt{z}ha(x) - z),$$

and an application of the resolvent estimates gives Theorem 3.

For the local smoothing estimate, we again formally take the Fourier transform in t:

$$P(\tau)\hat{u} = (-\tau - \Delta)\hat{u} = u_0,$$

and similarly for the local energy decay estimate. We then apply the resolvent estimates plus interpolation to prove Theorems 4 and 5.

1.6. Acknowledgements. This note is an expanded version of a talk given at the Courant-Colloquium "Göttingen trends in Mathematics", 12-14 October, 2007 at the Mathematisches Institut, Georg-August-Universität, Göttingen. The author gratefully acknowledges their hospitality and invitation to speak.

2. Main Results

2.1. General Problem. We assume M is a compact manifold, $P(h)$ (second-order), self-adjoint semiclassical pseudodifferential operator acting on half-densities. In the examples considered above, $P(h)$ is the semiclassical Laplace-Beltrami operator. That is, we can write the leading part of $P(h)$ in local coordinates as

$$P(h) = \frac{1}{2}\sum_{ij} hD_i a_{ij} hD_j + a_{ij} hD_i hD_j + \text{ lower order}$$

We assume the principal symbol p of $P(h)$ is real and independent of h. In our example,

$$p(x,\xi) = \sum_{ij} a_{ij}\xi_i\xi_j.$$

We assume

$$\begin{aligned} p = 0 &\implies dp \neq 0, \text{ and} \\ |\xi| \geqslant C &\implies |p| \geqslant \langle \xi \rangle^2 / C, \end{aligned}$$

and

$$\gamma \subset \{p = 0\} \text{ is a closed hyperbolic orbit for } \exp tH_p.$$

2.2. Main Result. Let $a \in \mathscr{C}^\infty(T^*X)$, $a \geqslant 0$, $a \equiv 0$ near γ, $a \equiv 1$ away from γ, and define a family of perturbations of $P(h)$:

$$Q(z,h) := P(h) - z - iha(x,hD),\ z \in [-\delta,\delta] + i[-c_0 h, c_0 h],$$

with $\delta, c_0 > 0$ fixed and independent of h. Our main estimate is given in the following Theorem.

Theorem 6 (Theorem 2 [7], Theorem 2 [6]). *Let M, $P(h)$, a, and $Q(z,h)$ be as above. Then there exist $C > 0$, $\delta > 0$, and $h_0 > 0$ such that*

$$\|Q(z,h)^{-1}\|_{L^2(X)\to L^2(X)} \leqslant C\frac{\log 1/h}{h},\ z \in [-\delta/2, \delta/2] \tag{2.1}$$

uniformly in $0 < h \leqslant h_0$.

Remark 7. The function a is added as an *absorption* term:

$$-\operatorname{Im} Q \geqslant c_1 h \quad \text{away from } \gamma,$$

so it suffices to prove (2.1) in a microlocal neighbourhood near γ. This will follow from the following estimate:

$$\|Q(z,h)^{-1}u\|_{L^2} \leqslant Ch^{-N}\|u\|_{L^2},\ z \in [-\delta,\delta] + i[-c_0 h, c_0 h] \tag{2.2}$$

if u is concentrated[(1)] near γ, and a semiclassical adaptation of the "three-line theorem" from complex analysis.

The rest of this note will be to present the main ideas from the proof of (2.2).

---

(1) We say u is concentrated near γ if $\exists \chi \in \mathscr{C}_c^\infty(T^*X)$, $\operatorname{supp}\chi \sim \gamma$, $\chi(x,hD)u = u + \mathscr{O}(h^\infty)$.

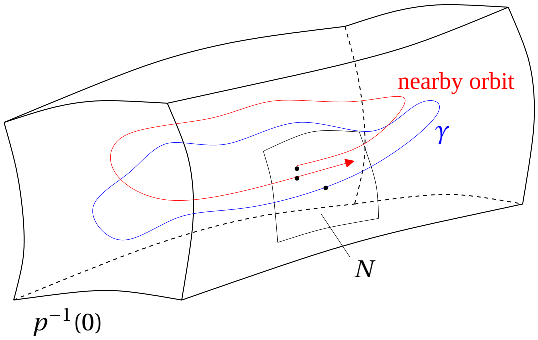


FIGURE 8. Poincaré section N and Poincaré map

3. Sketch of the Proof of (2.2)

The first idea we use in the proof is to reduce the study of $P(h)$ near γ to the study of a related operator on a lower dimensional space. The main advantage in this approach is that a neighbourhood of γ is not necessarily simply connected, so one would at the very least have to study $P(h)$ in a double-cover space - something we would like to avoid, while the lower dimensional space will be simply connected.

3.1. Poincaré Map. With $\gamma \subset \{p = 0\}$ we can define a Poincaré map and Poincaré section for γ. The Poincaré section $N \subset \{p = 0\}$ is a codimension 1 submanifold which is transversal to the flow (that is, if $\dim M = n$, we have $\dim T^*M = 2n$, $\dim\{p = 0\} = 2n - 1$, and $\dim N = 2n - 2$). We can define the Poincaré or *first return* map $S : N \to S(N)$ in the usual fashion (see Figure 8 for the basic picture in $n = 2$). It is well-known that S is symplectic and has a fixed point where γ intersects N. We implicitly identify N with a neighbourhood of $(0,0) \in \mathbb{R}^{2n-2}$, and write $S(0,0) = (0,0)$ for simplicity.

We will continue to sketch the proof and draw the figures in dimension 2 to fix the ideas, although the proof extends to any dimension. The assumption that γ is hyperbolic means $dS(0,0)$ is hyperbolic. That is, after a linear symplectic change of variables,

$$\begin{aligned} dS(0,0) &= \begin{pmatrix} \mu & 0 \\ 0 & 1/\mu \end{pmatrix}, \ \mu > 1 \\ &= \exp\begin{pmatrix} \lambda & 0 \\ 0 & -\lambda \end{pmatrix}, \ \lambda > 0. \end{aligned} \tag{3.1}$$

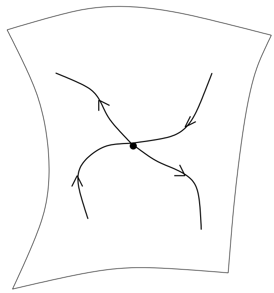


FIGURE 9. The unstable/stable manifolds $\Lambda_\pm$

We remark that in this example we have, for simplicity, assumed $\mu > 1$. The case $\mu < -1$ is also possible for a hyperbolic symplectic map, but presents nontrivial technical issues (this, along with the case where *some* of the eigenvalues lie on the unit circle, is handled in [**6**]).

With $dS(0,0)$ satisfying (3.1), we conclude

$$S = \exp H_q, \; q = \lambda x\xi + \mathscr{O}((x,\xi)^3). \tag{3.2}$$

We need also the notion of unstable/stable manifold. For S symplectic satisfying (3.1) there are two invariant (with respect to the action of S) Lagrangian submanifolds $\Lambda_\pm$ of $\mathbb{R}^{2n-2}$ on which S is expanding/contracting respectively (see Figure 9).

It remains to understand the quantization of S, control the error in (3.2), and relate estimates on $P(h)$ to estimates on the quantization of S.

3.2. Quantization of S. We quantize S as a microlocally defined h-Fourier Integral Operator $M : L^2(\mathbb{R}^{n-1}) \to L^2(\mathbb{R}^{n-1})$. The conjugation of a pseudodifferential operator $A = \mathrm{Op}\,(a)$ by M is

$$M^{-1}AM = \mathrm{Op}\,(S^* a) + \mathscr{O}(h^2).$$

Since S varies with energy level z, so does M, and we write $M = M(z)$. The following Theorem relates $P(h) - z$ in a microlocal neighbourhood of $\gamma \subset T^*M$ to $M(z)$ in a microlocal neighbourhood of $(0,0) \in T^*\mathbb{R}^{n-1}$.

Theorem 7 (Theorem 3 [6]). *With $P(h)$ and $M(z)$ as above, we have for u concentrated near γ:*

$$\|(P(h)-z)u\|_{L^2(M)} \geqslant C^{-1}h\|(I - M(z))Ru\|_{L^2(\mathbb{R}^{n-1})},$$

where Ru is the restriction of u to the projection of the Poincaré section. Further,

$$h\|u\|_{L^2(M)} \leqslant Ch\|Ru\|_{L^2(\mathbb{R}^{n-1})} + C\|(P(h)-z)u\|_{L^2(M)}.$$

Thus to prove (2.2), we need to show

$$\|(I - M(z))Ru\| \geqslant h^N \|Ru\| \text{ microlocally.}$$

Since this is just a sketch of the proof, now we cheat and write

$$M(z) = \exp\left(\frac{-i}{h}\operatorname{Op}(q-z)\right),$$

where q is defined in (3.2). This is close enough to being true to present the ideas of the rest of the proof. For G real-valued, to be determined, we write

$$e^{-G(x,hD_x)/h} M e^{G(x,hD_x)/h} = \exp\left(\frac{-i}{h} e^{-G/h}\operatorname{Op}(q-z)e^{G/h}\right).$$

Now

$$e^{-G/h}\operatorname{Op}(q)e^{G/h} = \operatorname{Op}(q) - ih\operatorname{Op}(H_q G) + \text{ error}$$

where now *error* is hard to estimate but controllable.

The rough idea now is a positive commutator argument: q is real-valued, but conjugating $Q = \operatorname{Op}(q)$ with $B = \operatorname{Op}(b)$ invertible gives

$$BQB^{-1} = Q - [Q,B]B^{-1}.$$

Q is self-adjoint, so $\exp(-i(Q-z)/h)$ is unitary. However if B is self-adjoint, the commutator $[Q,B]$ is skew-adjoint, so $\exp(-i[Q,B]/h)$ is not unitary. If we can arrange B in such a fashion that $[Q,B]$ has a definite sign, then $\exp(-i[Q,B]/h)$ is expanding or contracting and has spectrum away from 1. Hence we look at the commutator $[Q,B]$.

Our next guiding principle is that since S is hyperbolic, q roughly satisfies

$$H_q(\operatorname{dist}^2(\cdot,\Lambda_\mp)) \sim \pm\operatorname{dist}^2(\cdot,\Lambda_\mp).$$

It is a nontrivial fact that there is a symplectic choice of coordinates so that (see Figure 10)

$$\operatorname{dist}^2(\cdot,\Lambda_\mp) = \begin{cases} x^2 \\ \xi^2. \end{cases}$$

In these coordinates we would have

$$H_q(x^2 - \xi^2) = x^2 + \xi^2,$$

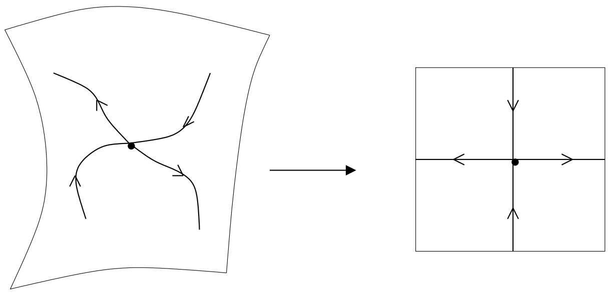


FIGURE 10. $\Lambda_\pm$ and the change of variables

which is the harmonic oscillator and hence is bounded below by h when quantized. Of course this doesn't quite work as we shall see. We also remark that it is important that all our changes of variables are symplectic, since a local symplectic transformation in phase space corresponds to a microlocally unitary h-FIO. Hence symplectic transformations in phase space preserve the spectral properties of the associated quantized operators.

After performing this change of variables, we show that in these coordinates, q takes the special form

$$q(x,\xi) = \lambda(x,\xi)x\xi.$$

That is, in any Taylor development of q near $(0,0)$, each term has $x\xi$ in it. Then

$$H_q(\mathrm{dist}^2(\cdot,\Lambda_-) - \mathrm{dist}^2(\cdot,\Lambda_+)) \sim \lambda(x,\xi)(x^2+\xi^2) + \text{ error},$$

where now *error* can be controlled.

From the remarks above, we then want to take

$$G = \mathrm{dist}^2(\cdot,\Lambda_-) - \mathrm{dist}^2(\cdot,\Lambda_+) = x^2 - \xi^2,$$

so that $H_q G \sim x^2 + \xi^2$ is the harmonic oscillator. However, the growth of G means we cannot take the exponential of G^w, so we try

$$G = \frac{h}{2}\log\left(\frac{h+x^2}{h+\xi^2}\right).$$

The growth of this G is acceptable, however it is in a bad calculus - that is, we lose $h^{-1/2}$ with each derivative.

The next idea is to use a special calculus with two parameters. That is, we rescale with an additional new parameter $\tilde{h} \geqslant h$:

$$X = \frac{\tilde{h}^{\frac{1}{2}}}{h^{\frac{1}{2}}} x, \ \Xi = \frac{\tilde{h}^{\frac{1}{2}}}{h^{\frac{1}{2}}} \xi,$$

and quantize with respect to $\tilde{h}$. If we prove estimates which are uniform in both $\tilde{h}$ and h, then we can freeze $\tilde{h}$ and conclude the estimates hold in $h \to 0$. We define G in these new coordinates:

$$G = \log\left(\frac{1+X^2}{1+\Xi^2}\right).$$

With this rescaling and choice of G,

$$H_q G \sim h\lambda\left(\frac{X^2}{1+X^2} + \frac{\Xi^2}{1+\Xi^2}\right),$$

which satisfies the same lower bound as the harmonic oscillator.

In the $\tilde{h}$ calculus,

$$\mathrm{Op}\,(H_q G) \geqslant \frac{h\tilde{h}}{C}, \ \tilde{h} > 0 \text{ fixed,}$$

and hence

$$-\operatorname{Im} e^{-G^w}\,\mathrm{Op}\,(q) e^{G^w} \geqslant \frac{h}{C}.$$

Thus

$$\begin{aligned} & \|e^{-G^w} M e^{G^w}\| \leqslant e^{-1/C} < 1 \\ \Longrightarrow \qquad & \|(I - e^{-G^w} M e^{G^w})\| \geqslant \frac{1}{C}. \end{aligned}$$

Rescaling back to h calculus we lose something because the operators are only h-tempered. However, this yields the desired estimate:

$$\|(I - M(z))Ru\| \geqslant h^N \|Ru\|$$

microlocally. Applying Theorem 7 gives (2.2).

References

[1] N. Burq – "Smoothing effect for Schrödinger boundary value problems", *Duke Math. J.* **123** (2004), no. 2, p. 403–427.

[2] N. Burq & M. Zworski – "Geometric control in the presence of a black box", *J. Amer. Math. Soc.* **17** (2004), no. 2, p. 443–471 (electronic).

[3] H. CHRISTIANSON – "Applications of cutoff resolvent estimates to the wave equation.", preprint.

[4] ———, "Cutoff resolvent estimates and the semilinear schrodinger equation", to appear in *Proc. Amer. Math. Soc.*

[5] ———, "Dispersive estimates for manifolds with one trapped orbit", to appear in *Commun. PDE.*

[6] ———, "Quantum monodromy and non-concentration near semi-hyperbolic orbits", Preprint.

[7] ———, "Semiclassical non-concentration near hyperbolic orbits", *J. Funct. Anal.* **246** (2007), no. 2, p. 145–195.

[8] S.-I. DOI – "Smoothing effects of Schrödinger evolution groups on Riemannian manifolds", *Duke Math. J.* **82** (1996), no. 3, p. 679–706.

[9] P. D. LAX, C. S. MORAWETZ & R. S. PHILLIPS – "The exponential decay of solutions of the wave equation in the exterior of a star-shaped obstacle", *Bull. Amer. Math. Soc.* **68** (1962), p. 593–595.

[10] G. LEBEAU – "Équation des ondes amorties", Algebraic and geometric methods in mathematical physics (Kaciveli, 1993), Math. Phys. Stud., vol. 19, Kluwer Acad. Publ., Dordrecht, 1996, p. 73–109.

[11] J. METCALFE & C. D. SOGGE – "Hyperbolic trapped rays and global existence of quasilinear wave equations", *Invent. Math.* **159** (2005), no. 1, p. 75–117.

[12] ———, "Global existence for Dirichlet-wave equations with quadratic nonlinearities in high dimensions", *Math. Ann.* **336** (2006), no. 2, p. 391–420.

[13] C. S. MORAWETZ – "The decay of solutions of the exterior initial-boundary value problem for the wave equation", *Comm. Pure Appl. Math.* **14** (1961), p. 561–568.

[14] C. S. MORAWETZ, J. V. RALSTON & W. A. STRAUSS – "Decay of solutions of the wave equation outside nontrapping obstacles", *Comm. Pure Appl. Math.* **30** (1977), no. 4, p. 447–508.

[15] S. NONNENMACHER & M. ZWORSKI – "Quantum decay rates in chaotic scattering", Seminaire: Equations aux Dérivées Partielles. 2005–2006, Sémin. Équ. Dériv. Partielles, École Polytech., Palaiseau, 2006, p. Exp. No. XXII, 8.

[16] J. RAUCH & M. TAYLOR – "Decay of solutions to nondissipative hyperbolic systems on compact manifolds", *Comm. Pure Appl. Math.* **28** (1975), no. 4, p. 501–523.

[17] Y. COLIN DE VERDIÈRE – "Ergodicité et fonctions propres du Laplacien", *Comm. Math. Phys.* **102** (1985), no. 3, p. 497–502.

[18] Y. COLIN DE VERDIÈRE & B. PARISSE – "Équilibre instable en régime semi-classique", Séminaire sur les Équations aux Dérivées Partielles, 1993–1994, École Polytech., Palaiseau, 1994, p. Exp. No. VI, 11.

[19] G. VODEV – "Local energy decay of solutions to the wave equation for nontrapping metrics", *Ark. Mat.* **42** (2004), no. 2, p. 379–397.

[20] S. ZELDITCH – "Uniform distribution of eigenfunctions on compact hyperbolic surfaces", *Duke Math. J.* **55** (1987), no. 4, p. 919–941.

Mathematisches Institut, Courant-Colloquium *Trends in Mathematics*
(R. Meyer, T. Schick, Y. Tschinkel, eds.), p. 47–56
Georg-August Universität Göttingen, 2008

FORMS IN MANY VARIABLES

Rainer Dietmann

Institut für Algebra und Zahlentheorie, Pfaffenwaldring 57, D-70569 Stuttgart, Germany
E-mail : dietmarr@mathematik.uni-stuttgart.de

Abstract. We give a survey about two new applications of the Hardy-Littlewood circle method from analytic number theory to arithmetic problems. These problems are finding a polynomial bound for the smallest integer solution of a quadratic Diophantine equation, and showing that the intersection of a system of rational cubic forms in sufficiently many variables admits a rational linear space of zeros of given dimension.

1. Introduction

One of the basic questions in number theory concerns the solubility of Diophantine equations: Let $P(X_1, \ldots, X_s) \in \mathbf{Z}[X_1, \ldots, X_s]$ be a polynomial with integer coefficients. Does the equation $P(\mathbf{x}) = 0$ have a solution in $\mathbf{Z}^s$? In this generality, this problem is hopelessly difficult. Indeed, answering Hilbert's tenth problem, Matijasevič ([**10**]) showed that there is no algorithm deciding this question for general P. For special classes of Diophantine equations, however, there are effective methods for deciding solvability or even finding a solution if there is one, and we will discuss two approaches in this direction. The first approach stems from the so-called *Hasse* (*Local-Global*) principle: Obviously, if $P(\mathbf{x}) = 0$ is soluble in integers, then also in reals and also all congruences $P(\mathbf{x}) \equiv 0 \pmod{m}$ have an integer solution, which is equivalent to $P(\mathbf{x}) = 0$ having solutions in all local rings $\mathbf{Z}_p$ of integral p-adic numbers, for all rational primes p. If these obvious necessary conditions

2000 ***Mathematics Subject Classification.*** 11D25, 11D72, 11D88, 11E76.

are also sufficient for P having an integer zero $\mathbf{x} \in \mathbf{Z}^s$, then one says that P satisfies the Hasse principle. For example, Minkowski showed the Hasse principle to hold true for quadratic forms. Since for a homogeneous form there is always the trivial zero, one has to modify the Hasse principle accordingly, either by removing the trivial zero or by rephrasing the conditions so that projective zeros are required. The Hasse principle for quadratic forms then is the following well known result: If $Q(X_1,\dots,X_s) \in \mathbf{Z}[X_1,\dots,X_s]$ is a quadratic form, then there is a zero $\mathbf{x} \in \mathbf{Z}^s \backslash \{\mathbf{0}\}$ of $Q(\mathbf{x}) = 0$ if and only if there is a zero $\mathbf{x} \in \mathbf{R}^s \backslash \{\mathbf{0}\}$ and if there are zeros $\mathbf{x} \in \mathbf{Z}_p^s \backslash \{\mathbf{0}\}$ for all p. Note that the real condition on Q just states that Q is not definite. Now for many polynomials it is quite easy to check the real and the local conditions (for example, by lifting solutions modulo p, where often only finitely many p have to be considered), so the Hasse principle then in particular gives an effective method for deciding if there is a solution or not. Unfortunately, the Hasse principle fails for many polynomials. One of the first counterexamples is due to Selmer ([**12**]), who has shown that the equation

$$3X3 + 4Y3 + 5Z3 = 0$$

has non-trivial real zeros, non-trivial p-adic zeros, but no non-trivial integer zero. In fact, this equation describes an algebraic curve of genus one, which is called an *elliptic curve*. Whereas the Hasse principle holds for conics (see above), which are algebraic curves of genus zero, it generally fails for higher genus curves, in particular for elliptic curves. We now describe a second approach to effectivity for Diophantine equations, which does not necessarily rely on the Hasse principle. Let H (the *'height'*) be the maximum of the moduli of the coefficients of the polynomial P. Then for a class of polynomials P in s variables with height H an effective function $\Lambda_s(H)$ is called a *search bound*, if the following holds true: If there is any integer solution $\mathbf{x} \in \mathbf{Z}^s$ of $P(\mathbf{x}) = 0$, then there is one with

$$|\mathbf{x}| \le \Lambda_s(H),$$

where $|\cdot|$ denotes the maximum norm. Clearly a search bound gives an algorithm not only for deciding if there is an integer solution, but also for finding one. Siegel ([**13**]) established that for quadratic (not necessarily homogeneous) P in any number of variables there exists a search bound $\Lambda_s(H)$. Making his argument effective, one finds that one may take

$$\Lambda_s(H) = \exp((2H)^{C_1(s)}),$$

where $C_1(s)$ is an effective constant. In the binary case $s = 2$, using Pell's equation one can show that this bound is essentially sharp and in particular cannot be replaced by a polynomial in H ([**8**]). For $s \geq 5$ and non-singular quadratic part of Q, later Kornhauser ([**7**]) obtained polynomial bounds. The remaining cases $s = 3$ and $s = 4$ and also a considerable improvement of Kornhauser's bounds have been established by the author in [**3**], Theorem 1:

Theorem 1. *Suppose that the quadratic part of Q is non-singular. Then one can take*

$$\Lambda_s(H) = \begin{cases} C_2 H^{2100} & s = 3 \\ C_3 H^{84} & s = 4 \\ C_4(s) H^{5s+100} & s \geq 5 \end{cases}$$

for effective constants C_2, C_3 *and* $C_4(s)$.

So from a qualitative point of view the question about the order of magnitude of search bounds for quadratic Diophantine equations has been settled: For $s = 2$ one needs an exponential bound, whereas for $s \geq 3$ polynomial bounds are possible.

2. Idea of proof

Let us now give a short sketch of the proof of Theorem 1. For simplicity, let us assume that there are no linear terms, so P is of the form

$$P(X_1, \ldots, X_s) = Q(X_1, \ldots, X_s) - n,$$

where Q is a non-singular integral quadratic form and n is an integer. Only the case of indefinite Q is interesting, since for definite Q it is easy to obtain good bounds even for all *real* solutions of $Q(\mathbf{x}) = n$.

Case I: $s \geq 5$ or $s = 4$ and $n \neq 0$. In this case we can apply the *Hardy-Littlewood circle method* from analytic number theory. This is an analytic method for counting solutions of Diophantine equations in some expanding box. Let

$$r(N) := \#\{\mathbf{x} \in \mathbf{Z}^s : |\mathbf{x}| \leq N \text{ and } P(\mathbf{x}) = 0\},$$

where N is a parameter. Now

$$\int_0 1 e(\alpha m)\, d\alpha = \begin{cases} 1 & m = 0 \\ 0 & m \neq 0, \end{cases}$$

where $e(\alpha)$ is the character $e(\alpha) = \exp(2\pi i\alpha)$. Hence, introducing the *exponential sum*

$$f(\alpha) = \sum_{\mathbf{x}\in\mathbf{Z}^s:|\mathbf{x}|\leq N} e(\alpha P(\mathbf{x})),$$

by interchanging the order of integration and summation we find that

$$r(N) = \int_0 1 f(\alpha)\, d\alpha.$$

Now if α is 'close' to a rational point $\frac{a}{q}$ with 'small' denominator q, then one has a very good approximation to $f(\alpha)$ involving so-called *Gauss sums*, because for $\alpha = \frac{a}{q}$ we just have to sort the $\mathbf{x}$ in residue classes modulo q, and for α close to such $\frac{a}{q}$ by partial summation still a good approximation by Gaussian sums holds true. Making this observation precise is the essence of the circle method, of which we used a recent modern version due to Heath-Brown ([**6**]). This way we get an asymptotic formula

$$r(N) = \mathfrak{S}\mathfrak{J}N^{s-2} + O_\varepsilon(N^{s/2-\delta+\varepsilon})$$

for any $\varepsilon > 0$ and $\delta = \frac{1}{2}$ for $s = 4$ and $n \neq 0$ (here a so-called *Kloosterman refinement* comes in) and $\delta = 0$ otherwise. Here $\mathfrak{S}$ is the so-called *singular series*, a measure for the density of p-adic solutions of the equation $P(\mathbf{x}) = 0$, and $\mathfrak{J}$ is the so-called *singular integral*, a measure for the density of real solutions of the equation $P(\mathbf{x}) = 0$. Since the asymptotic formula for $r(N)$ has a main term N^{s-2} exceeding the error term $N^{s/2-\delta+\varepsilon}$, we obtain a quantitative form of the Local-Global principle: If there are local (including real) solutions of $P(\mathbf{x}) = 0$, then $\mathfrak{S} > 0$ and $\mathfrak{J} > 0$, so for sufficiently large N we have $r(N) > 0$. Note that the asymptotic main term N^{s-2} is of the order of magnitude to be expected: We have order of magnitude N^s vectors under consideration, and the probability of being a zero of the quadratic equation $P(\mathbf{x}) = 0$ is N^{-2}. Since we want to establish a search bound, we may assume that there is an integer zero $\mathbf{x}$ of $P(\mathbf{x}) = 0$, hence there exist also real and p-adic zeros. One then can establish lower bounds for $\mathfrak{S}$ and $\mathfrak{J}$ which are explicit in the height H, and one can also give such explicit upper bounds for the implied O-constant. Since all the occurring terms are polynomial in H, one finds a polynomial bound on N in terms of H such that $r(N) \geq 1$ as soon as N exceeds this bound. This is exactly what we want, a polynomial search bound!

The circle method approach described above is also useful for another problem on search bounds for positive definite integral quadratic forms Q: Suppose that for an integer n the congruences

$$Q(x_1,\dots,x_s) \equiv n \pmod{m}$$

are soluble for all positive integers m. Then according to the philosophy of the Local-Global-principle, one should expect also an integer solution $\mathbf{x} \in \mathbf{Z}^s$ of the equation $Q(\mathbf{x}) = n$ to exist. For small n this may fail, but for 'sufficiently large' n and $s \geq 5$ it is true. Making 'sufficiently large' explicit is known as 'Tartakovski's problem'. Improving earlier results, in a joint paper with T.D. Browning ([**1**], Theorem 4) we could prove that for $s \geq 5$ the lower bound

$$n \gg_\varepsilon \left(|\det Q|^{\frac{s-2}{s-4}} ||Q||^{s+\varepsilon}\right)^{\frac{2}{s-3}}$$

is sufficient, where $||Q||$ is the height of Q. So in principle one has a method for finding *all* positive integers n which are represented integrally by Q: For large n, one has to check the congruence conditions, which is a finite problem, and for the remaining *finitely* many small n one could check by computer.

Case II: $s \in \{3,4\}$ and $n = 0$. In this case one deals with a homogeneous problem, for which lattice point methods from the geometry of numbers are available ([**2**]).

Case III: $s = 3$ and $n \neq 0$. This is the most difficult case, since the circle method breaks down here. By using reduction theory for quadratic forms, one can reduce the original problem of finding search bounds for quadratic Diophantine equations to the following more general and seemingly more difficult problem on finding search bounds for integral equivalence of integral ternary quadratic forms ([**3**], Theorem 4):

Theorem 2. *Let A and B be integral symmetric non-singular 3×3-matrices which are integral equivalent, so there is an unimodular integral 3×3-matrix R with*

$$B = R^T AR. \tag{1}$$

Then there is such R with

$$||R|| \ll (||A|| + ||B||)^{800}.$$

Here the height $||\cdot||$ is defined as maximum norm in the naive way. So in particular one has a means of deciding effectively if two integral ternary quadratic forms are equivalent or not. Problems of this kind have recently also found interest in cryptography ([**5**]).

Let us now give a short sketch of the proof of Theorem 2. First, it is much easier to deal with this problem first over the rationals. So we get a 'small' rational 3×3-matrix S with

$$B = S^T AS. \tag{2}$$

'Small' in the following means bounded by a polynomial in the heights of A and B, which is sufficient for our purposes, and the height of a rational matrix is again defined in the naive way as maximum norm for the nominators and the denominator, after clearing fractions. Then comparing (1) and (2) one finds that for

$$U = RS^{-1}$$

we have

$$U^T AU = A,$$

so U is a rational *automorph* of A. Fortunately, there is a parameterization of all rational automorphs of a ternary quadratic form going back to Hermite. So U must be of the form $U = U_{\mathbf{z}}$ for a parameter $\mathbf{z} \in \mathbf{Z}4$, where $U_{\mathbf{z}}$ is of the form

$$U_{\mathbf{z}} = \frac{1}{T(\mathbf{z})} \begin{pmatrix} T_1(\mathbf{z}) & \cdots & T_3(\mathbf{z}) \\ \vdots & & \vdots \\ T_7(\mathbf{z}) & \cdots & T_9(\mathbf{z}) \end{pmatrix}$$

for suitable integral non-singular quadratic forms $T, T_1, \ldots, T_9 \in \mathbf{Z}[X_1, \ldots, X_4]$ depending on A. The important thing now is that the T_i and T are *quaternary* quadratic forms with small height. Moreover, the denominator of U must be small because of $U = RS^{-1}$ with integral R and small S. An easy application of Hilbert's Nullstellensatz shows that there cannot be too much cancellation in $U_{\mathbf{z}}$, thus $T(\mathbf{z})$ must also be small. So we can use the circle method from Case I to find a small $\mathbf{z}' \in \mathbf{Z}4$ with

$$T(\mathbf{z}') = T(\mathbf{z})$$

and

$$\mathbf{z}' \equiv \mathbf{z} \pmod{(T(\mathbf{z}) \cdot \text{denominator } S)}.$$

Note that the additional congruence condition is a technicality making no serious problems for the circle method. Now let

$$U' := U_{\mathbf{z}'}.$$

Then clearly U' is small because $\mathbf{z}'$ and the T_i and T are. Moreover,

$$U'^T AU' = A,$$

because U' comes from the parameterization of all automorphs. Furthermore,

$$R' := U'S$$

is integral, because $R = US = U_{\mathbf{z}}S$ is and the congruence condition on $\mathbf{z}'$ ensured that all denominators get canceled, like for R. Finally,

$$R'^T AR' = S^T U'^T AU'S = S^T AS = B,$$

so R' has the desired transformation property, is integral and clearly small since U' and S are. This completes the proof of Theorem 2 and so also the proof of Theorem 1.

3. An application of the circle method to systems of cubic forms

We complete our exposition by describing another recent application of the circle method to a problem on higher degree forms. Let $C_1, \dots, C_r \in \mathbf{Z}[X_1, \dots, X_s]$ be cubic forms. Since for odd degree forms in sufficiently many variables there are always non-trivial real zeros, and since for systems of forms in sufficiently many variables there are always non-trivial p-adic zeros, one should expect according to the Local-Global-principle that any such system of r integral cubic forms has a non-trivial rational zero, providing that s is sufficiently large in terms of r. The best such bound currently known is due to Schmidt ([**11**]), using the circle method, who showed that $s \geq (10r)5$ is sufficient. It is a natural question to seek a generalization of this result to finding not only a common non-trivial rational zero to a system of rational cubic forms, but even a rational linear space of given dimension on which all those forms vanish. In this direction we could prove the following result ([**4**], Theorem 2):

Theorem 3. *Let $r, m \in \mathbf{N}$, and let $C_1, \dots, C_r \in \mathbf{Z}[X_1, \dots, X_s]$ be cubic forms where $s \gg r4m6 + r6m5$. Then the system of equations $C_1 = \dots = C_r = 0$ admits an m-dimensional rational linear space of zeros.*

For $r^{1/9} \ll m \ll r^{5/4}$ this improves earlier results due to Lewis/Schulze-Pillot ([**9**]) and Wooley ([**14**]). As mentioned above, the proof uses the circle method. Let us now give a short outline of the main arguments. We start by writing

$$\mathbf{y} = \begin{pmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_m \end{pmatrix}.$$

Now a simple Taylor-expansion argument shows that there are forms $\tilde{C}_1,\dots,\tilde{C}_R \in \mathbf{Z}[X_1,\dots,X_{sm}]$ with R of order of magnitude $rm3$ and the property that $\tilde{C}_i(\mathbf{y}) = 0$ $(1 \le i \le R)$ if and only if $C_1,\dots,C_r$ vanish on the rational linear space spanned by $\mathbf{x}_1,\dots,\mathbf{x}_m$. So we encoded the subspace property in a problem on finding rational zeros to a new system of cubic forms, for which we apply the circle method. So let

$$r(N) := \#\{\mathbf{y} \in \mathbf{Z}^{sm} : |\mathbf{y}| \le N \text{ and } \tilde{C}_i(\mathbf{y}) = 0\ (1 \le i \le R)\}.$$

Like in the quadratic case where we stipulated non-singularity, we need some geometric condition on the system of cubic forms in order to be able to bound the corresponding exponential sums occurring in the circle method approach. It turns out that we have to stipulate a *pencil condition* to the effect that no form of the rational pencil of $\tilde{C}_1,\dots,\tilde{C}_R$ vanishes on a rational linear space of too large dimension. If this pencil condition is satisfied, then the circle method goes through, yielding

$$r(N) = \mathfrak{S}\mathfrak{J}N^{sm-3R} + o(N^{sm-3R}),$$

where the singular integral $\mathfrak{J}$ is positive, because there are non-trivial real zeros since the degree is odd. Also the singular series $\mathfrak{S}$ is positive, since the number of variables is large enough and thus there are non-trivial p-adic zeros. In fact, using information on p-adic linear spaces on the original system $C_1,\dots,C_r$ rather than on p-adic points on $\tilde{C}_1,\dots,\tilde{C}_R$, we can get a quite good lower bound for $\mathfrak{S}$. Note that the main term N^{sm-3R} is of the order of magnitude to be expected heuristically. So there are many integer zeros $\mathbf{y}$ of the system $\tilde{C}_1 = \dots = \tilde{C}_R = 0$, and 'almost all' of them give rise to linearly independent $\mathbf{x}_1,\dots,\mathbf{x}_m$ since such linearly dependent $\mathbf{x}_i$ lie on a algebraic variety of small dimension and are much less than the points counted by $r(N)$. So if the pencil condition is satisfied, we are done. If the pencil condition is not satisfied, a form in the rational pencil of $\tilde{C}_1,\dots,\tilde{C}_R$ vanishes on a rational linear space of large dimension. Then one can show that the same is true for the original system of forms $C_1,\dots,C_r$. So without loss of generality C_1 vanishes on a rational linear space of high dimension. By parameterizing this space and substituting in the remaining forms $C_2,\dots,C_r$, we may proceed by induction on the number r of forms and again are done.

References

[1] T. BROWNING & R. DIETMANN – "On the representation of integers by quadratic forms.", (2008).

[2] J. CASSELS – "Bounds for the least solutions of homogeneous quadratic equations.", *Proc. Camb. Philos. Soc.* **51** (1954), p. 262–264.

[3] R. DIETMANN – "Small solutions of quadratic Diophantine equations.", *Proc. Lond. Math. Soc., III. Ser.* **86** (2003), no. 3, p. 545–582.

[4] ———, "Systems of cubic forms.", *J. Lond. Math. Soc., II. Ser.* **77** (2008), no. 3, p. 666–686.

[5] R. J. HARTUNG & C.-P. SCHNORR – *Public Key Identification Based on the Equivalence of Quadratic Forms*, Lecture Notes in Computer Science, Springer Berlin / Heidelberg, 2007.

[6] D. HEATH-BROWN – "A new form of the circle method, and its application to quadratic forms.", *J. Reine Angew. Math.* **481** (1996), p. 149–206.

[7] D. M. KORNHAUSER – "On small solutions of the general nonsingular quadratic diophantine equation in five or more unknowns.", *Math. Proc. Camb. Philos. Soc.* **107** (1990), no. 2, p. 197–211.

[8] ———, "On the smallest solution to the general binary quadratic diophantine equation.", *Acta Arith.* **55** (1990), no. 1, p. 83–94.

[9] D. LEWIS & R. SCHULZE-PILLOT – "Linear spaces on the intersection of cubic hypersurfaces.", *Monatsh. Math.* **97** (1984), p. 277–285.

[10] J. V. MATIJASEVIČ – "The Diophantineness of enumerable sets", *Dokl. Akad. Nauk SSSR* **191** (1970), p. 279–282.

[11] W. M. SCHMIDT – "On cubic polynomials. IV. Systems of rational equations", *Monatsh. Math.* **93** (1982), no. 4, p. 329–348.

[12] E. S. SELMER – "The Diophantine equation $ax^3 + by^3 + cz^3 = 0$", *Acta Math.* **85** (1951), p. 203–362 (1 plate).

[13] C. L. SIEGEL – "Zur Theorie der quadratischen Formen", *Nachr. Akad. Wiss. Göttingen Math.-Phys. Kl. II* (1972), p. 21–46.

[14] T. D. WOOLEY – "Forms in many variables", Analytic number theory (Kyoto, 1996), London Math. Soc. Lecture Note Ser., vol. 247, Cambridge Univ. Press, Cambridge, 1997, p. 361–376.

Mathematisches Institut, Courant-Colloquium Trends in Mathematics
(R. Meyer, T. Schick, Y. Tschinkel, eds.), p. 57–68
Georg-August Universität Göttingen, 2008

ON THE TOPOLOGY OF THE MODULI STACK OF STABLE CURVES

Johannes Ebert
Mathematical Institute, 24-29 St. Giles Street, Oxford OX1 3LB, United Kingdom
E-mail : ebert@maths.ox.ac.uk

Abstract. This note is an informal report on the joint paper **[4]** of the author with Jeffrey Giansiracusa, which grew out of the attempt to understand the topology of the moduli stack of stable curves. The main result is the construction of a map from the moduli stack to a certain infinite loop space, which is surjective on homology in a certain range. This shows the existence of many torsion classes in the homology of $\overline{\mathfrak{M}}_{g,n}$. We give a geometric description of some of the new torsion classes. Also, we give a new proof of an (old) theorem computing the second homology of the moduli stack.

The moduli space $\overline{\mathfrak{M}}_{g,n}$ of stable n-pointed curves of genus g is a compactification of the moduli space $\mathfrak{M}_{g,n}$ of smooth n-pointed curves. One adds a boundary $\partial\mathfrak{M}_{g,n}$ which contains singular curves of a certain type, namely *stable* ones. A singular curve C with n marked distinct smooth points $p_1,\ldots,p_n$ is called stable if all singularities are ordinary double points and if there is only a finite number of automorphisms of C which fix the p_i. Strictly speaking, due to the presence of automorphisms, one must study $\overline{\mathfrak{M}}_{g,n}$ as a *stack* and not as a space.
There is a coarse moduli space $\overline{\mathfrak{M}}_{g,n}^{coarse}$, which is the topological space usually referred to as the moduli space. There are two things to say about this coarse moduli space. First of all, the rational homology $H_*(\overline{\mathfrak{M}}_{g,n}^{coarse};\mathbb{Q})$ is isomorphic to the rational homology of the stack $\overline{\mathfrak{M}}_{g,n}$ (a concept explained below). Also, $\overline{\mathfrak{M}}_{g,n}^{coarse}$ it

2000 ***Mathematics Subject Classification.*** 32G15, 14H15, 22A22, 55R40.

The author is supported by a fellowship within the Postdoc-Programme of the German Academic Exchange Service (DAAD).

is a projective variety of complex dimension $3g-3+n$ and its singularities are of a very mild type (quotients of domains in a complex vector space by a finite group action).
It follows that $\overline{\mathfrak{M}}_{g,n}^{coarse}$ is a rational homology manifold, in other words, Poincaré duality with rational coefficients holds. However, if one wants to study topological invariants finer than rational homology, one is forced to consider the stack $\overline{\mathfrak{M}}_{g,n}$. For example, the integral homology of the coarse moduli space is not well-behaved at all.

1. A few words on stacks

Let us say a few words about stacks and how they can be studied by methods of algebraic topology. We will mainly consider stacks in the category of complex manifolds. As an excellent first introduction into the subject we recommend [**8**]; he only treats differentiable stacks, but almost all ideas carry over without much change. By definition, a stack is a very abstract object ("a lax sheaf of groupoids on the site of complex manifolds"), so let us discuss a relatively simple example, which helps to clarify the concept. We consider the stack $\mathfrak{M}_{g,n}$, the moduli stack of smooth n-pointed curves of genus g (alias Riemann surfaces). Let X be a complex manifold. We have to say what the groupoid $\mathfrak{M}_{g,n}(X)$ is. An object is a triple (E,π,j), where E is a complex manifold, $\pi: E \to X$ is a proper, surjective holomorphic submersion all of whose fibers are connected Riemann surfaces of genus g. The last piece of data is a holomorphic embedding $j: X\times\{1,\dots,n\}\to E$ such that $\pi\circ j$ is the projection onto X. If we forget about the complex structures, then Ehresmann's fibration theorem tells us that π is a differentiable fiber bundle with structure group $\mathrm{Diff}(F_g,(p_1,\dots,p_n))$. However, the complex structures on the fibers $\pi^{-1}(x)$ can vary with x. Experience shows that this is the appropriate notion of a holomorphic family of Riemann surfaces.

An isomorphism in the category $\mathfrak{M}_{g,n}(X)$ is the obvious thing: a biholomorphic map of the total spaces which commutes with the bundle maps and the embeddings.

For a holomorphic map $f: Y\to X$, we obtain a functor $f^*: \mathfrak{M}_{g,n}(X)\to\mathfrak{M}_{g,n}(Y)$. For two composable morphisms f_1, f_2, we do not quite have an equality $(f_2\circ f_1)^* = f_2^*\circ f_1^*$, but only up to "2-isomorphism". Finally, we can glue objects once we have a covering of a complex manifold and objects with suitably coherent isomorphisms on intersections.

It is a standard remark that the stack $\mathfrak{M}_{g,n}$ is not representable, i.e. that there does not exists a manifold M such that for any X, the groupoid $\mathfrak{M}_{g,n}(X)$ is equivalent to the set of holomorphic maps $X \to M$. However, in a certain precise sense, $\mathfrak{M}_{g,n}$ is not too far from being representable. The statement is formal, but the proof is not - it relies on Teichmüller theory (or geometric invariant theory, for those who like schemes). Let $\mathcal{T}_{g,n}$ be the Teichmüller space of n-pointed Riemann surfaces of genus g; it is a complex $3g-3+n$-dimensional complex manifold which is homeomorphic to $\mathbb{C}^{3g-3+n}$. Over $\mathcal{T}_{g,n}$, there is a universal family of Riemann surfaces, which gives an object in $\mathfrak{M}_{g,n}(\mathcal{T}_{g,n})$ which is, by abstract nonsense, a morphism of stacks $p : \mathcal{T}_{g,n} \to \mathfrak{M}_{g,n}$. This is an "atlas". The meaning of this phrase is that, whenever we have a complex manifold X and an object in $\mathfrak{M}_{g,n}(X)$ (alias a map $f : X \to \mathfrak{M}_{g,n}$), then we can find an open covering $(U_i)_{i\in I}$ of X, such that the restriction $f|_{U_i}$ admits a lift to $\mathcal{T}_{g,n}$. This is not hard to show (if and only if one knows Teichmüller theory): $\mathcal{T}_{g,n}$ is a classifying space for objects in $\mathfrak{M}_{g,n}$ with an additional piece of data: a homotopical trivialization of the underlying fiber bundle. For a general family of Riemann surfaces, such trivializations locally exist (by Ehresmann's theorem).

The atlas $\varphi : \mathcal{T}_{g,n} \to \mathfrak{M}_{g,n}$ has some additional properties which qualify the stack $\mathfrak{M}_{g,n}$ as a *complex-analytic Deligne-Mumford stack* or as a *complex orbifold* (which means the same).

To define the stack $\overline{\mathfrak{M}}_{g,n}$, we take not only holomorphic submersions as in the definition of $\mathfrak{M}_{g,n}$, but also suitably defined families of stable curves. This is also a Deligne-Mumford stack, but the construction of the atlas is far more technical as in the case of $\mathfrak{M}_{g,n}$. The reader is advised to consult either [**3**] and [**9**] (for an algebraic construction) or [**14**] for a differential-geometric perspective.

2. Homotopy theory of stacks

How do we extract homotopy theoretic information out of a stack? The following method makes sense in a more general situation, namely if we deal with topological stacks. Such a topological stack is a lax sheaf of groupoids on the site of topological spaces which admits an atlas (defined similarly as before). Let Stacks^{cpl} be the category of complex-analytic stacks and Stacks^{top} be the category of topological stacks. Given the very definition of a stack, one expects a functor

$\text{Stacks}^{top} \to \text{Stacks}^{cpl}$, but that does not happen. Let $\mathfrak{X}$ be a topological stack. Of course, we can restrict the sheaf defining $\mathfrak{X}$ to the subcategory of complex manifolds, but there is no reason why there should exist a complex-analytic atlas! Instead, there is a functor $\varphi : \text{Stacks}^{cpl} \to \text{Stacks}^{top}$ which extends the "underlying topological space functor" from complex manifolds to spaces. This is defined using an atlas, but it is a canonical construction whose result does not depend on that choice. However, given an analytic stack $\mathfrak{X}$, it may be very hard to describe the sheaf $\varphi(\mathfrak{X})$ explicitly. There are also differentiable stacks and similar remarks apply to this notion.

Given an atlas $X_0 \to \mathfrak{X}$ of a topological stack, the pullback $X_1 := X_0 \times_{\mathfrak{X}} X_0$ is again a space and there are suitable maps which define a topological groupoid $\mathbb{X}$ with object space X_0 and morphism space X_1. If $\mathfrak{X} = X /\!\!/ G$ is a quotient stack, then we obtain the translation groupoid of the group action $G \curvearrowright X$.
The following definition seems to be folklore.

Definition 2.1. Let $\mathfrak{X}$ be a topological stack and let $\mathbb{X}$ be the groupoid arising from an atlas of $\mathfrak{X}$. Then the *homotopy type* $\mathrm{Ho}(\mathfrak{X})$ of the stack $\mathfrak{X}$ is the homotopy type of $B\mathbb{X}$.

This definition has the obvious disadvantage that it is not clear that $\mathrm{Ho}(\mathfrak{X})$ is independent of the choice of the atlas. But in fact, it is.

Theorem 2.2. *The homotopy type* $\mathrm{Ho}(\mathfrak{X})$ *does not depend on the choice of the atlas. Moreover, it extends to a functor from the category of topological stacks to the homotopy category of spaces.*

The proof can be found in [**4**] and it is built on ideas from [**12**]. The second sentence is a quite strong statement, because it asserts that two different atlases do not merely give homotopy equivalent classifying spaces but also that all homotopy equivalences arising from different choices are mutually compatible.
If $\mathfrak{X} = X /\!\!/ G$ is a global quotient stack, then the homotopy type is the Borel-construction:

$$\mathrm{Ho}(X /\!\!/ G) = EG \times_G X.$$

A special case is the moduli space $\mathfrak{M}_{g,n}$, because the it is equivalent to the quotient of the Teichmüller space by the mapping class group[1] Γ_g^n. Because the $\mathcal{T}_{g,n}$ is contractible, we conclude that $\mathrm{Ho}(\mathfrak{M}_{g,n}) = B\Gamma_g^n$.
One can show that the homotopy type has the right (co)homology groups - there is a natural definition of the cohomology of a stack in terms of homological algebra and the result is that this homology is the same as the homology of the homotopy type. However, this remark does not apply to any homotopy-invariant functor, for example not to complex K-theory. Any good notion of complex K-theory should satisfy $K(X /\!\!/ G) = K_G(X)$ if G is a compact Lie group. But it is well-known that $K_G(X)$ and $K(EG \times_G X)$ are usually not isomorphic, see [**2**].

3. Pontrjagin-Thom maps

Let $f: M^m \to N^n$ be a proper smooth map of smooth manifolds, of codimension $d = n - m$ (which can be negative). The *normal bundle* is the stable vector bundle $\nu(f) := f^* TN - TM$ of virtual dimension d on M. As a stable vector bundle, it has a Thom spectrum $M^{\nu(f)}$. The Pontrjagin-Thom construction yields a stable homotopy class

$$\mathrm{PT}_f : \Sigma^\infty N_+ \to M^{\nu(f)}.$$

These Pontrjagin-Thom maps can be used to define umkehr maps in cohomology, once the normal bundle $\nu(f)$ is oriented. One defines f as the composition

$$H^*(M) \cong H^*(\Sigma^\infty M_+) \cong H^{*+d}(M^{\nu(f)}) \to H^{*+d}(\Sigma^\infty N_+) = H^{*+d}(N).$$

If we want to define umkehr maps also in the context of stacks, we need to construct Pontrjagin-Thom maps in the category of stacks. The problem is that the Whitney embedding theorem does not hold for stacks. But one can find a way around it and we can define the Pontrjagin-Thom map if $f: \mathfrak{X} \to \mathfrak{Y}$ is a representable proper map between complex-analytic stacks and $\mathfrak{Y}$ satisfies some mild technical conditions (this condition is satisfied for all orbifolds).

---

[1] This notation is traditional in the theory of mapping class groups. The group usually denoted by $\Gamma_{g,n}$ is closely related, but different.

4. Homotopy theory of smooth moduli spaces

The Pontrjagin-Thom construction played a crucial role in the modern homotopy theory of the moduli space $\mathfrak{M}_{g,n}$ which was developed by Tillmann, Madsen and Weiss [**10**], [**11**]. They studied the universal surface bundle(2) $\pi : \mathfrak{M}_{g,1} \to \mathfrak{M}_{g,0}$. The stable normal bundle $\nu(\pi)$ can be identified with the inverse of the vertical tangent bundle $T_v\pi$; the classifying map of $T_v\pi$ is a map $\mathfrak{M}_{g,1} \to BU(1)$. Thus the Pontrjagin-Thom construction yields a map

$$\alpha : \mathfrak{M}_g \to \Omega^\infty BU(1)^{-L}.$$

The main theorem of [**11**] is that α induces an isomorphism in integral homology in degrees $k \leqslant (g-2)/2$. A crucial ingredient of the proof is Harer's stability theorem [**7**] which says that $H_k(\mathfrak{M}_g;\mathbb{Z})$ does not depend on g if $k \leqslant (g-2)/2$. We will see below (see 6.1) that the homology of $\overline{\mathfrak{M}}_{g,n}$ does not satisfy any kind of stability. Therefore we cannot expect a result as elegant as the Madsen-Weiss theorem for $\overline{\mathfrak{M}}_{g,n}$.

5. The surjectivity theorem

There are several natural morphisms between the moduli stacks of stable curves. Namely, there are maps

1. $\xi_{g,n} : \overline{\mathfrak{M}}_{g-1,n+2} \to \overline{\mathfrak{M}}_{g,n}$
2. $\theta_{h,k} : \overline{\mathfrak{M}}_{h,k+1} \times \overline{\mathfrak{M}}_{g-h,n-k+1} \to \overline{\mathfrak{M}}_{g,n}$,
3. $\pi : \overline{\mathfrak{M}}_{g,n+1} \to \overline{\mathfrak{M}}_{g,n}$
4. $\sigma_* : \overline{\mathfrak{M}}_{g,n} \to \overline{\mathfrak{M}}_{g,n}; \sigma \in \Sigma_n$,

given by: identifying two smooth points to a node (1 and 2), forgetting the last point (3) or permuting the n marked points (4). These morphisms are representable morphisms of complex-analytic stacks; ξ and $\theta_{h,k}$ are proper immersions of codimension 1 and π can be interpreted as the universal family of stable curves (it has codimension -1). We are particularly interested in the morphisms ξ and $\theta_{h,k}$ and study their Pontrjagin-Thom maps. The normal bundles of these morphisms are easy to describe.

There are certain natural complex line bundles on $\overline{\mathfrak{M}}_{g,n}$: if $(C, p_1, \ldots, p_n)$ is an n-pointed stable curve, then p_i is a smooth point and hence $T_{p_i}C$ is defined; this

(2) The case $n > 0$ can easily be reduced to $n = 0$ using Harer stability.

gives line bundles $L_i \to \overline{\mathfrak{M}}_{g,n}$, $i = 1, \dots, n$.
The normal bundle of ξ is $L_{n+1} \otimes L_{n+2}$ and the normal bundle of $\theta_{h,k}$ is $L_{k+1} \otimes L_{n-k+1}$ (exterior tensor product). The morphism ξ is Σ_2-invariant and therefore induces $\tilde{\xi} : \overline{\mathfrak{M}}_{g-1,n+2} /\!\!/ \Sigma_2 \to \overline{\mathfrak{M}}_{g,n}$.

Let now $N(2) \subset U(2)$ be the normalizer of the standard maximal torus; there is a homomorphism $N(2) \to U(1)$ which multiplies the nonzero matrix entries. This induces a line bundle $V \to BN(2)$. The normal bundle of $\tilde{\xi}$ admits a bundle map and thus we obtain

$$\mathrm{PT}_{\tilde{\xi}} : \mathrm{Ho}(\overline{\mathfrak{M}}_{g,n}) \to \Omega^{\infty}\Sigma^{\infty} BN(2)^{V}.$$

Similarly, the normal bundle of $\theta_{h,0}$ admits a bundle map to the universal line bundle $L \to BU(1)$ and we obtain

$$\mathrm{PT}_{\theta_{h,0}} : \mathrm{Ho}(\overline{\mathfrak{M}}_{g,n}) \to \Omega^{\infty}\Sigma^{\infty} BU(1)^{L}.$$

The main result of [**4**] is the following.

Theorem 5.1. *The map* $\mathrm{PT}_{\tilde{\xi}}$ *induces an epimorphism in homology with field coefficients in degrees* $k \leqslant (g-2)/4$.
The map $\mathrm{PT}_{\theta_{h,0}}$ *induces an epimorphism in homology with field coefficients in degrees* $i \leqslant (g-2)/2(h+1)$.

The proof is based on the Harer-Ivanov stability theorem for the homology of the mapping class groups, on the Barratt-Priddy-Quillen theorem relating symmetric groups to infinite loop spaces and on the computation of the homology of the infinite loop space of the suspension spectrum of a space X in terms of the homology of X and the Dyer-Lashof algebra.
In section 6 below we will discuss the geometric meaning of some of the torsion classes provided by this theorem.
There is an important family of subrings $\mathscr{R}^*(\overline{\mathfrak{M}}_{g,n}) \subset H^*(\overline{\mathfrak{M}}_{g,n}; \mathbb{Q})$, the *tautological rings,* which is the smallest system of subalgebras which contain the classes $c_1(L_i)$ and which are closed under pullback and umkehr homomorphisms by the natural maps. It is easy to see that the Pontrjagin-Thom maps above map rational cohomology into the tautological ring. Therefore we can consider the cohomology classes induced by the Pontrjagin-Thom maps as an integral refinement of the tautological rings

6. The low-dimensional homology groups of $\overline{\mathfrak{M}}_g$

In this section, we present a short proof of the following theorem, which was first proven by Arbarello and Cornalba [**1**].

Theorem 6.1. *If $g > 4$, then $H_2(\overline{\mathfrak{M}}_g;\mathbb{Z})$ is a free Abelian group of rank $2+[g/2]$.*

Arbarello and Cornalba showed this using methods from algebraic geometry. Their argument showed also the apparently sharper result that any complex line bundle on $\overline{\mathfrak{M}}_g$ has a unique holomorphic structure. But the classical fact that $\pi_1(\overline{\mathfrak{M}}_g) = 0$ and an easy Hodge-theoretic argument show that Theorem 6.1 implies that as well.
The proof of 6.1 is based on the differential-topological notion of a Lefschetz fibration. In this framework, it is also easy to see that $\pi_1(\overline{\mathfrak{M}}_g) = 0$, using Dehn's theorem that Dehn twists generate the mapping class group. Consider the stack $\overline{\mathfrak{M}}_g$ as a differentiable stack. It quite difficult to describe this differentiable stack explicitly as a sheaf because a family of stable curves is not a bundle and when we pull back a family with an arbitrary smooth map, the resulting space becomes highly singular. But "up to concordance", the differentiable stack $\overline{\mathfrak{M}}_g$ is not too hard to understand. The notion of Lefschetz fibration is an old one in algebraic geometry, I learnt the following formulation from [**5**].

Definition 6.2. A *Lefschetz fibration* is a tuple (p,S,U,L,q), where $p: E^{k+2} \to B^k$ is a smooth map, $S \subset E$ is the subset of critical points of p and it is a submanifold of real codimension 4. One requires that $p|_S$ is an immersion with normal crossings. The normal bundle U of S in E is endowed with a complex structure and an embedding $j: U \to E$ as a tubular neighborhood; $L \to S$ is a complex line bundle, endowed with an immersion $i: L \to B$. $q: U \to L$ is a nondegenerate quadratic form and $p \circ j = i \circ q$. Finally, the fibers of p are oriented, connected stable surfaces.

For all $x \in B$, the nodes of the fiber $p^{-1}(x)$ are the points of $S \cap p^{-1}(x)$. Any component of S has a *type* $i \in \{0,1,\dots[g/2]\}$ (g is the genus of the fibers). Namely, a node can either be nonseparating ($i = 0$) or it can separate the surface into two parts of genus h and $g-h$ (if $h \leqslant g-h$, the type is h).
One can show that for any smooth manifold B, the set of concordance classes of Lefschetz fibrations is in bijection with the set of homotopy classes $[B;\mathrm{Ho}(\overline{\mathfrak{M}}_g]$. Details will appear elsewhere.
A Lefschetz fibration over a 1-manifold is nothing else than an oriented surface bundle; Lefschetz fibrations over oriented surfaces are also not hard to describe.

If F is a surface of genus g and $c \subset F$ a simple closed curve of type i (this is defined analogously to the type of a node), then there exists a Lefschetz fibration $p : E \to \mathbb{D}^2$ such that S consists of a single point s, $p(s) = 0$ of type i and the restriction $E|_{\mathbb{S}^1} \to \mathbb{S}^1$ is an oriented surface bundle whose monodromy is the Dehn twist around the curve c. If $E \to B$ is a Lefschetz fibration over a surface, then it is determined by the isomorphism class of the surface bundle $E|_{B \setminus p(S)}$ and by the monodromies around the points of $p(S)$.

Now we are ready for the proof of Theorem 6.1. We use the oriented bordism group $\Omega_2(\overline{\mathfrak{M}}_g)$ of Lefschetz fibrations, which is isomorphic to $H_2(\overline{\mathfrak{M}}_g;\mathbb{Z})$. We will establish an exact sequence

$$0 \longrightarrow \Omega_2(\mathfrak{M}_g) \longrightarrow \Omega_2(\overline{\mathfrak{M}}_g) \xrightarrow{\delta} \mathbb{Z}^{[g/2]+1} \longrightarrow 0. \tag{6.3}$$

The homomorphism δ is obtained by counting the singularities of a Lefschetz fibration, according to their type and with a sign which stems from orientation issues. This is invariant under oriented bordism.
To show that δ is surjective, we need to construct a Lefschetz fibration on an oriented surface with a single singularity of prescribed type. Take a Lefschetz fibration $E \to \mathbb{D}^2$ with a singularity of type i. The surface bundle $E|_{\mathbb{S}^1}$ is nullbordant in $\mathfrak{M}_g$, because $H_1(\mathfrak{M}_g;\mathbb{Z}) = 0$ for $g > 3$; this is a classical theorem by Powell [**13**]. Now take any nullbordism and glue in E. The result is a Lefschetz fibration with a single singularity.
An old theorem of Harer [**6**] states that $\Omega_2(\mathfrak{M}_g;\mathbb{Z}) \cong \mathbb{Z}$ if $g > 4$; an isomorphism is given by the following procedure: Take $a \in \Omega_2(\mathfrak{M}_g)$, which can be represented by an oriented closed surface M and a surface bundle $E \to M$. The signature of the oriented 4-manifold E is divisible by 4 and the assignment $[E \to M] \mapsto \frac{1}{4} sign(E)$ is an isomorphism. It follows immediately that the map $\Omega_2(\mathfrak{M}_g) \to \Omega_2(\overline{\mathfrak{M}}_g)$ induced by the inclusion is injective: if a surface bundle $E \to B$ is nullbordant when considered as a Lefschetz fibration, the manifold E is nullbordant and hence has signature 0. This show exactness of the sequence 6.3 on the left.
Exactness in the middle is shown by a simple surgery argument. If $E \to M$ is a Lefschetz fibration with $\delta(E \to M) = 0$, then the singular points of S of type i occur in pairs with opposite sign. If s_1, s_2 is such a pair, then we can cut out small discs in M around $p(s_1)$ and $p(s_2)$. The restriction of E to the boundary of any of the two

discs is a surface bundle and both are isomorphic (but the base has opposite orientations). Thus they are concordant as bundles over a cylinder. This cylinder can be glued in in and we obtain a new Lefschetz fibration, with the number of singularities reduced by 2. It represents the same bordism class as the original Lefschetz fibration. This finishes the proof of Theorem 6.1.

Remark 6.4. *The components of δ give cohomology classes $\delta_i \in H^2(\overline{\mathfrak{M}}_g;\mathbb{Z})$. They are related to our Pontrjagin-Thom maps as follows. Set $i = 0$, the other cases are similar. The Thom class of V is an element $u \in H^2(BN(2)^V;\mathbb{Z})$; it is suspended to $u' \in H^2(\Omega^\infty\Sigma^\infty BN(2)^V;\mathbb{Z})$. The class $\mathrm{PT}^*_{\tilde{\xi}}\, u'$ is precisely δ_0.*

7. An interesting class in $H^3(\overline{\mathfrak{M}}_g;\mathbb{F}_2)$

Theorem 5.1 states that the map $\overline{\mathfrak{M}}_{g,n} \to \Omega^\infty\Sigma^\infty BN(2)^V$ induces a surjection in homology with field coefficients. Equivalently, the map in cohomology with field coefficients is injective. Here we describe one of the torsion classes in $H^3(\overline{\mathfrak{M}}_{g,n};\mathbb{F}_2)$ geometrically. It is not hard to see that $H^*(BN(2);\mathbb{F}_2) \cong bF_2[x_1,x_2,w]/(w^3=0)$, where x_i is the image of (mod 2 reduction of) the Chern class $c_i \in H^(BU(2);\mathbb{F}_2)$ under the map induced from the inclusion $N(2) \subset U(2)$. The class $w \in H^1(BN(2);\mathbb{F}_2)$ comes from $BN(2) \to B\pi_0(N(2)) = B\mathbb{Z}/2$. Furthermore, the Euler class of the vector bundle V is $x_1 + w^2$. The Thom isomorphism is an isomorphism $\mathrm{th} : H^*(BN(2)) \cong H^{*+2}(BN(2)^V)$. Therefore, $H^3(BN(2)^V;\mathbb{F}_2) \cong \mathbb{F}_2$ and $\mathrm{th}(w)$ is a generator.
There is an (injective) homomorphism (of graded vector spaces, not of rings) $\sigma : H^*(BN(2)^V;\mathbb{F}_2) \to H^*(\Omega^\infty\Sigma^\infty BN(2)^V;\mathbb{F}_2)$, the cohomology suspension, and we want to describe $\mathrm{PT}^*_{\tilde{xi}}\,\sigma(\mathrm{th}(w)) \in H^3(\overline{\mathfrak{M}}_{g,n};\mathbb{F}_2)$. By 5.1, this is nonzero (if $g \geqslant 14$. By the universal coefficient theorem and by 6.1, $H^3(\overline{\mathfrak{M}}_{g,n};\mathbb{F}_2) \cong \mathrm{Hom}(H_3(\overline{\mathfrak{M}}_{g,n};\mathbb{Z});\mathbb{F}_2)$, and the latter is isomorphic to $\mathrm{Hom}(\Omega_3(\overline{\mathfrak{M}}_{g,n});\mathbb{F}_2)$.
Assume that B is a closed oriented 3-manifold and that $p : E \to B$ is a Lefschetz fibration which represents an element in $\Omega_3(\overline{\mathfrak{M}}_{g,n})$. Let $S \subset E$ be the singular locus, a 1-dimensional submanifold and let $S_0 \subset S$ be the open and closed subspace of singular points which are of type 0. Clearly, S_0 is a disjoint union of a finite number of circles $C_1,\dots C_k$. On any circle C_i, there is a twofold covering $q_i : \tilde{C}_i \to C_i$. Namely, for any $x \in C_i$, there exists a neighborhood $U \subset p^{-1}(p(x))$ such that $U \setminus x$ has precisely two components. These components are the elements of the fiber $q_i^{-1}(x)$.
Recall that there are exactly two equivalence classes of twofold coverings on a circle.

Let $a_i = 1$ if q_i is nontrivial and $a_i = 0$ if it is trivial. Define

$$\lambda(E \to B) := \sum_{i=1}^{k} a_i \in \mathbb{F}_2.$$

It is not hard to see that this is an additive bordism-invariant $\lambda : \Omega_3(\overline{\mathfrak{M}}_{g,n}) \to \mathbb{F}_2$ and hence a cohomology class $\lambda' \in H^3(\overline{\mathfrak{M}}_{g,n}; \mathbb{F}_2)$. More or less by unwinding the definitions, one can show that

$$\lambda' = \mathrm{PT}_{\xi}^{*}\, \sigma(\mathrm{th}(w)).$$

References

[1] E. Arbarello & M. Cornalba – "The Picard groups of the moduli spaces of curves", *Topology* **26** (1987), no. 2, p. 153–171.

[2] M. F. Atiyah & G. B. Segal – "Equivariant K-theory and completion", *J. Differential Geometry* **3** (1969), p. 1–18.

[3] P. Deligne & D. Mumford – "The irreducibility of the space of curves of given genus", *Inst. Hautes Études Sci. Publ. Math.* (1969), no. 36, p. 75–109.

[4] J. Ebert & J. Giansiracusa – "Pontrjagin-thom maps and the homology of the moduli stack of stable curves", preprint, arXiv:0712.0702.

[5] S. Galatius – "Homotopy theory of compactified moduli spaces", *Oberwolfach Reports* (2006), j.w.w. Y. Eliahberg.

[6] J. Harer – "The second homology group of the mapping class group of an orientable surface", *Invent. Math.* **72** (1983), no. 2, p. 221–239.

[7] J. L. Harer – "Stability of the homology of the mapping class groups of orientable surfaces", *Ann. of Math. (2)* **121** (1985), no. 2, p. 215–249.

[8] J. Heinloth – "Notes on differentiable stacks", Mathematisches Institut, Georg-August-Universität Göttingen: Seminars Winter Term 2004/2005, Universitätsdrucke Göttingen, Göttingen, 2005, p. 1–32.

[9] F. F. Knudsen – "The projectivity of the moduli space of stable curves. II. The stacks $M_{g,n}$", *Math. Scand.* **52** (1983), no. 2, p. 161–199.

[10] I. Madsen & U. Tillmann – "The stable mapping class group and $Q(\mathbb{C}\mathrm{P}^{\infty}_{+})$", *Invent. Math.* **145** (2001), no. 3, p. 509–544.

[11] I. Madsen & M. Weiss – "The stable moduli space of Riemann surfaces: Mumford's conjecture", *Ann. of Math. (2)* **165** (2007), no. 3, p. 843–941.

[12] B. Noohi – "Cohomology of pretopological stacks", unpublished manuscript, available at http://www.math.fsu.edu/~noohi/papers/papers.html, 2002.

[13] J. Powell – "Two theorems on the mapping class group of a surface", *Proc. Amer. Math. Soc.* **68** (1978), no. 3, p. 347–350.

[14] J. W. Robbin & D. A. Salamon – "A construction of the Deligne-Mumford orbifold", *J. Eur. Math. Soc. (JEMS)* **8** (2006), no. 4, p. 611–699.

Mathematisches Institut, Courant-Colloquium *Trends in Mathematics*
(R. MEYER, T. SCHICK, Y. TSCHINKEL, eds.), p. 69–96
Georg-August Universität Göttingen, 2008

THREE TROPICAL ENUMERATIVE PROBLEMS

Hannah Markwig

Mathematisches Institut, Universität Göttingen, Bunsenstr. 3-5, 37073 Göttingen
E-mail : markwig@umich.edu

***Abstract*.** In this survey, we describe three tropical enumerative problems and the corresponding moduli spaces of tropical curves. They have the structure of weighted polyhedral complexes. We observe similarities in the definitions of the weights, aiming at a better understanding of the tropical structure of the moduli spaces.

1. Introduction

In tropical geometry, algebraic varieties are degenerated to certain piece-wise linear objects called tropical varieties. This process loses a lot of information, but many properties of the algebraic variety can be read off from the tropical variety, and many theorems that hold on the algebraic side remarkably continue to hold on the tropical side. Since tropical varieties are piece-wise linear, they are in principle easier to understand than algebraic varieties and combinatorial methods apply. Thus there is hope that we can use tropical geometry to derive theorems in algebraic geometry.

One of the fields in which tropical geometry has had significant success recently is enumerative geometry. Enumerative geometry deals with the counting of geometric objects that are determined by certain incidence conditions. The conditions have to be chosen in such a way that only finitely many objects satisfy them. We will

2000 ***Mathematics Subject Classification*.** 14N10, 14N35, 51M20.

consider tropical analogues of the following three examples of enumerative numbers:

1. The numbers $N(d,g)$ of nodal degree d genus g plane curves through $3d+g-1$ points in general position.
2. The numbers $E(d,j)$ of nodal degree d elliptic (that is, genus 1) plane curves and with fixed j-invariant j through $3d-1$ points in general position.
3. The Hurwitz numbers $H_d^g(\eta,\nu)$ of genus g, degree d covers of $\mathbb{P}^1$, with specified ramification profiles η and ν over 2 fixed points in $\mathbb{P}^1$ and at most simple ramification over other points in $\mathbb{P}^1$.

Now we could instead count the corresponding tropical objects, defining numbers $N_{\text{trop}}(d,g)$, $E_{\text{trop}}(d,j)$ and $H^g_{d,\text{trop}}(\eta,\nu)$, and hope to end up with the same numbers. Each tropical object has to be counted with a certain tropical multiplicity that should reflect how many objects in the algebraic count degenerate to this tropical object. For the numbers $N(d,g)$, the Correspondence Theorem $N(d,g) = N_{\text{trop}}(d,g)$ has been shown in the pioneering work of Grigory Mikhalkin ([**12**]). The equality $E(d,j) = E_{\text{trop}}(d,j)$ was proved in [**10**] and $H_d^g(\eta,\nu) = H^g_{d,\text{trop}}(\eta,\nu)$ in [**3**].

The study of tropical enumerative numbers like the above requires an argument why the tropical count remains invariant under a deformation of the conditions, for instance, an argument why the numbers $N_{\text{trop}}(d,g)$ do not depend on the position of the $3d+g-1$ points (as long as they are in general position). The corresponding independence statements in algebraic geometry are a consequence of the fact that the enumerative numbers can be interpreted as intersection numbers of cycles on suitable moduli spaces, and that intersection products are invariant under deformation. In tropical geometry, we can construct analogues of the moduli spaces. Also, tropical intersection theory has been studied recently ([**13**],[**1**]). However, at this moment not all independence statements above can be deduced from general principles of tropical intersection theory. In fact, we can only use tropical intersection theory to prove independence statements for numbers of rational curves, that is, if the genus g is 0. For the numbers $N_{\text{trop}}(d,g)$, the independence was shown in [**12**] by relating the tropical numbers to classical ones for which the invariance is known. An alternative combinatorial proof determines the different possibilities of how a tropical curve can change when the points are deformed ([**7**]). For the numbers $E_{\text{trop}}(d,j)$ and $H^g_{d,\text{trop}}(\eta,\nu)$ the independence was shown in [**10**] and [**3**], respectively, using moduli space techniques. However, tedious case-by-case analyses and computations were necessary. The same techniques can also be used to prove the independence for the numbers $N_{\text{trop}}(d,g)$, and we outline this proof shortly in this survey since it cannot be found in the literature.

A main reason why we cannot prove the independence with tropical intersection theory is that the tropical moduli spaces we want to work with do not have a tropical structure yet. We can define them only as abstract weighted polyhedral complexes. The only case in which the tropical structure of the moduli space is well understood is for rational curves ([**14**], [**5**]). Since the corresponding moduli spaces in algebraic geometry are stacks rather than varieties, we expect that we need a rigorous definition of a tropical stack, which does not yet exist, before we can succeed in equipping the tropical moduli spaces with more structure. For genus 0, the notion of a tropical stack is avoided by introducing extra labelings that will remove automorphisms, see Section 7. Once we understand the tropical structure of the moduli spaces, we expect that tropical intersection theory should provide natural proofs of the independence statements, leading thus to a more rigorous set-up for tropical enumerative geometry.

The purpose of this survey is to describe the tropical moduli spaces used in the three enumerative problems mentioned above, and to observe similarities in their local structure. We hope a better understanding of tropical moduli spaces might eventually lead to a definition of a tropical structure for them.

We define three tropical moduli spaces that parametrize a larger set of objects than the ones we want to count. For the first enumerative problem (the numbers $N_{\text{trop}}(d,g)$) the space $\mathcal{M}_{g,n,\text{trop}}(\mathbb{R}^2,\Delta)$ parametrizes genus g, degree d plane tropical curves with $3d+g-1$ marked points. The space $\widetilde{\mathcal{M}_{1,n,\text{trop}}}(\mathbb{R}^2,\Delta)$ we use for the second problem (the numbers $E_{\text{trop}}(d,j)$) parametrizes degree d elliptic plane tropical curves with $3d-1$ marked points. The space $\mathcal{M}'_{g,0,\text{trop}}(\mathbb{R}^1,\Delta)$ for the third problem (the numbers $H^g_{d,\text{trop}}(\eta,\nu)$) parametrizes degree d, genus g tropical maps to $\mathbb{P}^1$. Then we define maps from these moduli spaces, for instance, the map that evaluates the $3d+g-1$ marked points for the first problem. The inverse image under this map of a point configuration consists of those degree d genus g tropical curves that pass through the point configuration. Inverse image points have to be counted with the suitable tropical multiplicity. The map for the second enumerative problem evaluates $3d-1$ points and associates the tropical j-invariant. The map for the third problem evaluates the position of the vertices of the tropical curve that can be thought of as branch points.

We start in Section 2 with an example which should motivate our definition of tropical curves. In Section 3, we define abstract tropical curves and parametrized tropical curves. The latter can be thought of as analogues of stable maps. We also define spaces parametrizing tropical curves that can be thought of as analogues of moduli spaces of stable maps. We equip those tropical moduli spaces with the structure of a weighted polyhedral complex in Section 4. In Section 5, we define maps from the tropical moduli spaces that are used to define the tropical enumerative problems. We explain how our definition of $N_{\text{trop}}(d,g)$ relates to Mikhalkin's original definition of $N_{\text{trop}}(d,g)$ ([**12**]). In Section 6, we give a short overview of the independence proofs that have to be shown for each of the three enumerative problems. We give only short outlines of proofs. For more details, or for more formal definitions, see [**10**], [**6**] or [**11**].

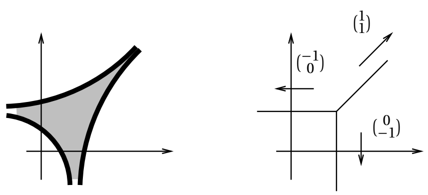
The author would like to thank Paul Johnson, Eric Katz, Thomas Markwig and Johannes Rau for helpful comments.

2. A motivating example

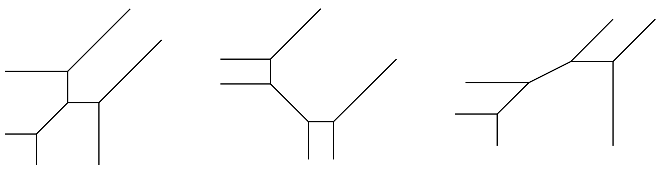
There are several ways to define the degeneration process which produces a tropical variety from an algebraic variety (see [**12**], [**16**], [**4**]). Here, we sketch just a basic example that motivates our later combinatorial definition of tropical curves. Let L be a projective line in $\mathbb{P}^2_{\mathbb{C}}$, and apply the map

$$\text{Log}\colon (\mathbb{C}^*)^2 \to \mathbb{R}^2, \quad (s,t) \mapsto (\log|s|, \log|t|)$$

to the restriction of L to $(\mathbb{C}^*)^2$. Let $(x:y:z)$ be the coordinates of $\mathbb{P}^2$, and identify $\mathbb{C}^2$ with the set $\{z \neq 0\}$. Then the map Log associates the point $(\log|\frac{x}{z}|, \log|\frac{y}{z}|) \in \mathbb{R}^2$ to a point $(x:y:z) \in \mathbb{P}^2$. The line L intersects the coordinate line $\{x=0\}$ in one point. When we move along the line L towards the intersection with $\{x=0\}$, the first coordinate of the image point under Log tends to $-\infty$. Also, when we move towards the intersection with the coordinate line $\{y=0\}$, the second coordinate of the image tends to $-\infty$. When we move towards the intersection with $\{z=0\}$, both coordinates become big and their difference tends to a constant. Furthermore, the image $\text{Log}(L) \subset \mathbb{R}^2$ (called the amoeba of L) should be 2-dimensional, as the complex line has two real dimensions. These observations suggest that the image looks similar to the left of the following picture:



A tropical line can be thought of as a limit of this amoeba after shrinking it to something one-dimensional, as on the right in the picture above. (For more details on the limit process, see [**12**] or [**4**].) The only information kept are the three infinite rays and their directions. Note that the primitive integer vectors pointing in these three directions sum up to 0. This is called the *balancing condition* and is important in our combinatorial definition. Now let $C \subset \mathbb{P}^2$ be a conic. It intersects $\{x = 0\}$ in two points, $(0 : p_0 : 1)$ and $(0 : p_1 : 1)$. We can move along C near the first point and the first coordinate of the image will tend to $-\infty$, whereas the second tends to $\log|p_0|$. For the second point, the first coordinate will again tend to $-\infty$, but the second to $\log|p_1|$. Thus the amoeba of a conic has two "tentacles" in each of the three directions $(-1,0)$, $(0,-1)$ and $(1,1)$. We can not say precisely what happens in the middle. When we shrink the amoeba to something 1-dimensional to get an idea of how a tropical conic should look like, there are indeed several possibilities of what can happen in the middle.



The picture shows three different types of a tropical conic.

In many places in the literature, an alternate degeneration process is given; namely, take the image of the valuation map from an algebraic variety over the field of Puiseux series K (or another field with a non-archimedean valuation). Since this definition does not require taking a limit, it is more useful for computations ([**2**]). We define plane tropical curves combinatorially (Definition 3.6). For plane curves, it is true that any such combinatorial object (roughly, a graph satisfying the balancing condition) comes from an algebraic curve under the degeneration process ([**17**]).

Remark 2.1. We can also apply the degeneration to higher-dimensional varieties. In the case of constant coefficients (that is, if the ideal defining the variety is an ideal of $\mathbb{C}[\underline{x}] \subset K[\underline{x}]$) the image under the valuation map is a polyhedral fan that satisfies (a higher-dimensional version of) the balancing condition ([**18**], Section 2.5). The role of the primitive integer vector pointing in the direction of an edge is played by the lattice in a top-dimensional cone of the fan. Combinatorially, higher-dimensional tropical varieties are defined roughly as polyhedral complexes obtained by gluing fans that satisfy the balancing condition ([**20**], [**5**]). Not every such polyhedral complex comes from an algebraic variety under the degeneration process.

3. Tropical $M_{g,n}(\mathbb{P}^r, d)$

We want to define a tropical analogue of $M_{g,n}(\mathbb{P}^r, d)$; that is, we want to define maps from abstract tropical curves to $\mathbb{R}^r$ such that the images look like the tropical curves we have seen in Section 2 above (that is, like graphs satisfying the balancing condition). The abstract tropical curves should be marked by n points. We have seen above that the unbounded edges (or, ends) of a tropical curve can be thought of as coming from the intersection with coordinate hyperplanes. In this sense, the ends are special points of the tropical curve. Thus we define the tropical analogue of marked points to be marked ends.

Let us first fix some notation we want to use for graphs. Let Γ be a graph. Unbounded edges (also called *ends*) are allowed. We denote the set of vertices by Γ^0 and the set of edges Γ^1. The subset of ends is called Γ^1_∞ and the subset of bounded edges Γ^1_0. We call a pair $F = (V, e)$ where e is an edge of Γ and $V \in \partial e$ a *flag* of Γ and think of it as a "directed edge"—an edge pointing away from its end vertex V. The genus of a connected graph Γ is the first Betti number of Γ, $h_1(\Gamma, \mathbb{Z})$, that is, the number of independent cycles.

Definition 3.1. An *abstract tropical curve* is a connected graph Γ whose vertices have valence at least 3 and whose bounded edges e are equipped with a length $l(e) \in \mathbb{R}_{>0}$.

The *genus* of an abstract tropical curve is the genus of Γ.

An *abstract n-marked tropical curve* is a tuple $(\Gamma, x_1, \dots, x_n)$ where Γ is an abstract tropical curve and $x_1, \dots, x_n \in \Gamma^1_\infty$ are distinct ends of Γ.

The set of all n-marked tropical curves with exactly n ends and of genus g is called $M_{g,n,\mathrm{trop}}$.

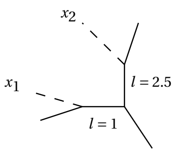
An *abstract tropical curve with labeled vertices* is an abstract tropical curve Γ where each vertex is labeled with $\mathrm{val}(V) - 2$ numbers such that the disjoint union of all vertex labelings equals $\{1,\dots,s-2+2g\}$, where s is the number of ends and g is the genus.

For $i \in \{1,\dots,s-2+2g\}$ we denote by V_i the vertex which has the label i. Note that for a curve with higher-valent vertices it is possible that $V_i = V_j$ for $i \neq j$ in this notation.

The *combinatorial type* α of an abstract tropical curve is the information left when dropping the lengths of the bounded edges.

Remark 3.2. We need vertex labelings only for the third enumerative problem, since tropical branch points of a map to tropical $\mathbb{P}^1$ are thought of as vertices of the underlying abstract tropical curve. Therefore we have to define two types of moduli spaces, one parametrizing curves with vertex labelings (but without marked ends) and one with marked ends.

Example 3.3. The following picture shows a 2-marked rational abstract tropical curve (without labeled vertices). Marked ends are drawn as dotted lines.



Remark 3.4. A connected graph of genus g has $\#\Gamma_0^1 = \#\Gamma_\infty^1 - 3 + 3g - \sum_V (\mathrm{val}\, V - 3)$ bounded edges. In particular, a 3-valent graph has $\#\Gamma_0^1 = \#\Gamma_\infty^1 - 3 + 3g$ bounded edges. A 3-valent graph of genus g has $\#\Gamma^0 = \#\Gamma_\infty^1 - 2 + 2g$ vertices. We need these relations for dimension counts later on.

Remark 3.5. For rational curves, the space $M_{0,n,\mathrm{trop}}$ is known to be the space of trees, or a quotient of the tropical Grassmanian ([**19**], [**14**], [**5**]). It is in fact equal to the tropicalization of $M_{0,n}$ (where $M_{0,n}$ is realized as a quotient of the Grassmanian) ([**8**], proposition 5.8). Therefore it is a fan satisfying the balancing condition as mentioned in remark 2.1.

Definition 3.6. A *(parametrized) tropical curve in* $\mathbb{R}^r$ *(with labeled vertices)* is a tuple (Γ, h) where Γ is an abstract tropical curve (with labeled vertices) and
$h\colon \Gamma \to \mathbb{R}^r$ is a continuous map satisfying:

1. h maps each edge e of length $l(e)$ affinely to a line segment with rational slope in $\mathbb{R}^r$, that is, if we identify the edge e with the interval $[0, l(e)]$ (or $[0,\infty)$ for an end),

h is of the form

$$h(t) = a + t \cdot v$$

for some $a \in \mathbb{R}^2$ and $v \in \mathbb{Z}^2$. The integral vector v occurring in this equation if $V \in \partial e$ is identified with 0 will be denoted $v(V,e)$ and called the *direction* of the flag (V,e). For an end e, we call its direction $v(e) = v(V,e)$ (where V is its only end vertex).

2. At every vertex $V \in \Gamma^0$, the *balancing condition* is fullfilled:

$$\sum_{e|V\in\partial e} v(V,e) = 0.$$

Note that $v(V,e) = -v(V',e)$ if $\{V,V'\} = \partial e$.

Definition 3.7. An *n-marked (parametrized) tropical curve in* $\mathbb{R}^r$ is a tuple

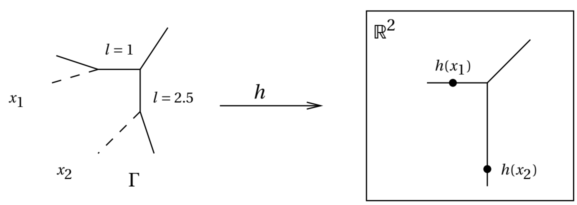
$$(\Gamma, h, x_1, \ldots, x_n)$$

where (Γ, h) is a tropical curve in $\mathbb{R}^r$, and $x_1, \ldots, x_n \in \Gamma^1_\infty$ are distinct ends of Γ that are mapped to a point in $\mathbb{R}^2$ by h (that is, $v(x_i) = 0$).

Definition 3.8.

1. The *genus* of a tropical curve in $\mathbb{R}^r$ is the genus of the underlying abstract tropical curve.
2. The *combinatorial type* of a tropical curve in $\mathbb{R}^r$ is given by the data of the combinatorial type of the underlying abstract tropical curve Γ together with the directions of all its edges (bounded edges as well as ends).
3. The *degree* of a tropical curve in $\mathbb{R}^r$ is the multiset $\Delta = \{v(e);\ e \in \Gamma^1_\infty \backslash \{x_1, \ldots, x_n\}\}$ of directions of its ends. If this degree consists of the vectors $-e_0, -e_1, \ldots, -e_r$, where $e_0 := -e_1 - \cdots - e_r$, and where each vector appears d times, we say that these curves have degree d.

Example 3.9. The following picture shows a rational tropical curve of degree 1 in $\mathbb{R}^2$ with two marked points.



Remark 3.10. Note that the direction vector $v(V,e)$ of a flag (V,e) (if it is nonzero) can be uniquely written as a product of a positive integer (called the weight of the edge e) and a primitive integer vector.

Remark 3.11. The map h of a tropical curve $(\Gamma, h, x_1, \ldots, x_n)$ does not need to be injective on the edges. It is allowed that $v(V,e) = 0$ for a flag (V,e), that is, the edge e is contracted to a point in $\mathbb{R}^2$. The remaining flags around the vertex V then satisfy the balancing condition. If V is a 3-valent vertex, this means that the two other flags (V,e_1) and (V,e_2) around V have to satisfy $v(V,e_1) = -v(V,e_2)$, that is, they point in opposite directions. Hence, the image $h(\Gamma)$ looks locally around $h(V)$ like a straight line.

This holds in particular for the marked ends $x_1, \ldots, x_n$, as they are required to be mapped to a point. Therefore, they can be seen as tropical analogues of the marked points of stable maps.

Note that the contracted bounded edges also lead to "hidden moduli parameters": if we vary the length of a contracted bounded edge, then we arrive at a family of different parametrized tropical curves whose images in $\mathbb{R}^2$ are all the same.

We are now ready to define the two types of moduli spaces mentioned in Remark 3.2.

Definition 3.12. For all g, $n \geqslant 0$ and Δ, let $M_{g,n,\mathrm{trop}}(\mathbb{R}^2, \Delta)$ be the set of all n-marked tropical curves $(\Gamma, h, x_1, \ldots, x_n)$ in $\mathbb{R}^2$ of degree Δ and genus $g' \leqslant g$.

For all $g \geqslant 0$ and Δ, let $M'_{g,0,\mathrm{trop}}(\mathbb{R}^1, \Delta)$ be the set of all tropical curves with labeled vertices (Γ, h) in $\mathbb{R}^1$ of degree Δ and genus $g' \leqslant g$.

We denote by $M^{\alpha}_{g,n,\mathrm{trop}}(\mathbb{R}^2, \Delta)$ and $M'^{\alpha}_{g,0,\mathrm{trop}}(\mathbb{R}^1, \Delta)$, respectively, the subsets of $M_{g,n,\mathrm{trop}}(\mathbb{R}^2, \Delta)$ and $M'_{g,0,\mathrm{trop}}(\mathbb{R}^1, \Delta)$ of tropical curves of combinatorial type α.

We need to include curves of lower genus here, since they appear in the boundary of types of genus g.

There are only finitely many combinatorial types in $M_{g,n,\mathrm{trop}}(\mathbb{R}^2, \Delta)$ and in $M'_{g,0,\mathrm{trop}}(\mathbb{R}^1, \Delta)$ ([**6**]).

Lemma 3.13. *The subsets $M^{\alpha}_{g,n,\mathrm{trop}}(\mathbb{R}^2, \Delta)$ and $M'^{\alpha}_{g,0,\mathrm{trop}}(\mathbb{R}^1, \Delta)$ are unbounded and open convex polyhedra in real vector spaces of dimension $2 + \#\Gamma^1_0$ and $1 + \#\Gamma^1_0$, respectively. They have, respectively, two or one coordinates $h(V)$ for the position of a root vertex V and coordinates $l(e)$ for the lengths of all bounded edges e. They are* *cut out by the inequalities that all lengths have to be positive and by the equations for the loops. If* Γ *is* 3*-valent, the expected dimensions are*

$$2+\#\Gamma_0^1-2g=\#\Delta-1+g \quad \textit{for} \quad M^{\alpha}_{g,n,\mathrm{trop}}(\mathbb{R}^2,\Delta),$$

and

$$1+\#\Gamma_0^1-g=\#\Delta-2+2g \quad \textit{for} \quad M'^{\alpha}_{g,0,\mathrm{trop}}(\mathbb{R}^1,\Delta).$$

*Proof***.** Given a curve of type α we can recover the map h from the data of the position of one root vertex. This is true because the directions are fixed by α and the lengths are fixed by the abstract curve. Thus $M^{\alpha}_{g,n,\mathrm{trop}}(\mathbb{R}^2,\Delta)$ and $M'^{\alpha}_{g,0,\mathrm{trop}}(\mathbb{R}^1,\Delta)$ are parametrized by the position $h(V_1)$ and the lengths of all bounded edges. The length coordinates have to satisfy the conditions that the g loops close up in the image in $\mathbb{R}^2$ (resp. $\mathbb{R}$). Each loop gives two (resp. one) conditions, but they do not have to be linearly independent. The statement about the expected dimension follows from Remark 3.4. □

A different choice of the root vertex or of the order of the bounded edges leads to a linear isomorphism on $M^{\alpha}_{g,n,\mathrm{trop}}(\mathbb{R}^2,\Delta)$ (resp. $M'^{\alpha}_{g,0,\mathrm{trop}}(\mathbb{R}^1,\Delta)$) of determinant ±1. This is obvious for the order of the bounded edges. If we choose another root vertex V', the difference $h(V)-h(V')$ of the images of the two vertices is given by $\sum_{(W,e)} l(e)\cdot v(W,e)$, where the sum is taken over a chain of flags leading from V to V'. This is obviously a linear combination of the lengths of the bounded edges. As these length coordinates themselves remain unchanged it is clear that the determinant of this change of coordinates is 1.

For any type α, the boundary of an open polyhedron $M^{\alpha}_{g,n,\mathrm{trop}}(\mathbb{R}^2,\Delta)$ or $M'^{\alpha}_{g,0,\mathrm{trop}}(\mathbb{R}^1,\Delta)$ consists of curves where some length coordinates are shrunk to 0. We can remove those edges and obtain a new tropical curve of a combinatorial type α', possibly of lower genus. The following picture shows how this can look like locally. The edges which tend to have length zero when we move towards the boundary are drawn in bold.

Thus we can glue the spaces $M^{\alpha}_{g,n,\mathrm{trop}}(\mathbb{R}^2,\Delta)$ or $M'^{\alpha}_{g,0,\mathrm{trop}}(\mathbb{R}^1,\Delta)$ along their boundaries.

4. The moduli spaces as weighted polyhedral complexes

Definition 4.1. Let $X_1,\dots,X_N$ be (possibly unbounded) open convex polyhedra in real vector spaces. A *polyhedral complex* with cells $X_1,\dots,X_N$ is a topological space X together with continuous inclusion maps $i_k\colon \overline{X_k} \to X$ such that X is the disjoint union of the sets $i_k(X_k)$ and the coordinate change maps $i_k^{-1} \circ i_l$ are affine (where defined) for all $k \neq l$. We usually drop the inclusion maps i_k in the notation and say that the cells X_k are contained in X.

The *dimension* $\dim X$ of a polyhedral complex X is the maximum of the dimensions of its cells. We say that X is of *pure dimension* $\dim X$ if every cell is contained in the closure of a cell of dimension $\dim X$. A point of X is said to be *in general position* if it is contained in a cell of dimension $\dim X$. For a point P in general position, we denote the cell of dimension $\dim X$ in which it is contained by X_P.

A *weighted polyhedral complex* is a polyhedral complex such that there is a weight $w(X_i) \in \mathbb{Q}$ associated to each cell X_i of highest dimension.

We want to glue the polyhedra $M^{\alpha}_{g,n,\mathrm{trop}}(\mathbb{R}^2,\Delta)$ or $M'^{\alpha}_{g,0,\mathrm{trop}}(\mathbb{R}^1,\Delta)$ to a weighted polyhedral complex. However, we want the polyhedral complex to be of the expected dimension, so in each case, we have to throw away certain strata. Later on we define maps from the moduli space to some other space that we use to impose conditions. We count tropical curves in the inverse image of a point. The strata whose dimension is too high are not mapped injectively and therefore do not contribute to the count. Thus we can drop them.

Also, we have to define weights for the top-dimensional strata. For this, we need the following definitions:

Definition 4.2. We call a type α in $M_{g,n,\mathrm{trop}}(\mathbb{R}^2,\Delta)$ or $M'_{g,0,\mathrm{trop}}(\mathbb{R}^1,\Delta)$ *regular* if the underlying graph is 3-valent and of genus g, and the g loops impose independent conditions.

Definition 4.3. Let α be a regular combinatorial type in $M_{g,n,\mathrm{trop}}(\mathbb{R}^2,\Delta)$ (resp. $M'_{g,0,\mathrm{trop}}(\mathbb{R}^1,\Delta)$). Pick g independent cycles of Γ, that is, generators of $H_1(\Gamma,\mathbb{Z})$. Each such generator is given as a chain of flags around the loop. Define a

$$2g \times 2 + \#\Gamma^1_0 = 2g \times n + \#\Delta - 1 + 3g$$

(resp. $g \times 1 + \#\Gamma^1_0 = g \times \#\Delta - 2 + 3g$) matrix A_α with two (one) columns for the position of a root vertex $h(V)$ and a column for each length coordinate, and with two (resp. one) rows for each cycle containing the equation of the loop in $\mathbb{R}^2$ (resp. $\mathbb{R}$)

(depending on the lengths of the bounded edges in the loop):

$$\sum_{(W,e)} v(W,e)\cdot l(e),$$

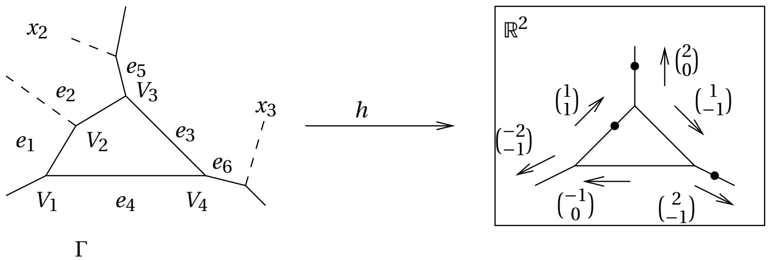
where the sum now goes over the chosen chain of flags around the loop. Then $A_\alpha\colon \mathbb{R}^{n+\#\Delta-1+3g} \to \mathbb{R}^{2g}$ (resp. $A_\alpha\colon \mathbb{R}^{\#\Delta-2+3g} \to \mathbb{R}^{g}$) is a linear map.

Denote by I_α the index of the sublattice
$A_\alpha(\mathbb{Z}^{n+\#\Delta-1+3g}) \subset \mathbb{Z}^{2g}$ (resp. $A_\alpha(\mathbb{Z}^{\#\Delta-2+3g}) \subset \mathbb{Z}^{g}$).

Note that $M^{\alpha}_{g,n,\mathrm{trop}}(\mathbb{R}^2,\Delta)$ (resp. $M'^{\alpha}_{g,0,\mathrm{trop}}(\mathbb{R}^1,\Delta)$) equals the intersection of $\mathbb{R}^2 \times (\mathbb{R}_{>0})^{\#\Gamma^1_0}$ (resp. $\mathbb{R}\times(\mathbb{R}_{>0})^{\#\Gamma^1_0}$) with the kernel of this map. This is true because we force the images of the cycles in $\mathbb{R}^2$ (resp. $\mathbb{R}$) to close up by requiring that the equations of the chains of flags are 0.

Note also that I_α does not depend on the chosen generators of $H_1(\Gamma,\mathbb{Z})$: If we choose another set of generators, these new generators are given as linear combinations with coefficients in $\mathbb{Z}$ of the old generators, so the rowspace of the matrix is not changed.

Example 4.4. The picture shows a regular curve C in $\mathscr{M}_{1,3,\mathrm{trop}}(\mathbb{R}^2,\Delta)$ (where $\Delta = \{(-2,-1),(0,2),(2,-1)\}$).



Choose the chain of flags $(V_1,e_1),\ldots,(V_4,e_4)$ around the cycle. The directions of those four flags are $(1,1)$, $(1,1)$, $(1,-1)$ and $(-1,0)$. Thus the map $A_\alpha\colon \mathbb{R}^8 \to \mathbb{R}^2$ is given by the following matrix:

$$\begin{pmatrix} 0 & 0 & 1 & 1 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 1 & -1 & 0 & 0 & 0 \end{pmatrix}$$

where the coordinates of $\mathbb{R}^8$ are $h(V_1)$, $l(e_1),\ldots,l(e_6)$.

4.1. The moduli space $\mathscr{M}_{g,n,\mathrm{trop}}(\mathbb{R}^2,\Delta)$ for the first enumerative problem, the numbers $N_{\mathrm{trop}}(d,g)$.

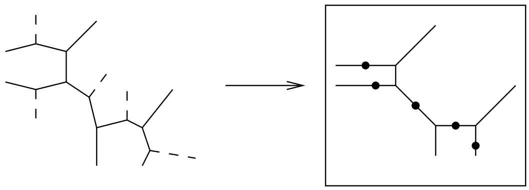
Definition 4.5. Let $C = (\Gamma, h, x_1, \ldots, x_n)$ be a tropical curve. If C has no contracted bounded edges (that is, no direction vector $v(e) = 0$ for $e \in \Gamma_0^1$), and if for all V such that there are two adjacent flags of the same direction $v(V, e_1) = v(V, e_2)$ the directions of the flags adjacent to V span $\mathbb{R}^2$, then C is called *relevant*. (In particular, every such vertex is at least 4-valent.)

We define $\mathscr{M}_{g,n,\mathrm{trop}}(\mathbb{R}^2, \Delta)$ to be the subset of $M_{g,n,\mathrm{trop}}(\mathbb{R}^2, \Delta)$ of relevant tropical curves which satisfy in addition the following property: if they are of genus $g' < g$, then they appear in the boundary of a relevant type of genus g.

The weight $w_1(\alpha)$ of a top-dimensional cell $M^{\alpha}_{g,n,\mathrm{trop}}(\mathbb{R}^2, \Delta)$ is defined to be the index I_α from Definition 4.3.

It follows from Proposition 4.1 in [**11**] that all types of top dimension in $\mathscr{M}_{g,n,\mathrm{trop}}(\mathbb{R}^2, \Delta)$ are regular. Thus the weight is well-defined.

The following picture shows an element of $\mathscr{M}_{0,5,\mathrm{trop}}(\mathbb{R}^2, 2)$:



4.2. The moduli space $\widetilde{\mathscr{M}}_{1,n,\mathrm{trop}}(\mathbb{R}^2, \Delta)$ for the second enumerative problem, the numbers $E_{\mathrm{trop}}(d, j)$. Let α be a combinatorial type in $M_{1,n,\mathrm{trop}}(\mathbb{R}^2, \Delta)$. The *deficiency* $\mathrm{def}(\alpha)$ is defined to be

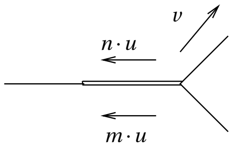
$$\mathrm{def}(\alpha) = \begin{cases} 2 \text{ if } g = 1 \text{ and the cycle is mapped to a point in } \mathbb{R}^2, \\ 1 \text{ if } g = 1 \text{ and the cycle is mapped to a line in } \mathbb{R}^2, \\ 0 \text{ otherwise.} \end{cases}$$

Since the loop imposes either two, one or no condition (depending on whether it spans $\mathbb{R}^2$, is mapped to a line or to a point), we can determine the dimension of $M^{\alpha}_{1,n,\mathrm{trop}}(\mathbb{R}^2, \Delta)$ exactly to be $\#\Delta + n + g - 1 - \sum_V (\mathrm{val}\, V - 3) + \mathrm{def}(\alpha)$ ([**10**]).

Definition 4.6. Remove from $M_{1,n,\mathrm{trop}}(\mathbb{R}^2, \Delta)$ the cells of dimension bigger than $\#\Delta + n$ and cells of rational curves which are not contained in the boundary of a cell corresponding to a genus 1 curve. The remaining subset of $M_{1,n,\mathrm{trop}}(\mathbb{R}^2, \Delta)$ is called $\widetilde{\mathscr{M}}_{1,n,\mathrm{trop}}(\mathbb{R}^2, \Delta)$. We associate the following weights to the strata of dimension $\#\Delta + n$:

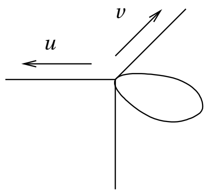
1. Assume $\operatorname{def}(\alpha) = 0$, and the curves of type α are of genus 1. Then we associate the weight $w_2(\alpha) = I_\alpha \cdot (\frac{1}{2})^r$, where r denotes the number of vertices V such that $\Gamma \setminus V$ has two connected components of the same combinatorial type (that is, for which both the abstract graph and the directions coincide).

2. Assume $\operatorname{def}(\alpha) = 1$. By the dimension count there is a 4-valent vertex. Assume first that the 4-valent vertex is adjacent to the cycle, that is, locally the curves look like the following picture:



In the notations above, $n \cdot u$, $m \cdot u$ and v denote the direction vectors of the corresponding edges (n and m are chosen such that their greatest common divisor is 1). If $n \neq m$, or if $n = m = 1$ and the cycle is formed by three edges due to the presence of a marked point, we associate the weight $w_2(\alpha) = |\det(u, v)|$. If $n = m = 1$ and no point is on the flat cycle, then we associate $w_2(\alpha) = \frac{1}{2}|\det(u, v)|$. (Due to the balancing condition this definition is not dependent of the choice of v.)
In case the 4-valent vertex is not adjacent to the cycle, we associate the weight 0.

3. Assume $\operatorname{def}(\alpha) = 2$. Assume first that the 5-valent vertex is adjacent to the cycle, that is, locally the curves look like this:

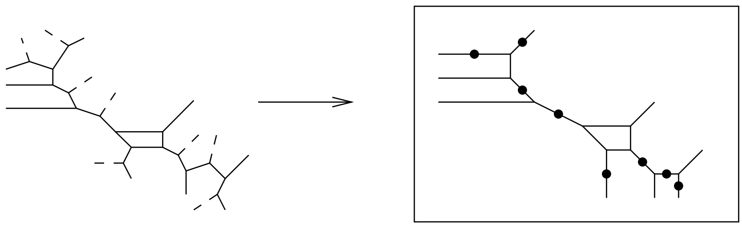


where u and v denote the direction vectors of the corresponding edges. We associate the weight $w_2(\alpha) = \frac{1}{2}(|\det(u, v)| - 1)$. (Note that due to the balancing condition this definition is independent of the choice of u and v.) In the case that there are two 4-valent vertices or that the 5-valent vertex is not adjacent to the cycle, we associate the weight 0.

The factor of $(\frac{1}{2})^r$ was left out in the original definition of $\widetilde{\mathcal{M}}_{1,n,\mathrm{trop}}(\mathbb{R}^2, \Delta)$ in [**10**]. The reason is that curves with vertices V such that $\Gamma \setminus V$ has two connected components of the same type count with multiplicity 0 later, since they are not mapped injectively.

Remark 4.7. Note that we include types in $\widetilde{\mathcal{M}}_{1,n,\mathrm{trop}}(\mathbb{R}^2,\Delta)$ which are not relevant and thus not included in $\mathcal{M}_{g,n,\mathrm{trop}}(\mathbb{R}^2,\Delta)$. The reason is that the map we want to use for the first enumerative problem only evaluates at different points, whereas the map for the second enumerative problem takes the cycle length of the tropical curve into account (see Definition 5.5). A cell corresponding to a non-relevant type like the one with $\mathrm{def}(\alpha) = 1$ above is not mapped injectively with just evaluations, because we can change the length coordinates in the cycle without changing the position of any marked point. It is mapped injectively in the second problem though, because a change of length coordinates in the cycle changes the cycle length. Therefore we have to consider it in the second problem, but not in the first one.

The following picture shows an element of $\widetilde{\mathcal{M}}_{1,8,\mathrm{trop}}(\mathbb{R}^2,3)$:



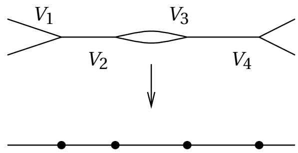
4.3. The moduli space $\mathcal{M}'_{g,0,\mathrm{trop}}(\mathbb{R}^1,\Delta)$ for the third enumerative problem, the numbers $H^g_{d,\mathrm{trop}}(\eta,\nu)$. As mentioned in Remark 3.2, we need vertex labels here.

Definition 4.8. Let $\mathcal{M}'_{g,0,\mathrm{trop}}(\mathbb{R}^1,\Delta)$ be the subset of $M'_{g,0,\mathrm{trop}}(\mathbb{R}^1,\Delta)$ containing all combinatorial types α such that if $M'^{\alpha}_{g,0,\mathrm{trop}}(\mathbb{R}^1,\Delta)$ is of dimension $\#\Delta-2+2g$ or bigger then α is regular and if $M'^{\alpha}_{g,0,\mathrm{trop}}(\mathbb{R}^1,\Delta)$ is of dimension less than $\#\Delta-2+2g$ then it is contained in a cell corresponding to a regular type.

In particular, the top dimension of $\mathcal{M}'_{g,0,\mathrm{trop}}(\mathbb{R}^1,\Delta)$ is $\#\Delta-2+2g$. We define the weight $w_3(\alpha)$ of a top-dimensional cell as the product of three types of factors:

– the index I_α;
– $\frac{1}{2}$ for every vertex V such that $\Gamma\setminus V$ has two connected components of the same combinatorial type;
– $\frac{1}{2}$ for every cycle which consists of two edges which have the same direction.

The following picture shows an element of $\mathcal{M}'_{1,0,\mathrm{trop}}(\mathbb{R}^1,2)$:



It is easy to show now that the three moduli spaces are indeed weighted polyhedral complexes.

Remark 4.9. Note that the weights in Definitions 4.5, 4.6 and 4.8 coincide. We do not need factors of $\frac{1}{2}$ in Definition 4.5 because a regular and relevant curve cannot have a vertex V such that $\Gamma \setminus V$ has two connected components of the same combinatorial type or a cycle consisting of two edges which have the same weight. Also, we do not need the special cases (2) and (3) of 4.6 in either of the two other definitions. They are not relevant. In the third enumerative problem, they are not of top dimension.

The factors of $\frac{1}{2}$ can be thought of as taking care of automorphisms (see also Section 7).

5. The tropical enumerative problems

Definition 5.1. A *morphism* between a weighted polyhedral complex X and a polyhedral complex Y is a continuous map $f\colon X \to Y$ such that for each cell $X_i \subset X$ the image $f(X_i)$ is contained in only one cell of Y, and $f|_{X_i}$ is a linear map (of polyhedra).

Assume $f\colon X \to Y$ is a morphism of weighted polyhedral complexes of the same pure dimension, and $P \in X$ is a point such that both P and $f(P)$ are in general position (in X and Y, respectively). Then locally around P the map f is a linear map between vector spaces of the same dimension. We denote by D_P the absolute value of the determinant of this linear map and define the *multiplicity* $\mathrm{mult}_f(P) = D_P \cdot w(X_P)$ of f at P to be D_P times the weight of the cell X_P, $w(X_P)$. Note that the multiplicity depends only on the cell X_P of X in which P lies. We call it the multiplicity of f in this cell.

A point $Q \in Y$ is said to be *in f-general position* if Q is in general position in Y and all points of $f^{-1}(Q)$ are in general position in X. Note that the set of points in f-general position in Y is the complement of a subset of Y of dimension at most $\dim Y - 1$; in particular it is a dense open subset. Now if $Q \in Y$ is a point in f-general position we define the *degree* of f at Q to be

$$\deg_f(Q) := \sum_{P \in f^{-1}(Q)} \operatorname{mult}_f(P).$$

Note that this sum is indeed finite: first of all there are only finitely many cells in X. Moreover, in each cell (of maximal dimension) of X where f is not injective (that is, where there might be infinitely many inverse image points of Q) the determinant of f is zero and hence so is the multiplicity for all points in this cell.

Moreover, since X and Y are of the same pure dimension, the cells of X on which f is not injective are mapped to a locus of codimension at least 1 in Y. Thus the set of points in f-general position away from this locus is also a dense open subset of Y, and for all points in this locus we have that not only the sum above but indeed the fiber of Q is finite.

Note that the definition of multiplicity $\operatorname{mult}_f(P)$ in general depends on the coordinates we choose for the cells. However, we will use this definition only for morphisms for which D_P, the absolute value of the determinant, does not depend on the chosen coordinates, if they are chosen in a natural way; in our case this means we choose lattice bases of the spaces $M^{\alpha}_{g,n,\mathrm{trop}}(\mathbb{R}^2,\Delta)$ and $M'^{\alpha}_{g,0,\mathrm{trop}}(\mathbb{R}^1,\Delta)$. Choosing a different lattice basis leads to a base change matrix of determinant ± 1 which does not change the multiplicity. Since D_P depends only on the cell and for us cells correspond to combinatorial types α, we will use the notation D_α.

As lattice bases are in general hard to compute, we use the following easier way to determine $\operatorname{mult}_f(C)$ for a morphism starting from one of our moduli spaces:

Construction 5.2. Let $f\colon \mathscr{M} \to Y$ be a morphism of weighted polyhedral complexes of the same pure dimension, where $\mathscr{M}$ is $\mathscr{M}_{g,n,\mathrm{trop}}(\mathbb{R}^2,\Delta)$, $\widetilde{\mathscr{M}}_{1,n,\mathrm{trop}}(\mathbb{R}^2,\Delta)$ or $\mathscr{M}'_{g,0,\mathrm{trop}}(\mathbb{R}^1,\Delta)$. For a regular type α the cell $M = M^{\alpha}_{g,n,\mathrm{trop}}(\mathbb{R}^2,\Delta)$ (resp. $M = M'^{\alpha}_{g,0,\mathrm{trop}}(\mathbb{R}^1,\Delta)$) is cut out of $V = \mathbb{R}^{2+\#\Gamma^1_0}$ (resp. $V = \mathbb{R}^{1+\#\Gamma^1_0}$) by the inequalities that lengths are positive and by $d = 2g$ (resp. $d = g$) independent equations for the loops. Pick a map $\tilde{f}_\alpha\colon V \to Y \times \mathbb{R}^d$ such that $\tilde{f}_\alpha|_M = f|_M \times A_\alpha$, where A_α is the map containing the equations for the loops as in 4.3.

Lemma 5.3. *For a map $\tilde{f}_\alpha$ (defined for a regular type α) from Construction* 5.2 *we have* $|\det(\tilde{f}_\alpha)| = I_\alpha \cdot D_\alpha = \operatorname{mult}_f(C)$, *where C is a curve of type α. In particular* $|\det(\tilde{f}_\alpha)|$ *does not depend on the choice of $\tilde{f}_\alpha$.*

This is basically a straightforward lattice index computation, which can be found in [**15**], Lemma 1.6. The index of a square integer matrix is just the absolute value of its determinant, and the index of a product of two maps $f \times g$ is equal to the index of $f|_{\ker g}$ times the index of g. Remember that $\tilde{f}_\alpha = f|_M \times A_\alpha$ and M (that is, the cell of type α) is the kernel of the map A_α (intersected with the conditions that lengths have to be positive) by Definition 4.3.

Example 5.4. For the maps we will use, we can choose a possible $\tilde{f}_\alpha$ just by choosing chains of flags to the marked points, respectively to the vertices. For the curve C in Example 4.4, choose V_1 to be the root vertex, and go from V_1 to x_1 via (V_1, e_1). Go to x_2 from V_1 via (V_1, e_1), (V_2, e_2) and (V_3, e_5) and to x_3 via (V_1, e_1), (V_2, e_2), (V_3, e_3) and (V_4, e_6). Thus $\tilde{f}_\alpha$ is the $2 + \#\Gamma_0^1 = 8$ times $2n + 2 = 8$ matrix

$$\begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 2 & 0 \\ 1 & 0 & 1 & 1 & 1 & 0 & 0 & 2 \\ 0 & 1 & 1 & 1 & -1 & 0 & 0 & -1 \\ 0 & 0 & 1 & 1 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 1 & -1 & 0 & 0 & 0 \end{pmatrix}$$

where the coordinates of $\mathbb{R}^8$ are $h(V_1)$, $l(e_1), \ldots, l(e_6)$. Note that we could for example also have gone to x_3 via (V_1, e_4) in which case the fifth and sixth row would be replaced by

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 1 & 0 & 2 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 \end{pmatrix}$$

This matrix differs from the other only by subtracting the seventh from the fifth and the eighth from the sixth line—that is, we subtract the two equations for the loop from one chain of flags to get to the other chain of flags. In particular choosing a different chain of flags does not change the absolute value of the determinant. We will see that the map $\tilde{f}_\alpha$ we describe here equals $\mathrm{ev}|_M \times A_\alpha$ when restricted to the cell $M = M^\alpha_{1,3,\mathrm{trop}}(\mathbb{R}^2, \Delta)$ (the map ev is defined in 5.5). Then by Lemma 5.3 we have $2 = |\det(\tilde{f}_\alpha)| = \mathrm{mult}_{\mathrm{ev}}(C)$.

Definition 5.5. Let

$$\mathrm{ev}_i : M_{g,n,\mathrm{trop}}(\mathbb{R}^2, \Delta) \to \mathbb{R}^2,\ (\Gamma, h, x_1, \ldots, x_n) \mapsto h(x_i)$$

denote the i-th evaluation map. By $\mathrm{ev} = \mathrm{ev}_1 \times \ldots \times \mathrm{ev}_n$ we denote the combination of all n evaluation maps.

The j-invariant of an elliptic curve tropicalizes to the cycle length ([**9**]). For a tropical curve $C = (\Gamma, h, x_1, \ldots, x_n)$ of genus 1, we pick a generator of $H_1(\Gamma, \mathbb{Z})$ given as a chain of flags. If we avoid passing any edge in two directions, it is unique up to orientation. We define the cycle length to be the sum of the lengths of the edges which are part of this cycle. This can also be expressed in terms of forgetful maps ([**10**]). We define a map $j : \widetilde{\mathcal{M}}_{1,n,\mathrm{trop}}(\mathbb{R}^2, \Delta) \to \mathbb{R}_{\geqslant 0}$ sending C to its cycle length. For a rational tropical curve, we say $j(C) = 0$. Define

$$\pi := \mathrm{ev} \times j : \widetilde{\mathcal{M}}_{1,n,\mathrm{trop}}(\mathbb{R}^2, \Delta) \to \mathbb{R}^{2n} \times \mathbb{R}_{\geqslant 0}.$$

We define the tropical branch map δ as

$$\delta\colon \mathscr{M}'_{g,0,\mathrm{trop}}(\mathbb{R}^1,\Delta) \to \mathbb{R}^{\#\Delta-2+2g}\colon (\Gamma,h) \mapsto (h(V_1),\dots,h(V_{\#\Delta-2+2g})),$$

where V_i is the vertex in Γ^0 with label i as defined in 3.1.

All those maps are morphisms of weighted polyhedral complexes. For example, the position $h(x_i)$ equals $h(V)+\sum v(V,e)\cdot l(e)$ where the sum goes over a chain of flags leading from V to x_i. This expression is linear in the coordinates $h(V)$ and $l(e)$.

Now we define the tropical enumerative numbers.

Definition 5.6.

1. Let $n = \#\Delta+g-1$. For a point configuration $\mathscr{P}\in\mathbb{R}^{2n}$ in ev-general position, define $N_{\mathrm{trop}}(\Delta,g) = \deg_{\mathrm{ev}}(\mathscr{P})$.
2. Let $n=\#\Delta-1$. For a point configuration $\mathscr{P}\in\mathbb{R}^{2n}$ in π-general position, define $E_{\mathrm{trop}}(\Delta,j)=\deg_\pi(\mathscr{P})$.
3. For a point configuration $\mathscr{P}\in\mathbb{R}^{\#\Delta-2+2g}$ in δ-general position, define $H^g_{d,\mathrm{trop}}(\eta,\nu)=\deg_\delta(\mathscr{P})$.

The question as posed in the introduction is now why those numbers do not depend on the point $\mathscr{P}$, that is, why the degrees of the three maps are constant. We give an outline of these proofs in Section 6.

Note that $N_{\mathrm{trop}}(\Delta,g)$ is defined differently in [**12**]: the tropical curves there are counted with a multiplicity $\mathrm{mult}(C)$ which is not defined via the evaluation map. We show that $\mathrm{mult}(C)$ for a relevant and regular curve C of type α coincides with $\mathrm{mult}_{\mathrm{ev}}(C)$ ([**11**]).

Definition 5.7. The multiplicity of a 3-valent vertex V is defined to be the absolute value of the determinant $\det(v_1,v_2)$, where v_1 and v_2 are two directions of flags adjacent to V. The balancing condition tells us that it makes no difference which two of the three flags adjacent to V we choose. The *multiplicity* $\mathrm{mult}(C)$ of a 3-valent tropical curve is defined to be the product of the multiplicities of all vertices ([**12**]).

Example 5.8. The multiplicity of the curve C from Example 4.4 equals $\mathrm{mult}(C)=2$. As we have seen in Example 5.4, $\mathrm{mult}_{\mathrm{ev}}(C)=2$, too.

A *string* in C is a subgraph of Γ homeomorphic either to $\mathbb{R}$ or to S^1 (that is, a "path" starting and ending with an unbounded edge, or a path around a loop) that does not intersect the closures $\overline{x_i}$ of the marked points.

Definition 5.9. For a tropical curve C of regular type α, pick a chain of flags for each marked point x_i leading from the root vertex V to x_i. Define a matrix $\widetilde{\mathrm{ev}}_\alpha : \mathbb{R}^{2+\#\Gamma_0^1} \to \mathbb{R}^{2n} \times \mathbb{R}^{2g}$ with two rows for each marked point containing the chain of flags and two rows for each loop containing the equation of the loop (as in Definition 4.3).

Remark 5.10. Note that $\widetilde{\mathrm{ev}}_\alpha = \mathrm{ev} \times A_\alpha$ on $M^\alpha_{g,n,\mathrm{trop}}(\mathbb{R}^2, \Delta)$, where A_α is defined in 4.3 and $M^\alpha_{g,n,\mathrm{trop}}(\mathbb{R}^2, \Delta)$ is the kernel of A_α intersected with the conditions that the lengths are positive. The map $\widetilde{\mathrm{ev}}_\alpha$ depends on the choices, but $|\det \widetilde{\mathrm{ev}}_\alpha|$ does not since $|\det \widetilde{\mathrm{ev}}_\alpha| = w(\alpha) \cdot D_\alpha = \mathrm{mult}_{\mathrm{ev}}(C)$ by Lemma 5.3. By abuse of notation, we still speak of *the map* $\widetilde{\mathrm{ev}}_\alpha$, even though its definition depends on the choices we made, and keep in mind that $|\det(\widetilde{\mathrm{ev}}_\alpha)|$ is uniquely determined.

Note that 5.4 gives an example of such a map $\widetilde{\mathrm{ev}}_\alpha$.

Lemma 5.11. *Let C be a curve of degree Δ and relevant and regular type α, which is marked by $\#\Delta + g - 1$ points. Then* $\mathrm{mult}_{\mathrm{ev}(C)}$ *is equal to* $\mathrm{mult}\, C$ *if C has no string.*

Note that curves with a string are not mapped injectively by ev (see [**6**], Remark 3.6), therefore they do not contribute to the count $\deg_{\mathrm{ev}}(\mathscr{P})$. Also, if we choose a configuration of points in general position, no curve with a string meets the points.

Proof. We show that $|\det(\widetilde{\mathrm{ev}}_\alpha)|$ equals $\mathrm{mult}\, C$, which is enough by Remark 5.10. The proof is by induction on the sum of the number of bounded edges and the genus. The induction beginning is shown in [**6**], Example 3.3.

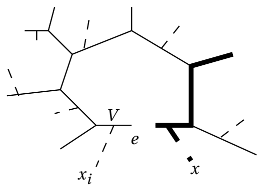
In the induction step, let us now assume C has k bounded edges, is a curve of genus g and degree Δ, and $k + g > 2$. Cut one of the bounded edges. That is, in the graph Γ, choose a bounded edge e and replace it by two ends, each being adjacent to one end vertex of e. Two things can happen:

1. The graph can decompose into two connected components.
2. The graph can stay connected, but a loop is broken. We denote the new connected graph of genus $g - 1$ by Γ_1. In this case, the edge e should be chosen such that it is adjacent to a marked point x_i. (Such a choice is possible as C has no string.)

We have to prove the statement for each of the two cases separately, as the arguments differ. The first case is shown in [**6**], Proposition 3.8. In the second case, Γ_1 has genus $g-1$, $\#\Delta+2$ ends that are not marked points, and

$$\#\Delta + g - 1 < (\#\Delta+2) + (g-1) - 1$$

marked points, therefore it has a string. This can be seen by removing $\overline{x_i}$ one by one, thus producing several connected components. Since we do not have enough marked points, we end up with less connected components than ends. We add a marked point x adjacent to one of the new ends. There is only one possibility to do this such that the new tropical curve has no string. The tropical curve C_1 of type α_1 defined in this way has genus $g-1$ and as many bounded edges as C. Therefore we can assume by induction that its multiplicity is equal to $|\det \widetilde{\mathrm{ev}}_{\alpha_1}|$. As $\mathrm{mult}(C) = \mathrm{mult}(C_1)$, it remains to show that $|\det \widetilde{\mathrm{ev}}_{\alpha}| = |\det \widetilde{\mathrm{ev}}_{\alpha_1}|$.



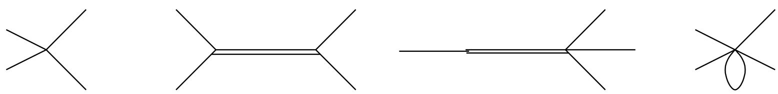
Choose coordinates to compare the two matrices of $\widetilde{\mathrm{ev}}_{\alpha}$ and $\widetilde{\mathrm{ev}}_{\alpha_1}$. Let V—the vertex adjacent to the marked point x_i—be the root vertex both for C and for C_1. Choose the same order of bounded edges, marked points and loops for the two curves. One of the loops of C, say L, is broken after the cutting of e. This loop corresponds to the last two lines of the matrix of $\widetilde{\mathrm{ev}}_{\alpha}$. For C_1, the last two lines are given by the marked point x. As chain of flags leading from V to x in C_1, we choose just the same chain of flags as for the loop L. The following table represents both matrices. The two matrices only differ by the $h(V)$-entries in the last two rows. In the table, each row represents two or more rows as before. Each matrix contains the first three rows, $\widetilde{\mathrm{ev}}_{\alpha}$ contains the fourth, and $\widetilde{\mathrm{ev}}_{\alpha_1}$ the fifth. E_2 denotes the two by two unit matrix.

	$h(V)$	bounded edges
the marked point x_i	E_2	0
other marked points	E_2	$*$
other loops	0	$*$
for $\widetilde{\mathrm{ev}}_{\alpha}$ the loop L	0	equation for L
for $\widetilde{\mathrm{ev}}_{\alpha_1}$ the new point x	E_2	equation for L

Note that both matrices are block matrices with a 2×2 block on the top left. Therefore, both determinants are equal to the determinant of the lower right block. But this block coincides for both matrices, because it does not involve the two numbers we changed from 0 to 1. □

6. The independence proof

To prove that $\deg_f(\mathscr{P})$ does not depend on $\mathscr{P}$ where f is one of our maps above, note first that the degree is locally constant on the subset of points in f-general position. This is true since at any curve that contributes to $\deg_f(\mathscr{P})$ the map f is a local isomorphism. The points in f-general position are the complement of a polyhedral complex of codimension 1, that is, they form a finite number of top-dimensional regions separated by "walls" that are polyhedra of codimension 1. To show that $\deg_f$ is constant it is therefore enough to consider a general point on such a wall and show that $\deg_f$ is locally constant at these points. Such a general point on a wall is the image under f of a general tropical curve C of a combinatorial type α such that the cell corresponding to α is of codimension 1. We have to classify all those types. For the first enumerative problem, this is done in [**11**] and for the second in [**10**]. (For the third, the wall-crossing statement is actually not necessary since all types contribute to the sum $\deg_\delta$, not depending on $\mathscr{P}$.) The following shows local pictures of codimension 1 types.



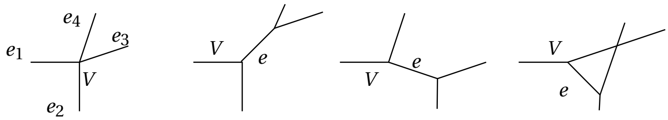
The pictures represent the abstract graph and the direction vectors at the same time: the double edge in the second and third picture from the right represents two edges of the graph Γ which are mapped to the same line segment of $\mathbb{R}^2$ since they are of the same direction. The loop in the picture on the right represents a loop of direction 0 (leading to a type of deficiency 2 which is of codimension 1 because of its 6-valent vertex).

For the first enumerative problem, the numbers $N_{\text{trop}}(d,g)$, we have to consider the left two pictures and the case of a 3-valent curve of genus $g-1$. The other ones are not relevant. For the second enumerative problem, the numbers $E_{\text{trop}}(d,j)$, we have to consider all the pictures above. The third picture from the left leads to several subcases depending on where the marked points are relative to the cycle.

Here, we want to present only the case corresponding to the leftmost picture. This is in fact the most important case, since it has to be considered in all enumerative problems. Also, it is the only case that has to be considered for rational curves. An analogous independence proof for rational curves appeared in [**6**]. We outline the proof only for the first enumerative problem, the numbers $N_{\text{trop}}(d,g)$, that is, we use the map ev: $\mathcal{M}_{g,n,\text{trop}}(\mathbb{R}^2,\Delta) \to (\mathbb{R}^2)^n$. For the other two maps, it is completely analogous.

Lemma 6.1. *Let $\mathscr{P} \in \mathbb{R}^{2n}$ be a configuration of points such that there is a curve C with one 4-valent vertex satisfying* $\text{ev}(C) = \mathscr{P}$*. Then the number of preimages near C under* ev *of a point $\mathscr{P}'$ near $\mathscr{P}$ (counted with multiplicity) does not depend on the choice of $\mathscr{P}'$.*

Proof. Let α be the type of C. The cell $M^{\alpha}_{g,n,\text{trop}}(\mathbb{R}^2,\Delta)$ is in the boundary of three top-dimensional cells, namely the ones where the 4-valent vertex is resolved.



We study the three matrices A_1, A_2 and A_3 of $\widetilde{\text{ev}}_{\alpha_1}$, $\widetilde{\text{ev}}_{\alpha_2}$ and $\widetilde{\text{ev}}_{\alpha_3}$. They differ only in the column corresponding to the edge e. Denote the four edges adjacent to the 4-valent vertex of C with $e_1,\dots,e_4$, and their respective directions with $v_1,\dots,v_4$. The root vertex is V as indicated in the picture. We assume that all choices of flags for evaluation and loops are made consistently. Then the three matrices only differ in the column which belongs to the new edge e. The following table represents all three matrices: Each matrix A_i contains the first block of columns (corresponding to the image $h(V)$ of the root vertex and the lengths l_i of the edges e_i) and the i-th of the last three columns (corresponding to the length of the edge e).

	$h(V)$	l_1	l_2	l_3	l_4	l^{α_1}	l^{α_2}	l^{α_3}
e_1	E_2	v_1	0	0	0	0	0	0
e_2	E_2	0	v_2	0	0	0	v_2+v_3	v_2+v_4
e_3	E_2	0	0	v_3	0	v_4+v_3	v_2+v_3	0
e_4	E_2	0	0	0	v_4	v_4+v_3	0	v_2+v_4
e_1, e_2	0	$-v_1$	v_2	0	0	0	v_2+v_3	v_2+v_4
e_1, e_3	0	$-v_1$	0	v_3	0	v_3+v_4	v_2+v_3	0
e_1, e_4	0	$-v_1$	0	0	v_4	v_3+v_4	0	v_2+v_4
e_2, e_3	0	0	$-v_2$	v_3	0	v_3+v_4	0	$-v_2-v_4$
e_2, e_4	0	0	$-v_2$	0	v_4	v_3+v_4	$-v_2-v_3$	0
e_3, e_4	0	0	0	$-v_3$	v_4	0	$-v_2-v_3$	v_2+v_4

The columns corresponding to the other bounded edges are not shown; it is enough to note here that they are the same for all three matrices. The size of the matrices is $2n+2g$ times $2+\#\Gamma_0^1$, and $2+\#\Gamma_0^1 = 2+\#\Delta-3+3g = n = \#\Delta+g-1+2g = 2n+2g$ because of remark 3.4 and definition 5.6. The first four rows correspond to the images of the marked points. The row labeled with e_i stands for the evaluations of marked points that can be reached from V via e_i. The last six rows correspond to the equations of the loops. The row labeled e_i, e_j stands for equations of loops that involve the two edges e_i and e_j. We get four different types of rows for the marked points depending on via which of the four edges e_i a marked point is reached from V. For the loops, we get six different types of rows depending on which two of the four edges $e_1,\dots,e_4$ are involved in a loop. Each row represents in fact two or more rows of the matrix, two rows for the two coordinates of the image of each marked point resp. two equations given by each loop. Loops that do not involve any of the four edges are not added, they do not change the computations. As det is linear in each column, $\det A_1 + \det A_2 + \det A_3$ is equal to the determinant of the following matrix, where we added the three last columns:

	$h(V)$	l_1	l_2	l_3	l_4	
e_1	E_2	v_1	0	0	0	0
e_2	E_2	0	v_2	0	0	$2v_2+v_3+v_4$
e_3	E_2	0	0	v_3	0	$2v_3+v_2+v_4$
e_4	E_2	0	0	0	v_4	$2v_4+v_3+v_2$
e_1 and e_2	0	$-v_1$	v_2	0	0	$2v_2+v_3+v_4$
e_1 and e_3	0	$-v_1$	0	v_3	0	$2v_3+v_2+v_4$
e_1 and e_4	0	$-v_1$	0	0	v_4	$2v_4+v_2+v_3$
e_2 and e_3	0	0	$-v_2$	v_3	0	v_3-v_2
e_2 and e_4	0	0	$-v_2$	0	v_4	v_4-v_2
e_3 and e_4	0	0	0	$-v_3$	v_4	v_4-v_3

Now we subtract the four columns for $l_1, \dots, l_4$ from the last column.

	$h(V)$	l_1	l_2	l_3	l_4	
e_1	E_2	v_1	0	0	0	$-v_1$
e_2	E_2	0	v_2	0	0	$v_2+v_3+v_4$
e_3	E_2	0	0	v_3	0	$v_3+v_2+v_4$
e_4	E_2	0	0	0	v_4	$v_4+v_3+v_2$
e_1 and e_2	0	$-v_1$	v_2	0	0	$v_2+v_3+v_4+v_1$
e_1 and e_3	0	$-v_1$	0	v_3	0	$v_3+v_2+v_4+v_1$
e_1 and e_4	0	$-v_1$	0	0	v_4	$v_4+v_2+v_3+v_1$
e_2 and e_3	0	0	$-v_2$	v_3	0	0
e_2 and e_4	0	0	$-v_2$	0	v_4	0
e_3 and e_4	0	0	0	$-v_3$	v_4	0

Due to the balancing condition $v_1+v_2+v_3+v_4=0$. We add v_1 times the $h(V)$-columns to the last column and get a matrix with a zero column whose determinant is 0. Therefore $\det A_1 + \det A_2 + \det A_3 = 0$.

Note that we assume here that the edges e_i are in fact all bounded. If this is not true, the argument needs to be changed slightly. If e_i is unbounded, then there can be no marked points that can be reached from V via e_i. That is, we do not have the corresponding rows.

For a given $i \in \{1,2,3\}$ let us now determine whether the combinatorial type α_i occurs in the inverse image under ev of a fixed point $\mathscr{P}'$ near $\mathscr{P}$. We may assume without loss of generality that the multiplicity of α_i is non-zero since other types are irrelevant for the statement of the proposition. Then A_i is an invertible matrix. There is therefore at most one inverse image point. The root vertex and length coordinates for a curve in the inverse image under ev of type α_i are given as $A_i^{-1}\cdot(\mathscr{P}',0)$, since $\widetilde{\mathrm{ev}}_{\alpha_i} = \mathrm{ev}\times A_{\alpha_i}$ on $M^{\alpha_i}_{g,n,\mathrm{trop}}(\mathbb{R}^2,\Delta)$ by Remark 5.10. In fact, this point exists in $M^{\alpha_i}_{g,n,\mathrm{trop}}(\mathbb{R}^2,\Delta)$ if and only if all coordinates of $A_i^{-1}\cdot(\mathscr{P}',0)$ corresponding to lengths of bounded edges are positive. By continuity this is obvious for all edges except the newly added edge !e, because in the boundary curve C all these edges had positive length. We conclude that there is a curve of type α_i mapping to $\mathscr{P}'$ if and only if the last coordinate (corresponding to the length of the newly added edge e) of $A_i^{-1}\cdot(\mathscr{P}',0)$ is positive. By Cramer's rule this last coordinate is $\det\widetilde{A}_i/\det A_i$, where $\widetilde{A}_i$ denotes the matrix A_i with the last column replaced by $(\mathscr{P}',0)$. But note that $\widetilde{A}_i$ does not depend on i since the last column was the only one where the matrices A_i differ. Hence whether there is a curve of type α_i or not depends only on the sign of $\det A_i$: either there are such inverse image points for exactly those i where $\det A_i$ is positive, or exactly for those i where $\det A_i$ is negative. But by the above the sum of the absolute values of the determinants satisfying this condition is the same in both cases. □

Remark 6.2. Note that we have to distinguish a case if we prove an analogous statement for other enumerative problems. If $\Gamma \setminus V$ has two connected components of the same combinatorial type, say the two components containing e_1 and e_2, then the type α_1 gets an extra factor of $\frac{1}{2}$. The types α_2 and α_3 are identical, so the statement is still true in this case. We do not have to consider this case for the numbers $N_{\text{trop}}(d,g)$ or $E_{\text{trop}}(d,j)$, because A_1 would not be injective since we can change length coordinates without changing the image.

7. Conclusion

We mentioned in the introduction that we believe that the moduli spaces we consider here should be equipped with tropical structure, and that once this is achieved the tedious case-by-case analysis (for each codimension 1 case) in the independence proof from Section 6 can be replaced by an easy intersection theory argument. This hope is in fact true in the case of rational curves: for rational curves (that is, for the numbers $N_{\text{trop}}(d,0)$ for example) the moduli space is known to be a fan satisfying a balancing condition as in Remark 2.1 ([**5**]) and thus a tropical variety. The tropical structure is derived from the tropical structure of $M_{0,n,\text{trop}}$ mentioned in Remark 3.5. A trick has been used to avoid the notion of a tropical stack here: the unmarked (that is, non-contracted) ends are labeled to make them distinguishable even if they have the same direction. Then there is a subgroup G of the symmetric group that acts on the moduli space with labeled ends by relabeling the non-contracted ends. The enumerative numbers we get have to be divided by $|G|$ to reflect the fact that we count each curve several times with different labels for the non-contracted ends ([**5**]). For a general type, there are $|G|$ ways to relabel the ends. For a type with vertices V such that $\Gamma \setminus V$ has two components of the same type, there are only $\frac{1}{2} \cdot |G|$ ways to label the ends. This enlightens why we include factors of $\frac{1}{2}$ for the weights in such a case in Definitions 4.6 and 4.8.

It is also known that the evaluation map $\text{ev}\colon \mathscr{M}_{0,n,\text{trop}}(\mathbb{R}^2,\Delta) \to (\mathbb{R}^2)^n$ is a morphism of tropical varieties ([**5**]). Since the space and the map are equipped with tropical structure we can use intersection theory arguments to deduce that $\deg_{\text{ev}}$ is constant for the case of rational curves ([**5**]).

The proof that the moduli space of rational tropical curves is a tropical variety, that is, balanced ([**14**], [**5**]), involves an argument which is very similar to the proof of Lemma 6.1 above. We need to consider a codimension 1 cell (that is, a cell corresponding to a curve with one 4-valent vertex) and consider the neighboring top-dimensional cells just as above. That means that for rational curves, the work to show an independence statement as above is hidden in the proof that the moduli space is a tropical variety. We hope that a similar statement can be shown for higher genus curves, too. We hope that the description of moduli spaces of tropical curves of higher genus as weighted polyhedral complexes used in this survey can contribute to the understanding of their tropical structure.

References

[1] L. Allermann & J. Rau – "Tropical intersection theory", Preprint, arXiv:0709.3705, 2007.

[2] T. Bogart, A. N. Jensen, D. Speyer, B. Sturmfels & R. R. Thomas – "Computing tropical varieties", *J. Symbolic Comput.* **42** (2007), no. 1-2, p. 54–73.

[3] R. Cavalieri, P. Johnson & H. Markwig – "Tropical Hurwitz numbers", Preprint, arXiv:0804.0579, 2008.

[4] A. Gathmann – "Tropical algebraic geometry", *Jahresber. Deutsch. Math.-Verein.* **108** (2006), no. 1, p. 3–32.

[5] A. Gathmann, M. Kerber & H. Markwig – "Tropical fans and the moduli space of rational tropical curves", Preprint, arXiv/0708.2268v1.

[6] A. Gathmann & H. Markwig – "Kontsevich's formula and the WDVV equations in tropical geometry", Advances in Mathematics (to appear). arXiv:math.AG/0509628, 2005.

[7] ______, "The numbers of tropical plane curves through points in general position", *J. Reine Angew. Math.* **602** (2007), p. 155–177.

[8] A. Gibney & D. Maclagan – "Equations for Chow and Hilbert Quotients", Preprint, arXiv:0707.1801, 2007.

[9] E. Katz, H. Markwig & T. Markwig – "The j-invariant of a plane tropical cubic", Preprint, arXiv:0709.3785, 2007.

[10] M. Kerber & H. Markwig – "Counting tropical elliptic plane curves with fixed j-invariant", Commentarii Mathematici Helvetici (to appear). arXiv:math.AG/0608472, 2006.

[11] H. Markwig – "The enumeration of plane tropical curves", Ph.D. Thesis, TU Kaiserslautern, 2006.

[12] G. Mikhalkin – "Enumerative tropical algebraic geometry in $\mathbb{R}^2$", *J. Amer. Math. Soc.* **18** (2005), no. 2, p. 313–377 (electronic).

[13] ______, "Tropical geometry and its applications", International Congress of Mathematicians. Vol. II, Eur. Math. Soc., Zürich, 2006, p. 827–852.

[14] ______, "Moduli spaces of rational tropical curves", Preprint, arXiv/0704.0839, 2007.

[15] J. RAU – "The index of a linear map of lattices", Preprint, TU Kaiserslautern, 2006.

[16] J. RICHTER-GEBERT, B. STURMFELS & T. THEOBALD – "First steps in tropical geometry", Idempotent mathematics and mathematical physics, Contemp. Math., vol. 377, Amer. Math. Soc., Providence, RI, 2005, p. 289–317.

[17] D. SPEYER – "Uniformizing tropical curves I: Genus zero and one", Preprint, arXiv:0711.2677.

[18] ———, "Tropical geometry", Ph.D. Thesis, University of California, Berkeley, 2005.

[19] D. SPEYER & B. STURMFELS – "Tropical mathematics", Preprint, math.CO/0408099, 2004.

[20] B. STURMFELS & J. TEVELEV – "Elimination theory for tropical varieties", Preprint, math.AG/0704.3471, 2007.

Mathematisches Institut, Courant-Colloquium Trends in Mathematics
(R. MEYER, T. SCHICK, Y. TSCHINKEL, eds.), p. 97–106
Georg-August Universität Göttingen, 2008

VOLUME AND L^2-BETTI NUMBERS OF ASPHERICAL MANIFOLDS

Roman Sauer
Department of Mathematics, University of Chicago, 5734 S. University Avenue, Chicago, Illinois 60637 • *E-mail :* `romansauer@member.ams.org`

Abstract. We give a leisurely account of the relationship between volume and L^2-Betti numbers on closed, aspherical manifolds based on the results in [**4**] – albeit with a different point of view. This paper grew out of a talk presented at the first colloquium of the Courant Center in Göttingen in October 2007.

1. Review of L^2-Betti numbers

The L^2-Betti numbers of a closed Riemannian manifold, as introduced by M. Atiyah, are analytical invariants of the long-time behavior of the heat kernel of the Laplacians of forms on the universal cover. We give a very brief review of these invariants; for extensive information the reader is referred to the standard reference [**3**].

Let $\widetilde{X} \to X$ be the universal cover of a compact Riemannian manifold, and let $\mathscr{F} \subset \widetilde{X}$ be a $\pi_1(X)$-fundamental domain. Then Atiyah defines the i-th L^2-Betti number in terms of the heat kernel on $\widetilde{X}$ as

$$b_i^{(2)}(X) = \lim_{t\to\infty} \int_{\mathscr{F}} \mathrm{tr}_{\mathbb{C}}\, e^{-t\Delta_i}(x,x)\, d\,vol(x).$$

Subsequently, simplicial and homological definitions of L^2-Betti numbers were developed by Dodziuk, Farber, and Lück. An important consequence of the equivalence of these definitions is the homotopy invariance of L^2-Betti numbers.

2000 ***Mathematics Subject Classification.*** 22D20,53C20,58J22.

Lück's definition is based on a dimension function $\dim_{\mathscr{A}}(M)$ for arbitrary modules M over a finite von Neumann algebra $\mathscr{A}$ with trace $\mathrm{tr}: \mathscr{A} \to \mathbb{C}$. For example, one has $\dim_{\mathscr{A}}(\mathscr{A}p) = \mathrm{tr}(p)$. Lück proceeds then to define $b_i^{(2)}(X)$ for an arbitrary space X with $\Gamma = \pi_1(X)$ as

$$b_i^{(2)}(X) = \dim_{L(\Gamma)} H_i\big(L(\Gamma) \otimes_{\mathbb{Z}\Gamma} C_*(\widetilde{X})\big) \in [0,\infty] \tag{1.1}$$

where $L(\Gamma)$ is the group von Neumann algebra of Γ. Some of the most fundamental properties of L^2-Betti numbers are:

– $\pi_1(X)$ finite $\Rightarrow b_i^{(2)}(X) = b_i(\widetilde{X})/|\pi_1(X)|$
– $\sum_{i\geq 0}(-1)^i b_i^{(2)}(X) = \chi(X) = \sum_{i\geq 0}(-1)^i b_i(X)$.
– $\bar{X} \to X$ d-sheeted cover $\Rightarrow b_i^{(2)}(\bar{X}) = d \cdot b_i^{(2)}(X)$.
– If X is aspherical and $\pi_1(X)$ amenable then $b_i^{(2)}(X) = 0$.
– If X is a $2n$-dimensional hyperbolic manifold then $b_i^{(2)}(X) > 0$ if and only if $i = n$.

2. Theorems relating volume and L^2-Betti numbers

Assumption 2.1. *Throughout this section, let M be an n-dimensional, closed, aspherical manifold.*

The inequality of Theorem 2.2 is stated by Mikhail Gromov [**2**]*Section 5.33 on p. 297 along with an idea(1) which he attributes to Alain Connes. We provide the first complete proof of that inequality [**4**]*Corollary to Theorem A. The rigorous implementation of Gromov's idea uses tools and ideas from Damien Gaboriau's theory of L^2-Betti numbers of measured equivalence relations and spaces with groupoid actions of such.

Theorem 2.2. *If (M,g) has a lower Ricci curvature bound* $\mathrm{Ricci}(M,g) \geq -(n-1)g$, *then*

$$b_i^{(2)}(M) \leq \mathrm{const}_n \,\mathrm{vol}(M,g) \quad \textit{for every } i \geq 0.$$

The *minimal volume* of a smooth manifold N is defined as the infimum of volumes of complete metrics on N whose sectional curvature is pinched between -1 and 1. We obtain the following

(1) We refer to this idea as *randomization.*

Corollary 2.3 (Minimal volume estimate).

$$b_i^{(2)}(M) \le \mathrm{const}_n \,\mathrm{minvol}(M).$$

The following theorem [**4**]*Theorem B is a generalization of a well-known vanishing result of Jeff Cheeger and Mikhail Gromov. Its connection to volume becomes apparent through its corollary.

Theorem 2.4. *If M is covered by open, amenable sets such that every point belongs to at most n sets, then*

$$b_i^{(2)}(M) = 0 \ \text{ for every } i \ge 0.$$

Here a subset $U \subset M$ is called *amenable* if $\pi_1(U)$ maps to an amenable subgroup of $\pi_1(M)$. There is also a version of this theorem for arbitrary spaces [**4**]*Theorem C. The following corollary is a non-trivial implication of the theorem above and work of Mikhail Gromov [**1**]*Section 3.4 where he constructs amenable coverings in the presence of small volume.

Corollary 2.5. *There is a constant $\varepsilon_n > 0$ only depending on n such that*

$$\mathrm{minvol}(M) < \varepsilon_n \Rightarrow b_i^{(2)}(M) = 0 \ \text{ for every } i \ge 0.$$

The results above are analogs of well-known theorems by Mikhael Gromov where *L^2-Betti numbers* are replaced by *simplicial volume*. Note however that the assumption of asphericity is crucial here unlike in the case of the simplicial volume.

3. Idea of proof of the main theorem

We describe some ideas involved in the proof of Theorems 2.2 and 2.4. In Subsection 3.1 we describe a general technique of bounding L^2-Betti numbers by constructing suitable equivariant coverings on the universal cover. Since the assumptions of our theorems are too weak to garantuee the existence of such covers we need substantially modify this technique; the new tool runs under the name *randomization*, and it is explained in Subsection 3.2. A full proof based on randomization is rather long and complicated; we explain instead an instructive toy example in Subsection 3.4. A crucial property of L^2-Betti numbers is described in Subsection 3.3. We conclude this sketch of proof in Subsection 3.5 with some remarks about other ingredients.

Throughout the section, we refer to Assumption 2.1.

3.1. How to bound L^2-Betti numbers by equivariant coverings in general. Let $\Gamma = \pi_1(M)$. Suppose we construct, under a certain geometrical assumption, a Γ-equivariant open covering $\mathscr{U}$ of the universal cover $\widetilde{M}$. Let us say that $\mathscr{U} = \{U_i\}_{i\in I}$ is indexed by a free Γ-set set I, and we have $\gamma U_i = U_{\gamma i}$. By a standard argument (partition of unity) one obtains a Γ-equivariant map f from $\widetilde{M}$ to the nerve of $\mathscr{U}$. The nerve is embedded in the full simplicial complex with index set I which we denote by $\Delta(I)$. Let

$$\Omega = \operatorname{map}\big(\widetilde{M}, \Delta(I)\big)$$

be the space of continuous maps with the natural Γ-action. We may view f as an element in Ω^Γ, the subspace of Ω consisting of Γ-equivariant maps. Next we argue that both the i-th Betti number and the L^2-Betti number are bounded from above by the number of equivariant i-simplices hit by $f(\widetilde{M})$.

Let $\mathscr{F}_i$ be a set of Γ-representatives of the i-skeleton $\Delta(I)^{(i)}$. For any $g \in \Omega$, let $C_i(g) \in \mathbb{N}$ be the number of i-simplices in $\mathscr{F}_i$ hit by $f(\widetilde{M})$. We think of C_i as a function

$$C_i : \Omega \to \mathbb{Z}.$$

Since $\widetilde{M}$ is contractible, M is a model of the classifying space $B\Gamma$, and the universal property of $E\Gamma$, the universal cover of $B\Gamma$, implies that there is an equivariant homotopy retract

$$\begin{tikzcd} \widetilde{M} \arrow[r, "f"'] & \Delta(I) \arrow[l, bend right] \end{tikzcd} .$$

Using the fact that the i-th L^2-Betti number is bounded by the number of equivariant i-simplices and the fact that the L^2-Betti number is some sort of dimension (with nice properties) of a certain homology module (see (1.1)), we easily obtain that

$$b_i^{(2)}(M) \le C_i(f).$$

By going to Γ-quotients we also obtain the same estimate for the usual Betti numbers. By Poincare duality it is actually enough to control $C_n(f)$, and we have

$$b_i(M), b_i^{(2)}(M) \le \mathrm{const}_n\, C_n(f) \tag{3.1}$$

for a constant const_n only depending on n. This follows from [**3**]*Example 14.28 on p. 498 since the fundamental class of M can be written as a sum of at most $C_n(f)$ singular simplices.

So to get a good bound on $b_i^{(2)}(M)$, we should find an equivariant cover $\mathscr{U}$ such that for the resulting map f to the nerve the quantity $C_n(f)$ is rather small.

3.2. Randomization. One directly sees the limitations of the above technique. The trivial estimate $C_n(f) \geq 1$ for any map $f \in \Omega$ prevents us from proving the vanishing of the L^2-Betti numbers. In particular, we cannot hope to prove Theorems 2.2 and 2.4 using it.

Next we phrase an idea of Mikhail Gromov (attributed to Alain Connes) in probabilistic terms that modifies the above technique.

By changing the point of view a bit, we regard a map $f \in \Omega^\Gamma$ that we sought to construct before as a Γ-invariant point measure on the Borel space Ω. Instead of trying to find a point measure f with small $C_n(f)$, Gromov suggests to look for Γ-invariant probability measures μ on Ω such that the *expected value*

$$\mathbb{E}_{(\Omega,\mu)}(C_n) = \int_\Omega C_n(f)\,d\mu(f) \text{ is sufficiently small.}$$

We refer to the problem of finding a suitable probability measure as the *randomization problem.* It turns out that in analogy to (3.1) one can actually show that

$$b_i^{(2)}(M) \leq \mathrm{const}_n \, \mathbb{E}_{(\Omega,\mu)}(C_n) \ \forall i \geq 0, \tag{3.2}$$

and that one can actually use the assumptions of Theorems 2.2 and 2.4 to construct a Γ-invariant probability measure μ s.t. $\mathbb{E}_{(\Omega,\mu)}(C_n)$ is smaller than $\mathrm{const}_n \mathrm{vol}(M)$ in the case of Theorem 2.2 and arbitrarily small in the case of Theorem 2.4, thus proving these theorems.

The construction of the latter will be explained in the toy case of $M = S^1$ in Subsection 3.4. A brief justification why (3.2) should hold follows next.

3.3. L^2-Betti numbers and actions on probability spaces. One would have to explain Damien Gaboriau's language of $\mathscr{R}$-simplicial complexes to give a proof of the estimate $b_i^{(2)}(M) \leq \mathbb{E}_{(\Omega,\mu)}(C_i)$. Instead, we want to at least point out that the L^2-Betti numbers of M can be computed by some sort of averaging over the probability space (Ω,μ). In Lück's algebraic definition averaging is reflected by interpreting $b_i^{(2)}(M)$ as the dimension of a certain induction of the homology of $\widetilde{M}$ with respect to a bigger von Neumann algebra, the so-called group measure construction of (Ω,μ) and Γ.

The *group measure space construction* $L^\infty(\Omega,\mu)\overline{\rtimes}\Gamma$ is defined as a completion of the algebraic crossed product $L^\infty(X) \rtimes \Gamma$ with respect to the trace

$$\mathrm{tr}\Big(\sum f_\gamma \gamma\Big) = \int_\Omega f_1(x)\,d\mu(x),$$

which is a sort of expected value. The group measure space construction contains the group von Neumann algebra $L(\Gamma)$ and $L^\infty(\Omega,\mu)$ as subalgebras. The crucial property is that

$$b_i^{(2)}(M) = \dim_{L^\infty(\Omega,\mu)\overline{\rtimes}\Gamma} H_i\big(L^\infty(\Omega,\mu)\overline{\rtimes}\Gamma \otimes_{\mathbb{Z}\Gamma} C_*(\widetilde{M})\big) \tag{3.3}$$

For a proof of $b_i^{(2)}(M) \le \mathrm{const}\,\mathbb{E}_{(\Omega,\mu)}(C_i)$ one would have to interpret the right hand side of (3.3) in Gaboriau's sense as L^2-Betti numbers of the $\mathscr{R}$-simplicial complex $\Omega\times\widetilde{M}$. For the better estimate (3.2) one needs a Poincare duality argument (see [**5**]).

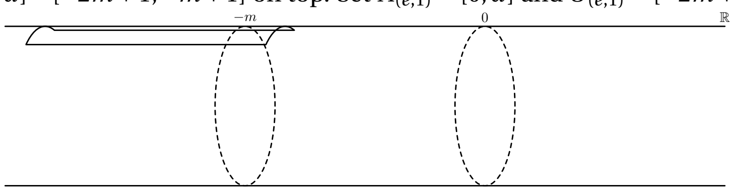
3.4. The toy case $M = S^1$. We want to outline the proof of Theorem 2.4, as presented in [**4**], for the example $M = S^1$. Of course, M itself is an amenable set, and we already know that its L^2-Betti numbers vanish. But we want to illustrate the construction of a $\mathbb{Z}$-invariant probability measure μ_ε on Ω such that the expected value $\mathbb{E}_{(\Omega,\mu_\varepsilon)}(C_i)$ is smaller than a given $\varepsilon > 0$.

Let $\Gamma = \mathbb{Z}$. For the index set I we take $I = \Gamma \times \{1,2\}$. The measure μ on $\Omega = \mathrm{map}(\widetilde{M},\Delta(I))$ will be obtained as the push-forward of the normalized Haar measure μ_{S^1} of S^1 under a certain Γ-equivariant map

$$\varphi_\varepsilon : S^1 \to \Omega.$$

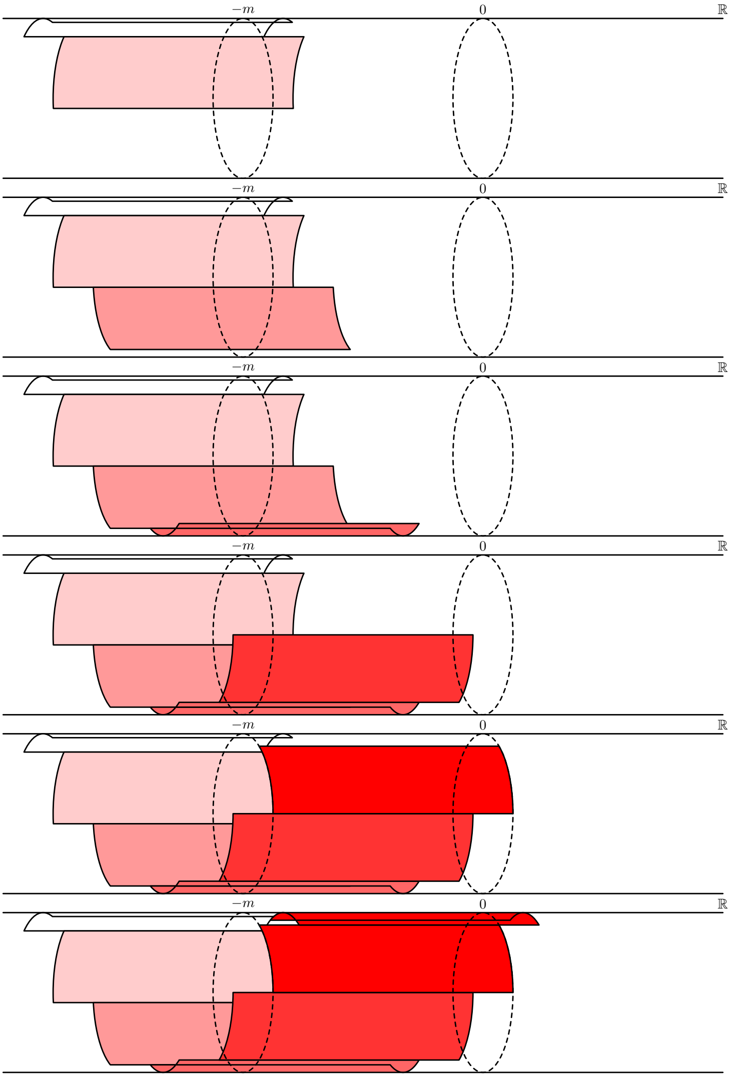
Let $m\in\mathbb{N}$ be larger than $2\varepsilon^{-1}$. Let $\alpha\in[0,1]$ be irrational with $0 < 1/m - \alpha < \frac{\varepsilon}{2m}$. Equip $S^1 = \mathbb{R}/\mathbb{Z}$ with the ergodic rotation given by addition of α. Next we define an equivariant cover $\mathscr{U} = \{A_i \times U_i\}_{i\in I}$ of $S^1\times\mathbb{R}$ such that $A_i \subset S^1$ are Borel sets and $U_i \subset \mathbb{R}$ are intervals of length m or 1. By definition, the map $\varphi_\varepsilon(z) : \mathbb{R}\to\Delta(I)$ is the nerve map associated to the cover $\{U_i;\ i\in I, z\in A_i\}$ for every $z\in S^1$.

To describe $\mathscr{U}$, consider the following picture(2) of $S^1\times\mathbb{R}$, where we see the tile $[0,\alpha]\times[-2m+1,-m+1]$ on top. Set $A_{(e,1)} = [0,\alpha]$ and $U_{(e,1)} = [-2m+1,-m+1]$.

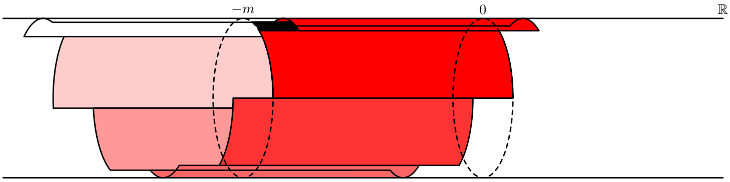


Next we consider the Γ-orbit $\{A_{(\gamma,1)}\times U_{(\gamma,1)}\}$ of the desribed tile in the following pictures.

(2) I am grateful to Clara Löh for programming the pictures in Metafont.



We almost obtain a partition of the cylinder $S^1 \times \mathbb{R}$ but because of $m\alpha < 1$ the translates do not quite close up after m steps. We have to introduce another tile (black in the picture) $[1 - m\alpha] \times [-m, -m+1]$ whose Γ-orbit $\{A_{(\gamma,2)} \times U_{(\gamma,2)}\}$ together with the orbit of the other tile partitions $S^1 \times \mathbb{R}$.



Finally we make the tiles just a little bit longer in the $\mathbb{R}$-direction to obtain the desired cover. We leave it to reader to verify that

$$\mathbb{E}_{(\Omega,(\varphi_\varepsilon)_*\mu_{S^1})}(C_1) < 1 - m\alpha + \alpha < \varepsilon.$$

3.5. Final remarks. In the actual proof of Theorems 2.2 and 2.4 one constructs suitable equivariant covers on the product of a Γ-probability space with $\widetilde{M}$, and then proceeds similarly as in Subsection 3.4 to obtain the desired probability measure on Ω. We want to mention the ingredients in the general case used to construct such covers.

In the case of Theorem 2.2 one can construct covers on $\widetilde{M}$ by balls of radius $0 < r < 1$ with multiplicity $< \mathrm{const}_n r^{-n}$ coming from maximal packings of concentric balls with smaller radii. This follows from the Bishop-Gromov inequality which provides packing inequalities in the presence of a lower Ricci curvature bound. In general, there is no way to obtain equivariant such covers. However, a suitable randomization in the sense of Subsection 3.2 of the problem of the existence of equivariant covers with small multiplicity can be solved, which leads to a proof of Theorem 2.2.

In the case of Theorem 2.4 one applies the generalized Rokhlin lemma from ergodic theory to construct covers similar to the one in the toy example over every of the amenable subsets and combines them to a cover on the product of a Γ-probablity space and $\widetilde{M}$.

References

[1] M. GROMOV – "Volume and bounded cohomology", *Inst. Hautes Études Sci. Publ. Math.* (1982), no. 56, p. 5–99 (1983).

[2] M. GROMOV – *Metric structures for Riemannian and non-Riemannian spaces*, Progress in Mathematics, vol. 152, Birkhäuser Boston Inc., Boston, MA, 1999, Based on the 1981 French original [MR0682063 (85e:53051)], With appendices by M. Katz, P. Pansu and S. Semmes, Translated from the French by Sean Michael Bates.

[3] W. LÜCK – *L^2-invariants: theory and applications to geometry and K-theory*, Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge. A Series of Modern Surveys in Mathematics [Results in Mathematics and Related Areas. 3rd Series. A Series of Modern Surveys in Mathematics], vol. 44, Springer-Verlag, Berlin, 2002.

[4] R. SAUER – "Amenable covers, volume and L^2-betti numbers of aspherical manifolds", to appear in J. Reine Angew. Math (Crelle).

[5] M. SCHMIDT – "L^2-betti numbers of $\mathscr{R}$-spaces and the integral foliated simplicial volume", Ph.D. Thesis, Universität Münster, 2005.

Mathematisches Institut, Courant-Colloquium Trends in Mathematics
(R. Meyer, T. Schick, Y. Tschinkel, eds.), p. 107–114
Georg-August Universität Göttingen, 2008

THE GLOBAL STRUCTURE OF AFFINE DELIGNE-LUSZTIG VARIETIES

Eva Viehmann
Mathematisches Institut der Universität Bonn, Beringstrasse 1, 53115 Bonn, Germany
E-mail: viehmann@math.uni-bonn.de

Abstract. We give an overview over current results on the global structure of affine Deligne-Lusztig varieties associated to a hyperspecial maximal compact subgroup. In particular, we discuss a formula for their dimensions and the set of connected components of the closed affine Deligne-Lusztig varieties.

1. Classical Deligne-Lusztig varieties

Deligne-Lusztig varieties were defined by Deligne and Lusztig in [**1**] as certain subvarieties of the flag manifold of a reductive group. They use finite coverings of these varieties to study representations of the reductive group over a finite field. Let us briefly recall their definition.

Let k be a finite field with $q = p^r$ elements and let $\overline{k}$ be an algebraic closure. We denote by $\sigma : x \mapsto x^q$ the Frobenius of $\overline{k}$ over k. Let G be a split connected reductive group over k and let A be a split maximal torus. Let B be a Borel subgroup containing A. Let W be the Weyl group of G. By the Bruhat decomposition its elements are in bijection with $B_{\overline{k}} \setminus G_{\overline{k}} / B_{\overline{k}}$.

For $w \in W$, the associated Deligne-Lusztig variety is

$$X_w = \{x \in G_{\overline{k}} / B_{\overline{k}} \mid x^{-1}\sigma(x) \in B_{\overline{k}} w B_{\overline{k}}\}.$$

2000 ***Mathematics Subject Classification.*** 20G25, 14G35, 14L05.

It is a smooth algebraic variety and equidimensional of dimension $l(w)$. The finite group $G(k)$ acts on X_w and hence also on its cohomology.

2. Affine Deligne-Lusztig varieties

For the definition of affine Deligne-Lusztig varieties one proceeds similarly, replacing the finite field k by a function field of characteristic p.

2.1. Definition and elementary properties. Let again k be a finite field with $q = p^r$ elements and $\overline{k}$ an algebraic closure. Let $F = k((t))$ and let $L = \overline{k}((t))$. Let $\mathcal{O}_F$ and $\mathcal{O}_L$ be the valuation rings. We denote by $\sigma : x \mapsto x^q$ the Frobenius of $\overline{k}$ over k and also of L over F.

Let G be a split connected reductive group over $\mathcal{O}_F$ and let A be a split maximal torus. Let B be a Borel subgroup containing A. Let $K = G(\mathcal{O}_L)$. In general one can also define affine Deligne-Lusztig varieties associated to other parahoric subgroups K of $G(L)$, another interesting case is the case of an Iwahori subgroup. Here we restrict our attention to K a hyperspecial maximal compact subgroup. Instead of subvarieties of the flag manifold, we consider now subschemes of the affine Grassmannian $X = G(L)/K$. It is an ind-scheme over $\mathrm{Spec}(\overline{k})$. The Bruhat decomposition used to define classical Deligne-Lusztig varieties is replaced by the Cartan decomposition. Let $X_*(A) = \mathrm{Hom}(\mathbb{G}_m, A)$. An element $\mu \in X_*(A)$ is called dominant if $\langle \alpha, \mu \rangle \geqslant 0$ for all positive roots α of G. Then the Cartan decomposition takes the form

$$G(L) = \bigcup_{\mu \in X_*(A) \text{ dominant}} K t^{\mu} K.$$

Example 1. We illustrate the above notions for the special case $G = GL_n$. For A we choose the diagonal torus and for B the subgroup of upper triangular matrices. Then $X_*(A) \cong \mathbb{Z}^n$, a tuple $\mu = (\mu_i)$ corresponds to the morphism mapping $x \in L^\times$ to the diagonal matrix with entries x^{μ_i} on the diagonal. The element μ is dominant (with respect to our choice of B) if $\mu_i \geqslant \mu_{i+1}$ for all $i < n$. The n-tuple μ can be visualized as the graph of the continous piecewise linear function $[0, n] \to \mathbb{R}$ mapping 0 to 0 and with slope μ_i on $[i-1, i]$. It is called the polygon associated to μ.

For $b \in G(L)$ and a dominant coweight $\mu \in X_*(A)$ the affine Deligne-Lusztig variety $X^G_\mu(b) = X_\mu(b)$ is the locally closed reduced $\overline{k}$-subscheme of X defined by

$$X_\mu(b)(\overline{k}) = \{g \in G(L)/K \mid g^{-1} b \sigma(g) \in K t^{\mu} K\}.$$

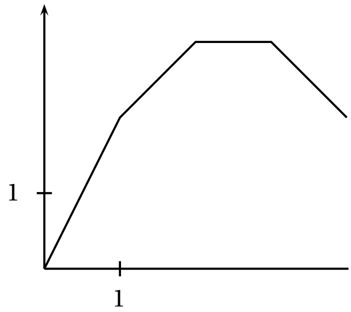


FIGURE 1. The polygon associated to $(2,1,0,-1)$

For dominant elements $\mu, \mu' \in X_*(A)$ we say that $\mu' \preceq \mu$ if $\mu - \mu'$ is a non-negative linear combination of positive coroots. The closed affine Deligne-Lusztig variety is the closed reduced subscheme of X defined by

$$X_{\preceq\mu}(b) = \bigcup_{\mu' \preceq \mu} X_{\mu'}(b).$$

Both $X_\mu(b)$ and $X_{\preceq\mu}(b)$ are locally of finite type.

Example 1. (continued) Let $G = GL_n$ and $\mu, \mu' \in X_*(A)$ dominant. Then $\mu \preceq \mu'$ if the polygon of μ lies below the polygon of μ' and if they have the same endpoint. In Figure 1, the set of $\mu \preceq (2,1,0,-1)$ consists of the three elements $(2,0,0,0)$, $(1,1,1,-1)$, and $(1,1,0,0)$.

Let

$$J = \{g \in G(L) \mid g \circ b\sigma = b\sigma \circ g\}.$$

Then there is a canonical J-action on $X_\mu(b)$ for each μ. The group J is the group of F-valued points of a reductive group over F which is an inner form of a Levi subgroup of G (compare [**5**], [**10**] 1.12, [**6**]).

Example 2. Let us consider the case where $b = t^\mu$ is central in G. It is one of very few cases where one can explicitly compute the affine Deligne-Lusztig variety. We obtain

$$\begin{aligned} X_\mu(b) &= \{g \in X \mid g^{-1}\sigma(g) \in K\} \\ &\cong \{g \mid g^{-1}\sigma(g) = 1\} \\ &= G(F)/G(\mathscr{O}_F). \end{aligned}$$

Here we used that every element of K can be written as $k^{-1}\sigma(k)$ for some element $k \in K$. For this b, the group J is equal to $G(F)$, thus J acts transitively on $X_\mu(b)$. Note that for general b the orbits of J on $X_\mu(b)$ are still zero-dimensional, in particular the action is not transitive.

By $[x]$ we denote the σ-conjugacy class of an element $x \in G(L)$. Left multiplication by $g \in G(L)$ induces an isomorphism between $X_\mu(b)$ and $X_\mu(gb\sigma(g)^{-1})$. Thus the isomorphism class of the affine Deligne-Lusztig variety only depends on $[b]$ and not on the representative b itself. This also explains why the element b did not occur in the classical situation: Over a finite field, every element is σ-conjugate to the identity.

The σ-conjugacy classes in $G(L)$ are classified by Kottwitz in [**5**] and [**6**]. The σ-conjugacy class of some b is determined by two invariants. The first is its Newton vector, a dominant element of $X_*(A)_{\mathbb{Q}}$. The second is an element $\kappa_G(b)$ of $\pi_1(G)$. Here $\pi_1(G)$ is the quotient of $X_*(A)$ by the coroot lattice of G. Let U be the unipotent radical of B. Then κ_G is the locally constant map $X \to \pi_1(G)$ mapping Ut^xK to the class of $x \in X_*(A)$ in $\pi_1(G)$.

Example 1. (continued) For $G = GL_n$, the Newton point $\nu = (\nu_i) \in \mathbb{Q}_+^n$ of b has the following elementary definition. We consider $b\sigma$ as a semilinear map $L^n \to L^n$. There are $m_i \in \mathbb{N}$ and $h_i \in \mathbb{N} \setminus \{0\}$ for $i = 1, \dots, n$ and a basis (x_i) of L^n such that

$$(b\sigma)^{h_i}(x_i) = t^{m_i} x_i.$$

We may assume that $\frac{m_i}{h_i} \geqslant \frac{m_{i+1}}{h_{i+1}}$ for all $i < n$. Then ν is given by $\nu_i = \frac{m_i}{h_i}$. For $G = GL_n$ we further have $\pi_1(G) \cong \mathbb{Z}$ and $\kappa_G(g) = v_t(\det g)$ where v_t denotes the usual valuation on $L = \overline{k}((t))$.

Remark 3. Using the same definition as above one can also define $X_\mu(b)$ as a set of points for $\mathbb{Q}_p$ instead of the function field $F = k((t))$. In this situation a scheme structure on $X_\mu(b)$ is not known in general. However, if $G = GL_n$ or GSp_{2n} and if μ is minuscule, then $X_\mu(b)$ can be identified with the set of $\overline{\mathbb{F}}_p$-valued points of a moduli space of p-divisible groups defined by Rapoport and Zink in [**10**].

2.2. Results on the global structure. In [**4**] Kottwitz and Rapoport give a criterion for $X_\mu(b)$ to be nonempty, compare also [**2**], Proposition 5.6.1. Then $\kappa_G(b) = \mu$ and $\mu - \nu$ is a positive linear combination of positive coroots with rational coefficients. From now on we only consider nonempty affine Deligne-Lusztig varieties.

The dimension of affine Deligne-Lusztig varieties is given by the following theorem. The formula has been conjectured by Rapoport in [**11**], Conjecture 5.10 in a different form and has been reformulated by Kottwitz in [**8**].

Theorem 4. *Let $X_\mu(b) \neq \emptyset$. Then*

$$\dim X_\mu = \dim X_{\leq\mu} = \langle \rho, \mu - \nu\rangle - \frac{1}{2}(\mathrm{rk}_F(G) - \mathrm{rk}_F(J)). \tag{1}$$

Here, ρ is the half-sum of the positive roots of G, and rk_F *denotes the rank of a maximal F-split torus of the corresponding group.*

Example 1. (continued) For $G = GL_h$, the nonemptyness of the affine Deligne-Lusztig variety is equivalent to the condition that the polygon associated to μ lies above the one associated to ν and that both have the same endpoint $(h, v_t(\det(b)))$. Then the right hand side of (1) is equal to the number of lattice points above ν and on or below μ. Note that $\langle \rho, \mu - \nu\rangle$ is in this case equal to the area between the two polygons.

Theorem 4 is proved by Görtz, Haines, Kottwitz, Reuman, and the author in the two articles [**2**] and [**13**]. The proof consists of two main steps.

Step 1. The theorem holds for the case that b is superbasic, i. e. no σ-conjugate of b is contained in a proper Levi subgroup of G.

Step 2. If the theorem holds for superbasic b, then it also holds in general.

The two parts of this subdivision correspond to the two articles cited above: Step 2 is carried through in [**2**], Step 1 in [**13**]. The proofs of the two steps use completely different methods. For Step 1 one can show (see [**2**], 5.9) that superbasic b essentially only occur if $G = GL_n$. Then one uses explicit σ-linear algebra to prove the theorem. The proof of Step 2 relates the dimension of affine Deligne-Lusztig varieties to the dimensions of certain orbit intersections which are known thanks to Mirkovic-Vilonen [**9**].

Rapoport also conjectures that the affine Deligne-Lusztig varieties are equidimensional. This part of the conjecture is still an open question.

For the closed affine Deligne-Lusztig varieties $X_{\le\mu}(b)$, we also know the set of connected components (see [**12**]). A first rough result is

Proposition 5. *The group J acts transitively on the set of connected components of $X_{\le\mu}(b)$.*

Again an important ingredient in the proof of this result is to consider the same Steps 1 and 2 as in the proof of Theorem 4.

Using this one can explicitly compute $\pi_0(X_{\le\mu}(b))$. To obtain a simpler formula, we first show that it is enough to consider data (G,μ,b) of a special form, using the Hodge-Newton decomposition. The Hodge-Newton decomposition is first shown in Katz's paper [**3**] for isocrystals with a σ^a-linear endomorphism. Kottwitz (see [**7**]) generalizes this to a result about affine Deligne-Lusztig varieties associated to any unramified reductive group (where Katz's result corresponds to the case of GL_n). His proof yields the following result.

Theorem 6. *Let $P = MN \subseteq G$ be a standard parabolic subgroup with $P_b \subseteq P$. If $\kappa_M(b) = \mu$, then the morphism $X_\mu^M(b) \hookrightarrow X_\mu^G(b)$ is an isomorphism.*

Here P_b is the standard parabolic subgroup of G whose Levi component is the centralizer of ν.

Example 1. (continued) We consider again the case $G = GL_n$. Standard parabolic subgroups P in GL_n correspond to ordered partitions $n = n_1 + n_2 + \cdots + n_l$ of n. The subgroup P_b corresponds to the partition associated to the first coordinates of the breakpoints of ν. Thus the condition $P_b \subseteq P$ is equivalent to $n_1, n_1+n_2, \dots, n_1+\cdots+n_{l-1}$ corresponding to breakpoints of ν. The condition $\kappa_M(b) = \mu$ means that each of these breakpoints of the Newton polygon ν also lies on the Hodge polygon μ. Thus in this case the conditions in Theorem 6 are the same as Katz's conditions.

We call a pair (μ, b) indecomposable with respect to the Hodge-Newton decomposition if for all standard parabolic subgroups P with $P_b \subseteq P = MN \subsetneq G$ we have $\kappa_M(b) \neq \mu$. Given G, μ, and b, we may always pass to a Levi subgroup M of G in which (μ, b) is indecomposable. For a description of the affine Deligne-Lusztig varieties it is therefore sufficient to consider pairs (μ, b) which are indecomposable with respect to the Hodge-Newton decomposition.

Let G_{ad} be the adjoint group of G. We denote the images of b and μ in G_{ad} also by b and μ. Then the sets of connected components of $X^G_{\preceq\mu}(b)$ and $X^{G_{\mathrm{ad}}}_{\preceq\mu}(b)$ can easily be computed from one another. The closed affine Deligne-Lusztig variety $X^{G_{\mathrm{ad}}}_{\preceq\mu}(b)$ is the product of closed affine Deligne-Lusztig varieties corresponding to the simple factors of G^{ad}. Hence it is enough to describe the set of connected components in the case that G is simple.

Theorem 7. *Let G, μ, and b be as above and indecomposable with respect to the Hodge-Newton decomposition. Assume that G is simple.*

1. *Either $\kappa_M(b) \neq \mu$ for all proper standard parabolic subgroups P of G with $b \in M$ or $[b] = [t^\mu]$ with t^μ central.*
2. *In the first case, κ_G induces a bijection $\pi_0(X_{\preceq\mu}(b)) \cong \pi_1(G)$.*
3. *In the second case, $X_\mu(b) = X_{\preceq\mu}(b) \cong J/(J\cap K) \cong G(F)/G(\mathcal{O}_F)$ is discrete.*

Note that (3) is the case considered in Example 2.

For the locally closed affine Deligne-Lusztig varieties $X_\mu(b)$ the set of connected components seems to be more difficult to compute. There are examples where J does not act transitively on $\pi_0(X_\mu(b))$, and also examples where it acts transitively, but where an assertion analogous to Theorem 7 (2) still does not hold.

References

[1] P. Deligne & G. Lusztig – “Representations of reductive groups over finite fields”, *Ann. of Math. (2)* **103** (1976), no. 1, p. 103–161.

[2] U. Görtz, T. J. Haines, R. E. Kottwitz & D. C. Reuman – “Dimensions of some affine Deligne-Lusztig varieties”, *Ann. Sci. École Norm. Sup. (4)* **39** (2006), no. 3, p. 467–511.

[3] N. M. Katz – “Slope filtration of F-crystals”, Journées de Géométrie Algébrique de Rennes (Rennes, 1978), Vol. I, Astérisque, vol. 63, Soc. Math. France, Paris, 1979, p. 113–163.

[4] R. Kottwitz & M. Rapoport – “On the existence of F-crystals”, *Comment. Math. Helv.* **78** (2003), no. 1, p. 153–184.

[5] R. E. Kottwitz – “Isocrystals with additional structure”, *Compositio Math.* **56** (1985), no. 2, p. 201–220.

[6] ______, “Isocrystals with additional structure. II”, *Compositio Math.* **109** (1997), no. 3, p. 255–339.

[7] ______, “On the Hodge-Newton decomposition for split groups”, *Int. Math. Res. Not.* (2003), no. 26, p. 1433–1447.

[8] ______, “Dimensions of Newton strata in the adjoint quotient of reductive groups”, *Pure Appl. Math. Q.* **2** (2006), no. 3, p. 817–836.

[9] I. MIRKOVIĆ & K. VILONEN – "Perverse sheaves on affine Grassmannians and Langlands duality", *Math. Res. Lett.* **7** (2000), no. 1, p. 13–24.

[10] M. RAPOPORT & T. ZINK – *Period spaces for p-divisible groups*, Annals of Mathematics Studies, vol. 141, Princeton University Press, Princeton, NJ, 1996.

[11] M. RAPOPORT – "A guide to the reduction modulo p of Shimura varieties", *Astérisque* (2005), no. 298, p. 271–318, Automorphic forms. I.

[12] E. VIEHMANN – "Connected components of affine Deligne-Lusztig varieties", to appear in Mathematische Annalen.

[13] ______, "The dimension of some affine Deligne-Lusztig varieties", *Ann. Sci. École Norm. Sup. (4)* **39** (2006), no. 3, p. 513–526.

Mathematisches Institut, Courant-Colloquium Trends in Mathematics
(R. MEYER, T. SCHICK, Y. TSCHINKEL, eds.), p. 115–140
Georg-August Universität Göttingen, 2008

LIE II THEOREM FOR LIE ALGEBROIDS VIA STACKY LIE GROUPOIDS

Chenchang Zhu

Mathematisches Institut, Universität Göttingen, Bunsenstr. 3-5, 37073 Göttingen
E-mail : chenchang.zhu@gmail.com

***Abstract*.** Unlike Lie algebras which one-to-one correspond to simply connected Lie groups, Lie algebroids (integrable or not) one-to-one correspond to a sort of étale stacky groupoids (W-groupoids). Following Sullivan's spacial realization of a differential algebra, we construct a canonical integrating Lie 2-groupoid for every Lie algebroid. Finally we discuss how to lift Lie algebroid morphisms to W-groupoid morphisms (Lie II). Examples of Poisson manifolds and symplectic stacky groupoids are provided. This paper contains essentially some ideas of proofs and examples, for a complete treatment please refer to **[29]** which also proves some connectedness result.

1. Introduction

Lie II theorem for Lie algebras studies how to lift morphisms of Lie algebras to morphisms of Lie groups. A Lie algebroid is the infinitesimal data of a Lie groupoid, as a Lie algebra for a Lie group. More precisely, for us, a Lie algebroid over a manifold M is a vector bundle $\pi : A \to M$ with a real Lie bracket $[,]$ on its space of sections $H^0(M, A)$ and a bundle map $\rho : A \to TM$ such that the Leibniz rule

$$[X, fY](x) = f(x)[X, Y](x) + (\rho(X)f)(x)Y(x)$$

holds for all $X, Y \in H^0(M, A)$, $f \in C^\infty(M)$ and $x \in M$. Hence when M is a point, a Lie algebroid becomes a Lie algebra. Also a tangent bundle $TM \to M$ is certainly a Lie algebroid with $[,]$ the Lie bracket of vector fields. The next example is a Poisson

---

2000 ***Mathematics Subject Classification*.** 58H99.

manifold P with $A = T^*P \to P$ and $[df, dg] = d\{f, g\}$ determined by the Poisson bracket $\{,\}$ (see the book [**21**] for a friendly introduction).

Thus Lie II theorem for Lie algebroids studies how to lift an infinitesimal morphism on the level of Lie algebroids to a global morphism. Its version having Lie groupoids as the global objects of Lie algebroids is well-known [**13**] [**16**]. However, unlike (finite dimensional) Lie algebras which always have their associated Lie groups, Lie algebroids do not always have their associated Lie groupoids [**2**] [**1**]. The complete integrability criteria is given in a remarkable work of Crainic and Fernandes [**6**]. But we claim the situation is not totally unsavable: if we are willing to enter the world of stacks, we do have the full one-to-one correspondence (Lie III) parallel to the classical one of Lie algebras [**24**],

$$\boxed{\text{Lie algebras}} \xleftrightarrow[\text{integration}]{\text{differentiation at identity}} \boxed{\text{Lie groups}}$$

$$\boxed{\text{Lie algebroids}} \xleftrightarrow[\text{integration}]{\text{differentiation at identity}} \boxed{\begin{array}{c}\text{“étale stacky Lie}\\ \text{groupoids”}\end{array}}$$

Here an *étale stacky Lie groupoid* $\mathscr{G} \Rightarrow M$ (which we also call *W-groupoid* for its existence is first conjectured by Weinstein [**26**] [**4**]) is a groupoid in the category of differentiable stacks with $\mathscr{G}$ an étale stack and M a manifold (see [**29**] for the exact definition), and we call the procedure of passing from an infinitesimal object to a global object “integration”. This problem is already very interesting, as shown by Cattaneo and Felder [**5**], in the case of Poisson manifolds: the object integrating $A = T^*P$ is the phase space of Poisson sigma model and it is further a *symplectic* W-groupoid [**23**].

Therefore my effort in this paper is to study the functoriality of this slightly wild W-groupoid, for example how to integrate a morphism $A \to B$ of Lie algebroids to a global morphism from a universal stacky groupoid of A to any stacky groupoid of B (Lie II). The results of the paper are positive. In case A is integrable[(1)], there is a unique source-simply connected Lie groupoid integrating A, which generalizes the corresponding theory of Lie algebras. However, in the general case, there are two “universal” étale stacky groupoids $\mathscr{G}(A)$ and $\mathscr{H}(A)$ associated with a Lie algebroid A, (see Section 2.1 or [**24**]). As shown in [**28**],

---

(1) That is A is the infinitesimal data of a Lie groupoid.

Theorem 1.1. *$\mathscr{G}(A)$ and $\mathscr{H}(A)$ are source-connected and simply connected and $\mathscr{G}(A)$ is furthermore source-2 connected, which means its source fibres have trivial homotopy groups $\pi_{\leqslant 2}$.*

This tells us that $\mathscr{G}(A)$ is more universal than $\mathscr{H}(A)$ hence we should expect a Lie II theorem using $\mathscr{G}(A)$. In fact A is integrable if and only if $\mathscr{H}(A)$ (*not* $\mathscr{G}(A)$) is representable. Using stacky groupoids, one should expect one further degree of connectedness. A simple example is the algebroid of the Poisson (even symplectic) manifold S^2, $A = T^*S^2$. In this case, A is integrable and $\mathscr{H}(A) = S^2 \times S^2$, but $\mathscr{G}(A) = \tilde{S}^2 \times \tilde{S}^2 / B\mathbb{Z}$. Here $\tilde{S}^2$ is the $B\mathbb{Z}$ gerbe on S^2 presented by the action groupoid $S^3 \times \mathbb{R} \Rightarrow S^3$ with $\mathbb{R}$ acting via the projection $\mathbb{R} \to S^1$ and the usual Hopf S^1 action on S^3. Analogously to simply-connected coverings, $\tilde{S}^2$ is the π_2-trivial covering of S^2 (See Example 4.1). Hence even with simple objects as S^2 we could expect further more interesting examples of $\mathscr{G}(A)$ to appear. Moreover the property of $\pi_2 = 0$ might also appeal to symplectic geometers.

For every Lie algebroid A, (notice that tangent bundles are Lie algebroids), we associate A a simplicial set $S(A) = [\dots S_2(A) \Rrightarrow S_1(A) \Rightarrow S_0(A)]$ with,

$$S_i(A) = \hom_{algd}(T\Delta^i, A) := \{\text{Lie algebroid morphisms } T\Delta^i \xrightarrow{f} A\}. \tag{1}$$

Here Δ^i is the i-dimensional standard simplex viewed as a smooth Riemannian manifold with boundary, hence it is isomorphic to the i-dimensional closed ball. Then the facial and degeneracy maps are induced by pullbacks of the tangent maps of natural maps $d_k : \Delta^{i-1} \to \Delta^i$ and $s_k : \Delta^i \to \Delta^{i-1}$. With this language, we also understand [**5**] and [**6**] in a fresh way: $S_1(A)$ and $S_2(A)$ are the space of fields and Hamiltonian symmetries respectively in Poisson sigma model [**5**] in the case of Poisson manifolds; or the space of A-paths $P_a A$ and of A-homotopies [**6**] respectively in general (see Section 2).

The simplicial set $S(A)$ is not entirely unknown to us: in the case of a Lie algebra $\mathfrak{g}$, let $\Omega^1(\Delta^n, \mathfrak{g})$ be the space of $\mathfrak{g}$-valued 1-form on Δ^n, then we have[(2)],

$$\begin{aligned}\hom_{algd}(T\Delta^n, \mathfrak{g}) =& \{\alpha \in \Omega^1(\Delta^n, \mathfrak{g}) | d\alpha = \frac{1}{2}[\alpha, \alpha]\} \\ =& \{\text{flat connections on the trivial } G\text{-bundle } G \times \Delta^n \to \Delta^n\},\end{aligned}$$

where G is a Lie group of $\mathfrak{g}$.

In fact for a differential algebra D, Sullivan [**22**] constructed a spatial realization of D, which is defined as the space of differential graded maps $\hom_{d.g.a.}(D, \Omega^*(\Delta^n))$,

(2) More precisely it is $-\alpha$ that is the connection.

and the simplicial set $S(A)$ is a parallel construction on the geometric side. Sullivan's construction also appears in the work of Ševera [**20**] to integrate a non-negatively graded super-manifold with a degree 1 vector field of square 0 (NQ-manifold), which was the original suggestion to our simplicial set $S(A)$. In fact a Lie algebroid can be viewed as a "degree 1" NQ-manifold, a "degree 2" symplectic NQ-manifold is a Courant algebroid [**12**] [**19**] which is now widely used in generalized complex geometry [**9**]. This construction also appears in the work of Getzler [**8**] and Henriques [**10**] to integrate an L_∞-algebra L, where the simplicial set is $\hom_{d.g.a.}(C^*(L), \Omega^*(\Delta^n))$ with $C^*(L)$ the Chevalley-Eilenberg cochains on L. This simplicial set is further proved to be a Kan simplicial manifold for nilpotent L_∞-algebras in [**8**] and for all L_∞-algebras in general in [**10**].

However it is not obvious that $S(A)$ is a simplicial manifold let alone a Kan simplicial manifold. Hence its 2-truncation being a Lie 2-groupoid is not immediate but proved in,

Theorem 1.2. *Given a Lie algebroid A, the 2-truncation of the simplicial set $S(A)$,*

$$S_2(A)/S_3(A) \Rrightarrow S_1(A) \rightrightarrows S_0(A),$$

is a Lie 2-groupoid that corresponds to the W-groupoid $\mathscr{G}(A)$ constructed in [**24**] *under the correspondence of Theorem 1.3 of* [**29**].

In this theorem $S_{\geqslant 1}(A)$'s are infinite dimensional spaces. As a respond of a question by Getzler and Roytenburg, it turns out that it is not necessary to take everything in the infinite dimensional space and we treat in this manner elsewhere [**28**]: the spirit is that this Lie 2-groupoid is Morita equivalent to a finite dimensional Lie 2-groupoid, arising in a fashion of local Lie groupoids of Pradines, $E \Rrightarrow P \rightrightarrows M$, where $\dim P = \dim A$ and $\dim E = 2\dim A - \dim M$ (See Remark 2.11).

Finally, after recognizing the existence of this more universal W-groupoid $\mathscr{G}(A)$, we have the expected

Theorem 1.3 (Lie II for Lie algebroids). *Let φ be a morphism of Lie algebroids $A \to B$, $\mathscr{G}$ a W-groupoid whose algebroid is B. Then up to 2-morphisms, there exists a unique morphism Φ of W-groupoids $\mathscr{G}(A) \to \mathscr{G}$ such that Φ induces the Lie algebroid morphism $\varphi : A \to B$.*

We prove this Theorem here for stacky groupoid and leave the treatment using Kanification of simplicial manifolds in [**28**].

Acknowledgments: I thank Alan Weinstein for his continuing guidance and support. I thank Andre Henriques for much helpful discussion and for communicating preliminary versions of his work on string groups. I am also very grateful to Ezra Getzler, Kirill Mackenzie, Pavel Ševera for very helpful discussions.

2. Lie 2-groupoids associated to Lie algebroids

In this section we construct a Lie 2-groupoid from a Lie algebroid A and prove this Lie 2-groupoid corresponds to the universal W-groupoid $\mathscr{G}(A)$ of A constructed in [**24**].

2.1. Construction of the universal W-groupoids. We first recall [**6**] that an A-*path* of a Lie algebroid A is a C^1 path $a(t) : \Delta^1 \to A$ with base map $\gamma(t) : \Delta^1 \to M$ satisfying

$$\rho(a(t)) = \dot{\gamma}(t), \tag{2}$$

where $\dot{\square}$ denotes the derivative of t. If it further satisfies boundary condition $a(0) = a(1) = \dot{a}(0) = \dot{a}(1) = 0$, we call it an A_0-path as in [**24**]. This notation is useful for technical reasons. We call $P_a A$ and $P_0 A$ the space of A-paths and A_0-paths respectively.

Remark 2.1. As shown in [**6**] and [**25**] respectively, they are infinite dimensional smooth manifolds, more precisely they process a structure of Banach manifolds, which we refer to [**11**] for the definition and properties. There are certain subtle differences comparing to the usual finite dimensional manifolds and we state here what we will use in this paper. A morphism $f : X \to Y$ between Banach manifolds is called a *submersion* if at every point $x \in X$ there exists a Banach chart (U, φ) and a Banach chart (V, ψ) at $f(x)$ such that φ gives an isomorphism of U to a product $U_1 \times V$ of open sets in some Banach spaces, and such that the map

$$\psi f \varphi^{-1} : U_1 \times V \to V$$

is the projection to the second factor. If $f : X \to Z$ is a submersion, then for any map $Y \to Z$ the fibre product $X \times_Z Y$ is a Banach manifold and the pull-back map $X \times_Z Y \to Y$ is again a submersion. We refer the reader to [**11**, Chapter II.2] for the proofs.

Moreover when we have a foliation $\mathscr{F}$ on a Banach manifold X, we can also form the *monodromy groupoid* of this foliation: The objects are points in the manifold, and arrows are paths within a leaf (up to homotopies) with fixed end points inside the leaf. Then we have the following lemma:

Lemma 2.2. *When a closed foliation $\mathscr{F}$ on a Banach manifold X has finite constant codimension*(3)*, the monodromy groupoid is a Banach groupoid over X, namely a groupoid where the space of arrows (not necessarily Hausdorff*(4)*) and the space of objects are Banach manifolds, all the structure maps are smooth morphisms between Banach manifolds, and the source and target are surjective submersions.*

Proof. By Frobenius Theorem for Banach manifolds [**11**, Chapter VI], we have also foliation charts $\{(h_i, U_i)\}_{i\in I}$ on X, namely charts with the property that, for each (i, j), the change of coordinates $h_j \circ h_i^{-1} : \mathbb{R}^k \times F \to \mathbb{R}^k \times F$ has the form

$$(h_j \circ h_i^{-1})(x, y) = (\varphi(x), \psi(x, y)),$$

where k is the codimension of the foliation and F is a Banach chart of the leaves of $\mathscr{F}$. Then the rest follows similarly as the proof of the finite dimensional case as in [**18**]. A typical Banach chart for the space of arrows is $\mathbb{R}^k \times F \times F$. Under these charts, the source and target maps are simply projections to $\mathbb{R}^k \times F$, hence they are submersions(5). □

There is an equivalence relation in $P_a A$, called A-homotopy [**6**].

Definition 2.3. Let $a(t, s)$ be a family of A-paths which is C^2 in s. Assume that the base paths $\gamma(t, s) := \rho \circ a(t, s)$ have fixed end points. For a connection ∇ on A, consider the equation

$$\partial_t b - \partial_s a = T_\nabla(a, b), \quad b(0, s) = 0. \tag{3}$$

Here T_∇ is the torsion of the connection defined by

$$T_\nabla(\alpha, \beta) = \nabla_{\rho(\beta)}\alpha - \nabla_{\rho(\alpha)}\beta + [\alpha, \beta].$$

Two paths $a_0 = a(0, \cdot)$ and $a_1 = a(1, \cdot)$ are homotopic if the solution $b(t, s)$ satisfies $b(1, s) = 0$.

(3) Here we only need the foliation to form a subbundle in the sense of [**11**, Chapter II]. Finite codimensionness guarantees this.

(4) To simplify our notation, all the manifolds in this paper are Hausdorff unless specially mentioned as here.

(5) It is obvious that they are surjective.

A-homotopies generate a closed foliation $\mathscr{F}$ of finite constant codimension on P_aA. Now the idea is to consider the monodromy groupoid $Mon(P_aA) \Rightarrow P_aA$ of this foliation. One could think of $Mon(P_aA)$ as the space of A-homotopies. The two maps from $Mon(P_aA)$ to P_aA assign to each A-homotopy the two paths at the ends. There are also two maps $P_aA \Rightarrow M$ which assign to each A-path its two end points respectively. Very similar to this, P_0A has also a foliation $\mathscr{F}_0$ by A-homotopies and the monodromy groupoid $Mon(P_0A)$ according to this foliation. In fact $\mathscr{F}_0$ is the restriction of $\mathscr{F}|_{P_0A}$, that is if two A_0-paths are A-homotopic in P_aA, then they are A-homotopic in P_0A. Moreover $Mon(P_aA) \Rightarrow P_aA$ is Morita equivalent to $Mon(P_0A) \Rightarrow P_0A$.

Sometimes to avoid dealing with infinite dimensional issues, we consider a variant $\Gamma \Rightarrow P$ of this groupoid obtained as follows (see for example [**24**] for details): take an open cover of P_aA, then P is the disjoint union of slices P_i over this cover that are transversal to the foliation $\mathscr{F}$. Then P is a smooth manifold, and the pullback groupoid by $P \to P_aA$, which we denote as $\Gamma \Rightarrow P$, is a finite dimensional Lie groupoid. What's even better is that it's an étale groupoid (i.e. the source and target are étale). The two groupoids $Mon(P_aA) \Rightarrow P_aA$ and $\Gamma \Rightarrow P$ are in fact Morita equivalent. Also, there are still two maps $P \Rightarrow M$.

The next step is clear: we take the quotient as stacks $[P_0A/Mon(P_0A)]$ and construct a stacky groupoid $\mathscr{G}(A) := [P_0A/Mon(P_0A)] \Rightarrow M$ where the two maps are endpoint maps. As in recent work [**3**], there is a 1-1 correspondence between Lie groupoids up to Morita equivalence and differentiable stacks. The Lie groupoid corresponds to a differentiable stack is called a *groupoid presentation*. Along these lines, $\mathscr{G}(A)$ has its groupoid presentation $Mon(P_0A) \Rightarrow P_0A$ or Morita equivalently $\Gamma \Rightarrow P$. Hence it is an étale differentiable stack. From the technical viewpoint, we take $\mathscr{G}(A)$ as an étale groupoid from now on, hence avoid infinite dimensional analysis. Moreover, the two maps from P_0A to M descend to the quotient, giving two maps $\mathbf{s}, \mathbf{t} : \mathscr{G}(A) \to M$. There are other maps: we define a multiplication $m : \mathscr{G} \times_{\mathbf{s},\mathbf{t}} \mathscr{G} \to \mathscr{G}$ by concatenation $\odot$ of paths, namely,

$$(a \odot b)(t) = \begin{cases} 2a(2t) & \text{when } t \in [0,1/2], \\ 2b(2t-1) & \text{when } t \in [1/2,1]; \end{cases} \tag{4}$$

we define an inverse $i : \mathscr{G} \to \mathscr{G}$, by reversing the orientation of a path; we define an identity section $e : M \to \mathscr{G}$ by considering constant paths. These maps are defined in detail in [**24**]. There, we prove that this makes $\mathscr{G} \Rightarrow M$ into a W-groupoid. A similar procedure using the holonomy groupoid $Hol(P_0A) \Rightarrow P_0A$ produces another natural W-groupoid $\mathscr{H}(A)$.

What is less obvious is to go back to a Lie algebroid from a W-groupoid $\mathscr{G} \Rightarrow M$. For this, we first have the following technical lemma (proved in [**29**, Section 3]),

Lemma 2.4. *For an immersion $\bar{e} : M \to \mathscr{G}$ from a manifold M to an étale stack $\mathscr{G}$, there is an étale chart G_0 of $\mathscr{G}$ such that $\bar{e}$ lifts to an embedding $e : M \to G_0$. We call such charts* good *charts and their corresponding groupoid presentations* good *presentations.*

Then given a W-groupoid $\mathscr{G} \Rightarrow M$, there is a neighborhood $U \subset G_0$ of M such that all the stacky groupoid structure maps descend to U and make $U \Rightarrow M$ a local Lie groupoid [**24**, Section 5][(6)], which resembles a Lie groupoid but multiplication and inverse are defined only locally. The structure of the local groupoid does not depend on the choice of U. We call it the local Lie groupoid of $\mathscr{G}$ and denote it by G_{loc}. The Lie algebroid of G_{loc} is defined to be *the Lie algebroid of W-groupoid $\mathscr{G}$*.

Remark 2.5. The étale chart $P = \sqcup_i P_i$ of $\mathscr{G}(A)$, made up by local transversals P_i of the foliation $\mathscr{F}$ on P_aA, is usually not a good étale chart directly.

We recall the construction of the local groupoid $G_{loc}(A)$ of A in [**6**]. Take a small open neighborhood $\mathscr{O}$ of M in the A-path space P_aA so that the foliation $\mathscr{F}$ restricted to $\mathscr{O}$ which we denote by $\mathscr{F}|_{\mathscr{O}}$ has good transversal sections, namely leaves in $\mathscr{F}|_{\mathscr{O}}$ intersect each transversal section only once. Then the quotient $\mathscr{O}/(\mathscr{F}|_{\mathscr{O}})$ is a local Lie groupoid over M, which is exactly the local Lie groupoid of $\mathscr{G}(A)$ as above (proven in [**24**, Section 5]).

The open set $\mathscr{O}/(\mathscr{F}|_{\mathscr{O}}) =: V$ can be also visualized by gluing $P_i|_{\mathscr{O}}$ together via the induced equivalence by $\mathscr{F}|_{\mathscr{O}}$ on $P_i|_{\mathscr{O}}$. Although P is not necessarily a good chart in the sense of Lemma 2.4, we join V to P, that is $P := V \sqcup (\sqcup_i P_i)$. Then P becomes good and the étale groupoid $\Gamma := P \times_{\mathscr{G}} P \Rightarrow P$ via $P \to \mathscr{G}$ is a good groupoid presentation. To avoid duplicated notation, from now on, $\Gamma \Rightarrow P$ is this good groupoid presentation which contains V.

A *W-groupoid morphism* $\Phi : (\mathscr{G} \Rightarrow M) \to (\mathscr{H} \Rightarrow N)$, is made up by a map $\Phi_1 : \mathscr{G} \to \mathscr{H}$ between stacks and $\Phi_0 : M \to N$ such that they preserve the W-groupoid

(6) Thanks to Lemma 2.4 we can construct the local groupoid at once and it is not necessary to divide M into pieces as therein.

structure maps up to 2-morphisms and these 2-morphisms satisfy again higher coherence conditions linking the 2-commutative diagrams of $\mathscr{G}$ and $\mathscr{H}$ (also see [**29**, Section 7]). Given such a morphism Φ, one can choose a good étale presentation $G_1 \Rightarrow G_0$ (resp. $H_1 \Rightarrow H_0$) of $\mathscr{G}$ (resp. $\mathscr{H}$) namely the one such that M (resp. N) embeds in G_0 (resp. H_0). Under the correspondence of differentiable stacks and Lie groupoids, morphisms of differentiable stacks are presented by H.S. (Hilsum-Skandalis) bibundles [**17**] [**15**], which are manifolds that both groupoids acts from left and right respectively with certain conditions (see for example [**25**, Section 3]). We denote E_Φ as the H.S. bibundle presenting the morphism $\Phi_1 : \mathscr{G} \to \mathscr{H}$. The restriction of $E_\Phi|_M$ presents the map $M \to \mathscr{G} \xrightarrow{\Phi} \mathscr{H}$ which is $M \xrightarrow{\Phi_0} N \xrightarrow{\bar{e}_{\mathscr{H}}} \mathscr{H}$ (up to a 2-morphism), therefore $E_\Phi|_M \cong M \times_{\Phi_0, N, e_{\mathscr{H}}} H_0 \to M$ admits a global section. Thus, extending this section on a local neighborhood $U(M) \subset G_0$ of M, we arrive at a section of $E_\Phi|_{U(M)}$ such that the composition $\Phi_{loc} : U(M) \to E_\Phi \to H_0$ extends the map $\Phi_0 : M \to N$. We then choose $U(M)$ close enough to M so that itself and $U(N) := \Phi_{loc}(U(M))$ have a local Lie groupoid structure as above. By construction we have a 2-commutative diagram,

$$\begin{tikzcd} U(M) \arrow[r] \arrow[d, "\Phi_{loc}"'] & \mathscr{G} \arrow[d, "\Phi_1"'] \\ U(N) \arrow[r] & \mathscr{H} \end{tikzcd} \tag{5}$$

Then Φ_{loc} preserves the local groupoid structures exactly because Φ preserves the W-groupoid structures. Hence Φ_{loc} is a local groupoid morphism and it induces an algebroid morphism $\varphi : A(\mathscr{G}) \to A(\mathscr{H})$, where $A(.)$ is the functor of taking the algebroid of a W-groupoid. If there are two morphisms Φ and Φ' differed by a 2-morphism in a compatible fashion with structure 2-morphisms of $\mathscr{G}$ and $\mathscr{H}$, then $E_\Phi \cong E_{\Phi'}$. Therefore $U(N) \cong U'(N)$ through this isomorphism and

$$\Phi_{loc} = \big(U(M) \xrightarrow{\Phi'_{loc}} U'(N) \cong U(N)\big).$$

Therefore the two W-groupoid morphisms induce the same algebroid morphism.

2.2. The construction of the Lie 2-groupoid. One uniform way to describe a (resp. Lie) n-groupoid is via its nerve: by requiring it to be a simplicial (resp. manifolds) sets [**14**] whose homotopy groups are trivial above π_n. We leave the readers to the introduction of [**29**] and the references therein for a general description for this and only recall briefly the definitions we need here.

A simplicial set (resp. manifold) X is made up by sets (resp. manifolds) X_n and structure maps

$$d_i^n : X_n \to X_{n-1} \text{ (face maps)} \quad s_i^n : X_n \to X_{n+1} \text{ (degeneracy maps)}, \;\; i \in \{0,1,2,...,n\}$$

that satisfy suitable coherence conditions. The first two examples are a simplicial m-simplex $\Delta[m]$ and a horn $\Lambda[m,j]$ with

$$(\Delta[m])_n = \{f : (0,1,...,n) \to (0,1,...,m) | f(i) \leqslant f(j), \forall i \leqslant j\},$$
$$(6) \qquad (\Lambda[m,j])_n = \{f \in (\Delta[m])_n | \{0,...,j-1,j+1,...,m\} \nsubseteq \{f(0),...,f(n)\}\}.$$

A simplicial set X is *Kan* if any map from the horn $\Lambda[m,j]$ to X ($m \geq 1$, $j = 1,..,m$), extends to a map of $\Delta[m]$. In the language of groupoids, the Kan condition corresponds to the possibility of composing various morphisms. In an n-groupoid, the only well defined composition law is the one for n-morphisms. This motivates the following definition.

Definition 2.6. A Lie n-groupoid X ($n \in \mathbb{N} \cup \infty$) is a simplicial manifold that satisfies $Kan(m,j)$ $\forall m \geq 1, 0 \leq j \leq m$ and $Kan!(m,j)$ $\forall m > n, 0 \leq j \leq m$.

$Kan(m,j)$: The restriction map $\hom(\Delta[m],X) \to \hom(\Lambda[m,j],X)$ is a surjective submersion.

$Kan!(m,j)$: The restriction map $\hom(\Delta[m],X) \to \hom(\Lambda[m,j],X)$ is a diffeomorphism.

Remark 2.7. A Lie n-groupoid X is determined by its first $(n+1)$-layers $X_0, X_1, \ldots, X_n$ and some structure maps. For example a Lie 1-groupoid is exactly determined by a Lie groupoid structure on $X_1 \Rightarrow X_0$. In fact a Lie 1-groupoid is the nerve of a Lie groupoid. A Lie 2-groupoid is exactly determined by $X_2 \Rrightarrow X_2 \Rightarrow X_0$ with a sort of 3-multiplications and face and degeneracy maps satisfying certain compatible condition. It is made precise in [**29**, Section 2]. Hence in this paper, we often write only the first three layers of a Lie 2-groupoid.

Now to a Lie algebroid A over a manifold M, we associate the simplicial set $S(A)$ of equation (1) in the introduction. The first three layers of $S(A)$ are actually familiar to us:

– it is easy to check that $S_0 = M$;

– S_1 is exactly the A-path space $P_a A$ since a map $T\Delta^1 \to A$ can be written as $a(t)dt$ with base map $\gamma(t) : \Delta^1 \to M$, it being a Lie algebroid map is equivalent to $\rho(a(t)) = \frac{d}{dt}\gamma(t)$ since the Lie bracket of $T\Delta^1$ is trivial and the anchor of it is identity;

– bigons in S_2 are exactly the A-homotopies in $P_a A$ since a bigon

$$f : T(d_2^2)^{-1}(Ts_0^1(T\Delta^0)) \to A$$

can be written as $a(t,s)dt+b(t,s)ds$ over the base map $\gamma(t,s)$ after a suitable choice of parametrization[(7)] of the disk $(d_2^2)^{-1}(s^1(\Delta^0))$. Then we naturally have $b(0,s) = f(0,s)(\frac{\partial}{\partial s}) = 0$ and $b(1,s) = f(1,s)(\frac{\partial}{\partial s}) = 0$. Moreover the morphism is a Lie algebroid morphism if and only if $a(t,s)$ and $b(t,s)$ satisfy equation (3) which defines the A-homotopy.

We take the "2-truncation" of this simplicial set, that is we take $X(A)$ to be

$$X_i(A) = S_i(A),\ i = 0,1, \quad \text{and } X_2(A) = S_2(A)/S_3(A),$$

where the quotient $S_2(A)/S_3(A)$ is formed by $\alpha \sim \beta$ if and only if they share the same boundary in $S_1(A)$ and they bound an element in $S_3(A)$, i.e. they are homotopic in the sense of [**14**].

Remark 2.8. We do not know whether $S(A)$ is a simplicial manifold or further a *Kan simplicial manifold*, namely a Lie ∞-groupoid as in Definition 2.6, though it is so for a Lie algebra [**10**]. If we have known that $S(A)$ is a Kan simplicial manifold then we could simply take the Lie 2-groupoid as the 2-truncation of $S(A)$, that is

$$\ldots S_3(A)/S_4(A) \to S_2(A)/S_3(A) \to S_1(A) \to S_0(A).$$

However, unlike that $S_1(A)$ being a Banach manifold involves solving an ODE, it is not clear how to solve directly the corresponding PDE for $S_2(A)$ to be a Banach manifold. This is one of the open questions left at the end of [**10**].

Also, although $S(A)$ has clear geometric meaning, it involves infinite dimensional manifold, and although $\mathscr{G}(A)$ is an étale SLie groupoid, the 2-truncation $X(A)$ will not be 2-étale in the sense of [**29**]. To achieve the étale version we need to use a sub-simplicial set based on the good étale covering P of $\mathscr{G}(A)$ in Remark 2.5, namely

$$S_0'(A) = M,\ S_1'(A) = P,\ S_i'(A) = S_i(A)|_P,\ \text{for } i \geqslant 2, \tag{7}$$

where $S_i(A)|_P$ is the subset of i-simplices in $S_i(A)$ whose 1-skeletons are made up by elements in P. See also Remark 2.11.

Therefore we have to use an alternative method, that is to use the 1-1 correspondence of Lie 2-groupoids of SLie groupoids. Recalling from [**29**], given an SLie

(7) We need the one that $\gamma(0,s) = x$ and $\gamma(1,s) = y$ for all $s \in [0,1]$.

groupoid $\mathscr{G} \Rightarrow M$ with a groupoid presentation G of $\mathscr{G}$, its associated Lie 2-groupoid Y is constructed by

$$Y_0 = M, \quad Y_1 = G_0, \quad Y_2 = E_m,$$

where E_m is the bibundle representing the multiplication $m : \mathscr{G} \times_M \mathscr{G} \to \mathscr{G}$ of the SLie groupoid.

Then Theorem 1.2 is equivalent to the following:

Theorem 2.9. *$X(A)$ is a Lie 2-groupoid and it corresponds to $\mathscr{G}(A)$ with the correspondence in* [**29**].

Proof. : Take the Banach chart $P_a A$ of $\mathscr{G}(A)$. The Lie 2-groupoid $Y(A)$ corresponding to $\mathscr{G}(A)$ has

$$Y_0(A) = S_0(A) = M, \quad Y_1(A) = S_1(A) = P_a A, \quad Y_2(A) = E_m = S_2(A)/\sim, \tag{8}$$

where two A-homotopies $a \sim b$ if and only if there is a path of A-homotopies $a(\epsilon)$ such that $a(0) = a$, $a(1) = b$, and paths $a(\epsilon) \in P_a A$ all have the same end points in A for $\epsilon \in [0,1]$. By Remark 2.1 and Lemma 2.2, $P_a A$ is a (Hausdorff) Banach manifold and $Mon(P_a A)$ is a Banach manifold (not necessarily Hausdorff). The source and target of $Mon(P_a A)$ to $P_a A$ are surjective submersions. All these hold for groupoid $Mon(P_0 A) \Rightarrow P_0 A$ too, therefore the multiplication bibundle $E^0_m := (P_0 A \times P_0 A) \times_{\odot, P_0 A, \mathbf{t}} Mon(P_0 A)$ is a Banach manifold and the left moment map $J_l : E^0_m \to P_0 A \times_M P_0 A$ is a surjective submersion. The Morita equivalence between $Mon(P_a A) \Rightarrow P_a A$ and $Mon(P_0 A) \Rightarrow P_0 A$ is provided by the map $f_\tau : P_a A \to P_0 A$ of parametrization

$$a(t) \mapsto a^\tau(t) := \dot{\tau}(t) a(\tau(t)), \tag{9}$$

using a fixed a smooth cut-off function

$$\tau : [0,1] \to [0,1], \quad \text{which satisfies } \tau'(0) = \tau'(1) = 0, \tau'(t) \geqslant 0. \tag{10}$$

Then $E_m = (P_a A \times_M P_a A) \times_{f_\tau, P_0 A \times P_0 A, J_l} E^0_m$ is a Banach manifold.

The face and degeneracy maps of $Y(A)$ and $X(A)$ are obviously the same. However what we mod out to form $Y_2(A)$ is not exactly the elements in $S_3(A)$ because it is not obvious that a path of A-homotopies can be connected to make a Lie algebroid morphism $T\Delta^3 \to A$. Hence to show $Y_2(A) = X_2(A)$ thus to give $X(A)$ a Lie 2-groupoid structure via the isomorphism to $Y(A)$, we need Proposition 2.10. This proposition is proven in [**29**]. $\square$

Proposition 2.10. *If there is a smooth path of Lie algebroid morphisms $\varphi(t)$: $T\Delta^2 \to A$ such that $\varphi(t)|_{\partial\Delta^2}$ stays the same when t varies in $[0,1]$, then there is a Lie algebroid morphism $\varphi : T\Delta^3 \to A$ such that $\varphi|_{d_i\Delta^3} = \varphi(i)$, for $i = 0,1$. Here d_i is the i-th face map $\Delta^3 \to \Delta^2$.*

Remark 2.11. The Lie 2-groupoid $X(A)$ is not 2-étale, that is $X_2(A) \to \hom(\Lambda[2,j], X(A))$ is not étale for $j = 0,1,2$. To obtain the étale version we shall use the simplicial set in (7). Then the replacement of $X(A)$ is

$$Z_0(A) = M, \quad Z_1(A) = P, \quad Z_2(A) := X_2(A)|_P,$$

where the restriction $X_2(A)|_P$ is the subset of equivalence classes of the 2-simplices in $S_i(A)$ whose 2-skeletons are made up by elements in P. Then $Z_2(A) \cong Y_2(A)|_P \cong E_m(\Gamma)$ which is the H.S. bibundle representing the multiplication m of $\mathscr{G}(A)$ for the étale groupoid presentation $\Gamma \Rightarrow P$ of $\mathscr{G}(A)$. Since $\Gamma \Rightarrow P$ is étale, the left moment map $E_m(\Gamma) \to \Gamma \times_{\mathbf{s},P,\mathbf{t}} \Gamma$ is étale. Since $Z_2(A) \to \hom(\Lambda[2,j], Z(A))$ are surjective submersions for all j, they are furthermore étale by dimensional counting. Hence $Z(A)$ is a 2-étale Lie 2-groupoid.

3. Lie II theorem

In this section we prove Theorem 1.3. We begin with

Lemma 3.1. *If φ is a Lie algebroid morphism $A \to B$, then it induces a W-groupoid morphism $\Phi : \mathscr{G}(A) \to \mathscr{G}(B)$ such that Φ gives back the Lie algebroid morphism φ. The same is true for W-groupoid $\mathscr{H}(\cdot)$.*

Proof. We prove it for $\mathscr{G}(\cdot)$ and the proof for $\mathscr{H}(\cdot)$ is similar. A Lie algebroid morphism $\varphi : A \to B$ induces a morphism $\varphi_\cdot : S(A) \to S(B)$. In particular, φ_1 and φ_2 give a morphism $P_aA \to P_aB$ and a morphism on the level of A-homotopies respectively. If we have two equivalent A-homotopies $\xi_0 \sim \xi_1$ of A, in the sense of (8), then their images $\varphi_2(\xi_1) \sim \varphi_2(\xi_2)$ are also equivalent. So φ_2 induces a homomorphism of groupoids $(Mon(P_aA) \Rightarrow P_aA) \to (Mon(P_aB) \Rightarrow P_aB)$, hence a morphism on the level of stacks $\Phi : \mathscr{G}(A) \to \mathscr{G}(B)$. This morphism gives a local groupoid morphism $\Phi_{loc} : G_{loc}(A) \to G_{loc}(B)$ which maps equivalence classes of A-paths in A to those in B via φ since the local groupoids can be understood as equivalence classes of A-paths. Therefore the corresponding Lie algebroid map is exactly φ. □

Proof of Thm. 1.3. To build the map Φ, by Lemma 3.1, we only have to treat the situation when $\varphi = id : A \to A$, that is, given a W-groupoid $\mathscr{G}$ whose algebroid is A, there is a W-groupoid morphism $\Psi : \mathscr{G}(A) \to \mathscr{G}$ lifting the $id : A \to A$. Both $\mathscr{G}(A)$ and $\mathscr{G}$ have an associated local groupoid whose Lie algebroid is A. They are isomorphic in a small enough neighborhood of M and we might as well assume they are the same and denote it by G_{loc}.

The basic idea is as following: given an A-path $a(t)$, we cut it into $n = 2^k$ small pieces,

$$a_l(t) = \frac{1}{n} a(\frac{t+i-1}{n}), \quad \text{for } t \in [0,1], \text{ and } l = 1, \dots, n.$$

The whole path a is a concatenation of these pieces, that is

$$a(t) = a_1 \odot a_2 \odot \cdots \odot a_n := n a_l(nt - (l-1)), \quad \text{for } t \in [\frac{l-1}{n}, \frac{l}{n}]. \tag{11}$$

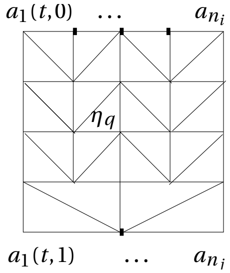
For a big enough n, we must have $a_l \in \mathscr{O}$ the neighborhood of M in $P_a A$ to define G_{loc} (see Remark 2.5). We denote g_l as the equivalence class that $a_l(t)$ represents in $\mathscr{O}/\mathscr{F}|_{\mathscr{O}}$. Then we define $\Psi([a(t)]) := (\dots((\pi(g_1) \cdot \pi(g_2)) \cdot \pi(g_3)) \cdot \dots) \cdot \pi(g_n)$, where π is the projection $G_0 \to \mathscr{G}$.

More precisely, we take the presentation $\Gamma \Rightarrow P$ of $\mathscr{G}(A)$ and a good presentation G of $\mathscr{G}$. We take also G to be such a good presentation that the multiplication bibundle E_m^G has a global section for the moment map $E_m^G \to G_0 \times_M G_0$, i.e. the multiplication is a strict map (otherwise we replace G by a finer cover). As we stated at the beginning, we also assume that $V \subset G_0$ (otherwise we can take a smaller open set $\mathscr{O}$ as explained in Remark 2.5). We only have to build a groupoid morphism $(\Gamma \Rightarrow P) \xrightarrow{f} G$. For $a(t) \in P$, we take a subdivision as above and define

$$f_0(a(t)) := (\dots((g_1 \cdot g_2) \cdot g_3) \cdot \dots) \cdot g_n.$$

Since $P = \sqcup_i P_i$, we might as well assume that the division of $a(t)$ has the same number of pieces on each P_i, otherwise we replace P by a finer transversal. This will not affect the local groupoid part $V \subset P$ since $f_0 = id$ on V. This makes f_0 smooth on each piece P_i, hence a smooth morphism on the disjoint union P.

To define $f_1 : \Gamma \to G_1$, we take an A-homotopy $a(t,s)$ between $a(t,0) \in P_i$ and $a(t,1) \in P_j$. Suppose that $a(t,0)$ is divided into $n_i = 2^{k_i}$ pieces and $a(t,1)$ into $n_j = 2^{k_j}$ pieces. Then the A-homotopy $a(t,s)$ gives rise to many small triangles η_q in the multiplication bimodule $E_m^{loc} \subset G_{loc} \times_M G_{loc}$ of G_{loc}, which is then in the multiplication bimodule E_m^G of G.



(12)

Then the associativity of E_m^G tells us how to compose these small triangles into a big bigon $\eta \in G_1 \subset E_m^G$ and the pentagon condition of the associator tells us that the result is independent of the order of composition. We define $\tilde{f}_1((a(t,s)) = \eta$. We make a fixed subdivision (12) locally, then $\tilde{f}_1$ is smooth. If two A-homotopies are linked by a path of A-homotopies, their subdivisions are also linked by a path of subdivisions. Hence $\tilde{f}_1$ descends to a map $f_1 : \Gamma \to G_1$.

Uniqueness of Φ follows from the next lemma. $\square$

Lemma 3.2. *If we have two morphisms of W-groupoids $\Phi, \Phi' : \mathscr{G}(A) \to \mathscr{H}$ corresponding to the same algebroid morphism $\varphi : A \to B$, then they differ by a 2-morphism.*

Proof. Since $\Phi, \Phi' : \mathscr{G}(A) \to \mathscr{H}$ correspond to the same algebroid morphism $\varphi : A \to B$, their corresponding morphisms of local groupoids are the same, i.e. $\Phi_{loc} = \Phi'_{loc}$. By diagram (5) and the fact that Φ and Φ' preserve the multiplication, we have the following 2-commutative diagram,

$$\begin{tikzcd} \mathscr{G}(A) \arrow[d, shift right, "\Phi"'] \arrow[d, shift left, "\Phi'"] & \sqcup_n G_{loc}(A)^{\times n} \arrow[l, "p_1"'] \arrow[d, "\Phi_{loc}=\Phi'_{loc}"] \\ \mathscr{H} & \sqcup_n G_{loc}^{\times n} \arrow[l, "p_2"'] \end{tikzcd}$$

where $G_{loc}(A)$ and G_{loc} are the local groupoids of $\mathscr{G}(A)$ and $\mathscr{H}$ respectively, and $p_1 : L := \sqcup_n G_{loc}(A)^{\times n} \to \mathscr{G}(A)$ is defined by $p_1(g_1, ..., g_n) = (...((\pi(g_1) \cdot \pi(g_2)) \cdot \pi(g_3)...) \cdot \pi(g_n)$, where $\pi : G_{loc}(A) \to P \to \mathscr{G}(A)$ with P the good étale chart of $\mathscr{G}(A)$. We define p_2 similarly. Here $\square^{\times n}$ denotes a n-fold fibre-product over M.

Therefore $\Phi' \circ p_1 \sim \Phi_{loc} \circ p_2 \sim \Phi \circ p_1$, where $f \sim g$ means that the two (1-)morphisms f and g differ by some 2-morphism. Since 2-morphism $\alpha : \Phi \to \Phi'$ is simply a natural transformation, which is a set of compatible arrows $\alpha(x) : \Phi(x) \to \Phi'(x)$ for every object x in the stack $\mathscr{G}(A)$. To show $\Phi \sim \Phi'$, one only has to show that p_1 is essentially surjective, namely p_1 projects to every object and every morphism up to isomorphisms. An essential surjection between stacks is an epimorphism. Thus we only have to show that for any object y over U in $\mathscr{G}(A)$, there is an open covering U_i of U such that there exists x_i over U_i in L (viewed as a category) and $p_1(x_i) \cong y|_{U_i}$. Take the good étale groupoid presentation $\Gamma \Rightarrow P$ of $\mathscr{G}(A)$. Then an object of $\mathscr{G}(A)$ over U is a groupoid principal bundle of $\Gamma \Rightarrow P$ over U. Take an open covering U_i of U so that $y|_{U_i}$ is trivial. Then y is decided by a map $U_i \to P$ and $y|_{U_i} = U_i \times_P \Gamma$ via this map.

On the other hand, the map $p_1 : L \to \mathscr{G}(A)$ is expressed as the composition of the following maps on the level of groupoids

(13)
$$\begin{tikzcd}
L \arrow[r] \arrow[dd, shift left] \arrow[dd, shift right] & \sqcup_n \Gamma^{\times n} \arrow[dd, shift left] \arrow[dd, shift right] & & \Gamma \arrow[dd, shift left] \arrow[dd, shift right] \\
& & \sqcup_n E_{m^{\times n}} \arrow[dl] \arrow[dr] & \\
L \arrow[r] & \sqcup_n P^{\times n} & & P
\end{tikzcd}$$

where $E_{m^{\times n}}$ is the bimodule presenting the map $m \circ (m \times id) \circ \ldots (m \times id \times \ldots id) : \mathscr{G}(A)^{\times n} \to \mathscr{G}(A)$. As recalled in Section 2.1, $E_m(\Gamma) \Rrightarrow P \Rightarrow M$ is a Lie 2-groupoid, so by $Kan(2,0)$ condition that it satisfies, the map $E_m(\Gamma) \to P \times_{\mathbf{t},M,\mathbf{t}} P$ is a surjective submersion. Composing this map with the projection $pr_2 : P \times_{\mathbf{t},M,\mathbf{t}} P \to P$, we have $J_r : E_m(\Gamma) \to P$ is a surjective submersion (see [**29**, Section 4.2]). Hence the right moment map of $E_{m^{\times n}}$ is also surjective submersion since the bibundle $E_{m^{\times n}}$ is formed by composing bibundles with the form $E_m(\Gamma) \times_M \Gamma^{\times k}$. For example

$$E_{m^{\times 2}} = (E_m(\Gamma) \times_M \Gamma \times_{P \times_M P} E_m(\Gamma))/\Gamma \times_M \Gamma,$$

and the right moment map of $E_{m^{\times 2}}$ to P comes from a composed map of J_r and pull-backs of J_r, namely $E_m(\Gamma) \times_M \Gamma \times_{P \times_M P} E_m(\Gamma) \to E_m(\Gamma) \to P$, which is a surjective submersion. This implies the descending map $E_{m^{\times 2}} \to P$ is also a surjective submersion.

Denote E the composed H.S. bibundle in (13) and it is simply $\sqcup_n E_{m^{\times} n}$ restricting to L via the left moment map. Since $L \to \sqcup_n P^{\times n}$ is an open embedding, $E \to P$ is a submersion, and it is furthermore surjective because an A-path can always be divided into small enough paths which lie in G_{loc}. Hence we have local sections $(P \supset) V_j \to E$ of this map. Then we take a pull-back covering U_{i_j} of U_i as in the following 2-commutative diagram

$$\begin{array}{ccccccc} U_{i_j} & \longrightarrow & V_j & \longrightarrow & E & \longrightarrow & L \\ \downarrow & & \downarrow & & \downarrow & & \downarrow \\ U_i & \longrightarrow & P & \xrightarrow{id} & P & \longrightarrow & \mathscr{G}(A) \end{array}$$

Then we have the composed map $U_{i_j} \to V_j \to E \to L$. Therefore $y|_{U_{i_j}} = U_{i_j} \times_{\mathscr{G}(A)} P \cong U_{i_j} \times_L L \times_{\mathscr{G}(A)} P = U_{i_j} \times_L E = p_1(U_{i_j} \to L)$. Therefore p_1 is an epimorphism. □

4. Examples

Now we show an example of a stacky groupoid $\mathscr{G}(A)$. In fact, for any manifold M with $\pi_2(M)$ nontrivial, $\mathscr{G}(TM)$ is *not* the traditional homotopy groupoid $\tilde{M} \times \tilde{M}/\pi_1(M)$ where $\tilde{M}$ is the simply connected cover of M. $\pi_2(M)$ will play a role too.

4.1. $\mathbb{Z}$-gerbes are S^1 bundles. $\mathbb{Z}$-gerbes and S^1-bundles over a manifold M are both characterized by $H^2(M, \mathbb{Z})$ via Čech cohomology and the Chern class respectively. Given an S^1 bundle S over M, let $\{U_i\}$ be a covering of M such that S trivializes locally as $U_i \times S^1$ with gluing function $g_{ij} : U_{ij} := U_i \cap U_j \to S^1$. As we know the map $H^1(M, S^1) \xrightarrow{\check{\delta}\circ\log} H^2(M, \mathbb{Z})$ gives us the Chern class $c_1(S) = [g_{...}]$ with $g_{...} = \check{\delta}(\log g_{..})$.

On the other hand, there is a stack $\mathscr{G}$ presented by groupoid $\sqcup U_{ij} \times \mathbb{Z} \Rightarrow \sqcup U_i$ with the groupoid multiplication $(x_{ij}, n) \cdot (x_{jk}, m) = (x_{ik}, n + m + g_{ijk})$ and the source and target maps inherited from the groupoid $\sqcup U_{ij} \Rightarrow \sqcup U_i$ which presents M. Here x_{ij} denotes a point in U_{ij}. An A-gerbe corresponds to an A-groupoid central extension for an abelian group A [**3**]. It is clear that $\mathscr{G}$ is a $\mathbb{Z}$-gerbe (or equivalently a $B\mathbb{Z}$-principal bundle [**27**]) over M from the following diagram of groupoid central extension:

$$\begin{array}{ccccccccc} 1 & \longrightarrow & \mathbb{Z} \times \sqcup U_i & \longrightarrow & \sqcup U_{ij} \times \mathbb{Z} & \longrightarrow & \sqcup U_{ij} & \longrightarrow & 1 \\ & & \downarrow\downarrow & & \downarrow\downarrow & & \downarrow\downarrow & & \\ & & \sqcup U_i & = & \sqcup U_i & = & \sqcup U_i & & \end{array}$$

The gerbe $\mathscr{G}$ is classified by the Čech class $[g_{...}] \in H^2(M, \mathbb{Z})$ which determines the groupoid multiplication (see also [**7**, Appendix]). From this class it is easy to go back to an S^1-bundle. Hence this gives the *1-1 correspondence* between $\mathbb{Z}$-gerbes and S^1-bundles.

Another equivalent way to obtain $\mathscr{G}$ from S is to realize $\mathscr{G}$ as a global quotient $[S/\mathbb{R}]$ where $\mathbb{R}$ acts(8) (left) on S via the S^1 action and the projection $pr : \mathbb{R} \to S^1$. We simply verify it by a Morita bibundle $\sqcup U_i \times \mathbb{R}$ between $\sqcup U_{ij} \times \mathbb{Z} \Rightarrow \sqcup U_i$ and the action groupoid $S \times \mathbb{R} \Rightarrow S$. Our convention is $\mathbf{t}(x,t) = x$ and $\mathbf{s}(x,t) = t \cdot x (=: x \cdot t^{-1})$. The moment maps are

$$J_l = pr_1 : \sqcup U_i \times \mathbb{R} \to \sqcup U_i, \quad J_r : \sqcup U_i \times \mathbb{R} \overset{id \times pr}{\to} \sqcup U_i \times S^1 \overset{\pi}{\to} S,$$

where pr_1 is the projection to the first component and $\pi : \sqcup U_i \times S^1 \to S$ is the chart projection since S is locally $U_i \times S^1$. The left and right actions are respectively

$$(x_{ij}, n) \cdot (x_j, a) = (x_i, n + a), \quad (x_i, a) \cdot (x = J_r(x_i, a), \lambda) = (x_i, a + \lambda),$$

which are free and transitive.

Example 4.1. Corresponding to the Hopf fibration $S^3 \to S^2$, there is a $\mathbb{Z}$ gerbe or $B\mathbb{Z}$ principal bundle denoted as $\tilde{S}^2$. Apply the long exact sequence of homotopy groups to the $B\mathbb{Z}$-fibration $\tilde{S}^2 \to S^2$. Since $\pi_n(B\mathbb{Z}) = 0$ except $\pi_1(B\mathbb{Z}) = \mathbb{Z}$, we have $\pi_1(\tilde{S}^2) = \pi_2(\tilde{S}^2) = 0$ and $\pi_{\geqslant 3}(\tilde{S}^2) = \pi_{\geqslant 3}(S^2)$. Hence we can view $\tilde{S}^2$ as a 2-connected "covering" of S^2.

4.2. Symplectic structure. Recall [**24**] that a *symplectic form on an étale differentiable stack* $\mathscr{X}$ is a G-invariant symplectic form on G_0, where G is an étale presentation of $\mathscr{X}$. Appearing on a non-étale presentation H of $\mathscr{X}$, the symplectic form could be an H-invariant pre-symplectic form ω on H_0, but we must have $\ker \omega = TO$, where O is the orbit of H_1 action on H_0.

Example 4.2. $\tilde{S}^2$ has a pull-back 2-form $\tilde{\pi}^* \omega$ with $\tilde{\pi} : \tilde{S}^2 \to S^2$ and ω the symplectic area form on S^2. As above, take the action groupoid $S^3 \times \mathbb{R} \Rightarrow S^3$ presenting $\tilde{S}^2$, where $\mathbb{R}$ acts on S^3 via the projection $\mathbb{R} \to S^1$. Then this 2-form $\tilde{\pi}^* \omega$ appears on S^3 as $\pi^* \omega$ where $\pi : S^3 \to S^2$. Indeed its kernel $\ker \pi^* \omega$ is only along the S^1 orbit in S^3. Hence it is a symplectic form on $\tilde{S}^2$. However, surprisingly, we will show that the de-Rham class $[\tilde{\pi}^* \omega] = 0 \in H^2(\tilde{S}^2, \mathbb{R})$.

(8) locally by $\lambda \cdot (x_i, [a]) = (x_i, [a + \lambda])$.

First of all we use groupoid de-Rham double complex $C^{p,q} = \Omega^q(S^3 \times \mathbb{R}^p)$ to calculate $H^\bullet(\tilde{S}^2, \mathbb{R})$.

$$
\begin{tikzcd}
 & \cdots & \cdots & \cdots & \cdots \\
 & C^{0,2} \arrow[u, "d"] \arrow[r, "\delta"] & C^{1,2} \arrow[u, "d"] \arrow[r, "\delta"] & C^{2,2} \arrow[u, "d"] \arrow[r, "\delta"] & \cdots \\
q \uparrow & C^{0,1} \arrow[u, "d"] \arrow[r, "\delta"] & C^{1,1} \arrow[u, "d"] \arrow[r, "\delta"] & C^{2,1} \arrow[u, "d"] \arrow[r, "\delta"] & \cdots \\
 & C^{0,0} \arrow[u, "d"] \arrow[r, "\delta"] & C^{1,0} \arrow[u, "d"] \arrow[r, "\delta"] & C^{2,0} \arrow[u, "d"] \arrow[r, "\delta"] & \cdots \\
 & & \xrightarrow{p} & &
\end{tikzcd}
\tag{14}
$$

The two differentials are the de-Rham $d = d^{p,q} : C^{p,q} \to C^{p,q+1}$ and the groupoid $\delta^{p,q} : C^{p,q} \to C^{p+1,q}$ with $\delta^{p,q} = \sum_{i=0}^{p}(-1)^{i+p}\delta_i^*$ and $\delta_i : S^3 \times \mathbb{R}^p \to S^3 \times \mathbb{R}^{p-1}$ by

$$\delta_0(x, t_1, \dots, t_p) = (x \cdot t_1, t_2, \dots, t_p),$$
$$\delta_i(x, t_1, \dots, t_p) = (x, t_1, \dots, t_i + t_{i+1}, \dots, t_p) \text{ for } 1 \leqslant i \leqslant p-1,$$
$$\delta_p(x, t_1, \dots, t_p) = (x, t_1, \dots, t_{p-1}).$$

Then $H^\bullet(\tilde{S}^2, \mathbb{R})$ is the total cohomology of $(C^{p,q}, d, \delta)$, i.e. the cohomology of the complex $(\oplus_{p+q=n} C^{p,q}, D = (-1)^p d^{p,q} + \delta^{p,q})$.

We use a spectral sequence $(E^{p,q}, d_r : E_r^{p,q} \to E_r^{p+r,q-r+1})$ to calculate the total cohomology of this double complex. We let the 0-th page $E_0^{p,q} = C^{p,q}$ and $d_0 = d$. Then the first page $E_1^{p,q} = H^q(S^3)$ which is 0 except for the 0th and the 3rd row being $\mathbb{R}$ everywhere. Hence $E_{\geqslant 2}^{2,0} = \mathbb{R}$ and $E_{\geqslant 2}^{1,1} = E_{\geqslant 2}^{0,2} = 0$. Hence $H^2(\tilde{S}^2, \mathbb{R}) = \mathbb{R}$.

In the double complex $C^{p,q}$, $[\tilde{\pi}^*\omega]$ is represented by

$$(0, 0, \pi^*\omega) \in C^{0,2} \oplus C^{1,1} \oplus C^{2,0}.$$

However, $(0, 0, \pi^*\omega)$ is exact under the total differential $D = \pm d + \delta$. In fact $D(t, \theta) = (0, 0, \pi^*\omega)$ where t is the coordinate on $\mathbb{R}$ and θ is the connection 1-form for the Hopf fibration $S^3 \to S^2$. This follows from the calculation $d^{0,1}\theta = \pi^*\omega$ and $\delta^{0,1}\theta = \delta_0^*\theta - \delta_1^*\theta$. Let X be the Reeb vector field on S^3. X is invariant under the S^1-action and $\theta(X) = 1$. Thus $\delta\theta(X) = 0$. Since $span\{X\} \oplus \ker\theta = TS^3$, we only have to care about $\frac{\partial}{\partial t} \in T\mathbb{R} \subset T(S^3 \times \mathbb{R})$. Since $T\delta_1 \frac{\partial}{\partial t} = 0$ and $T\delta_0 \frac{\partial}{\partial t} = X$, we have $\delta^{0,1}\theta(\frac{\partial}{\partial t}) = 1$. Hence $\delta^{0,1}\theta = d^{1,0}t$. Finally, it is not hard to see that

$$(\delta^{1,0}t)(x, t_1, t_2) = t_2 - (t_1 + t_2) + t_1 = 0.$$

This implies that $[\tilde{\pi}^*\omega] = [(0,0,\pi^*\omega)] = 0 \in H^2(\tilde{S}^2,\mathbb{R}) = \mathbb{R}$.

4.3. Symplectic W-groupoids. Recall [**23**] that a W-groupoid $\mathscr{G} \Rightarrow M$ is a *symplectic W-groupoid* if there is a symplectic form ω on $\mathscr{G}$ satisfying the following *multiplicative* condition:

$$m^*\omega = pr_1^*\omega + pr_2^*\omega, \tag{15}$$

on $\mathscr{G} \times_{\mathbf{s},M,\mathbf{t}} \mathscr{G}$, where pr_i is the projection onto the i-th factor.

Example 4.3 ($\mathscr{G}(T^*S^2)$). Take the symplectic manifold S^2 with its area form ω as above. Then the tangent Lie algebroid TS^2 is isomorphic to the Lie algebroid T^*S^2 of (S^2,ω) viewed as a Poisson manifold. The isomorphism is given by the contraction with ω, $\sharp\omega : TS^2 \to T^*S^2$. To obtain the symplectic W-groupoid of (S^2,ω), we can equally study $\mathscr{G}(TS^2)$.

The set of Lie algebroid morphisms $TN \to TM$ is equal to the set of smooth maps $Mor(N,M)$. We take the Lie 2-groupoid,

$$X_0 = M, \quad X_1 = Mor(\Delta^1,M), \quad X_2 = Mor(\Delta^2,M)/Mor(\Delta^3,M).$$

Recall that the quotient is by $\omega_1 \sim \omega_2$ if they have the same boundary and they bound an element in $Mor(\Delta^3,M)$. So the stack $\mathscr{G}(TM)$ is presented by the groupoid $G_1 \Rightarrow G_0$ with $G_1 = Mor(D^2,M)/Mor(D^3,M)$ the space of bigons (the quotient is similarly given by $\beta_1 \sim \beta_2$ if they have the same boundary and bound an element in $Mor(D^3,M)$), and $G_0 = Mor(I,M)$. Here D^k's are viewed as Δ^k with many degenerate faces. Then G_0 is simply the space of A-paths P_aTS and G_1 is the space of A-homotopies modding out of higher homotopies. Take two points L and R on D^2, the target and source are the morphisms restricted on the lower and upper arc from L to R respectively (see picture (17)).

G is Morita equivalent to the action groupoid $S^3 \times S^3 \times (\mathbb{R}\times\mathbb{R})/\mathbb{Z} \Rightarrow S^3 \times S^3$ where $(\mathbb{R}\times\mathbb{R})/\mathbb{Z}$ is a quotient group by the diagonal $\mathbb{Z}$ action $(r_1,r_2)\cdot n = (r_1+n, r_2+n)$, and the action of this quotient group is given by the projection

$$(\mathbb{R}\times\mathbb{R})/\mathbb{Z} \to \mathbb{R}/\mathbb{Z} \times \mathbb{R}/\mathbb{Z} = S^1 \times S^1$$

and the product of the S^1 action on S^3. Our convention of target and source maps are $\mathbf{t}(p,q,[r_1,r_2]) = (p,q)$ and $\mathbf{s}(p,q,[r_1,r_2]) = (p\cdot[-r_1], q\cdot[-r_2])$. The symplectic structure is $(\pi^*\omega, -\pi^*\omega)$ on $S^3 \times S^3$ with $\pi : S^3 \to S^2$.

To show this, we give the associated complex line bundle $L \to S^2$ of the S^1-principal bundle $S^3 \to S^2$ a Hermitian metric and a compatible connection. We denote $\xi//\gamma$ as the result of the parallel transportation of a vector $\xi \in L_{\gamma(0)}$ along a path γ in S^2 to $L_{\gamma(1)}$. Since parallel transportation is isometric: $L_{\gamma(0)} \to L_{\gamma(1)}$, it preserves the S^1 bundle $S^3 \subset L$ and the angle $ang(\xi_1, \xi_2)$ between ξ_1 and ξ_2. Here the angle $ang(\cdot,\cdot)$ $L \oplus L \to S^1$ is point-wise the usual angular map (or argument map) $\mathbb{C} \to S^1$. It satisfies

$$ang(\xi_1,\xi_2) + ang(\xi_2,\xi_3) = ang(\xi_1,\xi_3) \quad \text{and} \quad ang(\xi_1,\xi_2) = -ang(\xi_2,\xi_1).$$

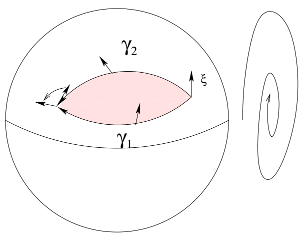
Therefore for two paths γ_1 and γ_2 sharing the same end points, we can define the angle $ang(\gamma_1,\gamma_2)$ between them to be

(16) $$ang(\gamma_1,\gamma_2) := ang(\xi//\gamma_1, \xi//\gamma_2), \quad \text{for } \xi \in T_{\gamma_1(0)}S^2.$$

Since parallel transportation preserves the angle, this definition dose not depend on the choice of ξ as the following calculation shows,

$$\begin{aligned} ang(\xi_1//\gamma_1,\xi_1//\gamma_2) = {} & ang(\xi_1//\gamma_1,\xi_2//\gamma_1) + \\ & ang(\xi_2//\gamma_1,\xi_2//\gamma_2) + ang(\xi_2//\gamma_2,\xi_1//\gamma_2) \\ = {} & ang(\xi_1,\xi_2) + ang(\xi_2//\gamma_1,\xi_2//\gamma_2) + ang(\xi_2,\xi_1) \\ = {} & ang(\xi_2//\gamma_1,\xi_2//\gamma_2). \end{aligned}$$

In fact $ang(\gamma_1,\gamma_2) = \int_D \omega_{area}$ where $\partial D = -\gamma_1 + \gamma_2$ and ω_{area} is the standard symplectic (area) form on S^2.



(17)

As shown in the above picture, because of $\pi_2(S^2)$, G_1 is not simply the paring groupoid $G_0 \times_{S^2\times S^2} G_0$—the set of paths with matched ends. To justify this, we use $ang : G_0 \times_{S^2\times S^2} G_0 \to S^1$ as in (16), then we have $G_1 = (G_0 \times_{S^2\times S^2} G_0) \times_{ang,S^1,pr} \mathbb{R}$ where $pr : \mathbb{R} \to S^1$ is the projection. For example (17) corresponds to (γ_1,γ_2,r) with $\lfloor r \rfloor = 2$. The pull-back groupoid of $G_1 \Rightarrow G_0$ by the projection (a surjective submersion) $S^3 \times_{\pi,S^2,\mathbf{t}} G_0 \times_{\mathbf{s},S^2,\pi} S^3 \to G_0$ with $\mathbf{s}(\gamma) = \gamma(0)$, $\mathbf{t}(\gamma) = \gamma(1)$, and $\pi : S^3 \to S^2$, is

$$(S^3 \times_{S^2} G_0 \times_{S^2} S^3 \times S^3 \times_{S^2} G_0 \times_{S^2} S^3) \times_{ang,S^1,pr} \mathbb{R} \Rightarrow S^3 \times_{S^2} G_0 \times_{S^2} S^3.$$

Now the pull-back groupoid of $S^3 \times S^3 \times (\mathbb{R} \times \mathbb{R})/\mathbb{Z} \Rightarrow S^3 \times S^3$ by the projection $S^3 \times_{S^2} G_0 \times_{S^2} S^3 \to S^3 \times S^3$ defined by $(\xi, \gamma, \xi') \mapsto (\xi'//\gamma, \xi//\gamma^{-1})$ is

(18)
$$(S^3 \times_{S^2} G_0 \times_{S^2} S^3 \times S^3 \times_{S^2} G_0 \times_{S^2} S^3) \times_{(S^3\times S^3)\times(S^3\times S^3)} S^3 \times S^3 \times (\mathbb{R}\times\mathbb{R})/\mathbb{Z} \Rightarrow S^3 \times_{S^2} G_0 \times_{S^2} S^3.$$

They are the isomorphic as Lie groupoids with the morphism from the second to the first by

$$(\xi, \gamma, \xi') \mapsto (\xi, \gamma', \xi')$$

on the base of the groupoid, where γ' is the path parallelly transported along the direction of $\xi \otimes \xi // \gamma(t)$ at[9] $\gamma(t)$ such that $ang(\gamma, \gamma') = ang(\xi, \xi'//\gamma)$, with the notice that $S^3 \otimes S^3 \subset TS^2$ is the sphere bundle. On the level of morphisms, we define

$$(\xi_1, \gamma_1, \xi_1', \xi_2, \gamma_2, \xi_2', [(r_1, r_2)]) \mapsto (\xi_1, \gamma_1', \xi_1', \xi_2, \gamma_2', \xi_2', r_2 - r_1). \tag{19}$$

We need to verify that this map does land on the correct manifold. By (18) we have $(\xi_1//\gamma_1^{-1}) \cdot [-r_2] = \xi_2//\gamma_2^{-1}$, thus

$$\begin{aligned} -[r_2] &= ang(\xi_1//\gamma_1^{-1}, \xi_2//\gamma_2^{-1}) \\ &= ang(\xi_1//\gamma_1^{-1}, \xi_2//\gamma_1^{-1}) + ang(\xi_2//\gamma_1^{-1}, \xi_2//\gamma_2^{-1}) \\ &= ang(\xi_1, \xi_2) + ang(\gamma_1^{-1}, \gamma_2^{-1}), \end{aligned}$$

hence $ang(\gamma_1, \gamma_2) = [r_2] + ang(\xi_1, \xi_2)$ and similarly $ang(\gamma_1, \gamma_2) = -[r_1] - ang(\xi_1', \xi_2')$. Therefore

$$[r_2 - r_1] = 2ang(\gamma_1, \gamma_2) - ang(\xi_1, \xi_2) + ang(\xi_1', \xi_2').$$

Since

$$\begin{aligned} &ang(\xi_1, \xi_1'//\gamma_1) + ang(\xi_1'//\gamma_1, \xi_2'//\gamma_1) + \\ &\qquad ang(\xi_2'//\gamma_1, \xi_2'//\gamma_2) + ang(\xi_2'//\gamma_2, \xi_2) = ang(\xi_1, \xi_2), \end{aligned}$$

we have

$$-ang(\xi_1, \xi_2) + ang(\xi_1', \xi_2') = -ang(\xi_1, \xi_1'//\gamma_1) + ang(\xi_2, \xi_2'//\gamma_2) - ang(\gamma_1, \gamma_2).$$

Hence

$$\begin{aligned} [r_2 - r_1] &= ang(\gamma_1, \gamma_2) + (-ang(\xi_1, \xi_1'//\gamma_1) + ang(\xi_2, \xi_2'//\gamma_2)) \\ &= ang(\gamma_1, \gamma_2) + ang(\gamma_1', \gamma_1) + ang(\gamma_2, \gamma_2') \\ &= ang(\gamma_1', \gamma_2') \end{aligned}$$

[9] that is $\dot\gamma'(t) = \dot\gamma(t)//\varphi_t(s)$ for some fixed $s \in \mathbb{R}$, where $\varphi_t(s)$ is the trajectory of $\xi \otimes \xi//\gamma(t)$. Since $\dot\gamma(0) = \dot\gamma(1) = 0$, γ' and γ have the same end points.

Therefore (19) is well-defined. It is not hard to see that it is indeed a groupoid isomorphism.

Therefore $G_1 \Rightarrow G_0$ and $S^3 \times S^3 \times (\mathbb{R}\times\mathbb{R})/\mathbb{Z} \Rightarrow S^3 \times S^3$ are Morita equivalent via this third groupoid (18). Therefore $\mathscr{G}(TS^2)$ is presented by $S^3 \times S^3 \times (\mathbb{R}\times\mathbb{R})/\mathbb{Z} \Rightarrow S^3 \times S^3$.

The next is to keep track with the symplectic form. The symplectic form on G_0 comes from the restriction of the symplectic form the whole path space PTS^2 [**5**],

$$\Omega((\delta\gamma_1, \delta a_1), (\delta\gamma_2, \delta a_2)) = \int_0^1 \omega(\delta\gamma_1(t), \delta a_2(t)) - \omega(\delta\gamma_2(t), \delta a_1(t)) dt,$$

where $(\delta\gamma_i(t), \delta a_i(t)) \in T(TS^2) \cong TS^2 \oplus TS^2$ after chosen a connection and $T(PTS^2) = PT(TS^2)$. Then the symplectic form on $S^3 \times S^3$ is induced by first pulling $\Omega|_{G_0}$ back to the Morita bibundle $S^3 \times_{S^2} G_0 \times_{S^2} S^3$ then pushing it to $S^3 \times S^3$ since it is G_1 invariant. In fact the symplectic form ω_1 on $S^3 \times S^3$ functions as

$$\omega_1((X_1, Y_1), (X_2, Y_2)) := \Omega((X_1, \delta\gamma_1, Y_1), (X_2, \delta\gamma_2, Y_2)),$$

where $\delta\gamma_i(0) = \pi_* X_1$ and $\delta\gamma_i(1) = \pi_* Y_1$. Comparing to the direct quotient $G_0/G_1 = S^2 \times S^2$ where $\Omega|_{G_0}$ descends to $(\omega, -\omega)$ on $S^2 \times S^2$, we can see that $\omega_1 = (\pi^*\omega, -\pi^*\omega)$. We also see that ω_1 is pre-symplectic and $\ker\omega_1$ is exactly the characteristic foliation of the groupoid $S^3 \times S^3 \times (\mathbb{R}\times\mathbb{R})/\mathbb{Z} \Rightarrow S^3 \times S^3$, that is the product of the S^1 orbits in S^3. Hence it gives a symplectic structure on $\mathscr{G}(TS^2)$.

As we seen before in Example 4.1, $S^3 \times \mathbb{R} \Rightarrow S^3$ presents the stack $\tilde{S}^2$. What we prove above is that the stacky groupoid $\mathscr{G}(TS^2)$ is not the usual one: $S^2 \times S^2$, but $\tilde{S}^2 \times \tilde{S}^2/B\mathbb{Z}$, to ensure that it has 2-connected source-fibre, which is a general result stated in Theorem 1.1. This resembles the construction of groupoid integrating TM when M is a non-simply connected manifold. The source simply connected groupoid of TM is $\tilde{M} \times \tilde{M}/\pi_1(M)$. Here for $\mathscr{G}(TS^2)$, the construction is comparable to this, but on a higher level—the aim is to kill π_2 of the source fibre. Further, using long exact sequence of homotopy groups, we have $\pi_2(\tilde{S}^2 \times \tilde{S}^2/B\mathbb{Z}) = \pi_1(B\mathbb{Z}) = \mathbb{Z}$ since $\pi(\tilde{S}^2 \times \tilde{S}^2) = 0$.

The symplectic structure on $\mathscr{G}(TS^2) = \tilde{S}^2 \times \tilde{S}^2/B\mathbb{Z}$ simply comes from the $(\tilde{\pi}^*\omega, -\tilde{\pi}^*\omega)$ as in Example 4.2 (notice that $B\mathbb{Z}$ is étale). Thus it is easy to see that it is multiplicative as in (15) and it makes $\mathscr{G}(TS^2)$ into a symplectic W-groupoid.

References

[1] R. ALMEIDA & P. MOLINO – "Suites d'Atiyah et feuilletages transversalement complets. (Atiyah sequences and transversally complete foliations).", (1985) (French).

[2] ———, "Suites d'Atiyah, feuilletages et quantification géométrique. (Atiyah sequences, foliations and geometric quantification).", (1985) (French).

[3] K. BEHREND & P. XU – "Differentiable stacks and gerbes", arxiv:math.DG/0605694.

[4] H. BURSZTYN & A. WEINSTEIN – "Picard groups in Poisson geometry.", *Mosc. Math. J.* **4** (2004), no. 1, p. 39–66 (English).

[5] A. S. CATTANEO & G. FELDER – "Poisson sigma models and symplectic groupoids.", Landsman, N. P. (ed.) et al., Quantization of singular symplectic quotients. Based on a research-in-pairs workshop, Oberwolfach, Germany, August 2–6, 1999. Basel: Birkhäuser. Prog. Math. 198, 61-93 (2001)., 2001.

[6] M. CRAINIC & R. L. FERNANDES – "Integrability of Lie brackets.", (2003) (English).

[7] G. FELDER, A. HENRIQUES, C. A. ROSSI & C. ZHU – "A gerbe for the elliptic gamma function.", *Duke Math. J.* **141** (2008), no. 1, p. 1–74 (English).

[8] E. GETZLER – "Lie theory for nilpotent L-infinity algebras", arxiv:math.AT/0404003.

[9] M. GUALTIERI – "Generalized complex geometr", arxiv:math.DG/0401221.

[10] A. HENRIQUES – "Integrating L-infinity algebras", arXiv:math.AT/0603563.

[11] S. LANG – *Differential and Riemannian manifolds. 3rd ed.; Repr. of the orig. 1972.*, Graduate Texts in Mathematics. 160. Heidelberg: Springer-Verlag. xiii, 364 p. DM 84.00; öS 655.20; sFr 81.00 , 1995 (English).

[12] Z. LIU, A. WEINSTEIN & P. XU – "Manin triples for Lie bialgebroids.", *J. Differ. Geom.* **45** (1997), no. 3, p. 547–574 (English).

[13] K. C. MACKENZIE & P. XU – "Integration of Lie bialgebroids.", *Topology* **39** (2000), no. 3, p. 445–467 (English).

[14] J. MAY – *Simplicial objects in algebraic topology.*, Chicago Lectures in Mathematics. Chicago: The University of Chicago Press. viii, 161 p. , 1992 (English).

[15] I. MOERDIJK – "Toposes and groupoids.", Categorical algebra and its applications, Proc. 1st Conf., Louvain-la- Neuve/Belg. 1987, Lect. Notes Math. 1348, 280-298 (1988)., 1988.

[16] I. MOERDIJK & J. MRČUN – "On integrability of infinitesimal actions.", *Am. J. Math.* **124** (2002), no. 3, p. 567–593 (English).

[17] J. MRČUN – "Stablility and invariants of Hilsum-Skandalis maps", Ph.D. Thesis, Utrecht University, 1996.

[18] J. PHILLIPS – "The holonomic imperative and the homotopy groupoid of a foliated manifold.", *Rocky Mt. J. Math.* **17** (1987), p. 151–165 (English).

[19] D. ROYTENBERG – "On the structure of graded symplectic supermanifolds and Courant algebroids.", Voronov, Theodore (ed.), Quantization, Poisson brackets and beyond. London Mathematical Society regional meeting and workshop on quantization, deformations, and new homological and categorical methods in mathematical physics, Manchester, UK, July 6–13, 2001. Providence, RI: American Mathematical Society (AMS). Contemp. Math. 315, 169-185 (2002)., 2002.

[20] P. ŠEVERA – "Some title containing the words "homotopy" and "symplectic", e.g. this one.", Molitor-Braun, Carine (ed.) et al., Proceedings of the 4th conference on Poisson geometry, Luxembourg, June 7–11, 2004. Luxembourg: Université du Luxembourg. Travaux Mathématiques 16, 121-137 (2005)., 2005.

[21] A. CANNAS DA SILVA & A. WEINSTEIN – *Geometric models for noncommutative algebras.*, Berkeley Mathematics Lecture Notes 10. Providence, RI: American Mathematical Society (AMS); Berkekey, CA: Berkeley Center for Pure and Applied Mathematics. xiv, 184 p. $ 20.00 , 1999 (English).

[22] D. SULLIVAN – "Infinitesimal computations in topology.", (1977) (English).

[23] H.-H. TSENG & C. ZHU – "Integrating Poisson manifolds via stacks.", Molitor-Braun, Carine (ed.) et al., Proceedings of the 4th conference on Poisson geometry, Luxembourg, June 7–11, 2004. Luxembourg: Université du Luxembourg. Travaux Mathématiques 16, 285-297 (2005)., 2005.

[24] ______, "Integrating Lie algebroids via stacks.", *Compos. Math.* **142** (2006), no. 1, p. 251–270 (English).

[25] ______, "Integrating Lie algebroids via stacks.", *Compos. Math.* **142** (2006), no. 1, p. 251–270 (English), arXiv:math.DG/0405003.

[26] A. WEINSTEIN – "Integrating the nonintegrable", unpublished note, 2003.

[27] A. WEINSTEIN & M. ZAMBON – "Variations on prequantization.", Molitor-Braun, Carine (ed.) et al., Proceedings of the 4th conference on Poisson geometry, Luxembourg, June 7–11, 2004. Luxembourg: Université du Luxembourg. Travaux Mathématiques 16, 187-219 (2005)., 2005.

[28] C. ZHU – "Lie II theorem for lie algebroids via stacky Lie groupoids", arxiv:math.DG/0701024.

[29] ______, "Lie n-groupoids and stacky lie groupoids", arxiv:math.DG/0609420.